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.** 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.

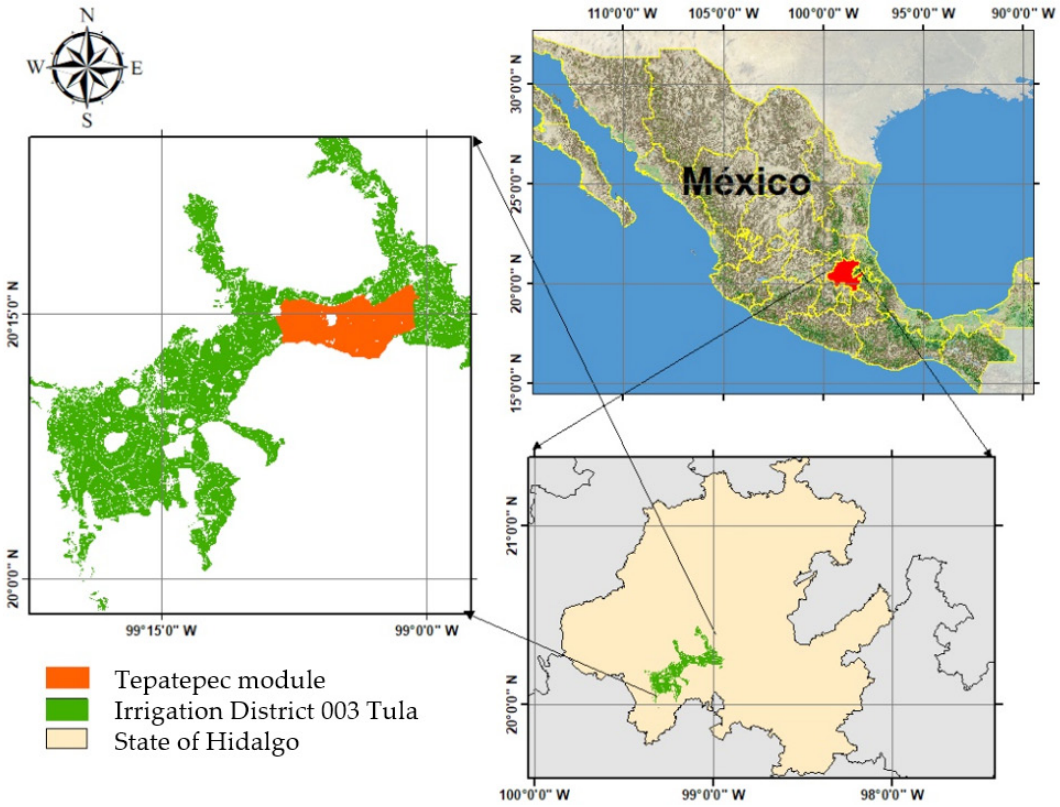

Figure 1. Location of the study area in Irrigation Module Five, Mezquital Valley, in the state of Hidalgo, Mexico.

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.

