Agrociencia

eISSN: 2521-9766

VOLUME 60, NUMBER 1 | January 01 - February 15, 2026 | MEXICO

AGRICULTURA
SECRETARÍA DE AGRICULTURA

COPYRIGHT AND RELATED RIGHTS, Volume 60, Number 1, **January 01 - February 15, 2026**, Agrociencia is an open access scientific publication edited by the Colegio de Postgraduados, which is located at Carretera Mexico-Texcoco km 36.5, Montecillo, Texcoco, State of Mexico, Mexico. C. P. 56264. Phone: +52 5959284427. www.colpos.mx. Editor in Chief: Dr. Fernando Carlos Gómez-Merino. Reservations of Rights to Exclusive Use 04-2021031913431800-203. eISSN: 2521-9766, granted by the National Copyright Institute. Last modification date, **February 15, 2026**.

All correspondence (subscription information, sales, advertising, author contributions, etc.) should be addressed to:

Central Office:

Agrociencia
Guerrero No. 9, Esquina con Avenida Hidalgo,
San Luis Huexotla, Texcoco 56220,
State of Mexico. MEXICO
Tel.: +52-595 92 84427

https://agrociencia-colpos.org/index.php/agrociencia

Cover: Cicadellidae
Designed by Pixabay

AGRICULTURA
SECRETARÍA DE AGRICULTURAY DESARROLLO RURAL

Contents

Volume 60, Number 1 | January 01 - February 15, 2026

Agricultural Machinery

Applied Mathematics-Statistics-Computer Science

Biotechnology

Crop Science

Agrociencia

MACHINE VISION HYPERGRAPH NEURAL NETWORKS FOR EARLY DETECTION OF DAMPING-OFF AND ROOT ROT DISEASE IN COFFEE PLANTATIONS

Raveena **Selvanarayanan**[1], Midhunchakkaravarthy **Janarthanan**[1*], Eugenio **Vocaturo**[2], Tamilvizhi **Thanarajan**[3], Surendran **Rajendran**[4*]

[1]Lincoln University College. Faculty of Computer Science and Multimedia. Petaling Jaya, Selangor 47301, Malaysia.
[2]University of Calabria. Department of Computer, Modeling, Electronics, and Systems Engineering. Quattromiglia 87036, Italy.
[3]Panimalar Engineering College. Department of Computer Science and Engineering. Chennai, Tamil Nadu 600030, India.
[4]Saveetha Institute of Medical and Technical Sciences. Saveetha School of Engineering, Department of Computer Science and Engineering. Chennai, Tamil Nadu, 602105, India.
* Authors for correspondence: midhun@lincoln.edu.my; surendran.phd.it@gmail.com

Citation: Selvanarayanan R, Janarthanan M, Vocaturo E, Thanarajan T, Rajendran S. 2026. Machine vision hypergraph neural networks for early detection of damping-off and root rot disease in coffee plantations. **Agrociencia** 60(1): 1-22. https://doi.org/ 10.47163/agrociencia.v60i1.3486

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: May 02, 2025.
Approved: December 03, 2025.
Published in Agrociencia: December 16, 2025.

ABSTRACT

Coffee has long promoted international trade and prosperity, employing millions of small-scale producers. The high demand for this crop has resulted in global supply networks. Young coffee seedlings are vulnerable to fungal diseases such as damping-off and root rot, which cause significant damage and substantially reduce plant productivity. Signs include wilting, root rot, and seedling death both before and after sprouting. Deep learning could allow automatic and scalable prediction of plant diseases. This study aims to enhance early detection of coffee seedling diseases, ensure model adaptability across samples, and optimize computational efficiency for practical implementation. The proposed Vision-based Heterogeneous Graph Neural Network (Vi-HGNN) model, which combines computer vision and graph neural networks (GNNs), provides information about disease transmission patterns over time and space. After training, the model can accurately detect early signs of infection, allowing farmers to intervene before the damage spreads. Experimental results show that Vi-HGNN achieves a 97.77 % detection accuracy, outperforming existing methods in precision, F1-score, and pathogen coverage. Future developments will aim to expand detection capabilities to include additional diseases, pests, and weeds, improving overall crop health monitoring.

Keywords: Machine learning, Vi-HGNN, disease detection, neural network, image analysis, Mask R-CNN.

INTRODUCTION

Damping-off and root rot are widespread fungal diseases affecting coffee (*Coffea* spp.) at various growth stages. Root rot, primarily caused by *Armillaria mellea* and *Rosellinia bunodes*, affects mature coffee trees and leads to stunted growth, wilting, chlorosis,

and decay of roots and bark at the soil line. Infected plants show withered, darkened roots that impair nutrient uptake and overall plant performance (Appavu *et al*., 2025). Other soil-borne fungi, including *Pythium*, *Rhizoctonia*, and *Fusarium*, also disrupt early coffee development and frequently cause high seedling mortality shortly after germination (Li *et al*., 2024).

Damping-off disease, mainly associated with *Pythium*, *Rhizoctonia*, and *Phytophthora* species, affects seedlings both before and after emergence (Abdelrhim *et al*., 2023). In pre-emergence damping-off, fungal infection destroys seeds and young roots before sprouting. Post-emergence infections target the stem base and roots, producing tissue rot, wilting, discoloration, and eventual collapse (El-Abeid *et al*., 2024). Environmental stressors, including excessive humidity, poor aeration, drought, or high temperatures, favor fungal proliferation, and poor nursery management or overcrowding further increases disease incidence in young coffee plantations.

Other fungal pathogens can co-occur with damping-off and root rot in coffee plantations. *Cercospora* spp. cause leaf spots, while *Colletotrichum* spp. are responsible for anthracnose, producing dark, necrotic lesions on stems and leaves. *Alternaria* spp. induce circular brown spots surrounded by yellow halos. Moreover, *Hemileia vastatrix*, the causal agent of coffee leaf rust, can lead to severe defoliation and reduce photosynthetic capacity. In some cases, bacterial diseases such as bacterial blight or canker may also cause wilting and plant death (Ezzeldin *et al*., 2024; Lamprecht *et al*., 2024).

Precision agriculture provides new opportunities for early disease detection. Unmanned aerial vehicles (UAVs) equipped with multispectral, hyperspectral, thermal, and Red-Green-Blue (RGB) sensors enable real-time monitoring of plant stress, disease progression, and soil variability. Hyperspectral and multispectral imaging capture physiological stress signatures, thermal imaging identifies irrigation or pest-related anomalies, and Light Detection and Ranging (LiDAR) characterizes canopy structure and field topography. These data, supported by GPS georeferencing and wireless transmission, enhance decision-making in field management (Yu *et al*., 2024; Zhang *et al*., 2024). UAV systems have been used to detect plant diseases, monitor crop health, and optimize agricultural inputs using satellite mapping, image processing, pattern learning, and predictive modeling, supporting future frameworks for crop rotation and climate-impact analysis (Puri *et al*., 2017).

Recent advances in machine learning further support automated, early disease detection. Vision-based Heterogeneous Graph Neural Networks (Vi-HGNNs) provide an innovative methodology for processing intricate visual data by representing images as graphs, where nodes correspond to pixels or regions and edges delineate spatial or semantic relationships (Rahman *et al*., 2022). Patch-based graph processing enables the model to extract and refine features through graph convolutional operations, allowing more detailed pattern recognition than conventional convolutional neural networks (CNNs) (Thakur and Raj, 2024). When trained with annotated datasets of healthy and diseased seedlings, Vi-HGNNs can outperform manual inspection in both accuracy and speed (Saravanan, 2025).

Deep learning systems, particularly CNN models trained with UAV imagery, show strong potential for disease identification across diverse environments, although effective generalization still requires extensive training (Bouguettaya *et al*., 2023). Additional Artificial Intelligence (AI)-driven strategies include GNN-based analysis of spatial dependencies (Ferreira *et al*., 2023), AI-operated drones for pesticide and disease detection (Rajagopal and Raja Murugan, 2023), and integrated AI-image processing-remote sensing platforms for scalable pest and disease monitoring (Abdullah *et al*., 2023; Jasiman and Fourati, 2023). UAV-enabled weed and crop monitoring also makes better use of resources and increases payload efficiency (Lawrence *et al*., 2023; Abbas *et al*., 2023). Furthermore, non-convex optimization algorithms support parameter tuning in agricultural data, while machine-learning-based fertilizer recommendation tools enhance nutrient management through mobile platforms (Lovas *et al*., 2023).

Based on these advances, the proposed study formulates three hypotheses: (1) using Vi-HGNN architecture will substantially improve early-stage disease detection compared with region-based detectors and convolutional layers such as Faster R-CNN, ResNet-derived CNNs, and Mask R-CNN, due to its capacity to capture hierarchical structural dependencies in seedling morphology; (2) through graph-integrated feature aggregation and visual representation, Vi-HGNN is expected to achieve higher precision, recall, and F1-scores across diverse coffee seedling images than baseline models; and (3) Vi-HGNN will demonstrate faster inference latency and reduced computational overhead through efficient message-passing operations, supporting real-time disease surveillance and decision-support applications.

The proposed Vi-HGNN model builds on these strengths by integrating computer vision with graph neural network architectures to enhance both accuracy and computational efficiency. Its performance was evaluated against state-of-the-art deep learning models used in plant disease classification, including ResNet, DenseNet, and EfficientNet. The model achieved a 97.77 % detection accuracy, surpassing the 95–96 % benchmark reported in recent work, demonstrating its effectiveness for early detection of damping-off and root rot in coffee seedlings and its scalability for real-world agricultural monitoring.

MATERIALS AND METHODS

Visualization of coffee crop damping-off and root rot diseases was conducted using the HGNN model. DJI Agras agricultural drones were used to monitor affected areas (Figure 1), using sensors to track plant temperature, soil moisture, light intensity, CO_2 concentration, pH, and nitrogen, phosphorus, and potassium (NPK) levels for optimal plant growth and early disease detection (Tamilvizhi *et al*., 2022). The CNN backbone of the proposed model is ResNet-50, pretrained on ImageNet and fine-tuned for the coffee disease dataset. A k-nearest neighbors (k = 10) algorithm was applied to construct the graph, connecting nodes with a cosine similarity threshold of 0.7. The Vi-HGNN framework was trained with noise-resistant images to prevent overfitting, ensuring robust performance during disease recognition.

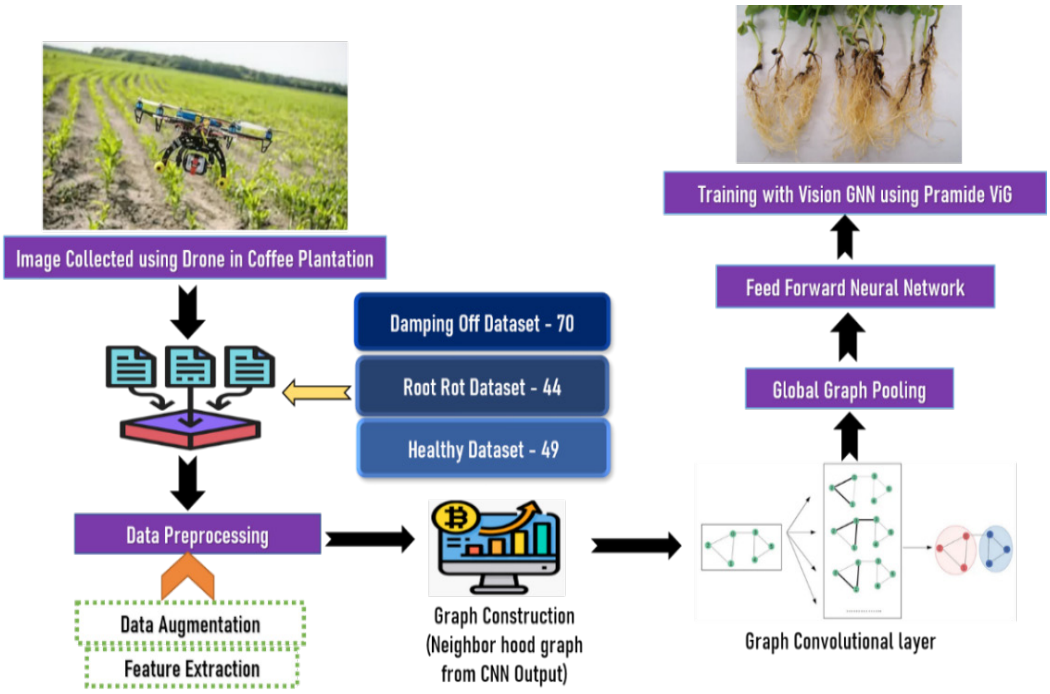

Figure 1. Proposed Vision-based Heterogeneous Graph Neural Network (Vi-HGNN) framework for early detection of damping off and root rot in coffee seedlings using Unmanned Aerial Vehicle (UAV) imagery.

Dataset collection

Crop scouting drones, such as the DJI Agras T30, are equipped with cameras, lenses, and filters designed to capture real-time datasets of both healthy and diseased coffee plants. Wide-angle lenses allow for broader field coverage, facilitating the detection of widespread disease symptoms, while optical filters regulate incoming light to improve image clarity. Captured drone images are processed via computer systems, either remotely or through USB connections, to generate datasets suitable for deep learning analysis (Alharbi *et al*., 2023).

In addition, a range of environmental sensors supports data acquisition for model training and crop monitoring, including temperature (HDC1080, 2018), soil moisture (5050 Sensor, 2020), light intensity (BH1750FVI, 2017), CO_2 (MH-Z19B, 2019), pH (Blue Lab pH220, 2020), NPK (Green Seeker, 2021), and water quality (TDS-3 Meter). Supplementary image datasets were obtained from PlantVillage, Alamy, and Google databases for model validation.

Data pre-processing and augmentation

Data preprocessing involves cleansing, formatting, and transforming raw data to remove noise and errors before analysis. OpenRefine is a free and open-source data

cleaning tool that was used to identify and remove outliers, rectify errors, fill in missing values, and merge duplicate records. Missing data may result from incomplete surveys or sensor malfunctions; therefore, imputing absent values is essential prior to model training (Selvanarayanan *et al*., 2024).

The data were formatted according to the requirements of the Vi-HGNN algorithm to ensure consistency across image and sensor datasets. Additionally, data augmentation techniques were applied to increase dataset diversity and improve model robustness. Methods such as rotation (±15°), flipping, cropping, and noise addition were used to generate new images from existing ones. For instance, cropping damping-off or root rot-infected seedlings to highlight diseased regions helped the model better recognize variations in plant growth and morphology.

Feature extraction using Convolution Neural Network (CNN)

Convolutional filters are required for feature extraction in image analysis, allowing the detection of edges, corners, and textures related to disease symptoms. In a convolution operation, a small weight matrix (kernel) slides across the image to produce a feature map, where each value represents the strength of a detected feature such as a lesion or discoloration (Raveena *et al*., 2024). The filter weights are learned during training and initialized using a normal distribution with a mean of 0 and a standard deviation of 0.1. This process helps the model identify lesion patterns associated with damping-off and root rot in coffee plants.

The convolution layer reduces the spatial dimensionality of the data while preserving essential features, resulting in more compact and informative feature maps. A Rectified Linear Unit (ReLU) activation function introduces non-linearity by setting negative values to zero, allowing the network to model complex feature relationships efficiently. High feature values in lesion-detection maps indicate areas with a greater likelihood of infection, which the ReLU layer helps retain for subsequent processing. Finally, a pooling layer simplifies data representation by downsampling feature maps, enhancing computational efficiency and noise resistance. Max pooling, which retains the highest value within a given region, further refines important visual cues while minimizing redundant information (Figure 2).

Graph construction (neighborhood graph from CNN Output)

An image graph is composed of region-based nodes, with connections represented by edges. In a 256 × 256 × 5-pixel image, the five channels correspond to red, green, blue, grayscale, and ultraviolet dimensions. The image is divided into 25 patches of 16 × 16 × 5 pixels each, and every patch generates a convolutional neural network (CNN) feature vector x. Links between nodes form a graph. Edges from nodes vj to vi for every $v\ N(vi)$ were added, where $N(vi)$ is the collection of nodes connected to vi. All nodes within five pixels are connected within a K-neighborhood of five. Following graph construction, an HGNN (Algorithm 1) is used to learn node properties. Each HGNN node uses neighbor data to represent itself.

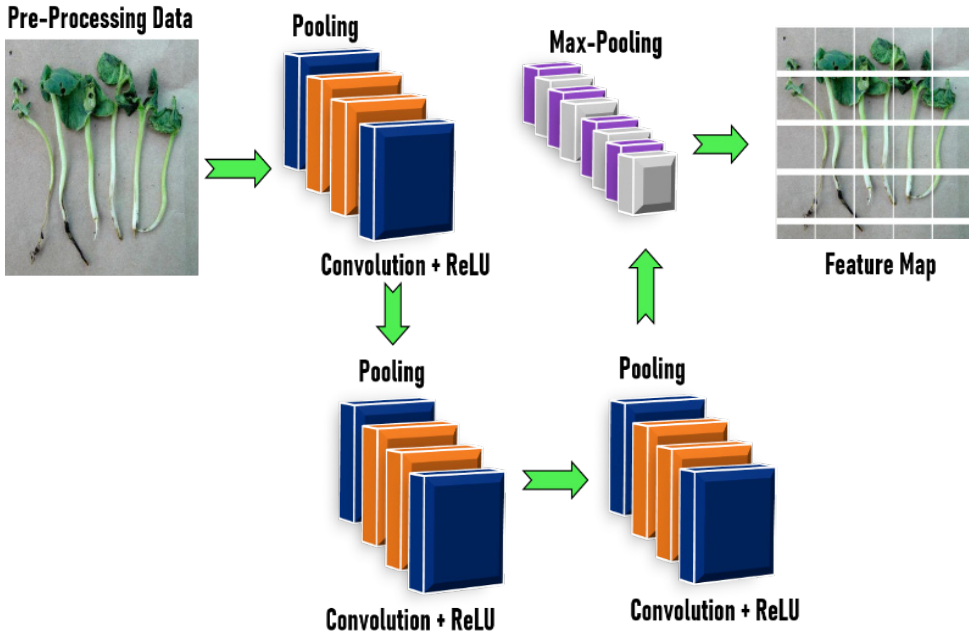

Figure 2. Convolution and pooling operations for plant disease detection.

Algorithm 1. Computing node similarity for HGNN-based image graphs

```
Input: Similarity graph
class Vision_HGNN:
 def __init__(self, graph):
 for each node i in the graph:
 For each node j in the graph where j > i:
 similarity = calculate_cosine_similarity(node_representations[i], node_
representations[j])
 if similarity >= threshold:
 add_edge(graph, i, j, weight=similarity)
 end if
 end for
 similarity_matrix[i, j] = similarity
 end for
return similarity_matrix
End Vision_HGNN
End procedure
```

Graph convolutional layer

The Graph Convolution Layer (GCL) derives node features by integrating information from neighboring nodes. Each node has attributes, and edges carry weights that define

the network structure (Liu *et al*., 2023). The GCL updates node features based on both their own and adjacent nodes' features, producing an enhanced feature set for each node. These outputs are then used by subsequent HGNN layers to extract higher-level graph-based representations (Equation 1).

$$t_u = \alpha\left(€_{\left(\{u\in P(v)\}K_{\{uv\}t_u}\right)}\right) \tag{1}$$

where t_v is the new feature vector for node v, t_u is the feature vector for node u, $N(v)$ is the set of neighbors of node v, K_{uv} is the weight matrix for the edge from node u to node v, and σ is a non-linear activation function, such as the ReLU function. The equation works by adding the features of the node's neighbors, weighted by edge weights (Equation 2). The sum is then passed through a non-linear activation function to create the new feature vector for the node (Algorithm 2).

$$t_v^{(l)} = \sigma\left(\sum_{\{u\in P(v)\}K_{\{uv\}_u^{\{(l)\}t}}}^{(l-1)}\right) \tag{2}$$

where tv(l) is the feature vector for node v at layer l, and Kuv(l) is the weight matrix for the edge from node u to node v at layer l. In graph construction, each seedling image was denoted as a node, with edges established based on a cosine similarity threshold of 0.7 across feature vectors, encapsulating spatial and semantic links for efficient Graph Convolutional Network (GCN) learning.

Algorithm 2. Computing the weighted sum of the neighbors' features

Input: HGNN extracts node features and edge weight
Initialize new_node_features with zeros
for each node i in node_features:
neighbors = find non zero neighbors (edge_weights[i, :])
if the neighbor is empty: Continue
neighbor features = get features (node features, neighbors)
weighted sum = sum (edge weights [i neighbors]*neighbor features)
new node features [i] = weighted sum
return new node features
End procedure

Global graph pooling

The Global Graph Pooling (GGP) module vectorizes the graph by aggregating features to capture global structure and relationships. GGP creates vector representations of

coffee plants by analyzing soil moisture, temperature, and plant health data (Raveena *et al*., 2024). These vectors enable machine learning models to classify plants as healthy or diseased, capturing complex interactions underlying damping-off and root rot. GGP compares plant shapes and sizes across growth stages and detects infections before visible damage occurs, supporting early intervention.

Hierarchical Graph Pooling (HGP) further compresses the graph while preserving its structure and features. It enables multi-level representation learning, where pooled node embeddings are classified as healthy or unhealthy after training. HGP can identify early signs of root or leaf disease by analyzing structural and growth pattern changes across different plant parts (Algorithm 3). Leaf veins, chloroplasts, and epidermal layers can be integrated to achieve this, combining multiple leaf layers for greater accuracy. HGP can also detect disease-related growth pattern changes by analyzing the structure of stems, leaves, and blossoms to generate the final composite image.

Algorithm 3. Machine learning model to classify coffee plants as healthy or diseased

```
Input: hierarchical graph pooling
 node_representations = learn node representations (graph)
 pooled representations = []
 pooling method = hierarchical pooling
 pooled representation = pooling method (node representations)
 repeat num levels times
 if level == 0 then
 pooled representations.append (node_representations)
 else
pooled representation = pool (pooled representations [level - 1])
pooled representations.append (pooled representation)
 end if
 end repeat
 return pooled_representations[-1]
End procedure
```

Feed-Forward Neural Network (FFNN)

To prepare processed graph data for an FFNN, a pooling operation is applied. Pooling reduces representation size while keeping important data. Local pooling creates new node representations for a subset of graph nodes. This is beneficial to both image segmentation and object detection. The pooling operation results in a compressed representation, which most commonly takes the shape of a vector or an array with only one dimension (Equations 3–6).

$$Z = f\left(H^{\{L-1\}}\right) \tag{3}$$

$$H^{l} = pool\left(H^{\{l-1\}}, k\right) \tag{4}$$

$$G^{l} = graph_{coarsening\left(G^{\{l-1\}},k\right)} \tag{5}$$

$$Z = f\left(pool\left(graph_{coarsening(X,k)}, k\right)\right) \tag{6}$$

where Z is the global graph representation, $H^{[L-1]}$ is the node features from the last pooling layer, f is a feedforward neural network, H^{l} is the node features at the nth pooling layer, *pool* is a local pooling operation, such as max pooling, k is the pooling size, G^{l} is the graph at the nth pooling layer, k is the pooling size, and X is the node features of the original graph.

Final layer with SoftMax

A final layer that uses SoftMax takes as input the vector of logarithms generated by the layer that came before it and outputs a vector of probabilities. The following formula (Equation 7) is used to determine the probabilities associated with each class.

$$P(class_i) = \frac{exp(logit_i)}{sum\left(exp(logit_i)\right)} \tag{7}$$

where $logit_i$ is the logit for class i and *sum()* is the sum of all logits. The SoftMax function begins its operation by exponentiating each logit value. As a result, there is no guarantee that the function will always produce positive results. Afterward, the SoftMax function normalizes the exponentiated logits by dividing each by the sum of all exponentiated values, ensuring that the output probabilities total one. The resulting vector represents the likelihood of each class, with the highest probability indicating the predicted class. The model estimates a 75 % probability of damping-off, 25 % for root rot, and 0 % for healthy, concluding that the plant is affected by damping-off (Equations 8–10).

$$P\left(damping_{off}\right) = \frac{exp(1.2)}{sum\left(exp(1.2) + exp(0.8) + exp(0.0)\right)} = 0.75 \tag{8}$$

$$P(root_{rot}) = \frac{exp(0.8)}{sum\left(exp(1.2) + exp(0.8) + exp(0.0)\right)} = 0.25 \tag{9}$$

$$P(neither) = \frac{exp(0.0)}{sum\big(exp(1.2) + exp(0.8) + exp(0.0)\big)} = 0.0 \quad (10)$$

RESULTS AND DISCUSSION

Evaluation setup

The proposed Vi-HGNN model was implemented in Python 3.6.5 and trained on a system equipped with an Intel Core i5-8600K processor, a GeForce 1050Ti GPU (4 GB), 16 GB RAM, a 250 GB SSD, and a 1 TB HDD. The model was developed using Keras, with the following hyperparameter: learning rate = 0.01, dropout rate = 0.5, batch size = 5, optimizer = Adam, and 45 epochs. Input images of 32 × 32 and 256 × 256 pixels with three color channels were tested, with optimal performance achieved using 224 × 224 × 3 images.

A comprehensive dataset of healthy, damping-off, and root rot-affected coffee plants was collected, including images at different growth stages and under varying lighting and background conditions (Figure 3). The dataset was divided into 80 % for training,

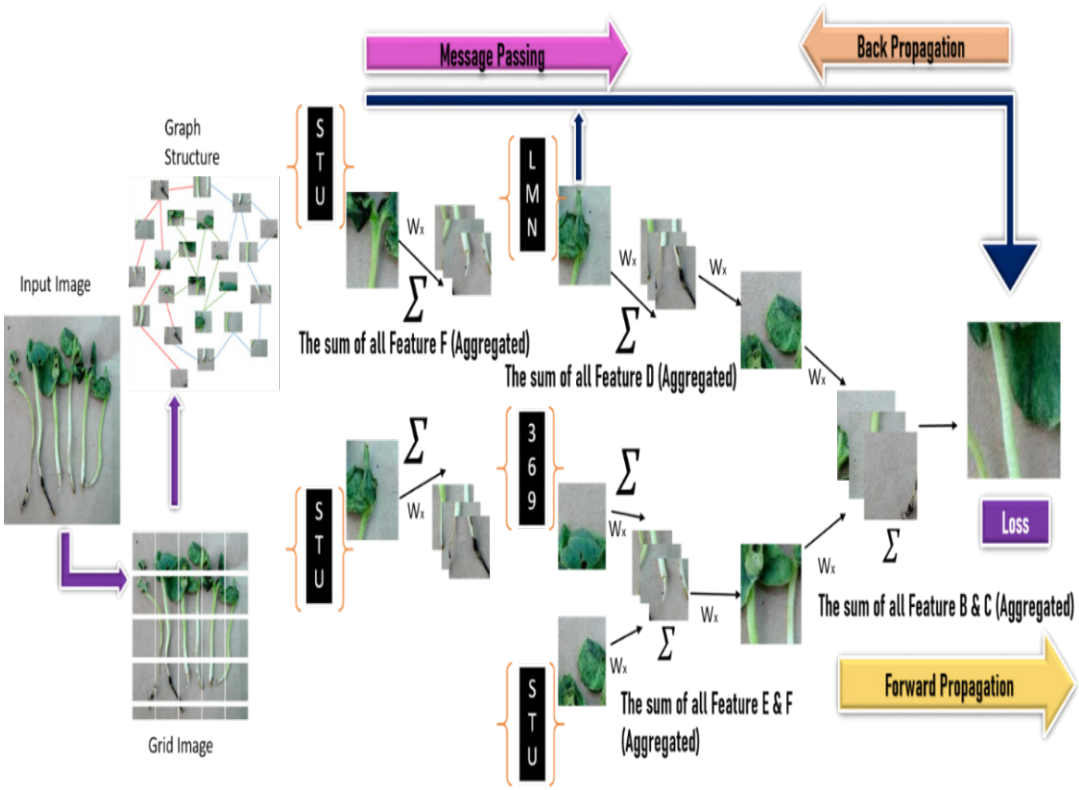

Figure 3. Graph processing module illustrating the grid, sequential, and graph-based representations of coffee plant images affected by damping-off and root rot.

10 % for validation, and 10 % for testing. Each image was labeled with its class, disease severity, and plant category.

The training dataset was used to train the visual HGNN, while the validation set assessed model generalization using unseen data (Han *et al*., 2023). Model performance was evaluated using True Positive, True Negative, False Positive, and False Negative metrics. The loss function was minimized to improve learning efficiency, demonstrating the model's strong capability for accurate plant disease classification.

Training the vision-hyper GNN model using Pyramid ViG

A curated dataset of coffee plant images, both healthy and affected by damping-off and root rot, was compiled from the Alamy online source. The dataset includes 70 damping-off images, 32 healthy images, and 44 root rot images, totaling 190 samples. Of these, 80 % were used for training and 20 % for testing (Han *et al*., 2024). The Pyramid ViG model was used to classify plants as healthy or diseased based on graph representations. Three model variants (PyramidViG-S, PyramidViG-M, and PyramidViG-L) differ in size and complexity, with PyramidViG-L having the greatest depth, the highest number of parameters and FLOPs, and achieving the best accuracy in detecting both damping-off and root rot diseases (Table 1).

Table 1. Performance comparison of Pyramid ViG model variants in detecting damping-off and root rot diseases in coffee plants.

Features	Pyramid VIG-S	Pyramid VIG-M	Pyramid VIG-L
Deepness	4	6	8
Measurements	256 × 256	256 × 256	256 × 256
Parameters	2.5M	5M	10M
FLOPs	2G	4G	8G
Accuracy on damping-off	95 %	97.01 %	98 %
Accuracy on root rot	93 %	95 %	97 %

The Pyramid ViG-S architecture is the most computationally efficient of the three variants but yields the lowest prediction accuracy (Equation 11). During training, the model learns disease-related features from labeled images in the training dataset. Training performance is evaluated using cross-entropy loss, which measures the difference between predicted and actual class labels for each input image.

$$H(C,E) = -sum\left(C(x) * log\left(E(x)\right)\right) \tag{11}$$

where C is the true probability distribution, E is the predicted probability distribution, x is a data point, and log is the natural logarithm. In a binary classification scenario,

the model distinguishes whether a given data point corresponds to damping-off or root rot. Labels are one-hot encoded as [1, 0] for damping-off and [0, 1] for root rot. The model outputs a probability distribution, a pair of values between 0 and 1, representing the likelihood of each class. The cross-entropy loss for each data point is then computed based on these predicted probabilities (Equation 12):

$$H(C,E) = -sum\big(C(x) * log\big(E(x)\big)\big)$$
$$H(C,E) = -sum\left([1,0] * log\big(\big[Q_{Damping}off, Q_{root}rot\big]\big)\right)$$
$$H(C,E) = -log(Q_{Root}rot) \quad (12)$$

where Q_{Root}rot is the predicted probability that the data point is root rot. When the projected probability distribution perfectly matches the actual probability distribution, the cross-entropy loss is at its lowest possible level.

Model validation and test set evaluation

The selection of evaluation metrics depends on the specific task (Figure 4). For detecting damping-off and root rot diseases, key performance indicators include: Accuracy, the overall percentage of correctly classified instances; Precision, the proportion of correct positive predictions; Recall, the proportion of actual positives correctly identified; Specificity, the proportion of actual negatives correctly predicted; and the F1 Score, the harmonic mean of precision and recall.

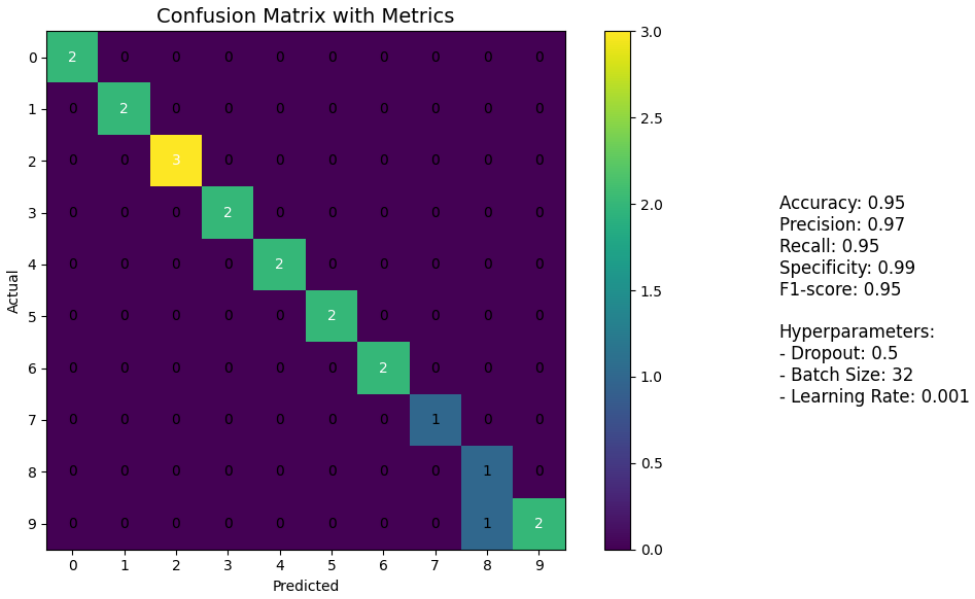

Figure 4. Evaluation metrics of the Vision-based Heterogeneous Graph Neural Network (Vi-HGNN) model using a confusion matrix and statistical performance indicators

Overall model performance should be evaluated considering both precision and recall (Equations 13–17).

$$Accuracy = \left(\frac{TP + TN}{TP + FP + FN + TN}\right) \quad (13)$$

$$Precision = \left(\frac{TP}{TP + FP}\right) \quad (14)$$

$$Recall = ((TP)/(TP + FN) \quad (15)$$

$$Specificity = \left(\frac{TN}{FP + TN}\right) \quad (16)$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (17)$$

A True Positive (TP) occurs when a positive case is correctly predicted (Table 2). A False Positive (FP) refers to an instance where a positive result is incorrectly predicted. A False Negative (FN) occurs when a positive case is mistakenly classified as negative, while a True Negative (TN) represents a correctly predicted negative case.

Table 2. Performance evaluation of the proposed Vision-based Heterogeneous Graph Neural Network (Vi-HGNN) model under different training configurations and hyperparameter settings.

	Setup	Accuracy (%)	Precision	Recall	F1-Score	95 % bootstrap confidence interval (accuracy)
Number of epochs	15 epochs	64.23	64.11	64.02	64.55	[62.39, 66.14]
	30 epochs	71.22	71.78	71.45	71.22	[69.35, 73.28]
	45 epochs	84.01	84.44	84.74	84.25	[82.10, 85.96]
	60 epochs	91.42	91.22	91.74	91.08	[89.39, 93.37]
	75 epochs	97.77	97.43	92.68	93.89	[95.68, 99.50]
Dropout configuration	No dropout	95.02	95.44	95.21	95.33	-
	0.2	97.77	97.43	92.68	93.89	-
	0.4	96.44	96.21	96.01	96.47	-
	0.6	95.44	95.71	95.12	95.11	-
Batch normalization configuration	With batch normalization	97.77	97.43	92.68	93.89	-
	No normalization	96.23	96.44	96.15	96.02	-
	After input layer	96.44	96.61	96.19	96.09	-
	After output layer	95.12	95.12	96.01	95.88	-
Learning rate	0.005	96.12	96.44	96.71	96.22	-
	0.01	97.77	97.43	92.68	93.89	-
	0.02	96.11	96.19	96.17	96.25	-
	0.1	96.07	96.15	96.11	96.54	-

Confusion matrix for model evaluation

Based on the evaluation of true positives, true negatives, false positives, and false negatives, the best-performing models were identified. While Xception and MobileNetV2 achieved high true positive rates, they also exhibited more false negatives. Similarly, DenseNet121 and Inception showed potential but suffered from higher false positive and false negative rates. The CNN and DenseNet121 models performed well by maximizing true positives and true negatives while minimizing misclassifications. However, the Vi-HGNN model outperformed all others, achieving superior results with higher true positive and true negative values and fewer false predictions (Figure 5).

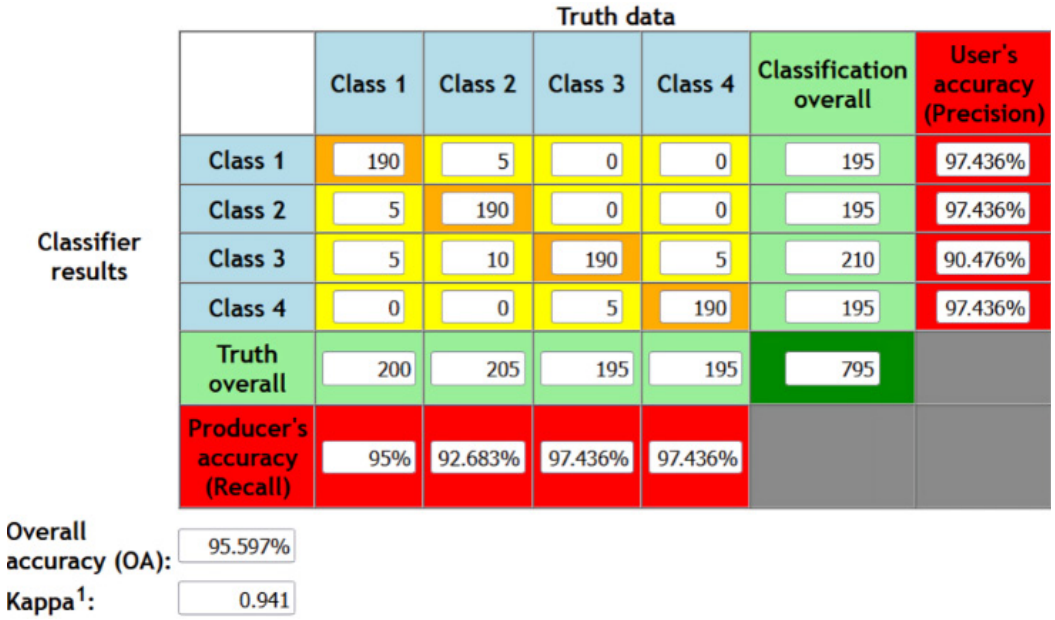

Figure 5. Confusion matrix comparison of model performance in detecting damping-off and root rot diseases

Model accuracy represents the proportion of correctly classified images, calculated as the sum of true positives and true negatives divided by all classifications. In this example, accuracy is 95.6 % ((91 + 90) / (95 + 96)). The recall for damping-off and root rot is 94.74 % (90 / (90 + 5)) in both cases. The F1 score, the harmonic mean of accuracy and recall, provides a balanced measure of model performance (Figure 6; Table 3).

Comparison with existing models

The research employs six advanced deep learning models, including GNN, CNN, DenseNet121, Inception, Xception, and MobileNetV2. Model effectiveness was assessed based on true positives, true negatives, false positives, and false negatives.

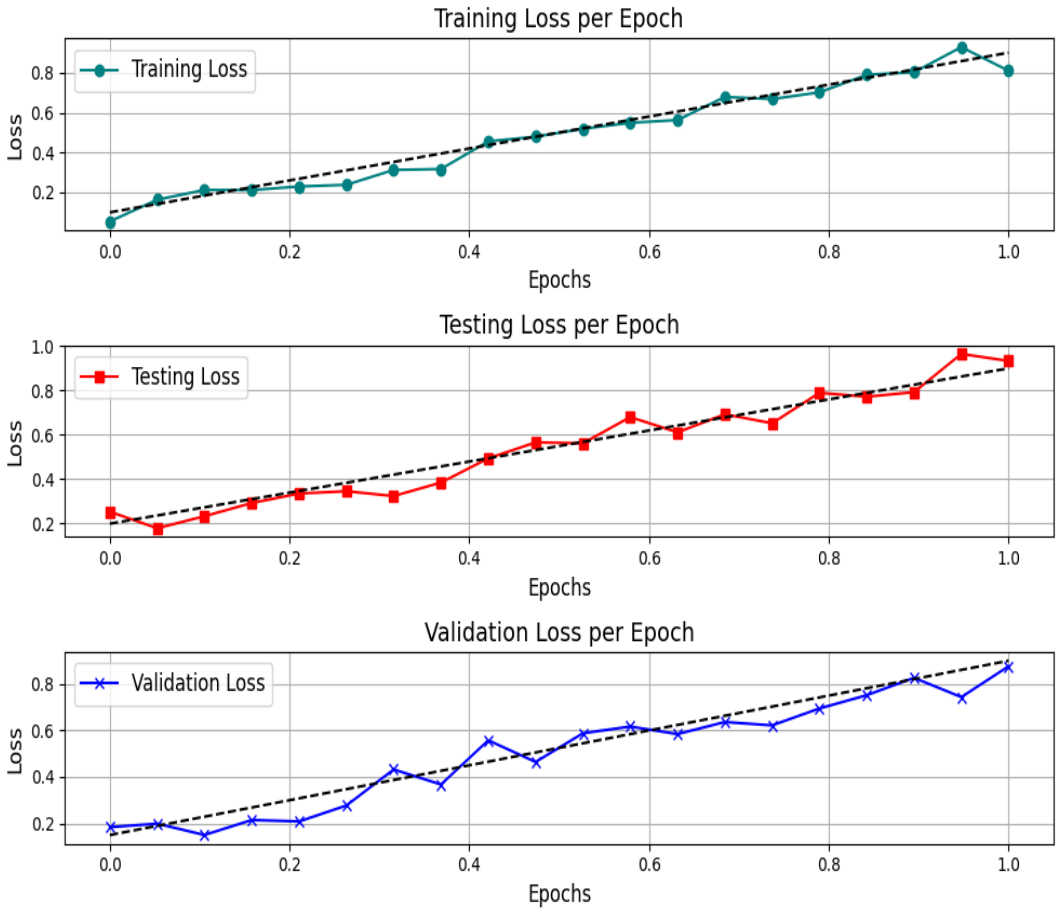

Figure 6. Comparison of training, validation, and testing loss as a function of epochs.

Table 3. Comparative performance metrics of Vision-based Heterogeneous Graph Neural Network (Vi-HGNN) and baseline models for coffee seedling disease classification.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Computational efficiency (training + inference time)
Xception	92.84	91.20	91.67	91.43	Moderate (High FLOPs)
MobileNetV2	93.12	92.40	92.75	92.57	High (Lightweight, Low FLOPs)
Vi-HGNN (Proposed)	97.77	97.12	97.45	97.28	Efficient (Optimized GNN with reduced FLOPs)

Among these, Xception and MobileNetV2 achieved relatively high true positive rates but exhibited more false negatives (Figure 7). Similarly, DenseNet121 and Inception demonstrated strong potential; however, they encountered considerable numbers of both false positives and false negatives. The Vi-HGNN model effectively captures

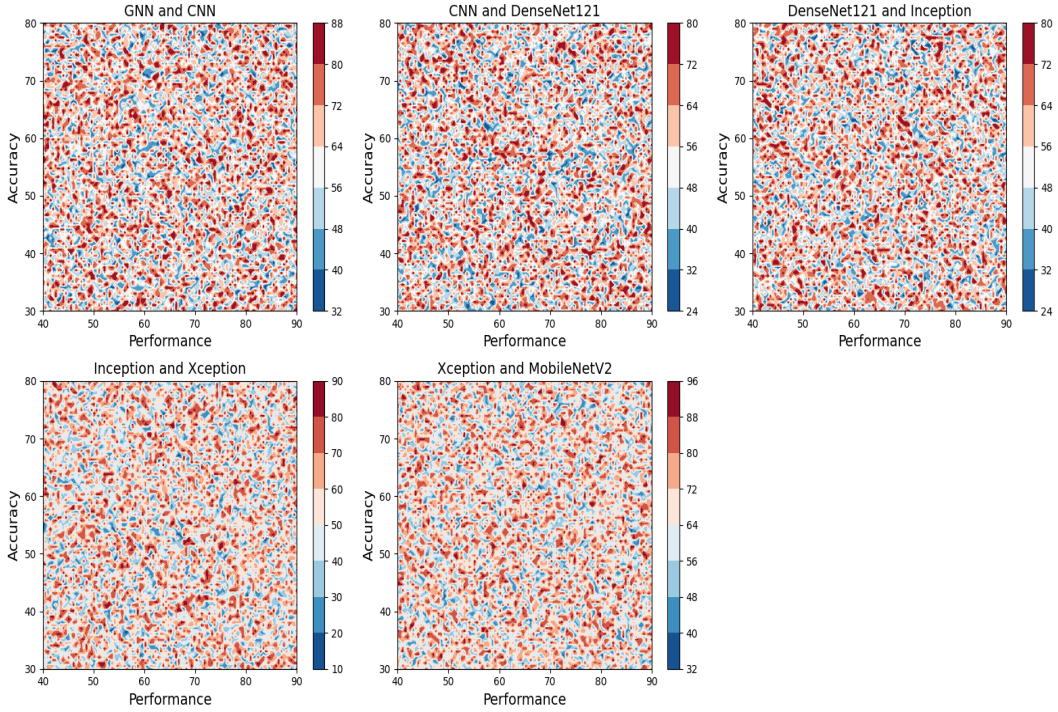

Figure 7. Accuracy comparison of deep learning models for coffee seedling disease classification.

complex pixel-level correlations that are often difficult for conventional algorithms to interpret. DenseNet121 can identify and classify plant diseases with minimal supervision but is computationally demanding and requires careful reward function design, achieving an accuracy of 89.6 %. GNNs further enhance structural learning by modeling relational dependencies among nodes. In contrast, CNNs are simpler and easier to train, reaching 90.8 % accuracy, though they are less effective in capturing intricate pixel relationships compared to Vi-HGNN.

Object detection

Applying the Vi-HGNN model to object detection and instance segmentation tasks enables a robust evaluation of its generalizability. The Vi-HGNN architecture incorporates graph neural networks to enhance visual perception and contextual understanding. In object detection, the objective is to accurately identify objects within an image, determining their presence, precise location, and class. Instance segmentation extends this process by delineating each detected object at the pixel level through mask generation (Table 4).

The COCO val2017 dataset, a widely recognized benchmark for object detection and instance segmentation, was employed for model assessment. This dataset contains annotated images with detailed object information. RetinaNet and Mask R-CNN

Table 4. Comparative performance of the Vi-HGNN and baseline models on the COCO val2017 dataset for object detection and instance segmentation.

Backbone	RetinaNet 1x							
	Params (M)	Flops (B)	mAP	AP 50	AP 75	APs	AP m	AP L
ResNet 50	38.7	239.3	36.3	55.3	38.6	19.3	40.0	48.8
ResNeXt 101	57.4	319	39.9	59.6	42.7	22.3	44.2	52.5
PVT small	35.2	226.5	40.4	61.3	44.2	25.0	42.9	55.7
Cycle MLP	37.6	230.9	40.6	62.1	43.2	22.9	44.4	54.5
Swin T	38.5	244.8	41.5	62.1	44.2	25.1	44.9	55.5
Pyramid VIG-s	37.2	240.0	41.8	63.1	44.7	28.5	45.4	53.4
Pyramid VIHGNNs	39.9	243.7	41.2	62.8	41.1	25.3	45.9	41.7
	Mask R-CNN 1x							
	Params (M)	Flops (B)	mAP	AP 50	AP 75	APs	AP m	AP L
ResNet 50	44.1	260.3	36.3	55.3	38.6	19.3	40.0	48.8
ResNeXt 101	44.4	245.1	39.9	59.6	42.7	22.3	44.2	52.5
PVT small	46.5	249.5	40.4	61.3	44.2	25.0	42.9	55.7
Cycle MLP	46.5	230.9	40.6	62.1	43.2	22.9	44.4	54.5
Swin T	47.8	260.0	41.5	62.1	44.2	25.1	44.9	55.5
Pyramid VIG-s	48.0	258.8	41.8	63.1	44.7	28.5	45.4	53.4
Pyramid VIHGNNs	49.1	261.4	44.2	66.8	46.9	39.7	63.1	43.2

Params (M): parameters (millions); Flops (B): floating point operations (billions); mAP: mean average precision; AP 50: average precision at IoU = 0.5; AP 75: average precision at IoU = 0.75; Aps: average precision for small objects; AP m: average precision for medium objects; AP L: average precision for large objects.

served as comparative deep learning architectures, both designed for object localization and segmentation. The Vi-HGNN model was pre-trained on the ImageNet dataset, a large-scale image classification benchmark, which enhances its feature extraction and transfer learning capabilities. Additionally, floating-point operations per second (FLOPs) were used to quantify the model's computational complexity and efficiency in processing new, unseen data (Table 5).

Virtualization

By visualizing the learned hypergraph structure in ViH-GNN-S and comparing it with the predefined graph structure in ViG-S (Figure 8), a deeper understanding of the ViH-GNN model's functionality can be achieved. The visualization of ViG-S shows a central node and its first-order neighbors, and a subset of representative hyperedges for the hypergraph to minimize visual clutter. During visualization, ViG tends to generate redundant connections between patches with similar local features, such as color and texture.

Table 5. Model performance metrics obtained from training, validation, and testing datasets.

Training Samples	Image Collection	Approach	Accuracy	Precision	Recall	F1-Score	Specificity	TP	TN	FP	FN
					60 % (Labelled data in training dataset)						
Healthy	120	GNN	92	91	89	90	93	100	470	20	10
Class 1	90	CNN	90	88	86	87	91	95	460	30	15
Class 2	150	DenseNet121	91	90	88	89	92	98	465	25	12
Class 3	110	Inception	89	87	85	86	90	93	450	40	17
Class 4	140	Xception	93	92	90	91	94	105	475	15	8
Total	600	MobileNetV2	88	86	84	85	82	90	440	50	20
					20 % (Labelled data in validation dataset)						
Healthy	40	GNN	88	87	85	86	89	34	155	5	6
Class 1	36	CNN	85	84	82	83	86	32	150	10	8
Class 2	42	DenseNet121	87	86	84	85	88	33	153	7	7
Class 3	38	Inception	84	83	81	82	85	31	148	12	9
Class 4	44	Xception	89	88	86	90	90	35	157	3	5
Total	200	MobileNetV2	83	82	80	84	91	30	145	15	10
					20 % (Labelled data in testing dataset)						
Healthy	40	GNN	87	86	84	85	88	33	154	6	7
Class 1	36	CNN	84	86	81	82	85	31	148	13	9
Class 2	42	DenseNet121	86	85	83	84	87	32	152	8	8
Class 3	38	Inception	83	82	80	81	84	30	146	14	10
Class 4	44	Xception	88	87	85	86	89	34	156	4	6
Total	200	MobileNetV2	82	81	79	80	83	29	143	17	11

TP: true positive; TN: true negative; FP: false positive; FN: false negative.

Although visually similar regions, such as sand on the right and leaves on the left, are assigned multiple redundant edges in ViG, these connections are often irrelevant to downstream tasks. Vi-HGNN reduces such redundancy and computational overhead by minimizing the number of hyperedges required to model these interactions. While ViG may generate noisy edges linking semantically unrelated yet visually similar patches, Vi-HGNN effectively captures higher-order relationships among patches and demonstrates robustness against noisy connections.

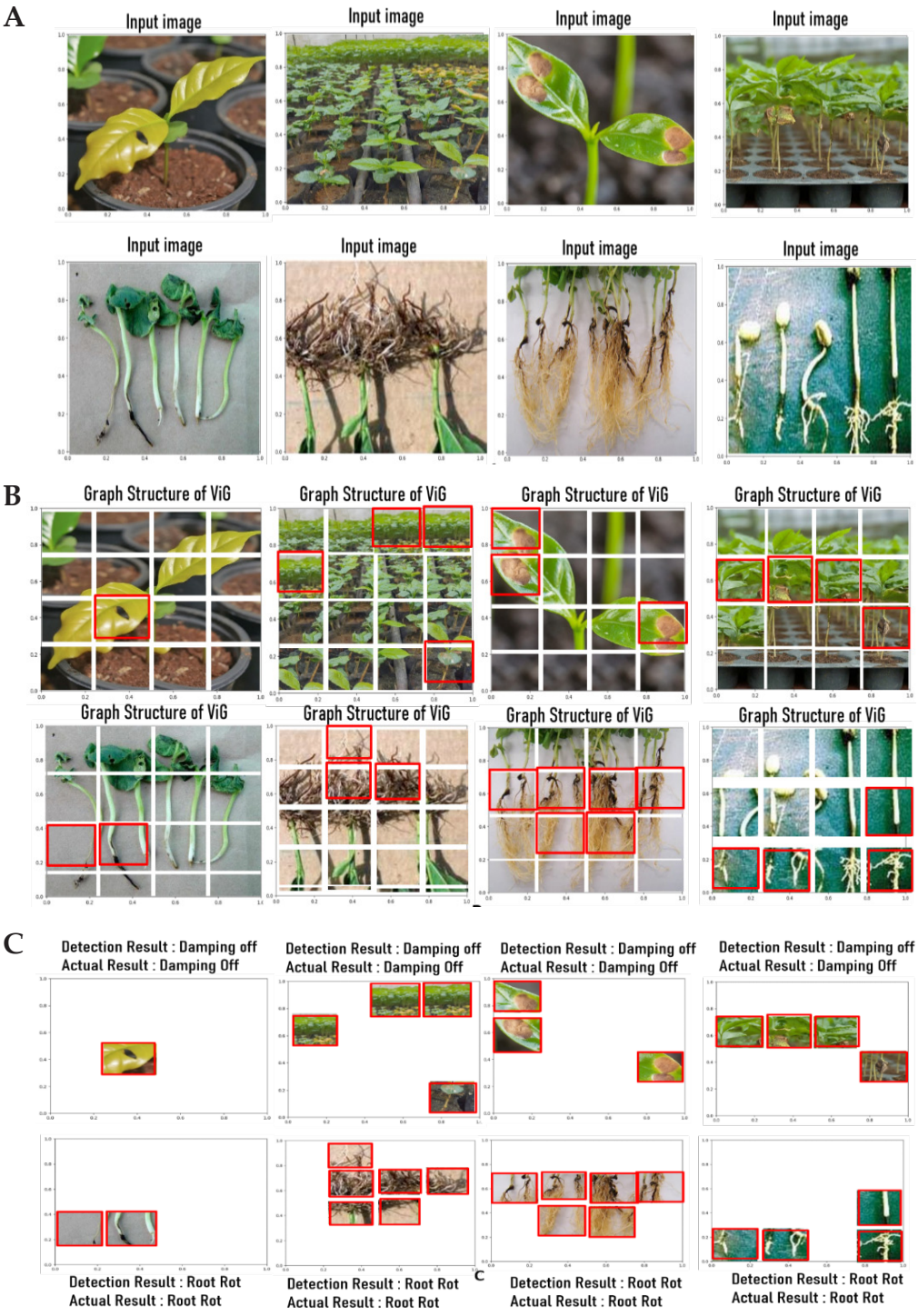

Figure 8. Visualization of the Vision-based Heterogeneous Graph Neural Network (Vi-HGNN) hypergraph structure. Each frame represents a hyperedge. A) Input image; B) graph structure of the Visual Graph (ViG); C) hypergraph structure of thee ViG-GNN.

CONCLUSIONS

The proposed Vision-based Heterogeneous Graph Neural Network (Vi-HGNN) represents a novel framework that models images as hypergraphs to capture complex inter-patch relationships beyond conventional convolutional or graph-based methods. The model effectively overcomes the limitations of traditional image representation methods by using image patches as nodes and applying hypergraph connectivity. The Vi-HGNN achieved 97.77 % accuracy in distinguishing damping-off and root rot diseases in coffee plants using the cross-entropy feature. Future research will aim to automate hypergraph structure optimization and develop hybrid Vi-HGNN architectures for more robust and scalable pest and disease management in agricultural systems.

ACKNOWLEDGMENTS

The authors extend their appreciation to King Saud University for funding the publication of this research through the Researchers Supporting Project number (RSPD2024R809), King Saud University, Riyadh, Saudi Arabia.

Data availability: Raveena S. 2024. Damping off and root rot disease (Data set). Zenodo. https://doi.org/10.5281/zenodo.11406946

Coding: Raveena S. 2024. Damping off and root rot disease (coding part). Zenodo. https://doi.org/10.5281/zenodo.11407023

REFERENCES

Abbas A, Zhang Z, Zheng H, Alami MM, Alrefaei AF, Abbas Q, Naqvi SA, Rao MJ, Mosa WF, Abbas Q, *et al*. 2023. Drones in plant disease assessment, efficient monitoring, and detection: A way forward to smart agriculture. Agronomy 13 (6): 1524–1544. https://doi.org/10.3390/agronomy13061524

Abdelrhim AS, Abdellatif YM, Hossain MA, Alamri S, Pessarakli M, Lessy AM, Dawood MF. 2023. Comparative study of three biological control agents and two conventional fungicides against coriander damping-off and root rot caused by *Rhizoctonia solani*. Plants 12 (8): 1694–1714. https://doi.org/10.3390/plants12081694

Abdullah HM, Mohana NT, Khan BM, Ahmed SM, Hossain M, Islam KS, Redoy MH, Ferdush J, Bhuiyan MA, Hossain MM, *et al*. 2023. Present and future scopes and challenges of plant pest and disease (P&D) monitoring: Remote sensing, image processing, and artificial intelligence perspectives. Remote Sensing Applications: Society and Environment 32: 100996. https://doi.org/10.1016/j.rsase.2023.100996

Alharbi M, Rajagopal SK, Rajendran S, Alshahrani M. 2023. Plant disease classification based on ConvLSTM U-Net with fully connected convolutional layers. Traitement du Signal 40 (1): 157–166. https://doi.org/10.18280/ts.400114

Appavu N. 2025. Detection and Classification of plant disease using hybrid AI Deep Learning techniques. *In* 2025 2025 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence.

Institute of Electrical and Electronics Engineers. Chennai, India, https://doi.org/10.1109/raeeucci63961.2025.11048289

Bouguettaya A, Zarzour H, Kechida A, Taberkit AM. 2023. A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images. Cluster Computing 26 (2) :1297–1317. https://doi.org/10.1007/s10586-022-03627-x

El-Abeid, Sozan E, Mosa MA, El-Tabakh MAM, Saleh AM, El-Khateeb MA, Haridy MSA. 2024. Antifungal activity of copper oxide nanoparticles derived from *Zizyphus spina* leaf extract against *Fusarium* root rot disease in tomato plants. Journal of Nanobiotechnology 22 (1). https://doi.org/10.1186/s12951-023-02281-8

Ezzeldin I, Ahmad AA, Abdo S, Bakr MA, Khalil MA, Abdallah Y, Ogunyemi SO. 2024. Suppression of root rot fungal diseases in common beans (*Phaseolus vulgaris* L.) through the application of biologically synthesized silver nanoparticles. Nanomaterials 14 (8): 710–725. https://doi.org/10.3390/nano14080710

Ferreira LVM, Leite RA, de Carvalho F, Andrade JFC, de Medeiros FHV, Moreira FMS. 2023. Rhizobacteria control damping-off and promote the growth of lima beans with and without co-inoculation with *Rhizobium tropici* CIAT899. Archives of Microbiology 205 (5). https://doi.org/10.1007/s00203-023-03555-3

Han B, Lu Z, Zhang J, Almodfer R, Wang Z, Sun W, Dong L. 2024. Rep-ViG-Apple: A CNN-GCN hybrid model for apple detection in complex orchard environments. Agronomy 14 (8): 1733. https://doi.org/10.3390/agronomy14081733

Han Y, Wang P, Kundu S, Ding Y, Wang Z. 2023. Vision HGNN: An image is more than a graph of nodes. *In* 2023 IEEE/CVF International Conference on Computer Vision. Institute of Electrical and Electronics Engineers. Paris, France, pp: 19878–19888. https://doi.org/10.1109/iccv51070.2023.01820

Jasiman F, Fourati LC. 2023. Agriculture 4.0 from IoT, artificial intelligence, drone, and blockchain perspectives. *In* 2023 15th International Conference on Developments in eSystems Engineering. Institute of Electrical and Electronics Engineers. Baghdad, Iraq, pp: 262–267. https://doi.org/10.1109/dese58274.2023.10099927

Lamprecht SC, Phasoana TJ, van Wyk W, Spies CFJ. 2024. First report of *Pythium* defense causing damping-off and root rot of soybean in South Africa. Journal of Plant Pathology 106 (2): 771. https://doi.org/10.1007/s42161-024-01592-5

Lawrence ID, Vijayakumar R, Agnishwar J. 2024. Dynamic application of unmanned aerial vehicles for analyzing the growth of crops and weeds for precision agriculture. *In* Gupta RK, Jain A, Wang J, Bharti SK, Patel S. (eds.), Artificial Intelligence Tools and Technologies for Smart Farming and Agriculture Practices. IGI Global Scientific Publishing: Hershey, PA, USA, pp: 115–132. https://doi.org/10.4018/978-1-6684-8516-3.ch007

Li J, Ai M, Hou J, Zhu P, Cui X, Yang Q. 2024. Plant pathogen interaction with root rot of *Panax notoginseng* as a model: Insight into pathogen pathogenesis, plant defense response, and biological control. Molecular Plant Pathology 25 (2): 13–27. https://doi.org/10.1111/mpp.13427

Liu W, Yan C, Li R, Chen G, Wang X, Wen Y, Zhang C, Wang X, Xu Y, Wang Y. 2023. *VqMAPK3*/*VqMAPK6*, *VqWRKY33*, and *VqNSTS3* constitute a regulatory node in enhancing resistance to powdery mildew in grapevine. Horticulture Research 10 (7). https://doi.org/10.1093/hr/uhad116

Lovas A, Lytras I, Rasonyi M, Sabanis S. 2023. Taming neural networks with Tusla: Nonconvex learning via adaptive stochastic gradient Langevin algorithms SIAM. Journal on Mathematics of Data Science 5 (2): 323–345. https://doi.org/10.1137/22m1514283

Puri V, Nayyar A, Raja L. 2017. Agriculture drones: A modern breakthrough in agriculture. Journal of Statistics and Management Systems 20 (4): 507–518. https://doi.org/10.1080/09720510.2017.1395171

Rahman MZ, Talukder MMR, Liberia MG, Rahman MA, Islam MR. 2022. Management of damping-off disease in bottle gourd seedlings under floating agriculture. Bangladesh Journal of Plant Pathology 38 (2): 9–14.

