STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–24.
Published online in International Academic Press (www.IAPress.org)

Machine Learning-Based Prediction and Multispectral Analysis for Precision Irrigation Management

Abdelaaziz Bellout [1,*], Azzedine Dliou [1,2], Rachid Latif [1], Amine Saddik [1,2], El Mehdi Cherrat [1,2], Rachid Bouharroud [3]

[1] *Laboratory of Systems Engineering and Information Technology LISTI, National School of Applied Sciences, Ibn Zohr University Agadir, Morocco*

[2] *Faculty of Applied Sciences, Ibn Zohr University, Ait Melloul, Morocco*

[3] *Regional Center of Agricultural Research of Agadir, National Institute of Agricultural Research, Avenue Ennasr, BP 415 Rabat Principale, 10090 Rabat, Morocco*

Abstract The objective of this work is to build a prediction system for normalized indices such as NDVI (Normalized Difference Vegetation Index), NDRE (Normalized Difference RedEdge index) and NDWI (Normalized Difference Water Index). Based on machine learning techniques, this prediction will allow us to compare various methods. Additionally, this prediction will allow us to precisely comprehend these three indices with a small amount of data. Multiple machine learning algorithms were trained and evaluated using appropriate parameters. For NDRE and NDWI prediction, the Support Vector Machine approach produced good results with Mean Squared Errors (MSE) of 0.0006 and 0.0012, respectively. On the other hand, the Random Forest approach performed better with a lower MSE of 0.0033 for predicting NDVI. Furthermore, patterns and trends in crop health, nutrient needs and water requirements were found by clustering analysis. The process of calculating and importing indices from TIFF data was made easier with the creation of a Graphical User Interface (GUI). The system provides an innovative approach for irrigation management, that support farmers in making well-informed decisions regarding irrigation and crop health.

Keywords Machine learning; Computer vision; Precision agriculture; NDVI NDWI; NDRE; Spectral data

DOI: 10.19139/soic-2310-5070-2178

1. Introduction

In today's rapidly growing world, the efficient management of irrigation systems is of crucial importance to maintain sustainable agriculture and effective resource utilization. This work aims to address this pressing demand by developing an innovative irrigation management system that leverages multispectral data analysis and advanced machine learning techniques. Traditional irrigation management methods, such as manual monitoring or sensor- based systems, have limitations in terms of cost, maintenance and provision of comprehensive and timely information on crop health and water requirements. To overcome these challenges, our study proposes an approach that harnesses the potential of multispectral data and advanced machine learning algorithms. Using remote sensing data, farmers can make informed decisions about different aspects of crop management, such as sowing, fertilization,

*Correspondence to: Abdelaaziz Bellout (Email: [abdelaaziz.bellout@edu.uiz.ac.ma]). Laboratory of Systems Engineering and Information Technology LISTI, National School of Applied Sciences, Ibn Zohr University Agadir, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

protection and cultivation. Better crop management, greater yields, and more profitability follow from this. In addition, by monitoring farmland using multitemporal remote sensing images, farmers can gain valuable insights into the efficiency of farming practices such as irrigation and fertilization, enabling them to maximize crop yields and reduce losses[1]. Vegetation indices, such as NDVI, play a critical role in precision agriculture and crop monitoring. These indices provide a reliable assessment of crop condition and health by capturing information on chlorophyll content, leaf area, canopy structure and water status. Monitoring these indices allows farmers to optimize prescription rates for precision farming practices such as variable fertilizer application, irrigation, and pesticide use. Farmers can target particular treatments to increase yield by identifying regions of poor performance or stress [2]. Recent studies have explored using multispectral and remote sensing data along with vegetation indices and machine learning techniques. Lee et al. (2020) used Sentinel-2 data and artificial neural networks to map forest vertical structure in South Korea [3]. Saddik et al. (2021) presented a real-time embedded system for calculating NDVI and NDWI from agricultural images, demonstrating the potential for real-time spectral index computation [4]. Lasko (2022) evaluated using Sentinel-1 and UAVSAR SAR data and texture metrics to fill gaps in Sentinel-2 NDVI and NDWI caused by clouds, achieving high accuracy with random forest models [5]. Nasiri et al. (2022) proposed combining Sentinel-2, UAV images and machine learning to map forest canopy cover, using UAV data for training [6]. These studies showcase the capabilities of integrating multi-source remote sensing data with vegetation indices and machine learning techniques for mapping vegetation structure and gap-filling obscured satellite imagery pixels.

Multiple studies integrate vegetation indices combined with machine learning in agriculture. SVM models using satellite data generate precise yield prediction maps [7]. Padarian et al employed support vector machines for soil moisture prediction using satellite data and climate variables [8]. By integrating diverse datasets, including weather, terrain, and soil data, machine learning significantly enhances the accuracy of crop predictions [9]. Transitioning from theory to practice, Sa et al. developed a novel framework that exemplifies the application of machine learning: using multispectral UAV imagery and deep neural networks for crop and weed segmentation in agricultural fields [10]. Using 9 spectral channels, they achieved improved classification accuracy over RGB alone, with AUCs of 0.839, 0.863 and 0.782 for background, crops and weeds respectively. For vegetation monitoring, machine learning techniques have shown excellent performance. Feng et al. used thirty remotely-sensed drought factors, including NDVI and NDWI, to reproduce a ground-based drought index (SPEI) using three machine learning models. The results showed that the bias-corrected random forest model (BRF) outperformed the other two models (Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP)), providing accurate and reliable drought maps for monitoring and management in South-Eastern Australia [11]. In a comprehensive analysis, Moussaid et al. compared various machine learning models, discovering that combining climatic data with phytosanitary treatment and fertilization details significantly boosts the accuracy of citrus crop yield predictions. The study also used vegetation indexes such as NDVI and GNDVI to extract spectral information from the satellite imagery. The orthonormal automatic pursuit algorithm gave good prediction scores of 0.2489 (MAE: Mean Absolute Error) and 0.0843 (MSE: Mean Squared Error) [12].

The application of NDVI, NDRE, and NDWI in precision agriculture has been well-documented, with each index offering unique insights into vegetation health, chlorophyll content, and water stress, respectively. Selecting them as the primary indices for this study is grounded in their distinct yet complementary capabilities in assessing vegetation health and water content. NDVI is widely recognized for its effectiveness in estimating vegetation density and health, making it a fundamental tool in agricultural and ecological studies. NDRE offers enhanced sensitivity to chlorophyll content, particularly useful in assessing crop health and nutrient status. NDWI, on the other hand, is instrumental in evaluating moisture content in vegetation, crucial for understanding water stress and irrigation management. However, the interrelationships among these indices and their collective impact on irrigation decision-making remain underexplored. This study posits that a deeper understanding of these relationships, facilitated by machine learning algorithms, can enhance the predictive accuracy of irrigation needs,

leading to more precise water management strategies. By leveraging the computational power of ML to analyze multispectral data, this research aims to offer a more nuanced understanding of crop health dynamics, thus contributing to the optimization of irrigation practices in precision agriculture. These indices require a multispectral camera capable of capturing specific bands (Green, Red, RedEdge, and NIR). However, not all cameras, such as Micasense RedEdge MX and Tetracam AWC, support all the necessary bands. The traditional method of capturing images from various bands for calculating indices is expensive and time-consuming compared to our machine learning approach, which requires only a few bands to obtain all the three vegetation indices, thus obtain a holistic view of the crop health.