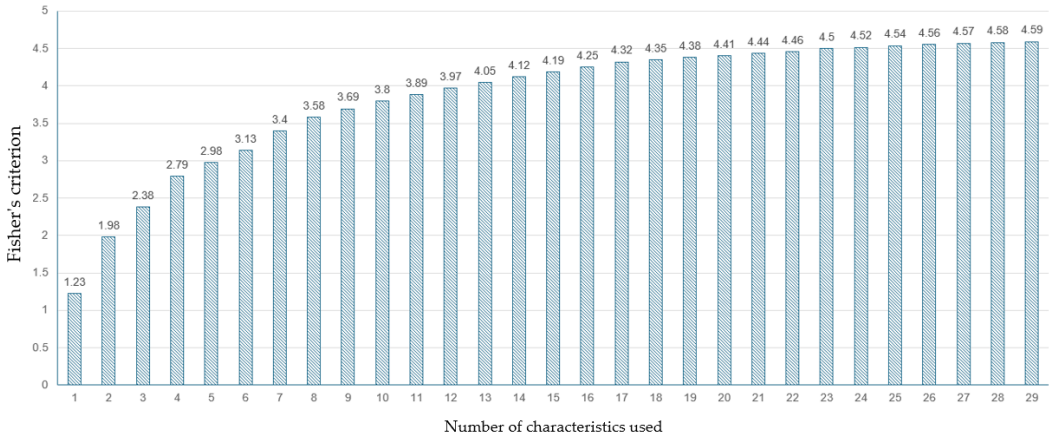

Figure 3. Evolution of Fisher's criterion during the execution of the Sequential Forward Selection (SFS) model.

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).

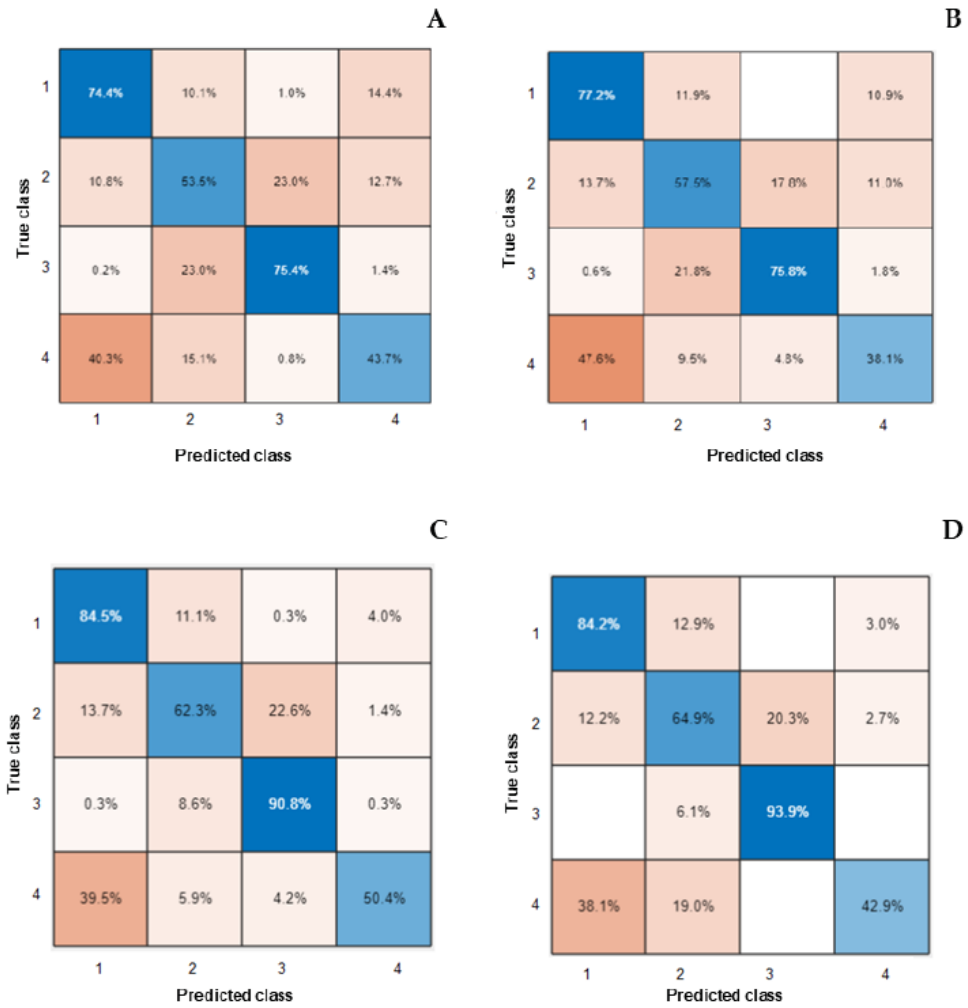

Figure 4. Confusion matrices of the Support Vector Machine (SVM) model. A: trained with all 86 features; B: tested with all 86 features; C: trained with 29 features; D: tested with 29 features.

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