Rajagopal MK, Bala Murugan MS. 2023. Artificial Intelligence-based drone for early disease detection and precision pesticide management in cashew farming. arXiv. https://doi.org/10.48550/arXiv.2303.08556

Raveena S, Surendran R, Khalaf AOI, Hamam H. 2024. Empowering coffee farming using counterfactual recommendation based RNN driven IoT integrated soil quality command system. Scientific Reports 14 (1): 6269–6280. https://doi.org/10.1038/s41598-024-56954-x

Saravanan MS. 2025. Smart plant disease management system integrating temporal data and deep learning techniques. *In* 2025 2025 6th International Conference on Data Intelligence and Cognitive Informatics. Institute of Electrical and Electronics Engineers. Tirunelveli, India, pp: 771–776. https://doi.org/10.1109/icdici66477.2025.11134823

Selvanarayanan R, Rajendran S, Alotaibi Y. 2024. Early detection of *Colletotrichum kahawae* disease in coffee cherry based on computer vision techniques. Computer Modeling in Engineering and Sciences 139 (1): 759–782. https://doi.org/10.32604/cmes.2023.044084

Tamilvizhi T, Surendran R, Anbazhagan K, Rajkumar K. 2022. Quantum behaved particle swarm optimization-based deep transfer learning model for sugarcane leaf disease detection and classification. Mathematical Problems in Engineering 3: 524–533. https://doi.org/10.1155/2022/3452413

Thakur S, Raj H. 2024. Synergistic strategies for sustainable crop protection: Harnessing soil solarization and biofumigants to combat damping-off pathogens in solanaceous vegetable crops. Journal of Plant Diseases and Protection 2: 1–10 https://doi.org/10.21203/rs.3.rs-3238787/v1

Yu C, Lv J, Xu H. 2024. Plant growth-promoting fungi and rhizobacteria control *Fusarium* damping-off in Mason pine seedlings by impacting rhizosphere microbes and altering plant physiological pathways. Plant and Soil 499 (1–2): 503–519. https://doi.org/10.1007/s11104-024-06475-3

Zhang B, Aidong C, Sun M, Zheng W, Zhang H, Fu Z, Zhao S, Zhang Z, Wang L, Zhang H, *et al*. 2024. Effect of five seed-coating agents on the germination rate and the evaluation of control effect on the damping-off disease of sugar beet. Sugar Tech 26 (6): 1746–1750. https://doi.org/10.1007/s12355-024-01360-w

Agrociencia

ELICITATION OF PRIOR DISTRIBUTIONS TO ESTIMATE VARIANCE COMPONENTS

Ignacio **Luna-Espinoza**[1,2]*, Paulino **Pérez-Rodríguez**[1], José **Crossa**[1,3], Osval Antonio **Montesinos-López**[4], Juan Manuel **Romero-Padilla**[1]

[1]Colegio de Postgraduados Campus Montecillo. Carretera México-Texcoco km 36.5, Montecillo, Texcoco, State of Mexico, Mexico. C. P. 56264.
[2]Universidad del Istmo. Ciudad Universitaria S/N, Ciudad Ixtepec, Oaxaca, Mexico. C. P. 70110.
[3]Centro Internacional de Mejoramiento de Maíz y Trigo. Carretera México-Veracruz km 45, El Batán, Texcoco, State of Mexico, Mexico. C. P. 56237.
[4]Universidad de Colima. Facultad de Telemática. Avenida Universidad 333, Las Víboras, Colima, Mexico. C. P. 28040.
* Author for correspondence: ignaciolunaespinoza@gmail.com

Citation: Luna-Espinoza I, Pérez-Rodríguez P, Crossa J, Montesinos-López OA, Romero-Padilla JM. 2026. Elicitation of prior distributions to estimate variance components. **Agrociencia** 60(1): 23-41. https://doi.org/ 10.47163/agrociencia.v60i1.3350

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: June 05, 2025.
Approved: January 12, 2026.
Published in Agrociencia: January 30, 2026.

ABSTRACT

In the Bayesian context, when variance components are modeled in normal hierarchical models, the inverted gamma distribution (IG) is typically used as the prior density for each component. However, the literature indicates that this prior density is highly informative, and thus the Half Cauchy distribution (HC) is recommended. The aim of this study was to evaluate, using simulation (in the context of high-dimensional data such as in the case of genomic selection applications), the suitability of the scaled inverse chi-squared $(\chi_{v,S}^{-2})$ distribution, which belongs to the family of scaled inverse gamma distributions, and HC as prior densities for the variance components in the Bayesian Ridge regression model. The evaluation was carried out when the number of observations in the response variable is greater than the number of predictor variables ($n > p$) as well as in high dimensions ($n << p$). The Bayesian learning of the posterior distribution was evaluated using the Hellinger distance (HD). The results of the Bayesian analysis were also compared with those obtained with the restricted maximum likelihood (REML). Results indicate that when $n > p$, the REML method underestimates the variance of the random effect, whereas in scenarios in which $n << p$, the method overestimates the same parameter when the variance of the error is large (greater than or equal to 6.0) and gives consistent estimations when the error variance is moderate (equal to 1.0). On the other hand, under prior distribution $(\chi_{v,S}^{-2})$ and in both scenarios ($n > p$) and $n << p$, it was observed that the parameters can be overestimated or underestimated, depending on the fixed values used to simulate the data. For the case of the HC prior distribution, the credibility intervals for both the variance of the effects of the predictor variables and the variance of the error contain the true values of the parameters, and their precisions increase with the sample size.

Keywords: Bayesian learning, scale inverse chi-squared distribution, Half Cauchy distribution, Hellinger distance.

INTRODUCTION

For the past two decades, the development of data-acquisition technology has enabled diverse gadgets to take anywhere between thousands and millions of measurements of one single unit of interest. For example, DNA microarray technology measures the expression levels of thousands of genes. Using the information gathered, it is possible to understand the biological regulation mechanisms and develop new drugs. In these data, the number p of predictor variables amounts to thousands, while the number n of individuals involved in the experiment is much lower, that is, $n << p$ (Giraud, 2015). Another example of data in high dimensions is raised when taking images and videos, where large image databases are being continually gathered, such as medical or astrophysical images. Each image can be made up of millions of pixels or voxels. In the case of medical images, the number p of pixels is generally much greater than the number n of patients in the study.

From the statistics and computer science point of view, the challenge is determining the function $\Phi(\cdot)$, in such a way that $\Phi(x_1, ..., x_p)$ predicts the response variable y exactly and accurately. Working with massive data is extremely challenging because separating useful information from noise is almost impossible (Giraud, 2015). This problem is known as the curse of dimensionality. Although the proposals found in the literature show the existence of countless functions $\Phi(\cdot)$, linear models $y_i = \mu + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i$ are some of the best to carry out predictions in high dimensions. Under the Bayesian approach, different prior distributions have been assigned to regression coefficients β_j, giving rise to the "Bayesian Alphabet" (Gianola, 2013).

In this study, errors were assumed to be independent with a normal distribution, with a mean of zero and a variance of ϕ_ϵ, that is, $\epsilon_i \sim \text{iid}N(0, \phi_\epsilon)$; whereas β_j regression coefficients are assigned normal independent prior distributions with a mean of zero and a variance of ϕ_β, that is, $\beta_j \sim \text{iid}N(0, \phi_\beta)$. It is worth raising the question of which are the prior densities that must be assigned to the ϕ_β and ϕ_ϵ variance components, in such a way that they enable a full Bayesian learning.

Until a little more than two decades ago, when variance components were modeled in normal hierarchical models (in models structured from frameworks that allow the analysis of data organized into levels or nested groups) (Gelman and Hill, 2007), the IG prior distribution with the shape parameter equal to the scale parameter IG (ε, ε) was typically used. According to Spiegelhalter *et al.* (1996), the frequent use of this distribution was due to the fact that, in the BUGS software, it was the only prior distribution available for modeling variance components. Its wide use can be explained by an argument provided by Fink (1997), who points out that before the general availability of computers, joint prior distributions were used in order for the product of the prior density function and the likelihood to be analytically treatable.

In this context, the full conditional distribution of each variance component is also IG, which facilitates the computational implementation of the Gibbs sampler to generate samples from the posterior distribution. Otherwise, if the full conditional distribution of the variance component is complex (when its kernel does not correspond with

that of a known univariate density function), the implementation of a Markov Chain Monte Carlo (MCMC) sampling technique must be implemented to obtain samples from the posterior distribution, which entails a high computational cost, such as the requirement for large amounts of computer time, particularly in high-dimensional settings. In contrast, assigning the IG prior to each variance component does not present the above issue.

Due to the widespread use of the IG prior distribution to model variance components, it is worth asking how appropriate the use of this prior distribution is for the modeling of variance components. A pioneering study evaluating this distribution as non-informative prior was conducted by Browne and Draper (2006). They point out that the use of this prior density must be taken with caution, since the IG distribution (ε, ε) has a peak near zero, and this can lead to problems when the random component variance is close to zero. Coincidentally, Gelman (2006) argues that the selection of small values for hyperparameter ε in the prior IG density (ε, ε) causes the posterior distribution of the variance components to produce small values; that is, a noticeable shrinkage occurs of the posterior distribution towards zero which causes an imperfect Bayesian learning.

On the other hand, Polson and Scott (2012) point out that it is not appropriate to assign the IG prior distribution to scale parameters in higher levels of normal hierarchical models; instead, they recommend the use of the HC density, particularly when proper prior distributions are required. The recommendation is supported by the fact that the HC prior distribution has a good performance close to zero and shows no drastic influence in other parts of the parametric space, which provides a solid justification for its routine use. The HC distribution is heavy-tailed, which implies the frequent occurrence of extreme values. This characteristic may facilitate the Bayesian learning of the posterior distribution of the corresponding variance component, as it leads to a broad exploration of the parametric space. Additionally, Polson and Scott (2012) point out that in contexts in which the regression coefficient vector is sparse, that is, where most coefficients are zero and only few have relevant effects, IG conjugated prior distribution can severely distort inferences.

In the framework of the Bayesian alphabet, Gianola *et al.* (2009) implicitly expose that prior distribution $(\chi_{\nu,s}^{-2})$ in the BayesA and BayesB models has a strong influence on the variance of the effects of the predictor variables. These results are ratified by Lehermeier *et al.* (2013), who point out the strong influence of hyperparameters in the predictive ability of the BayesA and BayesB models. The choice of a large value for the scale of the parameter of prior distribution $(\chi_{\nu,s}^{-2})$ assigned to the variance of the effects leads to an overfitting of the data. Meanwhile, a very small value leads to underfitting due to a noticeable shrinking of the effects. In both cases, the predictive ability of the models is reduced.

In the context of Bayesian learning, as the sample size increases, the influence of the prior distribution should disappear gradually, hence the prior distribution would have little importance in large samples (Bernardo and Smith, 1994). This result is valid

for parameters with maximum likelihood estimators that are identifiable (Gianola, 2013). However, in schemes in which $n << p$, Bayesian learning may arise only for n parameters or functions of n parameters, since $p - n$ parameters are not identifiable; hence, when $n << p$ the prior distributions will always be important, and their influence will never disappear, even with large samples (Gianola, 2013).

Motivated by the above arguments and by the properties of the HC distribution, the aim of this study is to evaluate, through simulation, the suitability of assigning the $(\chi_{v,S}^{-2})$ and HC prior distributions to the variance components in Bayesian Ridge regression. The results of both distributions were compared with the estimations using the classic frequentist method REML. The Bayesian learning capacity of the respective models was evaluated using the Hellinger distance (HD) between the marginal prior distribution assigned to the variance component and the corresponding posterior density. An HD nearer to one indicates that the posterior distribution moves away from the prior distribution, thus providing evidence of a robust Bayesian learning process. In contrast, an HD closer to zero will indicate a lack of Bayesian learning due to the influence of the prior distribution and its hyperparameters.

MATERIALS AND METHODS

The following section presents the sampling distribution and the structure of the three models evaluated: a) linear mixed model, b) a Bayesian model with a $(\chi_{v,S}^{-2})$ prior assigned to each variance component, and c) a Bayesian model with an HC prior assigned to the positive square root of each variance component.

Sampling model and prior distributions

Consider the standard linear model:

$$y = 1\mu + X\beta + \epsilon, \tag{1}$$

where y is a vector of dimension $n \times 1$, 1 is a vector of ones of dimension $n \times 1$, μ is an intercept, $X = \{x_{ij}\}$ is the incidence matrix of dimension $n \times p$, β is a vector of dimension $p \times 1$, and ϵ is a vector of errors of dimension $n \times 1$. The errors are assumed to be independent and normally distributed with mean zero and variance ϕ_ϵ, that is, $\epsilon \sim N(0, \phi_\epsilon I)$, where I is the identity matrix of order $n \times n$. The above elements define the following conditional distribution of the data:

$$\mathrm{p}(\boldsymbol{y}|\mu, \boldsymbol{\beta}, \phi_\epsilon) = \prod_{i=1}^{n} N\left(y_i|\mu + \sum_{j=1}^{p} x_{ij}\,\beta_j, \phi_\epsilon\right), \tag{2}$$

where the expression $N\left(y_i|\mu + \sum_{j=1}^{p} x_{ij}\,\beta_j, \phi_\epsilon\right)$ denotes the density function of a normal random variable for y_i, given the mean a and variance b, that is, $N(y_i|a, \mathrm{b})$.

Linear mixed model

In equation (1), if u = $X\beta$, where $\beta \sim N(0, \phi_\beta I)$, by the properties of the multivariate normal distribution, the following model is obtained:

$$y = 1\mu + \mathrm{u} + \epsilon, \tag{3}$$

where $\mathrm{u} \sim \mathrm{N}(0, \phi_\beta XX')$ and $\epsilon \sim N(0, \phi_\epsilon I)$. In the context of genomic selection, X is the matrix of markers coded into numerical values (for example, 0, 1, 2 to denote the number of times the major allele appears), XX' is proportional to the variance and covariance matrix of the random term u or what is known as the linear kernel in the RKHS non-parametric regression models (de los Campos *et al*., 2010). Equation (3) defines a standard linear mixed model where 1μ is a fixed term and u is a random effects vector.

Prior density for the intercept

The intercept was assigned a normal prior distribution with a mean zero and a variance $k = 1 \times 10^{10}$. This assignment is similar to assigning a flat prior distribution to the intercept.

Prior density for the regression coefficients

In the Ridge Bayesian regression model, the following prior distribution is assigned to the regression coefficients:

$$\mathrm{p}(\beta|\phi_\beta) = \prod_{j=1}^{p} \mathrm{N}(\beta_j|0, \phi_\beta) \tag{4}$$

Notice that the shrinking of the regression coefficients towards zero depends on variance component ϕ_β; the smaller variance ϕ_β, the greater the concentration of the prior density will be around zero.

A) Scaled inverse chi-squared prior to the variance components. Variance parameter ϕ_β is assigned a prior density $\chi^{-2}_{v_\beta, S_\beta}$, that is, $p(\phi_\beta|v_\beta, S_\beta) = \chi^{-2}(\phi_\beta|v_\beta, S_\beta)$, where $v_\beta > 0$ and $S_\beta > 0$ denote the degree of freedom and the scale parameter, respectively. This study uses the parametrization of prior distribution $\chi^{-2}(\phi_\beta|v_\beta, S_\beta)$ so that $E(\phi_\beta) = S_\beta/(v_\beta - 2)$ and $mode(\phi_\beta) = S_\beta/(S_\beta + 2)$. Similarly, ϕ_ϵ is assigned the prior distribution $\mathrm{p}(\phi_\epsilon|v_\epsilon, S_\epsilon) = \chi^{-2}(\phi_\epsilon|v_\epsilon, S_\epsilon)$. The joint prior distribution is expressed as follows (using the prior for the intercept, equation (4) and the prior for the variance components):

$$\mathrm{p}(\mu, \beta, \phi_\beta, \phi_\epsilon|H_1) = \mathrm{p}(\mu)\mathrm{p}(\beta|\phi_\beta)\mathrm{p}(\phi_\beta|v_\beta, S_\beta)\mathrm{p}(\phi_\epsilon|v_\epsilon, S_\epsilon), \tag{5}$$

where $H_1 = \{k, v_\beta, S_\beta, v_\epsilon, S_\epsilon\}$ is a set of values of the hyperparameters for the prior distributions. By applying Bayes' theorem, the posterior distribution is proportional to the product of the prior distribution (equation 5) and the likelihood (equation 2):

$$\begin{aligned} \mathrm{p}(\mu, \beta, \phi_\beta, \phi_\epsilon | y) \quad &\propto \quad \mathrm{N}(\mu|0,k)\left\{\prod_{j=1}^{p}\mathrm{N}(\beta_j|0,\phi_\beta)\right\}\chi^{-2}(\phi_\beta|v_\beta,S_\beta)\,\chi^{-2}(\phi_\epsilon|v_\epsilon,S_\epsilon) \\ &\times \quad \left\{\prod_{i=1}^{n} N\left(y_i|\mu+\sum_{j=1}^{p}x_{ij}\,\beta_j,\phi_\epsilon\right)\right\}. \end{aligned} \tag{6}$$

The kernel of the function of the posterior distribution in equation (6) does not correspond to that of any known distribution, but samples can be obtained using simulation procedures based on MCMC algorithms. The full conditional distributions are known, and it is therefore possible to obtain samples from the posterior distribution using the Gibbs sampler (Casella and George, 1992).

B) Half Cauchy distribution assigned to the standard deviation of the variance component. Following Gelman (2006), standard deviation $\psi_\beta = \sqrt{\phi_\beta}$ is assigned the standard HC prior density:

$$\mathrm{p}(\psi_\beta) = 2\pi^{-1}\left(1+\psi_\beta^2\right)^{-1}. \tag{7}$$

Notice that $\phi_\beta = \psi_\beta^2$ is the variance component. Using the transformation method (Casella and Berger, 2002), to transform from the standard deviation to the variance, the prior density for variance parameter ϕ_β is equal to:

$$\mathrm{p}(\phi_\beta) = \pi^{-1}\phi_\beta^{-1/2}\left(1+\phi_\beta\right)^{-1}, \tag{8}$$

for $\phi_\beta > 0$. Notice that ψ_β and ϕ_β have the same domain. Similarly, the HC prior density is obtained for variance component ϕ_ε, which is given by:

$$\mathrm{p}(\phi_\varepsilon) = \pi^{-1}\phi_\varepsilon^{-1/2}(1+\phi_\varepsilon)^{-1}, \tag{9}$$

for $\phi_\varepsilon > 0$. In addition, assuming independence, the joint prior distribution can be represented as follows (using equations 7, 8 and 9):

$$\mathrm{p}(\mu, \beta, \phi_\beta, \phi_\epsilon | H_2) = \mathrm{p}(\mu)\mathrm{p}(\beta|\phi_\beta)\mathrm{p}(\phi_\beta)\mathrm{p}(\phi_\epsilon). \tag{10}$$

Here, the set of values of the hyperparameters for the prior distributions has only one element, $H_2 = \{k\}$. By applying Bayes' theorem, the joint posterior distribution is proportional to the product of the prior distribution (equation 10) and the likelihood (equation 2):

$$\begin{aligned} \mathrm{p}(\mu, \beta, \phi_\beta, \phi_\epsilon | y) \quad &\propto \quad \mathrm{N}(\mu|0,k)\left\{\prod_{i=1}^{p} \mathrm{N}(\beta_i|0,\phi_\beta)\right\}\phi_\beta^{-\frac{1}{2}}\phi_\varepsilon^{-\frac{1}{2}}(1+\phi_\beta)^{-1}(1+\phi_\varepsilon)^{-1} \\ &\times \quad \left\{\prod_{i=1}^{n} N\left(y_i|\mu+\sum_{j=1}^{p} x_{ij}\,\beta_j, \phi_\epsilon\right)\right\}. \end{aligned} \tag{11}$$

The kernel of the function given in equation (11) does not correspond with any known distribution, but samples can be obtained using simulation procedures based on MCMC algorithms. The full conditional distributions for the intercept and the regression coefficients can be verified to be normal. In this context, to perform an inference from the posterior distribution, a hybrid mechanism was implemented, using the Gibbs sampler for sampling the intercept and regression coefficients, and the Metropolis algorithm with a random walk for the variance components.

Selection of hyperparameters for prior distributions $\chi^{-2}(\boldsymbol{\phi_\beta}|\boldsymbol{\nu_\beta}, \boldsymbol{S_\beta})$ and $\chi^{-2}(\boldsymbol{\phi_\epsilon}|\boldsymbol{\nu_\epsilon}, \boldsymbol{S_\epsilon})$
To select the hyperparameters of prior densities $\mathrm{p}(\phi_\beta|\nu_\beta, S_\beta)$ and $\mathrm{p}(\phi_\epsilon|\nu_\epsilon, S_\epsilon)$, the guidelines recommended by Pérez-Rodríguez and de los Campos (2014) and Pérez-Rodríguez *et al*. (2018) were followed. The recommendations establish that the variance of response variable y_i can be divided into two components: the one of the linear predictor and the one of the error:

$$\begin{aligned} Var(y_i) \quad &= \quad Var\left(\sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i\right) = Var(\mathrm{g}_i) + Var(\epsilon_i) \\ &= \quad \sum_{j=1}^{p} x_{ij}^2 Var(\beta_j) + Var(\epsilon_i) = \phi_\beta \sum_{j=1}^{p} x_{ij}^2 + \phi_\epsilon, \end{aligned} \tag{12}$$

for $i = 1, \ldots, n$, where $\mathrm{g}_i = \sum_{j=1}^{p} x_{ij}\beta_j$. On the other hand, the total variance due to the linear predictor is equal to $\sum_{i=1}^{n} Var(g_i) = \phi_\beta \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2$. Thus, Pérez-Rodríguez *et al.* (2018) showed that the average prior variance due to the linear predictor in equation (12) is equal to:

$$V_g = \frac{1}{n}\sum_{i=1}^{n} Var(g_i) = \frac{1}{n}\phi_\beta \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2 = \phi_\beta MS_x$$

where $MS_x = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2$ and the variance of response variable V_y is equal to

$$V_y = V_g + V_\epsilon = \phi_\beta MS_x + \phi_\epsilon, \tag{13}$$

where $V_\epsilon = var(\epsilon_i) = \phi_\epsilon$. From equation (13), it follows that

$$\phi_\beta = \frac{V_g}{MS_x}. \tag{14}$$

By replacing ϕ_β by the mode of the density $\chi^{-2}(\phi_\beta|v_\beta, S_\beta)$ in equation (14) and solving for scale parameter S_β, $S_\beta = V_g(v_\beta + 2)/MS_x$. Moreover, using the concept of heritability in the context of quantitative genetics, $h^2 = V_g/V_y$, it follows that $V_g = h^2 V_y$ (Falconer and Mackay, 1996). Therefore,

$$S_\beta = h^2 V_y(v_\beta + 2)/MS_x = \frac{R^2 V_y\left(v_\beta + 2\right)}{MS_x}$$

where R^2 is the proportion of the variance that is explained a priori by the predictor variables. By default, similar to Pérez-Rodríguez *et al.* (2018), it is considered that R^2 = 0.5. Therefore, for fixed v_β and given the y data vector and matrix X, the terms V_y, MS_x, S_β are calculated.

A similar procedure (v_β, S_β) was used to obtain the hyperparameters (v_ϵ, S_ϵ) (Pérez-Rodríguez *et al.*, 2018). The result of the procedure is the hyperparameters $(v_\epsilon, S_\epsilon) = (v_\epsilon, \left(1 - R^2\right)V_y(v_\epsilon + 2)\right)$. On the other hand, to determine the respective degrees of freedom v_β and $v_\in$, the recommendations by Pérez-Rodríguez and de los Campos (2014) were followed, of taking small degrees of freedom to reduce the influence of the prior density on the posterior density. This study considered $v_\beta = v_\epsilon = 5$, since they are the minimum degrees of freedom that ensure finite prior mean and variance.

Discrepancy in the prior and posterior distributions

In Bayesian learning, it is important to evaluate how much useful information is obtained as progress is made in the learning process; in other words, what the discrepancy is between the prior and posterior densities as new evidence, generated with uncertainty under the sampling model, is incorporated. To evaluate the sensitivity of the models, the HD was used. This distance, unlike other measurements used to evaluate the influence of the *f* prior distribution on the posterior *g*, such as the

Kullback-Leibler divergence measurement, has the characteristic of being completely defined in situations in which one of the densities reaches the value of zero and the other one does not. This property, for example, does not appear in the measurement based on the Kullback-Leibler divergence, because when one of the densities reaches the value of zero and the other one does not, the Kullback-Leibler measurement becomes infinite (Roos and Held, 2011). The HD between densities f and g, which is denoted by $H(f, g)$, is defined by:

$$H(f,g) = \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\left(\sqrt{f(u)} - \sqrt{g(u)}\right)^2 du} = \sqrt{1 - \int_{-\infty}^{\infty}\sqrt{f(u)g(u)}du}$$

It is possible to verify that $0 \leq H(f,g) \leq 1$; in addition, $H(f,g) = 0$ if, and only if, $f = g$ (Roos and Held, 2011). On the other hand, $H(f,g) = 1$ if density f assigns a probability of 0 to each set that g assigns a positive probability and vice-versa.

When the posterior distribution is highly influenced by the election of the hyperparameters that are assigned to the prior distributions, the posterior distribution does not deviate from the prior distribution, which results in a lack of Bayesian learning (Lehermeier *et al*., 2013). In this case, HD will tend towards zero. By contrast, if the distance $H(f, g)$ between prior density f and posterior density g tend towards one, the knowledge of the state of nature is continually modified as new data is acquired. The HD is small if the perfect prior density is used. In this case, a small HD is not indicative of the lack of Bayesian learning. However, Lehermeier *et al*. (2013) claim that selecting the perfect prior density has a probability close to zero, particularly if there is a lack of prior information.

To evaluate the effect of the prior density of each variance component, the HD between the prior marginal density and the posterior density was determined numerically. The posterior marginal density was estimated from the MCMC samples with an estimation of non-parametric densities with a Gaussian kernel and a bandwidth parameter given by Silverman's criterion (Silverman, 1998).

Simulated data and fitting of the models

Both the response variable and the predictor variables were simulated, creating complex scenarios similar to what is done in genomic selection (Meuwissen *et al.*, 2001; Pérez-Rodríguez and de los Campos, 2014) in the field of quantitative genetics, where the prediction of a continuous variable (phenotype) based on thousands of covariables (molecular markers) is of interest. The data were simulated using the XSim software implemented in the Julia package (Bezanson *et al*., 2017), based on the method proposed by Cheng *et al*. (2018), which records the positions and the origins of the founders of each chromosome segment.

First, the chromosomes of the founders are labelled using identifiers, and each chromosome is represented by two vectors: one that indicates the origin of the

founders and the other, the crossing positions. The states of the alleles of the founding genomes are generated, whether from map positions defined by the user and the allele frequency or from real haplotypes or data sequences. In meiosis, the new gamete is formed with chromosomic segments of the paternal and maternal gametes of the parents, with segments introduced in both sides of the crossing.

The simulation considered a chromosome length of 0.1 cM, 10 QTLs for each chromosome, and a population of 500 unrelated founders. The 10 QTLs distributed at random in each chromosome made up the significant regression coefficients in the Bayesian Ridge regression model. In independent simulations, chromosome numbers 4, 5, 7, 9, 12, 24, and 48 were taken. These chromosome numbers were chosen in order to observe the behavior of the fittings as the number of predictors gradually increases. Each predictor variable takes the values of 0, 1, or 2, according to the frequency of the major allele. The predictor variables with no variation were eliminated from their respective set. This framework produced sets of predictor variables with 349, 456, 616, 785, 1037, 2059, and 4090 variables, respectively. Additionally, the sample sizes considered were 200, 400, 600, 800, 1000, 2000, 4000, 6000, and 8000 (Figure 1), which enables the evaluation of the models in scenarios $n > p$, and $n < p$, including $n << p$.

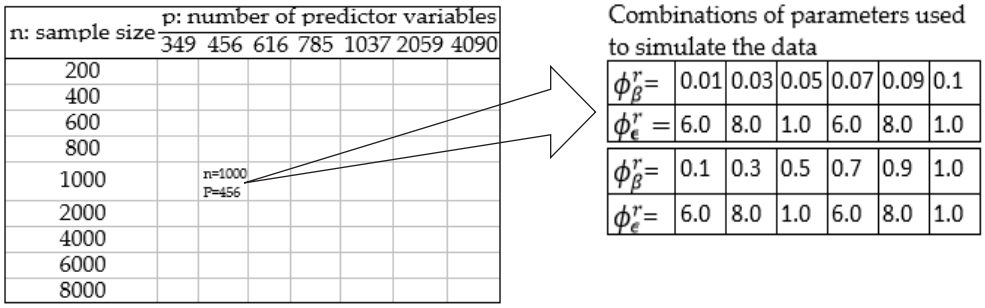

Figure 1. Simulation grid. Sample size (n) × number of predictor variables (p), and combinations of parameters used to simulate the data in each point of the grid $n \times p$.

To evaluate the influence of the prior densities against small and large variations in the variance components, in each point of the $n \times p$ grid, the response variable was simulated with the combination of the pairs of fixed values $(\phi_\beta^r, \phi_\epsilon^r)$ = (0.01, 6.0), (0.03, 8.0), (0.05, 1.0), (0.07, 6.0), (0.09, 8.0), (0.1, 1.0), (0.1, 6.0), (0.3, 8.0), (0.5, 1.0), (0.7, 6.0), (0.9, 8.0), (1.0, 1.0). Values of $0.01 \leq \phi_\beta^r < 0.1$ helped observe the performance of the fittings against ϕ_β^r values close to zero, whereas with values of $0.1 \leq \phi_\beta^r \leq 1$, showed the behavior of the fittings against values of ϕ_β^r far away from zero. On the other hand, the value of $\phi_\epsilon^r = 1.0$ helps observe the performance of the fittings against little noise in the error, whereas the values of $\phi_\epsilon = 6.0, 8.0$ display the performance of the adjustments in the face of much noise in the error. The combination of parameters $(\phi_\beta^r, \phi_\epsilon^r)$ considered to simulate the data in each point of the $n \times p$ grid is illustrated

(Figure 1). Only additive genetic effects were simulated, and an absence of dominance was assumed.

The models were fitted with the simulated response and predictor variables. The Bayesian models were fitted using routines written in the C and R programming languages that implement the algorithms of the Gibbs sampler and random walk. This allowed the optimization of the process to obtain MCMC samples, making it faster, combining compiled and interpreted code. On the other hand, the estimation with the REML method of the variance of the linear mixed model defined in equation (3) was carried out with the *mixed.solve* function of R's rrBLUP package (Endelman, 2011).

The Gelman-Rubin diagnosis (Gelman and Rubin, 1992) verified the convergence of the MCMC chains towards their stationary distribution. For this purpose, two 100 000-sized chains were used, considering half of each chain as burning and implementing a thinning every 10 observations. If the statistical value of the Gelman-Rubin (DG) test is lower than 1.2, it is concluded that the convergence is acceptable. The summary of statistics of the posterior estimated densities was calculated with the average of the corresponding statistics given for each one of the two MCMC chains. To facilitate the discussion of results, fixed values were used to simulate the data, which were denoted by ϕ_β^r y ϕ_ϵ^r, called true parameters.

RESULTS AND DISCUSSION

Scenario $n > p$

The performance of the fittings in scenarios $n > p$ is exposed in detail using the combination of $n = 8000$ and $p = 349$ (Table 1) as an example. For other scenarios, similar results were obtained (not shown). The MCMC chains of each variance component quickly converged, except with the pair of values $\phi_\beta^r = 0.01$ and $\phi_\epsilon^r = 6.0$ and $\phi_\beta^r = 0.03$ and $\phi_\epsilon^r = 8.0$, for which chain convergence was slow. For example, when the data were generated with $\phi_\beta^r = 0.03$ and $\phi_\epsilon^r = 8.0$, $n = 8000$ and $p = 349$, 100 000 realizations of the chain under prior HC were needed to reach their stationary distribution; in contrast, under the same prior distribution and with data generated with real parameters $\phi_\beta^r = 0.01$ and $\phi_\epsilon^r = 1.0$, $n = 8000$ and $p = 349$, only 10 000 iterations of the chain were needed to reach its stationary distribution. With 100 000 observations, the statistical value of the DG test was always lower than 1.2, which indicates convergence in each one of the chains generated (Table 1). Due to this, the generation of 100 000 observations of each chain was considered so that all were of the same size and their convergences were ensured at the same time.

Under prior density $(\chi_{v,S}^{-2})$ the credibility intervals (CI) of o.95 for 0.95 for ϕ_ϵ generally contained the true values of parameters ϕ_ϵ (Table 1, Figures 2B and 2D). These results contrast with the CI obtained for ϕ_β. In the intervals built for ϕ_β four behaviors are observed, linked to the values of ϕ_β^r and ϕ_ϵ^r. First, when value ϕ_β^r is near zero $(\phi_\beta^r \leq 0.3)$ and ϕ_ϵ^r is large $(\phi_\epsilon^r \geq 6.0)$, the posterior distribution of ϕ_β is biased towards

Table 1. Summary of statistics after fitting the models with scaled inverse chi-squared ($\chi_{v,S}^{-2}$) and Half Cauchy (HC) priors, with n = 8000 and p = 349 (four chromosomes). Fixed values (ϕ_β^r, ϕ_ϵ^r) to simulate the data (Pars).

Pars		REML		Prior	Posterior ϕ_β							Posterior $\phi_\in$						
ϕ_β^r	ϕ_ϵ^r	$\hat{\phi}_\beta$	$\hat{\phi}_\epsilon$		M	Me	SD	Q0.025	Q0.975	HD	DG	M	Me	SD	Q0.025	Q0.975	HD	DG
0.01	6	0.00005	5.9361	$(\chi_{v,S}^{-2})$	0.2944	0.2908	0.0387	0.2250	0.3759	1.0000	1.02	5.9028	5.9034	0.0921	5.7254	6.0724	0.8921	1.00
				HC	0.0110	0.0090	0.0089	0.0008	0.0332	0.8207	1.03	5.9351	5.9334	0.0959	5.7532	6.1350	0.9524	1.00
0.1	6	0.00063	5.9347	$(\chi_{v,S}^{-2})$	0.3368	0.3330	0.0444	0.2610	0.4295	1.0000	1.00	5.9115	5.9077	0.0946	5.7305	6.0954	0.8887	1.00
				HC	0.0776	0.0750	0.0229	0.0407	0.1301	0.8077	1.00	5.9428	5.9472	0.0921	5.7648	6.1275	0.9521	1.00
0.03	8	0.00020	7.9649	$(\chi_{v,S}^{-2})$	0.4025	0.3985	0.0526	0.3141	0.5147	1.0000	1.02	7.9212	7.9231	0.1277	7.6787	8.1660	0.8843	1.01
				HC	0.0162	0.0104	0.0169	0.0002	0.0607	0.7958	1.01	7.9717	7.9706	0.1239	7.7264	8.2203	0.9546	1.00
0.3	8	0.00251	7.9662	$(\chi_{v,S}^{-2})$	0.6107	0.6078	0.0794	0.4635	0.7758	0.9996	1.00	7.9397	7.9389	0.1279	7.6974	8.1974	0.8880	1.00
				HC	0.2972	0.2911	0.0583	0.1972	0.4227	0.8002	1.01	7.9830	7.9838	0.1279	7.7310	8.2297	0.9538	1.00
0.05	1	0.00045	0.9965	$(\chi_{v,S}^{-2})$	0.0909	0.0904	0.0123	0.0685	0.1164	0.9987	1.02	0.9952	0.9950	0.0165	0.9627	1.0271	0.9755	1.00
				HC	0.0558	0.0553	0.0096	0.0386	0.0767	0.8655	1.01	0.9977	0.9978	0.0163	0.9676	1.0301	0.9406	1.00
0.5	1	0.00400	0.9968	$(\chi_{v,S}^{-2})$	0.5700	0.5681	0.0571	0.4737	0.6909	0.7742	1.00	0.9993	0.9993	0.0165	0.9668	1.0323	0.8797	1.01
				HC	0.5524	0.5502	0.0579	0.4553	0.6712	0.8481	1.00	0.9980	0.9973	0.0154	0.9693	1.0299	0.9430	1.01
1.0	1	0.0078	0.9975	$(\chi_{v,S}^{-2})$	1.1223	1.1180	0.1064	0.9284	1.3343	0.6950	1.01	0.9995	0.9990	0.0162	0.9667	1.0300	0.8890	1.01
				HC	1.1173	1.1108	1.1062	0.9307	1.3362	0.8508	1.01	0.9984	0.9985	0.0157	0.9682	1.0293	0.9421	1.00
0.07	6	0.00043	5.9346	$(\chi_{v,S}^{-2})$	0.3189	0.3176	0.0410	0.2506	0.4079	1.0000	1.01	5.9075	5.9087	0.0946	5.7330	6.0931	0.8901	1.00
				HC	0.0515	0.0494	0.0184	0.0200	0.0932	0.8032	1.01	5.9381	5.9357	0.0957	5.7543	6.1309	0.9507	1.01
0.7	6	0.00449	5.9362	$(\chi_{v,S}^{-2})$	0.7378	0.7302	0.0939	0.5744	0.9400	0.9935	1.00	5.9281	5.9282	0.0914	5.7574	6.0969	0.8883	1.00
				HC	0.5612	0.5553	0.0800	0.4192	0.7271	0.8170	1.00	5.9401	5.9428	0.0959	5.7568	6.1229	0.9515	1.00
0.09	8	0.00066	7.9666	$(\chi_{v,S}^{-2})$	0.4449	0.4420	0.0590	0.3409	0.5673	1.0000	1.00	7.9348	7.9338	0.1240	7.7079	8.1819	0.8824	1.01
				HC	0.0756	0.0742	0.0286	0.0246	0.1345	0.7711	1.01	7.9740	7.9782	0.1229	7.7358	8.2187	0.9545	1.00
0.9	8	0.00808	7.9612	$(\chi_{v,S}^{-2})$	1.2335	1.2274	0.1516	0.9789	1.5465	0.9815	1.00	7.9637	7.9594	0.1269	7.7252	8.2152	0.8868	1.00
				HC	1.0044	0.9865	0.1384	0.7639	1.3088	0.8177	1.00	7.9676	7.9676	0.1269	7.7242	8.2350	0.9538	1.00

REML: restricted maximum likelihood estimations; M: mean; Me: median; SD: standard deviation; Q0.025: quantile 0.025; Q0.975: quantile 0.975; HD: Hellinger distance; DG: Gelman-Rubin diagnosis.

the right of the true value ϕ_β^r , which exposes its overestimation (Figures 2A and 2C). This shows the difficulty of estimating the variance component ϕ_β when the variability of the error is large and the effects of the predictor variables are small. In scenarios where ($\phi_\beta^r \leq 0.3$) and ($\phi_\epsilon^r \geq 6.0$), the HD distances under the prior ($\chi_{v,S}^{-2}$) are near 1.0, indicating a discrepancy between the prior and posterior densities. However, Bayesian learning is poor because the posterior densities fail to capture the true values of the parameters ϕ_β^r (Figures 2A and 2C).

Second, when the variance component ϕ_β tends towards the value of 1.0 ($\phi_\beta^r \geq 0.7$), the value ϕ_ϵ^r = 6.0 (ϕ_ϵ^r = 8.0) leads parameter ϕ_β^r to be overestimated if the size of the sample is insufficient (n = 8000). Under this scenario, at least n = 8000 registers are needed for the posterior distribution of ϕ_β to center its mass around the true value ϕ_β^r. This performance was also verified when calculating the HD distance, whose value close to 1.0 indicates a complete distancing between the prior and posterior densities.

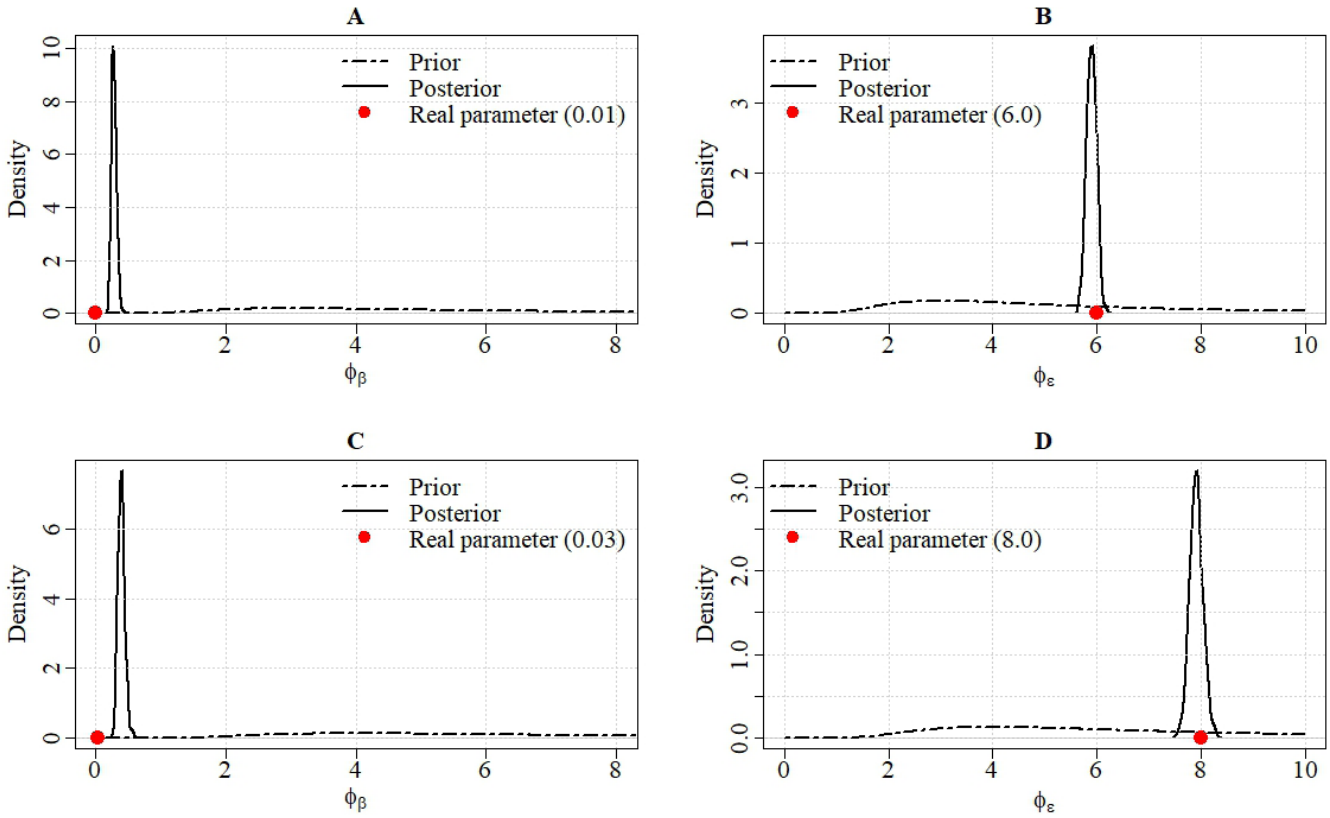

Figure 2. Posterior densities for ϕ_β and ϕ_ϵ with scaled inverse chi-squared ($\chi^{-2}_{v,S}$) prior, $n > p$ (n = 8000, p = 349) under two scenarios. A, B: (ϕ^r_β = 0.01, ϕ^r_ϵ = 6.0); C, D: (ϕ^r_β = 0.03, ϕ^r_ϵ = 8.0).

Third, when ϕ^r_ϵ = 1.0, the posterior distribution of ϕ_β was found around the true value of parameter ϕ^r_β (ϕ^r_β = 0.5, 1.0). However, this behavior does not express a substantial Bayesian learning because the prior distribution is found around the true values of the parameters (ϕ^r_β = 0.5, 1.0), which is why the HD distance only reached values below 0.8 (0.7742 and 0.6950). In this context, a relevant prior density is being used, in which case an HD distance relatively far from the value of 1.0 does not imply a lack of Bayesian learning. Finally, it was observed that although ϕ^r_β moved away from the origin (ϕ^r_β = 0.9), the high noise induced by the variance of the error (ϕ^r_ϵ = 8.0) generally causes the overestimation of parameter ϕ^r_β.

When standard HC prior distribution is used, almost all CI at a probability of 0.95 for ϕ_ϵ captured the different values assigned to parameter ϕ^r_ϵ (Table 1, Figures 3B and 3D). These results are similar to those obtained under prior ($\chi^{-2}_{v,S}$). However, for component ϕ_β, the performance of the posterior distribution contrasts with the performance of the posterior density when considering ($\chi^{-2}_{v,S}$) prior. In general, each one of the CI at 0.95 probability for ϕ_β contains the true value of the parameter (Figures 3A and

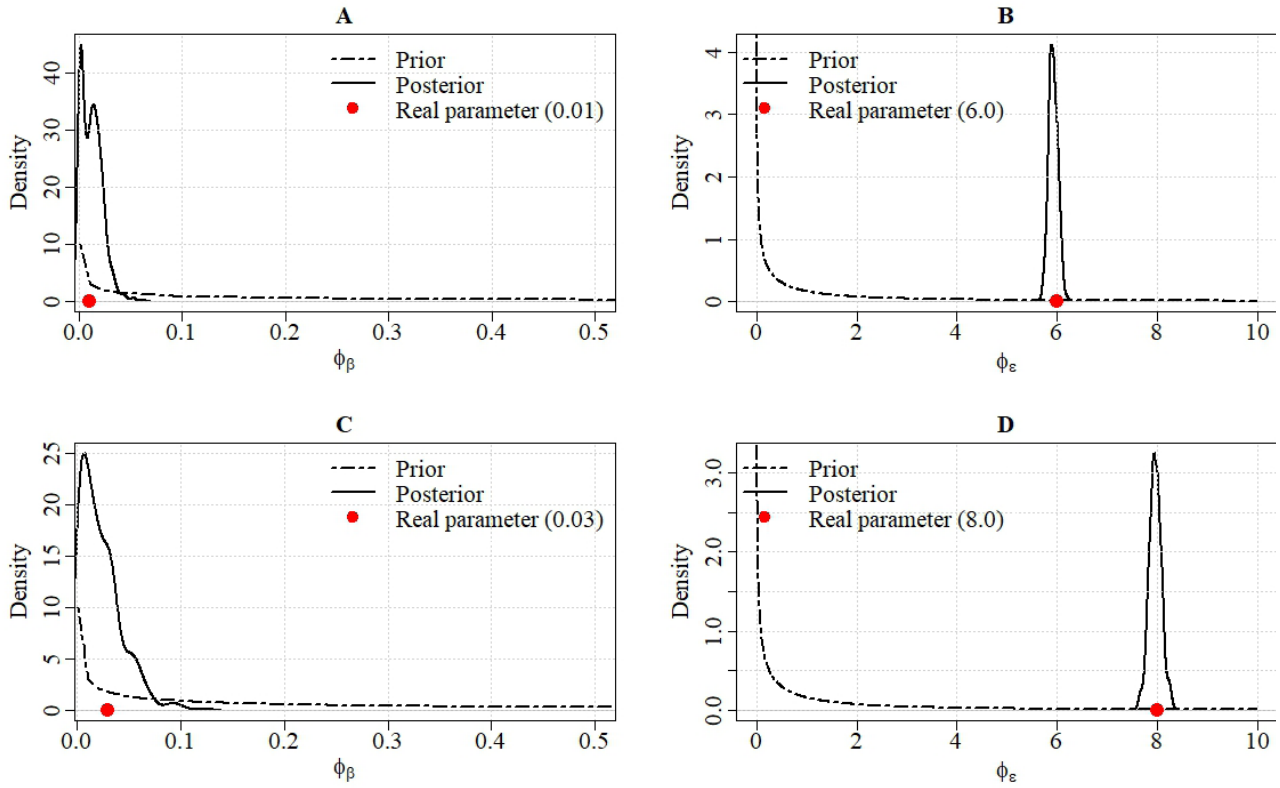

Figure 3. Posterior densities for ϕ_β and ϕ_ϵ with Half Cauchy prior (HC), $n > p$ (n = 8000, p = 349) under two scenarios. A, B: (ϕ_β^r = 0.01, ϕ_ϵ^r = 6.0); C, D: (ϕ_β^r = 0.03, ϕ_ϵ^r = 8.0).

3C), alongside the fact that in each simulation scenario, the mean of the posterior distribution is relatively near the real value of the parameter.

Regarding the effectiveness of Bayesian learning, the prior distribution for ϕ_β accumulated its greatest mass near the origin. One could suspect that this property allows the posterior density to effectively capture the true value of parameter ϕ_β^r when it is near zero. However, when the real value of parameter ϕ_β^r in the data-generating model moves away from the origin (ϕ_β^r = 1.0), the greatest discrepancy is created between prior and posterior densities (Table 1). The posterior distribution of ϕ_β concentrates its mass around the true value of the parameter (ϕ_β^r = 1.0). This behavior is also reflected in the posterior distribution of ϕ_ϵ, because when the standard HC prior distribution is assigned to the positive square root of variance component ϕ_ϵ, the prior density for ϕ_ϵ also accumulates its largest mass near zero.

This characteristic of the prior distribution for ϕ_ϵ has no influence in the learning of the posterior distribution, which concentrates around the true value of parameter ϕ_ϵ^r, regardless of it moving away from the origin (Figures 3B and 3D). Thus, the posterior density is effectively updated towards the true value of the parameter, whether ϕ_ϵ^r

= 1.0, 6.0, or 8.0, as applicable (Table 1). This behavior is also expressed in the HD distances between the prior and posterior densities, which surpass the value of 0.943, indicating an efficient and effective Bayesian learning.

The results of fitting the linear mixed model were reported in equation (3) using the REML method (Table 1). The adjustment shows the inability of the method to correctly estimate the variance component ϕ_β^r, which was underestimated in each simulation scenario. An important positive aspect of the method is its precise estimation of variance component ϕ_ϵ^r, which should be emphasized.

Scenario $n << p$

The performance of the fittings in scenarios $n << p$ is exposed in detail with $n = 200$ and $p = 349$ (Table 2). Similar results were obtained for other scenarios (not shown). The value of the DG statistic obtained in each chain diagnosis was always near the value of

Table 2. Summary of posterior statistics when fitting the models with the scaled inverse chi-squared ($\chi_{v,s}^{-2}$) and Half Cauchy (HC) priors, with $n = 200$ and $p = 349$ (four chromosomes). Fixed values (ϕ_β^r, ϕ_ϵ^r) to simulate the data (Pars).

Pars		REML		Prior	Posterior ϕ_β							Posterior ϕ_ϵ						
ϕ_β^r	ϕ_ϵ^r	$\hat{\phi}_\beta$	$\hat{\phi}_\epsilon$		M	Me	SD	Q0.025	Q0.975	HD	DG	M	Me	SD	Q0.025	Q0.975	HD	DG
0.01	6	0.2275	6.3076	$(\chi_{v,s}^{-2})$	1.9487	1.8772	0.5930	1.0405	3.3452	0.7308	1.00	5.4021	5.3246	0.6638	4.2412	6.8612	0.6395	1.00
				HC	0.4022	0.2618	0.4129	0.0021	1.4026	0.5410	1.00	6.2458	6.1744	0.7396	5.0096	7.8028	0.8620	1.00
0.1	6	<0.0001	6.2279	$(\chi_{v,s}^{-2})$	1.1901	1.1414	0.3177	0.7245	1.9346	0.8306	1.00	4.5427	4.5210	0.5110	3.6001	5.6098	0.6623	1.00
				HC	0.1403	0.0716	0.1886	0.0003	0.6513	0.6248	1.01	4.9041	4.8810	0.5195	3.9745	6.0137	0.8652	1.00
0.03	8	0.2237	7.7108	$(\chi_{v,s}^{-2})$	2.2131	2.1204	0.6317	1.2347	3.7102	0.7650	1.01	6.7496	6.6938	0.8207	5.3649	8.5626	0.6447	1.00
				HC	0.3566	0.2165	0.3828	0.0064	1.3704	0.5614	1.11	7.6705	7.6344	0.8350	6.1926	9.4141	0.8747	1.00
0.3	8	<0.0001	8.2110	$(\chi_{v,s}^{-2})$	1.8041	1.7300	0.5050	1.0513	2.9742	0.8541	1.00	7.7970	7.7091	0.8790	6.3354	9.7150	0.6680	1.00
				HC	0.1184	0.0601	0.1484	0.0006	0.5081	0.6499	1.06	8.2006	8.1574	0.8104	6.7311	10.062	0.8818	1.01
0.05	1	0.0449	0.9737	$(\chi_{v,s}^{-2})$	0.2887	0.2811	0.0795	0.1633	0.4543	0.7601	1.00	0.8601	0.8570	0.1014	0.6764	1.0677	0.6476	1.00
				HC	0.0678	0.0498	0.0669	0.0017	0.2433	0.6963	1.09	0.9726	0.9660	0.1122	0.7606	1.2089	0.8316	1.00
0.1	1	0.2430	0.8681	$(\chi_{v,s}^{-2})$	0.4012	0.3900	0.1170	0.2168	0.6854	0.6452	1.01	0.8119	0.8029	0.1092	0.6325	1.0573	0.6136	1.00
				HC	0.2268	0.2169	0.1328	0.0160	0.5136	0.6425	1.00	0.9069	0.8965	0.1365	0.6574	1.1926	0.6048	1.00
0.5	1	0.6857	0.8070	$(\chi_{v,s}^{-2})$	0.7564	0.7360	0.1946	0.4431	1.2107	0.4998	1.00	0.8186	0.8139	0.1212	0.6074	1.0710	0.6055	1.00
				HC	0.7082	0.7068	0.2111	0.3266	1.1693	0.7093	1.03	0.8308	0.8200	0.1370	0.5983	1.1174	0.7977	1.01
1.0	1	0.9288	1.3627	$(\chi_{v,s}^{-2})$	1.0629	1.0447	0.2629	0.6050	1.6496	0.5635	1.00	1.3618	1.3558	0.1886	1.0209	1.7529	0.6107	1.00
				HC	0.9401	0.9079	0.2861	0.4600	1.5646	0.7065	1.02	1.3845	1.3663	0.2024	1.0380	1.8006	0.8124	1.00
0.07	6	0.2507	5.7746	$(\chi_{v,s}^{-2})$	1.6935	1.6122	0.4811	0.9593	2.7936	0.7603	1.00	5.1111	5.0594	0.6036	4.0211	6.5185	0.6480	1.00
				HC	0.2998	0.1975	0.3135	0.0033	1.0564	0.5760	1.09	5.8007	5.7680	0.6313	4.6230	7.0731	0.8656	1.01
0.7	6	1.2383	4.7563	$(\chi_{v,s}^{-2})$	2.1189	2.0505	0.6153	1.1409	3.5246	0.6610	1.00	4.4630	4.4246	0.5587	3.4763	5.6572	0.6264	1.01
				HC	1.1893	1.0960	0.6333	0.1965	2.6282	0.5699	1.00	4.8899	4.8401	0.6927	3.6580	6.3870	0.8404	1.02
0.09	8	<0.0001	8.4050	$(\chi_{v,s}^{-2})$	2.3169	2.1960	0.7028	1.2862	4.0000	0.7581	1.00	7.2775	7.2057	0.8767	5.8100	9.0892	0.6471	1.00
				HC	0.1344	0.0306	0.2366	0.0007	0.7694	0.6410	1.03	8.4020	8.3493	0.8498	6.8836	10.243	0.8807	1.04
0.9	8	0.5396	8.5494	$(\chi_{v,s}^{-2})$	2.9830	2.8282	0.9100	1.6091	5.1945	0.6838	1.00	7.3021	7.2452	0.9541	5.5731	9.2495	0.6168	1.00
				HC	0.5606	0.3302	0.6740	0.0009	2.4129	0.5199	1.00	8.6582	8.6184	1.0601	6.6583	10.799	0.8661	1.00

REML: restricted maximum likelihood estimations; M: mean; Me: median; SD: standard deviation; Q0.025: quantile 0.025; Q0.975: quantile 0.975; HD: Hellinger distance; DG: Gelman-Rubin diagnosis.

1.0, suggesting that both chains originate from the marginal posterior distribution of each variable, regardless of the prior density considered (Table 2). On the other hand, when n = 200 and p = 349, every HD distance between the prior and posterior densities was less than its corresponding HD when n = 800 and p = 349.

This result indicates the consistency of the Bayesian learning as the model is updated, adding new available information (Tables 1 and 2). The posterior density of ϕ_ϵ registered more learning under the HC prior than under $(\chi^{-2}_{v,S})$ prior. In contrast, for ϕ_β, the learning under prior density HC, was only greater in the data generated with parameters (ϕ^r_β = 0.5, ϕ^r_ϵ = 1.0) and (ϕ^r_β = 1.0, ϕ^r_ϵ = 1.0). These greater HD distances expose the versatility of the prior HC density to efficiently learn in high dimensions, even when the true value of parameter ϕ^r_β moves away from zero, which is the region in which the prior density accumulates its greatest mass.

On the other hand, the CIs built from the posterior density for ϕ_β with prior $(\chi^{-2}_{v,S})$ in contexts of great uncertainty, which is induced by ϕ^r_ϵ = 6.0 (8.0), tended to overestimate parameter ϕ^r_β. Nevertheless, when the variance of the error uncertainty reduced (ϕ^r_ϵ = 1.0), the CI at a probability of 0.95 correctly captured true value ϕ^r_β when true parameters ϕ^r_β were relatively far from the origin (ϕ^r_β = 0.5, 1.0). This result shows the difficulty of correctly estimating parameter ϕ^r_β if its real value is near the origin, regardless of the noise induced by component ϕ^r_ϵ. Results also indicate that parameter ϕ^r_ϵ tended to be correctly estimated, regardless of its value ϕ^r_ϵ = 1.0, 6.0, 8.0.

In contrast with the CIs built under $(\chi^{-2}_{v,S})$, the CIs built under HC prior contained the real values of the parameters, regardless of the real values of parameters ϕ^r_β or ϕ^r_ϵ (Table 2, Figure 4). However, in accordance with the true parameters (ϕ^r_β, ϕ^r_ϵ) used in the data-generating mechanism, three substantial behaviors can be distinguished. First, when the variance of the error is large (ϕ^r_ϵ = 6.0, 8.0) and variance component ϕ^r_β tends towards zero, the posterior distribution of ϕ^r_β was observed to present a positive bias, taking value ϕ^r_β as a reference, which shows the influence of parameter ϕ^r_ϵ to attract the distribution of ϕ^r_β towards value ϕ^r_ϵ (Figures 4A and 4B). Second, when the variance of the error decreases (ϕ^r_ϵ = 1), the posterior distribution of ϕ_β usually centers its mass around true value ϕ^r_β; under these circumstances, the posterior means and medians of ϕ_β practically coincide with true values ϕ^r_β used to simulate the data (Table 2). Third, when parameter ϕ^r_β moves away from the origin (ϕ^r_β = 0.5, 0.7, 0.9, 1.0), the posterior distribution of ϕ_β concentrates its mass around the value ϕ^r_β, regardless of the location of the posterior distribution of ϕ_ϵ (Figure 4). Regardless of the density of the prior density assigned to the variance components and the noise induced by ϕ^r_ϵ, the CIs tended to be shorter as the sample size increased.

Regarding the estimation of the variance components according to the linear mixed model via REML, two substantial behaviors can be distinguished. In scenarios in which ϕ^r_β is relatively near zero (ϕ^r_β = 0.01, 0.03, 0.07, 0.09) and ϕ^r_ϵ is relatively large (ϕ^r_β = 6.0, 8.0), component ϕ^r_β was significantly overestimated, whereas ϕ^r_ϵ was estimated correctly (Table 2). In general terms, the total variability (ϕ^r_β + ϕ^r_ϵ) was appropriately captured. However, the method is incapable of breaking the total variability down

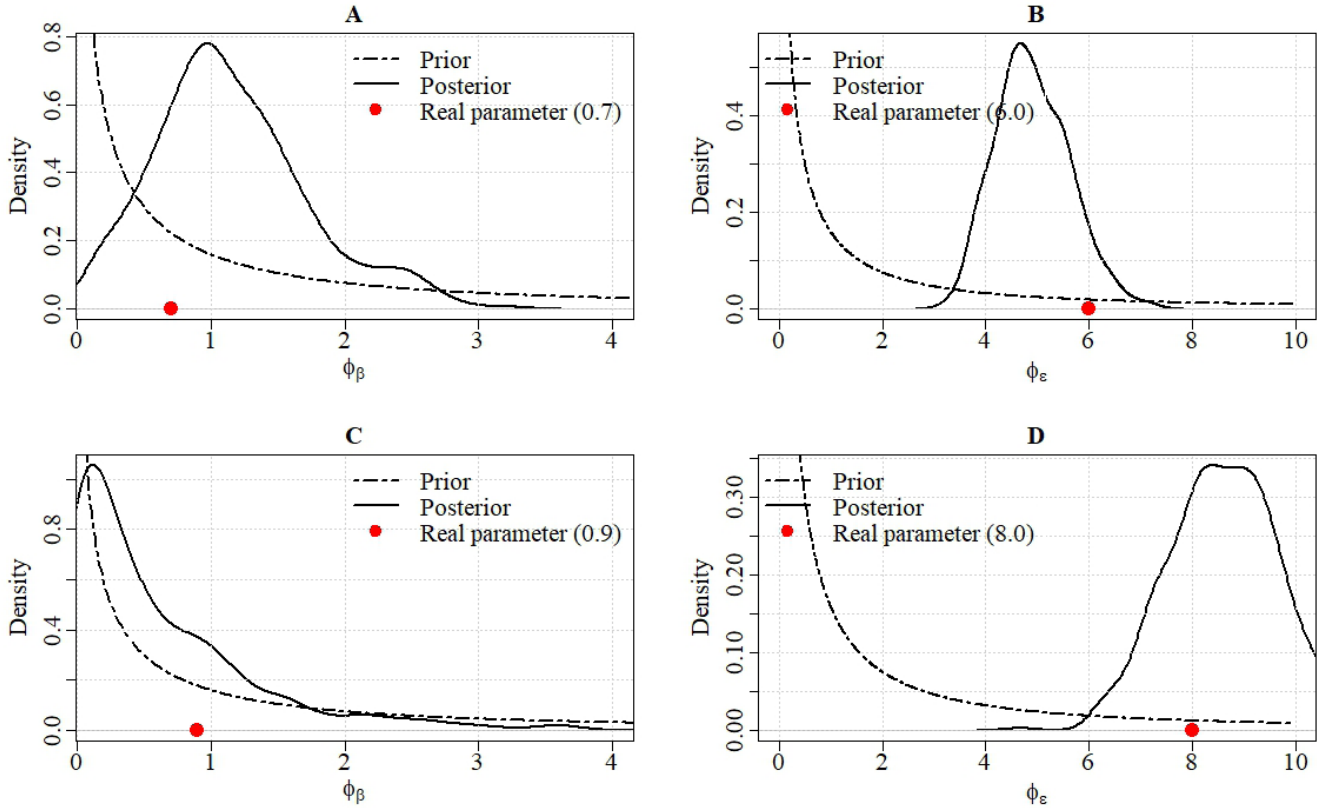

Figure 4. Posterior densities for ϕ_β and ϕ_ϵ with a Half Cauchy prior (HC), $n << p$ ($n = 200$, p = 349) under two scenarios. A, B: ($\phi_\beta^r = 0.07$, $\phi_\epsilon^r = 6.0$); C, D: ($\phi_\beta^r = 0.09$, $\phi_\epsilon^r = 8.0$).

into its individual components, ϕ_β^r and ϕ_ϵ^r. When $\phi_\epsilon^r = 1.0$, the REML method correctly estimated each variance component, ϕ_β^r and ϕ_ϵ^r, regardless of the true value ϕ_β^r used to generate the data. In addition, the results obtained with $n > p$ contrast with those obtained with $n << p$. In contexts in which $n > p$, the REML method did not detect component ϕ_β^r (Table 1), whereas when the number of records is lower than the number of parameters ($n < p$), ϕ_β^r was overestimated when $\phi_\epsilon^r = 6.0$, 8.0 and correctly estimated when $\phi_\epsilon^r = 1.0$ (Table 2).

CONCLUSIONS

The influence of $(\chi_{v,s}^{-2})$ prior and standard Half Cauchy (HC) distributions on the posterior density of each variance component in the Bayesian Ridge regression model was evaluated in contexts in which there are more observations than predictor variables and in high dimensions. Based on the results of the genetic evaluation simulation, the following conclusions can be drawn: a) The $(\chi_{v,s}^{-2})$ prior density hyperparameters

have a strong influence over the learning of the posterior density of the variance of the predictor variables, particularly when true parameter ϕ_β^r is near the value of zero and the variance of the error is large ($\phi_\epsilon^r \geq 6.0$). b) Under prior density ($\chi_{v,s}^{-2}$), when the variance of the error is large ($\phi_\epsilon^r \geq 6.0$), the CIs at 0.95 overestimate variance component ϕ_β. c) Under prior density ($\chi_{v,s}^{-2}$), the CIs for parameter ϕ_ϵ^r are almost always exact and their accuracy increases with the sample size. d) If the HC prior density is assigned to every variance component, the CIs of 0.95 for ϕ_β and ϕ_ϵ are almost always exact and their accuracy increases with the sample size. e) The problem of the curse of dimensionality does not interfere in the precision of the CIs, neither for ϕ_β^r nor for ϕ_ϵ^r. f) Regardless of the prior density considered in the parametric scenarios explored in this study, the CIs for ϕ_β^r are highly inaccurate when $n \leq 600$. The accuracy of the CIs for ϕ_β^r worsens specifically under prior density ($\chi_{v,s}^{-2}$), where the CIs are also usually inexact. However, the CIs for both ϕ_β^r and ϕ_ϵ^r tend to be shorter as the sample size increases, regardless of the prior density assigned to the variance component.