Throughout this work, we will examine the technical aspects of our methodology, present the results of our experiments and analyses, and discuss the potential implications and future directions of this research. In general, this study successfully demonstrates the application of machine learning techniques in analyzing multispectral data for precision irrigation management. Indeed, four of the most widely used models were compared, namely Support Vector Machine (SVM), Polynomial Regression, Neural Network and Random Forest. Our findings highlight the remarkable predictive accuracy of the SVM model, which excelled in predicting both the NDRE and NDWI with mean squared errors (MSE) of just 0.0006 and 0.0012, respectively. Furthermore, the Random Forest model showed exceptional performance in predicting NDVI, achieving an MSE of 0.0033. The integration of these algorithms into a user-friendly graphical interface not only enhances the system's usability but also empowers farmers with a robust tool for informed decision-making in irrigation and crop health management. This advancement in precision agriculture promises significant impacts on sustainable agricultural practices, optimizing resource use, and enhancing crop health monitoring.

The objective of this research is to model the relationships between key vegetation indices NDVI, NDRE and NDWI using multispectral remote sensing data and machine learning algorithms. The specific contributions are:

- Analyzing relationships among NDVI, NDRE, and NDWI via statistical analyis on a dataset comprising computed index values;
- Training and evaluating supervised regression machine learning models to predict each index from the other two;
- Applying clustering techniques to group samples based on index value patterns;
- Developing a graphical user interface for easy extraction of indices from multispectral bands;
- Providing predictive capabilities when certain bands are missing to support index calculation.

The article's remaining sections are organized as follows: Section 2 describes the importance of precision agriculture and the use of vegetation indices, multispectral data analysis, and machine learning techniques for agriculture. In Section 3, the study area, dataset, and methods for calculating the vegetation index are presented. Results and discussion are presented in Section 4, along with model training and comparison, clustering analysis, and user interface design. Section 5 concludes by summarizing the findings, constraints, and prospects for the future.

2. Remote sensing in precision agriculture

Precision agriculture encompasses a suite of technologies and practices to optimize crop production and increase efficiency. Also known as precision farming or satellite farming, it involves collecting high-resolution data on variability within agricultural fields and implementing targeted interventions tailored to sub-field scales [13]. This contrasts with traditional uniform application of inputs like irrigation, fertilizers, and pesticides across entire fields regardless of localized conditions. Numerous studies have demonstrated the benefits of precision agriculture technologies in optimizing yields, maximizing profits,

and minimizing environmental impacts. For example, variable rate nitrogen fertilizer based on mapping crop needs using aerial imagery increased yields in corn by over 28% [14]. Robertson et al. found that variable rate irrigation matched to soil maps reduced water use by 15% with no yield loss [15]. Targeted application of pesticides lowered agricultural chemical use by up to 70% [16]. Bongiovanni and Lowenberg-DeBoer showed that variable rate inputs and planting increased average returns per acre for corn and soybean [17]. These studies highlight the diverse applications where precision agriculture can improve crop management. However, there are also challenges and limitations. Detailed data collection requires significant technical expertise and investments. Data analytics capabilities are essential to generate actionable insights [18]. Variable rate technologies like smart sprayers are costly. The highly fragmented nature of land holdings can impede adoption, as small farms lack technical skills and resources. Nevertheless, the demonstrated benefits show that precision agriculture can play an important role in sustainably meeting global food demands.

2.1. Vegetation Indices

Analysis of reflectance data from agricultural lands is a key application of remote sensing in precision agriculture [19]. Vegetation indices derived from different spectral bands are particularly important for extracting information on crop conditions and development [20]. These indices provide quantitative measures that indicate specific vegetation characteristics based on the sensitivity of spectral bands to plant properties. For instance, leaf pigments strongly absorb visible light for photosynthesis while cell structure reflects near-infrared signals. Such indices leverage the spectral response signatures of plants to infer their biophysical and structural parameters [21]. The NDVI is one of the most widely used indices for agricultural applications, adopted globally for crop monitoring, yield estimation, and cropland mapping using satellite data [22]. NDVI exploits the contrast between strong chlorophyll absorption in the red band and high reflectance in the near-infrared band, which is affected by leaf mesophyll structure [23].

However, NDVI can saturate in high density vegetation. The NDRE addresses NDVI saturation by replacing red with the narrow red edge band (680-740 nm) which is more sensitive to chlorophyll content [24]. The Soil Adjusted Vegetation Index (SAVI) incorporates a soil adjustment factor to minimize background influences on reflectance [25]. The Enhanced Vegetation Index (EVI) improves sensitivity for high biomass regions by adding blue band information [26]. These and other indices provide capabilities to assess specific crop characteristics. In addition to visible and near-infrared wavelengths, vegetation water content can be assessed using shortwave-infrared bands. The NDWI indicates vegetation water status and drought stress [27] while the Normalized Difference Infrared Index (NDII) is strongly linked to leaf water potential and transpiration [28]. Thermal infrared bands can detect crop water stress and complement spectral indices [29].

These measures distill key information from raw spectral bands to quantify crop status and condition parameters. Analysis of multiple indices provides deeper insights compared to individual indices like NDVI alone. However, interrelationships between spectral indices are not fully characterized. Examining connections between indices like NDVI, NDRE and NDWI can more fully elucidate overall crop conditions. Table 1 presents the equations corresponding to some various precision agriculture indices.

2.2. Multispectral Data Analysis

Multispectral sensors measure reflectance at specific wavelengths across the electromagnetic spectrum. For vegetation monitoring, typical bands include blue, green, red, near-infrared and shortwave infrared. Multispectral data enables derivation of vegetation indices for agricultural applications [30]. Satellite systems like Landsat, Sentinel-2 and MODIS provide free moderate resolution data. For very high resolution, aerial platforms like manned aircraft and UAVs can be equipped with specialized multispectral cameras [31].Analysis of multispectral data requires radiometric and atmospheric corrections to retrieve surface reflectance. Preprocessing may involve steps like cloud masking and image compositing [32]. Once corrected data is available, various information extraction techniques can be applied. Vegetation indices are commonly calculated to quantify crop characteristics. Statistical methods or machine learning can

Table 1. Vegetation indices equation

Index	Equation	Reference	Sensor	Application and description
NDVI	$\frac{B_{\text{NIR}}-B_{\text{Red}}}{B_{\text{NIR}}+B_{\text{Red}}}$	Tucker - 1979	Spectrometer	Data from a spectrometer sensor has been acquired to measure the biomass of plants.
NDRE	$\frac{B_{\text{NIR}}-B_{\text{Rededge}}}{B_{\text{NIR}}+B_{\text{Rededge}}}$	Horler et al. - 1983	Spectral sensor	The data format was derived from a spectral sensor used to identify variations in leaf in a maize farming region.
NDWI	$\frac{B_{\text{G}}-B_{\text{NIR}}}{B_{\text{G}}+B_{\text{NIR}}}$	McFeeters - 1996	Multispectral sensor	The sensor is attached to a satellite that provides hyperspectral and multispectral data to monitor the irrigation system in expansive agricultural regions.
NDMI	$\frac{B_{\text{NIR}}-B_{\text{SWIR}}}{B_{\text{NIR}}+B_{\text{SWIR}}}$	Hardisky et al. - 1983	Radiometer	The monitoring of Spartina alterniflora canopies is conducted using a radiometer-based technique.
GNDVI	$\frac{B_{\text{NIR}}-B_{\text{G}}}{B_{\text{NIR}}+B_{\text{G}}}$	Candiago et al. - 2015	Hyperspectral sensor	A satellite was outfitted with a hyperspectral sensor camera to observe the pigment composition in vegetation.
BSI	$\frac{(B_{\text{R}}+B_{\text{SWIR}})-(B_{\text{NIR}}+B_{\text{B}})}{(B_{\text{R}}+B_{\text{SWIR}})+(B_{\text{NIR}}+B_{\text{B}})}$	Rikimaru et al. - 2002	Hyperspectral sensor	Forest canopy density was extracted using a hyperspectral sensor.
SAVI	$\frac{B_{\text{NIR}}-B_{\text{R}}}{B_{\text{NIR}}+B_{\text{R}}+L}+(1+L)$	Huete - 1988	Hyperspectral sensor	The monitored crops included cotton and grass, and the data was gathered utilizing a satellite-mounted sensor.

relate the dataset variables to ground parameters of interest, such as yield or soil moisture. Multitemporal data can reveal crop growth patterns over the season. Multispectral imagery enables monitoring crop conditions at finer spatial and temporal scales compared to traditional scouting.

3. Materials and Methods

The objective of the current study was to establish a relationship between the three vegetation indices (NDVI, NDRE, NDWI) using supervised machine learning models. We implemented four different models: the support vector regressor (SVM), polynomial regression, the neural network, and the random forest. To begin, we used a multispectral dataset and extracted vegetation indices from this dataset. These clues were recorded in a CSV file, which served as a dataset for our machine learning models. In order to obtain the best prediction performance, we used techniques such as cross-validation and comprehensive hyperparameter optimization for the four monitored models. This process enabled us to pinpoint the optimal parameters for each model, thereby enhancing their capability to accurately model the complex relationships among the vegetation indices.

Next, we performed clustering on the CSV file to detect trends and data groupings. This allowed to better understand the underlying structures of the data and to identify possible similarities between the samples. Finally, we developed a user-friendly graphical interface (GUI) enabling users to import multispectral bands and compute the corresponding vegetation indices. When some necessary bands are not available, the interface also offers the use of machine learning models to predict the value of the missing index.

3.1. Study Area and Dataset

The AgEagle team conducted a crop inspection of a maize-planted area in the state of São Paulo, Brazil, and made the dataset publicly available [33]. The inspection covered an area of 274 hectares and was carried out at an altitude of 120 meters (The drone was flying at an altitude of 120 meters) during the end of the vegetative phase of the crop (February 2021), one of the most sensitive phases for maize cultivation. The AgEagle dataset is an extensive compilation of UAV data obtained from external users of AgEagle's drones, cameras, and sensors . The dataset, originally developed for prospective customers to assess AgEagle's products, has attracted interest from researchers due to its exceptional quality [34, 35].

The dataset is composed of 4 folders and each data contain four images each with a single band (green, nir, red border, red), which have been distributed among the folders. The inspection drone used is an AgEagle eBee Ag, a fixed-wing drone that offers a flight time of up to 55 minutes and can cover 160 hectares in a single flight with standard battery. It was equipped with AgEagle eBee Duet M, a dual-purpose high-resolution RGB and 4-band multispectral mapping camera. The eBee Ag drone is equipped with a wingspan of 116 cm and is constructed of expanded polypropylene. With a weight of 0.8 kg, this drone is propelled by an extremely quiet electric motor. When the drone is at a height of 42 meters, the image resolution is 1.1 cm/pixel for the RGB camera and 4 cm/pixel for the multispectral camera. The battery capacity is 3700 mah. It includes both an RGB camera and a multispectral camera, with green, red, red edge and near-infrared multispectral bands. The multispectral camera in the system is the Sequoia+ sensor. The RGB camera, known as S.O.D.A., provides high-quality visual images. The system lens has a maximum aperture of F/2.8-11 and a focal length of 10.6 mm. The system has a maximum resolution of 5,472 x 3,648 pixels in an aspect ratio of 3:2. It can capture images in JPEG and TIFF formats.

Multispectral data are presented as TIFF format in grayscale. The purpose of TIFF is to describe and store raster data. TIFF describe image data in bilevel, grayscale, palette-color, and full-color in multiple color spaces. A TIFF file begins with an 8-byte header that points to an Image File Directory (IFD). This IFD contains information about the image (metadata), as well as pointers to the actual image data. Grayscale images are a generalization of black and white images. Bilevel images can only store black-and-white image data, but grayscale images can also store grayscales. The actual image data pointed by the IFD is a matrix or grid of pixel values. In the case of grayscale images, the matrix is a single channel of pixel values, where each pixel represents a grayscale from 0 (black) to 255 (white). The width and height of the matrix are specified in the IFD. Figure 1 shows four figures extracted from the dataset. Each figure represents a single spectral band (red edge, red, near-infrared and green) and shows the same area of land seen from different perspectives. Each of these images provides different information, which will be used elsewhere to calculate vegetation indices. In grayscale, high values are represented in white and low values in black.

For each of the 4 folders, 517 images (1280 x 920 pixels) were considered.

3.2. Algorithm study

Our proposed algorithm is dedicated to predicting one vegetation index based on two others based on the available image band. It is composed of many function blocs. After acquiring the images of all bands, the first functional block as described in Algorithm 1 will rename all the band images to simplify the process, then the three indices (NDRE, NDVI, NDWI) are calculated using Algorithm 2 and stored in CSV file. After that, four machine learning models are trained and compared for each index to choose the best model to use for predicting each one. To get meaningful information for the vegetation indices, the K-means model is trained based on the calculated indices to extract insightful data and classify the crop areas based on that values. At the end, a User interface has been developed to facilitate the farmer's task and offer him an elegant solution for monitoring the condition of his farm.

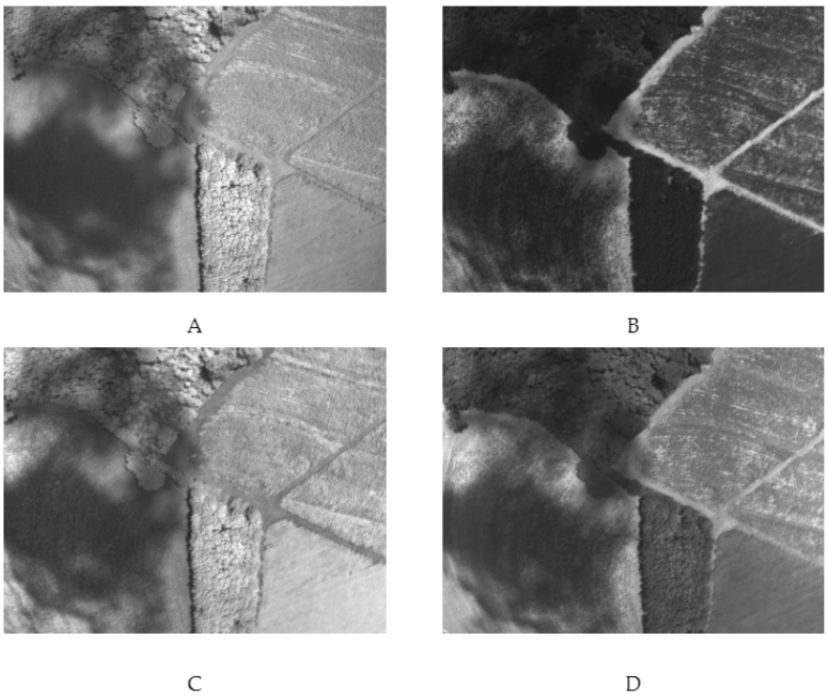

Figure 1. Example of band images: (A) RED EDGE , (B) RED ,(C) NEAR INFRA RED , (D) GREEN

3.3. Vegetation Index Calculation

First of all, each piece of data was browsed one by one. Each piece of data is divided into four folders, so we get four images for each iteration. For each image, we extract the pixel matrix from the TIFF file. In total, we have four matrices. Then, the null values were cleaned from the matrices to avoid division by zero. Using the matrices thus obtained, we calculated the three vegetation indices (NDVI, NDRE and NDWI). Each index uses two matrices to be calculated. Any negative values have been removed because they do not represent vegetation areas and may introduce accuracy issues. Each matrix was represented by its average value and then stored in a row in our CSV file. Each row contained the average value of NDVI, NDRE and NDWI. At the end, the CSV file contained 517 records.