The results show the suitability of the HC distribution as a prior density to model variance components in the Bayesian Ridge regression model. Its election must be preferred over the REML method or, in Bayesian contexts, over distribution ($\chi_{v,s}^{-2}$) as the prior of the variance components.

ACKNOWLEDGEMENTS

To the National Council for Humanities, Science and Technology (CONAHCYT) for the economic support granted for the doctorate studies of the main author.

REFERENCES

Bernardo JM, Smith AFM. 2000. Bayesian theory. John Wiley and Sons: Hoboken, NJ, USA. 586 p.

Bezanson J, Edelman A, Karpinski S, Shah VB. 2017. Julia: A fresh approach to numerical computing. SIAM Review 59 (1): 65–98. https://doi.org/10.1137/141000671

Browne WJ, Draper D. 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis 1 (3): 473–514. https://doi.org/10.1214/06-ba117

Casella G, Berger, RL. 2002 Statistical Inference (Second edition). Duxbury Press: Belmont, CA, USA. 686 p.

Casella G, George EI. 1992. Explaining the Gibbs sampler. The American Statistician 46 (3): 167–174. https://doi.org/10.1080/00031305.1992.10475878

Cheng H, Garrick D, Fernando R. 2015. XSim: Simulation of descendants from ancestors with sequence data. G3: Genes|Genomes|Genetics 5 (7): 1415–1417. https://doi.org/10.1534/g3.115.016683

de los Campos G, Gianola D, Rosa GJ, Weigel KA, Crossa J. 2010. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genetics Research 92 (4): 295–308. https://doi.org/10.1017/s0016672310000285

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4 (3): 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics (Fourth edition). Longman: Harlow, UK. 448 p.

Fink D. 1997. A Compendium of conjugate priors. Montana State University: Bozeman, MT, USA. 47 p.

Gelman A, Hill J. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press: Cambridge, UK. 625 p.

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4). https://doi.org/10.1214/ss/1177011136

Gelman A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1 (3). https://doi.org/10.1214/06-ba117a

Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R. 2009. Additive genetic variability and the Bayesian alphabet. Genetics 183 (1): 347–363. https://doi.org/10.1534/genetics.109.103952

Gianola D. 2013. Priors in whole-genome regression: The Bayesian alphabet returns. Genetics 194 (3): 573–596. https://doi.org/10.1534/genetics.113.151753

Giraud C. 2015. Introduction to high-dimensional statistics. Universite Paris Saclay: Paris, France. 345 p.

Lehermeier C, Wimmer V, Albrecht T, Auinger HJ, Gianola D, Schmid VJ, Schön CC. 2013. Sensitivity to prior specification in Bayesian genome-based prediction models. Statistical Applications in Genetics and Molecular Biology 12 (3). https://doi.org/10.1515/sagmb-2012-0042

Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157 (4): 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Pérez-Rodríguez P, Acosta-Pech R, Pérez-Elizalde S, Cruz CV, Espinosa JS, Crossa J. 2018. A Bayesian genomic regression model with skew normal random errors. G3 Genes|Genomes|Genetics 8 (5): 1771–1785. https://doi.org/10.1534/g3.117.300406

Pérez-Rodríguez P, de los Campos G. 2014 Genome-wide regression and prediction with the BGLR statistical package. Genetics 198 (2): 483–495. https://doi.org/10.1534/genetics.114.164442

Polson NG, Scott JG. 2012. On the half-cauchy prior for a global scale parameter. Bayesian Analysis 7 (4). https://doi.org/10.1214/12-ba730

Roos M, Herd L. 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. Bayesian Analysis 6 (2): 259–278. https://doi.org/10.1214/11-ba609

Silverman BW. 1998. Density estimation for statistics and data analysis. Chapman and Hall: London, UK. 22 p.

Spiegelhalter DJ, Thomas A, Best NG, Gilks, WR. 1996. BUGS 0.5 examples. MRC Biostatistics Unit: Cambridge, UK. 75 p.

Agrociencia

CLASSIFICATION OF ALFALFA (*Medicago sativa* L.) PHENOLOGY USING MACHINE LEARNING METHODS

Álvaro **Murguia-Cozar**[1], Antonia **Macedo-Cruz**[1*], Demetrio Salvador **Fernández-Reynoso**[1]

[1]Colegio de Postgraduados Campus Montecillo. Carretera México-Texcoco km 36.5, Montecillo, Texcoco, State of Mexico, Mexico. C. P. 56264.
* Author for correspondence: macedoan@colpos.mx

Citation: Murguia-Cozar A, Macedo-Cruz A, Fernández-Reynoso DS. 2026. Classification of alfalfa (*Medicago sativa* L.) phenology using machine learning methods. **Agrociencia** 60(1): 42-61. https://doi.org/ 10.47163/agrociencia.v60i1.3441

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: June 02, 2025.
Approved: January 23, 2026.
Published in Agrociencia: January 30, 2026.

ABSTRACT

Alfalfa (*Medicago sativa* L.) is an important crop for food security and livestock sustainability. The accurate identification of its phenological stages, based on subjective observations, can be improved using machine learning methods that enable objective and efficient classification based on field data and remote sensors. The purpose of this study was to classify four phenological stages of alfalfa using Sentinel-2 images and machine learning models such as Support Vector Machine (SVM) and Multilayer Perceptron (MLP) neural networks. To this end, a dimensionality reduction process based on correlation analysis and Sequential Forward Selection (SFS) of features was integrated to optimize accuracy and computational efficiency. In this study, 41 Sentinel-2 images corresponding to 72 alfalfa plots during one agricultural cycle were analyzed. From the images, 86 texture, color, and vegetation index characteristics were extracted; subsequently, a correlation analysis was applied to eliminate redundant variables, reducing the set to 50 independent characteristics. On this subset, the SFS method was implemented with a gradient stopping criterion, which allowed the identification of the 29 variables with the highest discriminating power. As a result, the SVM model improved its accuracy from 70.1 to 82.2 % after the reduction of characteristics, while the MLP network achieved the highest overall accuracy (85 %, k = 0.77) with a configuration of 50-50 neurons in two hidden layers. The combination of correlation analysis and feature selection reduced dimensionality by 58 % without loss of accuracy. The MLP network outperformed the SVM in its generalization ability. This approach constitutes a low-cost operational alternative for the phenological monitoring of perennial crops using freely available satellite imagery.

Key words: Support vector machine, multilayer perceptron neural network, forward sequential selection, alfalfa monitoring.

INTRODUCTION

The world population is projected to reach 8.1 billion people by 2030, indicating a substantial rise in food demand. Therefore, it will be essential to increase the area under cultivation, improve agricultural productivity, and intensify crop rotation (FAO, 2002). In Mexico, between 2006 and 2018, the area under cultivation grew by

1.5 %, reaching a total of 20.27 million ha (SIAP, 2019). Food security, both locally and globally, will depend on the ability to increase food production.

Constant crop monitoring can optimize agricultural productivity, which in turn leads to increased production. This procedure facilitates estimations regarding yield, crop health, and requirements for water and nutrients (Bouni *et al*., 2024). However, traditional methods that rely on field studies are often costly and time-consuming. As an alternative, the use of satellite imagery offers an efficient solution for large-scale crop monitoring, providing valuable biophysical data (Thenkabail *et al*., 2012).

Satellite images record the Earth's surface through spectral signatures that quantify reflected light, allowing statistical and machine learning models to extract key information about crops. Sentinel-2 images stand out by their high resolution of 10 m in the visible spectrum and near-infrared band, rendering them an effective instrument for agricultural research. Various studies have used satellite imagery and machine learning techniques to improve agricultural productivity and predict parameters such as crop yield, water demand, and phenology. This approach has been used to estimate corn production (Kayad *et al*., 2019), classify crops (Solano-Correa *et al*., 2019), analyze rice phenology (Supriatna *et al*., 2020), and determine the peak flowering of rapeseed crops (Han *et al*., 2020).

Recent studies have expanded this framework. Sadri *et al*. (2022) developed a machine learning and remote sensing model to estimate daily irrigation requirements on farms. Similarly, de la Fuente *et al*. (2023) used Sentinel-2 satellite images and machine learning algorithms to predict grape crop yields by analyzing time series of vegetation indices, with overall accuracies greater than 90 %. Longchamps and Philpot (2023) proposed a phenological monitoring method that uses a two-dimensional space for chlorophyll and canopy water indexes. Shojaeezadeh *et al*. (2025) combined optical and radar data to estimate more than 13 phenological stages with machine learning. A study on alfalfa was able to estimate crop height from Sentinel-2 images and machine learning techniques (Random Forest and XGBoost), with a coefficient of determination of 0.79 and a mean absolute error of 4 cm (Bahrami *et al*., 2025).

Crop monitoring using satellite imagery allows for estimating yield, water demand, and agricultural productivity. In recent years, various studies have been conducted to identify crops established in agricultural regions through the analysis of satellite images. Ashourloo *et al*. (2018) proposed a model that establishes a threshold using the relationship between the red, green, blue, and near-infrared spectral bands of Landsat-8 images to identify alfalfa and other crops. Li *et al*. (2021) implemented a classification method that combines generative adversarial networks, convolutional neural networks, and short- and long-term memory neural networks on Tier-1 Landsat-8 images to discriminate between corn and soybean crops.

Minallah *et al*. (2020) merged Planetscope and Sentinel-2 images to classify five types of land cover (wheat, tobacco, other vegetation, water bodies, and urban areas) by training a convolutional neural network. However, the phenological classification of crops from satellite images is a largely unexplored field, especially in the case of alfalfa.

Fatemeh and Hossein (2022) applied a combination of optical satellite data and radar images to discriminate crops such as alfalfa, wheat, sugar beet, apples, and grapes by training machine learning algorithms using the Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Enhanced Vegetation Index (EVI), and Leaf Area Index (LAI). Similarly, Chen and Zhang (2023) conducted related studies by integrating Landsat 7/8, Sentinel-2, and Sentinel-1 radar images to monitor the growth of alfalfa, corn, and soybean crops.

Phenological observations in agriculture and horticulture furnish farmers with basic information for informed decision-making concerning optimal operational programs, including planting, fertilization, irrigation, crop protection, and forecasting phenological phases (Chmielewski, 2023). Phenological data can be used in the design of models to predict biophysical parameters of crops such as grain yield, harvest, fertilization, and pesticide application. In recent years, various research projects have focused on identifying phenological stages in crops such as rice, wheat, apple, and olive (Milicevic *et al*., 2020; Velumani *et al*., 2020; Yang *et al*., 2020), using high-resolution images taken with unmanned aerial vehicles (drones), whose limitation is their high cost.

Selecting an optimal subset of spectral, color, and texture features reduces dimensionality and improves computational efficiency without compromising the accuracy of machine learning models in identifying crop phenological stages. Therefore, the objective of this study was to classify four phenological stages of alfalfa using Sentinel-2 images and machine learning models (SVM and MLP), integrating a dimensionality reduction process based on correlation analysis and sequential feature selection to optimize accuracy and computational efficiency. The main contribution of this research lies in the application of supervised classification models on Sentinel-2 satellite images to identify the phenological stages of alfalfa, which constitutes an efficient and low-cost tool for large-scale crop monitoring.

MATERIALS AND METHODS

Study area

This study was conducted on agricultural plots belonging to Irrigation Module Five "Usuarios y Productores Unidad Tepatepec, A.C." in Irrigation District 003 "Tula," in the southwestern part of the state of Hidalgo, Mexico (Figure 1). The area is located within the Mezquital Valley region (20° 14′ 42.29" N and 99° 5′ 24.45" W), taking the town of "Tepatepec" as a reference point. Irrigation Module Five is supplied by the Alto Requena, Principal Requena, Endhó, and Dren irrigation canals, which carry wastewater from the Mexico Valley Basin, the Tula River, and the Salado River. During one agricultural cycle, 72 alfalfa plots in the Tepatepec Irrigation Module were monitored to determine the phenological stages of the crop throughout the 2019 agricultural cycle.

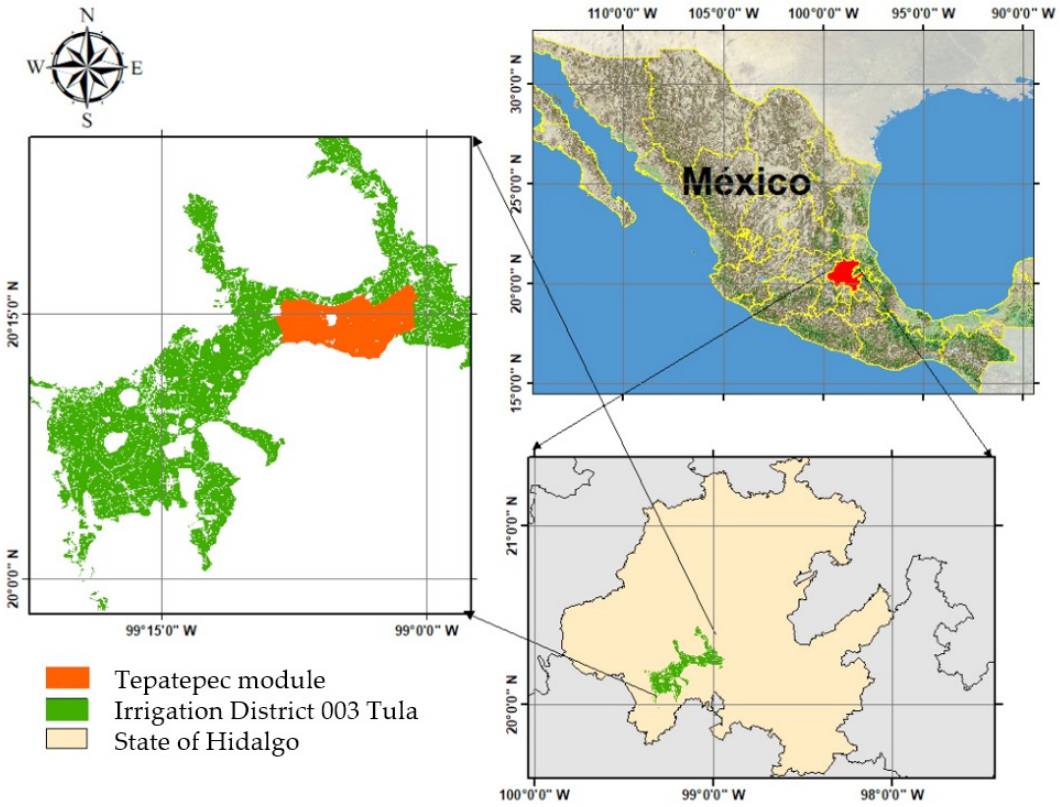

Figure 1. Location of the study area in Irrigation Module Five, Mezquital Valley, in the state of Hidalgo, Mexico.

Remote sensor data acquisition

To monitor crop phenology, Sentinel-2 satellite images of the study area were used, captured by a high-resolution multispectral camera (available at https://browser.dataspace.copernicus.eu). The images were downloaded in spectral bands 02 (blue), 03 (green), 04 (red), and 08 (near infrared), corresponding to the period from March to November 2019. Images with cloud cover of less than 50 % were selected. The images were processed using the QGIS program (QGIS Development Team; Essen, Germany) to form four-band spectral composites with a spatial resolution of 10 m and a temporal resolution of 5 d. A total of 41 spectral composites were generated.

Sample extraction

The samples used to train the machine learning models were extracted from the spectral compositions by cropping the pixels corresponding to the plots studied. These were stored in (.mat) files and labeled with the date, plot number, and phenological stage. Four main stages were identified (regrowth, medium development, maximum development, and cutting and drying), each with an approximate interval of 20 to 30 d (Figure 2). Regrowth occurs a few days after harvesting, followed by medium

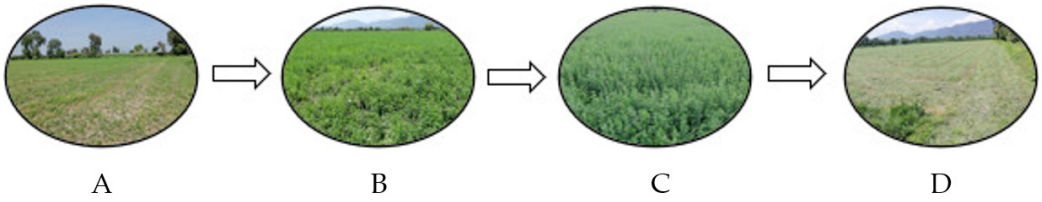

Figure 2. Monitoring of control plots. A: regrowth stage; B: average growth; C: full growth; D: cutting and drying.

development, characterized by intermediate stem growth. Maximum development corresponds to the greatest growth of the crop, with a leafy structure and the presence of some flowers. Finally, in the cutting and drying stage, the crop is cut and left on the ground until it reaches a moisture content of 14 to 18 % for baling.

The sample extraction process yielded 2405 plot clippings of different sizes, ranging from 6 to 43 pixels per side, corresponding to the control plots studied. Of these, 676 belong to the regrowth stage, 490 to the medium development stage, 1099 to the maximum development stage, and 140 to the cutting and drying stage.

Calculation and extraction of features

From the images, 86 texture, color, and vegetation features were extracted and programmed in Matlab version 2017 (The MathWorks Inc.; Natick, MA, USA). The feature extraction methods implemented were Moran's Local Spatial Autocorrelation Index (LISA), Local Binary Pattern (LBP), and Leaf Area Index (LAI). The extracted features were used to create a data matrix with a total of 2405 samples, corresponding to the different phenological stages (Table 1).

Moran's Local Index of Spatial Autocorrelation (LISA)

This index measures the spatial association between a data point and its neighbors (Anselin, 1995) and is suitable for spatial data such as satellite images. The method analyzes all pixels in the samples using 3 × 3 sliding windows and calculates the LISA value (I_i) for each pixel using a weight matrix that quantifies the contribution of its neighbors according to the following formula:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j\neq i}^{n} w_{ij}\left(x_j - \bar{X}\right)$$

where x_i is the pixel analyzed, $\bar{X}$ is the global mean of the pixels contained in the sample, S_i^2 is the variance of the sample, x_j are the neighboring pixels, w_{ij} is the square matrix of 8 × 8 pixels with weight equal to one divided by the number of neighboring pixels, and n is the total number of pixels in the sample

Table 1. Set of texture, color, and vegetation features extracted from Sentinel-2 satellite images of the study area.

Type	Indicator	Features	Number of Features
Texture	LISA	lisa_rm, lisa_rv, lisa_gm, lisa_gv, lisa_bm, lisa_bv, lisa_nirm, lisa_nirv, lmorl*_m, lmorl*_v, lmora*_m, lmora*_v, lmorb*_m, lmorb*_v, lmory*_m, lmory*_v, lmori*_m, lmori*_v, lmorq*_m, lmorq*_v, lmorh*_m, lmorh*_v, lmors*_m, lmors*_v, lmorv*_m, lmorv*_v,	26
Texture	LBP	lbp_rm, lbp_rv, lbp_gm, lbp_gv, lbp_bm, lbp_bv, lbp_nirm, lbp_nirv, lbpl*_m, lbpl*_v, lbpa*_m, lbpa*_v, lbpb*_m, lbpb*_v, lbpy*_m, lbpy*_v, lbpi*_m, lbpi*_v, lbpq*_m, lbpq*_v, lbph*_m, lbph*_v, lbps*_m, lbps*_v, lbpv*_m, lbpv*_v	26
Color	RGB and NIR	red_m, red_v, green_m, green_v, blue_m, blue_v, nir_m, nir_v	8
Color	L*a*b*	l*_m, l*_v, a*_m, a*_v, b*_m, b*_v	6
Color	YIQ	y*_m, y*_v, i*_m, i*_v, q*_m, q*_v,	6
Color	HSV	h*_m, h*_v, s*_m, s*_v, v*_m, v*_v	6
Vegetation	Color ratios	Ratio_rm, Ratio_rv, Ratio_gm, Ratio_gv, Ratio_bm, Ration_bv	6
Vegetation	LAI	lai_m, lai_v	2

m: mean of the pixels in the region of interest; v: variance; LISA: Local Spatial Autocorrelation Index; LBP: Local Binary Pattern; RGB: Red, Green, Blue; NIR: Near Infrared; L*a*b*: lightness (L*), green-red channel (a*), and blue-yellow channel (b*); YIQ: luminance (Y), in-phase component (I), and quadrature component (Q); HSV: hue, saturation, and value; LAI: Leaf Area Index.

Local binary pattern (LBP)

The LBP indicator extracts texture features from images (Ojala *et al.*, 2000) and is calculated as follows:

$$LBP = \sum_{P=1}^{P=8} s(g_P - g_c)2^P, \quad s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

where $P = 8$ is the number of neighbors analyzed, g_p and g_c are the values of the central and neighboring pixels, respectively.

Leaf area index (LAI)

LAI quantifies the leaf area per unit of surface area developed by the crop (Bastiaanssen, 1998). The indicator uses the soil-adjusted vegetation index (SAVI) as an explanatory variable. The model for estimating the index is calculated using the following expression:

$$LAI = -\frac{ln\left(\frac{0.69 - SAVI}{0.59}\right)}{0.91}$$

where $SAVI = \frac{(1+L)(NIR-RED)}{L+NIR+RED}$ and indicates the ground-adjusted vegetation index with L = 0.5, which is the factor that describes vegetation density, ranging from zero for areas with high vegetation to one for areas with sparse vegetation; *NIR* is the near-infrared spectral band, and *RED* is the red spectral band.

Color indices

Color models, or color spaces, are methodologies for representing color. The RGB (Red, Green, Blue) model is the most widely used for representing digital images. Color spaces can be projected onto each other using transformation models based on the RGB color space (Chaki and Dey, 2021). In this research, conversions were performed between the HSV (hue, saturation, value) models, the Commission on Illumination (CIE) L*a*b* scale (Robertson, 1976), and YIQ (Luminance (Y), In-phase Quadrature), which separates color from brightness.

Color ratios

The color ratios (r) are vegetation indicators that relate the spectral bands of the RGB model. This type of indicator has been used in land cover classification (Appice and Malerba, 2019), the identification of leaf diseases in alfalfa crops (Qin *et al*., 2016), and the estimation of vegetation cover (García-Martínez *et al*., 2020). They are calculated as follows:

$$r = \frac{R}{R+G+B};\ \ g = \frac{G}{R+G+B};\ \ b = \frac{B}{R+G+B}$$

A feature extraction algorithm was designed that, in the first instance, applies the programmed indicator processes to the samples, considers the plot as the region of interest, and determines the mean and variance of the pixels contained within the region of interest. Finally, it stores the information in a data matrix with 2405 rows (samples) and 86 columns (extracted features).

Correlation analysis of features

Although neural networks can handle correlated inputs, several studies indicate that redundancy between input variables can generate duplicate information, increase model complexity, and even affect its stability or computational efficiency (Zhang *et al*., 2018; Lagari *et al*., 2021; Chan *et al*., 2022). Therefore, in this study, a correlation analysis was performed between the extracted features to reduce redundancies before training. The correlation matrix was obtained, and variables with correlations greater than 0.8 were eliminated, retaining only one per group. This process reduced the set from 86 to 50 features (Table 2).

Table 2. Selected features without high correlation (final set for analysis).

Type	Indicator	Features	Number of features
Texture	LISA	lisa_rm, lisa_rv, lisa_gm, lisa_gv, lisa_bm, lisa_nirm, lisa_nirv, lmora*_m, lmora*_v, lmorb*_m, lmorb*_v, lmori*_m, lmori*_v, lmorh*_m, lmors*_m, lmors*_v	16
Texture	LBP	lbp_rm, lbp_rv, lbp_gm, lbp_gv, lpb_bv, lbp_nirm, lbp_nirv, lbpa*_m, lbpa*_v, lbpb*_m, lbpb*_v, lbpi*_m, lbpi*_v, lbph*_m, lbph*_v, lbps*_m, lbps*_v,	17
Color	RGB and NIR	red_m, red_v, nir_m, nir_v	4
Color	L*a*b*	a*_m, a*_v, b*_m, b*_v	4
Color	HSV	h*_m, h*_v, s*_m, s*_v	4
Vegetation	Color ratios	Ratio_rv, Ratio_bm, Ration_bv	3
Vegetation	LAI	lai_m, lai_v	2

m: mean of the pixels in the region of interest; v: variance; LISA: Local Indicators of Spatial Association; LBP: Local Binary Pattern; RGB: Red, Green, Blue; NIR: Near Infrared; L*a*b*: lightness (L*), green-red channel (a*) and blue-yellow channel (b*); HSV: hue, saturation, and value; LAI: Leaf Area Index.

Feature selection

The Stepwise Forward Selection (SFS) algorithm was used to identify the characteristics with the greatest discriminatory power, reducing the set from 50 to 29 variables. This method uses Fisher's criterion, which maximizes the variance between classes and within each class. Its application was intended to optimize the performance of the classifiers, reducing the number of features without compromising their performance. To evaluate the separability of the features, the following equations were used (Mery, 2015):

$$C_b = \sum_k p_k(\bar{z}_k - \bar{z})(\bar{z}_k - \bar{z})^T$$

$$C_w = \sum_{k=1}^{k} p_k C_k$$

$$C_k = \frac{1}{N_k - 1}\sum_{j=1}^{N_k}\left(z_{kj} - \bar{z}_k\right)\left(z_{kj} - \bar{z}_k\right)^T$$

$$J = spur(C_w^{-1}C_b)$$

where C_b is the intraclass variability, p_k is the prior probability of the kth class, $\bar{z}_k$ and $\bar{z}$ are the means of the kth class and overall, C_w is the interclass variability, z_{kj} is the

*j*th vector of selected features of the kth class, and N_k is the number of samples in the kth class. Fisher's criterion *(J)* determines which feature is the best; the higher *J* is, the greater the separability.

The algorithm used is a modification of the one proposed by Mery (2011), concluding the selection process upon reaching a predetermined number of features. The proposed algorithm uses $J_i - J_{i-1} \geq 0.01$ as the stopping criterion. When a feature increases the value of *J* above 0.01, the algorithm adds it to the subset and continues searching for more. Otherwise, it terminates execution and returns the subset of selected features as the result. The results of the SFS model selected 29 of the 50 features (Table 3).

Table 3. Set of features selected using the Sequential Forward Selection (SFS) algorithm for classification.

Type	Indicator	Features	Number of features
Texture	LISA	lisa_bm, lisa_nirm, lisa_nirv, lmorb*_v, lmors*_v	5
Texture	LBP	lbp_gm, lbps*_m, lbpi*_v, lbpb*_m, lbp_nirm, lbp_rv, lbpi*_m, lbp_rm	8
Color	RGB and NIR	red_m, red_v, nir_m, nir_v	4
Color	L*a*b*	a*_m, a*_v, b*_m, b*_v,	4
Color	HSV	h*_m, h*_v, s_m, s_v	4
Vegetation	Color ratios	Ratio_rv, Ratio_bm	2
Vegetation	LAI	lai_m, lai_v	2

m: mean of the pixels in the region of interest; v: variance; LISA: Local Indicators of Spatial Association; LBP: Local Binary Pattern; RGB: Red, Green, Blue; NIR: Near Infrared; L*a*b*: lightness (L*), green-red channel (a*) and blue-yellow channel (b*); HSV: hue, saturation, and value; LAI: Leaf Area Index.

Support Vector Machine (SVM) classification model

Two SVM models with quadratic kernel functions (quadratic SVM) were trained, one with the initial 86 features and the other with the 29 selected by SFS. The models were evaluated by cross-validation with 10 partitions, and the training, validation, and test accuracies, as well as the accuracies per class, were calculated. The overall and class-specific accuracies were determined using the following expression:

$$PG = \frac{TP}{TP + FP}; \quad PC_i = \frac{TP_i}{TP_i + FP_i}$$

where *PG* is the overall accuracy, *TP* are the true positives (corresponding to correctly predicted samples), *FP* are the false positives (incorrectly predicted samples), *PCi* is the accuracy of class *i*, *TPi* is the true positives for class *i* (correctly predicted samples of a defined class), and *FPi* is the false positives for class i (incorrectly predicted samples of a defined class).

Multilayer Perceptron Neural Networks (MLP)

Ten MLPs with different topological configurations (from 5 × 5 to 50 × 50 neurons in hidden layers) were trained and evaluated using the *patternnet* tool in Matlab 2017. A total of 2405 samples with the 29 selected features were used: 676 for regrowth, 490 for average development, 1099 for maximum development, and 140 for cutting and drying. The total set was divided into three subsets: training (70 %), validation (15 %), and testing (15 %). The training and validation stages allowed the classifier to learn to discriminate between classes, while the testing stage evaluated its performance with previously unused data.

The configuration of the networks studied included the use of the *trainscg* (Scaled Conjugate Gradient, SCG) error correction model, with the *crossentropy* function to evaluate performance as a measure of discrepancy between the predicted output and the actual tags. A total of 1000 training epochs were established, with a limit of 100 failures for verification, a minimum gradient of 1×10^{-7}, a learning factor of 0.1, a decay factor of 0.1, and a growth factor of 10.

The performance of the topological structures was assessed, and the one with the highest accuracy was selected. The performance evaluation of the networks was carried out by calculating the training, validation, test, and overall accuracy of each network, as well as the Kappa coefficient (k):

$$k = \frac{(P_0 - P_e)}{(1 - P_e)}$$

where P_0 is the observed probability of agreements and P_e is the expected probability of agreements by chance.

Finally, two MLP neural networks were trained with the chosen topological configuration. The first network was trained with the 86 features initially extracted, and the second with the 29 selected features.

RESULTS AND DISCUSSION

Feature selection using the SFS model

The evolution of the performance criterion during sequential forward selection showed a sustained increase in Fisher's value (J) as new variables were added, demonstrating a progressive improvement in the model's discriminatory capacity (Figure 3). The growth was more pronounced in the first iterations (variables 1 to 10), where relevant information was incorporated more quickly; from variable 20 onward, the curve tended to stabilize, indicating increasingly smaller contributions. This trend confirms the effectiveness of the method in prioritizing the most influential variables and supports the use of the stopping threshold based on the J gradient, which avoids integrating redundant characteristics.

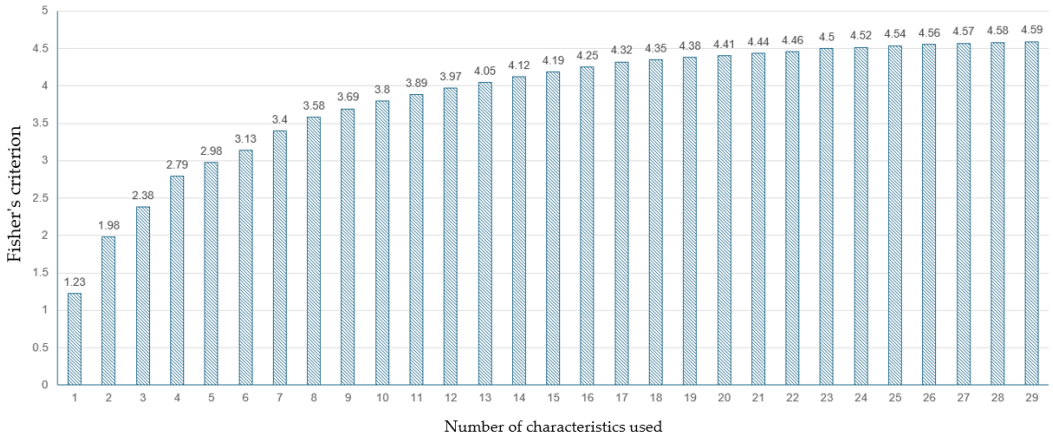

Figure 3. Evolution of Fisher's criterion during the execution of the Sequential Forward Selection (SFS) model.

Compared to other approaches, floating algorithms (such as Sequential Floating Forward Selection, SFFS) allow previous combinations to be reviewed and avoid the nesting problem of classic SFS, which usually produces subsets with greater discriminatory power (Nakariyakul and Casasent, 2009). Filter methods, such as Minimum Redundancy Maximum Relevance (mRMR) and RELevance In Estimating Features (RELIEF), select relevant and non-redundant variables using statistical criteria without requiring model training, making them efficient in high dimensionality (Pudjihartono *et al*., 2022). Embedded methods, such as Least Absolute Shrinkage and Selection Operator (LASSO) and Support Vector Machine Recursive Feature Elimination (SVM-RFE), integrate selection into the training process and reduce model complexity through regularization or recursive variable elimination (Pudjihartono *et al*., 2022).

The proposed approach maintains the interpretative simplicity of SFS but integrates an adaptive criterion that halts the process before performance stabilization, achieving a balance between precision and parsimony. In contexts with abundant spectral variables and multispectral indices, where collinearity is frequent, this gradient-based control allows for the construction of more compact and stable models, reducing the risk of overfitting and the computational cost of training.

Although SFS presents the nesting problem, where variables selected at the beginning cannot be discarded later, this effect can be mitigated by performing a preliminary correlation analysis. By eliminating highly correlated features, initial redundancy is reduced and the probability of the algorithm incorporating features with duplicate information is decreased. Thus, even though SFS maintains its nested structure, the negative impact of this limitation is mitigated, and a more stable and representative selection of the data set is obtained.

Results of the quadratic SVM model with 86 features

The quadratic SVM model had an overall accuracy of 76.1 % during the training phase. The confusion matrix shows that the phenological stages of regrowth and maximum development were classified correctly, while the stages of medium development and cutting and drying had lower accuracy levels, with accuracies of 54 and 44 %, respectively. Even with these variations between classes, the overall performance of the model is considered acceptable, especially when compared to results reported in similar studies on rice phenology (Han *et al*., 2020; Yang *et al*., 2020). In the testing phase, the model achieved an overall accuracy of 70.1 %, maintaining a distribution of correct predictions per class similar to that observed during training, which demonstrates a moderate capacity for generalization (Figure 4A).

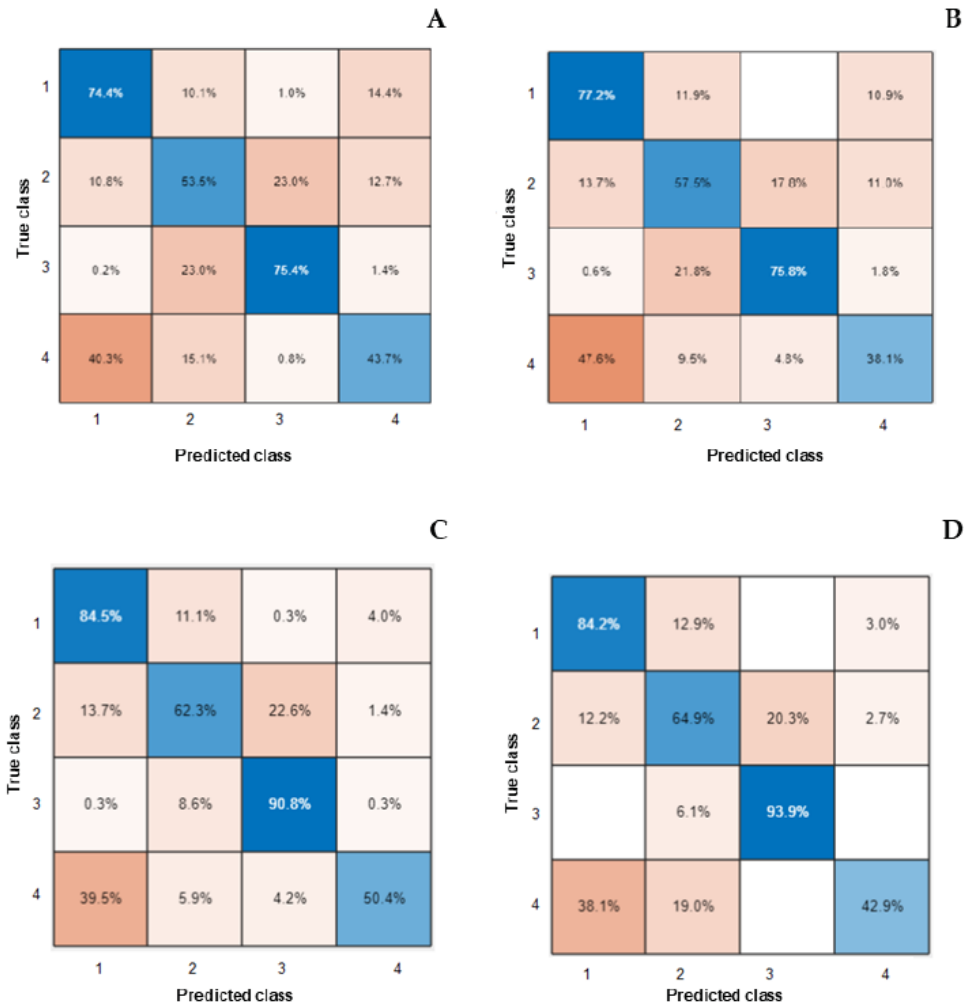

Figure 4. Confusion matrices of the Support Vector Machine (SVM) model. A: trained with all 86 features; B: tested with all 86 features; C: trained with 29 features; D: tested with 29 features.

Results of the quadratic SVM model with 29 features

The SVM model trained with the 29 selected features achieved an overall accuracy of 81 % in the training phase and 82.2 % during the testing phase. Reducing the feature space from 86 to 29 variables not only decreased the dimensionality of the model but also improved performance, increasing accuracy by 4 % for training and 10 % for testing compared to the original model. Accuracy by class also showed a significant improvement, especially in classes 1 and 3, where discrimination levels exceeded 80 %. However, difficulties persist in the classification of classes 2 and 4, suggesting that feature reduction favored model generalization but did not completely resolve the imbalance and spectral similarity associated with these classes.

Unlike annual crops such as rice or wheat, alfalfa has a cyclical phenological pattern, with continuous regrowth and more gradual structural changes between stages. This characteristic reduces spectral separability between intermediate phases, especially between medium and maximum development, which partially explains the confusion observed in classes 2 and 4. Therefore, the performance obtained reflects not only the accuracy of the model but also the phenological complexity of the crop analyzed.

The use of SFS with gradient-based stopping criteria not only reduced the dimensionality of the dataset but also allowed for the selection of features with greater discriminating power without sacrificing accuracy. This result coincides with recent studies reporting that variable reduction using supervised methods improves model stability and generalization capacity, especially in contexts with high spectral redundancy (Pudjihartono *et al*., 2022; Chan *et al*., 2022). In contrast to purely statistical filter methods, sequential selection allowed interactions between features relevant to phenological differentiation to be captured, which is reflected in the superior performance of the SVM model after reduction.

Evaluation of topological arrangements

Ten topological configurations of MLP neural networks were evaluated using the 29 selected features. The network with 50 neurons in each of the two hidden layers achieved the best overall performance, recording the highest accuracy values during the validation (86 %) and testing (85 %) stages. This structure demonstrated adequate generalization capacity and stability between the training and validation sets and was therefore considered the optimal configuration for the proposed model (Table 4).

MLP network with 86 features

The MLP network, trained with 86 original features and the chosen topological structure, exhibited a training accuracy of 85 %, a validation accuracy of 80 %, and a testing accuracy of 81 %, as indicated by the confusion matrices (Figure 5). The accuracy per class for labels 2 and 4 are low compared to labels 1 and 3, which can be attributed to the imbalance in the number of samples between categories, as these classes are significantly underrepresented. This imbalance directly affects the training process of the MLP model, causing minority classes to contribute less to the loss function and

Table 4. Evaluation of topological arrangements of the Multilayer Perceptron Neural Network (MLP).

Topological arrangement*	Training	Validity	Test	General	Kappa	Time (s)
5_5	79.20	78.12	78.12	78.88	0.67	9.24
10_10	83.78	85.04	82.55	83.78	0.75	19.33
15_15	81.52	81.44	79.78	81.25	0.71	16.18
20_20	83.01	80.89	81.44	82.45	0.73	14.12
25_25	83.84	78.12	81.99	82.70	0.74	12.60
30_30	82.23	83.10	81.16	82.20	0.73	12.03
35_35	85.32	80.61	80.61	83.91	0.76	17.36
40_40	81.94	83.66	78.12	81.62	0.72	11.97
45_45	86.04	80.89	79.22	84.24	0.76	24.73
50_50	84.79	86.15	85.04	85.03	0.77	20.49

*The topological arrangement refers to the size of the two hidden layers in the MLP network.

be underrepresented during weight optimization (He and García, 2009; Johnson and Khoshgoftaar, 2019).

Several studies have shown that monitored learning models tend to be biased toward majority classes, which reduces sensitivity in less frequent classes and generates higher confusion errors (Guo *et al*., 2017; Branco *et al*., 2016). In this case, the lower accuracy obtained for classes 2 and 4 does not reflect a deficiency in the network structure but rather an expected consequence of data imbalance, suggesting the advisability of applying class balancing techniques, such as subsampling or loss weighting, to improve the model's discriminative capacity in future adjustment stages.

MLP network with 29 features

The MLP network trained with 29 features achieved an overall accuracy of 85 % and a Kappa coefficient of 77 %. Consistent performance was observed in all phases, with high values on the main diagonal, indicating correct classification of most samples (Figure 6). The most frequent classes have accuracy rates above 85 %, while confusion errors are mainly concentrated between classes 2 and 4, reflecting a certain spectral or structural similarity between them. In the test matrix, the model maintains stable performance with an average accuracy of over 80 %, demonstrating adequate generalization capacity and the absence of significant overfitting. Overall, the consolidated matrix shows an average accuracy of 86 %, with an acceptable balance between sensitivity and specificity per class, confirming the robustness of the model and the relevance of the selected features.

Although the SVM model performed adequately, the MLP outperformed it in terms of accuracy in all evaluation phases, suggesting that the relationship between

Figure 5. Confusion matrices of the Multilayer Perceptron (MLP) Neural Network for each stage of development (86 × 50 × 50 × 4).

the extracted characteristics and phenological classes is not strictly linear. Neural networks are capable of modeling nonlinear decision boundaries and highly complex relationships between texture, color, and vegetation indices, which explains their superior ability to discriminate stages with gradual transitions (Guo *et al*., 2017; Johnson and Khoshgoftaar, 2019). This result is consistent with studies in which neural networks have outperformed maximum margin algorithms in phenological classification based on satellite images (Solano-Correa *et al*., 2019; Shojaeezadeh *et al*., 2025).

Overall, the results confirm that the combination of Sentinel-2 multispectral images, a monitored feature selection scheme, and an optimized neural architecture allows for

Figure 6. Confusion matrices of the Multilayer Perceptron (MLP) Neural Network for each stage of development (29 × 50 × 50 × 4).

highly accurate classification of the phenology of a perennial, multi-component crop such as alfalfa. The contribution lies not only in the final performance of the model but also in demonstrating that it is possible to reduce the dimensionality of the variables by 58 % without significant loss of accuracy, providing empirical evidence in favor of more compact, explainable, and computationally efficient models for agricultural remote sensing.

Compared to studies that have applied machine learning to annual crops such as rice, wheat, or corn, this study differs in three relevant methodological aspects. First, it addresses a perennial crop with multiple harvests, which implies a more continuous phenological dynamic and less spectral separability than the annual cycles studied by

Han *et al*. (2020) or Yang *et al*. (2020). Second, while most studies use only vegetation indices or spectral variables, this study incorporates combined characteristics of texture, color, and indices, which improved the model's discriminating capacity after variable selection. Finally, unlike research that exclusively applies SVM or RF, here models with different mathematical natures (maximum margins vs. neural networks) are compared, demonstrating the superiority of MLP in scenarios of gradual phenological transition. These differences allow the results to be interpreted not only in numerical terms, but also within a broader methodological comparative framework.

CONCLUSIONS

Multilayer Perceptron (MLP) neural networks proved to be the most efficient model for the phenological classification of alfalfa based on Sentinel-2 images. Variable selection contributes to improving the stability, interpretability, and efficiency of machine learning models in scenarios with high spectral redundancy. The proposed approach represents a low-cost operational alternative for phenological monitoring of perennial crops and lays the foundation for the generation of early warning systems and estimation of biophysical variables from satellite data. However, the differentiated performance between classes suggests that the incorporation of data balancing strategies or the use of complementary sensors could improve the discrimination of phenological stages with low spectral separability.

REFERENCES

Anselin L. 1995. Local indicators of spatial association—LISA. Geographical Analysis 27 (2): 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Appice A, Malerba D. 2019. Segmentation-aided classification of hyperspectral data using spatial dependency of spectral bands. ISPRS Journal of Photogrammetry and Remote Sensing 147: 215–231. https://doi.org/10.1016/j.isprsjprs.2018.11.023

Ashourloo D, Shahrabi HS, Azadbakht M, Aghighi H, Matkan A, Radiom S. 2018. A novel automatic method for alfalfa mapping using time series of Landsat-8 OLI data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (11): 4478–4487. https://doi.org/10.1109/jstars.2018.2874726

Bahrami H, Chokmani K, Homayouni S, Adamchuk VI, Albasha R, Saifuzzaman M, Leduc M. 2025. Machine learning-based alfalfa height estimation using Sentinel-2 multispectral imagery. Remote Sensing 17 (10): 1759. https://doi.org/10.3390/rs17101759

Bastiaanssen W. 1998. Remote sensing in water resources management: The state of the art. International Water Management Institute: Colombo, Sri Lanka. 118 p.

Bouni M, Hssina B, Douzi K, Douzi S. 2024. Integrated IoT approaches for crop recommendation and yield prediction using machine learning. IoT 5 (4): 634–649. https://doi.org/10.3390/iot5040028

Branco P, Torgo L, Ribeiro RP. 2016. A survey of predictive modeling under imbalanced distributions. ACM Computing Surveys 49 (2): 31. https://doi.org/10.1145/2907070

Chaki J, Dey N. 2021. Image color feature extraction techniques: Fundamentals and applications. Springer: Warsaw, Poland. 83 p. https://doi.org/10.1007/978-981-15-5761-3

Chan JYL, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong ZW, Chen YL. 2022. Mitigating the multicollinearity problem and its machine learning approach: A review. Mathematics 10 (9): 1283. https://doi.org/10.3390/math10081283

Chen J, Zhang Z. 2023. An improved fusion of Landsat-7/8, Sentinel-2, and Sentinel-1 data for monitoring alfalfa: Implications for crop remote sensing. International Journal of Applied Earth Observation and Geoinformation 124: 103533. https://doi.org/10.1016/j.jag.2023.103533

Chmielewski FM. 2013. Phenology in agriculture and horticulture. *In* Schwartz MD. (ed.), Phenology: An Integrative Environmental Science. Springer: Dordrecht, Netherlands, pp: 539–561. https://doi.org/10.1007/978-94-007-6925-0_29

de la Fuente D, Rivilla E, Tena A, Vitorino J, Navascués E, Tabasco A. 2023. Yield estimation using machine learning from satellite imagery. BIO Web of Conferences 68: 01013. https://doi.org/10.1051/bioconf/20236801013

FAO (Food and Agriculture Organization). 2002. Agua y cultivos: logrando el uso óptimo del agua en la agricultura. Roma, Italia. https://agua.org.mx/wp-content/uploads/2010/11/118_agua_y_cultivos.pdf (Retrieved: April 2025).

Fatemeh K, Hossein Y. 2022. Crop classification based on phenology information by using time series of optical and synthetic-aperture radar images. Remote Sensing Applications: Society and Environment 27: 100812. https://doi.org/10.1016/j.rsase.2022.100812

García-Martínez H, Flores-Magdaleno H, Khalil-Gardezi A, Ascencio-Hernández R, Tijerina-Chávez L, Vázquez-Peña MA, Mancilla-Villa OR. 2020. Estimación de la fracción de cobertura de la vegetación en maíz (*Zea mays*) mediante imágenes digitales tomadas por un vehículo aéreo no tripulado (UAV). Revista de Fitotecnia Mexicana 43 (4): 399–409. https://doi.org/10.35196/rfm.2020.4.399

Guo H, Li Y, Shang J, Gu M, Huang Y, Gong B. 2017. Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications 73: 220–239. https://doi.org/10.1016/j.eswa.2016.12.035

Han J, Shi L, Yang Q, Huang K, Zha Y, Yu J. 2020. Real-time detection of rice phenology through convolutional neural network using handheld camera images. Precision Agriculture 22 (1): 154–178. https://doi.org/10.1007/s11119-020-09734-2

He H, García EA. 2009. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering 21 (9): 1263–1284. https://doi.org/10.1109/tkde.2008.239

Johnson JM, Khoshgoftaar TM. 2019. Survey on deep learning with class imbalance. Journal of Big Data 6 (1): 27. https://doi.org/10.1186/s40537-019-0192-5

Kayad A, Sozzi M, Gatto S, Marinello F, Pirotti F. 2019. Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. Remote Sensing 11 (23): 2873. https://doi.org/10.3390/rs11232873

Lagari P, Tsoukalas L, Safarkhani S, Lagaris I. 2021. Eliminating multicollinearity issues in neural network ensembles: Incremental, negatively correlated, optimal convex blending. arXiv. https://doi.org/10.48550/arXiv.2104.14715

Li J, Shen Y, Yang C. 2021. An adversarial generative network for crop classification from remote sensing timeseries images. Remote Sensing 13 (1): 65. https://doi.org/10.3390/rs13010065

Longchamps L, Philpot W. 2023. Full-season crop phenology monitoring using two-dimensional normalized difference pairs. Remote Sensing 15 (23): 5565. https://doi.org/10.3390/rs15235565

Mery D. 2011. BALU: A Matlab toolbox for computer vision, pattern recognition and image processing. Santiago, Chile. http://dmery.ing.puc.cl/index.php/balu (Retrieved: April 2025).

Mery D. 2015. Computer vision for X-ray testing: Imaging, systems, image databases, and algorithms. Springer: London, UK. 607 p.

Milicevic M, Zubrinic K, Grbavac I, Obradovic I. 2020. Application of deep learning architectures for accurate detection of olive tree flowering phenophase. Remote Sensing 12 (13): 2120. https://doi.org/10.3390/rs12132120

Minallah N, Tariq M, Aziz N, Khan W, Rehman A, Belhaouari SB. 2020. On the performance of fusion based planetscope and Sentinel-2 data for crop classification using inception inspired deep convolutional neural network. PLOS ONE 15 (9): e0239746. https://doi.org/10.1371/journal.pone.0239746

Nakariyakul S, Casasent DP. 2009. An improvement on floating search algorithms for feature subset selection. Pattern Recognition 42 (9): 1932–1940. https://doi.org/10.1016/j.patcog.2008.11.018

Ojala T, Pietikäinen M, Mäenpää T. 2000. Gray scale and rotation invariant texture classification with local binary patterns. *In* Computer Vision - ECCV 2000. Springer: Berlin, Germany, pp: 404–420. https://doi.org/10.1007/3-540-45054-8_27

Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. 2022. A review of feature selection methods for machine learning-based disease risk prediction. Frontiers in Bioinformatics 2: 927312. https://doi.org/10.3389/fbinf.2022.927312

Qin F, Liu D, Sun B, Ruan L, Ma Z, Wang H. 2016. Identification of alfalfa leaf diseases using image recognition technology. PLOS ONE 11 (12): e0168274. https://doi.org/10.1371/journal.pone.0168274

Robertson AR. 1976. The CIE 1976 color-difference formulae. Color Research and Application 2 (1): 7–11. https://doi.org/10.1002/j.1520-6378.1977.tb00104.x

Sadri S, Famiglietti JS, Pan M, Beck HE, BergA, Wood EF. 2022. FarmCan: A physical, statistical, and machine learning model to forecast crop water deficit for farms, Hydrology and Earth System Sciences 26 (20): 5373–5390. https://doi.org/10.5194/hess-26-5373-2022

Shojaeezadeh SA, Elnashar A, Weber TKD. 2025. A novel fusion of Sentinel-1 and Sentinel-2 with climate data for crop phenology estimation using Machine Learning. Science of Remote Sensing 11: 100227. https://doi.org/10.1016/j.srs.2025.100227

SIAP (Sistema de Información Agroalimentaria y Pesquera). 2019. Producción agrícola. Gobierno de México. Sistema de Información Agroalimentaria y Pesquera. Ciudad de México, México. http://infosiap.siap.gob.mx/gobmx/datosAbiertos.php (Retrieved: October 2019).

Solano-Correa YT, Bovolo F, Bruzzone L. 2019. A semi-supervised crop-type classification based on Sentinel-2 NDVI satellite image time series and phenological parameters. *In* 2019 IEEE International Geoscience and Remote Sensing Symposium. Institute of Electrical and Electronics Engineers. Yokohama, Japan, pp: 457–460. https://doi.org/10.1109/igarss.2019.8897922

Supriatna R, Wibowo A, Shidiq IPA. 2020. Spatial analysis of rice phenology using Sentinel 2 and UAV in Parakansalak, Sukabumi district, Indonesia. International Journal of GEOMATE 19 (72): 205–210. https://doi.org/10.21660/2020.72.5621

Thenkabail PS, Knox JW, Ozdogan M, Gumma MK, Congalton RG, Wu Z, Milesi C, Finkral A, Marshall M, Mariotto I. 2012. Assessing future risks to agricultural productivity, water resources and food security: How can remote sensing help? Photogrammetric Engineering and Remote Sensing 78 (8): 773–782

Velumani K, Madec S, de Solan B, López-Lozano R, Gillet J, Labrosse J, Jezequel S, Comar A, Baret F. 2020. An automatic method based on daily in situ images and deep learning to date wheat heading stage. Field Crops Research 252: 107793. https://doi.org/10.1016/j.fcr.2020.107793

Yang Q, Shi L, Han J, Yu J, Huang K. 2020. A near real-time deep learning approach for detecting rice phenology based on UAV images. Agricultural and Forest Meteorology 287: 107938. https://doi.org/10.1016/j.agrformet.2020.107938

Zhang Z, Zhang Y, Li Z. 2018. Removing the feature correlation effect of multiplicative noise. arXiv. https://doi.org/10.48550/arXiv.1809.07023

Agrociencia

ENZYMATIC ACTIVITY OF THREE STRAINS FROM TWO *Schizophyllum* SPECIES GROWN ON DIFFERENT SUBSTRATES

Alma Rosa **Agapito-Ocampo**[1,2], Mariel **Fabian-Jurado**[1], Ma. de Lourdes **Acosta-Urdapilleta**[1], Silvia **Capello-García**[3], Denis **Castro-Bustos**[1], Maura **Téllez-Téllez**[1*]

[1]Universidad Autónoma del Estado de Morelos. Centro de Investigaciones Biológicas. Avenida Universidad 1001, Chamilpa, Cuernavaca, Morelos, Mexico. C. P. 62209.
[2]Universidad Autónoma del Estado de Morelos. Doctorado en Ciencias Agropecuarias y Desarrollo rural. Avenida Universidad 1001, Chamilpa, Cuernavaca, Morelos, Mexico. C. P. 62209.
[3]Universidad Juárez Autónoma de Tabasco. División Académica de Ciencias Biológicas. Avenida Universidad s/n, Magisterial, Tabasco, Villahermosa, Mexico. C. P. 86040.
* Author for correspondence: maura.tellez@uaem.mx

Citation: Agapito-Ocampo AR, Fabian-Jurado M, Acosta-Urdapilleta ML, Capello-García S, Castro-Bustos D, Téllez-Téllez M. 2026. Enzymatic activity of three strains from two *Schizophyllum* species grown on different substrates. **Agrociencia** 60(1): 62-75. https://doi.org/ 10.47163/agrociencia.v60i1.3404

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: January 15, 2025.
Approved: January 21, 2026.
Published in Agrociencia: February 03, 2026.

ABSTRACT

Species of the genus *Schizophyllum* have been used in morphogenesis research and in the production of polysaccharides and enzymes. In this study, the growth and enzymatic activity (laccases, amylases, cellulases, pectinases, and xylanases) of two strains of *Schizophyllum commune* and one strain of *Schizophyllum radiatum* were evaluated on different agro-industrial substrates (cedar sawdust, jacaranda sawdust, pine sawdust, peanut shell, coconut fiber, corn stubble, and corn cobs). Growth rate, mycelial characteristics, and enzymatic activity were assessed in Petri dishes. All strains grew on the seven substrates, with higher mycelial density on peanut shells, corn stubble, and corn cobs. The highest enzymatic activity was observed on corn stubble and peanut shell, followed by jacaranda sawdust for amylase, pectinase, and xylanase. *Schizophyllum radiatum* showed greater mycelial extension but lower enzymatic activity than *S. commune* strains. Substrates with lower lignin content (peanut shells, corn stubble, and corn cobs) enhanced growth and enzymatic activity in all strains, indicating that these agro-industrial residues are suitable substrates for obtaining enzyme cocktails from *Schizophyllum* species.

Keywords: corn cobs, corn stubble, laccases, peanut shells, pectinases, xylanases.

INTRODUCTION

Interest in edible and medicinal mushrooms has increased due to their nutritional and nutraceutical value, thereby strengthening research on basidiomycete species for industrial applications. This demand requires improved strategies to optimize the production of fungal metabolites. The culture system influences fungal physiology, growth, and metabolic pathways, modifying the synthesis and secretion of specific metabolites and enzymes. Evidence indicates that solid-state culture can yield higher levels of fungal enzymes at the laboratory scale (Gomes *et al*., 2018).

White-rot fungi can degrade lignin, whereas other fungal groups have limited or no ability to break down lignin but can rapidly decompose other components of plant biomass (Kameshwar and Qin, 2017). Brown-rot fungi primarily break down the polysaccharides of lignocellulose, including cellulose and hemicellulose, while leaving lignin only slightly modified, resulting in a dry brown residue. These fungi have lost the genes encoding class II peroxidases, and depolymerization of plant cell wall polysaccharides occurs mainly through non-enzymatic Fenton reactions generated outside the hyphae (Veloz-Villavicencio *et al*., 2020). Fungal enzymes are essential for the efficient conversion of plant residues. The extracellular enzyme system includes hydrolases (e.g., cellulases, amylases, xylanases) involved in polysaccharide decomposition and phenoloxidases that degrade lignin and open phenyl rings (e.g., laccase, manganese peroxidase, lignin peroxidase, aryl alcohol oxidase). These enzymes have broad biotechnological applications in the food, paper, textile, bioremediation, and cosmetics industries, among others (Ergun and Urek, 2017).

The wood-degrading fungus *Schizophyllum commune* Fr. 1815 is globally distributed and has been recognized as a model organism for research on morphogenesis, gene regulation, and its metabolites (Ohm *et al*., 2010; Pelkmans *et al*., 2016). *Schizophyllum radiatum* Fr. 1851 and *S. commune* were previously considered conspecific based on morphological and genetic similarities (ITS and rDNA). However, multigene analysis has demonstrated that they are separate species that exhibit similar traits and are closely related (Mišković *et al*., 2023). The fruiting bodies of *S. commune* are consumed by various ethnic groups across Asia, Africa, and the Americas (Kamalebo *et al*., 2018). In Mexico, it is traditionally consumed in six southern states (Cappello-García *et al*., 2018). In Tabasco, its consumption and local sale have been documented in the municipalities of Teapa and Macuspana (Ruán-Soto and Cifuentes-Blanco, 2011).

Schizophyllum commune is the most extensively researched species within its genus regarding its biological activity and the characterization of significant molecules. The antitumor activity of polysaccharides from *S. radiatum* has also been examined (López-Legarda *et al*., 2021). *Schizophyllum commune* exhibits a restricted capacity for lignin degradation due to the absence of genes encoding peroxidases, akin to the traits of brown-rot fungi (Ohm *et al*., 2010). This species demonstrates elevated cellulase, xylanase, and pectinase activities, accompanied by diminished laccase activity (Zhu *et al*., 2016). This study assessed the growth and enzymatic activity of three *Schizophyllum* strains cultivated on seven agro-industrial residues to identify fungal strains with significant enzyme production potential.

MATERIALS AND METHODS

Microorganism

Two *S. commune* strains (HEMIM-98 and HEMIM-99) and one *S. radiatum* (HEMIM-107) strain were used. All strains were isolated from wild specimens and were preserved

in the Morelos Mycological Herbarium (HEMIM) collection at the Autonomous University of the State of Morelos. The strains were propagated and maintained on 90 × 15 mm Petri dishes with potato dextrose agar. After the mycelium colonized about 80 % of the surface area of the medium, the cultures were utilized as inoculum.

Culture conditions

Seven agro-industrial wastes were selected based on their lignocellulosic composition: cedar sawdust, jacaranda sawdust, pine sawdust, peanut shells, coconut fiber, corn stubble, and corn cobs. The chopped substrates, approximately 0.5 cm in size, were hydrated for 30 min, resulting in a moisture content of around 76.7, 72, 76.8, 63, 74, 69.2, and 71 % for cedar sawdust, jacaranda sawdust, pine sawdust, peanut shell, coconut fiber, corn stover, and corn cob, respectively. Each substrate was placed into 90 x 15 mm Petri dishes, filling 80 % of the plate volume, and sterilized at 121°C for 20 min. The plates were inoculated with a 5 mm diameter piece of mycelium and incubated in complete darkness at 23–25°C. All assays were performed in triplicate.

Growth speed and mycelial characterization

The mycelial growth velocity (Vc) was determined as the slope of the straight-line equation, obtained from radius measurements taken every 24 h until the mycelium fully colonized the Petri dish. Mycelial characterization was performed on the 17-day-old cultures, assessing texture, density (abundant, regular, or scarce), coloration, and mycelial type (aerial or creeping) for each strain (Sobal *et al*., 2007). All assays were conducted in triplicate.

Enzymatic extracts

Enzymatic extracts (EA) were obtained by adding 25 mL of sterile distilled water per gram of dry substrate colonized by mycelium after 17 d of culture. The mixtures were shaken at 100 rpm for 20 min and refrigerated at 4 °C for 24 h. After this period, they were centrifuged at 20 000 × g for 10 min at 2 °C using a HERMLE Z36HK centrifuge, and then stored at -20 °C until use.

Enzymatic activities

Laccase activity was evaluated in triplicate using 2,6-dimethoxyphenol (DMP) as the substrate. The reaction mixture comprised 950 µL of the substrate (2 mM DMP in 0.1 M acetate buffer, at pH 4.0, 4.5, 5.0, 5.5) and 50 µL of EA, incubated for 5 min at 40 °C. The absorbance was subsequently measured at 468 nm using a UV-Vis spectrophotometer (L6S). The activity unit (U) of laccases was defined as the quantity of enzyme that generates an increase of one absorbance unit per minute in the reaction mixture. Activity was expressed as U gX^{-1} (U per gram of dry agro-industrial waste colonized by mycelium) (Téllez-Téllez *et al*., 2008).

The activities of amylases, cellulases, pectinases, and xylanases were assessed using starch, carboxymethylcellulose, polygalacturonic acid, and birch xylan, with

all substrates dissolved at 1 % in 0.1 M acetate buffer at pH 5.0. Reaction mixtures consisting of 950 µL of substrate and 50 µL of EA were incubated at 50 °C for 45 min. The release of reducing sugars was quantified using the dinitrosalicylic acid (DNS) method, with absorbance measured at 575 nm (Miller, 1959). In all cases, one U was defined as the amount of enzyme that released one µmol of reaction product per minute under assay conditions (in U gX^{-1}). All tests were conducted in triplicate.