3.3.1. Images pre-processing

The following algorithm describes the process of renaming our files. We had 4 folders, each containing about 517 images. The algorithm iterates over the folders and performs the renaming of the files.

3.3.2. The index extraction algorithm

The algorithm 2 described above is applied to our image dataset in order to calculate the vegetation indices.

The process traverses the dataset and imports the matrices of the individual frames of each band, then calculates the denominator and numerator. For the denominator, if a value is zero, we approximate it with small positive values (close to zero), and for negative values, we replace them with zero in order to capture only the vegetation and obtain more accurate and representative clues. When we calculate the average values, we do not take into account the zero values (corresponding to non-vegetative areas).

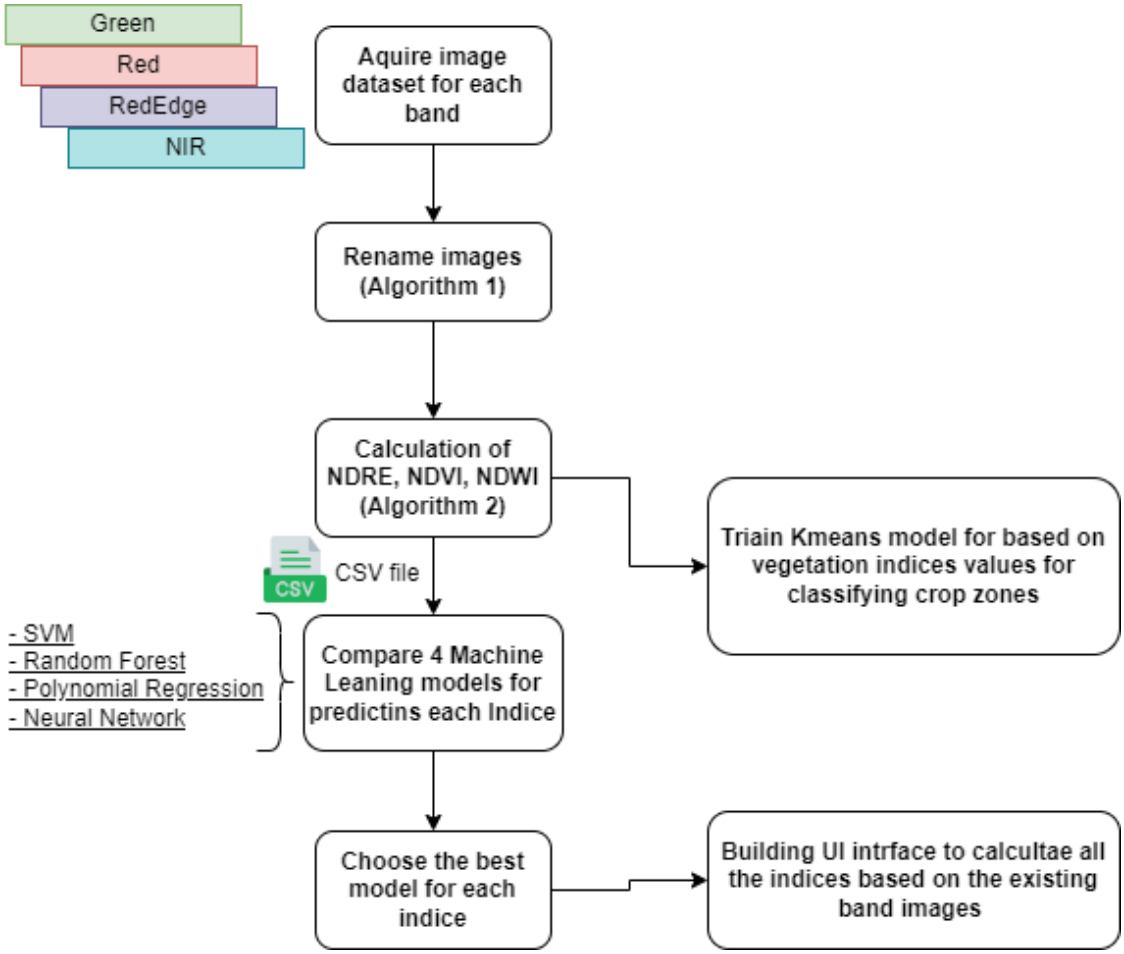

Figure 2. Main algorithm

Algorithm 1: Rename Image

Require: source_dir_green, source_dir_nir, source_dir_red, source_dir_reg
Ensure: Spectral band repositories exist

```
Function RENAME_DIR()
    RENAME_IMAGES(source_dir_green, prefix = "img_green_")
    RENAME_IMAGES(source_dir_nir, prefix = "img_nir_")
    RENAME_IMAGES(source_dir_red, prefix = "img_red_")
    RENAME_IMAGES(source_dir_reg, prefix = "img_reg_")
Function RENAME_IMAGES(directory, prefix)
    files ← get_list_of_files(directory)
    index ← 1
    foreach file in files do
        newName ← prefix + str(index) + ".tiff"
        rename(file, newName)
        index ← index + 1
```

Finally, we save the records in a CSV file and keep them for each sample (each sample is composed of 4 bands). Then we will try to predict each of the three indexes based on the two others. To achieve this,

Algorithm 2: Calculation of NDRE, NDVI, NDWI and checking the accuracy of decimal values

```
Require: source_dir_green, source_dir_nir, source_dir_red, source_dir_reg, csv_file_name
Ensure: CSV file containing the tuples (NDRE, NDVI, NDWI)
Function Calculate_Indices()
    data_file ← open_file(csv_file_name, mode = "write");
    writer ← csv.writer(data_file);
    writer.writerow(['NDRE', 'NDVI', 'NDWI']);
    for i in range(1, 518) do
        gre ← source_dir_green + "/img_green_" + str(i) + ".tiff";
        nir ← source_dir_nir + "/img_nir_" + str(i) + ".tiff";
        red ← source_dir_red + "/img_red_" + str(i) + ".tiff";
        reg ← source_dir_reg + "/img_reg_" + str(i) + ".tiff";

        gre_image ← READ_IMAGE(gre);
        red_image ← READ_IMAGE(red);
        nir_image ← READ_IMAGE(nir);
        reg_image ← READ_IMAGE(reg);

        ndre_denominator ← nir_image + reg_image;
        ndre_denominator[ndre_denominator == 0] ← 0.001;
        NDRE ← (nir_image - reg_image) / ndre_denominator;

        ndvi_denominator ← nir_image + red_image;
        ndvi_denominator[ndvi_denominator == 0] ← 0.001;
        NDVI ← (nir_image - red_image) / ndvi_denominator;

        ndwi_denominator ← gre_image + nir_image;
        ndwi_denominator[ndwi_denominator == 0] ← 0.001;
        NDWI ← -(gre_image - nir_image) / ndwi_denominator;

        NDWI[NDWI ¡ 0] ← 0;
        NDVI[NDVI ¡ 0] ← 0;
        NDRE[NDRE ¡ 0] ← 0;

        avg_ndvi ← mean(NDVI[NDVI ≠ 0]);
        avg_ndwi ← mean(NDWI[NDWI ≠ 0]);
        avg_ndre ← mean(NDRE[NDRE ≠ 0]);

        writer.writeRow(avg_ndre, avg_ndvi, avg_ndwi);
    data_file.close();
```