Statistical analysis

All analyses were performed in triplicate, and mean values were used for statistical analysis. Data were analyzed using SigmaStat software (Systat Software Inc., Palo Alto, CA, USA). Differences among treatments were assessed by one-way analysis of variance (ANOVA) and Tukey's test ($p < 0.05$).

RESULTS AND DISCUSSION

Mycelial growth rate and characterization

In general, the *S. radiatum* strain exhibited higher Vc values than both *S. commune* strains (Table 1). Corn stubble exhibited the highest Vc values for all three strains, whereas pine sawdust resulted in the lowest Vc values. The three lignin-rich substrates exhibited the lowest Vc values.

Regarding the morphological characteristics observed (Table 2), the mycelium on peanut shells, corn stover, and corn cobs exhibited a white hue, a cottony texture, and a density ranging from regular to abundant. In contrast, all three sawdust substrates, as well as the coconut fiber, exhibited hyaline mycelium with an inconspicuous texture and very low (near-zero) density.

Table 1. Mycelial growth velocity of the evaluated *Schizophyllum* strains on various agro-industrial waste substrates.

Substrate	Mycelial growth rate (mm h^{-1})*		
	S. commune HEMIM-98	*S. commune* HEMIM-99	*S. radiatum* HEMIM-107
Cedar sawdust	0.152 (0.004)c	0.178 (0.002)b	0.205 (0.001)c
Jacaranda sawdust	0.178 (0.008)b	0.171 (0.003)b	0.261 (0.015)b
Pine sawdust	0.149 (0.006)c	0.165 (0.016)b	0.194 (0.008)c
Peanut shells	0.181 (0.006)b	0.172 (0.005)b	0.264 (0.022)b
Coconut fiber	0.178 (0.006)b	0.177 (0.012)b	0.226 (0.012)c
Corn stubble	0.206 (0.002)a	0.203 (0.006)a	0.302 (0.005)a
Corn cob	0.173 (0.003)b	0.185 (0.006)b	0.225 (0.007)c

*Mean values per column with different letters are statistically different ($p \leq 0.05$). Standard error is shown in parentheses.

Table 2. Mycelium characteristics of the evaluated *Schizophyllum* strains grown on various agro-industrial substrates.

Substrate	Characteristics				
	Color	Texture	Type of mycelium	Density	Hyphal aggregation
		S. commune HEMIM-98			
Cedar sawdust	Hyaline	Absent	Aerial	Low	Absent
Jacaranda sawdust	Hyaline	Absent	Aerial	Low	Present
Pine sawdust	Hyaline	Absent	Aerial	Low	Present
Peanut shell	White	Cottony	Aerial	Middle	Present
Coconut fiber	Hyaline	Absent	Creeping	Low	Absent
Corn stubble	White	Cottony	Aerial	High	Present
Corn cob	White	Cottony	Aerial	Middle	Absent
		S. commune HEMIM-99			
Cedar sawdust	Hyaline	Absent	Aerial	Low	Absent
Jacaranda sawdust	Hyaline	Absent	Aerial	Low	Absent
Pine sawdust	Hyaline	Absent	Aerial	Low	Absent
Peanut shell	White	Cottony	Aerial	Middle	Present
Coconut fiber	Hyaline	Absent	Creeping	Low	Absent
Corn stubble	White	Cottony	Aerial	High	Present
Corn cob	White	Cottony	Aerial	Middle	Absent
		S. radiatum HEMIM-107			
Cedar sawdust	Hyaline	Absent	Aerial	Low	Absent
Jacaranda sawdust	Hyaline	Absent	Creeping	Middle	Absent
Pine sawdust	Hyaline	Cottony	Creeping	Middle	Absent
Peanut shell	Hyaline	Cottony	Aerial	High	Present
Coconut fiber	Hyaline	Absent	Creeping	Low	Absent
Corn stubble	White	Cottony	Aerial	High	Absent
Corn cob	White	Cottony	Aerial	High	Absent

The best substrates for the three strains, based on mycelial density, were peanut shells, corn stubble and corn cob (Figures 1, 2 and 3), with total invasion occurring after 6–10 d. In the other substrates, total invasion occurred after 7–12 d. *Schizophyllum commune* HEMIM-98 presented fruiting bodies and spores after 20 d of growth (Figure 4).

Previous studies have reported the growth of *S. commune* on agro-industrial substrates incubated at 28 °C under dark conditions. The fastest growth occurred in cocoa shells (0.4 mm h^{-1}), banana leaves (0.41 mm h^{-1}), and mixed substrates of coconut-cocoa (0.42 mm h^{-1}), cocoa-banana (0.41 mm h^{-1}), coconut-cocoa (0.42 mm h^{-1}), and cocoa-banana (0.41 mm h^{-1}), with no statistically significant differences observed (Carreño-Ruiz *et al*., 2014).

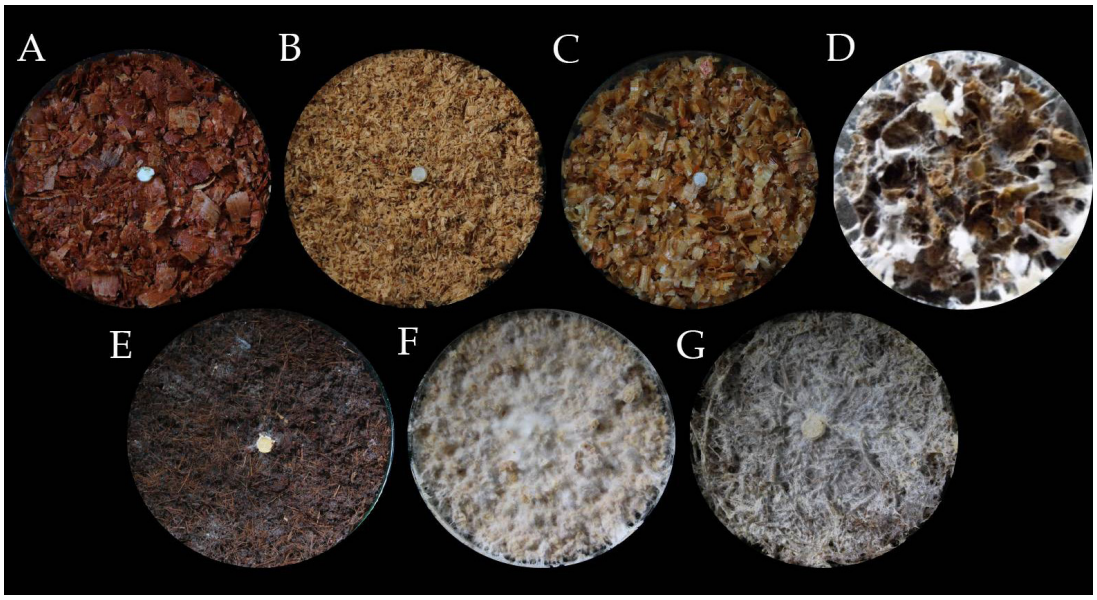

Figure 1. Mycelial growth of *Schizophyllum commune* HEMIM-98 on different substrates. A: cedar sawdust; B: jacaranda sawdust; C: pine sawdust; D: peanut shell; E: coconut fiber; F: corn stubble; G: corn cob.

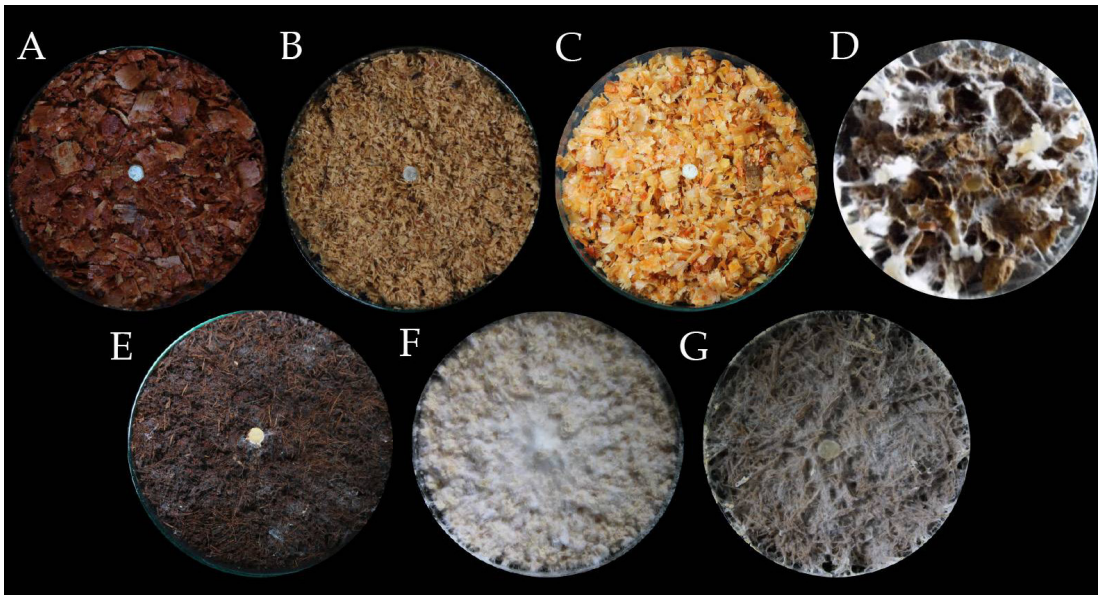

Figure 2. Mycelial growth of *Schizophyllum commune* HEMIM-99 on different substrates. A: cedar sawdust; B: jacaranda sawdust; C: pine sawdust; D: peanut shell; E: coconut fiber; F: corn stubble; G: corn cob.

Figure 3. Mycelial growth of *Schizophyllum radiatum* HEMIM-107 on different substrates. A: cedar sawdust; B: jacaranda sawdust; C: pine sawdust; D: peanut shell; E: coconut fiber; F: corn stubble; G: corn cob.

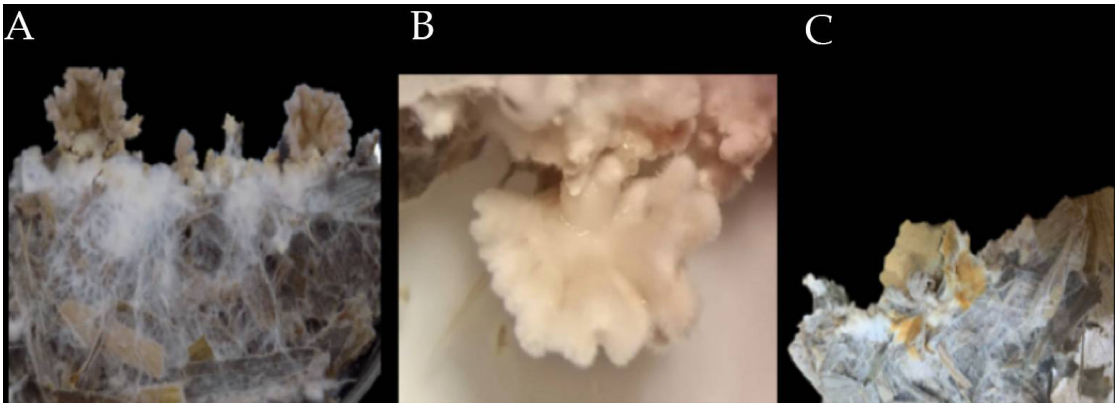

Figure 4. Fruiting bodies and sporulation of *Schizophyllum commune* HEMIM-98 grown on corn stover in Petri dishes. A, B: Fruiting body; C: spores.

In another study, Carreño-Ruiz *et al*. (2020) cultivated four *S. commune* strains on cocoa shells, banana leaves, corn leaves, and mulatto sticks. All strains showed accelerated growth on mulatto stick (Vc = 0.41 mm h^{-1}). One strain attained a Vc value of 0.195 mm h^{-1} on banana leaves, whereas the other three strains cultivated on corn leaves exhibited Vc values of 0.056, 0.098, and 0.17 mm h^{-1}. The authors proposed that *S. commune* exhibited greater growth on mulatto sticks due to their status as a natural host. However, its use as a substrate is not feasible due to the ecological and economic implications of removing the tree from the natural environment (Ruan-Soto *et al*., 2004).

Enzymatic activity

All strains presented enzymatic activities for the five tests assessed, with xylanase activity typically yielding the highest values. The *S. radiatum* strain showed the lowest laccase and pectinase activities but ranked second in xylanase activity. The three strains demonstrated amylase activity (Figure 5A), with *S. commune* HEMIM-99 exhibiting the highest value and *S. radiatum* the lowest, both cultivated on jacaranda sawdust (10.26 and 0.27 U gX^{-1}, respectively). However, *S. radiatum* reached its maximum amylase activity when grown on corn stover (6.77 U gX^{-1}). For cellulase activity, the best substrates for all strains were peanut shells and corn stubble (Figure 5B). The HEMIM-99 strain grown on the three types of sawdust did not exhibit cellulase activity and had minimal activity when grown on coconut fiber and corn cobs. Meanwhile, *S. commune* HEMIM-98 and *S. radiatum* had the highest cellulase activity on corn stover (14.97 and 6.69 U gX^{-1}, respectively), whereas HEMIM-99 reached its maximum cellulase activity on peanut shells (9.02 U gX^{-1}).

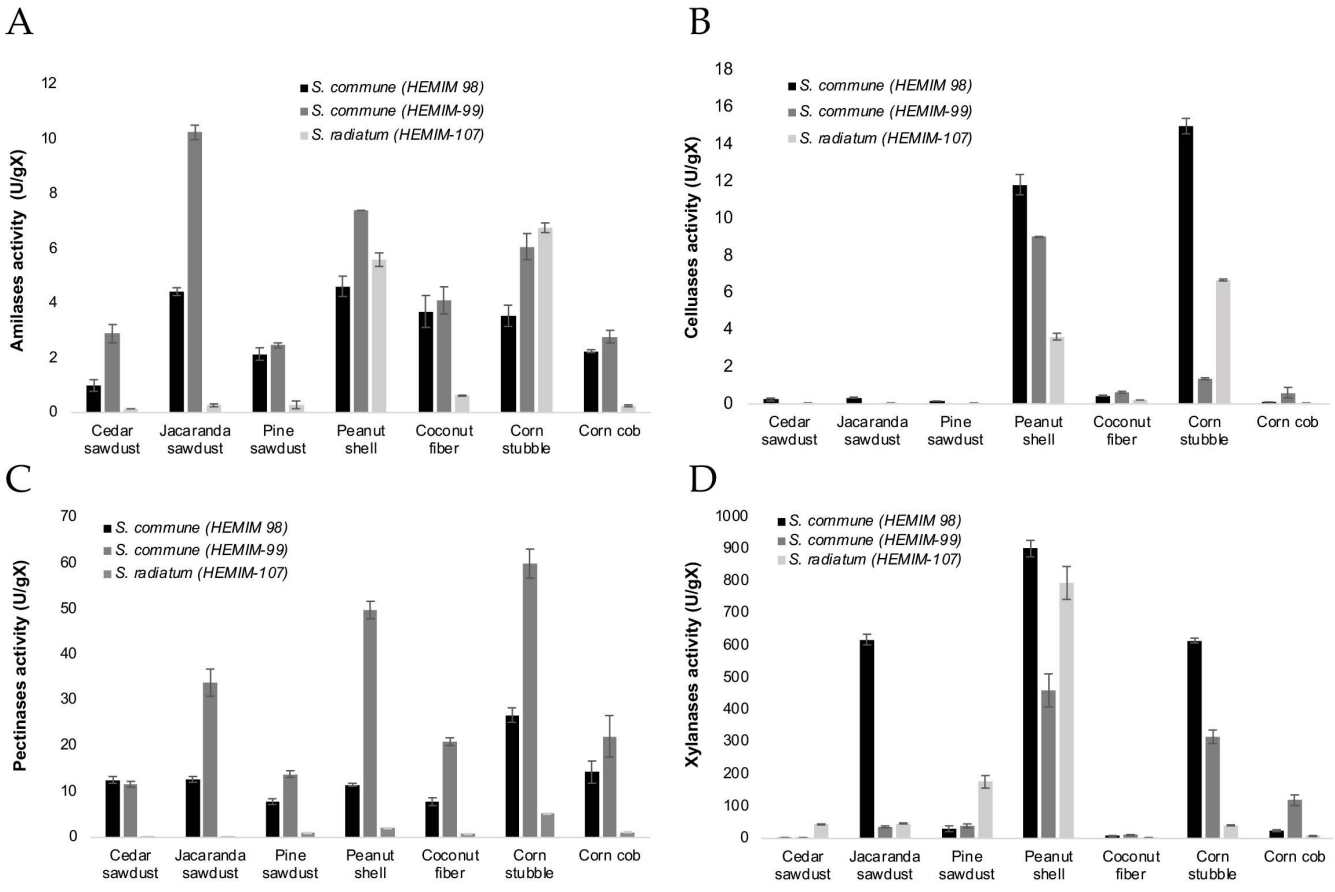

Figure 5. Enzymatic activity of the evaluated *Schizophyllum* strains grown on different agro-industrial residues. A: Amylases; B: cellulases; C: pectinases; D: xylanases.

The pectinase activity of the *S. radiatum* strain was minimal to undetectable on all evaluated substrates, while both *S. commune* strains exhibited measurable activity when grown on all substrates; HEMIM-98 had values between 10 and 26 U gX^{-1}, and HEMIM-99 had its highest value at 59.83 U gX^{-1} on corn stubble (Figure 5C). Peanut shells served as the best substrate for xylanase activity across all three strains, yielding values of 900, 459, and 795 U gX^{-1} for HEMIM-98, HEMIM-99, and HEMIM-107, respectively (Figure 5D). Corn stubble also provided significant xylanase activity, with values of 615, 317, 176 U gX^{-1} for HEMIM-98, HEMIM-99, and HEMIM-107, respectively. The lowest xylanase activity was found for HEMIM-107 (4.3 U gX^{-1}) grown on coconut fiber, while *S. commune* grown on cedar sawdust achieved values of 2.86 and 2.09 U gX^{-1} for HEMIM-98 and HEMIM-99, respectively.

Regarding laccase activity, the three strains demonstrated elevated activity at pH 5.0 and 5.5; *S. commune* HEMIM 98 exhibited the highest activity, while *S. radiatum* had the lowest (Figure 6). The highest laccase activity recorded for the three strains occurred on corn stubble, reaching 23.6 U gX^{-1} (HEMIM-98), 17.7 U gX^{-1} (HEMIM-99), and 10.29 U gX^{-1} (HEMIM-107), followed by coconut fiber with 1.46, 2.33, and 0.8 U gX^{-1}, respectively. On all other substrates, laccase activity was less than 2 U gX^{-1} (Figure 6).

In general, the substrates that favored fungal growth, mycelium density, and hydrolytic enzymatic activity (amylases, pectinases, cellulases and xylanases) were those with the lowest lignin content. Corn stubble contains 6.8 % lignin, 27.6 % hemicellulose, and 45.5 % cellulose (Costa *et al*., 2015); corn cobs contain 15 % lignin, 35 % hemicellulose, and 45 % cellulose (Garrote *et al*., 2007); and peanut shells contain 27 % lignin, 30 % hemicellulose, and 45 % cellulose (Gatani *et al*., 2010). In the case of laccase activity, corn stubble and coconut fiber were also the most favorable substrates. Coconut fiber contains 41 % lignin, 16.3 % hemicellulose, and 31.2 % cellulose (Abdullah *et al*., 2021).

In recent decades, several studies have explored the use of agro-industrial waste as raw materials for the production of value-added products. Fungal enzymes participate in the degradation of these solid substrates to supply a source of carbon and mineral nutrients for fungal growth. Cellulases, xylanases, pectinases, and ligninases participate in the degradation of lignocellulosic substrates (Gomes *et al*., 2018). In *S. commune*, the production of several industrially important enzymes has been reported, including cellulases, xylanases, pectinases, lipases, laccases, manganese peroxidase, and lignin peroxidase (Kam *et al*., 2016; Sornlake *et al*., 2017; Gautam *et al*., 2018; Kumar *et al*., 2018; Mehmood *et al*., 2018).

The genome of *S. commune* contains 16 genes encoding lignin oxidative enzymes (FOLymes), including one cellobiose dehydrogenase, one aryl alcohol oxidase, one alcohol oxidase, two laccases, one glyoxal oxidase, and four benzoquinone reductases (Ohm *et al*., 2010). It also has vast enzymatic machinery to degrade cellulose, pectin, and hemicellulose. In total, 240 glycoside hydrolases, 75 glycosyltransferases, 16 polysaccharide lyases, and 30 carbohydrate esterases have been identified, totaling 366 carbohydrate-active enzymes (CAZymes), of which 106 are presumed to be involved in plant polysaccharide degradation. Although *S. commune* is sometimes classified as

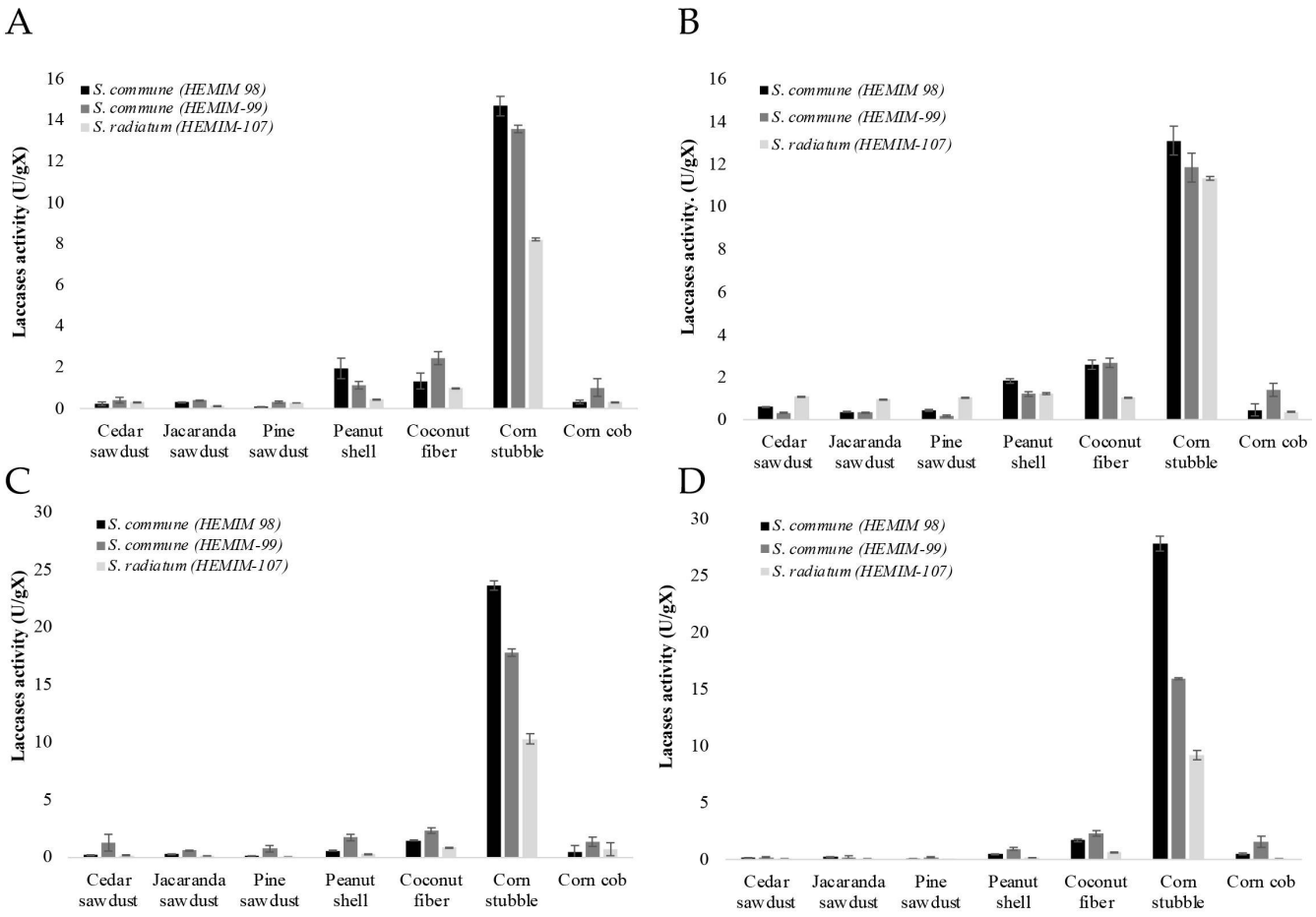

Figure 6. Laccase activity of the evaluated *Schizophyllum* strains grown on different agro-industrial residues. A: pH 4.0; B: pH 4.5; C: pH 5.0; D: pH 5.5.

a white-rot fungus, it does not significantly degrade lignin *in vitro*. The genome lacks class II peroxidases and contains a reduced set of enzymes with cellulose-binding modules. However, it encodes 22 lytic polysaccharide monooxygenases. Due to these characteristics, *S. commune* exhibits a decay mode between dark-rot fungi and has been considered an intermediate between white- and dark-rot fungi (Riley *et al*., 2014). Zhu *et al*. (2016) reported that *S. commune* excretes a broad set of extracellular enzymes involved in the degradation of plant cell wall components. Their study identified increased activities of several enzymes during cultivation, including endoglucanase, cellobiohydrolase, β-glucosidase, polygalacturonase, β-xylosidase, and xylanase. These findings suggest that lignocellulose degradation by *S. commune* involves a hydroxyl radical-mediated mechanism for lignocellulose modification, in parallel with a synergistic system of several enzymes that degrade polysaccharides. Sornlake *et al*.

(2017) evaluated the hydrolytic efficiency of enzymes produced by *S. commune* G-135, which contain glycosyl hydrolases. One mutant (Avicel-PH101) efficiently hydrolyzed several lignocellulosic residues, with the highest values reported in corn cobs (98 %), and a significant improvement in xylan conversion. In this study, overall corn stubble yielded the highest enzymatic activities, although corn cob was suitable for producing amylases and pectinases.

Arunrattanamook *et al*. (2022) reported that the maximum enzymatic activity of *S. commune* was reached on the second day of cultivation and remained stable until day seven. This period coincided with the onset of extracellular polysaccharide production. At the highest levels of these polysaccharides, xylanase activity reached its minimum, indicating that this activity is not directly associated with fungal growth. Singh *et al*. (2017) reported that nutrient depletion began after 5 d of liquid cultivation of *S. commune*. The lignocellulolytic enzyme system was activated, increasing protein content and enabling enzymatic degradation of cellulose and hemicellulose, thereby increasing the concentration of reducing sugars in the medium and favoring the growth of the fungus.

Kondaveeti *et al*. (2020) reported two cellobiohydrolase enzymes from *S. commune* KMJ820 obtained by liquid culture of approximately 50 and 150 kDa; these enzymes stand out compared to other cellobiohydrolases for their high enzymatic activity and their potential for large-scale production of glucose or ethanol by biological methods. Faheem *et al*. (2023) purified and characterized a ρ-diphenol oxidase from *S. commune* (MF-O5); the enzyme demonstrated tolerance to salt, metal ions, organic solvents, and surfactants, highlighting its potential for use in various industrial applications.

CONCLUSIONS

Schizophyllum strains grow on economically accessible substrates, enabling the production of enzymatic cocktails for diverse biotechnological applications. Peanut shells, corn stubble, and corn cobs are abundant agro-industrial wastes, making them accessible. Their use as substrates for fungal growth and enzyme production offers a sustainable alternative for generating value-added products while preventing them from contributing to organic matter accumulation. This study constitutes the first report on enzyme production by *S. radiatum*, demonstrating its potential as an alternative to *S. commune*. Although *S. radiatum* exhibits comparatively lower enzymatic activity, its rapid growth makes it a viable and efficient option for biotechnological applications.

REFERENCES

Abdullah NA, Sainorudin MH, Rani MSA, Mohammad M, Kadir NH, Asim N. 2021. Structure and thermal properties of microcrystalline cellulose extracted from coconut husk fiber. Polimery 66 (3): 187–192. https://doi.org/10.14314/polimery.2021.3.4

Arunrattanamook N, Sornlake W, Champreda V. 2022. Co-production of schizophyllan and cellulolytic enzymes from bagasse by *Schizophyllum commune*. Bioscience, Biotechnology, and Biochemistry 86 (8): 1144–1150. https://doi.org/10.1093/bbb/zbac091

Cappello-García S, Carreño-Ruíz SD, Gaitán-Hernández R. 2018. Fruit body production of *Schizophyllum commune*. *In* Sánchez JE, Mata G, Royse DJ. (eds.), Updates on Tropical Mushrooms. Basic and Applied Research. El Colegio de la Frontera Sur: San Cristóbal de las Casas, Mexico, pp: 95–104.

Carreño-Ruiz SD, Cappello-García S, Gaitán-Hernández R, Cifuentes-Blanco J, Rosique-Gil E. 2014. Growth of three tropical edible fungi in culture mediums and agricultural waste. Revista Mexicana de Ciencias Agrícolas 5 (8): 1447–1458. https://doi.org/10.29312/remexca.v5i8.822

Carreño-Ruiz SD, Cappello-García S, Gaitán-Hernández R, Torres-de la Cruz M, Gaspar-Génico JA, Rosique-Gil JE. 2020. Producción de basidiomas de Schizophyllum commune (Fungi: Basidiomycota) en subproductos agrícolas de Tabasco, México. AgroProductividad 13 (5): 65–71.

Costa LAS, Assis DJ, Gomes GVP, da Silva JBA, Fonsêca AF, Druzian JI. 2015. Extraction and characterization of cellulose nanocrystals from corn stover. Cellulose Chemistry and Technology 49 (2): 127–133. https://doi.org/10.1016/j.matpr.2015.04.045

Ergun SO, Urek RO. 2017. Production of ligninolytic enzymes by solid state fermentation using *Pleurotus ostreatus*. Annals of Agrarian Science 15 (2): 273–277. https://doi.org/10.1016/j.aasci.2017.04.003

Faheem M, Bokhari SAI, Malik MA, Ahmad B, Riaz M, Zahid N, Hussain A, Ghani A, Ullah H, Shah W, *et al*. 2023. Production, purification, and characterization of p-diphenol oxidase (PDO) enzyme from lignolytic fungal isolate *Schizophyllum commune* MF-O5. Folia Microbiologica 68 (6): 867–888. https://doi.org/10.1007/s12223-023-01056-w

Garrote G, Falqué E, Domínguez H, Parajó JC. 2007. Autohydrolysis of agricultural residues: Study of reaction byproducts. Bioresource Technology 98 (10): 1951–1957. https://doi.org/10.1016/j.biortech.2006.07.049

Gatani M, Argüello R, Sesín S. 2010. Effect of chemical treatments on the mechanical properties of peanut shell and cement blends. Materiales de Construcción 60 (298): 137–147. https://doi.org/10.3989/mc.2010.46908

Gautam A, Kumar A, Bharti AK, Dutt D. 2018. Rice straw fermentation by *Schizophyllum commune* ARC-11 to produce high level of xylanase for its application in pre-bleaching. Journal of Genetic Engineering and Biotechnology 16 (2): 693–701. https://doi.org/10.1016/j.jgeb.2018.02.006

Gomes E, da Silva R, de Cassia Pereira J, Ladino-Orjuela G. 2018. Fungal growth on solid substrates: A physiological overview. *In* Pandey A, Larroche C, Soccol CR. (eds.), Current Developments in Biotechnology and Bioengineering. Elsevier: Amsterdam, Netherlands, pp: 31–56. https://doi.org/10.1016/B978-0-444-63990-5.00003-7

Kam YC, Hii SL, Sim CY, Gaik L, Ong A. 2016. *Schizophyllum commune* lipase production on pretreated sugarcane bagasse and its effectiveness. International Journal of Polymer Science 2016 (1): 2918202. https://doi.org/10.1155/2016/2918202

Kamalebo HM, Wa Malale HNS, Ndabaga CM, Degreef J, Kesel AD. 2018. Uses and importance of wild fungi: Traditional knowledge from the Tshopo province in the Democratic Republic of the Congo. Journal of Ethnobiology Ethnomedicine 14 (1). https://doi.org/10.1186/s13002-017-0203-6

Kameshwar AKS, Qin W. 2017. Qualitative and quantitative methods for isolation and characterization of lignin-modifying enzymes secreted by microorganisms. Bioenergy Research 10 (1): 248–266. https://doi.org/10.1007/s12155-016-9784-5

Kondaveeti S, Patel SKS, Woo J, Wee JH, Kim SY, Al-Raoush RI, Kim IW, Kalia VC, Lee JK. 2020. Characterization of cellobiohydrolases from *Schizophyllum commune* KMJ820. Indian Journal of Microbiology 60 (2): 160–166. https://doi.org/10.1007/s12088-019-00843-9

Kumar B, Bhardwaj N, Alam A, Agrawal K, Prasad H, Verma P. 2018. Production, purification and characterization of an acid/alkali and thermo tolerant cellulase from *Schizophyllum commune* NAIMCC-F-03379 and its application in hydrolysis of lignocellulosic wastes. AMB Express 8 (1): 173. https://doi.org/10.1186/s13568-018-0696-y

López-Legarda X, Rostro-Alanis M, Parra-Saldivar R, Villa-Pulgarín JA, Segura-Sánchez F. 2021. Submerged cultivation, characterization and *in vitro* antitumor activity of polysaccharides from *Schizophyllum radiatum*. International Journal of Biological Macromolecules 186: 919–932. https://doi.org/10.1016/j.ijbiomac.2021.07.084

Mehmood T, Saman T, Irfan M, Anwar F, Salman M, Tabassam Q. 2018. Pectinase production from *Schizophyllum commune* through central composite design using citrus waste and its immobilization for industrial exploitation. Waste Biomass Valorization 10 (9): 2527–2536. https://doi.org/10.1007/s12649-018-0279-9

Miller GL. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Analytical Chemistry 31 (3): 426–428. https://doi.org/10.1021/ac60147a030

Mišković J, Rašeta M, Krsmanović N, Karaman M. 2023. Update on mycochemical profile and selected biological activities of genus *Schizophyllum* Fr. 1815. Microbiology Research 14 (1): 409–429. https://doi.org/10.3390/microbiolres14010031

Ohm RA, de Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, de Vries RP, Record E, Levasseur A, Baker SE, *et al*. 2010. Genome sequence of the model mushroom *Schizophyllum commune*. Nature Biotechnology 28 (9): 957–963. https://doi.org/10.1038/nbt.1643

Pelkmans J, Lugones L, Wösten HA. 2016. Fruiting body formation in basidiomycetes. *In* Wendland J. (ed.), Growth, Differentiation and Sexuality. The Mycota, vol 1. Springer: Cham, Switzerland, pp: 387–405. https://doi.org/10.1007/978-3-319-25844-7_15

Riley R, Salamov AA, Brown D, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard V, Morin E, Otillar R, *et al*. 2014. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. Proceedings of the National Academy of Sciences 111 (27): 9923–9928. https://doi.org/10.1073/pnas.1400592111

Ruán-Soto F, Cifuentes-Blanco J. 2011. Notas etnomicológicas del poblado de Teapa, Tabasco. *In* López-Hernández ES, Ruiz-Macdonel MA. (eds.), Educación Ambiental para la Conservación de la Biodiversidad. El Colegio de Investigadores de Tabasco A.C. Villahermosa, México, pp: 249–256

Ruan-Soto F, Garibay-Orijel R, Cifuentes-Blanco J. 2004. Conocimiento micológico tradicional en la planicie costera del Golfo de México. Revista Mexicana de Micología 19: 57–70.

Singh MK, Kumar M, Thakur IS. 2017. Proteomic characterization and schizophyllan production by *Schizophyllum commune* ISTL04 cultured on *Leucaena leucocephala* wood under submerged fermentation. Bioresource Technology 236: 29–36. https://doi.org/10.1016/j.biortech.2017.03.170

Sobal M, Martínez-Carrera D, Morales P, Roussos S. 2007. Classical characterization of mushroom genetic resources from temperate and tropical regions of Mexico. Micología Aplicada Internacional 19 (1): 15–23.

Sornlake W, Rattanaphanjak P, Champreda V, Eurwilaichitr L, Kittisenachai S, Roytrakul S, Fujii T, Inoue H. 2017. Characterization of cellulolytic enzyme system of *Schizophyllum commune* mutant and evaluation of its efficiency on biomass hydrolysis. Bioscience, Biotechnology and Biochemistry 81 (7): 1289–1299. https://doi.org/10.1080/09168451.2017.1320937

Téllez-Téllez M, Fernández FJ, Montiel-González AM, Sánchez C, Díaz-Godínez G. 2008. Growth and laccase production by *Pleurotus ostreatus* in submerged and solid-state fermentation. Applied Microbiology Biotechnology 81 (4): 675–679. https://doi.org/10.1007/s00253-008-1628-6

Veloz-Villavicencio E, Mali T, Mattila HK, Lundell T. 2020. Enzyme activity profiles produced on wood and straw by four fungi of different decay strategies. Microorganisms 8 (1): 73. https://doi.org/10.3390/microorganisms8010073

Zhu N, Liu J, Yang J, Lin Y, Yang Y, Ji L, Li M, Yuan H. 2016. Comparative analysis of the secretomes of *Schizophyllum commune* and other wood-decay basidiomycetes during solid-state fermentation reveals its unique lignocellulose-degrading enzyme system. Biotechnology for Biofuels 9 (42): 1–22. https://doi.org/10.1186/s13068-016-0461-x

Agrociencia

EXOPOLYSACCHARIDE SYNTHESIS BY *Bacillus thuringiensis* HA1 USING CARBON SOURCES FROM THE SUGARCANE AGROINDUSTRY

Jesús David **Castilla-Marroquín**[1], Francisco **Hernández-Rosas**[1], José Andrés **Herrera-Corredor**[1], Neith **Pacheco**[2], Ricardo **Hernández-Martínez**[3*]

[1]Colegio de Postgraduados Campus Córdoba. Carretera Federal Córdoba-Veracruz km 348, Manuel León, Amatlán de los Reyes, Veracruz, Mexico. C. P. 94953.
[2]CIATEJ (Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco Tablaje Catastral 31264 Km 5.5 Carretera Sierra Papacal Chuburna Puerto Parque Científico Tecnológico de Yucatán, Yucatán, Mérida, Mexico. C. P. 97302.
[3]SECIHTI-Colegio de Postgraduados Campus Córdoba. Carretera Federal Córdoba-Veracruz km 348, Manuel León, Amatlán de los Reyes, Veracruz, Mexico. C. P. 94953.2
* Author for correspondence: odracirhema@gmail.com

Citation: Castilla-Marroquín JD, Hernández-Rosas F, Herrera-Corredor JA, Pacheco N, Hernández-Martínez R. 2025. Exopolysaccharide synthesis by *Bacillus thuringiensis* HA1 using carbon sources from the sugarcane agroindustry. **Agrociencia** 60(1): 76-89. https://doi.org/ 10.47163/agrociencia.v60i1.3524

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: June 13, 2025.
Approved: January 12, 2026.
Published in Agrociencia: January 15, 2026.

ABSTRACT

Exopolysaccharides are biopolymers produced by bacteria and have characteristics that make them suitable for applications in the pharmaceutical, environmental, and food industries. However, exopolysaccharide production faces challenges like high production costs. Therefore, strategies such as culture conditions improvement, strain selection, and the use of low-cost carbon sources have emerged as alternatives to improve exopolysaccharide production. In this work, the capability of *Bacillus thuringiensis* HA1 to produce exopolysaccharides using low-cost carbon sources (commercial sucrose, molasses, and panela) was explored. The production conditions were evaluated as follows: fermentation time (0–86 h), initial pH (5–9), temperature (31–43 °C), carbon sources (commercial sucrose, molasses, and panela), and concentration of carbon sources (50–350 g L^{-1}). The settled conditions to assess the carbon sources were 60 h, 37 °C, and pH 7.5. Exopolysaccharide production was higher using commercial sucrose (23.54 mg mL^{-1}), followed by molasses (8.62 mg mL^{-1}) and panela (6.37 mg mL^{-1}). The sucrose sample showed similarity to a glucan-type exopolysaccharide, since the presence of peaks at 1000–1200 is characteristic of C–O–C glycosidic linkages, while the molasses sample showed similarity to the standard levan. These results were achieved without pretreating the carbon sources, thus allowing the process to be economically feasible. To date, *Bacillus thuringiensis* has not been reported as a producer of two types of exopolysaccharides using different carbon sources.

Keywords: levan, biopolymers, molasses, commercial sucrose, submerged culture.

INTRODUCTION

Exopolysaccharides (EPSs) are high-molecular-weight carbohydrate biopolymers produced by microorganisms (Nadzir *et al*., 2021). EPSs function as protective mechanisms against environmental factors and are key for adaptation, survival, and

other functionalities (Osemwegie *et al*., 2020). These polysaccharides are known for their biocompatibility, nontoxicity, and unique functionalities, making them valuable in healthcare and biomedical applications such as drug delivery systems, healing and tissue engineering, collagen stimulants and anti-aging agents, antioxidants and skin care agents, cholesterol or triglyceride reducers, antibiotics to promote cytotoxic activity for colon and breast cancer, immune-stimulatory agents, and stabilizers in formulations.

Probiotic-derived EPSs exhibit health-promoting properties by improving the human digestive system, thus contributing to health and well-being (Aziz *et al*., 2022; Pourjafar *et al*., 2022; Wu *et al*., 2022; Ahuja *et al*., 2023; Waoo *et al*., 2023). In the food industry, EPSs are applied as texture enhancers, viscosity agents, and moisture retainers in products such as low-fat dairy, gluten-free bakery items, and fermented meats (Abarquero *et al*., 2021; Pourjafar *et al*., 2023). In the environmental field, EPSs have been used in remediation strategies to address heavy metal contamination and organic pollutant bioremediation (Balíková *et al*., 2022).

EPS production can be achieved by bacterial fermentation (Saadat *et al*., 2019). Microorganisms from the genera *Leuconostoc*, *Zymomonas*, *Halomonas*, and *Bacillus* have been confirmed as EPS producers (Braga *et al*., 2022; Erkorkmaz *et al*., 2022; Vega-Vidaurri *et al*., 2022). Several *Bacillus* species, such as *B. subtilis*, *B. licheniformis*, and *B. megaterium*, have been identified as EPS producers, most notably of levan, a fructan biopolymer with significant prebiotic and antioxidant properties (Díaz-Cornejo *et al*., 2022). However, EPS production faces challenges such as high production costs, low yields, and purity constraints (Erkorkmaz *et al*., 2022; Zhang *et al*., 2023). Therefore, approaches including strain selection, genetic engineering, culture condition optimization (initial pH, temperature, and fermentation time), and the use of cheap carbon sources have emerged as alternatives to improve EPS production (Nguyen *et al*., 2020; Zhang *et al*., 2023).

Many agro-industrial co-products such as fruit pomace and husk, lignocellulosic biomass, and molasses have been used for EPS production (Wang *et al*., 2025). However, the pretreatment of carbon sources like fruit waste or lignocellulosic biomass is necessary to ensure sugar availability, which increases EPS production costs (Pérez-Contreras *et al*., 2025). Although pretreatment may improve yields, Wang *et al*. (2025) found that out of 65 studies, 37 did not use pretreatments, and submerged fermentation was employed in 63 studies. Therefore, unpretreated sugar agro-industrial products and by-products such as sucrose, molasses, and panela can serve as alternatives for EPS production, as they are abundant and constitute important raw materials (Ni *et al*., 2022; Venkatesh *et al*., 2023).

Sucrose is a natural sweetener widely used for human consumption and has been studied as an enzyme production inducer, leading to the production of bioproducts like sucrose isomers, oligosaccharides, and monosaccharides (Ni *et al*., 2022). Molasses is a dark-brown, viscous liquid rich in proteins, inorganic salts, trace elements, and high sugar concentrations (Liang *et al*., 2022). Panela, known in Latin America as a

traditional unrefined sweetener or non-centrifugal sugar, contains phenolic acids, flavonoids, minerals, and bioactive compounds (Zidan and Azlan, 2022). Therefore, the objective of this work was to investigate the capability of *Bacillus thuringiensis* HA1 to synthesize EPSs using cheap carbon sources (commercial sucrose, molasses, and panela) produced by the sugarcane agroindustry.

MATERIALS AND METHODS

Exopolysaccharides (EPSs) were produced using *Bacillus thuringiensis* HA1, obtained from the Laboratory of Applied Microbial Biotechnology at the Postgraduate College Campus Córdoba. The strain was initially cultured on nutrient agar (23 g L^{-1}) in slanted tubes and incubated at 37 °C for 24 h. Subsequently, the reactivated culture was transferred to 250 mL Erlenmeyer flasks containing nutrient broth (8 g L^{-1}) and incubated at 37 °C for an additional 24 h under continuous agitation at 150 rpm. The resulting biomass was used as the inoculum (Gayosso-Sánchez *et al*., 2024).

Strain conservation was performed by cultivating *B. thuringiensis* HA1 as described above, followed by biomass recovery through centrifugation at 7500 × g and two washes with sterile water. The biomass was then resuspended in sterile water (5 mL), and 1 mL aliquots were transferred to cryovials containing 30 % glycerol and sterile crystal spheres. The cryovials were stored at 4 °C until use (Vega-Vidaurri *et al*., 2022). Prior to EPS production, the conserved *B. thuringiensis* HA1 strain was reactivated.

Exopolysaccharide production

EPS production was carried out using commercial sucrose under submerged culture conditions in 120 mL glass vessels of identical shape. The kinetics of EPS production were evaluated in a formulated medium based on commonly reported components (g L^{-1}): commercial sucrose (100), yeast extract (5), sodium chloride (5), dipotassium phosphate (1), and magnesium sulphide (0.2). The culture conditions were established at an initial pH of 7.5 using 0.1 M sodium phosphate buffer, an inoculum concentration of 1 × 10^6 CFU mL^{-1}, a temperature of 37 °C, and agitation at 150 rpm. Samples were collected at regular intervals of 0, 12, 24, 36, 48, 60, 72, and 84 h (Long *et al*., 2024).

One factor at a time approach for exopolysaccharide production

The influence of several factors on EPS production by *B. thuringiensis* HA1 was evaluated through a series of experiments. Temperature effects were assessed at 33, 35, 37, 39, 41, and 43 °C under an initial pH of 7. Subsequently, the effect of initial pH was evaluated at 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, and 9, using the optimal temperature determined in the previous assays (Long *et al*., 2024). The effect of carbon sources on EPS production was evaluated using commercial sucrose, molasses, and panela. Carbon source concentrations were analyzed at 50, 100, 150, 200, 250, 300, and 350 g L^{-1}, maintaining the temperature and pH conditions determined previously (Gudiña *et al*., 2022). All experiments were conducted in triplicate.

Exopolysaccharide recovery

The submerged culture was centrifuged at 7500 × g for 15 min at 4 °C, and EPSs were recovered from the supernatant by precipitation with chilled ethanol at a ratio of 1:3 (v/v). The mixture was stored overnight at -4 °C to achieve maximum precipitation. The EPSs were resuspended in distilled water and washed three times with cold acetone to remove monosaccharide residues, followed by a second precipitation with chilled ethanol. EPSs were then recovered by centrifugation at 7500 × g for 15 min and lyophilized at -50 °C for 8 h under 0.18 bar using a Labconco FreeZone 4.5 system. EPSs were quantified by dry weight (Zhang *et al*., 2025).

Sugar profiles of carbon sources by chromatographic analysis

Sugar quantification was performed using a high-resolution liquid chromatograph (HPLC) Thermo Finnigan Surveyor equipped with a Surveyor LC system, autosampler, and RI Surveyor Plus detector. Separation was achieved using a Phenomenex Rezex RNM-Carbohydrate Na^{+2} column (300 × 7.8 mm) with Milli-Q double-distilled water as the mobile phase. The column temperature was maintained at 80 °C, the flow rate at 0.4 mL min^{-1}, and the detector temperature at 37 °C. Prior to injection, samples were diluted and filtered through PHENEX PTFE Acrodisc filters (25 mm, 0.2 μm pore). Sugar concentrations were determined using calibration curves of sucrose, glucose, and fructose (Sigma-Aldrich) in the range of 200–1000 ppm. Results are reported in g L^{-1} (Gayosso-Sánchez *et al*., 2024).

Total phenolic content determination on carbon sources

The total phenolic content (TPC) of the carbon sources was spectrophotometrically determined using the Folin-Ciocalteu method, with gallic acid (Sigma-Aldrich, St. Louis, MO, USA) as the standard. Results were expressed as milligrams of gallic acid equivalents per gram (mg GAE g^{-1}). The reaction mixture was prepared by combining 250 μL of 1 N Folin-Ciocalteu reagent with 250 μL of the carbon source sample. After 8 min, 1250 μL of 7.5 % Na_2CO_3 and 480 μL of distilled water were added, and the mixture was incubated in the dark for 30 min. Absorbance was measured at 760 nm using a UV-Vis spectrophotometer (Thermo Fisher Scientific Biomate 3S UV-Vis, WI, USA) (Jiménez-Morales *et al*., 2024).

Exopolysaccharide characterization

The EPSs produced by *B. thuringiensis* HA1 under submerged culture conditions were characterized by Fourier-transform infrared spectroscopy (FT-IR) using attenuated total reflectance (ATR) with a zinc selenide crystal and a resolution of 4 cm^{-1}. Spectra were recorded in the mid-infrared region from 400 to 4000 cm^{-1}. Levan from *Erwinia herbicola* (Sigma) and inulin from *Dahlia* tubers (Sigma) were used as reference standards. The obtained spectra were processed using Origin 6.1 software (OriginLab Corporation, USA) (Zhang *et al*., 2025).

Statistical analysis

To evaluate the effect of each independent variable (temperature, initial pH, carbon source type, and concentration), differences among treatment means were analyzed for significance using analysis of variance (ANOVA) followed by Tukey's test in Minitab 17.

RESULTS AND DISCUSSION

Exopolysaccharide production

EPS production over time was initially evaluated using commercial sucrose (100 g L^{-1}). EPS synthesis began after 12 h (0.79 mg mL^{-1}) and reached a maximum at 60 h (1.77 mg mL^{-1}) (Figure 1). Biomass production also started after 12 h (0.09 mg mL^{-1}), reaching its highest value at 24 h (0.16 mg mL^{-1}), and *B. thuringiensis* HA1 growth remained constant until 72 h (0.164 mg mL^{-1}).

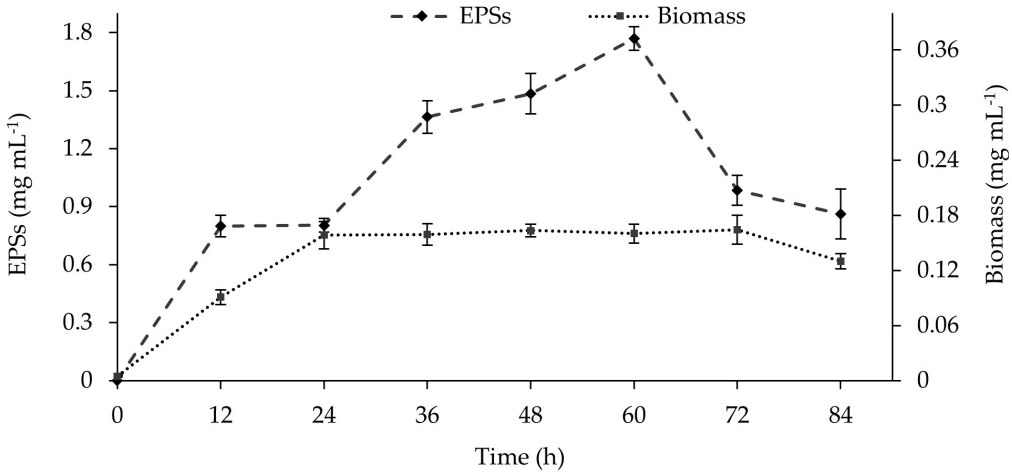

Figure 1. Kinetics of exopolysaccharide (EPS) production by *Bacillus thuringiensis* HA1 in submerged culture using commercial sucrose as carbon source (bars indicate standard error).

The EPS production obtained in this study was higher than that reported by Zhang *et al*. (2021), who observed a maximum of 0.609 g L^{-1} using *Lactobacillus paracasei* as the inoculum. Similarly, Mıdık *et al*. (2020) reported a maximum EPS production of 0.515 g L^{-1} after 120 h using *Lactobacillus plantarum* MF460, while Kumar *et al*. (2020) achieved a maximum EPS production of 1.3 mg mL^{-1} after 72 h with *Pediococcus acidilactici* NCDC 252. Song *et al*. (2021) reported EPS production of 1.62 g L^{-1} after 72 h of fermentation using *P. acidilactici* M76 and black raspberry extract as the carbon source. However, the values obtained in this study were lower than those reported by Gudiña *et al*. (2022), who achieved 6.1 g L^{-1} of EPSs after 144 h using *Rhizobium viscosum* CECT908. These differences in EPS production may be attributed to the selected strain and its

intrinsic capacity for EPS synthesis. Notably, *B. thuringiensis* HA1 achieved maximum EPS production in a shorter fermentation time compared to most reported studies.

Influence of temperature on exopolysaccharide production

The highest EPS production was observed at 37 °C (3.62 mg mL^{-1}), followed by 39 °C (3.19 mg mL^{-1}). At 33, 35, 41, and 43 °C, EPS production values were 1.14, 1.15, 1.2, and 1.12 mg mL^{-1}, respectively, while the lowest production was recorded at 31 °C (0.8 mg mL^{-1}). Biomass production reached its maximum at 33 °C (0.53 mg mL^{-1}) and was lowest at 41 and 43 °C (0.15 mg mL^{-1}) (Figure 2).

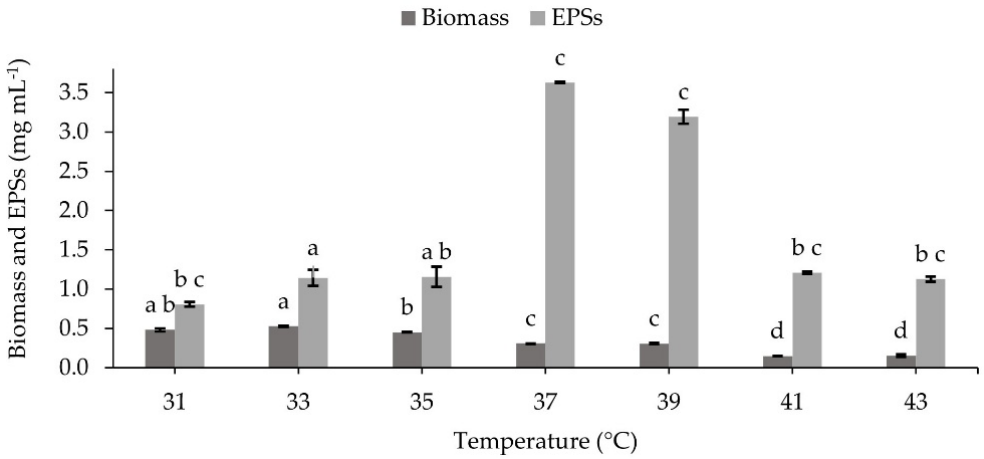

Figure 2. Effect of temperature on exopolysaccharide (EPS) production using *Bacillus thuringiensis* HA1 (bars indicate standard error).

These results were higher than those reported by Zhao and Liang (2023), who achieved a maximum of 186.77 mg L^{-1} at 37 °C using *Lactiplantibacillus plantarum* M5, but lower than those reported by Upadhyaya *et al*. (2024), who reported 9.99 g L^{-1} at 30 °C using *Bacillus tequilensis*. These results confirm that temperature is a key factor influencing EPS production and that optimal temperatures vary among bacterial strains (Upadhyaya *et al*., 2024).

Influence of pH on exopolysaccharide production

The maximum EPS production (3.14 mg mL^{-1}) was achieved at pH 7.5, whereas the lowest (0.11 mg mL^{-1}) occurred at pH 5. In contrast, maximum biomass production (1.09 mg mL^{-1}) was observed at pH 7, while the lowest (0.04 mg mL^{-1}) was also recorded at pH 5 (Figure 3). These demonstrate how pH has a direct influence on the metabolic processes involved in EPS production.

These results were higher than that reported by Cheng *et al*. (2019), who achieved a maximum of 1.5 g L^{-1} at pH 7.2 using *L. plantarum* LPC-1 as the inoculum. However,

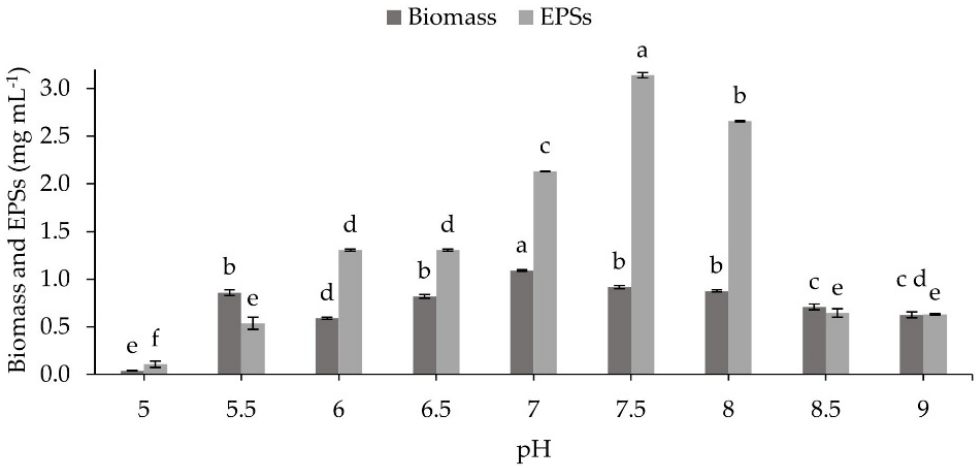

Figure 3. Effect of initial pH on exopolysaccharide (EPS) production by *Bacillus thuringiensis* HA1 (bars indicate standard error).

the values were lower than those reported by Asgher *et al*. (2021), who reached a maximum EPS production of 9.006 g L^{-1} at pH 7 using *Bacillus licheniformis* under optimized conditions. These comparisons indicate a wide variation in optimal pH for EPS production, reflecting the strain-specific capacity to adapt to different environmental conditions (Mıdık *et al*., 2020).

These conditions have an influence not only on biomass growth but also on metabolite production, directing fermentation toward two possible outcomes: programmed cell death, a genetically controlled process of self-destruction triggered by unfavorable conditions for growth, or enhanced EPS synthesis through enzyme expression, enabling biofilm formation that protects microorganisms from the fermentation environment (Ju *et al*., 2022; Naseem *et al*., 2024).

Influence of carbon sources and their concentration on exopolysaccharide production

The influence of carbon source type (commercial sucrose, molasses, and panela) and concentration on EPS and biomass production was evaluated. The maximum EPS production (23.54 mg mL^{-1}) was achieved using sucrose at 50 g L^{-1} (Figure 4). When molasses was used as the carbon source, the highest EPS production (8.62 mg mL^{-1}) occurred at 100 g L^{-1}, whereas panela yielded a maximum EPS production of 6.37 mg mL^{-1} at 50 g L^{-1}. Regarding biomass production, the highest value was observed with molasses at 250 g L^{-1} (3.87 mg mL^{-1}), followed by panela at 300 g L^{-1} (4.18 mg mL^{-1}). In contrast, sucrose resulted in a lower maximum biomass production of 1.06 mg mL^{-1} at 300 g L^{-1}.

EPS production using commercial sucrose was higher than that reported by Santos and Cruz, who obtained 19.8 g L^{-1} using *Leuconostoc mesenteroides* with sucrose as the

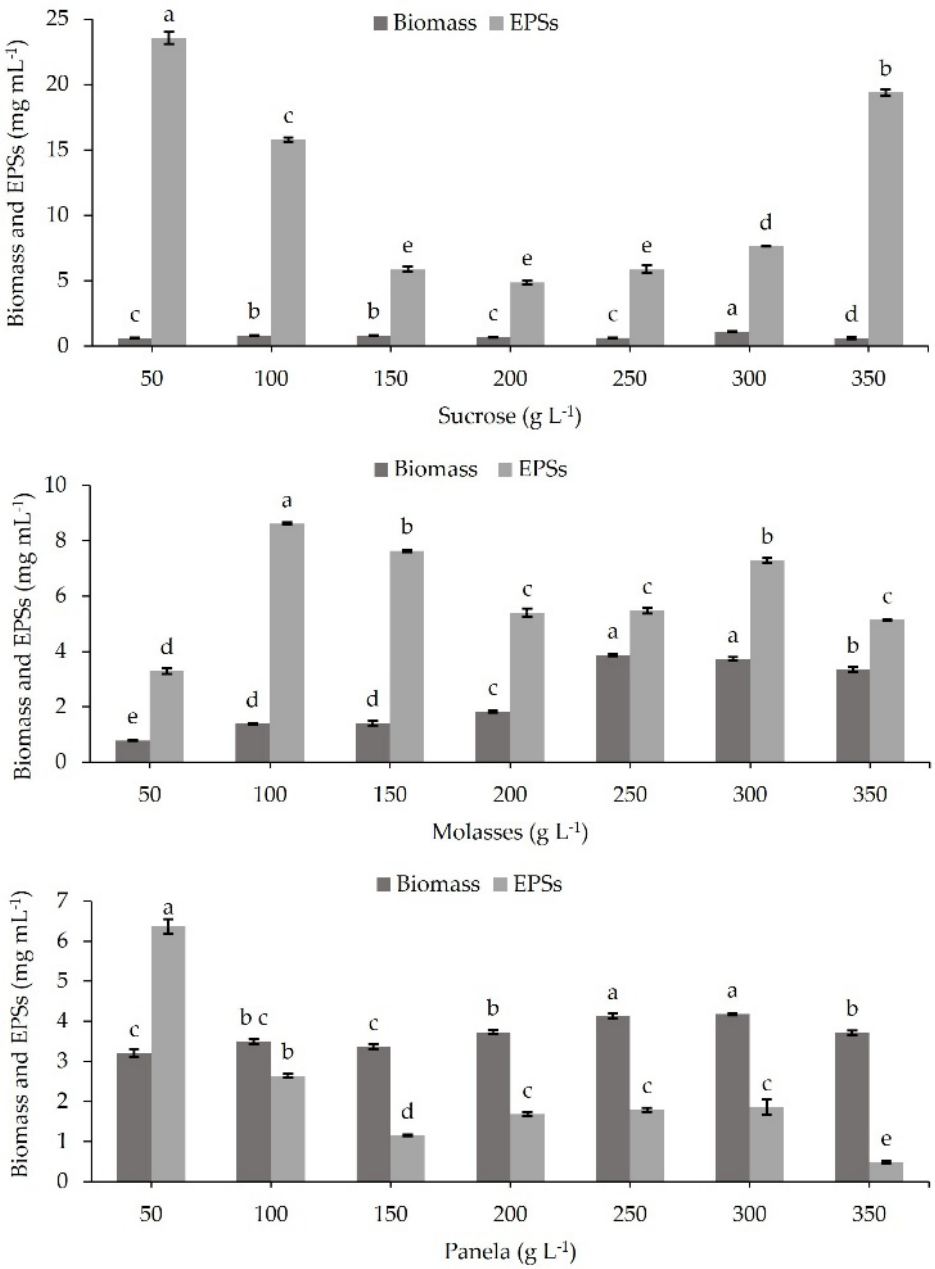

Figure 4. Effect of the carbon sources and their concentration on exopolysaccharide (EPS) production by *Bacillus thuringiensis* HA1 (bars indicate standard error).

carbon source. Similarly, Erkorkmaz *et al*. reported a maximum EPS production of 24.5 g L^{-1} using industrial sugar beet sucrose, which is comparable to the production observed in the present study without carbon source pretreatment.