we will train four machine learning models (**Support Vector Machine (SVM), Polynomial Regression, Neural Network and Random Forest**) and compare the obtained results for each index.

3.4. Machine Learning Models

3.4.1. Support Vector Machine (SVM)

The Support Vector Machine (SVM) model is a powerful, non-probabilistic binary linear classifier that constructs a hyperplane or set of hyperplanes in a high-dimensional space. For linear SVM, the goal is to find the optimal hyperplane that separates the classes with the maximum margin. This is achieved by solving a quadratic optimization problem to minimize $\|\mathbf{w}\|^2$ subject to $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$, where y_i are the labels, $\mathbf{w}$ is the normal vector to the hyperplane, and $\mathbf{b}$ is the bias term. In non-linear cases, SVM uses kernel functions to transform the input space into a higher-dimensional space where a hyperplane can be used for separation. The SVM regressor aims to find a function that best fits the training data by minimizing the error between the predicted values and the actual values. To do this, he seeks to find a hyperplane that separates the data optimally. A hyperplane is a subspace of dimension d 1 in a space of dimension d. In SVM, the hyperplane is defined by the equation :

$$f(x) = w^T x + b \tag{1}$$

where x is the characteristic vector of a sample, w is the vector normal to the hyperplane (T for transpose of w), and b the bias. The function $f(x)$used to predict regressive values. The objective of the SVM regressor is to find the optimal values of w and b by minimizing an objective function and respecting constraints.

3.4.2. Polynomial Regression

Polynomial regression is a statistical technique that extends linear regression by introducing non-linear relationships between the independent and dependent variables. This is achieved by including terms with higher powers of the independent variables in the model. Compared to the straight-line relationship of linear regression, polynomial regression can capture more complex, curvilinear patterns in the data. It is expressed by the following equation:

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_n x^n + \varepsilon \tag{2}$$

Where y is the dependent variable, x is the independent variable. θ_i are the regression coefficients representing the linear and non-linear contributions of each term, respectively, n is the degree of the polynomial, defining the complexity of the non-linearity and ε is the error term accounting for random noise in the data. While polynomial regression can model diverse real-world relationships and interpret its coefficients, but careful model selection and evaluation are needed to avoid overfitting (memorizing training data instead of learning patterns) and multicollinearity. Thus, selecting the right polynomial degree, regularization, and rigorous performance evaluation (e.g., cross-validation) are essential for scientific study.

3.4.3. Neural Network

Neural Networks (NNs) consist of layers of neurons that transmit signals to one another. Feedforward neural networks were among the first and most powerful learning algorithms. They are also called deep networks, multilayer perceptron (MLP) or simply neural networks. As data passes through the artificial mesh of the network, each layer processes an aspect of the data, filters out outliers, identifies familiar entities, and produces the final output. Neural networks are composed of the following:

- Input layer: This layer consists of neurons that receive inputs and transmit them to the other layers. The number of neurons in the input layer must be equal to the attributes or characteristics in the dataset.
- Output layer: Is the predicted characteristic and depends on the type of model.
- Hidden layers: Between the input layer and the output layer, there are hidden layers depending on the type of model. The hidden layers contain a large number of neurons that apply transformations to the inputs before transmitting them. As the network is trained, weights are updated to become more predictive.
- Weight of neurons: Weights are the strength or amplitude of a connection between two neurons. Weights are often initialized to small random values, such as values in the range of 0 to 1.

3.4.4. Random Forest

Random Forest is supervised algorithms that combine several decision trees to address challenges related to regression or classification. In remote sensing land cover classification, the most often used algorithm is random forest. Heterogeneous areas can be classified using RF, a non-parametric machine learning classifier. One decision tree is vulnerable to data variations and overfitting. This issue is usually avoided by using a Random Forest. It yields a large variety of decision trees, each using a distinct sample

of the full dataset and a set of randomly chosen predictors at each node. The researchers have found that RF outperforms decision trees in terms of accuracy (i.e., 92%) [36]. The many classification trees on subsets of the training data are responsible for RF's outstanding accuracy. RF is less affected by noisy datasets and outliers. It performs well at handling high-dimensional, multi-source datasets. Simple to use, RF just requires the user to provide two input parameters: the number of split variables (n-tree) and the number of trees (m-tree).

In order to assess the performance of each model, we use the following metrics: coefficient of determination (R-squared), mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE).

3.4.5. Clustering : K-means

In the clustering phase of this study, we employed a well utilized unsupervised machine learning algorithm, K-means, to categorize the computed vegetation indices according to their respective values. The K-means algorithm stands as a prime example of an efficient and straightforward clustering method, extensively utilized across various applications[37]. At its core, the algorithm initiates by selecting initial clustering centers randomly. It then computes the Euclidean distance of each sample point to these centers. Based on the nearest criterion, points are assigned to the class that shows the greatest similarity to the clustering center. This process involves a continuous update of the clustering centers, iterating until convergence is achieved based on the objective criterion function. This convergence is determined by the mean value of all sample points within each category.

The algorithm's workflow is as follows: The input includes the number of clusters (K) and a database containing N objects. The output is a set of K clusters that minimize the Mean Squared Error (MSE). The process begins with the user specifying the value of K and selecting K random points from the sample as the initial clustering centers. The algorithm then traverses all sample points, calculating the distance from each point to every clustering center based on the nearest distance principle. Points are accordingly assigned to their respective classes. Subsequently, it recalculates the clustering centers by computing the average value of all objects in each class to establish new centers. This recalibration and reassignment process repeat until the target criterion function stabilizes, indicating that the clustering centers no longer change, thereby completing the clustering process.

4. Results and discussion

4.1. Vegetation indices calculation and analysis

4.1.1. Vegetation indices calculation results

The vegetation indices calculation results were stored on CSV file. This file will be used to model the relationship between the NDRE, NDVI and NDWI vegetation indices. We reduced the accuracy of index values as a preprocessing task, as machine learning models are sensitive to the high accuracy of the data. In some cases, when the input values were highly accurate, especially with many digits after the decimal point, this could lead to an increase in the complexity and size of the model, as well as a longer calculation time. In addition, it could introduce unwanted noise or fluctuations into the models.

4.1.2. Visualization of the calculated indicess

Figure 3 shows an RGB (red, green, blue) image of an area of the field. As mentioned earlier, the drone is equipped with a dual camera that simultaneously captures RGB and multispectral images.