Production values using molasses were higher than that reported by Asgher *et al*. (2021), who achieved 2.855 g L^{-1} using sugarcane molasses and *B. licheniformis* mutants. Under optimized conditions, the same authors reported 9.006 g L^{-1}, which is close to the maximum EPS production obtained in this study using unpretreated molasses. Gudiña *et al*. (2022) reported a maximum EPS production of 6.1 g L^{-1} using sugarcane molasses and *R. viscosum* CECT908. In contrast, Mehta *et al*. (2023) reported 18.5 g L^{-1} using pretreated sugarcane molasses and *Bacillus megaterium* KM3, while Liang *et al*. (2022) reported the highest EPS production (48.45 g L^{-1}) using sugarcane molasses and *Leuconostoc citreum* B-2. These results indicate that EPS yield is not exclusively dependent on the carbon source, as strain selection is also a determining factor in EPS production.

Overall, previous studies demonstrate that sugarcane molasses is a suitable alternative carbon source for reducing EPS production costs. For instance, Gudiña *et al*. (2022) reported that their process was approximately 30 times less expensive than a conventional production process using glucose as the carbon source.

Total phenolic contents and sugar profile determination of carbon sources

Sugar profile and total phenolic content were analyzed to evaluate the response of *B. thuringiensis* HA1 to the different carbon sources. Commercial sucrose contained 94 % disaccharide and showed no detectable phenolic compounds. In contrast, molasses consisted of 18 % sucrose, 1 % glucose, and 6 % fructose, with a phenolic content of 19.5 mg GAE g^{-1}. Panela exhibited a soluble sugar composition of 45 % sucrose, 22 % glucose, and 33 % fructose, and a phenolic content of 2.4 mg GAE g^{-1}.

The sugar profile indicated a direct correlation between sucrose concentration and EPS production, as maximum EPS synthesis was achieved using commercial sucrose. This can be attributed to the role of sucrose as a metabolic inducer, promoting the expression of enzymes involved in the polymerization of homopolysaccharides for the synthesis of α-glucans or β-fructans (Jurášková *et al*., 2022; Kaur and Dey, 2022; Yu *et al*., 2022). In contrast, the presence of glucose and fructose in molasses and panela may have negatively affected EPS production. This phenomenon may be due to monosaccharides being primarily used by bacteria as a carbon source for biomass production, rather than EPS synthesis (Bruni and Terrel, 2022). Consistently, the highest biomass production was observed when molasses and panela were used as carbon sources, suggesting a negative correlation between microbial growth and EPS production due to the redirection of metabolic flux toward primary metabolism and glycolysis (Qiu *et al*., 2023).

It was also observed that in the carbon sources where phenolic compounds were found, the production of EPSs was lower. Although polyphenols can stimulate probiotic bacterial growth (Chen *et al*., 2023), the combined presence of readily assimilable monosaccharides and phenolic compounds may explain the reduced EPS production observed when molasses and panela were used as carbon sources.

Exopolysaccharide characterization by Fourier-transform spectroscopy

Spectra obtained from EPSs produced using different carbon sources were compared with those of inulin from *Dahlia* tubers and levan from *E. herbicola* (Figure 5). The spectrum showing the highest similarity corresponded to levan from *E. herbicola*. Characteristic levan bands include the O–H stretching vibration at 3400 cm^{-1}, the C–H stretching vibration at 2900 cm^{-1}, indicating the presence of aliphatic chains, the C=O stretching vibration at 1650 cm^{-1}, the C–O–C and C–O–H stretching vibrations associated with glycosidic linkages in the 1050–1150 cm^{-1} region, and the fructofuranosyl ring vibrations at 900–950 cm^{-1}.

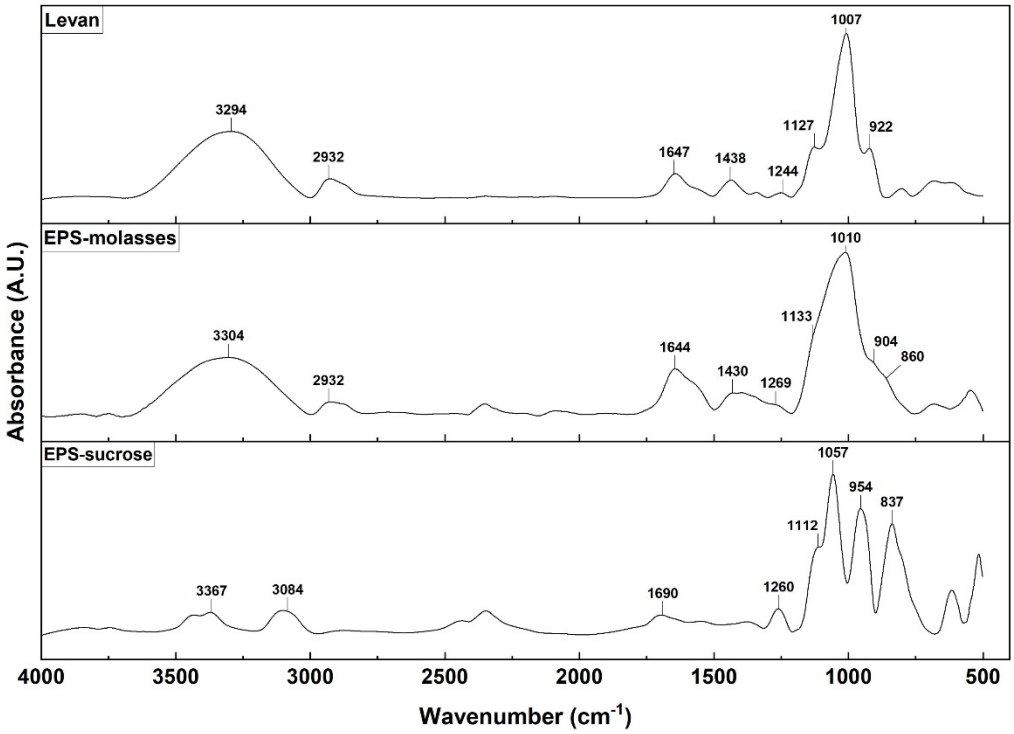

Figure 5. Fourier transform infrared spectra of the exopolysaccharides (EPS) produced by *Bacillus thuringiensis* HA1 compared to levan standard reference.

The EPS produced using molasses as the carbon source exhibited these characteristic bands, with the O–H band at 3304 cm^{-1}, the C–H band at 2932 cm^{-1}, the C=O band at 1644 cm^{-1}, the C–O–C and C–O–H stretching vibrations at 1133 and 1010 cm^{-1}, and the fructofuranosyl ring vibration at 904 cm^{-1}. These distinctive bands are consistent with those reported by Hertadi *et al*. (2020), Thakham *et al*. (2020), and Bruni and Terrell (2022), in which levan produced by *Escherichia coli* BL21 (DE3), *Bacillus siamensis*, and *Gluconobacter japonicus*, respectively, was characterized using FT-IR spectroscopy.

The EPS produced using sucrose showed an O–H band at 3367 cm^{-1}, a slightly shifted C–H band at 3084 cm^{-1}, a C=O band at 1690 cm^{-1}, C–O–C and C–O–H stretching vibrations at 1112 and 1057 cm^{-1}, a fructofuranosyl ring vibration at 954 cm^{-1}, and an additional band at 837 cm^{-1}. Although these bands correspond to an EPS, comparison with dextran and inulin standards did not reveal spectral similarity.

CONCLUSIONS

Bacillus thuringiensis HA1 exhibited remarkable metabolic flexibility, a trait of significant value for industrial biotechnology applications. This study demonstrated that the biosynthetic output of the strain can be directed by the composition of the carbon source, with the ability to produce levan when sugarcane molasses was used. In contrast, the use of commercial sucrose shifted the metabolism toward the production of a distinct glucan-like exopolysaccharide. These findings highlight both the influence of carbon source selection and the capacity of *B. thuringiensis* HA1 to valorize agro-industrial coproducts into diverse value-added products.

ACKNOWLEDGEMENTS

Jesús David Castilla-Marroquín thanks the Secretariat of Science, Humanities, Technology, and Innovation (SECIHTI), for scholarship number 858814 and to the IxM project number 331.

REFERENCES

Abarquero D, Renes E, Fresno JM, Tornadijo ME. 2021. Study of exopolysaccharides from lactic acid bacteria and their industrial applications: A review. International Journal of Food Science and Technology 57 (1): 16–26. https://doi.org/10.1111/ijfs.15227

Ahuja V, Bhatt AK, Banu JR, Kumar V, Kumar G, Yang YH, Bhatia SK. 2023. Microbial exopolysaccharide composites in biomedicine and healthcare: Trends and advances. Polymers 15 (7): 1801. https://doi.org/10.3390/polym15071801

Asgher M, Rani A, Khalid N, Qamar SA, Bilal M. 2021. Bioconversion of sugarcane molasses waste to high-value exopolysaccharides by engineered Bacillus licheniformis. Case Studies in Chemical and Environmental Engineering 3: 100084. https://doi.org/10.1016/j.cscee.2021.100084

Aziz T, Naveed M, Sarwar A, Makhdoom SI, Mughal MS, Ali U, Yang Z, Shahzad M, Sameeh MY, Alruways MW, *et al*. 2022. Functional annotation of *Lactiplantibacillus plantarum* 13-3 as a potential starter probiotic involved in the food safety of fermented products. Molecules 27 (17): 5399. https://doi.org/10.3390/molecules27175399

Balíková K, Vojtková H, Duborská E, Kim H, Matúš P, Urík M. 2022. Role of exopolysaccharides of *Pseudomonas* in heavy metal removal and other remediation strategies. Polymers 14 (20): 4253. https://doi.org/10.3390/polym14204253

Braga A, Gomes D, Amorim C, Silvério SC, Alves J, Rainha J, Cardoso BB, Rodrigues JL, Rodrigues LR. 2022. One-step production of a novel prebiotic mixture using *Zymomonas mobilis* ZM4. Biochemical Engineering Journal 183: 108443. https://doi.org/10.1016/j.bej.2022.108443

Bruni GO, Terrell E. 2022. A review on the production of C4 platform chemicals from biochemical conversion of sugar crop processing products and by-products. Fermentation 8 (5): 216. https://doi.org/10.3390/fermentation8050216

Chen X, Lan W, Xie J. 2023. Natural phenolic compounds: Antimicrobial properties, antimicrobial mechanisms, and potential utilization in the preservation of aquatic products. Food Chemistry 440: 138198. https://doi.org/10.1016/j.foodchem.2023.138198

Cheng X, Huang L, Li K. 2019. Antioxidant activity changes of exopolysaccharides with different carbon sources from *Lactobacillus plantarum* LPC-1 and its metabolomic analysis. World Journal of Microbiology and Biotechnology 35 (5). https://doi.org/10.1007/s11274-019-2645-6

Díaz-Cornejo S, Otero MC, Banerjee A, Gordillo-Fuenzalida F. 2022. Biological properties of exopolysaccharides produced by *Bacillus* spp. Microbiological Research 268: 127276. https://doi.org/10.1016/j.micres.2022.127276

Erkorkmaz BA, Kırtel O, Abaramak G, Nikerel E, Öner ET. 2022. UV and chemically induced *Halomonas smyrnensis* mutants for enhanced levan productivity. Journal of Biotechnology 356: 19–29. https://doi.org/10.1016/j.jbiotec.2022.07.005

Gayosso-Sánchez A, Hernández-Martínez R, Pacheco-López N, Herrera-Corredor J, Valdivia-Rivera S, Herrera-Pool I. 2024. Effect of the carbon-nitrogen ratio on the co-production of polyhydroxyalkanoates and exopolysaccharides by *Enterobacter soli*. Revista Mexicana de Ingeniería Química 23 (2): 1–17. https://doi.org/10.24275/rmiq/bio24211

Gudiña EJ, Couto MR, Silva SP, Coelho E, Coimbra MA, Teixeira JA, Rodrigues LR. 2022. Sustainable exopolysaccharide production by *Rhizobium viscosum* CECT908 using corn steep liquor and sugarcane molasses as sole substrates. Polymers 15 (1): 20. https://doi.org/10.3390/polym15010020

Hertadi R, Amari MMS, Ratnaningsih E. 2020. Enhancement of antioxidant activity of levan through the formation of nanoparticle systems with metal ions. Heliyon 6 (6): e04111. https://doi.org/10.1016/j.heliyon.2020.e04111

Jiménez-Morales K, Herrera-Pool E, Ayora-Talavera T, Cuevas-Bernardino J, García-Cruz U, Pech-Cohuo S, Pacheco N. 2024. Bioactive compounds preservation in functional flour production from *Cordia dodecandra* A. DC fruit: Impact of drying method and pretreatment. Revista Mexicana de Ingeniería Química 23 (2): 1–18. https://doi.org/10.24275/rmiq/alim24200

Ju Y, Shan K, Liu W, Xi C, Zhang Y, Wang W, Wang C, Cao R, Zhu W, Wang H, Zhao Y, Hao L. 2022. Effect of different initial fermentation pH on exopolysaccharides produced by *Pseudoalteromonas agarivorans* Hao 2018 and identification of key genes involved in exopolysaccharide synthesis via transcriptome analysis. Marine Drugs 20 (2): 89. https://doi.org/10.3390/md20020089

Jurášková D, Ribeiro SC, Silva CCG. 2022. Exopolysaccharides produced by lactic acid bacteria: From biosynthesis to health-promoting properties. Foods 11 (2): 156. https://doi.org/10.3390/foods11020156

Kaur N, Dey P. 2022. Bacterial exopolysaccharides as emerging bioactive macromolecules: From fundamentals to applications. Research in Microbiology 174 (4): 104024. https://doi.org/10.1016/j.resmic.2022.104024

Kumar R, Bansal P, Singh J, Dhanda S. 2020. Purification, partial structural characterization and health benefits of exopolysaccharides from potential probiotic *Pediococcus acidilactici* NCDC 252. Process Biochemistry 99: 79–86. https://doi.org/10.1016/j.procbio.2020.08.028

Liang L, Xu M, Pan L, Zhou Z, Han Y. 2022. Structural characterization of exopolysaccharide produced by *Leuconostoc citreum* B-2 cultured in molasses medium and its application in set yogurt. Processes 10 (5): 891. https://doi.org/10.3390/pr10050891

Long X, Hou X, Li S, Chen A, Zhang Z, Shen G. 2024. Fermentation optimization and in-vitro antioxidant activity of exopolysaccharides produced by *Leuconostoc suionicum* LSBM1 using sugar beet molasses. Sugar Tech 26 (5): 1405–1414. https://doi.org/10.1007/s12355-024-01396-y

Mehta K, Shukla A, Saraf M. 2023. Production kinetics and structural characterization of levan derived from *Bacillus megaterium* KM3 using pretreated cane molasses. Journal of Polymers and the Environment 32 (4): 1602–1618. https://doi.org/10.1007/s10924-023-03054-y

Mıdık F, Tokatlı M, Elmacı SB, Özçelik F. 2020. Influence of different culture conditions on exopolysaccharide production by indigenous lactic acid bacteria isolated from pickles. Archives of Microbiology 202 (4): 875–885. https://doi.org/10.1007/s00203-019-01799-6

Nadzir MM, Nurhayati RW, Idris FN, Nguyen MH. 2021. Biomedical applications of bacterial exopolysaccharides: A review. Polymers 13 (4): 530. https://doi.org/10.3390/polym13040530

Naseem M, Chaudhry AN, Jilani G, Alam T, Naz F, Ullah R, Zahoor M, Zaman S, Sohail N. 2024. Exopolysaccharide-producing bacterial cultures of *Bacillus cereus* and *Pseudomonas aeruginosa* in soil augment water retention and maize growth. Heliyon 10 (4): e26104. https://doi.org/10.1016/j.heliyon.2024.e26104

Nguyen P, Nguyen T, Bui D, Hong P, Hoang Q, Nguyen H. 2020. Exopolysaccharide production by lactic acid bacteria: the manipulation of environmental stresses for industrial applications. AIMS Microbiology 6 (4): 451–469. https://doi.org/10.3934/microbiol.2020027

Ni D, Chen Z, Tian Y, Xu W, Zhang W, Kim B, Mu W. 2022. Comprehensive utilization of sucrose resources via chemical and biotechnological processes: A review. Biotechnology Advances 60: 107990. https://doi.org/10.1016/j.biotechadv.2022.107990

Osemwegie OO, Adetunji CO, Ayeni EA, Adejobi OI, Arise RO, Nwonuma CO, Oghenekaro AO. 2020. Exopolysaccharides from bacteria and fungi: Current status and perspectives in Africa. Heliyon 6 (6): e04205. https://doi.org/10.1016/j.heliyon.2020.e04205

Pérez-Contreras S, Hernández-Rosas F, Lizardi-Jiménez MA, Herrera-Corredor JA, Baltazar-Bernal O, Ávalos-de la Cruz DA, Hernández-Martínez R. 2025. Sugarcane industry by-products: A decade of research using biotechnological approaches. Recycling 10 (4): 154. https://doi.org/10.3390/recycling10040154

Pourjafar H, Ansari F, Sadeghi A, Samakkhah SA, Jafari SM. 2022. Functional and health-promoting properties of probiotics' exopolysaccharides; isolation, characterization, and applications in the food industry. Critical Reviews in Food Science and Nutrition 63 (26): 8194–8225. https://doi.org/10.1080/10408398.2022.2047883

Qiu S, Yang A, Zeng H. 2023. Flux balance analysis-based metabolic modeling of microbial secondary metabolism: Current status and outlook. PLoS Computational Biology 19 (8):e1011391. https://doi.org/10.1371/journal.pcbi.1011391

Saadat, Y. R.; Khosroushahi, A. Y.; Gargari, B. P. A comprehensive review of anticancer, immunomodulatory and health beneficial effects of the lactic acid bacteria exopolysaccharides. *Carbohydrate Polymers* **2019**, *217*, 79–89. https://doi.org/10.1016/j.carbpol.2019.04.025.

Santos VAQ, Cruz CHG. 2017. *Zymomonas mobilis* immobilized on loofa sponge and sugarcane bagasse for levan and ethanol production using repeated batch fermentation. Brazilian Journal of Chemical Engineering 34 (2): 407–418. https://doi.org/10.1590/0104-6632.20170342s20150350

Song Y, Lee C, Lee S, Baik S. 2021. Evaluation of probiotic properties of *Pediococcus acidilactici* M76 producing functional exopolysaccharides and its lactic acid fermentation of black raspberry extract. Microorganisms 9 (7): 1364. https://doi.org/10.3390/microorganisms9071364

Thakham N, Thaweesak S, Teerakulkittipong N, Traiosot N, Kaikaew A, Lirio GA, Jangiam W. 2020. Structural characterization of functional ingredient levan synthesized by *Bacillus siamensis* isolated from traditional fermented food in Thailand. International Journal of Food Science 2020: 1–12. https://doi.org/10.1155/2020/7352484

Upadhyaya C, Patel H, Patel I, Ahir P, Upadhyaya T. 2024. Development of biological coating from novel halophilic exopolysaccharide exerting shelf-life-prolonging and biocontrol actions for post-harvest applications. Molecules 29 (3): 695. https://doi.org/10.3390/molecules29030695

Vega-Vidaurri JA, Hernández-Rosas F, Ríos-Corripio MA, Loeza-Corte JM, Rojas-López M, Hernández-Martínez R. 2022. Coproduction of polyhydroxyalkanoates and exopolysaccharide by submerged fermentation using autochthonous bacterial strains. Chemical Papers 76 (4): 2419–2429. https://doi.org/10.1007/s11696-021-02046-3

Venkatesh T, Nandhu LA, Silpa V, Dharmalingam B, Ishwarya P, Reshma M, Sajeev M, Pandiselvam R, Kothakota A. 2023. Current production strategies and sustainable approaches towards the resurgence of non-centrifugal cane sugar production – a review. Sustainable Food Technology 1 (2): 200–214. https://doi.org/10.1039/d2fb00032f

Wang Z, Zheng Y, Guo J, Lai Z, Liu J, Li N, Li Z, Gao M, Qiao X, Yang Y, *et al*. 2025. Recent advance on the production of microbial exopolysaccharide from waste materials. Bioprocess and Biosystems Engineering. https://doi.org/10.1007/s00449-025-03169-7

Waoo AA, Singh S, Pandey A, Kant G, Choure K, Amesho KT, Srivastava S. 2023. Microbial exopolysaccharides in the biomedical and pharmaceutical industries. Heliyon 9 (8): e18613. https://doi.org/10.1016/j.heliyon.2023.e18613

Wu J, Han X, Ye M, Li Y, Wang X, Zhong Q. 2022. Exopolysaccharides synthesized by lactic acid bacteria: Biosynthesis pathway, structure-function relationship, structural modification and applicability. Critical Reviews in Food Science and Nutrition 63 (24): 7043–7064. https://doi.org/10.1080/10408398.2022.2043822

Yu L, Qian Z, Ge J, Du R. 2022. Glucansucrase produced by lactic acid bacteria: Structure, properties, and applications. Fermentation 8 (11): 629. https://doi.org/10.3390/fermentation8110629

Zhang J, Hu T, Ma Y, Ma Y, Sun Q. 2025. Characterisation and bioactivity analysis of exopolysaccharides from *Lactiplantibacillus plantarum* L3. Carbohydrate Polymer Technologies and Applications 10: 100830. https://doi.org/10.1016/j.carpta.2025.100830

Zhang R, Zhou Z, Ma Y, Du K, Sun M, Zhang H, Tu H, Jiang X, Lu J, Tu L, *et al*. 2023. Production of the exopolysaccharide from *Lactiplantibacillus plantarum* YT013 under different growth conditions: Optimum parameters and mathematical analysis. International Journal of Food Properties 26 (1): 1941–1952. https://doi.org/10.1080/10942912.2023.2239518

Zhang Y, Dai X, Jin H, Man C, Jiang Y. 2021. The effect of optimized carbon source on the synthesis and composition of exopolysaccharides produced by *Lactobacillus paracasei*. Journal of Dairy Science 104 (4): 4023–4032. https://doi.org/10.3168/jds.2020-19448

Zhao X, Liang Q. 2023. Optimization, probiotic characteristics, and rheological properties of exopolysaccharides from *Lactiplantibacillus plantarum* MC5. Molecules 28 (6): 2463. https://doi.org/10.3390/molecules28062463

Zidan D, Azlan A. 2022. Non-centrifugal sugar (NCS) and health: A review on functional components and health benefits. Applied Sciences 12 (1): 460. https://doi.org/10.3390/app12010460

Agrociencia

SURVEY OF LEAFHOPPERS (HEMIPTERA: CICADELLIDAE) AND THEIR SEASONAL ABUNDANCE IN BERRY EXPORTS IN MICHOACAN, MEXICO

Laura Delia **Ortega-Arenas**[1*], Juan Andres **Lara-García**[1], Jorge Manuel **Valdez-Carrasco**[1]

[1]Colegio de Postgraduados Campus Montecillo. Postgrado en Fitosanidad-Entomología y Acarología. Carretera México-Texcoco km 36.5, Montecillo, Texcoco, State of Mexico, Mexico, C. P. 56264.
* Author for correspondence: ladeorar@colpos.mx

Citation: Ortega-Arenas LD, Lara-García JA, Valdez-Carrasco JM. 2026. Survey of leafhoppers (Hemiptera: Cicadellidae) and their seasonal abundance in berry exports in Michoacan, Mexico. **Agrociencia** 60(1): 90-105. https://doi.org/ 10.47163/agrociencia.v60i1.3483

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: June 05, 2025.
Approved: January 21, 2026.
Published in Agrociencia: January 29, 2026.

ABSTRACT

Leafhoppers are among the most threatening pests of berries. Given the pest status of leafhoppers in berry production and the magnitude and importance of berries for export in Mexico, this study was conducted to survey, identify, and determine the seasonal abundance of leafhopper fauna in commercial blackberry and blueberry orchards in Michoacan. It was hypothesized that Cicadellidae species and their abundance, collected in commercial berry orchards, will vary by crop, phenological stage, and management. The leafhoppers were collected on a monthly basis using yellow traps, vacuum, and netting techniques from June through December 2020. A total of 7512 specimens representing six subfamilies, 18 tribes, 35 genera, and 45 species of leafhoppers were identified. Overall, a higher number of species and specimens were predominantly captured using yellow traps in comparison to the vacuum and netting methods. The largest number of species was concentrated in the subfamilies Cicadellinae and Deltocephalinae. *Graphocephala rufimargo* Walker (68 %) and *Scaphytopius nitridus* (DeLong) (5.07 %) were the predominant species. The dynamics and abundance were correlated with the fruiting periods, reduced precipitation, and management practices. Organic blackberries harbored more leafhoppers than those subjected to conventional management. The data collected in this survey, along with findings from related studies conducted in Mexico, should provide a foundation for additional research on the vector capacity of the identified species. It will also aid in developing management strategies to mitigate the risk of *Xylella fastidiosa* transmission in berry production systems.

Keywords: blueberry, blackberry, pests, fastidious bacterium vectors.

INTRODUCCIÓN

The production of berries (strawberries, raspberries, blueberries, and blackberries) is under constant threat due to pests such as leafhoppers (Hemiptera: Cicadellidae). These insects can destroy or obstruct the feeding site, extract nutrients, or transmit pathogens that cause serious diseases (Hail *et al*., 2010; Pérez-Mejía *et al*., 2020; Ortega-Arenas *et al*., 2022). Leafhoppers interact with *Xylella fastidiosa* (Wells), a fastidious

bacterium that poses a significant phytosanitary risk due to its broad host range, which includes blackberry (Elbeaino *et al*., 2014) and blueberry (di Genova *et al*., 2020). Several diseases, including Pierce's disease in grapevines, citrus variegated chlorosis, and leaf scorch in coffee, are caused by *X. fastidiosa* (Hail *et al*., 2010). This bacterium is transmitted by insect vectors, mainly leafhoppers (Cicadellidae) and spittlebugs (Cercopidae) (Janse and Obradovic, 2010; Camacho-Aguilar *et al*., 2019; Ortega-Arenas *et al*., 2022). To date, 39 species and 19 genera of leafhoppers capable of transmitting *X. fastidiosa* have been identified in the Americas (EFSA, 2015). In California alone, there are at least 20 vector species associated with the bacterium found on grapevines, with the most significant being *Graphocephala atropunctata* (Signoret), *Draeculacephala minerva* (Ball), and *Xyphon* (Carneocephala) *fulgidum* (Nottingham) (Redak *et al*., 2004). In Mexico, 40 species of leafhoppers associated with blueberries were recorded in the production regions of Jalisco, including *D. minerva* and *Homalodisca insolita* (Walker) (Pérez-Mejía *et al*., 2020), which are considered potential vectors of *X. fastidiosa*, but no tests were performed to verify their transmission ability. In Yucatan, the presence of *Oncometopia clarior* (Walker), *Hortensia similis* (Walker), *Phera obtusifrons* (Fowler), and *Homalodisca* sp. (Stel) as potential vectors has been documented in citrus (Blanco-Rodriguez *et al*., 2015), whereas *G. atropunctata*, *Phylaenus spumaris* (Stel) (Purcell *et al*., 2014), *Homalodisca vitripennis* (Germar) (Camacho-Aguilar *et al*., 2019), *D. minerva*, and *X. fulgidum* have been reported in grapevines (Redak *et al*., 2004), and *H. vitripennis* in blueberry (Burbank *et al*., 2020).

Considering that references in the American continent are pointing to leafhopper species as potential vectors of *X. fastidiosa* and that these could constitute a significant threat to agriculture in Mexico, it was hypothesized that these leafhoppers could be present in Mexican berry orchards and that the species collected through different methods and their abundance will vary by crop, phenological stage, and management. Therefore, this study aimed to survey, identify, and determine the seasonal abundance of leafhopper fauna in commercial blackberry and blueberry orchards in Michoacan, Mexico.

MATERIALS AND METHODS

Location of the experiment

This study was conducted from June to December 2020 in blackberry orchards under micro-tunnels located in "Tres Parajes" (19.6582 N, 102.4356 W, 1551 m altitude), "Cuatro Parajes" (19.5485 N, 102.4382 W, 1588 m altitude), and a blueberry site (19.6658 N, 102.4354 W, 1599 m altitude) within the Atapan locality, in the municipality of Los Reyes, Michoacan. The climate of the region is humid, semi-warm temperate [type (A) C(m)(f)] with summer rains and an average temperature of 22 °C (García, 1998).

In the "Tres Parajes" and "Cuatro Parajes" orchards, blackberry 'Dasha' was transplanted directly into the soil with mulch in the rows. This variety was established

under organic management in sectors 1, 3, and 4, while conventional management was applied in sector 6. Additionally, in the "Cuatro Parajes" orchard, blueberry 'Arana' was transplanted in bags with coconut substrate, utilizing a semi-hydroponic system and total mulch. This variety was established under conventional management in sector 10. The company defined the management practices, which included pruning, fertilization, and the application of both chemical and organic products for pest control.

Adult sampling

To estimate the presence and abundance of leafhoppers, samples were collected through yellow traps, vacuum, and netting from June to December 2020 and revised monthly. Five rectangular yellow traps (12.5 × 21.5 cm) covered on both sides with a light layer of poly-isobutylene-based adhesive fused at high heat (Ferommis, Mexico) were installed in each orchard. Each trap was geo-referenced and placed at the average height of the plant, fastened to the tunnel support with black raffia. The 15 traps were replaced every 30 d. Leafhoppers were vacuumed with a blower-vacuum cleaner with a 26-cc gasoline engine (Truper, Mexico), adapted to suck the insects and collect them in an internal capture net. Five vacuums were made per orchard and date, and at each point, the machine was operated for 2 min at medium speed, directing the suction to the plant and herbaceous vegetation near the base of the plant. Insect collection by netting consisted of 20 blows with an entomological net at five points on the orchard, including near the plant's base. The collected material was stored in 96 % ethyl alcohol and kept at -4°C until processing.

Species determination

Leafhoppers captured in the traps were removed by submerging them in white gasoline for 10 min, after which they were filtered and washed three times with 96 % ethyl alcohol. The specimens collected in vacuum and netting were stored in sellable bags, with 20 mL of 96 % ethyl alcohol, for transportation to the laboratory. Samples were separated and counted by morphological similarity and sex using a Leica EZ4 stereoscopic microscope (Leica Geosystems, Mexico) and preserved in 96 % ethyl alcohol.

For morphological identification, adults were mounted in triangles, and male genitalia were extracted according to Acevedo-Reyes *et al.* (2019). The mounted structures were used to identify the genus and species levels based on the taxonomic keys of Nielson (1968), DeLong and Freytag (1974), DeLong and Hamilton (1974), Freytag (1992), and Young (1977). Females were placed at the genus level. Specimens were deposited in the entomological collection of vector insects under the supervision of PhD Laura Delia Ortega-Arenas from the Phytosanitary, Entomology, and Acarology Program of the Postgraduate College.

Relative importance index

The importance of the species captured in the study was determined using the relative importance index (RII), which allowed the weighting of the presence of different taxonomic entities over time. RII values were obtained according to the following equation:

$$RII = (Ni / Nt) * (Mi / Mt) * 100$$

where *Ni* represents the number of individuals of species *i*, *Nt* denotes the total number of individuals captured across all species, *Mi* indicates the number of samples in which species *i* appears, and *Mt* refers to the total number of samples that were analyzed. According to Paradell *et al*. (2014), this index weighs the ratio of individuals of each species (*Ni*/*Nt*) by the importance it represents throughout the sampling cycle (*Mi*/*Mt*), with expected values ranging from 0 to 100. Species with RII values of ≤1, ≥1.1 and ≤5, ≥5.1 and ≤20, and ≥20.1 were considered occasional, less frequent, frequent, and dominant, respectively (Pérez-Mejía *et al*., 2020).

Data analyses

Monthly populations and their relationship with phenological stages and environmental variables were analyzed using descriptive statistics. The Pearson correlation coefficient was calculated between population density and average temperature and precipitation, using the SAS program (v. 9.0).

RESULTS AND DISCUSSION

Determination of species and relative importance

In the blackberry- and blueberry-producing regions of Atapan, Los Reyes, Michoacan, a total of 7512 specimens representing six subfamilies, 18 tribes, 35 genera, and 45 species of Cicadellidae were collected using three methods (Figures 1, 2, and 3). The Cicadellinae and Deltocephalinae subfamilies had the highest number of species. *Graphocephala rufimargo* (68 %), *Scaphytopius nitridus* (5.07 %), *Empoasca* sp. 1 (3.86 %), *Empoasca* sp. 2 (3.37 %), *Graminella cognita* (3.18 %), *G. sonora* (3.15 %), *Agallia quadripunctata* (2.76 %), and *Dalbulus maidis* (2.53 %) were the most abundant species, representing 88 % of the collected population. In general, a greater number of species and specimens were captured using yellow traps (n = 35) (86.18 %), compared to vacuum (n = 27) (9.73 %), and netting (n = 18) (4.09 %). Organic blackberries harbored more leafhoppers (73 %) than those subjected to conventional management (27 %) (Table 1).

The species found in this study were classified as three dominant species, 10 frequent species, 10 less frequent species, and 22 occasional species. The latter were represented by one genus and one to three species each (Table 1). In both organic and conventional

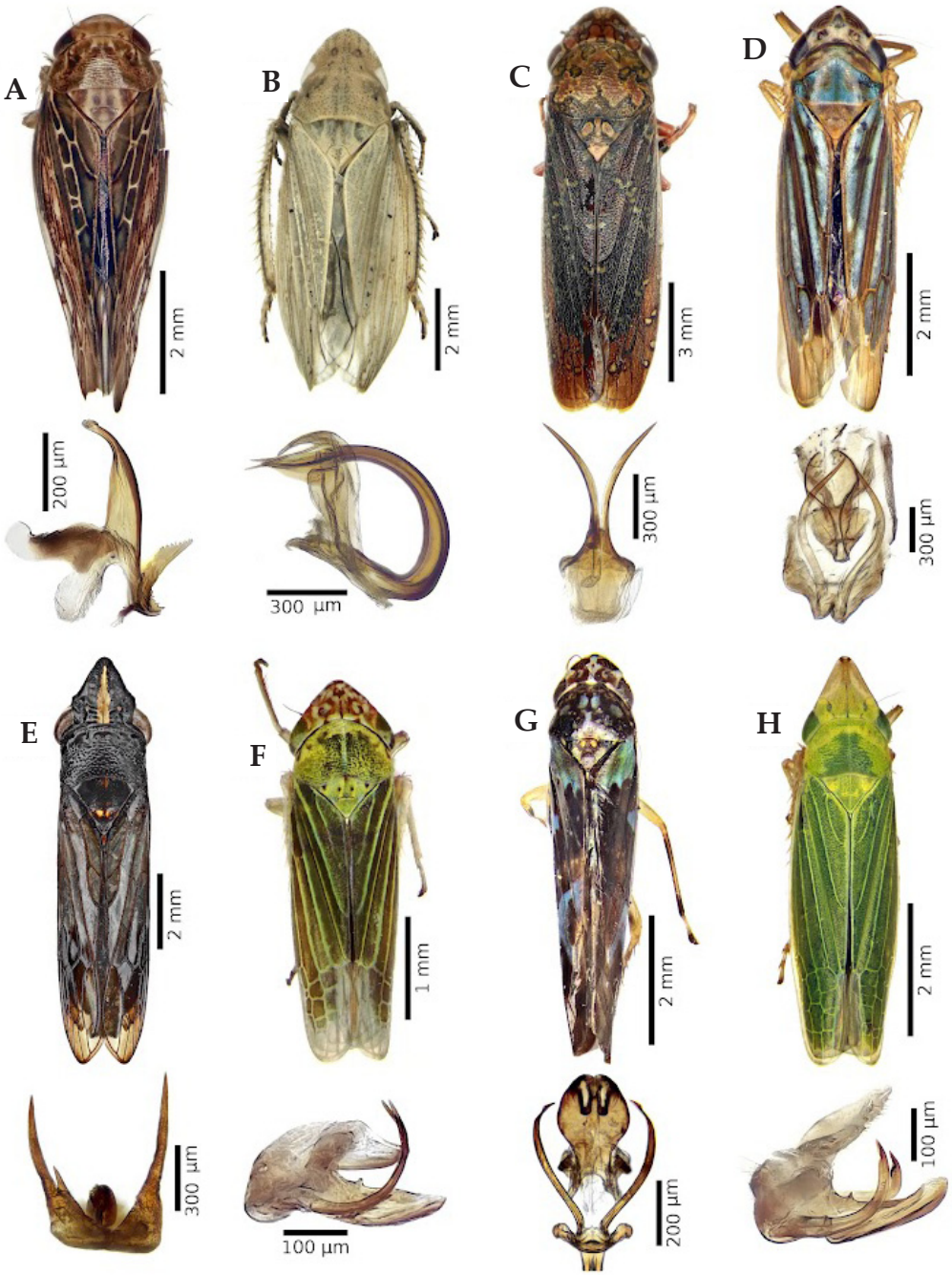

Figure 1. Cicadellidae species and male genitalia collected in blackberry and blueberry orchards in Los Reyes, Michoacan, Mexico. A: *Acinopterus angulatus*; B: *Acuera ultima*; C: *Paraulacizes figurata*; D: *Chlorogonalia coeruleovittata*; E: *Phera centrolineata*; F: *Xyphon reticulatum*; G: *Sibovia recta*; H: *Draeculacephala minerva*.

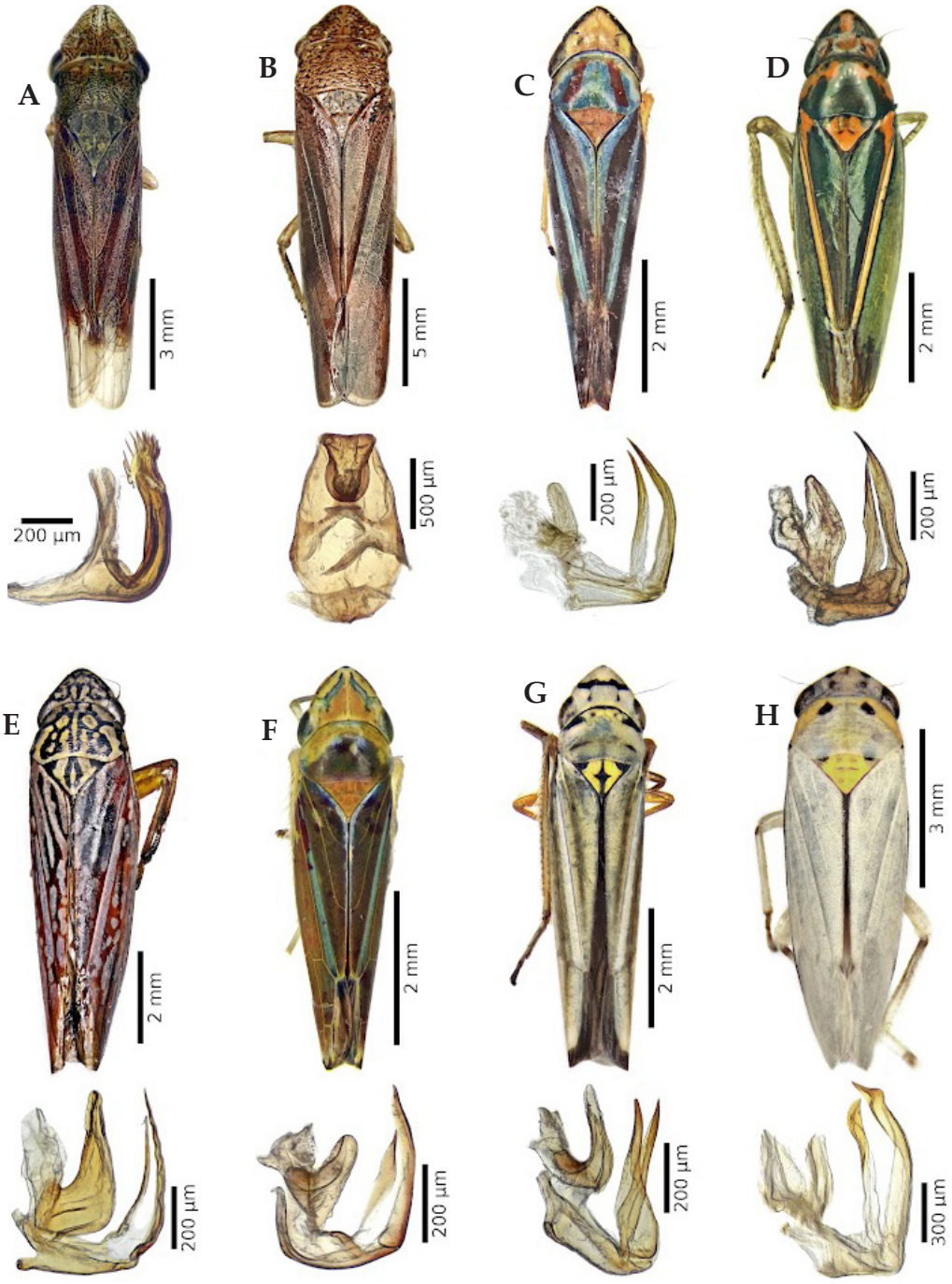

Figure 2. Cicadellidae species and male genitalia collected in blackberry and blueberry orchards in Los Reyes, Michoacan, Mexico. A: *Homalodisca insolita*; B: *Cyrtodisca major*; C: *Graphocephala fennahi*; D: *Graphocephala aurolineata*; E: *Graphocephala punctulata*; F: *Graphocephala rufimargo*; G: *Graphocephala flavovittata*; H: *Graphocephala marathonensis*.

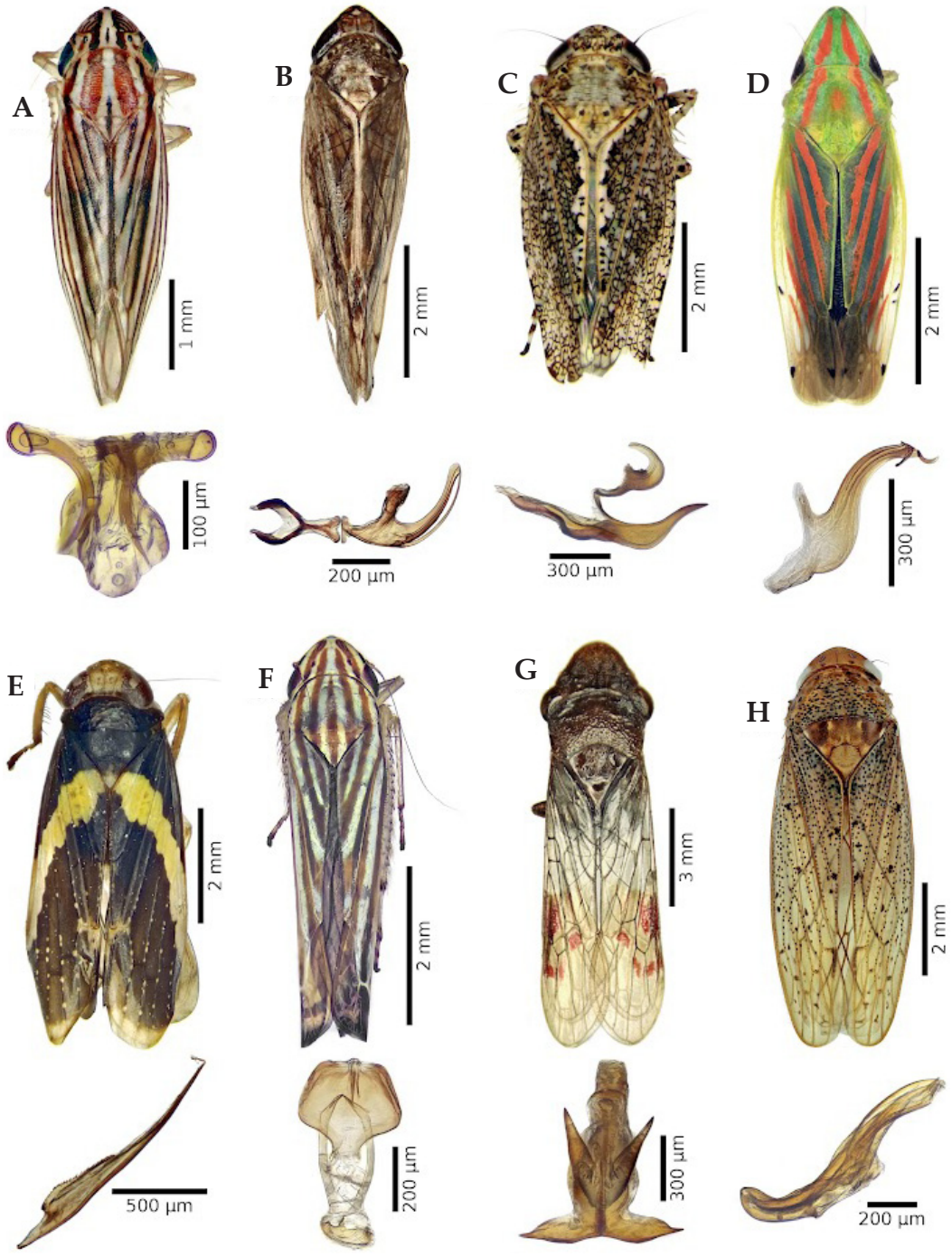

Figure 3. Cicadellidae species and male genitalia collected in blackberry and blueberry orchards in Los Reyes, Michoacan, Mexico. A: *Plesiommata mollicula*; B: *Osbornellus rarus*; C: *Texananus* sp.; D: *Spanbergiella mexicana*; E: *Omanolidia bistyla*; F: *Sibovia compta*; G: *Homalodisca ichthyocephala*; H: *Ponana woodruffi*.

Table 1. Relative importance of Cicadellidae species collected by different methods in commercial blackberry and blueberry orchards under organic and/or conventional management in Atapan, Los Reyes, Michoacan, Mexico.

Subfamilia / Tribu / Especie		Yellow traps						Vacuum						Netting					
	A	Organic blackberry		Conventional blackberry		Conventional blueberry		Organic blackberry		Conventional blackberry		Conventional blueberry		Organic blackberry		Conventional blackberry		Conventional blueberry	
		RII	C	RII	C	RII	C	RII	C	RII	C	RII	C	RII	C	RII	C	RII	C
Cicadellinae																			
Cicadellini																			
Chlorogonalia coeruleovittata	4	0	_	0	_	0	-	0.42	O	0	-	0.16	O	0	-	0	-	0	-
Dilobopterus sp	4	0.01	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Draeculacephala minerva	80	0.50	O	0.38	O	0.33	O	3.75	LF	4.49	LF	0.64	O	5.10	F	7.94	F	3.72	LF
Graphocephala aurolineata	32	0.08	O	0.73	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Graphocephala fennahi	10	0.08	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Graphocephala flavovittata	83	1.30	LF	1.31	LF	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Graphocephala marathonensis	7	0	-	0	-	1.15	LF	0	-	0	-	0	-	0	-	0	-	0	-
Graphocephala punctulata	43	0.14	O	0.73	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Graphocephala rufimargo	4726	80.7	D	48.95	D	10.91	F	1.56	LF	0.42	O	1.44	LF	0	-	0	-	0	-
Plesiommata mollicula	1	0	-	0	-	0	-	0	-	0	-	0.08	O	0	-	0	-	0	-
Sibovia compta	36	0.01	O	0.38	O	6.22	F	0	-	0	-	1.92	LF	0	-	0	-	0	-
Sibovia recta	15	0.05	O	0.19	O	0	-	0	-	0	-	0.16	O	0	-	0	-	0	-
Xyphon reticulatum	10	0.003	O	0.01	O	0.22	O	0.10	O	0.28	O	0.24	O	0	-	0	-	0	-
Proconini																			
Cyrtodisca major	2	0.01	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Homalodisca ichthyocephala	2	0.01	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Homalodisca insolita	2	0	-	0.02	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Paraulacizes figurata	6	0	-	0.07	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Phera centrolineata	119	1.83	LF	0.49	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Oncometopia sp	5	0.002	O	0.05	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Coelidiinae																			
Coelidiini																			
Omanolidia bistyla	3	0.01	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Deltocephalinae																			
Acinopterini																			
Acinopterus angulatus	5	0.003	O	0	-	0	-	0	-	0	-	0.32	O	0	-	0	-	0.35	O
Athysanini																			
Coladonus beameri	56	0.87	O	0	-	0	-	0.10	O	0.84	O	0.08	O	0.24	O	0.40	O	1.42	LF
Chiasmini																			
Exitianus picatus	61	0.01	O	0	-	0	-	7.50	F	1.97	LF	4.07	LF	4.37	LF	2.38	LF	4.26	LF
Deltocephalini																			
Amblysellus necopinus	94	0	-	0	-	2.29	LF	10	F	0.14	O	8.31	F	12.62	F	4.76	LF	8.51	F
Daltonia blacki	22	0	-	0	-	0	-	0.83	O	0	-	4.31	LF	0	-	0	-	0	-
Graminella cognita	239	0	-	0	-	0	-	25.42	D	35.96	D	23.32	D	13.59	F	16.67	F	6.91	F
Graminella sonora	237	0	-	0	-	0	-	26.67	D	25.28	D	25.88	D	18.45	F	10.71	F	6.74	F
Planicephalus flavicosta	25	0	-	0	-	0	-	4.17	LF	1.12	LF	0.96	O	0.97	O	1.59	LF	0.18	O
Sanctanus fasciatus	11	0	-	0	-	0.05	O	0	-	0.14	O	0.32	O	0	-	0.40	O	0.71	O
Hecalini																			
Spanbergiella mexicana	4	0	-	0	-	0	-	0.10	O	0	-	0.48	O	0	-	0	-	0	-
Macrostelini																			
Balclutha mexicana	5	0	-	0	-	0	-	0.21	O	0	-	0.24	O	0	-	0	-	0	-
Dalbulus maidis	190	0.57	O	6.53	F	8.72	F	0	-	0	-	0	-	3.64	LF	3.57	LF	0.71	O
Phelepsiini																			
Texananus hosanus	4	0	-	0	-	0	-	0	-	0.14	O	0.24	O	0	-	0	-	0	-
Scaphoideini																			
Osbornellus rarus	66	0.72	O	0.71	O	0.22	O	0	-	0	-	0.48	O	0	-	0	-	2.66	LF

Table 1. Continue.

Subfamilia / Tribu / Especie	A	Yellow traps						Vacuum						Netting					
		Organic blackberry		Conventional blackberry		Conventional blueberry		Organic blackberry		Conventional blackberry		Conventional blueberry		Organic blackberry		Conventional blackberry		Conventional blueberry	
		RII	C	RII	C	RII	C	RII	C	RII	C	RII	C	RII	C	RII	C	RII	C
Scaphytopiini																			
Scaphytopius nitridus	381	3.88	LF	8.49	F	12.10	F	0.42	O	0	-	0.16	O	0.24	O	0	-	0.35	O
Stenometopiini																			
Stirellus bicolor	30	0	-	0.01	O	0	-	1.04	O	0.28	O	3.83	LF	0.24	O	0.79	O	4.96	LF
Iassinae																			
Gyponini																			
Acuera ultima	3	0.01	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Ponana woodruffi	11	0.01	O	0.08	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-
Megophthalminae																			
Agallini																			
Agallia quadripunctata	207	1.09	O	4.98	LF	12.76	F	5	LF	5.06	LF	2.56	LF	4.37	LF	6.35	F	8.51	F
Typhlocybinae																			
Alebrini																			
Trypanalebra maculata	49	0	-	0	-	9.16	F	0.21	O	0	-	1.60	LF	0.97	O	0	-	0.89	O
Dikraneurini																			
Alconeura sp	34	0.15	O	0.22	O	1.64	LF	0	-	0	-	0	-	0	-	0	-	0	-
Typhlocybella minima	30	0	-	0	-	0	-	1.04	O	1.40	LF	1.60	LF	3.64	LF	2.38	LF	0.35	O
Empoascini																			
Empoasca sp 1	290	2.55	LF	6.20	F	0.65	O	1.04	O	6.18	F	0.24	O	8.74	F	11.11	F	14.18	F
Empoasca sp 2	253	0.97	O	7.92	F	3.71	LF	2.92	LF	5.06	LF	0.16	O	7.28	F	7.94	F	11.35	F
Empoasca sp 3	15	0.02	O	0.19	O	0	-	0	-	0	-	0	-	0	-	0	-	0	-

A: Total abundance of leafhoppers; RII: Relative importance index values; C: Classification by presence and absence; D: dominant species; F: frequent species; LF: less frequent species; O: occasional species; -: absent species.

blackberries, the dominant species captured in yellow traps was *G. rufimargo*, which denotes an established association with the crop. However, the dominance favored *G. sonora* and *G. cognita* in the collections made using the vacuum method. Particularly, the organically managed plantations exhibited the highest abundance of rare species.

Seasonal abundance

Leafhoppers were present throughout the study period, and the growth rates were similar across the three orchards. The highest adult infestation was recorded in orchards managed organically, as indicated by yellow traps, occurring from November to December in blackberry and from September to December in blueberry (Figure 4). This situation was observed in all three orchards and coincided with the fruiting and harvesting periods.

The population of females and males was similar in the crops studied (Figure 4), with a slight tendency in favor of males captured by yellow traps in organic blackberry. The correlation analysis indicated that the prevailing rainfall in the study area negatively

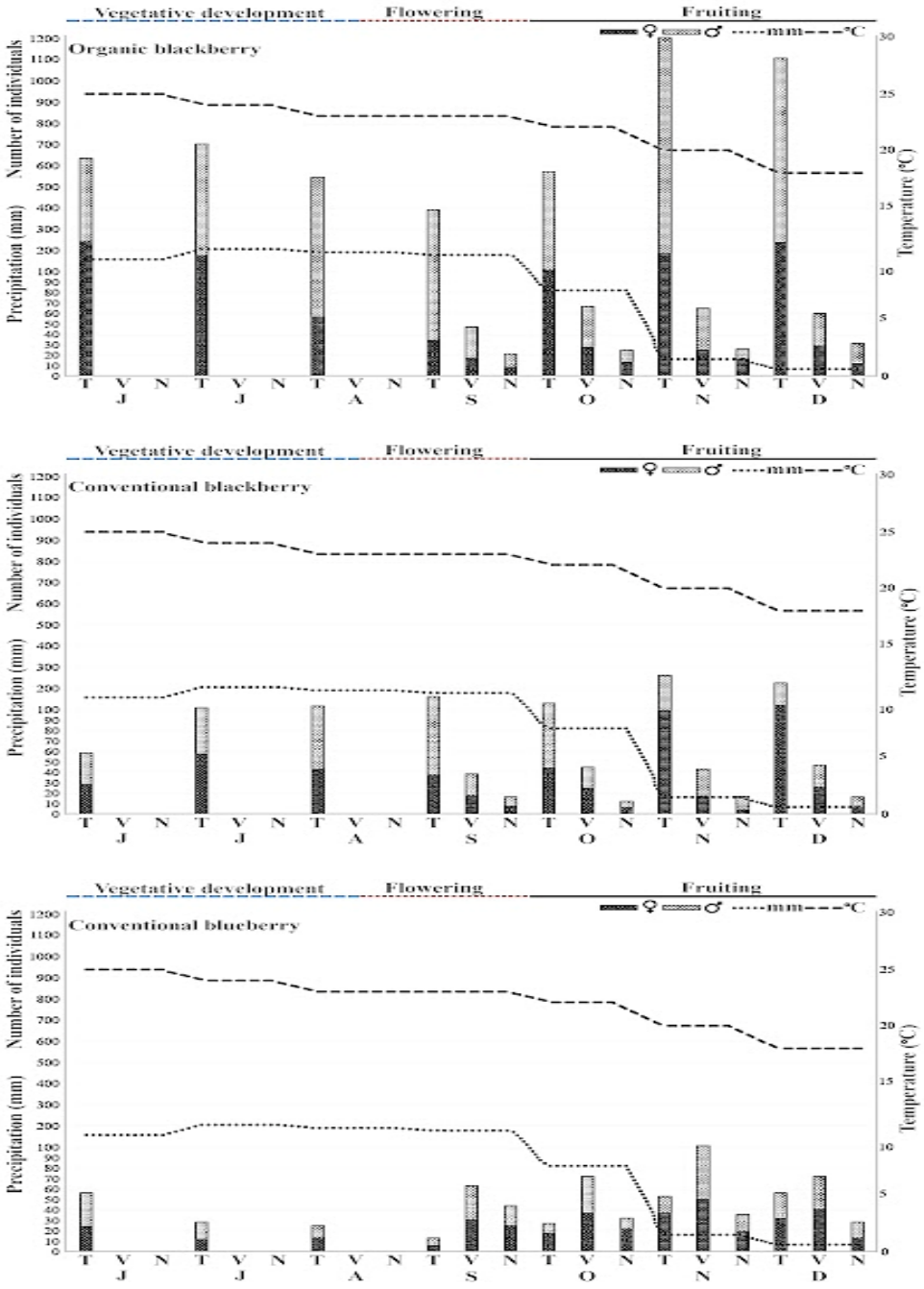

Figure 4. Seasonal abundance of leafhoppers captured by yellow traps (T), vacuum (V), and netting (N) in organically managed blackberry, conventionally managed blackberry, and conventionally managed blueberry orchards from June to December 2020 in Los Reyes, Michoacan, Mexico.

impacted the abundance of leafhoppers ($p \leq 0.05$). During periods of higher rainfall, specifically from June to August, the leafhopper population tended to decrease (Table 2).

Table 2. Pearson's correlation coefficients for temperature and precipitation on the seasonal abundance of Cicadellidae, collected from blackberry and blueberry orchards under organic and conventional management in Los Reyes, Michoacan, Mexico.

Orchard	Variable	Total leafhoppers		
		Yellow traps	Vaccum	Netting
Conventional blueberry	Precipitation	-0.639	-0.628	0.819
	Probability	0.121	0.371	0.180
	Temperature	-0.403	-362	0.7458
	Probability	0.369	0.637	0.254
Conventional blackberry	Precipitation	-0.767	-0.644	-0.092
	Probability	0.043*	0.355	0.907
	Temperature	-0.858	-0.767	-0.375
	Probability	0.013*	0.232	0.624
Organic blackberry	Precipitation	-0.817	-0.735	-0.889
	Probability	0.024*	0.264	0.110
	Temperature	-0.761	-0.387	-0.959
	Probability	0.046*	0.612	0.040*

*Statistically significant Pearson's correlations with probability of $p \leq 0.05$.

The preference of leafhoppers for yellow suggests that traps of this color are an effective and reliable method for capturing them (Krüger and Fiore, 2019). However, it is important to note that if the objective of the capture is to perform molecular analyses for species identification or to evaluate potential vectors, vacuum and netting methods are more appropriate for effectively preserving biological material (Sánchez-Romero *et al*., 2019). Additionally, the low capture rates of leafhoppers in yellow traps used in blueberry orchards can be related to the management practices of the planting system, which inhibit the growth of herbaceous vegetation between crop rows and in turn affect the capture of leafhoppers on the plants located at the edges of the orchards.

The change in species dominance among collecting methods may be due to the ability of the vacuum method to collect specimens related to grasses growing in neighboring crops (Pinedo-Escatel and Moya-Raygoza, 2018). An organic system favors the presence of herbaceous vegetation between and around the crops and acts as a harbor for the adult leafhoppers, which move towards the crop for feeding (Pérez-Mejía *et al*., 2020). For this reason, some growers consider it convenient to remove herbaceous vegetation that serves as alternative hosts for leafhoppers, while others advocate for maintaining this vegetation and planting flowering plants near crops to promote the presence of pollinators and beneficial organisms that help suppress leafhopper populations (Salas-Figueroa, 2020).

In this study, 43 of the 45 species of Cicadellidae found were reported for the first time for blackberry 'Dasha' in Mexico, and 15 were added to the list of species recorded by Pérez-Mejía *et al.* (2020) in blueberry in Jalisco. Of the 45 species recorded in this study, *D. minerva* (Lopes *et al*., 2009) and *H. insolita* (Sanderlin and Melanson, 2010) were reported as vectors of *X. fastidiosa* on grapevine and pecan, respectively. However, references are pointing to species of the genera *Graphocephala* (Ranieri *et al*., 2020), *Xyphon* (Redak *et al*., 2004), *Dilobopterus*, *Oncometopia* (Alves *et al*., 2008), *Plesiommata* (Dellapé *et al*., 2016), and *Sibovia* (Müller *et al*., 2021) in various crops as potential vectors of the bacterium.

Population peaks, especially during fruiting periods, are attributed to increased availability and concentration of nutrients in the host (Brodbeck *et al*., 1990) and to higher enzymatic metabolism in the xylem (Coudron *et al*., 2007), which define the preference and selectivity of leafhoppers. The slight trend of male populations in organic blackberry contrasts with the studies of Chen *et al*. (2010) and Pérez-Mejía *et al*. (2020), who found a greater abundance of females, which was related to prolonged feeding times and longer life expectancy during migration (Swenson, 1971; Beanland *et al*., 2000).

The prevailing temperatures in the study area (18–25 °C) were generally favorable for leafhopper development, except in organically managed blackberry plantations (van Nieuwenhove *et al*., 2016). However, correlation analyses indicated a negative relationship between leafhopper abundance and precipitation, and overall climatic variables did not exert a determining influence on population levels (Pérez-Mejía *et al*., 2020). In blueberry production systems, the use of micro-tunnels protects plants from direct rainfall and generates a favorable microclimate that facilitates leafhopper establishment (Cruz-Andres *et al*., 2018). Periodic increases in leafhopper influx may also be explained by the presence of abundant surrounding vegetation, which serves as a feeding source (Almeida and Nunney, 2015). In addition, orchard-specific management practices can modulate population abundance.

The identification of 11 potential species of Cicadellidae as vectors of *X. fastidiosa* in blackberry and blueberry crops suggests a possible risk to crop health. However, the presence of these species in orchards does not necessarily imply their active participation in the transmission of the pathogen (Weintraub and Beanland, 2006; Ortega-Arenas *et al*., 2022); therefore, further studies are required to confirm its ability to transmit *X. fastidiosa* through laboratory assays or molecular analysis. The increased presence of herbaceous vegetation in organic systems could potentially explain the higher abundance of leafhoppers in these environments. This vegetation serves as both a refuge and a food source for insect populations. However, its presence does not necessarily heighten the risk of pathogen transmission if these species are not effective vectors under such conditions.

CONCLUSIONS

A total of 7512 leafhopper specimens were identified, representing six subfamilies, 18 tribes, 35 genera, and 45 species. Yellow traps were the most effective sampling method, capturing more species and individuals than vacuum sampling and netting. Species richness was mainly concentrated in the subfamilies Cicadellinae and Deltocephalinae, with *Graphocephala rufimargo* (68 %) and *Scaphytopius nitridus* (5.07 %) as the predominant species. Leafhopper abundance was higher in organically managed blackberry plantations than in conventionally managed ones, with population dynamics closely associated with fruiting periods, reduced precipitation, temperature conditions, and management practices. These results highlight the need for complementary studies to assess the vector capacity of the detected species and to develop management strategies aimed at reducing the risk of *Xylella fastidiosa* transmission in these production systems.

ACKNOWLEDGMENTS

The authors are grateful to PhD Edith Blanco-Rodriguez for her great assistance during taxonomic species identification and to Eng. Filiberto Guerrero-Manzo for his assistance during field trials in the orchards of his property.