Figure 3. Field RGB Image

As with TIFF files, index matrices are 1280 × 960 pixels (one cell per pixel). We use these matrices before calculating averages so that we can visualize them. We can clearly see that the indices really reflect the actual RGB image data.

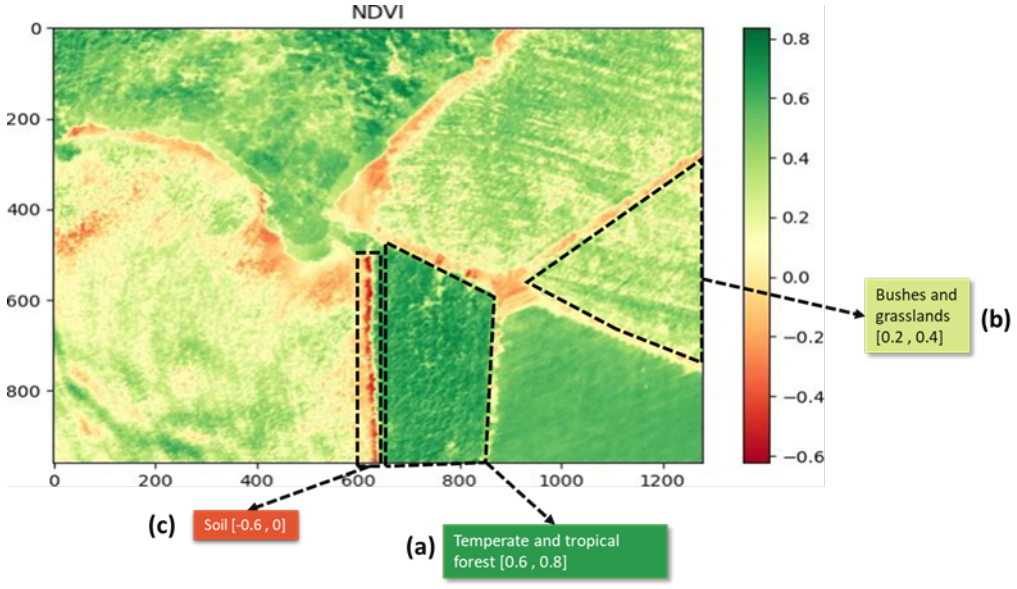

Figure 4. Field NDVI Image

The NDVI map in Figure 4 showcases varying levels of vegetation health, where the spectrum of colors represents the density of plant growth. Dark green patches correspond to lush, thriving vegetation typical of forests, indicating high NDVI values between 0.6 and 0.8 (a). These are areas with abundant chlorophyll activity, reflecting ample NIR light. Areas depicted in lighter green and yellow tones denote grasslands and shrubs with moderate vegetative density, reflected by NDVI scores ranging from 0.2 to 0.4 (b), suggesting less vigorous plant life. The reddish hues represent minimal to non-existent vegetation,

with NDVI values approaching zero or even negative, from -0.6 to 0 (c), typically associated with exposed soil or areas devoid of vegetation. This gradation from red through yellow to green provides a clear visual representation of the varying vegetation conditions across the landscape.

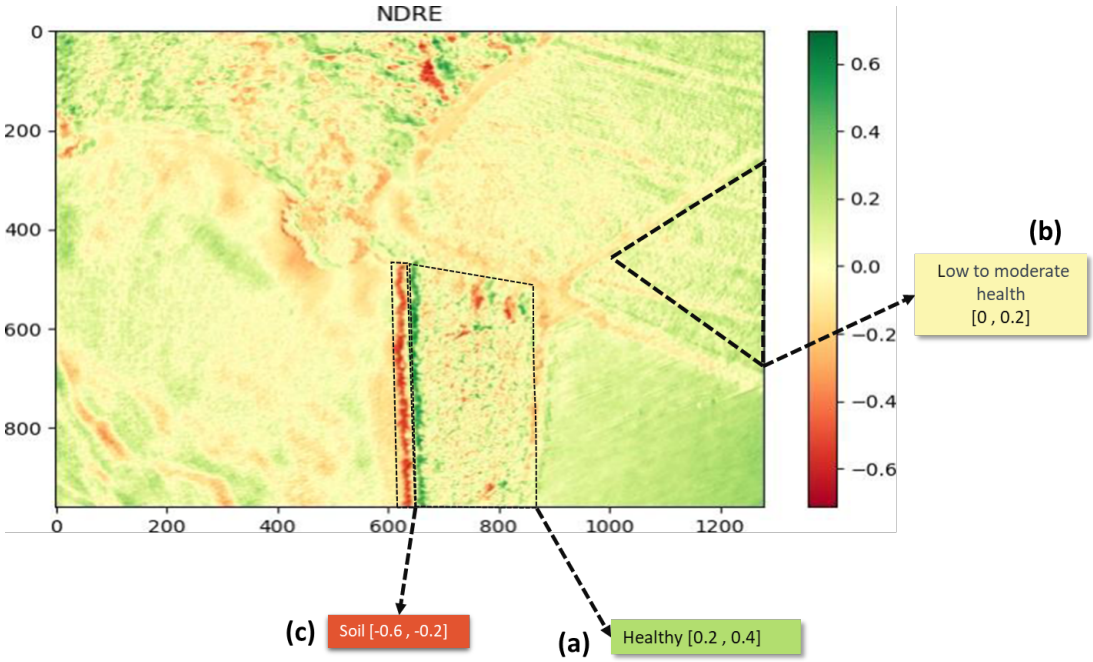

Figure 5. Field NDRE Image

The display of the NDRE index highlights the diverse conditions of crop health present in the agricultural environment (Figure 5). A subtle color spectrum is used to depict the gradient of NDRE values, which shows the health and nutrient status of the vegetation. As illustrated in, zone (a) showed a great value [0.2 , 0.4] which signifies optimal chlorophyll concentration, suggesting that crops are flourishing with potential for high yields. Shifting to Zone (b), with NDRE values ranging between [0, 0.2], we see a transition to paler greens to yellow, pointing to fields with moderate crop vitality. These regions could represent agricultural lands in their growth phase, where the vegetation is healthy but not as dense or mature as in zone (a). Zone (c) is characterized by reddish tinge, pointing to NDRE values between [-0.6, -0.2]. These lower values are typically representative of bare soil or severely stressed vegetation, calling for immediate remedial actions to address possible deficiencies in nutrients or water.