REFERENCES

Acevedo-Reyes N, Zetina DH, Blanco-Rodríguez E, López-Buenfil JA, Martínez-Rosas R. 2019. Méndez-Herrera Technique: New clearing technique proposed for immature stages and internal structures of adult insects. Southwestern Entomologist 44 (2): 519–522. https://doi.org/10.3958/059.044.0218

Almeida RP, Nunney L. 2015. How do plant diseases caused by *Xylella fastidiosa* emerge? Plant Disease 99 (11): 1457–1467. https://doi.org/10.1094/pdis-02-15-0159-fe

Alves E, Leite B, Marucci RC, Pascholati SF, Lopes JRS, Andersen PC. 2008. Retention sites for *Xylella fastidiosa* in four sharpshooter vectors (Hemiptera: Cicadellidae) analyzed by scanning electron microscopy. Current Microbiology 56 (5): 531–538. https://doi.org/10.1007/s00284-008-9119-7

Beanland L, Hoy CW, Miller SA, Nault LR. 2000. Influence of aster yellows phytoplasma on the fitness of aster leafhopper (Homoptera: Cicadellidae). Annals of the Entomological Society of America 93 (2): 271–276. https://doi.org/10.1603/0013-8746(2000)093[0271:ioaypo]2.0.co;2

Blanco-Rodríguez E, Romero-Nápoles J, Lomelí-Flores R, Mora-Aguilera G, Dietrich C. 2015. Leafhoppers associated with citrus orchards in the Peninsula of Yucatan, Mexico. Entomología Mexicana 2 (1): 830–834. https://doi.org/10.3958/059.047.0108

Brodbeck BV, Mizell RF, French WJ, Andersen PC, Aldrich JH. 1990. Amino acids as determinants of host preference for the xylem feeding leafhopper, *Homalodisca coagulata* (Homoptera: Cicadellidae). Oecologia 83 (1): 338–345. https://doi.org/10.1007/bf00317557

Burbank LP, Sisterson MS, O'Leary M. 2020. Infection of blueberry cultivar 'Esmeral' with a California Pierce´s disease strain of *Xylella fastidiosa* and acquisition by glassy-winged sharpshooter. Plant Disease 104 (1): 154–160. https://doi.org/10.1094/pdis-05-19-1126-re

Camacho-Aguilar II, Hernández-Castillo FD, González-Gallegos E, Blanco-Rodríguez E, Flores-Olivas A, García-Martínez O. 2019. Host and vectors of *Xylella fastidiosa* in Parras, Coahuila vineyards, Mexico. Revista Bio Ciencias 6: e413. https://doi.org/10.15741/revbio.06.e413

Chen WL, Leopold RA, Boetel MA. 2010. Host plant effects on development and reproduction of the glassy-winged sharpshooter, *Homalodisca vitripennis* (Homoptera: Cicadellidae). Environmental Entomology 39 (5): 1545–1553. https://doi.org/10.1603/EN10098

Coudron TA, Brandt SL, Hunter WB. 2007. Molecular profiling of proteolytic and pectin transcripts in *Homalodisca vitripennis* (Hemiptera: Auchenorrhyncha: Cicadellidae) feeding on sunflower and cowpea. Archives of Insect Biochemistry and Physiology 66 (2): 76–88. https://doi.org/10.1002/arch.20200

Cruz-Andres OR, Pérez-Herrera A, Martínez-Gutiérrez GA, Morales I. 2018. Macro tunnels coverings and their effect on the nutraceutical properties of "Chile de agua". Revista Fitotecnia Mexicana 41 (4a): 555–558.

Dellapé G, Paradell S, Semorile L, Delfederico L. 2016. Potential vectors of *Xylella fastidiosa*: A study of leafhoppers and treehoppers in citrus agroecosystems affected by Citrus Variegated Chlorosis. Entomologia Experimentalis et Applicata 161 (2): 92–103. https://doi.org/10.1111/eea.12491

DeLong DM, Freytag PH. 1974. Studies of the Gyponinae: The genus *Acuera* (Homoptera: Cicadellidae). The Ohio Journal of Science 74 (3): 185.

DeLong DM, Hamilton KGA. 1974. The genus *Amblysellus* (Homoptera: Cicadellidae): A key to the known species, with descriptions of eight new species. The Canadian Entomologist 106 (8): 841–849. https://doi.org/10.4039/ent106841-8

di Genova D, Lewis KJ, Oliver JE. 2020. Natural infection of southern highbush blueberry (*Vaccinium corymbosum* interspecific hybrids) by *Xylella fastidiosa* subsp. *fastidiosa*. Plant Disease 104 (10): 2595–2605. https://doi.org/10.1094/pdis-11-19-2477-re

EFSA (European Food Safety Authority). 2015. Scientific opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction options. EFSA Journal 13 (1). https://doi.org/10.2903/j.efsa.2015.3989

Elbeaino T, Yassen T, Valentini F, Ben Moussa IE, Mazzoni V, D'onghia M. 2014. Identification of three potential insect vectors of *Xylella fastidiosa* in southern Italy. Phytopathologia Mediterranea 53 (2): 328–332. https://doi.org/10.14601/Phytopathol_Mediterr-14113

Freytag PH. 1992. Two new species of *Acinopterus* (Homoptera: Cicadellidae) from Central America. Journal of Kansas Entomological Society 65 (4): 459–461.

García E. 1998. Modificaciones al Sistema de Clasificación Climática de Köppen. Universidad Autónoma de México. Instituto de Geografía. Ciudad de México, México. 90 p.

Hail D, Mitchell F, Lauzière I, Marshall P, Brady J, Bextine B. 2010. Detection and analysis of the bacterium, *Xylella fastidiosa*, in glassy-winged sharpshooter, *Homalodisca vitripennis*, populations in Texas. Journal of Insect Science 10 (1): 1–11. https://doi.org/10.1673/031.010.14128

Janse JD, Obradovic A. 2010. *Xylella fastidiosa*: Its biology, diagnosis, control and risks. Journal of Plant Pathology 92 (1): 35–48.

Krüger K, Fiore N. 2019. Sampling methods for leafhopper, planthopper, and psyllid vectors. *In* Musetti R, Pagliari L. (eds.), Methods in Molecular Biology. Humana Press: New York, NY, USA, pp: 37–52. https://doi.org/10.1007/978-1-4939-8837-2_4

Lopes JRS, Daugherty MP, Almeida RPP. 2009. Context-dependent transmission of a generalist plant pathogen: Host species and pathogen strain mediate insect vector competence.

Entomologia Experimentalis et Applicata 131 (2): 216–224. https://doi.org/10.1111/j.1570-7458.2009.00847.x

Müller C, Esteves MB, Kleina HT, Nondillo A, Botton M, Lopes JRS. 2021. First sharpshooter species proven as vectors of *Xylella fastidiosa* subsp. *multiplex* in *Prunus salicina* trees in Brazil. Tropical Plant Pathology 46 (3): 386–391. https://doi.org/10.1007/s40858-021-00430-8

Nielson MW. 1968. The leafhopper vectors of phytopathogenic virus (Homoptera, Cicadellidae), taxonomy, biology and virus transmission. Technical Bulletin 1382. Agricultural Research Service. Washington, DC, USA. 368 p.

Ortega-Arenas LD, Blanco-Rodríguez E, Pinedo-Escatel JA, Aranda-Ocampo S. 2022. Chicharritas y la enfermedad de Pierce. *In* Infante GS (ed.), Insectos y Ácaros Vectores de Fitopatógenos. México, pp: 127–156.

Paradell S, Defea B, Dughetti A, Zarate A, Remes LMM. 2014. Diversity of Auchenorrhyncha (Hemiptera: Cicadellidae: Delphacidae) associated with *Vicia villosa* in Southern Buenos Aires Province, Argentina. Florida Entomologist 97 (2): 674–684. https://doi.org/10.1653/024.097.0247

Pérez-Mejía FA, Ortega-Arenas LD, Bautista-Martínez N, Blanco-Rodríguez E, López-Buenfil JA. 2020. Leafhoppers associated with blueberry in Jalisco, Mexico. Southwestern Entomologist 45 (1): 275–288. https://doi.org/10.3958/059.045.0129

Pinedo-Escatel JA, Moya-Raygoza G. 2018. Diversity of leafhoppers (Hemiptera: Cicadellidae) associated with border grasses and maize during the wet and dry seasons in Mexico. Environmental Entomology 47 (2): 282–291. https://doi.org/10.1093/ee/nvx204

Purcell AH, Porcelli F, Cornara D, Bosco D, Picciau L. 2014. Characteristics and identification of xylem-sap feeders. Regione Puglia: Bari, Italy. 27 p.

Ranieri E, Zitti G, Riolo P, Isidoro N, Ruschioni S, Brocchini M, Almeida RPP. 2020. Fluid dynamics in the functional foregut of xylem-sap feeding insects: A comparative study of two *Xylella fastidiosa* vectors. Journal of Insect Physiology 120: 103995. https://doi.org/10.1016/j.jinsphys.2019.103995

Redak RA, Purcell AH, Lopes JRS, Blua MJ, Mizel RF, Andersen PC. 2004. The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology. Annual Review of Entomology 49 (1): 243–270. https://doi.org/10.1146/annurev.ento.49.061802.123403

Salas-Figueroa CA. 2020. Mix florales y su contribución al control biológico. Revista Red Agrícola. https://doi.org/10.13140/rg.2.2.22084.76160

Sánchez-Romero MI, García-Lechuz Moya JM, González-López JJ, Orta-Mira N. 2019. Collection, transport and general processing of clinical specimens in microbiology laboratory. Enfermedades Infecciosas y Microbiología Clínica 37 (2): 127–134. https://doi.org/10.1016/j.eimc.2017.12.002

Sanderlin RS, Melanson RA. 2010. Insect transmission of *Xylella fastidiosa* to Pecan. Plant Disease 94 (4): 465–470. https://doi.org/10.1094/pdis-94-4-0465

Swenson KG. 1971. Relation of age, sex, and mating of *Macrosteles fascifrons* to transmission of aster yellows. Phytopathology 61 (6): 657–659. https://doi.org/10.1094/Phyto-61-657

van Nieuwenhove GA, Frias EA, Virla EG. 2016. Effects of temperature on the development, performance and fitness of the corn leafhopper *Dalbulus maidis* (DeLong) (Hemiptera: Cicadellidae): implications on its distributions under climate change. Agricultural and Forest Entomology 18 (1): 1–10. https://doi.org/10.1111/afe.12118

Weintraub PG, Beanland L. 2006. Insect vectors of phytoplasms. Annual Review of Entomology 51 (1): 91–111. https://doi.org/10.1146/annurev.ento.51.110104.151039

Young DA. 1977. Taxonomic study of the Cicadellinae (Homoptera: Cicadelllidae) Part 2. New World Cicadellini and the genus *Cicadella*. North Carolina Agricultural Experiment Station: Raleigh, NC, USA. 1135 p.

Agrociencia

ANALYSIS OF BIOECONOMY ASPECTS IN AGRICULTURE AND BIOLOGICAL SCIENCES WITHIN AN INTERNATIONAL CONTEXT

Artemio **Martínez-Jazmin**[1], Roselia **Servín-Juárez**[1], Dora Angélica **Ávalos-de la Cruz**[1], José Luis **Spinoso-Castillo**[1], Venancio **Cuevas-Reyes**[2*]

[1]Colegio de Postgraduados Campus Córdoba. Carretera Córdoba-Veracruz km 348, Manuel León, Amatlán de los Reyes, Veracruz, Mexico. C. P. 94953.
[2]Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Campo Experimental Valle de México. Carretera Texcoco-Los Reyes km 13.5, Coatlinchán, Texcoco, State of Mexico, Mexico. C. P. 56250.
* Author for correspondence: cuevas.venancio@gmail.com

Citation: Martínez-Jazmín A, Servín-Juárez R, Ávalos-de la Cruz DA, Spinoso-Castillo JL, Cuevas-Reyes V. 2026. Analysis of bioeconomy aspects in agriculture and biological sciences within an international context. **Agrociencia** 60(1): 106-124. https://doi.org/ 10.47163/agrociencia.v60i1.3254

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: June 22, 2025.
Approved: January 21, 2026.
Published in Agrociencia: February 04, 2026.

ABSTRACT

The objective of this research was to identify key aspects of the bioeconomy by examining multiple international studies, particularly in the areas of agriculture and biological sciences from 2008 to 2023. A bibliographic source analysis was conducted using Bibliometrix tools from R Studio and VOSviewer to analyze a database extracted from Scopus. During this period, the bioeconomy experienced significant growth in published research and its increasing relevance to the global scientific community. The number of citations and articles reflects the impact of bioeconomy research in academia, with countries such as Finland, Germany, and Italy standing out for their publication volume. The study identified three main categories defining current trends in the bioeconomy: sustainable development; forestry and production; and innovation, biomass, and biotechnology. There is a global pursuit of an environmentally friendly economic model. Therefore, the identified areas can inform future research and contribute to the development of public policies for specific contexts and the advancement of the bioeconomy.

Keywords: trends, circular economy, sustainable development, biotechnology.

INTRODUCTION

Biology and economics converge when addressing socioeconomic and environmental problems, giving rise to the concept of bioeconomy, which highlights the economic value of natural resources (Wang *et al*., 2022). The academic and scientific community has focused on developing an economy based on biological principles, which includes revaluing natural resources previously considered waste and unlocking their potential through an integrated and highly efficient approach. This makes it possible to reduce dependence on fossil fuels and mitigate the adverse effects of climate change (Carbonell *et al*., 2021). Contrary to the classical economic model that exploits agro-ecosystems for food and energy, the bioeconomic model seeks to reduce biological loss and pollution resulting from agri-food processes (Mougenot and Doussoulin, 2022).

This research summarizes the findings on the bioeconomy from various international bodies, such as the 2018 Global Bioeconomy Summit (GBS), which explored the concept of a bioeconomy through the lens of a sustainable economy that respects nature, generates economic benefits, and aligns with the Sustainable Development Goals. Similarly, the Economic Commission for Latin America and the Caribbean (ECLAC) highlights four pillars for the bioeconomy: sustainable development, action on climate change, social inclusion, and value-added innovations. The European Union views the bioeconomy as a crucial driver of green economic growth and the adoption of new technologies.

The bioeconomy concept is internationally relevant, yet its definitions vary by country or organization. This variation allows for the identification of common elements among these definitions and various international research studies. Analyzing these elements is essential for understanding the bioeconomy's scope and potential, as well as for developing effective strategies for its implementation to foster a more sustainable future (IICA, 2020). However, previous studies indicate that in Mexico, the momentum behind policies promoting the bioeconomy remains fragile (de la Cruz-Borrego and Caballero-Rico, 2021).

The objective of this research was to identify the key aspects of the bioeconomy by examining various international studies, particularly focusing on agriculture and biological sciences from 2008 to 2023. This timeframe was chosen because it marks the beginning of a notable increase in publications on the subject. The research hypothesizes that there are common analytical themes at the international level that have influenced the development of the bioeconomy.

MATERIALS AND METHODS

The analysis of bibliographic sources was carried out using two open-access programs: 1) VOSviewer, which is a software tool for constructing and visualizing bibliometric networks, and 2) Bibliometrix, which is an R Studio tool used to perform analyses. The analysis covered the period from April 2008 to April 2023. Data was collected from bibliographic records, which included authors, document titles, publication years, source titles, volume numbers and pages, citation counts, source and document types, publication stages, DOIs, and open access status. Additionally, bibliographic reports were generated, providing information on affiliations, series identifiers, PubMed IDs, publishers, languages, correspondence addresses, abbreviated source titles, abstracts, keywords, and indexed keywords for the documents. Funding details included information such as funding numbers, acronyms, sponsors, and funding texts.

Elsevier's Scopus was selected as the search engine due to its extensive coverage, which spans a wide array of disciplines and publications on a global scale. Other databases, such as Dimensions, OpenAlex, PubMed, and Lens, were not selected for specific reasons. For instance, PubMed is more focused on biomedical fields, while Lens and Dimensions, despite their broad scope, contain significant amounts of patent data and gray literature that do not align with the objectives of this study. Additionally,

OpenAlex is still in the early stages of adoption and lacks the comprehensive bibliometric tools that Scopus provides.

In April 2023, a search was conducted in the Scopus database using the keyword "bioeconomy" in article titles, limited to the fields of agricultural and biological sciences. This initial search identified 1792 unique documents. The dataset was then refined by narrowing the search to the subject area of agricultural and biological sciences, resulting in a total of 412 documents, which included articles, books, book chapters, and conference abstracts. From this collection, 184 articles were selected for analysis using the VOSviewer and Bibliometrix tools.

Analysis with Bibliometrix

The Bibliometrix package (http://www.bibliometrix.org) provides a comprehensive set of tools for conducting quantitative research in bibliometrics and scientometrics, all developed in the R programming language. Selected articles were sourced from the Scopus database. After downloading, the data was imported into the working environment, where a normalization process was implemented to address inconsistencies in author names, affiliations, and keywords. This process entailed unifying author names according to specific guidelines, standardizing affiliations using controlled vocabularies, and consolidating synonyms into keywords. Additionally, duplicate records were removed through algorithms that utilized DOI, titles, dates, and numerical fields. The analysis was ultimately conducted with tools such as word clouds and relationship maps, alongside the identification of sources, countries, and the most cited authors.

Analysis with VOSviewer

The bioeconomy analysis was conducted using VOSviewer software. The selected articles were exported in CSV format for further processing. The analysis involved several stages. First, a bibliographic map was created in VOSviewer, using the keyword co-occurrence option as the unit of analysis. Next, the data were imported, and the map was configured with a minimum threshold of six co-occurrences to identify general trends in the research (van Eck and Waltman, 2014). Fifty keywords were selected to facilitate a more detailed analysis, removing those that did not contribute relevant information to the study. Ultimately, the results concentrated on visualizing the co-occurrence network and the overlap of keywords, which facilitated the identification of research patterns within the field of bioeconomy.

RESULTS AND DISCUSSION

Scientific production, countries, authors, and journals

Scientific output for the bioeconomy within the fields of agriculture and biological sciences has demonstrated a growing trend since its inception. In 2008, there was only one article related to this topic, which increased to 41 documents by 2022 (Figure 1).

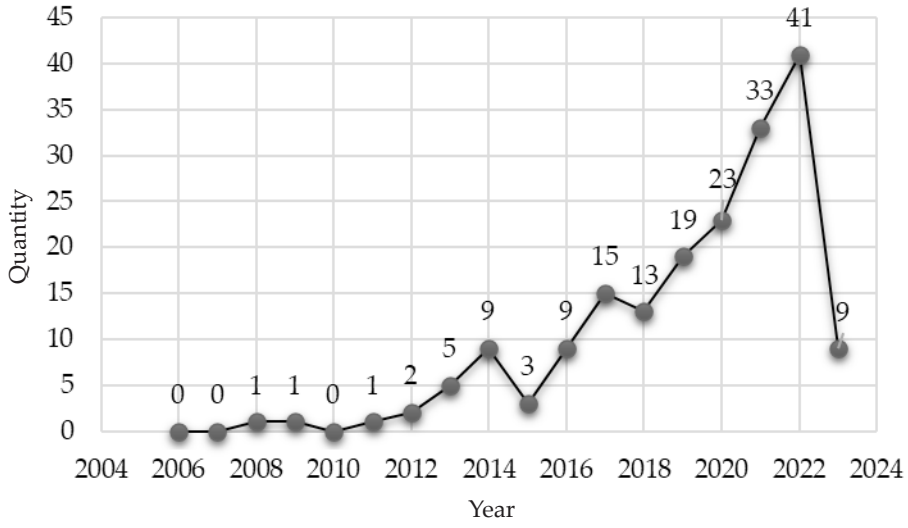

Figure 1. Scientific production of articles on bioeconomy (2008–2023) located in the Scopus database in the areas of agricultural and biological sciences.

A total of 3445 citations were found, with an average of 18.73 per article. The maximum number of citations was recorded in 2020 (675). The 10 most cited articles (Table 1) recorded a total of 1224 citations. These articles constitute 35.8 % of all scientific research in this emerging field, with an average of 122.4 citations per article.

With 187 citations, Arajo *et al*. (2021) holds the highest citation count. This research emphasizes the productive potential of algae in Europe and its role in promoting sustainable development. In second place is Dahmen *et al*. (2019), which has 179 citations and analyzes the organization of lignocellulosic biomass supply, which has great potential for the future as a regional biomass source for European lignocellulosic biorefineries. Lastly, Pülzl *et al*. (2014) examined the bioeconomy and its impact on forestry, along with classical discourses such as sustainable forest management, forest biodiversity, and climate change.

The analysis of countries with the highest citations of articles on the bioeconomy (see Figure 2) reveals that Finland leads with 772 citations, followed by Germany with 675 citations and Italy with 446 citations. A total of 54 countries have published related documents, but only two countries have published 30 or more articles (21.3 %). Seven countries fall within the range of 11 to 20 documents (30 %), while 45 countries have published 10 or fewer articles (48.5 %).

Finland and Germany are the countries with the highest number of published documents on bioeconomy (Table 2); however, Germany has more citations (772), with an average of 33.8 citations per article, compared to Finland, which has 675 citations and an average of 28.6 citations per article. This finding contrasts with Biancolillo *et*

Table 1. Most cited articles related to scientific research in bioeconomy in agriculture and biological sciences in the Scopus database.

Rank	Author (year)	Article title	Journal	Number of citations
1	Araújo *et al*. (2021)	Current status of the algae production industry in Europe: An emerging sector of the blue bioeconomy	Frontiers in Marine Science	187
2	Dahmen *et al*. (2019)	Integrated lignocellulosic value chains in a growing bioeconomy: Status quo and perspectives	GCB Bioenergy	179
3	Pülzl *et al*. (2014)	Bioeconomy - an emerging meta-discourse affecting forest discourses?	Scandinavian Journal of Forest Research	152
4	D'Amato *et al*. (2020)	Towards sustainability? Forest-based circular bioeconomy business models in Finnish SMEs	Forest Policy and Economics	146
5	Lewandowski *et al*. (2016)	Progress on optimizing miscanthus biomass production for the European bioeconomy: Results of the EU FP7 project OPTIMISC	Frontiers in Plant Science	134
6	Kleinschmit *et al*. (2014)	Shades of green: A social scientific view on bioeconomy in the forest sector	Scandinavian Journal of Forest Research	130
7	Kröger and Raitio (2017)	Finnish forest policy in the era of bioeconomy: A pathway to sustainability?	Forest Policy and Economics	83
8	Pätäri *et al*. (2016)	Global sustainability megaforces in shaping the future of the European pulp and paper industry towards a Bioeconomy	Forest Policy and Economics	80
9	van Lancker *et al*. (2016)	Managing innovation in the bioeconomy: An open innovation perspective	Biomass and Bioenergy	77
10	Toppinen *et al*. (2018)	The future of wooden multistory construction in the forest bioeconomy – A Delphi study from Finland and Sweden	Journal of Forest Economics	76

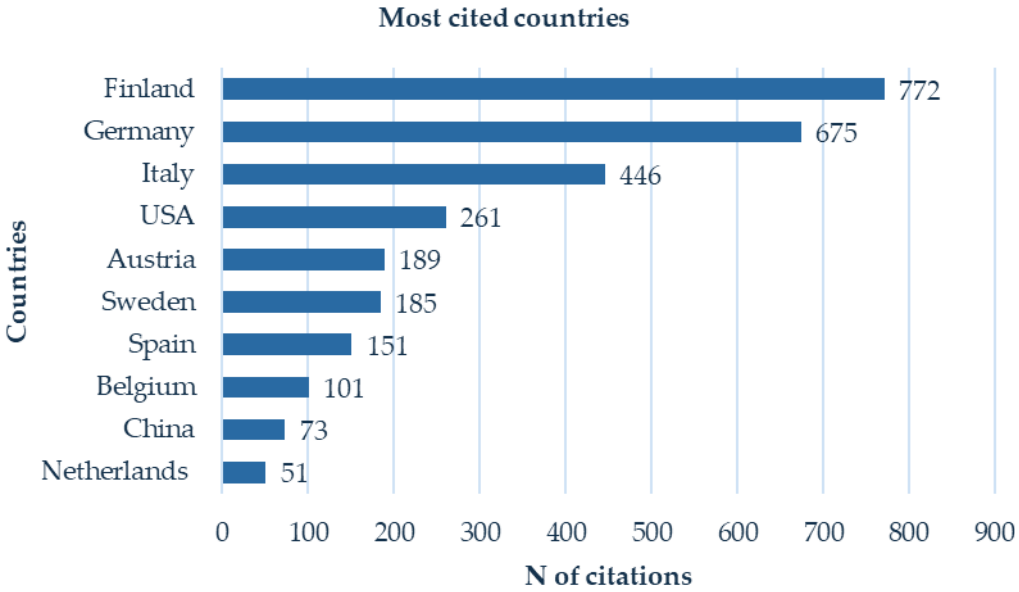

Figure 2. Countries of origin with the most citations of bioeconomy articles found in the Scopus database in the fields of agricultural and biological sciences (2008–2023).

Table 2. Average number of citations for countries with the highest number of articles published on the bioeconomy in the fields of agricultural and biological sciences (2008–2023).

Rank	Country	Number of citations	Average number of citations per article
1	Germany	772	28.6
2	Finland	675	33.8
3	Italy	446	34.4
4	USA	261	32.6
5	Austria	189	31.5
6	Sweden	185	37.0
7	Spain	151	25.2
8	Belgium	101	33.7

al. (2020), who reported that Finland and Canada have more scientific publications on bioeconomy.

As for the journals with the highest number of published documents (Table 3), Forest Policy and Economics stands out with 28 publications, followed by Forest with 13 scientific articles, and Biomass and Bioenergy with 11 documents.

Table 3. Journals publishing articles on bioeconomy in the fields of agricultural and biological sciences (2008–2023).

Rank	Journal	Number of articles	Citations	CiteScore	Impact factor	Science category	Position in the category	Quartile in category
1	Forest Policy and Economics	28	812	9.0	4.0	Forestry	4	Q1
2	Forests	13	77	4.4	2.4	Forestry	21	Q1
3	Biomass and Bioenergy	11	272	11.5	5.8	Energy and fuels	58	Q2
4	Land Use Policy	7	168	13.7	6.0	Environmental studies	25	Q1
5	GCB Bioenergy	6	229	10.3	5.9	Energy and fuels	57	Q2
6	AgBioForum	5	37			Agronomy and crop science	44	Q2
7	Bio-based and Applied Economics	5	75	2.8	1.8	Economics	241	Q2
8	Scandinavian Journal of Forest Research	5	370	3.0	1.8	Forestry	33	Q2
9	International Forestry Review	4	63		1.8	Forestry	40	Q2
10	Journal of Agricultural and Environmental Ethics	4	23	4.3	2.2	History and philosophy of science	11	Q1

Among the authors with the highest number of articles published over the last 15 years (Table 4), 10 studies by Toppinen A. stand out, followed by Korhonen J., with five publications. Toppinen *et al*. (2018) discuss sustainable development and the use of wood in construction, concluding that domestic consumers in Finland and Sweden are concerned about environmental sustainability.

Table 4. Leading authors in the publication of articles on bioeconomy in the fields of agricultural and biological sciences (2008–2023).

Rank	Author	Articles	Citations	Hirsch Index	Affiliated institution	Area of expertise
1	Anne Toppinen	10	529	49	University of Helsinki, Helsinki, Finland	Sustainability science, forest, economics, circular economy, bioeconomy and corporate responsibility
2	Jaana Elina Korhonen	7	151	18	Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA	Forest-based sector, bioeconomy, sustainability, strategic management and mixed methods
3	Daniela Kleinschmit	4	341	23	University of Freiburg, Tennenbacher Str., Germany	Forest governance, bioeconomy, policy integration, participation and political communication
4	Satu Pätäri	4	195	21	School of Business and Management, Lappeenranta University of Technology, Lappeenranta, Finland	Strategic management, corporate responsibility, firm financial and innovation performance, Delphi method
5	Dalia D'Amato	4	162	26	University of Helsinki, Helsinki, Finland	Sustainability transformations
6	Alex Giurca	4	119	18	Chair of Forest and Environmental Policy, University of Freiburg, Tennenbacher Str., Freiburg, Germany	Bioeconomy, sustainability transformations, forest governance, timber trade, environmental communication
7	Stefan Bringezu	3	141	55	Wuppertal Institute for Climate, Environment and Energy, Germany	Socio-industrial metabolism, sustainability assessment, systems analysis, footprints, resource policies
8	Alessandro Paletto	4	31	29	Council for Agricultural Research and Analysis of the Agricultural Economy (CREA), Italy; University of Florence, Italy	Environmental impact assessment, environment, sustainability, natural resource management, ecology, biodiversity ecosystem ecology, mapping conservation and water quality
9	Markus Kröger	3	89	30	Department of Political and Economic Studies, Faculty of Social Sciences, University of Helsinki, Finland	Natural resource politics, social movements, Latin America, India and the Arctic
10	Isabella de Meo	3	29	23	Council for Agricultural Research and Analysis of the Agricultural Economy (CREA), Italy; University of Florence, Italy	Coarse woody debris; dead wood; forestry, sustainable development; environmental economics; circular economy, ecosystem service, land use, natural resource

Network display

A word cloud was generated showing the most frequently occurring words with a larger diameter or thickness (Figure 3), indicating the degree of interest among researchers (defined as strands). The words "bioeconomy," "forest," and "development" are displayed in the center, with the words "sustainable" and "production" shown to a lesser extent.

Figure 3. Word cloud on the concept of bioeconomy in the fields of agricultural and biological sciences (2008–2023) from the articles analyzed, created using Bibliometrix.

Main keyword co-occurrences

The relevance of words, determined by weighted percentages, is illustrated in the following values (Figure 4): bioeconomy (2.12 %), sustainable (0.9 %), production (0.73 %), development (0.72 %), and innovation (0.67 %). The remaining 96 identified words have lower representation, ranging from 0.15 to 0.34 %, with none exceeding 260 occurrences. Konstantinis *et al*. (2018) emphasize the fundamental role of technology in advancing the bioeconomy. The branched map includes the terms "environmental" and "agriculture" as relevant to the search. Abad-Segura *et al*. (2021) also talk about the words "transition" and "governance." They stress the need for support policies to help the economy move from one that wastes resources to one that uses them efficiently.

The relevance of the words "bioeconomy," "sustainable development," "forestry biomass," and "biotechnology," obtained in the VOSviewer program, showed a significant total link strength, primarily for the first three, with a strength greater than 200 (Table 5).

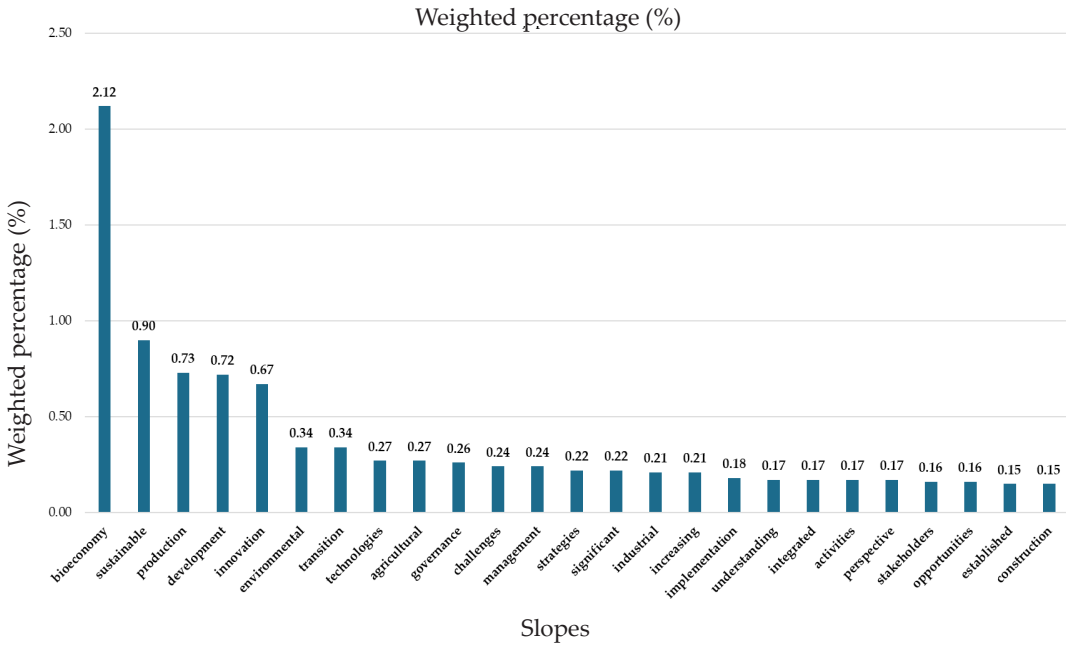

Figure 4. Co-occurrence of the top 25 keywords in bioeconomy articles in the fields of agricultural and biological sciences (2008–2023).

Table 5. Analysis of the top 25 co-occurrences of keywords from articles on bioeconomy in the fields of agricultural and biological sciences (2008–2023).

Rank	Key word	Co-occurrences	Total amplitude
1	Bioeconomy	92	318
2	Sustainable development	61	280
3	Forestry	47	239
4	Biomass	25	104
5	Biotechnology	18	69
6	Bioenergy	17	76
7	Economics	17	92
8	Circular economy	16	57
9	Climate change	16	88
10	Europe	14	79
11	Innovation	14	57
12	Biofuel	13	50
13	Circular bioeconomy	13	35
14	European union	13	78
15	Finland	13	77
16	Wood	13	69
17	Environmental economics	12	71
18	Timber	12	89
19	Commerce	11	68

Table 5. Continue.

Rank	Key word	Co-occurrences	Total amplitude
Rank	Key word	Co-occurrences	Total amplitude
20	Economic analysis	11	64
21	Forest management	11	77
22	Biodiversity	10	49
23	Development	10	56
24	Land use	10	58
25	Bio-based	9	54

Main aspects of the bioeconomy

Co-occurrence analyses reveal the relevance of the words "sustainable," "development," "forest," "production," "biomass," "innovation," and "biotechnology." However, these findings contrast with those found by Biancolillo *et al*. (2020), who identified three main concepts related to the bioeconomy: sustainable development, bioenergy production, and climate change mitigation. To contrast the analysis of keywords, a classification based on degrees was carried out (Table 6). The concept of bioeconomy is closely related to sustainable and development; in the second degree, to production and forestry; and in the third degree, to the concepts of innovation, biomass, and biotechnology.

Table 6. Classification of the main aspects of the bioeconomy in articles in the fields of agricultural and biological sciences (2008–2023).

Rank	Word	Software count
1	Bioeconomy	3
1	Sustainable	3
1	Development	3
2	Production	2
2	Forestry	2
3	Innovation	1
3	Biomass	1
3	Biotechnology	1

The evolution of the bioeconomy concept (Figure 5) shows a growing concern for climate change in 2018. In contrast, the European Union's research focus shifted in 2019 and 2020 to promoting innovation and public policies in agriculture, thereby supporting sustainable development. Finally, in 2021, the main research trends included the circular economy, development, crop production, and the creation of biorefineries.

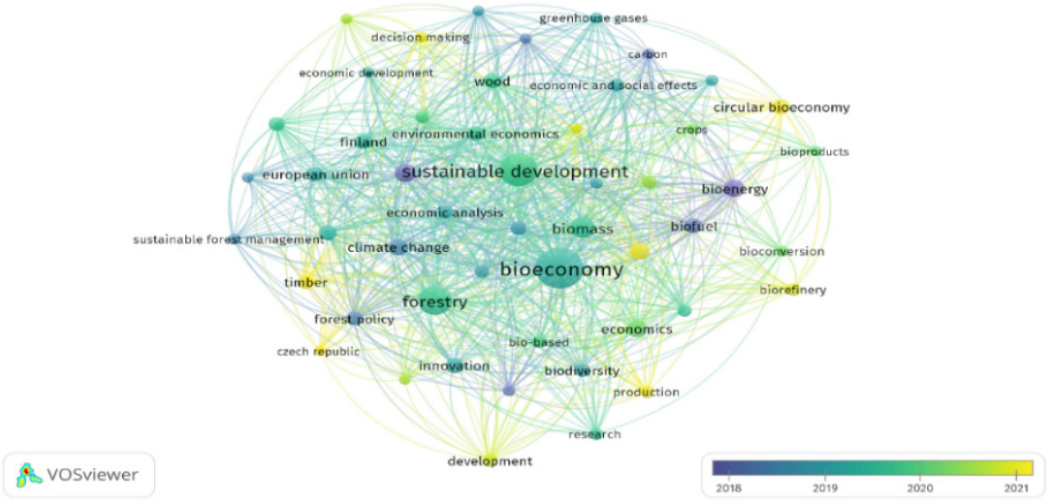

Figure 5. Overlay display of the evolution of the concept of bioeconomy from 2018 to 2021.

Analysis of the main aspects of the bioeconomy

First-degree concepts of importance

The words "sustainable" and "development" reflect the construction of the bioeconomy as a strategy that balances the sustainable conservation of natural resources with the promotion of economic development. The incorporation of sustainable practices in the use of biodiversity and natural resources guarantees the preservation of economic, environmental, and social assets for future generations, while promoting responsible economic development.

Sustainable. The pursuit of a balanced use of biological resources closely links the bioeconomy and sustainability. Given the concerns about climate change and responsible use of biological resources, it is crucial to develop models for the sustainable management of these resources (Mougenot and Doussoulin, 2022). The strategic integration of the bioeconomy and the circular economy promotes the sustainable use of natural resources through a systemic approach that enhances resource efficiency and minimizes environmental impact (Abad-Segura *et al*., 2021). The bioeconomy is an evolving economic model, characterized by varying degrees of technological progress (Papadopoulou *et al*., 2021).

Development. According to the Royal Spanish Academy (RAE, 2023), development is considered to be the "evolution of an economy towards better standards of living," while the United Nations (UN, 1997) describes it as "a multidimensional endeavor to achieve a better quality of life for all peoples."

Development in the bioeconomy is considered an essential concept in the advancement of agriculture, agribusiness, and various sectors involved in the production and utilization of bio-based raw materials (Wang *et al*., 2022). The bioeconomy helps development by creating new markets based on biological resources (Borrego and Rico, 2021), increasing value (Papadopoulou *et al*., 2021), and supporting regional development (Sanz-Hernández *et al*., 2019).

Sustainable development. The challenge of achieving sustainability while promoting economic growth and technological advancement was first introduced on a global scale in 1972, during the United Nations Conference on the Human Environment in Stockholm. It was established that "Man is entitled to freedom, equality, and the enjoyment of adequate living conditions in an environment of such quality that he can lead a dignified life and enjoy well-being, and he has the solemn obligation to protect and improve the environment for present and future generations" (UN, 1973).
The Global Bioeconomy Convention defines sustainable development as "the production, use, and conservation of biological resources, including related knowledge, science, technology, and innovation, to provide information, products, processes, and services in all economic sectors, and move toward a sustainable economy" (GBS, 2018). The bioeconomy initially emerged as a strategy to utilize biological technologies, particularly biotechnology, more effectively. However, this concept has since evolved into a broader vision centered on sustainable development, aligning with the Sustainable Development Goals (IICA, 2020).

Second-degree concepts of importance

Currently, the production and management of natural resources are crucial to global well-being, which is why the bioeconomy vision presents significant opportunities in agri-food production. Production and forestry are the two areas that hold significant importance.

Production. Humanity is confronting significant challenges, including the depletion of fossil fuel supplies, the degradation of natural resources, and the acceleration of climate change. The bioeconomic model presents a transformative approach focused on developing renewable biological resources and converting them into high-value products. This strategy aims to address these challenges sustainably and effectively (Mougenot and Doussoulin, 2022). Achieving this goal requires both interdisciplinary and multidisciplinary approaches, as well as new insights into the use of by-products and waste. This shift drives the growth of bioenergy (Biancolillo *et al*., 2020) and biofuel production, paving the way toward fulfilling the Sustainable Development Goals (Nazari *et al*., 2021).

Forestry. Silviculture is a vital discipline within forestry sciences that focuses on forest management and regeneration. It utilizes established strategies and techniques for

regenerating tree species, based on knowledge of species autoecology as well as the ecology and dynamics of forests (Bannister *et al*., 2016). This approach has resulted in the publication of 225 documents by 567 organizations across 44 countries, concentrating on the forest bioeconomy. This bioeconomy is closely associated with three key concepts: sustainable development, bioenergy generation, and climate change mitigation (Biancolillo *et al*., 2020).

Third-degree concepts of importance

Aspects related to the concepts of innovation, biomass, and biotechnology have been identified, specifically focusing on tools for transforming waste from traditional production systems, now commonly referred to as biomass. The convergence of bioeconomy with these concepts creates a dynamic environment ripe for change. Innovations in developing processes and products not only diversify sources of raw materials, thereby reducing dependence on non-renewable resources, but also foster the emergence of new industries and job opportunities.

Innovation. The concept of innovation was first articulated by the esteemed economist Joseph Alois Schumpeter in his work "The Theory of Economic Development." He suggested that economic development is shaped by research and new knowledge, which in turn leads to innovations that create new techniques and production methods (Schumpeter, 2017). The Oslo Manual further defines innovation as the introduction of a new or significantly enhanced product, service, or good; a process; a marketing strategy; or a novel organizational method within a company's internal operations (OECD, 2005).

In the face of rapid climate change, it is vitally important to rethink the way biological resources are produced, consumed, transformed, stored, recycled, and disposed of. Advances in bioeconomy research and the integration of innovative technologies will enhance the management of renewable biological resources and foster the development of new markets for food and bioproducts (López-Feria and Barrero-Domínguez, 2021).

Biomass. The European Union defines biomass as "the biodegradable fraction of products, waste, and residues from agriculture, including vegetable and animal substances" (Olmo, 2018). Several centuries ago, during the Industrial Revolution, biomass was the primary source of energy worldwide. However, today, dependence on oil is evident. Due to environmental concerns, a return to cleaner energy sources is being considered to reduce pollution. The bioeconomy model leads to a change focused on a more sustainable economy, which can be achieved by accepting and implementing circular production models. The bioeconomy is becoming increasingly important as biomass is recognized as a key resource that integrates processes and enables the creation of high-value-added products (Carbonell *et al*., 2021). Within the Mexican framework, a wide range of by-products from agriculture, livestock, fisheries, forestry, and agro-industry can be utilized in the generation of biomass for

the sustainable production of biological resources and to meet the nation's energy demands (SADER, 2019).

Biotechnology. The connection between biotechnology and bioeconomy is essential, as biotechnology enables the management of biological organisms and processes for practical purposes. The bioeconomy benefits from biotechnology to enhance production efficiency, innovate in the development of new products, and use biological resources more effectively. It is argued that advances in biotechnology have driven the emergence of the bioeconomy (Konstantinis *et al*., 2018). Conversely, the bioeconomy offers an opportunity for the development of biotechnology (de la Cruz-Borrego and Caballero-Rico, 2021) and the creation of new products to enhance the efficient use of biomass generated in production processes, thereby promoting sustainability and innovation.

Europe is mapping out a unified path in the field of bioeconomy, based on research and technological innovation in multiple biotechnology applications (from de Besi and McCormick, 2015). Globally, several countries have integrated bioeconomy into their strategic policies (Carbonell *et al*., 2021). The three areas identified show us the lines of work pursued by the bioeconomy in the field of agricultural and biological sciences, requiring strategies and policies in different geographical areas (local, state, regional, national, and global), as well as the integration of multiple actors, both public and private, to achieve sustainable and permanent development for future generations.

Political importance of bioeconomy

Bioeconomy shares a common goal with sustainable policies, which is the need to establish a regulatory framework that promotes and facilitates practices to achieve an "environmentally friendly" economic model. Kröger and Raitio (2017) analyzed forestry policy in terms of sustainability goals and how to harmonize different dimensions. Bioeconomy emerges as a pillar in the search for sustainable and efficient economic growth to transition to a society with a strong dependence on renewable biological resources (Kleinschmit *et al*., 2014; Lombeyda, 2020). This scenario stipulates that related public policies should foster interdisciplinary approaches to study, innovation, and knowledge generation (Paltaki *et al*., 2021). The bioeconomy is becoming the central axis of progress, with continuous growth, which has already been adopted in the national development plans of several countries (Sanz-Hernández *et al*., 2019).

In this context, Mexico shows a degree of weakness in promoting public policies that encourage bioeconomy, as well as in making informed decisions regarding products derived from forest resources. At this turning point, it is crucial to capitalize on the opportunities presented by Mexico to develop a new economic model that aligns with environmental sustainability (de la Cruz-Borrego and Caballero-Rico, 2021).

CONCLUSIONS

Three general levels or strands were identified that outline the guidelines for bioeconomy in the field of agriculture and biological sciences: the first level is sustainable development; the second is forestry and production; and the third is innovation, biomass, and biotechnology. Intertwining these elements forms a holistic approach that aims to balance responsible biological resource exploitation with economic growth and sustainability. The analysis carried out confirmed that research on bioeconomy in the area of agriculture and biological sciences shares the same aspects.

One limitation of the study was that only scientific articles within the fields of agriculture and biological sciences were considered. Research on bioeconomy should be expanded in a more general context, and the findings should be linked to public policy to support more informed decision-making for the proper implementation of bioeconomy in various sectors of the economy.

The insights derived from this analysis can direct subsequent research, influence the formulation of context-specific public policies, and promote the advancement of the bioeconomy paradigm. The relationship between bioeconomy and public policy is crucial, as well-designed policies can promote the adoption of bioeconomic practices and foster markets centered on renewable biological resources. As the bioeconomy continues to evolve, it plays a vital role in seeking innovative and sustainable solutions to address both current and future challenges.

ACKNOWLEDGEMENTS

This work was supported by the Postgraduate College (COLPOS) Campus Córdoba within the Master's Degree in Sustainable Agroalimentary Innovation Science and by the Secretariat of Science, Humanities, Technology, and Innovation (SECIHTI).

REFERENCES

Abad-Segura E, Batlles-de la Fuente A, González-Zama MD, Belmonte-Ureña LJ. 2021. Implications for sustainability of the joint application of bioeconomy and circular economy: A worldwide trend study. Sustainability 13 (13): 7182. https://doi.org/10.3390/su13137182

Araújo R, Calderón FV, López JS, Azevedo IC, Bruhn A, Fluch S, Tasende MG, Ghaderiardakani F, Ilmjärv T, Laurans M, *et al*. 2021. Current status of the algae production industry in Europe: An emerging sector of the blue bioeconomy. Frontiers in Marine Science 7: 626389. https://doi.org/10.3389/fmars.2020.626389

Bannister JR, Donoso PJ, Mujica R. 2016. La silvicultura como herramienta para la restauración de bosques templados. Bosque 37 (2): 229–235. https://doi.org/10.4067/S0717-92002016000200001

Biancolillo I, Paletto A, Bersier J, Keller M, Romagnoli M. 2020. A literature review on forest bioeconomy with a bibliometric network analysis. Journal of Forest Science 66 (7): 265–279. https://doi.org/10.17221/75/2020-jfs

Carbonell SAM, Cortez LAB, Madi LFC, Anefalos LC, Baldassin JR, Leal RLV. 2021. Bioeconomy in Brazil: Opportunities and guidelines for research and public policy for regional development. Biofuels, Bioproducts and Biorefining 15 (6): 1675–1695. https://doi.org/10.1002/bbb.2263

Dahmen N, Lewandowski I, Zibek S, Weidtmann A. 2019. Integrated lignocellulosic value chains in a growing bioeconomy: Status quo and perspectives. GCB Bioenergy 11 (1): 107–117. https://doi.org/10.1111/gcbb.12586

D'Amato D, Veijonaho S, Toppinen A. 2020. Towards sustainability? Forest-based circular bioeconomy business models in Finnish SMEs. Forest Policy and Economics 110: 101848. https://doi.org/10.1016/j.forpol.2018.12.004

de Besi M, McCormick K. 2015. Towards a bioeconomy in Europe: National, regional and industrial strategies. Sustainability 7 (8): 10461–10478. https://doi.org/10.3390/su70810461

de la Cruz-Borrego JG, Caballero-Rico FC. 2021. Bioeconomía una alternativa para México. Análisis bibliométrico a la Web of Science. Dilemas Contemporáneos: Educación, Política y Valores 9 (1): 00046. https://doi.org/10.46377/dilemas.v9i1.2878

GBS (Global Bioeconomy Summit). 2018. Comunicado Cumbre Global de Bioeconomía 2018. Innovación en la bioeconomía global para la transformación sostenible e inclusiva y el bienestar. Berlín, Alemania. 20 p.

IICA (Instituto Interamericano de Cooperación para la Agricultura). 2020. Bioeconomía: potencial y retos para su aprovechamiento en América Latina y el Caribe: manual de capacitación. San José, Costa Rica. 115 p.

Kleinschmit D, Lindstad BH, Thorsen BJ, Toppinen A, Roos A, Baardsen S. 2014. Shades of green: A social scientific view on bioeconomy in the forest sector. Scandinavian Journal of Forest Research 29 (4): 402–410. https://doi.org/10.1080/02827581.2014.921722

Konstantinis A, Rozakis S, Maria EA, Shu K. 2018. A definition of bioeconomy through the bibliometric networks of the scientific literature. AgBioForum 21 (2): 64–85.

Korhonen J, Honkasalo A, Seppälä J. 2018. Circular economy: The concept and its limitations. Ecological Economics 143: 37–46. https://doi.org/10.1016/j.ecolecon.2017.06.041

Kröger M, Raitio K. 2017. Finnish forest policy in the era of bioeconomy: A pathway to sustainability? Forest Policy and Economics 77: 6–15. https://doi.org/10.1016/j.forpol.2016.12.003

Lewandowski I, Clifton-Brown J, Trindade LM, van der Linden GC, Schwarz KU, Müller-Sämann K, Anisimov A, Chen CL, Dolstra O, Donnison IS, *et al*. 2016. Progress on optimizing miscanthus biomass production for the European bioeconomy: Results of the EU FP7 project OPTIMISC. Frontiers in Plant Science 7: 1620. https://doi.org/10.3389/fpls.2016.01620

Lombeyda MB. 2020. Bioeconomía: una alternativa para la conservación. Letras Verdes. Revista Latinoamericana de Estudios Socioambientales 27: 13–30. https://doi.org/10.17141/letrasverdes.27.2020.3984

López-Feria S, Barrero-Domínguez B. 2021. Estrategia de bioeconomía y sostenibilidad en Dcoop a través de la Innovación. C3-BIOECONOMY: Circular and Sustainable Bioeconomy 2: 101–114. https://doi.org/10.21071/c3b.vi2.13745

Mougenot B, Doussoulin JP. 2022. Conceptual evolution of the bioeconomy: A bibliometric analysis. Environment, Development and Sustainability 24 (1): 1031–1047. https://doi.org/10.1007/s10668-021-01481-2

Nazari MT, Mazutti J, Bass LG, Colla LM, Brandli L. 2021. Biofuels and their connections with the sustainable development goals: A bibliometric and systematic review. Environment,

Development and Sustainability 23 (8): 11139–11156. https://doi.org/10.1007/s10668-020-01110-4

OECD (Organization for Economic Co-operation and Development). 2018. Oslo manual 2018: Guidelines for collecting, reporting and using data on innovation (Fourth edition). *In* The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing: Paris, France. https://doi.org/10.1787/9789264304604-en

Olmo IR. 2018. Directiva (UE) 2018/2001 del Parlamento Europeo y del Consejo de 11 de diciembre de 2018 relativa al fomento del uso de energía procedente de fuentes renovables. Diario Oficial de la Unión Europea 87: 141–144.

Paltaki A, Michailidis A, Chatzitheodoridis F, Zaralis K, Loizou E. 2021. Bioeconomy and livestock production nexus: A bibliometric network analysis. Sustainability 13 (22): 12350. https://doi.org/10.3390/su132212350

Papadopoulou CI, Loizou E, Melfou K, Chatzitheodoridis F. 2021. The knowledge based agricultural bioeconomy: A bibliometric network analysis. Energies 14 (20): 6823. https://doi.org/10.3390/en14206823

Pätäri S, Tuppura A, Toppinen A, Korhonen J. 2016. Global sustainability megaforces in shaping the future of the European pulp and paper industry towards a bioeconomy. Forest Policy and Economics 66: 38–46. https://doi.org/10.1016/j.forpol.2015.10.009

Pülzl H, Kleinschmit D, Arts B. 2014. Bioeconomy - an emerging meta-discourse affecting forest discourses? Scandinavian Journal of Forest Research 29 (4): 386–393. https://doi.org/10.1080/02827581.2014.920044

RAE (Real Academia Española). 2023. Desarrollo. Madrid, España. https://dle.rae.es/desarrollo?m=form (Retrieved: October 2023)

SADER (Secretaría de Agricultura y Desarrollo Rural). 2019. Biomasa, creación ecológica de energía. Gobierno de México. Secretaría de Agricultura y Desarrollo Rural. Ciudad de México, México. https://www.gob.mx/agricultura/articulos/biomasa-creacion-ecologica-de-energia (Retrieved: February 2025).

Sanz-Hernández A, Esteban E, Garrido P. 2019. Transition to a bioeconomy: Perspectives from social sciences. Journal of Cleaner Production 224: 107–119. https://doi.org/10.1016/j.jclepro.2019.03.168

Schumpeter JA. 2017. Theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle. Routledge: New York, NY, USA. 320 p. https://doi.org/10.4324/9781315135564

Toppinen A, Röhr A, Pätäri S, Lähtinen K, Toivonen R. 2018. The future of wooden multistory construction in the forest bioeconomy - A Delphi study from Finland and Sweden. Journal of Forest Economics 31: 3–10. https://doi.org/10.1016/j.jfe.2017.05.001

UN (United Nations). 1973. Report of the United Nations conference on the human environment. Stockholm, Sweden. 89 p.

UN (United Nations). 1997. Development program. A/RES/51/240. New York, NY, USA. 67 p.

van Eck NJ, Waltman L. 2014. Visualizing bibliometric networks. *In* Ding Y, Rousseau R, Wolfram D. (eds.), Measuring Scholarly Impact. Springer: Cham, Switzerland, pp: 285–320. https://doi.org/10.1007/978-3-319-10377-8_13

van Lancker J, Wauters E, van Huylenbroeck, G .2016. Managing innovation in the bioeconomy: An open innovation perspective. Biomass and Bioenergy 90: 60–69. https://doi.org/10.1016/j.biombioe.2016.03.017

Wang T, Yu Z, Ahmad R, Riaz S, Khan KU, Siyal S, Chaudhry MA, Zhang T. 2022. Transition of bioeconomy as a key concept for the agriculture and agribusiness development: An extensive review on ASEAN countries. Frontiers in Sustainable Food Systems 6: 998594. https://doi.org/10.3389/fsufs.2022.998594

Agrociencia

CLIMATE CHANGE AND SUSTAINABLE RESOURCE MANAGEMENT IN SAUDI ARABIA: STRATEGIC ADAPTATION

Maher **Toukabri**[1], Antar **Chaabi**[2*]

[1]Northern Border University, College of Business Administration. P.O. Box 1312-1431, Arar, Saudi Arabia, 91431.
[2]Taif University. Khurma University College. P.O. Box 11099, Taif, Saudi Arabia. 21944.
* Author for correspondence: chaabiantar@yahoo.fr

Citation: Toukabri M, Chaabi A. 2026. Climate change and sustainable resource management in Saudi Arabia: Strategic adaptation. **Agrociencia** 60(1): 125-137. https://doi.org/ 10.47163/ agrociencia.v60i1.3364
Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: May 24, 2025.
Approved: January 12, 2026.
Published in Agrociencia: February 10, 2026.

ABSTRACT

This study investigates the complex impacts of climate change on Saudi Arabia's ecosystems, focusing on two major challenges: biodiversity loss and water scarcity. Using structural equation modeling (SEM), the research evaluates the effectiveness of national adaptation strategies that integrate biodiversity conservation, water resource management, and climate policy. The analysis examines sustainable agricultural practices, biodiversity protection programs, and advanced water conservation technologies. The results identified water scarcity as the most critical issue, with renewable water resources expected to decline by 20–30 % by mid-century. Biodiversity loss, particularly among endemic species such as the Arabian oryx, also emerged as a severe threat. Results point out the need to expand desalination capacity, promote agroecological farming, and strengthen ecosystem restoration initiatives, alongside public awareness and environmental education to foster long-term resilience. Aligning adaptation strategies with Saudi Arabia's Vision 2030 framework is essential to support economic diversification, safeguard natural resources, and enhance ecological sustainability.

Keywords: water scarcity, biodiversity loss, sustainable agriculture, ecosystem resilience.

INTRODUCTION

Climate change represents one of the most critical global challenges, with profound effects on ecosystems, economies, and societies. In Saudi Arabia, the consequences are particularly acute due to the country's arid climate, limited freshwater resources, and dependence on oil-based economic structures. Situated in one of the driest regions worldwide, the Kingdom faces intensifying water scarcity, biodiversity degradation, and growing threats to sustainable development.

The scarcity of freshwater remains one of Saudi Arabia's most pressing environmental challenges. Recurrent droughts, rising temperatures, and over-extraction of groundwater have accelerated resource depletion, while energy-intensive desalination contributes to carbon emissions (Chowdhury and Al-Zahrani, 2013; Rashed, 2025). Climate change intensifies desertification, especially in areas like the Rub' al-Khali,

resulting in soil degradation, decreased agricultural productivity, and the extinction of species adapted to arid environments (Shayanmehr *et al*., 2022; Toukabri and Chaabi, 2025). Coastal ecosystems, especially coral reefs and mangroves along the Red Sea and the Arabian Gulf, face additional stress from sea-level rise, ocean acidification, and extreme weather events, threatening marine biodiversity and fisheries (Cooper and Hiscock, 2024).

These environmental pressures have far-reaching socioeconomic implications. Agriculture and food security are highly vulnerable to climate variability, as water-intensive crops experience yield declines and dependence on food imports increases (Zhang *et al*., 2025). Public health is also affected by rising temperatures and the spread of vector-borne diseases (Ebi and Semenza, 2008). From an economic perspective, the dual challenges posed by climate impacts and the global energy transition highlight the pressing need for economic diversification, which is a key goal of Saudi Vision 2030 (Namdar *et al*., 2021).

In response, the Kingdom has accelerated its transition toward renewable energy and sustainable resource management. Large-scale solar projects, green hydrogen initiatives, and innovations such as solar-powered desalination and wastewater recycling are reshaping the national energy and water landscape (Rashed, 2025). Concurrently, initiatives such as the Green Saudi Initiative and the Green Middle East Initiative demonstrate increasing regional leadership in combating desertification and promoting carbon neutrality. Technological advancements, including artificial intelligence and blockchain, further enhance monitoring, forecasting, and efficiency across energy and agricultural systems (Hughes, 2003).

While there is an increasing amount of research on climate change in arid regions, there is a lack of empirical studies that quantitatively explore the relationships among biodiversity loss, water scarcity, and adaptation strategies in Saudi Arabia using advanced modeling techniques. The effectiveness of current climate adaptation strategies remains insufficiently evaluated, as existing studies have largely focused on sector-specific impacts with limited integration across biodiversity, water management, and climate policy.

This study addresses this gap by assessing the effectiveness of Saudi Arabia's adaptation strategies using structural equation modeling (SEM). The analysis identifies the main pathways of ecosystem vulnerability to climate change, evaluates the performance of current adaptation measures in key sectors such as energy, agriculture, and water, and proposes innovative and integrated strategies to strengthen ecological resilience in alignment with Vision 2030.

MATERIALS AND METHODS

Research hypotheses

Climate change exerts multifaceted environmental and socio-economic impacts, particularly in arid regions such as Saudi Arabia. Previous studies have established

theoretical relations between climatic shifts and ecosystem vulnerability, emphasizing effects on heatwave frequency, water availability, and biodiversity integrity (Thomas *et al*., 2004; Milly *et al*., 2005; IPCC, 2021). Empirical evidence further shows that extreme temperatures and water stress increase energy demand and mortality risks, reinforcing the need for adaptive approaches such as integrated water resources management (Blinda and Thivet, 2009; GWP, 2000).

Building on these theoretical and empirical foundations, this study formulates a set of hypotheses to systematically examine the direct and indirect relationships among climate change, water scarcity, biodiversity loss, and adaptation measures in the Saudi context: Climate change is expected to increase both the frequency and intensity of extreme heatwaves (H1). These heatwaves will increase mortality rates (H2a) and lead to greater energy consumption (H2b). Climate change is expected to exacerbate water scarcity (H3), which requires the adoption of new water management strategies (H4). Climate change is believed to contribute directly to biodiversity loss (H5). Integrated water management solutions are considered necessary to address these issues and reduce water scarcity (H6). The increasing adoption of water-saving technologies is also considered a key adaptation strategy under climate change conditions (H7). Lastly, it is hypothesized that the relationship between adaptation strategies and climate change is mediated by water scarcity (H8).

Research design

This study utilized a quantitative research design to investigate the effects of climate change on Saudi Arabia's ecological systems and to assess the effectiveness of adaptation strategies. The quantitative approach is particularly appropriate for examining large-scale environmental datasets, identifying causal and correlational relationships among variables, and providing statistical evidence to support policy recommendations. Such an approach allows for the systematic evaluation of how climate variables influence biodiversity, water resources, and socio-economic factors.

This design is consistent with established methodologies used in climate impact research and adaptation assessments, particularly in arid and semi-arid regions (Fussel, 2009; Mertz *et al*., 2009). Furthermore, it builds upon recent empirical studies that utilize quantitative modeling to analyze ecosystem vulnerability and resilience under changing climatic conditions (El-Rawy *et al*., 2023; Scicluna and Galdies, 2025).

Data collection

This study used both primary and secondary data to provide comprehensive, multi-perspective insights into the impacts of climate change and the adaptation strategies in Saudi Arabia. A structured questionnaire was used to collect primary data from 327 people, including government officials, environmental experts, and community representatives. The survey incorporated both closed-ended questions (Likert-scale and multiple-choice) and open-ended questions, enabling quantitative and qualitative assessment of adaptation practices and their perceived effectiveness (Creswell and Creswell, 2018).

A stratified random sampling approach was used to ensure representation from key sectors (agriculture, water resources, energy, and biodiversity conservation) as well as across major regions. A pilot test with 20 participants was conducted to refine the questionnaire and ensure clarity, reliability, and construct validity (DeVellis, 2017). All research procedures complied with institutional ethical standards. Approval was obtained from the relevant ethics committee, informed consent was secured from all participants, confidentiality was maintained, participation was voluntary, and respondents were free to withdraw at any stage without penalty (Andersson and Keskitalo, 2018; Rana *et al*., 2021).

Secondary data were obtained from reputable national and international sources, including the Saudi Ministry of Environment, Water and Agriculture (MEWA, 2024), the General Authority for Statistics (GASTAT, 2024), the World Bank (2024) open data, and the United Nations Environment Program (UNEP, 2024). These datasets provided indicators related to temperature trends, CO_2 emissions, water and energy consumption, agricultural output, and renewable resource utilization.

The strategic adaptation framework was structured around four core variables. Climate change, as the independent variable, was measured using indicators of temperature change, drought frequency, heatwaves, and precipitation patterns (Fussel, 2009; IPCC, 2021). Adaptation strategies, modeled as a mediating variable, encompassed water management innovations, agroecological practices, and biodiversity conservation initiatives (El-Rawy *et al*., 2023; Scicluna and Galdies, 2025). Ecological vulnerability, considered as the dependent variable, was assessed through land degradation, water scarcity, and biodiversity loss (Pörtner *et al*., 2023), while ecosystem resilience was evaluated using indicators of recovery capacity, biodiversity regeneration, and post-disturbance agricultural productivity (Adger, 2005; Smit and Wandel, 2006).

Data analysis

Data were analyzed using IBM SPSS Statistics 27.0 and AMOS 26.0 (IBM Corp., Armonk, NY, USA). The analytical framework integrated descriptive statistics, correlation analysis, multiple regression, and structural equation modeling (SEM) to test the proposed hypotheses and examine relationships among climate change variables, adaptation strategies, and ecosystem resilience.

Descriptive statistics were used to summarize participants' demographic characteristics and overall trends in stakeholder perceptions of climate impacts and adaptation responses (Creswell and Creswell, 2018). Pearson's correlation analysis then identified significant associations between climate indicators, such as temperature increases and precipitation variability, and measures of ecological vulnerability, including water scarcity and biodiversity loss, assessing both the strength and direction of linear relationships (Gbetibouo, 2009; Mertz *et al*., 2009).

Multiple regression analysis was used to evaluate the predictive effects of climate change variables on ecological vulnerability and to assess the moderating role of adaptation strategies and policy interventions (Field, 2013). To test the hypothesized

causal pathways, SEM was conducted using AMOS to estimate direct and indirect effects and examine interdependencies among climate change, heatwave frequency, energy consumption, mortality, water scarcity, biodiversity loss, and ecosystem resilience. SEM is widely used for testing theoretical models in environmental and climate research (Bollen, 1989; Byrne, 2010).

Instrument reliability and validity were assessed. Internal consistency was evaluated using Cronbach's alpha, with values above 0.7 indicating acceptable reliability (Cronbach, 1951). Confirmatory factor analysis (CFA) was performed to validate the measurement model and confirm the construct validity of latent variables (Hair *et al*., 2010). Overall, this analytical approach provided a statistically robust basis for interpreting the effects of climate change on ecological vulnerability and the role of adaptation and policy measures to enhance ecosystem resilience in Saudi Arabia.

RESULTS AND DISCUSSION

Model fit, validity, and robustness of the structural model

The proposed structural equation model demonstrated strong overall fit to the observed data (Table 1). The Comparative Fit Index (CFI = 0.96) and Tucker–Lewis Index (TLI = 0.94) exceeded the recommended threshold of 0.9, indicating a well-fitting and parsimonious model. Similarly, the Root Mean Square Error of Approximation (RMSEA = 0.05) and the Standardized Root Mean Square Residual (SRMR = 0.04) were below the accepted cutoff of 0.08, reflecting minimal residual variance and close correspondence between observed and predicted covariance matrices. The chi-square to degrees of freedom ratio (CMIN/df = 2.45) also fell within the acceptable range (<3.0).

Together, these indicators confirm that the structural model is statistically robust and theoretically coherent, supporting reliable interpretation of the causal relationships among climate change, water scarcity, biodiversity loss, adaptation strategies, and ecosystem resilience. The high CFI and TLI values, in particular, suggest that the

Table 1. Model fit indices for the proposed structural equation model.

Fit index	Value	Threshold	Interpretation
CFI	0.96	>0.90	Good
TLI	0.94	>0.90	Good
RMSEA	0.05	<0.08	Acceptable
SRMR	0.04	<0.08	Good
CMIN/df	2.45	<3.00	Good

CFI: comparative fit index; TLI: Tucker-Lewis index; RMSEA: root mean square error of approximation; SRMR: standardized root mean square residual; CMIN/df: chi-square divided by degrees of freedom.

hypothesized interdependencies provide an empirically valid representation of climate-ecosystem dynamics in Saudi Arabia.

Discriminant validity was assessed using the Fornell-Larcker criterion. The square root of the average variance extracted (AVE) for each construct (Table 2) exceeded its correlations with other constructs, confirming adequate discriminant validity (Fornell and Larcker, 1981). These results indicate that climate change, heatwave frequency, water scarcity, adaptation strategies, and mortality represent empirically distinct constructs that contribute uniquely to the model, reinforcing the reliability of the measurement framework.

Table 2. Fornell-Larcker discriminant validity assessment for the measurement model constructs.

		CC	HEAT	WS	AS	MORT
AVE		0.72	0.68	0.75	0.70	0.67
Square Root of AVE		0.85	0.82	0.87	0.84	0.82
Correlations with other constructs	HEAT	0.65				0.52
	WS	0.72				
	MORT		0.52			
	EC		0.47			
	AS			0.60		
	WST				0.55	

CC: climate change; HEAT: heatwave frequency; WS: water scarcity; AS: adaptation strategies; MORT: mortality; AVE: average variance extracted; EC: energy consumption; WST: water-saving technologies.

To further test model stability, a sensitivity analysis was conducted by systematically adjusting key parameters related to climate change intensity, water scarcity, and policy interventions (Table 3). A 10 % increase in climate change intensity produced only a minor increase (+0.05) in the path coefficient linking climate change to ecosystem vulnerability, indicating stability under moderate stress conditions. Reducing water scarcity by 10 % resulted in a small decline (-0.07) in the coefficient associated with adaptation strategies, suggesting that improved water availability slightly reduces adaptive pressure without altering its significance. The inclusion of policy interventions strengthened the effects of adaptation strategies and energy consumption (+0.08), reflecting the moderating role of governance and climate policy. Overall, these findings confirm the internal consistency and robustness of the model.

Prior to hypothesis testing, descriptive statistics were examined to contextualize stakeholder perceptions of climate change impacts and adaptation measures, as well as prevailing environmental conditions during the study period. This analysis

Table 3. Sensitivity analysis of structural model parameters under climate and policy scenario adjustments.

Parameter adjustment	Path coefficient change	Conclusion
Increase in climate change intensity (10 %)	+0.05	No significant change in model fit
Decrease in water scarcity (10 %)	-0.07	Small decrease in path coefficient for adaptation strategies
Introduction of policy interventions	+0.08	Increased effect on adaptation strategies and energy consumption

revealed baseline patterns in awareness, perceived environmental stress, and resource management practices, providing an interpretive foundation for subsequent SEM results. These perceptions showed the interpretation of direct, mediating, and moderating effects examined in the following analyses.

Effects of climate change on ecological and socio-economic outcomes

The structural model revealed strong and statistically significant direct relationships among all hypothesized variables ($p < 0.001$), with standardized path coefficients ranging from 0.47 to 0.75 (Table 4). Climate change showed a substantial positive dependence on heatwave frequency and intensity ($\beta = 0.65$), supporting H1 and confirming that broader climatic shifts directly explain variability in heatwave occurrence. This finding aligns with global climate projections predicting intensified heat extremes under global warming scenarios (Seneviratne *et al*., 2012).

Table 4. Hypothesis testing results and path coefficients for the structural model.

Hypothesis	Estimate (β)	Standardized estimates	Standard error	Critical Ratio	*p*-value	Conclusion
H1	0.65	0.70	0.12	5.42	<0.001	Supported
H2a	0.52	0.60	0.11	4.73	<0.001	Supported
H2b	0.47	0.55	0.09	5.22	<0.001	Supported
H3	0.72	0.75	0.10	5.14	<0.001	Supported
H4	0.60	0.65	0.08	4.62	<0.001	Supported
H5	0.58	0.62	0.09	5.27	<0.001	Supported
H6	0.70	0.75	0.06	7.00	<0.001	Supported

CC: climate change; HEAT: heatwave frequency and intensity; WS: water scarcity; AS: adaptation strategies; MORT: mortality rate; EC: energy consumption; WST: water-saving technologies.

Heatwaves exhibited a significant positive dependence on mortality rates (β = 0.52), supporting H2a and demonstrating a clear relationship between climatic extremes and public health outcomes (McMichael *et al*., 2006; Arbuthnott and Hajat, 2017). Heatwaves also showed a direct dependence on energy consumption (β = 0.47), confirming H2b and indicating that rising temperatures drive increased cooling demand and climate-sensitive energy use (Li *et al*., 2015).

A strong dependence between climate change and water scarcity was observed (β = 0.72), supporting H3 and demonstrating the role of temperature, precipitation variability, and evaporation in water availability (Vörösmarty *et al*., 2000; Kundzewicz, 2008). Water scarcity, in turn, significantly influenced the adoption of adaptation strategies (β = 0.6), which validates H4 and indicates that worsening resource constraints directly stimulate adaptive management responses (Liu *et al*., 2022).

Climate change also showed a significant dependence on biodiversity loss (β = 0.58), confirming H5 and highlighting ecosystem sensitivity to climatic stressors, particularly in arid environments (IPCC, 2014; Pimm *et al*., 2014). Finally, adaptation strategies relied heavily on using water-saving technologies (β = 0.7), which supports H6 and emphasizes the critical role of effective governance and planning in facilitating sustainable technological advancements (Liu *et al*., 2022).

Mechanisms shaping adaptation responses

The mediation analysis confirmed that water scarcity fully mediates the relationship between climate change and adaptation strategies (indirect effect: β = 0.55, $p < 0.001$). This shows that climate change does not trigger adaptation directly; instead, its influence operates through increasing water stress, which compels the implementation of adaptive resource management practices. Water scarcity thus emerges as a central mechanism linking climatic stress to policy and technological responses.

In addition, policy interventions were found to significantly moderate the relationship between climate change and heatwave frequency (β = 0.4, $p < 0.001$). Regions with stronger policy frameworks and mitigation measures had a weaker climate-heatwave relationship, showing the buffering role of governance and regulatory capacity. This finding is consistent with prior research emphasizing institutional effectiveness as a key determinant of climate resilience (Adger *et al*., 2005).

Theoretical implications

The path analysis, mediation and moderation tests, and sensitivity analysis show how climate change, ecological vulnerability, and adaptation strategies are connected in Saudi Arabia. Climate change directly drives heatwave frequency, water scarcity, biodiversity loss, and energy consumption, while indirectly shaping adaptive responses through water stress. Policy interventions play an important moderating role by reducing the severity of climatic impacts, particularly heat extremes.

The vulnerabilities associated with heatwaves, water scarcity, and biodiversity loss confirm the heightened sensitivity of Saudi Arabia's arid environment to climate

stressors (Chowdhury and Al-Zahrani, 2013; Shayanmehr *et al*., 2022; Rashed, 2025; Toukabri and Chaabi, 2025). The mediating role of water scarcity supports theoretical claims that environmental pressures function as key drivers of adaptation (Blinda and Thivet, 2009; Taabni and El Jihad, 2012; Liu *et al*., 2022), while the moderating effect of policy interventions reinforces the importance of governance in enhancing resilience (Adger *et al*., 2005; Namdar *et al*., 2021). The results also align with the proposed need for integrated, multi-dimensional adaptation strategies combining technological, ecological, and policy measures (Fussel, 2009; El-Rawy *et al*., 2023; Scicluna and Galdies, 2025).

Policy implications and practical applications

The findings provide clear guidance for climate policy in Saudi Arabia. Given the mediating role of water scarcity, integrated water management should remain central to national resilience strategies, including renewable-energy-powered desalination, precision irrigation, and wastewater recycling to reduce freshwater dependence and carbon emissions (Blinda and Thivet, 2009; Liu *et al*., 2022; Rashed, 2025).

The moderating effect of policy interventions demonstrates the value of governance capacity in mitigating climate risks. Effective measures include early-warning systems for extreme heat, climate-responsive urban design, and incentives for energy-efficient cooling technologies, which enhance public health preparedness and energy sustainability (Adger *et al*., 2005; Namdar *et al*., 2021).

The strong dependence of adaptation strategies on water-saving technologies further shows the need for innovation-oriented policies. Promoting drought-resistant crops, smart irrigation, and ecosystem monitoring tools, supported by digital technologies such as artificial intelligence, can enhance predictive capacity and optimize resource allocation (El-Rawy *et al*., 2023; Scicluna and Galdies, 2025).

Aligning these measures with Saudi Arabia's Vision 2030 ensures an adequate approach to climate resilience, ecological protection, and socio-economic development. Effective implementation will require strengthened coordination among the water, energy, and agricultural sectors. Overall, the combined information here shows how governance, technology, and ecosystem management can work together to improve our ability to adapt to growing climate risks.

CONCLUSION

This study quantitatively evaluates the impacts of climate change on Saudi Arabia's ecosystems and the effectiveness of adaptation strategies. The results show that climate change intensifies heatwaves, water scarcity, and biodiversity loss, while policy interventions and adaptive measures significantly mitigate these effects. The mediating role of water scarcity and the moderating role of governance highlight the importance of strong institutions and coherent policy frameworks for climate resilience.

The findings empirically validate links between ecological vulnerability, adaptation behavior, and policy effectiveness in arid contexts and support Saudi Vision 2030 by emphasizing integrated adaptation policies that combine renewable energy, ecosystem conservation, and technological innovation. Limitations include reliance on secondary data, which may not capture local or temporal variability. Future research should apply longitudinal designs and real-time environmental data to improve predictive capacity.

ACKNOWLEDGEMENTS

The researchers would like to acknowledge the Deanship of Scientific Research, Taif University for funding this work.

REFERENCES

Adger WN, Arnell NW, Tompkins EL. 2005. Adapting to climate change: Perspectives across scales. Global Environmental Change 15 (2): 75–76. https://doi.org/10.1016/j.gloenvcha.2005.03.001

Andersson E, Keskitalo ECH. 2018. Adaptation to climate change? Why business-as-usual remains the logical choice in Swedish forestry. Global Environmental Change 48: 76–85. https://doi.org/10.1016/j.gloenvcha.2017.11.004

Arbuthnott KG, Hajat S. 2017. The health effects of hotter summers and heat waves in the population of the United Kingdom: A review of the evidence. Environmental Health 16 (1): 119. https://doi.org/10.1186/s12940-017-0322-5

Blinda M, Thivet G. 2009. Ressources et demandes en eau en Méditerranée: situation et perspectives. Sécheresse 20 (1): 9–16. https://doi.org/10.1684/sec.2009.0162

Bollen KA. 1989. Structural equations with latent variables. John Wiley and Sons: New York, NY, USA. 514 p. https://doi.org/10.1002/9781118619179

Byrne BM. 2010. Structural equation modeling with AMOS: Basic concepts, applications, and programming (Second Edition). Routledge: New York, NY, USA. 396 p. https://doi.org/10.4324/9780203805534

Chowdhury S, Al-Zahrani M. 2013. Implications of climate change on water resources in Saudi Arabia. Arabian Journal for Science and Engineering 38 (8): 1959–1971. https://doi.org/10.1007/s13369-013-0565-6

Cooper RJ, Hiscock KM. 2024. Groundwater resources: Challenges and solutions. Cambridge Prisms: Water 3. https://doi.org/10.1017/wat.2024.15

Creswell JW, Creswell JD. 2018. Research design: Qualitative, quantitative, and mixed methods approaches (Fifth edition). Sage Publications: Los Angeles, CA, USA. 438 p.

Cronbach LJ. 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16 (3): 297–334. https://doi.org/10.1007/BF02310555

DeVellis RF. 2017. Scale development: Theory and applications (Fourth edition). Sage Publications: Thousand Oaks, CA, USA. 280 p.

Ebi KL, Semenza JC. 2008. Community-based adaptation to the health impacts of climate change. American Journal of Preventive Medicine 35 (5): 501–507. https://doi.org/10.1016/j.amepre.2008.08.018

El-Rawy M, Batelaan O, Al-Arifi N, Alotaibi A, Abdalla F, Gabr ME. 2023. Climate change impacts on water resources in arid and semi-arid regions: A case study in Saudi Arabia. Water 15 (3): 606. https://doi.org/10.3390/w15030606

Field A. 2013. Discovering statistics using SPSS (Fourth edition). Sage Publications: Thousand Oaks, CA, USA. 952 p.

Fornell C, Larcker DF. 1981. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 18 (1): 39–50. https://doi.org/10.2307/3151312

Fussel H. 2009. Review and quantitative analysis of indices of climate change exposure, adaptive capacity, sensitivity, and impacts. Potsdam Institute for Climate Impact Research: Postdam, Germany. 34 p.

GASTAT (General Authority for Statistics). 2024. Statistical yearbook. Riyadh, Saudi Arabia. https://www.stats.gov.sa/en/home (Retrieved: December 2025).

Gbetibouo GA. 2009. Understanding farmers' perceptions and adaptations to climate change and variability: The case of the Limpopo Basin, South Africa. International Food Policy Research Institute: Washington, DC, USA. 52 p.

GWP (Global Water Partnership). 2000. Integrated water resources management. Stockholm, Sweden.

Hair JF, Black WC, Babin BJ, Anderson RE. 2010. Multivariate data analysis (Seventh edition). Pearson: New York, NY, USA. 761 p.

Hughes L. 2003. Climate change and Australia: Trends, projections, and impacts. Australian Journal of Ecology 28 (4): 423–443. https://doi.org/10.1046/j.1442-9993.2003.01300.x

IPCC (Intergovernmental Panel on Climate Change). 2014. Climate change 2014: Impacts, adaptation, and vulnerability. Cambridge University Press: Cambridge, UK.

IPCC (Intergovernmental Panel on Climate Change). 2021. Sixth assessment report (AR6): Climate change 2021: The physical science basis. Cambridge University Press: Cambridge, UK.

Kundzewicz ZW. 2008. Climate change impacts on the hydrological cycle. Ecohydrology and Hydrobiology 8 (2–4): 195–203. https://doi.org/10.2478/v10104-009-0015-y

Li M, Shi J, Guo J, Cao J, Niu J, Xiong M. 2015. Climate impacts on extreme energy consumption of different types of buildings. PLoS ONE 10 (4): e0124413. https://doi.org/10.1371/journal.pone.0124413

Liu W, Liu X, Yang H, Ciais P, Wada Y. 2022. Global water scarcity assessment incorporating green water in crop production. Water Resources Research 58 (1). https://doi.org/10.1029/2020WR028570

McMichael AJ, Woodruff RE, Hales S. 2006. Climate change and human health: Present and future risks. The Lancet 367 (9513): 859–869. https://doi.org/10.1016/S0140-6736(06)68079-3

Mertz O, Mbow C, Reenberg A, Diouf A. 2009. Farmers' perceptions of climate change and agricultural adaptation strategies in rural Sahel. Environmental Management 43 (5): 804–816. https://doi.org/10.1007/s00267-008-9197-0

MEWA (Ministry of Environment, Water and Agriculture). 2024. Environmental indicators report. Riyadh, Saudi Arabia. 146 p.

Milly PCD, Dunne KA, Vecchia AV. 2005. Global pattern of trends in streamflow and water availability in a changing climate. Nature 438 (7066): 347–350. https://doi.org/10.1038/nature04312

Namdar R, Karami E, Keshavarz M. 2021. Climate change and vulnerability: The case of MENA countries. ISPRS International Journal of Geo-Information 10 (11): 794. https://doi.org/10.3390/ijgi10110794

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science 344 (6187). https://doi.org/10.1126/science.1246752

Pörtner HO, Scholes RJ, Arneth A, Barnes DKA, Burrows MT, Diamond SE, Duarte CM, Kiessling W, Leadley P, Managi S, *et al*. 2023. Overcoming the coupled climate and biodiversity crises and their societal impacts. Science 380 (6642). https://doi.org/10.1126/science.abl4881

Rana J, Dilshad S, Ahsan MA. 2021. Ethical issues in research. *In* Farazmand A. (ed.), Global Encyclopedia of Public Administration, Public Policy, and Governance. Springer: Cham, Switzerland. https://doi.org/10.1007/978-3-319-31816-5_462-1

Rashed AA. 2025. Climate change challenges in Saudi Arabia: Strategies for mitigation. Journal of Umm Al-Qura University for Appllied Sciences 2025. https://doi.org/10.1007/s43994-025-00278-7

Scicluna BM, Galdies C. 2025. Assessing the impact of temperature and precipitation trends of climate change on agriculture based on multiple global circulation model projections in Malta. Big Data and Cognitive Computing 9 (4): 105. https://doi.org/10.3390/bdcc9040105

Seneviratne SI, Nicholls N, Easterling D, Goodess CM, Kanae S, Kossin J, Luo Y, Marengo J, McInnes K, Rahimi M, *et al*. 2012. Changes in climate extremes and their impacts on the natural physical environment. *In* Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner GK, Allen SK, Tignor M, Midgley PM. (eds.), Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press: Cambridge, UK, pp: 109-230.

Shayanmehr S, Henneberry SR, Ali EB, Sabouni MS, and Foroushani NS. 2022. Climate change, food security, and sustainable production: A comparison between arid and semi-arid environments of Iran. Environment Development and Sustainability 26: 359–391. https://doi.org/10.1007/s10668-022-02712-w

Smit B, Wandel J. 2006. Adaptation, adaptive capacity, and vulnerability. Global Environmental Change 16 (3): 282–292. https://doi.org/10.1016/j.gloenvcha.2006.03.008

Taabni M, El Jihad MD. 2012. Eau et changement climatique au Maghreb: quelles stratégies d'adaptation? Les cahiers d'Outre-Mer 65 (260): 493–518. https://doi.org/10.4000/com.6718

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, de Siqueira MF, Grainger A, Hannah L, *et al*. 2004. Extinction risk from climate change. Nature 427 (6970): 145–148. https://doi.org/10.1038/nature02121

Toukabri M, Chaabi A. 2025. Strengthening ecosystems and biodiversity conservation in response to climate change in Saudi Arabia. European Journal of Sustainable Development 14 (1): 357–372. https://doi.org/10.14207/ejsd.2025.v14n1p357

UNEP (United Nations Environment Programme). 2024. Global Resources Outlook 2024. Bend the trend: Pathways to a liveable planet as resource use spikes. Nairobi, Kenya. 161 p.

Vörösmarty CJ, Green P, Salisbury J, Lammers RB. 2000. Global water resources: Vulnerability from climate change and population growth. Science 289 (5477): 284–288. https://doi.org/10.1126/science.289.5477.284

World Bank. 2024. World Development Indicators. Washington, DC, USA. https://databank.worldbank.org/source/world-development-indicators (Retrieved: December 2025).

Zhang S, Zhang H, Xie F, Wu D. 2025. Climate change and sustainable agriculture: Assessment of climate change impact on agricultural resilience. Sustainability 17 (16): 7376. https://doi.org/10.3390/su17167376

Agrociencia

MACHINE LEARNING-BASED CROP RECOMMENDATION SYSTEM INTEGRATING SOIL PROPERTIES AND WEATHER CONDITIONS WITH IOT-DRIVEN DATA COLLECTION

Vivek **Balaji**[1], Karuppaiya Sathaiah **Balamurugan**[2*], Tamilvizhi **Thanarajan**[3], Arun Mozhi Selvi **Sundarapandi**[4]

[1]Saveetha Institute of Medical and Technical Sciences. Saveetha School of Engineering, Department of Computer Science and Engineering. Chennai, Tamil Nadu 602105, India.
[2]Karpaga Vinayaga College of Engineering and Technology. Department of Electronics and Communication Engineering. Chengalpattu, Tamil Nadu 603308, India.
[3]Panimalar Engineering College. Department of Computer Science and Engineering, Chennai, Tamil Nadu 600123, India.
[4]Holycross Engineering College. Department of Computer Science and Engineering. Thoothukudi, Tamil Nadu 628851, India.
* Author for correspondence: profksbala@gmail.com

Citation: Balaji V, Balamurugan KS, Thanarajan T, Sundarapandi4 AMS. 2026. Machine learning-based crop recommendation system integrating soil properties and weather conditions with IoT-driven data collection. **Agrociencia** 60(1): 138-154. https://doi.org/ 10.47163/agrociencia.v60i1.3428

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: March 28, 2025.
Approved: January 12, 2028.
Published in Agrociencia: February 11, 2026.

ABSTRACT

Agriculture forms the backbone of human civilization by ensuring food security, economic growth, and rural development. However, farmers face significant challenges in selecting suitable crops due to variability in soil nutrients, pH levels, and unpredictable weather conditions, often leading to reduced productivity and soil degradation. Indian farmers experience seasonal yield losses due to inappropriate crop selection and limited scientific guidance. Existing crop recommendation systems largely rely on static datasets or conventional machine learning models with limited integration of real-time data, resulting in moderate accuracy levels of 80–90 % and limited adaptability to field variability. To overcome these limitations, the proposed system integrates soil properties and weather conditions using Internet of Things (IoT)-driven data collection. Soil and weather sensors connected through a Long Range (LoRa) gateway collect real-time environmental data, which are processed using the XGBoost algorithm in a cloud environment for accurate crop prediction. The developed system achieved 99 % accuracy, outperforming Decision Tree, Random Forest, and Artificial Neural Network (ANN) models, and provides a reliable, scalable, and sustainable decision-support tool for data-driven precision agriculture.

Keywords: machine learning, Internet of Things, LoRa gateway, soil sensor, weather sensor, crop recommendation system.

INTRODUCTION

Agriculture is essential for national wealth and plays a significant role in economic development. In India, a predominantly agrarian country with a population exceeding 1.2 billion, approximately 70 % of the population is engaged in agriculture, producing a diverse range of food products and raw materials, such as jute and cotton, which

are inputs for industries that manufacture everyday goods. Farming increases food production and provides materials needed for commercial manufacturing. Traditional agricultural practices remain widely used and include hand planting, crop rotation, harvesting, agroforestry, terracing, seed saving, polyculture, and subsistence farming. While these practices are based on long-standing knowledge, they are labor-intensive, time-consuming, and often inefficient.

Farmers face numerous challenges arising from both natural and human-induced factors, such as soil erosion, climate change, biodiversity loss, limited capital and labor, and water scarcity. One of the most significant challenges is the selection of suitable crops and fertilizers based on specific soil conditions. Advances in technology and data science have contributed to the progress of precision agriculture. The Internet of Things (IoT) consists of interconnected networks of smart devices and sensors that facilitate continuous data communication and exchange. These systems empower farmers to make data-driven decisions using advanced analytics and machine learning algorithms, resulting in improved yields, optimized resource utilization, and more sustainable farming practices.

Proper crop selection based on soil type is essential for sustainable farming and optimal productivity, as factors such as soil nutritional content, pH, drainage capacity, and texture significantly influence crop growth and development. Sivasubramanian *et al*. (2025) created an application designed to identify the crops most suited to the specific soil characteristics of a given location. This innovation enables farmers to enhance yields while simultaneously reducing input costs and minimizing environmental impacts. Additionally, Kumar *et al*. (2025) introduced a model for predicting weather using multivariate time-series data that combines a Secondary Decomposition Network with a Self-Attentional Spatio-Temporal Learning Network (SASTLNet). Their approach utilizes Singular Spectrum Decomposition with Fuzzy Entropy (SSD-FE) to filter out irrelevant components and isolate essential meteorological information.

In the field of remote sensing, Alotaibi *et al*. (2024) introduced the Dipper-Throated Optimization Deep Convolutional Neural Network Crop Classification (DTODCNN-CC) framework for crop classification using imagery. This framework employs GoogleNet-based Deep Convolutional Neural Networks (DCNNs) to extract high-level visual features. The hyperparameters are optimized through the Dipper-Throated Optimization algorithm, while an Extreme Learning Machine classifier, enhanced by a Modified Sine Cosine Algorithm, achieves improved classification accuracy. Additionally, Selvanarayanan *et al*. (2024) presented an RNN-IoT model that integrates artificial intelligence (AI) and IoT sensors to monitor soil health in coffee plantations. This model utilizes real-time soil and weather data analyzed through Recurrent Neural Networks (RNNs) and Gated Recurrent Units (GRUs) to enhance prediction accuracy, monitoring efficiency, and decision-making compared to traditional methods.

Several studies have focused on integrated machine learning and IoT-based crop recommendation systems. Abdullahi *et al*. (2024) proposed an ensemble machine learning and IoT-based approach aimed at enhancing agricultural productivity in

Somalia, identifying the Decision Tree classifier as the most effective method for crop suggestion. Sravanthi and Moparthi (2024) developed a Multi-Kernel Kronecker Guided Pelican CNN for crop recommendation based on soil data, in addition to a Combined Graph Sample and Aggregate Network designed for disease prediction, fertilizer recommendation, and disease management. Gopi and Karthikeyan (2023) introduced the Red Fox Optimization-Ensemble RNN (RFOERNN-CRPY) framework for crop recommendation and yield prediction for comparative performance analysis. Agrawal *et al*. (2024) utilized a Multi-Criteria Decision Analysis (MCDA) approach through the Analytical Hierarchy Process (AHP) to recommend crops such as mustard and wheat, considering soil, climatic, and topographic factors. Tanaka *et al*. (2024) applied Support Vector Machine (SVM), Random Forest, and Artificial Neural Network (ANN) algorithms trained on datasets encompassing eight soil types to support crop and fertilizer recommendations. Finally, Asadi and Shamsoddini (2024) assessed machine learning, object-based, and deep learning methods for crop type mapping using radar and optical imagery.

Beyond crop selection, several studies have explored broader agricultural decision-support systems. Chelliah *et al*. (2022) introduced a machine learning framework aimed at predicting market demand and improving coordination between farmers and financial institutions, facilitating better utilization of government schemes. Momenpour *et al*. (2024) conducted a bibliometric analysis of machine learning-based crop harvest forecasting, examining knowledge structures, research trends, and collaboration networks. Attri *et al*. (2023) examined the applications of machine learning in agriculture, classifying them into categories including weed and pest identification, plant disease detection, stress alleviation, smart farming, and crop harvesting diagnostics.

Chipatela *et al*. (2024) used artificial intelligence to analyze soybean performance with climate and edaphic data and employing a Random Forest classifier with predictors like soil phosphorus, pH, organic matter, precipitation, and texture. Sakthipriya and Naresh (2024) developed a CNN-based automated image classification method and compared its accuracy against established architectures, including ResNet, DenseNet, VGGNet, and AlexNet. Zhao *et al*. (2024) proposed a backpropagation neural network optimized using Sand Cat Swarm Optimization (SCSO), Hunter-Prey Optimization (HPO), and Golden Jackal Optimization (GJO) to predict the performance of summer maize based on meteorological, soil, and crop data from northern China.

Other contributions highlighted the importance of practical deployment and system integration. Elbasi *et al*. (2023) investigated how machine learning and IoT can be used in modern agriculture to minimize resource waste by selecting suitable crops according to water availability, climate, and soil conditions, utilizing real-time sensor data. Krishnan *et al*. (2023) created a forward-thinking web application that optimizes direct crop trading between farmers and retailers. The app has price prediction, offline navigation, and ongoing support to help farmers make more money. Reddy *et al*. (2023) found that Naïve Bayes demonstrated superior accuracy and performance compared to CNN models. Musanase *et al*. (2023) proposed a comprehensive crop and

fertilizer recommendation system for Rwanda that integrates machine learning with rule-based fertilization. Islam *et al*. (2023) created an IoT-enabled machine learning system that employs soil and environmental sensors to monitor nutrients and provide real-time crop recommendations.

This study introduces a novel method for crop and fertilizer recommendations based on machine learning, utilizing integrated soil and climatic data. It builds upon existing approaches that emphasize soil and weather conditions. In developing countries like India, farmers often rely on traditional knowledge or guesswork to select crops. This approach frequently results in low yields and inefficient resource use. A dedicated, evidence-based system that leverages modern technologies, including IoT and machine learning, is necessary to provide crop recommendations tailored to real-time soil and climatic conditions. This research is driven by the desire to bridge the gap between technology and agriculture, equipping farmers with tools that facilitate intelligent, data-driven decision-making.

This study combined real-time data from soil and weather sensors with IoT-based communication via Long Range (LoRa) gateways and utilized cloud computing for data storage, preprocessing, and model training. An intelligent recommendation model was evaluated using the XGBoost algorithm. The hypothesis posited that the amalgamation of sensor data, cloud computing, and IoT communication would enhance the precision and promptness of agricultural recommendations. Additionally, it was expected that XGBoost would outperform traditional prediction methods under varying soil and climatic conditions to enhance resource efficiency, productivity, and sustainability. The system is deployable on actual farmland, provides real-time crop guidance through a user-friendly interface, and is scalable across diverse soils and climates. There are also potential future extensions to include fertilizer recommendations, pest forecasting, and yield prediction.

MATERIALS AND METHODS

This research utilizes a systematic experimental approach to validate the hypotheses. The investigation considers the significant challenges farmers encounter when choosing crops under diverse soil and climatic conditions, which are key factors affecting crop productivity. Existing forecasting and recommendation systems have limitations, such as inappropriate technique selection, weak attribute selection, and suboptimal analytical performance, which frequently lead to diminished crop productivity.

A novel technique is proposed in which crop recommendations are based on climate and soil health. Soil and weather data are transmitted to the cloud via a LoRa gateway. In the cloud, the data undergo preprocessing and are divided into training and testing sets in a 70 to 30 % ratio. The XGBoost algorithm then processes the data to update the soil and weather conditions, generating appropriate crop recommendations. These recommendations are delivered to farmers through a user interface (Figure 1).

It is important to consider the relationships between soil physical and chemical properties and crop suitability (Table 1). Various soil textures, such as sandy loam,

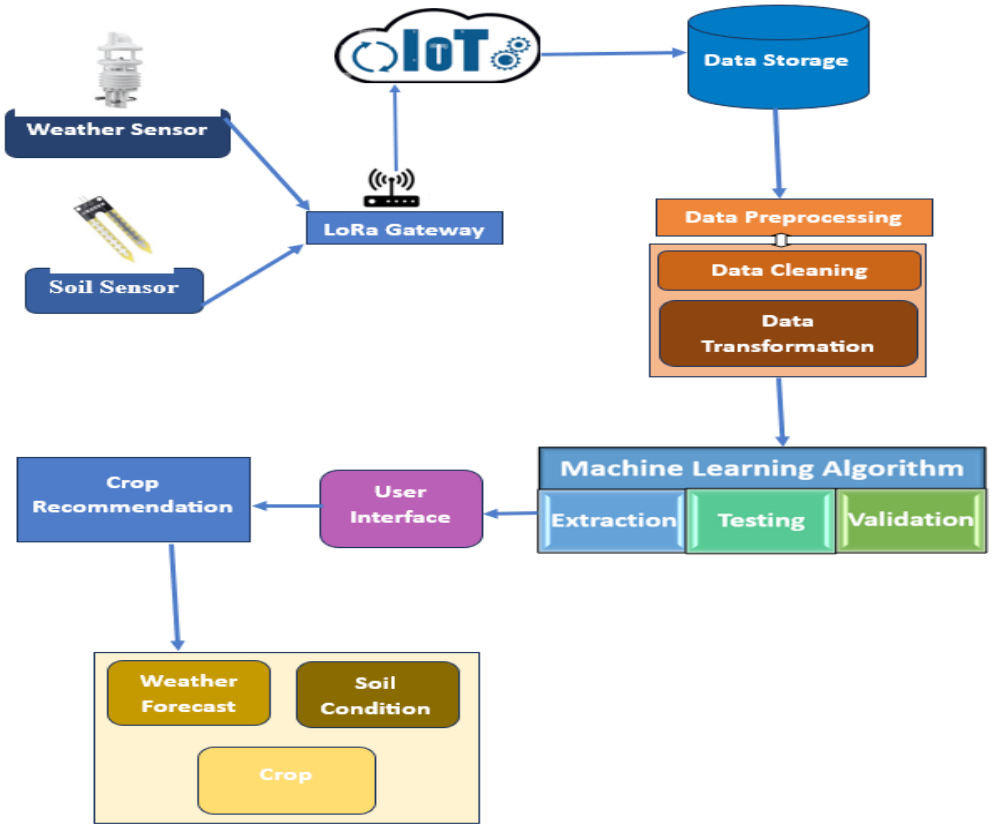

Figure 1. Architecture of the proposed IoT-XGBoost-based crop recommendation system.

Table 1. Optimal soil physical and chemical parameters for crop suitability.

Parameter	Suitable value range	Suitable crops
	Soil texture	
Sandy loam	45 % sand, 35 % silt, 20 % clay	Carrots, potatoes, peanuts
Silty loam	50 % silt, 25 % sand, 25 % clay	Wheat, soyabean, corn
Clay loam	40 % clay, 30 % silt, 30 % sand	Rice, broccoli, kale
Soil texture	Aggregated and well-drained	Most crops
	Soil pH	
Acidic	5.0–6.0	Blueberries, potatoes, sweet potatoes
Slightly acidic	6.0–6.5	Tomatoes, corn, soyabeans
Neutral	6.5–7.5	Wheat, oats, barley
Alkaline	7.5–8.0	Asparagus, spinach, beets
Nitrogen(N)	20–40 mg kg^{-1}	Leafy greens, corn, wheat
Phosphorus (P)	30–50 mg kg^{-1}	Beans, peas, root vegetables
Potassium (K)	100–200 mg kg^{-1}	Tomatoes, carrots, potatoes
Organic matter	3–6 %	Most crops
Soil moisture	20–30 % field capacity	Rice, corn, beans

silty loam, and clay loam, are suitable for specific crops like carrots, wheat, and rice, respectively. Most crops benefit from well-aggregated and well-drained soils, while optimal soil pH ranges from acidic to slightly alkaline. Furthermore, adequate nutrient contents, including nitrogen, phosphorus, and potassium, as well as levels of organic matter and drainage conditions, must be provided to ensure soil fertility and support proper crop growth.

Dataset

The data utilized in this project was obtained from publicly accessible Indian databases that encompass soil and weather conditions (Crop Recommendation Dataset, available at Kaggle, https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset). This dataset comprises 2200 records featuring seven essential agronomic parameters, including nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall from various agricultural regions across India, ensuring representation of diverse soil and climatic conditions.

To maintain data integrity, a comprehensive preprocessing workflow was implemented. Duplicate and incomplete entries were eliminated, outliers were detected using the interquartile range (IQR) method, and missing values were addressed through median imputation. Additionally, all numerical features were normalized using min-max scaling to a range of 0–1, ensuring uniformity across the dataset. These procedures guaranteed that the dataset utilized for training and testing the XGBoost model was clean, consistent, and statistically reliable.

LoRa gateway

The data were transmitted to the cloud through bidirectional Long Range (LoRa) gateways that allow users to receive information for real-time analysis and monitoring. A LoRa gateway usually receives data from several field-installed sensors that have LoRa capabilities. These sensors measure various parameters, including temperature, soil moisture, humidity, and other meteorological factors. The gateway features a high-sensitivity receiver and a powerful transceiver that operates in unlicensed ISM bands, typically at 868 MHz.

Once collected, the data are forwarded to the cloud via backhaul connections such as Ethernet, Wi-Fi, or cellular networks. The data are packaged into internet-compatible formats, such as MQTT or HTTP, and securely transmitted to cloud servers. In a cloud environment, data is stored, processed, and analyzed to generate actionable insights related to agricultural conditions. This supports optimized irrigation scheduling, crop management strategies, and overall productivity. When farmers request information or send commands, the process initiates through the cloud platform, which processes the request and transmits the results back to the LoRa gateway as downlink data. The gateway subsequently forwards the processed information to authorized users via the user interface.

Data preprocessing

Several procedures are involved in preprocessing data for the XGBoost method to ensure efficient training and accurate predictions. Initially, soil data are collected from multiple sources, including remote sensing, laboratory analyses, and soil surveys. The dataset typically consists of both numerical and categorical variables. During the data cleaning stage, outliers and missing values are addressed. If not properly handled, outliers can distort the training process and negatively impact model performance. Missing values are replaced using appropriate techniques, such as mean or median imputation, to ensure the dataset is complete and consistent. The preprocessing stage prepares a clean and reliable dataset for subsequent model training and evaluation.

For feature x with mean μ and standard deviation σ, the normalization feature x' can be described as:

$$x' = \frac{x - \mu}{\sigma}$$

In regression tasks, the Mean Squared Error (MSE) is commonly used to appraise the model's performance. For predicted values $\hat{y}_i$ and actual values y_i, MSE is determined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

Data splitting

Training a robust machine learning model requires both the training and testing sets to be highly representative (Figure 2). To achieve this, a fair split ratio and the sampling strategy were established in advance. While there is still no consensus on the ideal train-test split ratio, the Pareto principle serves as the foundation for the commonly recommended 80:20 ratio.

XGBoost algorithm

The XGBoost algorithm effectively handles complex interactions and patterns in data, making it an ideal choice for crop recommendations based on soil conditions. In this study, soil-related attributes, including pH, nutrient content, texture, drainage capacity, historical crop yields, and crop compatibility across soil types, were utilized to train the model. Through this training process, XGBoost identified the relationships and trends that connect soil characteristics to agricultural productivity.

To enhance model reliability and prevent overfitting, XGBoost's built-in regularization mechanisms were utilized. To discourage too much model complexity and encourage sparsity in feature weights, both L1 (Lasso) and L2 (Ridge) regularization terms were added to the objective function. Early stopping was used to halt training when

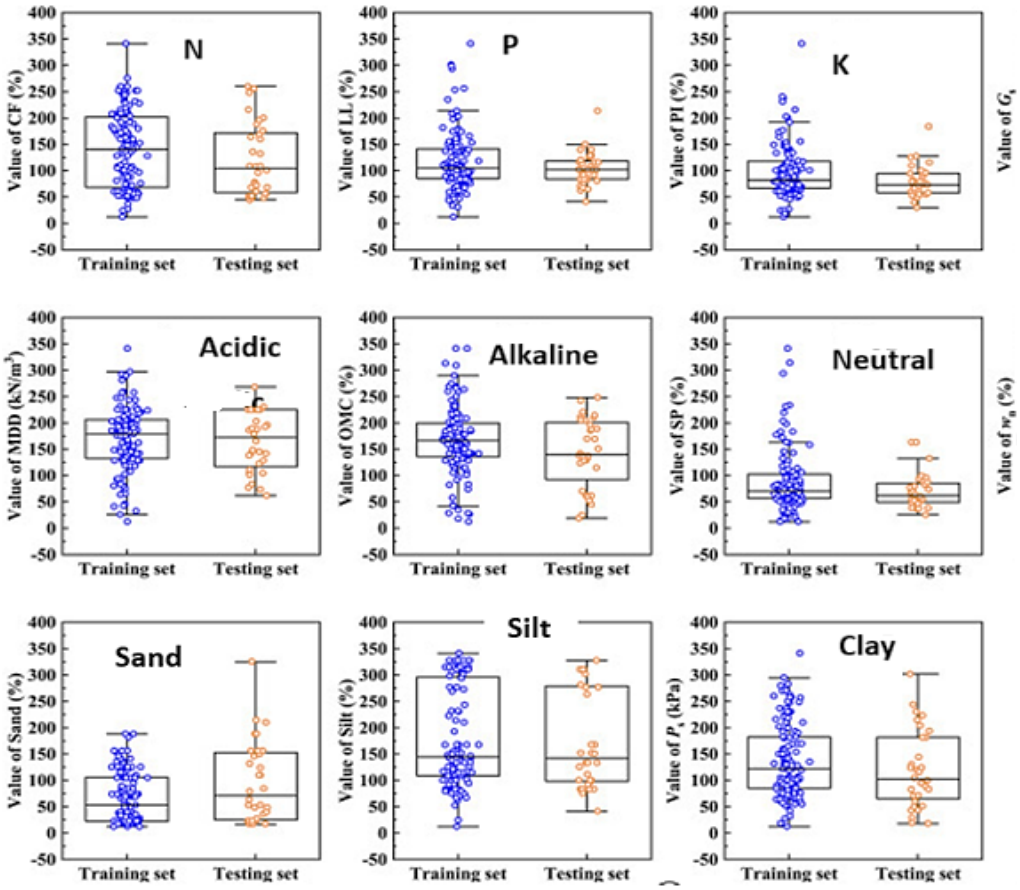

Figure 2. Distribution of soil and climatic parameters across training and testing datasets.

the validation loss did not decrease for 20 consecutive rounds, ensuring controlled learning. Additionally, 10-fold cross-validation was implemented to verify model stability by rotating training and testing folds and calculating mean performance scores across iterations. Hyperparameters such as maximum depth, learning rate, subsample, and column sample per tree were optimized through grid search. These strategies established an effective validation framework, mitigated overfitting, and ensured that the reported performance accurately reflects true predictive capability rather than mere memorization of the training data.

Cross-validation

To ensure that the trained XGBoost model generalized effectively to unseen data, 10-fold cross-validation was used during the training phase. The dataset was randomly divided into 10 equal subsets. In each iteration, nine subsets were used for training, while one subset was reserved for testing. This process was repeated until each subset had served as the validation fold once. The final performance metrics were calculated

by averaging the results across all folds, which helped reduce overfitting and provided a stable and reliable estimation of performance indicators.

After training and validation, the model was used to recommend suitable crops for specific soil types. Using soil data from a given location, the trained XGBoost model predicted either crop suitability or expected yield. This capability depended on the model's ability to generalize from the training data by effectively leveraging the relationships identified between soil properties and crop growth. XGBoost optimized an objective function that integrated a loss term with a regularization component:

$$Objective = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where $L(y_i, \hat{y}_i)$ is the loss function that calculates the variation across the values that are real ($\hat{y}_i$) and expected (y_i), $\Omega(f_k)$ is the is the regularization term that prevents excessive fitting by penalizing intricate models, and f_k symbolizes the collective's k-th tree.

For every training instance, the reduction function's gradient and Hessian were calculated and expressed as:

$$Gradient\ (g_i) = \frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i}$$

$$Hessian\ (h_i) = \frac{\partial^2 L(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}$$

where L represents loss function, y_i represents true label, and y_i represents the predicted value. During the enhancement process, decision forests are constructed using these gradients and Hessians.

Tree building

XGBoost constructs trees in a sequential manner, where each new tree aims to correct the errors made by the previous ensemble. During the tree-building phase, optimal split points are selected based on factors such as the gain or reduction in the objective function. The overall prediction of the XGBoost model was obtained by aggregating the forecasts from each tree, which are often adjusted by a shrinkage variable. The final prediction of the XGBoost model results from this aggregation of outputs from all trees.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)$$

Experimental setup

The prototype of the proposed method integrates soil and weather sensors connected via a LoRa gateway for real-time data transmission and analysis. The proposed crop recommendation method, which is based on soil and weather conditions, was implemented using the Luster Leaf 1601 Rapitest soil kit. This kit measures soil pH and nutrient levels, including nitrogen, phosphorus, and potassium. Weather parameters, such as temperature, humidity, wind direction and speed, rainfall, and barometric pressure, were collected using the SwitchDoc Labs WeatherRack2 sensor. Data transmission between field sensors and the cloud was facilitated through a LoRa gateway. The system required a portable computer equipped with an Intel i5 processor for local coordination tasks. Python and Google Colab were utilized for data processing, model training, and analysis.

Key performance indicators

Performance metrics were utilized to assess the effectiveness of the XGBoost model to predict crop suitability based on soil conditions. Commonly used metrics included accuracy, precision, recall, F1-score, and mean squared error, which quantifies the average squared difference between predicted and actual values. The metrics for accuracy, precision, recall, and F1-score were calculated as follows:

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

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

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

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

where *TP* is true positive, *TN* is true negative, *FP* is false positive, and *FN* is false negative.

Accuracy measured the proportion of correct predictions, while precision assessed the reliability of positive predictions. Recall indicated the model's ability to identify

all relevant instances, whereas specificity measured its effectiveness in correctly identifying non-relevant cases. The F1-score offered a balanced evaluation of model performance by representing the harmonic mean of precision and recall.

RESULTS AND DISCUSSION

The reported results represent the average performance obtained through 10-fold cross-validation, following hyperparameter tuning and early stopping. This approach ensures that the high accuracy values reflect robust model generalization rather than overfitting. Model performance was derived from the averaged outcomes of the cross-validation process, which incorporated XGBoost's built-in L1/L2 regularization and early stopping, thus supporting reliable and generalizable predictions.

The integration of a weather sensor (SwitchDoc Labs WeatherRack2), a soil sensor (Luster Leaf 1601 Rapitest), and a LoRa gateway facilitated continuous data transmission to the cloud. In this environment, the XGBoost algorithm analyzed both real-time and historical data to generate optimal crop recommendations. The weather sensor provided real-time measurements of temperature, humidity, rainfall, and solar radiation, while the soil sensor captured essential parameters such as moisture, pH, and nutrient availability.

Data were transmitted via the LoRa gateway to cloud services for analysis, and farmers accessed the resulting insights through a user-friendly mobile interface. This enabled informed decisions regarding crop selection, planting time, and resource management. Overall, this integrated framework enhanced agricultural productivity by improving yield potential and resource efficiency while supporting sustainable, location-specific farming practices, achieving an overall accuracy of approximately 99 %.

The XGBoost model achieved 100 % accuracy on the training data and 99.1–99.6 % accuracy during validation. However, the narrow performance gap was carefully examined to rule out overfitting. Learning-curve analysis indicated steady convergence between training and validation accuracies, demonstrating stable generalization. Additionally, 10-fold cross-validation yielded consistent results across folds, with variance remaining below 0.5 %. The combination of early stopping, hyperparameter regularization (max_depth = 6, subsample = 0.8, colsample_bytree = 0.8), and feature normalization helped mitigate memorization effects. The high accuracy was attributed to the well-structured, low-noise Kaggle dataset, which featured clear labeling and balanced crop classes, rather than to model overfitting.

The confusion matrix (Figure 3) offers a detailed overview of the model's predictive performance by contrasting actual and predicted classifications. It identifies four possible outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

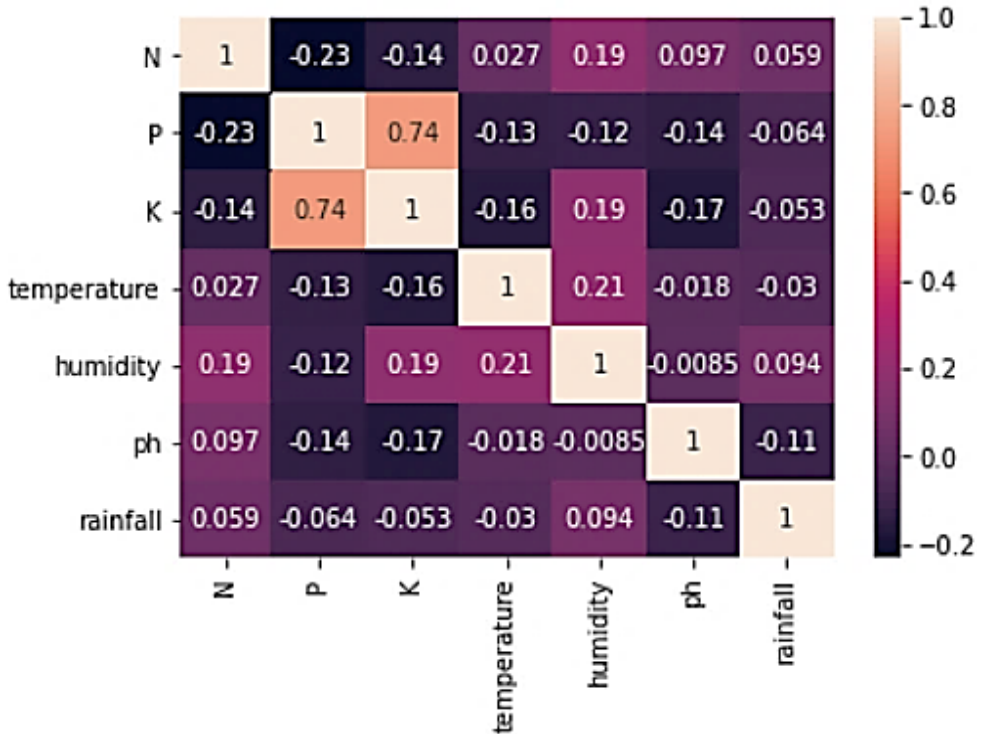

Figure 3. Confusion matrix for crop recommendation model performance.

True positives refer to instances when the model accurately identifies a crop as suitable based on current weather and soil conditions. These conditions include factors such as temperature, humidity, pH, rainfall, and soil nutrients like nitrogen, potassium, and sodium. True negatives denote when the model correctly determined that a crop was not suitable. False positives arise when the model mistakenly predicts a crop to be suitable, while false negatives indicate situations where a suitable crop is incorrectly classified as unsuitable. These results serve as the foundation for calculating performance metrics for the XGBoost-based model for recommendations, including accuracy, precision, recall (sensitivity), specificity, and the F1-score (Table 2).

All crops reached a training accuracy score of 100 %, while validation demonstrated consistently high performance, ranging from 99.1 to 99.6 % (Figure 4).

The performance evaluation of the crop recommendation system indicated that the XGBoost-based model performed exceptionally well, achieving 99 % accuracy and F1-score, along with 98 % precision and recall for crop recommendations based on soil conditions (Table 3). These performance metrics reflect the final optimized XGBoost model, which was trained and validated on the full Kaggle dataset, following hyperparameter tuning and 10-fold cross-validation, and therefore represent the model's best achievable performance using the full set of soil and weather features.

The superior performance of the proposed XGBoost model, when compared to

Agrociencia 2026. DOI: https://doi.org/10.47163/agrociencia.v60i1.3428
Scientific Article

Table 2. Performance metrics of the XGBoost-based crop recommendation model.

	Precision	Recall	F1-score	Trained accuracy (%)	Validated accuracy (%)
Cotton	0.99	0.99	0.99	100	99.6
Pomegranate	0.99	0.99	0.99	100	99.2
Papaya	0.99	0.99	0.99	100	99.1
Orange	0.99	0.99	0.99	100	99.6
Muskmelon	0.99	0.99	0.99	100	99.2
Mango	0.99	0.99	0.99	100	99.6
Maize	0.99	0.99	0.99	100	99.4
Mung bean	0.99	0.95	0.99	100	99.6
Grapes	0.99	0.99	0.99	100	99.3
Kidney beans	0.99	0.99	0.99	100	99.2
Jute	0.93	0.99	0.95	100	99.2
Coffee	0.99	0.99	0.99	100	99.1
Apple	0.99	0.99	0.99	100	99.6
Black gram	0.99	0.99	0.99	100	99.2
Pigeon pea	0.99	0.99	0.99	100	99.4

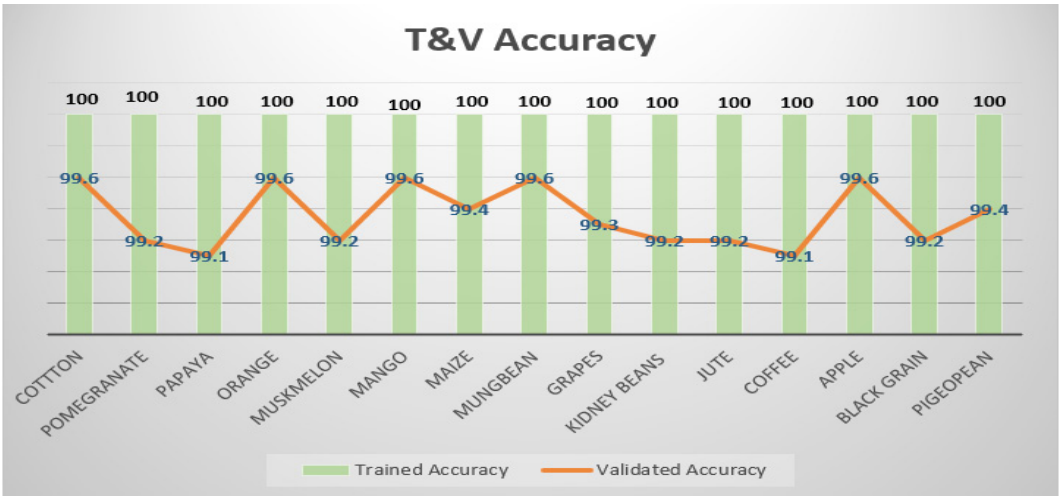

Figure 4. Training and validation accuracy obtained across crop categories.

Decision Tree, Logistic Regression, ANN, GB Boost, Random Forest, and RNN, can be attributed to its ensemble learning architecture and optimized regularization framework. Unlike single-tree approaches, such as Decision Tree, which are prone to overfitting, or linear models like logistic regression, which struggle to adequately capture nonlinear relationships, XGBoost merges multiple weak learners through gradient boosting to effectively lower both bias and variance. In contrast to traditional

Table 3. Comparative performance evaluation of the proposed XGBoost model and existing crop recommendation techniques.

Other techniques	Accuracy	Precision	Recall	F1-Score
Decision Tree	91.01	84.39	88.17	85.34
Logistic Regression	94.23	94.41	93.81	94.03
ANN	88.72	83.16	85.74	-
IDCSO-WLSTM	93.69	91.78	92.66	-
MKGPCNN	98	98.55	98.6	98
Proposed (XGBoost)	99	98.80	98.9	99

ANN: Artificial Neural Network; IDCSO-WLSTM: Improved Dragonfly-Cat Swarm Optimization with Weighted Long Short-Term Memory; MKGPCNN: Multi-Kernel Kronecker Guided Pelican Convolutional Neural Network.

neural network models (ANN and RNN), which often require large datasets and extended training times to achieve convergence, XGBoost operates efficiently on moderate-sized, structured datasets. Its weighted decision-tree boosting, combined with L1 and L2 regularization, column subsampling, and early stopping, enhances generalization and minimizes sensitivity to noise. Furthermore, XGBoost effectively captures interactions between soil nutrients and environmental variables, including nitrogen-temperature and pH-rainfall relationships, leading to consistently higher precision, recall, and F1-scores compared to the competing methods.

The proposed XGBoost approach demonstrated strong overall performance, achieving an accuracy and specificity of 0.9, a sensitivity of 0.88, and an F1-score of 0.9 (Table 4). Both GB Boost and Random Forest also showed competitive results, with each attaining a specificity of 0.9 and an F1-score of 0.91. However, GB Boost slightly outperformed Random Forest in accuracy, with values of 0.89 and 0.87, respectively. In contrast, the RNN model had the weakest performance across all metrics, particularly in sensitivity (0.85) and F1-score (0.83). All models were assessed under a standardized benchmark setting that utilized a reduced feature subset and maintained consistent parameter configurations to ensure fairness in comparison. The marginally lower accuracy range

Table 4. Comparative performance evaluation of the proposed XGBoost Model and other machine learning techniques.

Prediction techniques	Accuracy	Sensitivity	Specificity	F1-Score
Proposed XGBoost	0.9	0.88	0.9	0.9
GB Boost	0.89	0.89	0.9	0.91
Random Forest	0.87	0.88	0.9	0.91
RNN	0.86	0.85	0.85	0.83

(0.88–0.9) observed for the proposed model is attributed to this constrained comparison framework, rather than indicating any decline in model robustness or reliability.

CONCLUSIONS

The XGBoost algorithm for crop recommendation demonstrated superior performance. This technology enables the real-time monitoring of various environmental factors, including soil moisture, pH, temperature, and humidity, to ensure accurate environmental assessments. However, while the Kaggle dataset is representative, it may not adequately capture seasonal or region-specific soil variability across the diverse agroclimatic zones of India. The proposed system mainly focuses on temperature, humidity, and soil nutrients, overlooking other important factors such as pest infestations, irrigation frequency, and crop rotation patterns. Additionally, the real-time field deployment was simulated using laboratory sensors, which may not accurately reflect actual field conditions.

Future research can explore various extensions of this work. Integrating Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) with IoT-based data collection could enhance crop prediction accuracy and enable early disease detection. Additionally, satellite imaging and remote sensing technologies may offer improved soil and climate analysis, as well as spatial and temporal assessments. The incorporation of smart irrigation modules, fertilizer optimization tools, and real-time mobile advisory applications can further enhance the system's utility for farmers. Researchers should also consider integrating economic, environmental, and sustainability metrics into a comprehensive agricultural decision-support framework. These advancements will contribute to a more adaptive, scalable, and intelligent precision agriculture platform.

REFERENCES

Abdullahi MO, Jimale AD, Ahmed YA, Nageye AY. 2024. Revolutionizing Somali agriculture: Harnessing machine learning and IoT for optimal crop recommendations. Discover Applied Sciences 6 (3): https://doi.org/10.1007/s42452-024-05739-y

Agrawal N, Govil H, Kumar T. 2024. Agricultural land suitability classification and crop suggestion using machine learning and spatial multicriteria decision analysis in semi-arid ecosystem. Environment, Development and Sustainability 27 (6): 13689–13726. https://doi.org/10.1007/s10668-023-04440-1

Alotaibi Y, Rajendran B, Rani KG, Rajendran S. 2024. Dipper throated optimization with deep convolutional neural network-based crop classification for remote sensing image analysis. PeerJ Computer Science 10: e1828. https://doi.org/10.7717/peerj-cs.1828

Asadi B, Shamsoddini A. 2024. Crop mapping through a hybrid machine learning and deep learning method. Remote Sensing Applications: Society and Environment 33: 101090. https://doi.org/10.1016/j.rsase.2023.101090

Attri I, Awasthi LK, Sharma TP. 2023. Machine learning in agriculture: A review of crop management applications. Multimedia Tools and Applications 83 (5): 12875–12915. https://doi.org/10.1007/s11042-023-16105-2

Chelliah BJ, Latchoumi TP, Senthilselvi A. 2022. Analysis of demand forecasting of agriculture using machine learning algorithm. Environment, Development and Sustainability 26 (1): 1731–1747. https://doi.org/10.1007/s10668-022-02783-9

Chipatela FM, Khiari L, Jouichat H, Kouera I, Ismail M. 2024. Advancing toward personalized and precise phosphorus prescription models for soybean (*Glycine max* (L.) Merr.) through machine learning. Agronomy 14 (3): 477. https://doi.org/10.3390/agronomy14030477

Elbasi E, Zaki C, Topcu AE, Abdelbaki W, Zreikat AI, Cina E, Shdefat A, Saker L. 2023. Crop prediction model using machine learning algorithms. Applied Sciences 13 (16): 9288. https://doi.org/10.3390/app13169288

Gopi PSS, Karthikeyan M. 2023. Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model. Multimedia Tools and Applications 83 (5): 13159–13179. https://doi.org/10.1007/s11042-023-16113-2

Islam MR, Oliullah K, Kabir MM, Alom M, Mridha MF. 2023. Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. Journal of Agriculture and Food Research 14: 100880. https://doi.org/10.1016/j.jafr.2023.100880

Krishnan N, Surendran R, Nathan M. 2023. Crop tracker - A web application to sell or buy crops and predict crop price using machine learning. IET Conference Proceedings 2022 (26): 152–156. https://doi.org/10.1049/icp.2023.0386

Kumar M, Subramanian P, Surendran R. 2025. Multivariate time series weather forecasting model using integrated secondary decomposition and self-attentive spatio-temporal learning network. Global NEST Journal 27 (4). https://doi.org/10.30955/gnj.06195

Momenpour SE, Bazgeer S, Moghbel M. 2024. A bibliometric analysis of the literature on crop yield prediction: Insights from previous findings and prospects for future research. International Journal of Biometeorology 68 (5): 829–842. https://doi.org/10.1007/s00484-024-02628-2

Musanase C, Vodacek A, Hanyurwimfura D, Uwitonze A, Kabandana I. 2023. Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. Agriculture 13 (11): 2141. https://doi.org/10.3390/agriculture13112141

Reddy P, Surendran R, Venkatraman M. 2023. Enhancing accuracy in crop yield estimation through the integration of naive bayes and convolutional neural networks. *In* 2023 Intelligent Computing and Control for Engineering and Business Systems. Institute of Electrical and Electronics Engineers. Chennai, India. https://doi.org/10.1109/iccebs58601.2023.10449096

Sakthipriya S, Naresh R. 2024. Precision agriculture based on convolutional neural network in rice production nutrient management using machine learning genetic algorithm. Engineering Applications of Artificial Intelligence 130: 107682. https://doi.org/10.1016/j.engappai.2023.107682

Selvanarayanan R, Rajendran S, Algburi S, Khalaf OI, Hamam H. 2024. Empowering coffee farming using counterfactual recommendation based RNN driven IoT integrated soil quality command system. Scientific Reports 14 (1). https://doi.org/10.1038/s41598-024-56954-x

Sivasubramanian S, Gopalsamy Venkatesan SK, ThanarajanT, Rajendran S. 2025. Al-Biruni earth radius optimization for enhanced environmental data analysis in remote sensing imagery. Agrociencia 59 (5). https://doi.org/10.47163/agrociencia.v59i5.3380

Sravanthi G, Moparthi NR. 2024. An efficient IoT based crop disease prediction and crop recommendation for precision agriculture. Cluster Computing 27 (5): 5755–5782. https://doi.org/10.1007/s10586-023-04246-w

Tanaka TST, Heuvelink GBM, Mieno T, Bullock DS. 2024. Can machine learning models provide accurate fertilizer recommendations? Precision Agriculture 25 (4): 1839–1856. https://doi.org/10.1007/s11119-024-10136-x

Zhao L, Qing S, Li H, Qiu Z, Niu X, Shi Y, Chen S, Xing X. 2024. Estimating maize evapotranspiration based on hybrid back-propagation neural network models and meteorological, soil, and crop data. International Journal of Biometeorology 68 (3): 511–525. https://doi.org/10.1007/s00484-023-02608-y

Agrociencia

USE OF DESALINATION BRINES IN THE CULTIVATION OF HALOPHYTES: A VISION OF CIRCULAR ECONOMY

Arlett Leticia **Ibarra-Villarreal**[1], Germán Eduardo **Dévora-Isiordia**[1*], Rosario **Montoya-Pizeno**[1], Edgar Omar **Rueda-Puente**[2], Rafael Enrique **Cabanillas-López**[3], Marco Antonio **Gutiérrez-Coronado**[4]

[1]Instituto Tecnológico de Sonora. Departamento de Ciencias del Agua y Medio Ambiente. Calle 5 de Febrero 818 Sur, Ciudad Obregón, Sonora, Mexico. C. P. 85000.
[2]Universidad de Sonora. Departamento de Agricultura y Ganadería. Boulevard Luis Encinas y Rosales S/N, Colonia Centro, Hermosillo, Sonora, Mexico. C. P. 83000.
[3]Universidad de Sonora. Departamento de Ingeniería Química. Boulevard Luis Encinas y Rosales S/N, Colonia Centro, Hermosillo, Sonora, Mexico. C. P. 83000.
[4]Instituto Tecnológico de Sonora. Departamento de Biotecnología y Ciencias Ambientales. Calle 5 de Febrero 818 Sur, Ciudad Obregón, Sonora, Mexico. C. P. 85000.
* Author for correspondence: german.devora@itson.edu.mx

Citation: Ibarra-Villarreal AL, Dévora-Isiordia GE, Montoya-Pizeno R, Rueda-Puente EO, Cabanillas-López RE, Gutiérrez-Coronado MA. 2026. Use of desalination brines in the cultivation of halophytes: A vision of circular economy. **Agrociencia** 60(1): 155-174. https://doi.org/ 10.47163/agrociencia.v60i1.3491

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: May 08, 2025.
Approved: December 04, 2025.
Published in Agrociencia: December 15, 2025.