For the NDWI thus, depicted in Figure 6 illustrates varied levels of water content within vegetation and soil moisture, as indicated by the range of colors across the landscape. Vibrant green areas signify vigorous and hydrated vegetation, likely indicative of forests or well-watered fields, with NDWI values between 0.4 and 0.8 (a). These regions reflect high levels of water content within the plant leaves and are suggestive of healthy, well-hydrated vegetation. Transitioning to lighter green and yellowish regions, we observe areas representative of bushes and grasslands with moderate water content, where NDWI values range from -0.2 to 0.2 (b). This variation suggests a gradient in plant hydration levels, which could be attributed to different vegetation types or varying degrees of soil moisture. The reddish and orange tones, particularly noticeable in the bottom left corner, correspond to low NDWI values, ranging from -0.4 to -0.6 (c). These colors typically denote bare soil or regions with low soil moisture and minimal to no vegetation, which could be due to dry soil conditions or areas where vegetation is sparse or non-existent. For the NDWI, we multiply the index by -1 to obtain values similar to the other indices and to avoid negative values in the CSV file. This is due to the fact that healthy and dense vegetation reflects more in

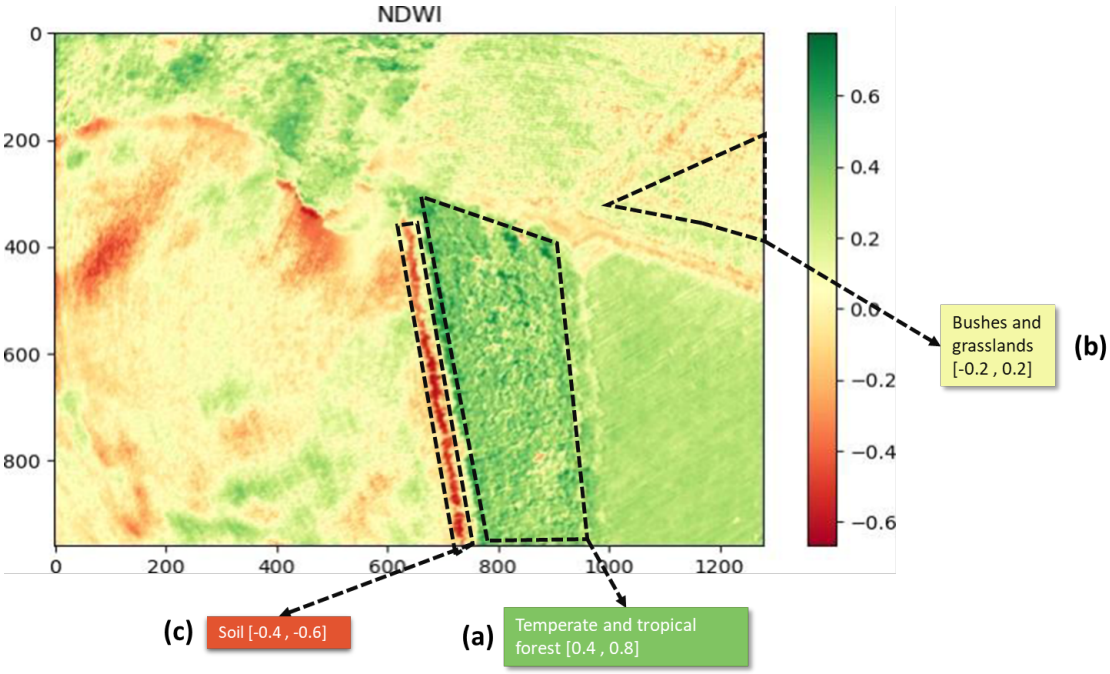

Figure 6. Field NDWI Image

the near-infrared than in the green band, which would give a negative value for the expression (green - near-infrared). However, it is important to note that the interpretation of index values and color densities may vary depending on the context and specific objectives of the study.

4.1.3. The correlation matrix

The correlation matrix reveals significant associations between vegetation indices (Figure 7). NDVI and NDWI have a correlation of 0.81, indicating a positive relationship between vegetation density and soil moisture content.

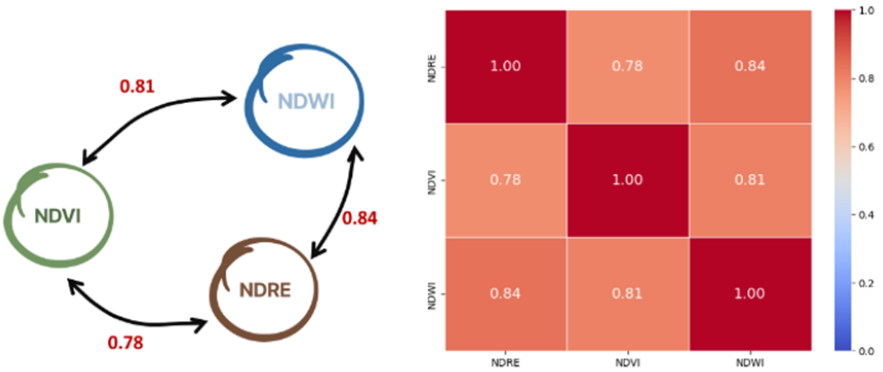

Figure 7. Correlation matrix of the three indices

NDVI and NDRE are strongly correlated at 0.78, highlighting a positive association between vegetation density measured by these indices. In addition, NDRE and NDWI show a correlation of 0.84, highlighting their positive relationship. These results confirm the interaction between vegetation density and soil moisture content, influencing vegetation index measurements.

4.1.4. Compact Line Chart

This graph clearly illustrates the strong linear relationship between the three vegetation indices (Figure 8), which gives us an idea of the machine learning algorithms we can use later.

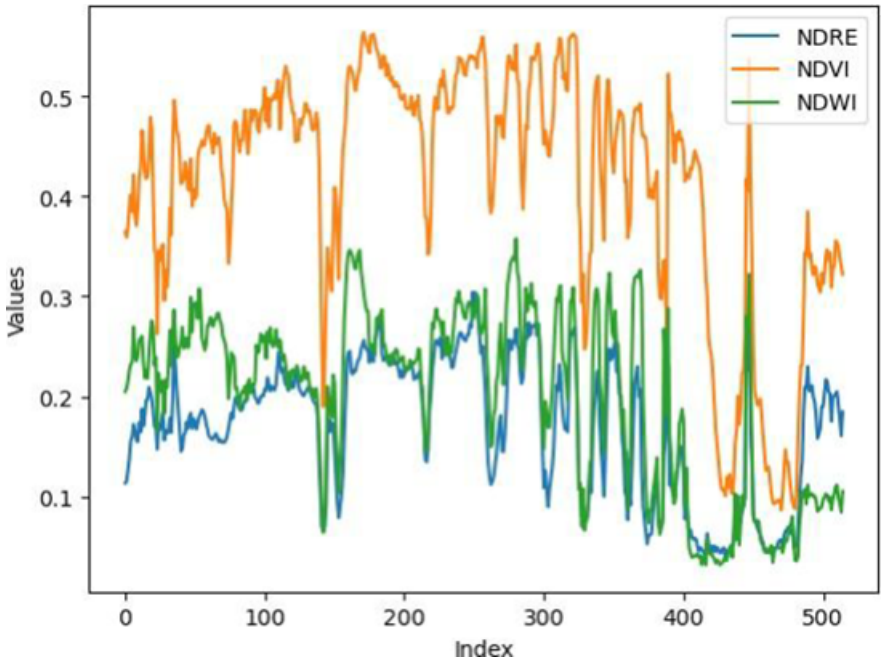

Figure 8. The Compact line chart of the three indices

It highlights the correlation between NDRE, NDVI and NDWI index values, which can be exploited in the choice and development of machine learning models for vegetation data analysis.

4.2. Models training and comparison

4.2.1. NDRE model prediction

We will train the four machine learning models (**Support Vector Machine (SVM), Polynomial Regression, Neural Network and Random Forest**) to obtain a function capable of predicting the value of the NDRE index from the other two indices, NDVI and NDWI. We consider the vector X = (NDVI, NDWI) as our two-dimensional "features" and Y = NDRE as our "target".

The results of the different models showed in Figure 9 are as follows:

Random Forest: The model shows a good fit with a coefficient of determination (R^2) of 0.8514, indicating that 85.14% of the variance of the data is explained. The mean square error (RMS) is low at 0.0007, which means that the predictions are very close to the real values. The model also displays a square root of mean square error (RMSE) of 0.0272 and a mean absolute error (MAE) of 0.0196.