ABSTRACT

Water scarcity and distribution constitute a problem driven by population growth and industrial overexploitation. To secure water supply, desalination technologies for seawater and brackish water have been adopted, becoming critically important. Reverse osmosis is the highest-rated technology for this process and generates two output streams: permeate water and brine, the latter characterized by a high concentration of total dissolved solids (TDS). When untreated, brine is discharged into water bodies and soils, causing ecological damage. To mitigate this impact, the circular economy proposes reusing part of the brine in agriculture through halophyte plants, which offer the advantage of growing under high salt concentrations. The objective of the research was to document the salinity tolerance of halophyte species and the potential use of water rejected from the desalination process as irrigation water, with a circular economy approach. Brine reuse represents an opportunity to reduce waste and generate environmental, social, and economic benefits. Among the main halophyte species capable of tolerating brine above 30 000 mg L^{-1} are *Suaeda salsa* (L.) Pall., *Salicornia bigelovii* Torr., *Rhizophora mangle* L., and *Chenopodium quinoa* Willd. *Salicornia europaea* L. is classified as a halophyte species with medium tolerance (10 000–30 000 mg L^{-1}). Species with low salinity tolerance (5000–10 000 mg L^{-1} TDS) include *Atriplex nummularia* Lindl., *Zoysia japonica* Steud., and *Crithmum maritimum* L. These plants also possess significant nutritional and pharmaceutical properties and can be used as livestock feed, human food, for oil extraction, soil remediation, and other applications.

Keywords: Reverse osmosis, *Rhizophora mangle* L., *Salicornia bigelovii* Torr., salinity tolerance, *Suaeda salsa* (L.) Pall.

INTRODUCTION

Rapid population growth and industrialization have increased demand on water resources, making it difficult to meet basic human needs such as drinking water, food production, and electricity. Although most of the planet is covered by water, only about 2.5 % is available as fresh water, and three-quarters of this volume is frozen in glaciers and ice caps, making it practically inaccessible in the short term (Ríos-Arriola *et al*., 2022; Dévora-Isiordia *et al*., 2023). Currently, about 2.2 billion people lack access to safe drinking water (Salehi, 2022). Climate change is expected to worsen this situation and threaten sustainable development, highlighting the urgent need to preserve existing water sources and develop new ones (Dévora-Isiordia *et al*., 2021).

The development and implementation of desalination technologies has increased. Desalination removes salts, contaminants, and minerals from seawater or brackish water to produce fresh water for human, agricultural, or industrial use. Its main advantage is that it provides water independently of the hydrological cycle. However, the process generates two output streams: desalinated water and reject water or brine (Musie and Gonfa, 2023; Saleem *et al*., 2023). Brine is characterized by causing environmental problems due to its high concentration of salts, organic matter, and other substances, disrupting the osmotic balance of marine species and causing dehydration and mortality (Valdés *et al*., 2021).

Research by Gil-Trujillo and Sadhwani Alonso (2023) showed that optimization of diffusers can improve dilution performance in seawater. Soil contamination is also a concern, as salinity degrades physical, chemical, and microbiological soil properties, reducing crop yield (Chi *et al*., 2025). To prevent soil damage, several disposal methods for brine have been proposed, such as discharge into surface waters or sewers, deep-well injection, evaporation ponds, soil application, zero-liquid discharge, and aquaculture farms (Dévora-Isiordia *et al*., 2017). The Yaoba oasis in China is an example of this problem, where the accumulation of salts from irrigation water quality has limited agricultural development and deteriorated groundwater quality (Lu *et al*., 2025).

Because soil salinity causes yield losses, halophyte crops have been selected for irrigation with brine, reducing the negative effects of disposal and applying reuse techniques under circular economy principles (Al-Tamimi *et al*., 2023). Gómez-Bellot *et al*. (2021) evaluated brine from reverse osmosis for irrigating two forage halophyte species, *Crithmum maritimum* L. and *Atriplex halimus* Lindl., observing no negative effects and considering brine a viable option for revegetation or soil conservation. Oron *et al*. (2023) proposed a zero-discharge replacement system using brine sequentially in duckweed and fish ponds to produce nutrient-rich wastewater for halophyte irrigation, demonstrating efficient food production and reduced environmental impact.

Further applications show promising results. Robertson *et al*. (2019) evaluated *Salicornia bigelovii* Torr. production for fresh tips using brine enriched with nutrients through aquaculture systems, tripling yield to 23.7 Mg ha^{-1} and increasing economic return despite higher per-hectare costs. Li *et al*. (2019) reported that *Suaeda salsa* (L.) Pall. and *Suaeda glauca* Bunge retain ions in their roots, while Li *et al*. (2023) found

that *S. glauca* and *Limonium aureum* (L.) Hill accumulate heavy metals and NaCl in their tissues, regulating soil salinity. Al-Tamimi *et al*. (2023) also evaluated *Salicornia* at different concentrations and nutrient contents using three irrigation types, obtaining yields of 16 kg m^{-2} with brine at 25 641 mg L^{-1}. Higher forage production (2–2.8 kg m^{-3}) occurred with pressure-compensated drip irrigation and subsurface irrigation. The economic water productivity analysis showed that brine from aquaculture tanks combined with pressure drip irrigation costs USD 5.8–6.2 m^{-3}, higher than desalination costs (USD 1.5 m^{-3}), but they concluded that the greatest economic benefits come from pass-through aquaculture irrigation and drip or subsurface systems.

The scientific contribution of this work lies in the synergy between brine reuse from desalination processes and the cultivation of halophyte plants of potential commercial value within a circular economy framework. This approach reduces environmental impacts from brine discharge, promotes waste valorization, and presents a sustainable management proposal for desalination brine. This review aims to raise awareness of sustainable brine disposal and emphasizes the potential of halophyte cultivation for this purpose. Therefore, the objective was to document the salinity tolerance of halophyte species and evaluate the potential use of reject water from desalination as irrigation water under a circular economy approach. To achieve this objective, the following research questions were established: What is the most widely used desalination system? Which crops can be irrigated with brine water? What is the salt tolerance of these crops? What is their industrial potential? Is it possible to apply the circular economy to the synergy between saline waste from desalination plants and the cultivation of halophytic plants?

MATERIALS AND METHODS

To ensure a focused and current foundation for the study, a systematic search strategy was implemented. Literature selection was restricted to research articles published within the last 10 years up to 2025. The search for relevant literature was carried out using the SCOPUS database, one of the most comprehensive digital sources of peer-reviewed scientific articles and widely used in bibliometric analysis (Sweileh, 2021). The process began with general concepts related to brines from desalination processes and subsequently incorporated more specific terms to focus the search on reverse osmosis, circular economy, and halophytes, applying the Boolean operators “AND” and “OR” (Maçaira *et al*., 2018).

The *h*-index and number of citations were used to assess the productivity and impact of researchers through their publications. An acceptable minimum average *h*-index of ≥ 10 and an average of 1000 citations were considered for all authors. The *h*-index measures research impact by combining productivity and influence: articles were ranked from highest to lowest number of citations, and the *h*-index corresponds to the number of publications that have received at least that same number of citations.

RESULTS AND DISCUSSION

Bibliometric analysis

This systematic review with statistical data analysis examines the current status of the use of brine from desalination processes as irrigation water for the cultivation of halophyte plants. According to Boolean search engines in SCOPUS, the combination circular economy "AND" "OR" reverse osmosis brine yielded 1032 articles; halophyte "AND" "OR" circular economy yielded 14 articles; and halophyte "AND" "OR" reverse osmosis brine yielded eight articles published between 2015 and 2025.

Individual productivity and global interdisciplinary collaboration are essential for developing, researching, and identifying halophyte species capable of tolerating brine residues from desalination systems. This study, based on the establishment of keywords, resulted in 1054 articles that contributed to answering the research questions. The three most cited authors were Cárdenas-Pérez, Araus, and Oron, with 22 824, 10 400, and 4105 citations and *h*-indexes of 71, 53, and 36, respectively. Of the 70 research articles used for this study, the average number of citations was 1379, with an average *h*-index of 11.79. Israel and Spain were the countries with the highest number of publications (57).

It is essential to continue researching and obtaining information from additional databases, as global scientific output is constantly changing. The results of the last two search combinations show that research in the field of halophytes, reverse osmosis brines, and their synergistic use within a circular economy framework remains low. This highlights the importance of continuing research and establishing the current state of the art, which was a key motivation for this work. More publications are needed to understand the adaptation of halophytes under irrigation schemes using brine, including evaluations of yields, phenology, and related parameters.

Desalination systems and reverse osmosis

Several systems have been developed for seawater desalination, including reverse osmosis, electrodialysis, and distillation (Table 1). Each method has different characteristics in terms of energy requirements, membrane use, and operating pressure, with the objective to reduce the salt content of seawater or brackish water to obtain water suitable for human, agricultural, and industrial use. Each method presents specific advantages and disadvantages, and the choice depends on factors such as energy consumption, operating costs, and environmental impact (Ashraf *et al*., 2022).

Currently, reverse osmosis is the most widely used desalination technology worldwide due to its high production capacity, low energy consumption, high permeate quality, and lower cost compared with thermal desalination systems (Feria-Díaz *et al*., 2021; Montoya-Pizeno *et al*., 2023). Ongoing advances in low-pressure membrane materials and energy-recovery systems have further reduced reverse osmosis costs, reaching USD 0.4–0.8 m^{-3} with specific energy consumption of 2.2–3 kWh m^{-3} for seawater

Table 1. Classification of desalination processes by separation mechanism (Feria-Díaz *et al*., 2021; Ashraf *et al*., 2022).

	Process	Technique	System
Separation of salts in water	Membrane	Pressure	Reverse osmosis
	Resin	Chemical attraction	Ionic exchange
	Membrane selective	Electric charge	Electrodialysis
Separation of water from salts	Evaporation	Steam	Thermal vapor compression Mechanical vapor compression Multi-effect vertical tube distillation Multi-effect distillation of horizontal tubes Solar distillation Simple flash distillation Submerged tube distillation Multistage flash distillation
	Crystallization	Cold	Freezing Hydrate formation

(Armendáriz-Ontiveros *et al*., 2022; Feo-García *et al*., 2024). In contrast, thermal technologies, although more robust under high salinity, maintain significantly higher energy demands (6–10 kWh m^{-3}) and total costs of USD 1.2–2 m^{-3}, limiting their application mainly to regions with residual thermal sources or abundant solar energy (Ríos-Arriola *et al*., 2022; Benahmed *et al*., 2025)..

The reverse osmosis process (Figure 1) begins with physicochemical pretreatment tailored to the characteristics of the feedwater to remove undissolved solids, organic matter, and microorganisms, preventing membrane fouling (Ahmed *et al*., 2023). After

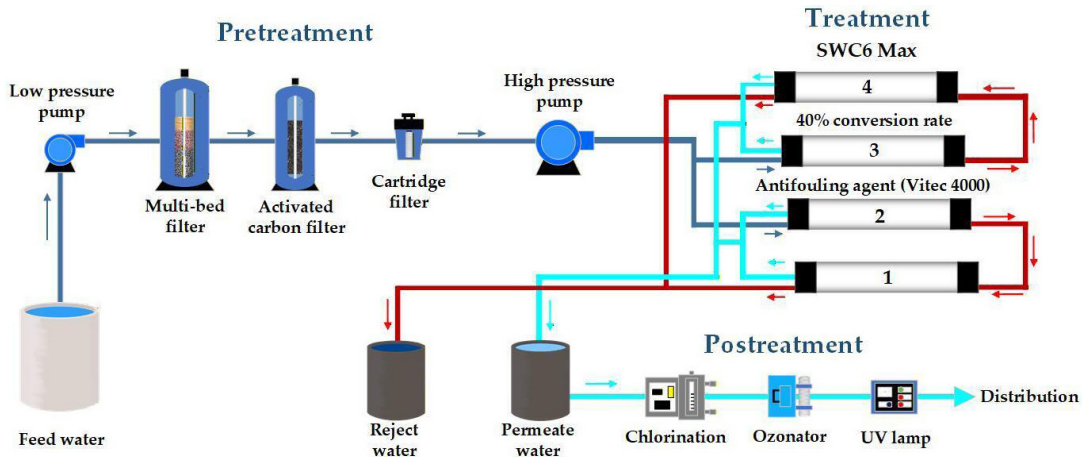

Figure 1. Reverse osmosis process flow diagram.

pretreatment, brackish or seawater is pressurized and passed through a semipermeable membrane that separates unwanted particles to obtain water that meets established quality standards (Feria-Díaz *et al*., 2021). At the end of the process, permeate water and brine are obtained. Because the osmotic pressure of seawater is 2.68 MPa, the system requires a high-pressure pump applying at least twice that value (5.51 MPa). The permeate exits at 0.3 MPa, sufficient for distribution, while the reject stream exits at 5.24 MPa and is reutilized through energy-recovery devices, which contributes to the higher energy efficiency of reverse osmosis compared with thermal methods.

The semipermeable membrane used consists of three polymer layers: a thin polyamide film (0.2 μm), polysulfone (40 μm), and polyester (120 μm) (Álvarez-Sánchez *et al*., 2025). The permeate contains low salt concentrations and undergoes post-treatment (remineralization and disinfection using chlorination, ultraviolet treatment, or ozone) to meet quality requirements for use and consumption (Valdés *et al*., 2021). The brine retains impurities and process chemicals such as anti-scalants (polyphosphates), biocides (chlorine), and pH-adjusting agents (HCl or NaOH) and is characterized by a high concentration of total dissolved solids (TDS) (Gómez-Bellot *et al*., 2021; Shokry *et al*., 2025).

The brine problem

The reject water, also known as brine, has specific characteristics depending on the source and quality of the feed water, the pretreatment applied, the desalination process, and the percentage of recovered water. In some cases, TDS exceeds 55 000 mg L^{-1}; this increase, known as the concentration factor, establishes the number of times salts and contaminants are concentrated in the reject stream compared to the feed water. It is calculated based on conversion and is a key indicator, since higher conversion results in a higher concentration factor (Gil-Trujillo and Sadhwani-Alonso, 2023).

Brine contains ions such as Na^{+}, K^{+}, Ca^{2+}, Mg^{2+}, Cl^{-}, SO_4^{2-} , HCO^{3-} and PO_4^{3-}. In theory, all ions in seawater and brine should concentrate proportionally according to the concentration factor, but in practice they do not, due to physicochemical processes within the reverse osmosis membrane, solubility and precipitate formation (concentration polarization), and the electrochemical properties of the ions (Zolghadr-Asli *et al*., 2023; Myung *et al*., 2025). This type of water is generally discharged directly into the sea, generating negative environmental impacts on aquatic flora and fauna. High salinity, temperature, and metal content can cause eutrophication, reduced dissolved oxygen, and species extinction, especially in sensitive ecosystems such as coral reefs (Gómez-Bellot *et al*., 2021; Cornejo-Ponce *et al*., 2022).

In 2022, 20 956 desalination plants worldwide produced more than 115.62 million m^3 d^{-1} of brine (Ríos-Arriola *et al*., 2022). The main sources were Saudi Arabia, the United Arab Emirates, Kuwait, and Qatar (Jones *et al*., 2019; Lee *et al*., 2024). In China, 30 160 million Mg of brine wastewater with concentrations of 1000–20 000 mg L^{-1} were treated in 2022; 26.4 % originated from the coal and petrochemical industry, 20.7 % from the chemical industry, and 2.3 % from desalination technologies.

In reverse osmosis processes, brine production is directly related to the recovery percentage. For example, a seawater feed of 35 000 mg L^{-1} and a flow rate of 103 600 m^3 d^{-1} at 41.6 % recovery produces 60 000 m^3 d^{-1} of reject water at 60 000 mg L^{-1} (Liu *et al*., 2024). In this study, the focus was on saline waste from reverse osmosis desalination plants; therefore, the concentration of reject water depends on the characteristics of the feed water (concentration, flow rate, and conversion). Brines also differ in concentration according to their origin: brackish water (1000–10 000 mg L^{-1}) yields reject water at about 20 000 mg L^{-1}, while seawater at 35 000 mg L^{-1} results in a reject concentration of 60 000 mg L^{-1}.

Case studies and economic viability

The appropriate management of brine is critical to minimize its environmental and financial impact, as the cost of brine disposal ranges from 5 to 33 % of the total cost of a desalination plant (Cornejo-Ponce *et al*., 2022). Because of this, its reuse is implemented in different areas. In China, 1508 million m^3 are reused annually for the irrigation of crops tolerant to saline concentrations, such as *S. bigelovii*. Its production cost is estimated at USD 878.30 ha^{-1}, while its selling price is USD 4.73 kg^{-1}. Cárdenas-Pérez *et al*. (2021) report that the biomass yield of *S. bigelovii* is 35 Mg ha^{-1}. Financial analyses indicate that, under optimal conditions, its cultivation can generate a profit of USD 76 000 ha^{-1}, which is higher than other halophytes such as *Distichlis spicata* (L.) Greene and *Sporobolus virginicus* (L.) Kunth (Rotter *et al*., 2025).

Another example of reuse is the production of salt blocks obtained from the crystallization of salts (Figure 2). These blocks are used as livestock feed due to their nutritional value. Their monthly production is around 544 000 blocks of 10–25 kg,

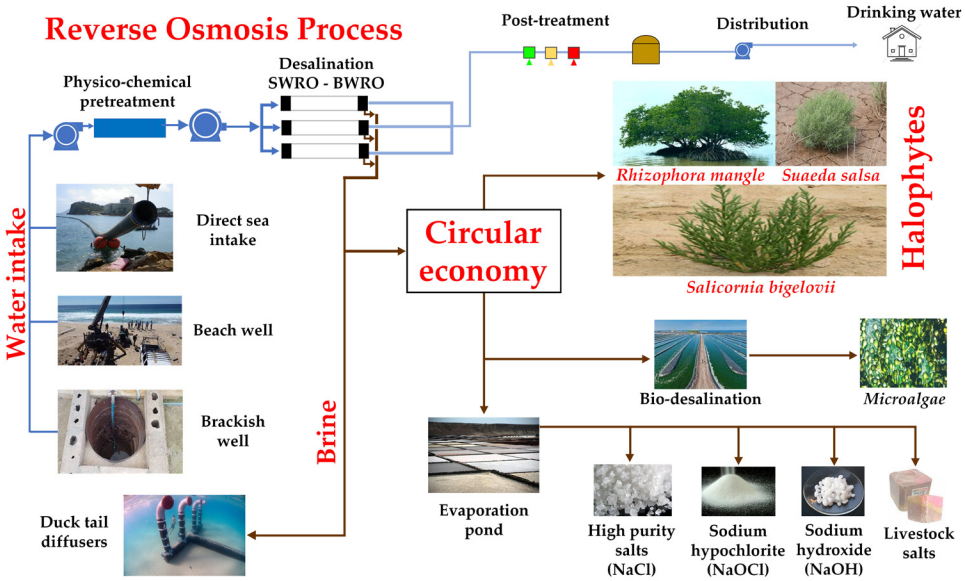

Figure 2. Conceptual transition from linear to circular economy models (Mallick, 2022).

with 108 988 Mg of salt derived from a brine source of 60 000 mg L^{-1}. This reduces the release of brine into the environment and generates an economic benefit through commercialization, with regional selling prices ranging from USD 25 to 45. Similarly, sodium chloride, sodium sulfate, and acid and alkaline solutions are obtained from salt crystallization processes by evaporation (Liu *et al*., 2024).

Even so, the most commonly used alternatives for brine disposal include ocean discharge, evaporation ponds, canal disposal, and deep well injection. Although ocean discharge is the most widely used method, it can negatively affect marine flora and fauna (Lee *et al*., 2024). An example is *Posidonia oceanica* (L.) Delile, which is sensitive to changes in salinity in the Mediterranean Sea and is affected by brine spills and is therefore used as an indicator of environmental impact (Blanco-Murillo *et al*., 2023).

Petersen *et al*. (2018) also report significant impacts on the reef-building coral species *Stylophora pistillata* Esper, *Acropora tenuisy* Dana, and *Pocillopora verrucosa* Ellis due to a 10 % increase in salinity combined with antifoulants (0.2 mg L^{-1}) in the Gulf of Aqaba in the Red Sea. The presence of amphipods and species tolerant to hypersaline spills, such as marine worms and polychaetes, has also been compromised along the coasts of Spain near brine discharges (Grammatiki *et al*., 2025). In another study, de Serio *et al*. (2025) define a tolerance threshold of 5000 mg L^{-1} and exposure of less than 48 h to avoid physiological stress in benthic communities. Key species like crustaceans and mollusks, crucial to the nutrient cycle and the stability of the food web, would suffer otherwise.

The circular economy solution

The brine produced during the reverse osmosis process is typically discarded due to its high salt concentration, leading to significant environmental impacts and elevated disposal costs (Gómez-Bellot *et al*., 2021). However, this resource can now be managed under circular economy principles, transforming it into a valuable input. The circular economy is an advanced model compared to the linear economy (Figure 3). Its objective is sustainable development, seeking to optimize waste management through reuse by

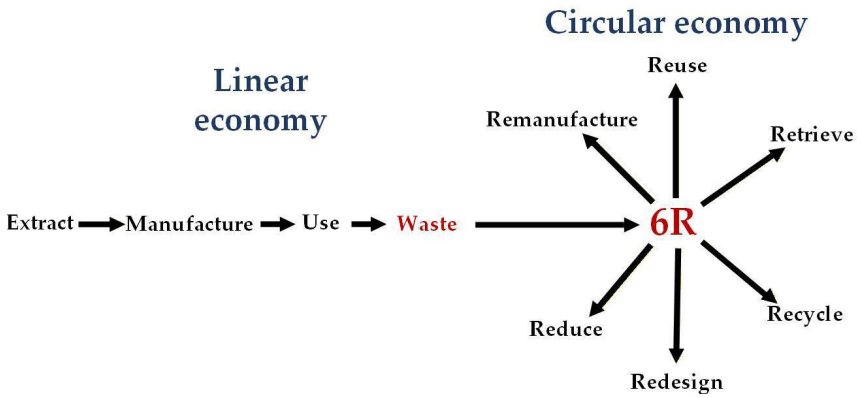

Figure 3. Transition from linear economy to circular economy (Mallick, 2022).

applying principles that include preserving and enhancing natural capital, optimizing resource use, and promoting system efficiency (Espinoza, 2023; Raudales-García *et al*., 2024).

This approach emphasizes reducing environmental impacts throughout the product life cycle by implementing the 6Rs: reduce, reuse, recycle, recover, remanufacture, and redesign (Mallick, 2022). The circular economy provides economic, environmental, and social benefits by promoting business competitiveness and pollution reduction (Espinoza, 2023). The potential use of brine as irrigation water for halophyte cultivation aligns with these principles through resource reuse and proper management (Cornejo-Ponce *et al*., 2022; Mallick, 2022; Raudales-García *et al*., 2024). In this context, waste from one process becomes an input for another sector, such as agriculture (Gómez-Bellot *et al*., 2021; Mallick, 2022).

Use of brine for the cultivation of halophyte plants

Circular economy is a relevant model for arid regions where water is a scarce resource (Cornejo-Ponce *et al*., 2022). These regions face severe water scarcity and salinity problems that generate abiotic stress and put agricultural production at risk. Agriculture consumes 70 % of available water, making it the activity with the highest demand due to low irrigation efficiency and the high water requirements of crops. In this context, it is necessary to explore strategies to use non-conventional water resources and marginal lands (Araus *et al*., 2021).

The reuse of brine in agriculture is an option that contributes to water conservation through sustainable practices (Oron *et al*., 2023). Plants differ in their tolerance to salinity; the most tolerant are halophyte plants, which withstand electrical conductivity (EC) values over 30 dS m^{-1}. Above this threshold, most crops cannot survive, and only crops with thresholds above 4–8 dS m^{-1} are classified as very tolerant. Therefore, the volume of brine that can be applied depends on plant type and brine properties (Lee *et al*., 2024).

The low salinity tolerance of many crops limits the use of saline water for irrigation, making it essential to select flora that naturally grow in saline soils. A soil is classified as saline when its EC exceeds 4 dS m^{-1}, with an exchangeable sodium percentage below 15 and a sodium adsorption ratio below 13. These soils, also known as white alkaline soils or Solonchaks, generally reduce germination and development in most crops. Therefore, it is necessary to consider crop alternatives adapted to adverse conditions, such as halophyte plants (Biswas and Naher, 2019; Gómez-Bellot *et al*., 2021; Sparks *et al*., 2024).

Halophyte plants constitute a small group representing about 1 % of global flora, encompassing around 2500 genera that include shrubs, bushes, mangroves, and grasses capable of growing in environments with high salt concentrations (Sánchez and Matos, 2018; Lee *et al*., 2024). These plants have evolved to tolerate high salinity and require salt exposure for optimal growth, completing their life cycle in salinities of 200 mM NaCl or higher. However, a precise threshold cannot be defined due to their

taxonomic and ecological complexity (Behera and Ramachandran, 2021; Cárdenas-Pérez *et al*., 2021; Teotia and Chaudhary, 2024).

Based on their degree of salinity tolerance, halophyte plants are divided into obligate halophytes and facultative halophytes. Obligate halophytes grow and develop in highly saline environments, while facultative halophytes can develop in habitats with low or no salinity. Their classification remains ambiguous because different authors use diverse criteria such as anatomy, tolerance mechanisms, and salt tolerance intensity (Rahman *et al*., 2021).

From a physiological perspective, considering salinity tolerance and ion accumulation and transport, they can be classified into three types: recritohalophytes, which secrete salt through external salt glands; euhalophytes, which accumulate salt in the vacuoles of green succulent tissues of leaves and stems, reducing the ionic content in the cytoplasm; and salt-exclusion halophytes, which accumulate large amounts of salt in tissue vacuoles (Lu *et al*., 2021; Rahman *et al*., 2021).

In functional terms, halophyte plants have developed specialized mechanisms to tolerate salt stress and maintain cellular osmotic balance. These include sequestration of Cl^- and Na^+ ions in vacuoles, accumulation of osmoprotectants to reduce osmotic potential, selective membrane transporters to regulate salt adsorption and exclusion, and the production of defense metabolites such as superoxide and glutathione reductase to neutralize reactive oxygen species (Mzoughi and Majdoub, 2021; Rahman *et al*., 2021).

There is a wide range of salt concentrations in different types of water, from ultrapure water to seawater (Table 2). This classification is important when assessing the salinity tolerance of halophyte plants, as each species has a specific threshold for salt exposure. Ultrapure and pure water have extremely low TDS concentrations and are therefore unsuitable for irrigating halophytes, especially non-facultative species that require saline conditions for optimal growth. In contrast, brackish, saline, and seawater contain TDS levels consistent with the conditions tolerated by halophytes, enabling the use of saline discharges from desalination plants for irrigating species adapted to high salinity.

Table 2. Electrical conductivity (EC) and total dissolved solids (TDS) across different water types (Sarathe *et al*., 2022).

Water	EC (dS m^{-1})	TDS (mg L^{-1})
Ultrapure	0–0.00004	0–0.03
Pure	0.00004–0.00046	0.03–0.30
Deionized	0.00046–0.00465	0.30–3.00
Potable	0.00465–1.55270	3.00–1000
Brackish	1.5527–15.5279	1000–10 000
Saline	15.5279–46.5838	10 000–30 000
Sea water	46.5838–77.6397	30 000–50 000

Salinity tolerance of halophyte plants

Salinity tolerance is a key criterion for selecting halophyte species suitable for reusing brine from desalination processes, thereby reducing environmental impact and supporting circular economy strategies. Halophyte species exhibit varying tolerance levels (high, medium, and low) depending on their physiological adaptations (Table 3).

Table 3. Salinity tolerance classification of halophyte crops based on electrical conductivity (EC) and total dissolved solids (TDS) values.

Crop	Salinity tolerance		Tolerance level TDS (mg L^{-1})	Reference
	EC (dS m^{-1})	TDS (mg L^{-1})		
Suaeda salsa (L.) Pall.	62.11	40 000	High >30 000	Wang *et al*. (2023)
Salicornia bigelovii Torr.	55	35 420		Cárdenas-Pérez *et al*. (2021), García-Galindo *et al*. (2021)
Rhizophora mangle L.	54.34	35 000		Lopes *et al*. (2023)
Chenopodium quinoa Willd.	54	34 776		Sánchez and Matos (2018)
Salicornia europaea L.	40	25 760	Medium 10 000–30 000	Araus *et al*. (2021)
Paspalum vaginatum Sw.	15	9960		Chavarria *et al*. (2021)
Puccinellia spp.	13.6	8766		Vaziriyeganeh *et al*. (2022)
Stenotaphrum secundatum (Walter) Kuntze	10	6440	Low 5000–10 000	Ascencios *et al*. (2019)
Zoysa japonica Steud.	10	6440		Hooks *et al*. (2022) Christiansen *et al*. (2022)
Atriplex nummularia Lindl.	9.07	5844		
Atriplex halimus L.	7.9	5087		Gómez-Bellot *et al*. (2021)
Crithmum maritimun L.	7.9	5087		Gómez-Bellot *et al*. (2021)

Cultivation of halophyte plants with different salinity tolerance

Highly tolerant species such as *Suaeda salsa* (L.) Pall. (Figure 4A), *Salicornia bigelovii* Torr. (Figure 4B*), Rhizophora mangle* L. (Figure 4C), and *Chenopodium quinoa* Willd. (Figure 4D) are ideal crops for areas with high salinity levels (Table 3). These halophyte species are adapted to extreme salinity and can be cultivated in arid or coastal zones where saline or brackish water predominates (Sánchez and Matos, 2018). For example, *S. salsa* tolerates salinity up to 40 000 mg L^{-1}, making it suitable for saline soils irrigated with brackish water and useful as forage in arid regions (Wang *et al*., 2023).

Species such as *Salicornia europaea* L. (Figure 4E) are suitable for coastal areas with moderate salinity (Table 3). This crop shows better growth when irrigated with water at 40 dS m^{-1} and can be used for forage, biodiesel, and applications in the pharmaceutical and cosmetic industries (Araus *et al*., 2021). In contrast, species with lower salinity tolerance (Figure 4F–L) perform better in slightly saline soils (Table 3). *Crithmum maritimum* L. (Figure 4F) and *Atriplex halimus* L. (Figure 4G) can be irrigated with reverse osmosis brine at 4.7–7.9 dS m^{-1} without adverse effects on growth,

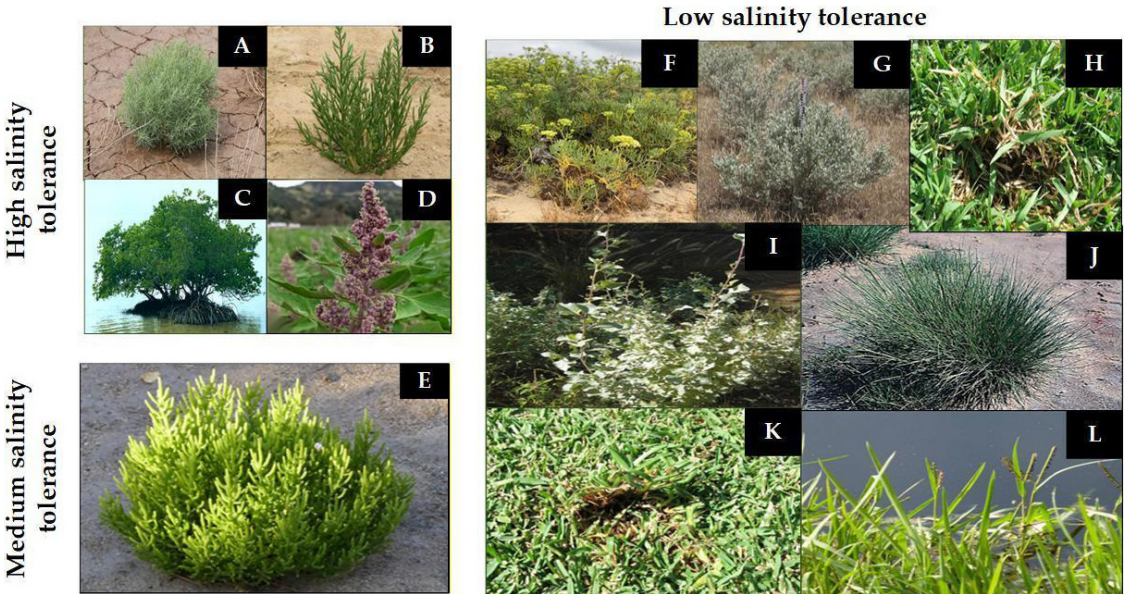

Figure 4. Representative halophyte species with different levels of salt tolerance. A: *Suaeda salsa* (L.) Pall. (Gao *et al*., 2022); B: *Salicornia bigelovii* Torr. (García-Galindo *et al*., 2021); C: *Rhizophora mangle* L. (Lopes *et al*., 2023); D: *Chenopodium quinoa* Willd. (Pizo-Ossa *et al*., 2024); E: *Salicornia europaea* L. (Araus *et al*., 2021); F: *Crithmum maritimum* L. (Roxo *et al*., 2023); G: *Atriplex halimus* L. (Gómez-Bellot *et al*., 2021); H: *Stenotaphrum secundatum* (Walter) Kuntze (Ascencios *et al*., 2019); I: *Atriplex nummularia* Lindl. (Christiansen *et al*., 2022); J: *Puccinellia ciliata* Bor (Çınar and Ünay, 2024); K: *Zoysia japonica* Steud. (Hooks *et al*., 2022); L: *Paspalum vaginatum* Sw. (Chavarria *et al*., 2021).

supporting the use of brine for halophyte irrigation across salinity levels ranging from brackish water to seawater (Gómez-Bellot *et al*., 2021). Likewise, *Atriplex nummularia* Lindl. (Figure 4I) maintains good bromatological quality, particularly in crude protein (14.36 %), organic matter (69.91 %), mineral matter (30.09 %), and dry matter (22.43 %), when irrigated with desalination brine (Christiansen *et al*., 2022). Species of *Atriplex* have been widely studied for their tolerance to saline effluents from brackish water desalination (Sánchez and Matos, 2018).

These crops have significant potential for agriculture and the rehabilitation of saline-affected areas. Using brine from desalination processes represents an alternative irrigation source that reduces discharge volumes and provides added value. However, factors such as brine properties, the irrigation method, volume and frequency, and soil characteristics, including texture, moisture, and ion concentrations, must be considered to assess economic feasibility and prevent environmental risks (Lee *et al*., 2024).

Halophyte species with agroindustrial potential

The production potential of halophytes varies according to species and the environmental conditions. Species with the greatest agroindustrial relevance are

characterized by high salinity tolerance and high biomass production. Some halophyte crops can reach yields of 10–20 Mg of dry biomass per hectare when irrigated with seawater (Table 4). These traits position them as valuable resources for multiple industries (Lee *et al*., 2024).

In the food industry, *Salicornia bigelovii* and *Suaeda salsa* stand out for their high protein content (16–18 %), essential oils (up to 30 %), and carbohydrates, making them valuable for various industrial applications. The oil extracted from *S. bigelovii* is rich in polyunsaturated fatty acids, making it an ideal input for the production of edible oils and nutritional supplements. In addition, *S. bigelovii* and *S. europaea* have been described as potential substitutes for common salt, as consuming salt derived from these plants, with equivalent sodium levels, has little effect on blood pressure. In countries such as Korea and Japan, *Salicornia* is marketed as a gourmet product (Cárdenas-Pérez *et al*., 2021; Wang *et al*., 2023). *Chenopodium quinoa* has a high protein content of 14.1 %, surpassing crops such as rice (8.1 %), wheat (7.2 %), and corn (9.9 %). Its seeds contain essential amino acids, vitamins (A, B2, and E), and minerals such as calcium, magnesium, iron, copper, zinc, and lithium (Yousfi *et al*., 2025), allowing its use in bread, desserts, and other dishes prepared from its grains and leaves (Tourajzadeh *et al*., 2024).

Halophyte species also play an important role in livestock nutrition due to their ability to produce high-quality biomass under saline conditions. Several halophyte species,

Table 4. Advantages and disadvantages of halophyte species of agricultural interest (Hameed *et al*., 2024).

	Advantages	Disadvantages	References
Suaeda salsa (L.) Pall. *Suaeda fruticosa* (L.) Forssk.	Provide cardioprotective foods, support cultivation of salt-tolerant plants, act as nitrogen adsorbents in saline agriculture, and accumulate heavy metals for soil remediation.		Li *et al*. (2023)
Atriplex nummularia Lindl. *Atriplex halimus* L.	Useful for livestock feed, human food, oil extraction, and soil remediation; adaptable to saline conditions and suitable for commercial-scale crops.	Vulnerable to habitat loss, climate change, and invasive species; limited knowledge on mechanisms, optimal conditions, and economic value; barriers to consumer acceptance and commercial expansion.	Navarro-Torre *et al*. (2023)
Chenopodium quinoa Willd. *Crithmum maritimum* L.	Highly salt-tolerant; grow adequately in arid and coastal areas with saline or brackish water; *C. quinoa* provides grains and leaves widely used in food products.		Navarro-Torre *et al*. (2023) Sánchez and Matos (2018) Tourajzadeh *et al*. (2024)
Salicornia bigelovii Torr. *Salicornia europaea* L.	Reduce soil salinity by absorbing sodium; valuable in food and nutraceutical industries as salt substitutes; suitable for commercial cultivation under saline irrigation.		Cárdenas-Pérez *et al*. (2021) Jiang *et al*. (2025) Navarro-Torre *et al*. (2023)

such as *A. nummularia*, *S. salsa*, and *S. bigelovii,* are valuable in animal feed because of their high protein and fiber content (Cárdenas-Pérez *et al*., 2021), while *S. europaea* has also been recognized as a nutritious fodder for livestock (Araus *et al*., 2021). In the energy sector, *Salicornia* seeds are used in biodiesel production due to their oil content (26–33 %), which exceeds that of cotton (15–24 %) and soybeans (17–21 %). Their oil yield can reach 11 442 kg ha^{-1}, surpassing conventional crops and offering a sustainable biofuel option. Species of the genus *Atriplex* have also shown potential for bioethanol production (Cárdenas-Pérez *et al*., 2021). For the pharmaceutical and cosmetics industry, *S. bigelovii*, *S. europaea*, and *S. salsa* contain flavonoids, antioxidants, and phenolic compounds with anti-inflammatory and antibacterial properties, enabling their use in health and skincare products (Araus *et al*., 2021; Wang *et al*., 2023).

From an environmental perspective, genera such as *Suaeda* and *Atriplex* are effective in phytoremediation, as they absorb nitrogen and phosphorus from the soil. Species such as *A. halimus*, *S. bigelovii*, and *S. europaea* can accumulate heavy metals including cadmium, zinc, copper, selenium, nickel, and lead, while *C. quinoa* can accumulate chromium in its tissues (Cárdenas-Pérez *et al*., 2021). Regarding soil salinity mitigation, plants such as *S. fruticosa*, *S. salsa*, and *Sesuvium portulacastrum* L. reduce salinity and alkalinity by absorbing sodium ions and storing salts in their tissues (Jiang *et al*., 2025).

CONCLUSIONS

Water scarcity continues to drive the expansion of desalination technologies, making the sustainable management of resulting brine increasingly important. Utilizing brine as an input for halophyte cultivation offers a circular-economy pathway that can diversify agricultural production and create new economic and food opportunities. However, scientific output on the combined topics of halophytes, reverse osmosis brine, and circular-economy applications remains low. Future research should focus on identifying halophytes with high nutritional and economic value, determining their tolerance thresholds to maximize biomass yield, and assessing viable economic activities that justify cultivation costs. Exploring plant-microorganism interactions that enhance early germination and establishment will be essential to improve halophyte production systems.

ACKNOWLEDGEMENTS

The authors thank the Technological Institute of Sonora through the PROFAPI program academic body 2025-11 and PROFAPI program individual 2025-007 and 2025-035, for funding this Project. We also thank the community of Punta Chueca, Sonora, Mexico, for their contribution to the circular economy perspective of *S. bigelovii*. In addition, we thank the University of Sonora for funding the publication. The images were processed by ITSON chemical engineering student Emilio Medina-Bojórquez.

REFERENCES

Ahmed MA, Amin S, Mohamed AA. 2023. Fouling in reverse osmosis membranes: Monitoring, characterization, mitigation strategies and future directions. Heliyon 9 (4): e14908. https://doi.org/10.1016/j.heliyon.2023.e14908

Al-Tamimi M, Green S, Dahr WA, Al-Muaini A, Lyra D, Ammar K, Dawoud M, Kenyon P, Kemp P, Kennedy L, *et al*. 2023. Drainage, salt-leaching impacts, and the growth of *Salicornia bigelovii* irrigated with different saline waters. Agricultural Water Management 289: 108512. https://doi.org/10.1016/j.agwat.2023.108512

Álvarez-Sánchez J, Dévora-Isiordia GE, Villegas-Peralta Y, Chaparro-Valdez LE, Meza-Tarin SA, Muro-Urista CR, Sánchez-Duarte RG, Pérez-Sicairos S, Medina-Bojorquez E, Rascon-Leon S. 2025. Analysis and prediction of concentration polarization in a pilot reverse osmosis plant with seawater at different concentrations using Python software. Processes 13 (10): 3139. https://doi.org/10.3390/pr13103139

Araus JL, Rezzouk FZ, Thushar S, Shahid M, Elouafi IA, Bort J, Serret MD. 2021. Effect of irrigation salinity and ecotype on the growth, physiological indicators and seed yield and quality of *Salicornia europaea*. Plant Science 304: 110819. https://doi.org/10.1016/j.plantsci.2021.110819

Armendáriz-Ontiveros MM, Dévora-Isiordia GE, Rodríguez-López J, Sánchez-Duarte RG, Álvarez-Sánchez J, Villegas-Peralta Y, Martínez-Macias MR. 2022. Effect of temperature on energy consumption and polarization in reverse osmosis desalination using a spray-cooled photovoltaic system. Energies 15 (20): 7787. https://doi.org/10.3390/en15207787

Ascencios D, Montalvo N, Meza K, Cifuentes A. 2019. Irrigation coefficients of turfgrass (*Stenotaphrum secundatum*) in arid conditions, under water stress using subsurface drip irrigation. Scientia Agropecuaria 10 (2): 207–216. https://doi.org/10.17268/sci.agropecu.2019.02.06

Ashraf HM, Al-Sobhi SA, El-Naas MH. 2022. Mapping the desalination journal: A systematic bibliometric study over 54 years. Desalination 526: 115535. https://doi.org/10.1016/j.desal.2021.115535

Behera SS, Ramachandran S. 2021. Potential uses of halophytes for biofuel production: Opportunities and challenges. *In* Sustainable Biofuels. Academic Press: New York, NY, USA, pp: 425–448. https://doi.org/10.1016/B978-0-12-820297-5.00015-3

Benahmed A, Bessedik M, Abdelbaki C, Mokhdar SA, Goosen MFA, Höllerman B, Zouhiri A, Badr N. 2025. Investigating the long-term economic sustainability and water production costs of desalination plants: A case study from Chatt Hilal in Algeria. Egyptian Journal of Aquatic Research 51 (1): 31–38. https://doi.org/10.1016/j.ejar.2024.11.011

Biswas JC, Naher UA. 2019. Soil nutrient stress and rice production in Bangladesh. *In* Advances in Rice Research for Abiotic Stress Tolerance. Woodhead Publishing, London, UK, pp: 431–445. https://doi.org/10.1016/B978-0-12-814332-2.00021-6

Blanco-Murillo F, Marín-Guirao L, Sola I, Rodríguez-Rojas F, Ruiz JM, Sánchez-Lizaso JL, Sáez CA. 2023. Desalination brine effects beyond excess salinity: Unravelling specific stress signaling and tolerance responses in the seagrass *Posidonia oceanica*. Chemosphere 341: 140061. https://doi.org/10.1016/j.chemosphere.2023.140061

Cárdenas-Pérez S, Piernik A, Chanona-Pérez JJ, Grigore MN, Perea-Flores MJ. 2021. An overview of the emerging trends of the *Salicornia* L. genus as a sustainable crop. Environmental and Experimental Botany 191: 104606. https://doi.org/10.1016/j.envexpbot.2021.104606

Chavarria M, Wherley B, Jessup R, Chandra A. 2021. Physiological responses to salinity among warm-season turfgrasses of contrasting salinity tolerance. Journal of Agronomy and Crop Science 207 (4): 669–678. https://doi.org/10.1111/jac.12501

Chi Y, Fan M, Zhang Z, Qu Y. 2025. Zoning the soil salinization levels in the northern China's coastal areas based on high-resolution soil mapping. Ecological Indicators 172: 113303. https://doi.org/10.1016/j.ecolind.2025.113303

Christiansen AHC, Norman HC, Andreasen C. 2022. Utilization of the halophytic shrubs *Atriplex nummularia* Lindl. and *Rhagodia preissii* Moq. as crops in salt-affected semi-arid regions: How temperature, salinity, seed weight and size affect seed germination. Frontiers in Plant Science 13. https://doi.org/10.3389/fpls.2022.989562

Çınar VM, Ünay A. 2024. Responses of alkali grass (*Puccinellia ciliata* Bor) genotypes to geothermal water. Türk Tarım ve Doğa Bilimleri Dergisi 11 (2): 566–573. https://doi.org/10.30910/turkjans.1445527

Cornejo-Ponce L, Vilca-Salinas P, Arenas MJ, Lienqueo-Aburto H, Moraga-Contreras C. 2022. Use of saline waste from a desalination plant under the principles of the circular economy for the sustainable development of rural communities. *In* Zhang T. (ed.), The Circular Economy. Recent Advances in Sustainable Waste Management. IntechOpen. https://doi.org/10.5772/intechopen.105409

de Serio F, de Padova D, Chiaia G, Ben Meftah M, Mossa M. 2025. Brackish water vs. brine outfall: Impact of desalination plant discharge in vulnerable coastal sites. Desalination 615: 119291. https://doi.org/10.1016/j.desal.2025.119291

Dévora-Isiordia GE, Canales-Elorduy AG, Chávez-Guillén R. 2021. Brackish groundwater and solar energy for desalination plants. Sustainable Water Resources Management 8 (1). https://doi.org/10.1007/s40899-021-00591-z

Dévora-Isiordia GE, Cásares-de la Torre CA, Morales-Mendívil DP, Montoya-Pizeno R, Velázquez-Limón N, Aguilar-Jiménez JA, Ríos-Arriola J. 2023. Evaluation of concentration polarization due to the effect of feed water temperature change on reverse osmosis membranes. Membranes 13 (1): 3. https://doi.org/10.3390/membranes13010003

Dévora-Isiordia GE, Robles Lizárraga A, Fimbres Weihs GA, Álvarez Sánchez J. 2017. Comparación de métodos de descarga para vertidos de salmueras, provenientes de una planta desalinizadora en Sonora, México. Revista Internacional de Contaminacion Ambiental 33 (2): 45–54. https://doi.org/10.20937/rica.2017.33.esp02.04

Espinoza HA. 2023. Economía circular: una aproximación a su origen, evolución e importancia como modelo de desarrollo sostenible. Revista de Economia Institucional, 25 (49): 109–134. https://doi.org/10.18601/01245996.v25n49.06

Feo-García J, Pulido-Alonso A, Florido-Betancor A, Florido-Suárez NR. 2024. Cost studies of reverse osmosis desalination plants in the range of 23,000–33,000 m^3/day. Water 16 (6): https://doi.org/10.3390/w16060910

Feria-Díaz JJ, Correa-Mahecha F, López-Méndez MC, Rodríguez-Miranda JP, Barrera-Rojas J. 2021. Recent desalination technologies by hybridization and integration with reverse osmosis: A review. Water 13 (10). https://doi.org/10.3390/w13101369

Gao J, Guan B, Ge M, Eller F, Yu J, Wang X, Zuo J. 2022. Can allelopathy of *Phragmites australis* extracts aggravate the effects of salt stress on the seed germination of *Suaeda salsa*? Frontiers in Plant Science 13. https://doi.org/10.3389/fpls.2022.990541

García-Galindo E, Nieto-Garibay A, Troyo-Diéguez E, Lucero-Vega G, Murillo-Amador B, Ruiz-Espinoza FH, Fraga-Palomino HC. 2021. Germination of *Salicornia bigelovii* (Torr.) under shrimp culture effluents and the application of vermicompost leachate for mitigating salt stress. Agronomy 11 (3): 424. https://doi.org/10.3390/agronomy11030424

Gil-Trujillo A, Sadhwani Alonso JJ. 2023. Evaluation of brine management strategies from the perspective of the sustainable development goals: Application in the canary islands. Desalination 554: 116483. https://doi.org/10.1016/j.desal.2023.116483

Gómez-Bellot MJ, Lorente B, Ortuño MF, Medina S, Gil-Izquierdo Á, Bañón S, Sánchez-Blanco MJ. 2021. Recycled wastewater and reverse osmosis brine use for halophytes irrigation: Differences in physiological, nutritional and hormonal responses of *Crithmum maritimum* and *Atriplex halimus* plants. Agronomy 11 (4): 627. https://doi.org/10.3390/agronomy11040627

Grammatiki K, de Jonge N, Nielsen JL, García-Gomez SC, Avramidi E, Lymperaki MM, Marcou M, Ioannou G, Papatheodoulou M, Dargent O, *et al*. 2025. eDNA metabarcoding of marine invertebrate communities at reverse osmosis desalination plant outfalls in Cyprus. Marine Pollution Bulletin 214: 117609. https://doi.org/10.1016/j.marpolbul.2025.117609

Hameed A, Hussain S, Rasheed A, Ahmed MZ, Abbas S. 2024. Exploring the potentials of halophytes in addressing climate change-related issues: A synthesis of their biological, environmental, and socioeconomic aspects. World 5 (1): 36–57. https://doi.org/10.3390/world5010003

Hooks T, Masabni J, Ganjegunte G, Sun L, Chandra A, Niu G. 2022. Salt tolerance of seven genotypes of zoysiagrass (*Zoysia* spp.). Technology in Horticulture 2 (1): 1–7. https://doi.org/10.48130/tih-2022-0008

Jiang H, Okoye CO, Ezenwanne BC, Wu Y, Jiang J. 2025. Synergistic potential of halophytes and halophilic/halotolerant plant growth-promoting bacteria in saline soil remediation: Adaptive mechanisms, challenges, and sustainable solutions. Microbiological Research 298: 128227. https://doi.org/10.1016/j.micres.2025.128227

Jones E, Qadir M, van Vliet MTH, Smakhtin V, Kang S. 2019. The state of desalination and brine production: A global outlook. Science of The Total Environment 657: 1343–1356. https://doi.org/10.1016/j.scitotenv.2018.12.076

Lee CH, Ho HJ, Chen WS, Iizuka A. 2024. Total resource circulation of desalination brine: A review. Advanced Sustainable Systems 8 (7). https://doi.org/10.1002/adsu.202300460

Li CY, He R, Tian CY, Song J. 2023. Utilization of halophytes in saline agriculture and restoration of contaminated salinized soils from genes to ecosystem: *Suaeda salsa* as an example. Marine Pollution Bulletin 197: 115728. https://doi.org/10.1016/j.marpolbul.2023.115728

Li J, Hussain T, Feng X, Guo K, Chen H, Yang C, Liu X. 2019. Comparative study on the resistance of *Suaeda glauca* and *Suaeda salsa* to drought, salt, and alkali stresses. Ecological Engineering 140: 105593. https://doi.org/10.1016/j.ecoleng.2019.105593

Liu J, Bai X, Bai Y. 2024. Exploration and case analysis of treatment processes and reuse pathways for industrial brine wastewater in China. Water Cycle 5: 278–285. https://doi.org/10.1016/j.watcyc.2024.08.001

Lopes DMS, Lopes AS, Falqueto AR, Gontijo ABPL, Rogalski M, Tognella MMP. 2023. Photosynthetic and gene expression analyses in *Rhizophora mangle* L. plants growing in field conditions provide insights into adaptation to high-salinity environments. Trees 37 (3): 733–747. https://doi.org/10.1007/s00468-022-02380-3

Lu C, Yuan F, Guo J, Han G, Wang C, Chen M, Wang B. 2021. Current understanding of role of vesicular transport in salt secretion by salt glands in recretohalophytes. International Journal of Molecular Sciences 22 (4): 1–13. https://doi.org/10.3390/ijms22042203

Lu T, Luo P, Wang J, Lu Y, Huo A, Liu L. 2025. Soil salinity accumulation and groundwater degradation due to overexploitation over recent 40-year period in Yaoba Oasis, China. Soil and Tillage Research 248: 106398. https://doi.org/10.1016/J.STILL.2024.106398

Maçaira PM, Tavares Thomé AM, Cyrino Oliveira FL, Carvalho Ferrer AL. 2018. Time series analysis with explanatory variables: A systematic literature review. Environmental Modelling and Software 107: 199–209. https://doi.org/10.1016/j.envsoft.2018.06.004

Mallick S. 2022. Sustainable circular economy design in 2050 for water and food security using renewable energy. *In* Circular Economy and Sustainability, Volume 2: Environmental Engineering. Elsevier: Cambridge, MA, USA, pp: 509–521. https://doi.org/10.1016/B978-0-12-821664-4.00013-3

Montoya-Pizeno R, Morales-Mendivil DP, Cabanillas-López RE, Dévora-Isiordia GE. 2023. The impact of solar intermittency on the concentration polarization factor, water quality and specific energy consumption in the reverse osmosis process. Water 15 (17). https://doi.org/10.3390/w15173022

Musie W, Gonfa G. 2023. Fresh water resource, scarcity, water salinity challenges and possible remedies: A review. Heliyon 9 (8): e18685. https://doi.org/10.1016/j.heliyon.2023.e18685

Myung J, Choi WY, Oh S, Jang K, Kim E, Lee S, Park J, Lee D. 2025. Impact of brine composition on the selective removal of Ca^{2+} and Mg^{2+} via carbonate formation. Chemical Engineering Journal 507: 160537. https://doi.org/10.1016/j.cej.2025.160537

Mzoughi Z, Majdoub H. 2021. Pectic polysaccharides from edible halophytes: Insight on extraction processes, structural characterizations and immunomodulatory potentials. International Journal of Biological Macromolecules 173: 554–579. https://doi.org/10.1016/j.ijbiomac.2021.01.144

Navarro-Torre S, Garcia-Caparrós P, Nogales A, Abreu MM, Santos E, Cortinhas AL, Caperta AD. 2023. Sustainable agricultural management of saline soils in arid and semi-arid Mediterranean regions through halophytes, microbial and soil-based technologies. Environmental and Experimental Botany 212: 105397. https://doi.org/10.1016/j.envexpbot.2023.105397

Oron G, Appelbaum S, Guy O. 2023. Reuse of brine from inland desalination plants with duckweed, fish and halophytes toward increased food production and improved environmental control. Desalination 549: 116317. https://doi.org/10.1016/j.desal.2022.116317

Petersen KL, Paytan A, Rahav E, Levy O, Silverman J, Barzel O, Potts D, Bar-Zeev E. 2018. Impact of brine and antiscalants on reef-building corals in the Gulf of Aqaba – Potential effects from desalination plants. Water Research 144: 183–191. https://doi.org/10.1016/j.watres.2018.07.009

Pizo-Ossa MA, Montes-Rojas C, Aguilera-Arango GA, Ramos-Villafañe YP, López-Hoyos JL, Morillo-Coronado Y. 2024. Evaluation of the phenological cycle and morphological characterization of quinoa genotypes (*Chenopodium quinua* Willd.). Tropical and Subtropical Agroecosystems 27 (2). https://doi.org/10.56369/tsaes.5397

Rahman MM, Mostofa MG, Keya SS, Siddiqui MN, Ansary MMU, Das AK, Rahman MA, Tran LSP. 2021. Adaptive mechanisms of halophytes and their potential in improving salinity tolerance in plants. International Journal of Molecular Sciences 22 (19). https://doi.org/10.3390/ijms221910733

Raudales-Garcia EV, Acosta-Tzin JV, Aguilar-Hernández PA. 2024. Circular economy: A systematic and bibliometric review. Región Científica 3 (1): 2024192. https://doi.org/10.58763/rc2024192

Ríos-Arriola J, Velázquez N, Aguilar-Jiménez JA, Dévora-Isiordia GE, Cásares de la Torre CA, Corona-Sánchez JA, Islas S. 2022. State of the art of desalination in Mexico. Energies 15 (22): 8434. https://doi.org/10.3390/en15228434

Robertson SM, Lyra DA, Mateo-Sagasta J, Ismail S, Akhtar MJU. 2019. Financial analysis of halophyte cultivation in a desert environment using different saline water resources for irrigation. *In* Hasanuzzaman M. Nahar K, Öztürk M. (eds.), Ecophysiology, Abiotic Stress Responses and Utilization of Halophytes. Springer: Singapore, pp: 347–364. https://doi.org/10.1007/978-981-13-3762-8_17

Rotter A, Rinkevich B, Deniz I, Reddy MM, Girão M, Carvalho MF, Gunde-Cimerman N, Gostinčar C, Cueto M, Díaz-Marrero AR, *et al*. 2025. Innovative solutions for valorization of desalination brine. Water Research X 28: 100372. https://doi.org/10.1016/j.wroa.2025.100372

Roxo G, Brilhante M, Moura M, de Sequeira MM, Silva L, Costa JC, Vasconcelos R, Talhinhas P, Romeiras MM. 2023. Genome size variation within *Crithmum maritimum*: Clues on the colonization of insular environments. Ecology and Evolution 13 (4). https://doi.org/10.1002/ece3.10009

Saleem H, Abounahia N, Siddiqui HR, Zaidi SJ. 2023. Qatar desalination research: An overview. Desalination 564: 116802. https://doi.org/10.1016/j.desal.2023.116802

Salehi M. 2022. Global water shortage and potable water safety; Today's concern and tomorrow's crisis. Environment International 158: 106936. https://doi.org/10.1016/j.envint.2021.106936

Sánchez AS, Matos ÂP. 2018. Desalination concentrate management and valorization methods. *In* Sustainable Desalination Handbook: Plant Selection, Design and Implementation. Butterworth-Heinemann: Oxford, UK, pp: 351–399. https://doi.org/10.1016/b978-0-12-809240-8.00009-5

Sarathe S, Baredar PV, Dwivedi G, Tapdiya S, Gaurav A. 2022. Review of various types of renewable-powered desalination technologies with economic analysis. Materials Today: Proceedings 56: 326–335. https://doi.org/10.1016/j.matpr.2022.01.175

Shokry F, Elshikh MR, Elganzoury RME, Komber AO, Shehata AS, Zarraa RM. 2025. Design and optimization of a reverse osmosis system for reject water treatment in a petrochemical plant. Egyptian Journal of Chemistry 68 (12): 419–433. https://doi.org/10.21608/ejchem.2025.362711.11343

Sparks DL, Singh B, Siebecker MG. 2024. The chemistry of saline and sodic soils. *In* Environmental Soil Chemistry (Third edition). Academic Press: Cambridge, MA, USA, pp: 411–438. https://doi.org/10.1016/B978-0-443-14034-1.00010-1

Sweileh WM. 2021. Bibliometric analysis of peer-reviewed literature on antimicrobial stewardship from 1990 to 2019. Globalization and Health 17 (1). https://doi.org/10.1186/s12992-020-00651-7

Teotia N, Chaudhary DR. 2024. Application of halophyte microbiome for development of salt tolerance in crops. *In* Improving Stress Resilience in Plants: Physiological and Biochemical Basis and Utilization in Breeding. Academic Press: Cambridge, MA, USA, pp: 143–164. https://doi.org/10.1016/B978-0-443-18927-2.00001-7

Tourajzadeh O, Piri H, Naserin A, Cahri MM. 2024. Effect of nano biochar addition and deficit irrigation on growth, physiology and water productivity of quinoa plants under salinity conditions. Environmental and Experimental Botany 217: 105564. https://doi.org/10.1016/j.envexpbot.2023.105564

Valdés H, Saavedra A, Flores M, Vera-Puerto I, Aviña H, Belmonte M. 2021. Reverse osmosis concentrate: Physicochemical characteristics, environmental impact, and technologies. Membranes 11 (10): 753. https://doi.org/10.3390/membranes11100753

Vaziriyeganeh M, Carvajal M, Du N, Zwiazek JJ. 2022. Salinity tolerance of halophytic grass *Puccinellia nuttalliana* is associated with enhancement of aquaporin-mediated water transport

by sodium. International Journal of Molecular Sciences 23 (10). https://doi.org/10.3390/ijms23105732

Wang N, Zhao Z, Zhang X, Liu S, Zhang K, Hu M. 2023. Plant growth, salt removal capacity, and forage nutritive value of the annual euhalophyte *Suaeda salsa* irrigated with saline water. Frontiers in Plant Science 13. https://doi.org/10.3389/fpls.2022.1040520

Yousfi S, Shahid M, Thushar S, Ferreira JP, Serret MD, Araus JL. 2025. Effect of irrigation salinity on yield and quality of seeds in different quinoa genotypes. Agricultural Water Management 312: 109413. https://doi.org/10.1016/j.agwat.2025.109413

Zolghadr-Asli B, McIntyre N, Djordjevic S, Farmani R, Pagliero L. 2023. The sustainability of desalination as a remedy to the water crisis in the agriculture sector: An analysis from the climate-water-energy-food nexus perspective. Agricultural Water Management 286: 108407. https://doi.org/10.1016/j.agwat.2023.108407

VOLUME 60, NUMBER 1 | January 01 - February 15, 2026 | MEXICO