Neural Network: The model shows a less strong fit with a coefficient of determination (R^2) of 0.7453, explaining about 74.53% of the variance of the data. The mean square error (RMS) is higher at 0.0013,

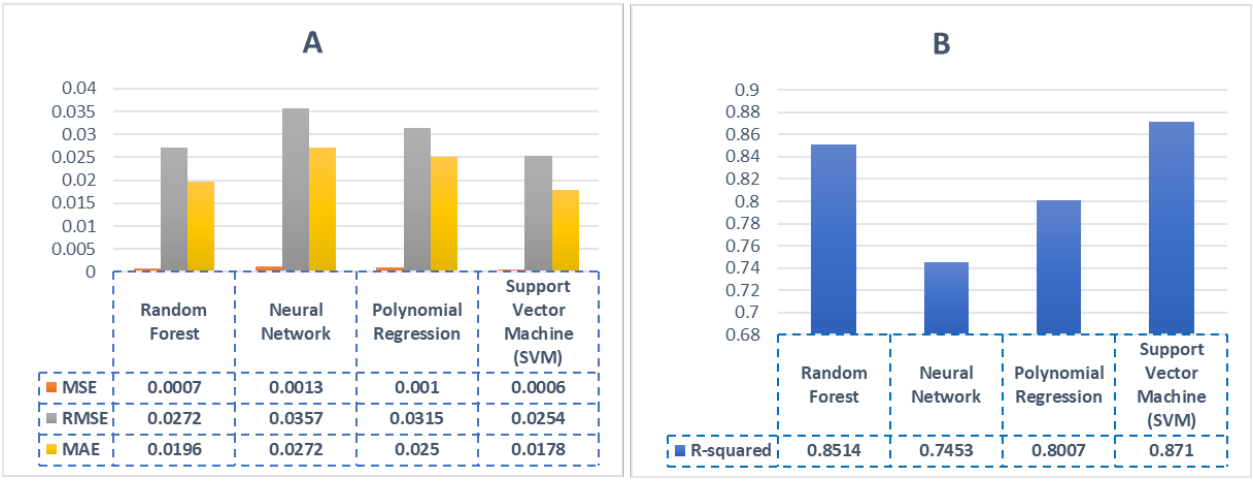

Figure 9. Model results comparison for NDRE :(A) Errors results, (B) R-squared results

indicating less accurate predictions compared to the Random Forest model. The RMSE is 0.0357 and the MAE is 0.0272.

Polynomial Regression: The model displays a reasonable fit with a coefficient of determination (R^2) of 0.8007, explaining about 80.07% of the variance of the data. The mean squared error (RMS) is 0.0010, indicating relatively accurate predictions. The RMSE is 0.0315 and the MAE is 0.0250.

Support Vector Machine (SVM): The model shows a good fit with a coefficient of determination (R^2) of 0.8710, explaining 87.10% of the variance of the data. The mean square error (MSE) is low at 0.0006, indicating predictions very close to the true values. The RMSE is 0.0254 and the MAE is 0.0178.

These results indicate that the SVM model has the best performance with the highest coefficient of determination and the lowest errors, while the Neural Network model shows the lowest performance with a lower coefficient of determination and higher errors. The Random Forest and Polynomial Regression models fall somewhere in between in terms of performance. In our graphical user interface (GUI), we will use the model with the highest predictions for our applications.

4.2.2. NDWI model prediction

Using the same four models (the support vector regressor (SVM), polynomial regression, neural network and random forest), we will choose X = (NDVI, NDRE) and Y = NDWI.

The outcomes of the various models illustrated in Figure 10 are listed below:

Random Forest: The model has a coefficient of determination (R^2) of 0.8438, which means that it explains about 84.38% of the variance of the data. The mean square error (MSE) is 0.0013, indicating that the predictions are relatively close to the actual values. The RMSE is 0.0360 and the MAE is 0.0250.

Neural network: The model displays a coefficient of determination (R^2) of 0.6513, explaining about 65.13% of the variance of the data. The mean square error (MSE) is higher at 0.0029, indicating less accurate predictions compared to the Random Forest model. The RMSE is 0.0538 and the MAE is 0.0436.

Polynomial regression: The model has a coefficient of determination (R^2) of 0.7510, explaining about 75.10% of the variance of the data. The mean squared error (MSE) is 0.0021, indicating relatively accurate predictions. The RMSE is 0.0454 and the MAE is 0.0351.

Support Vector Machine (SVM): The model displays a coefficient of determination (R^2) of 0.8530, explaining about 85.30% of the variance of the data. The mean square error (MSE) is 0.0012, indicating predictions relatively close to the actual values. The RMSE is 0.0349 and the MAE is 0.0266.

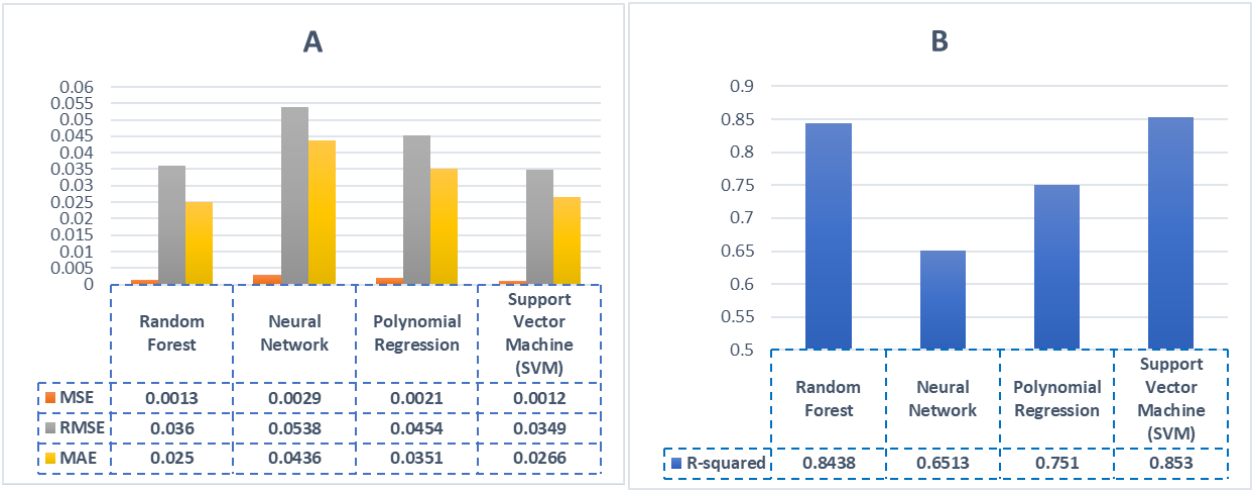

Figure 10. Model results comparison for NDWI :(A) Errors Performance Metrics, (B) R-squared Performance Metric

These results indicate that the SVM model has the best performance with the highest coefficient of determination and the lowest errors. The neural network model shows the lowest performance with a lower coefficient of determination and higher errors. The Random Forest and Polynomial Regression models fall somewhere in between in terms of performance.

4.2.3. *NDVI model prediction*

At the end, we have chosen X = (NDRE, NDWI) and Y = NDVI to train the four models (support vector regressor (SVM), polynomial regression, neural network and random forest).

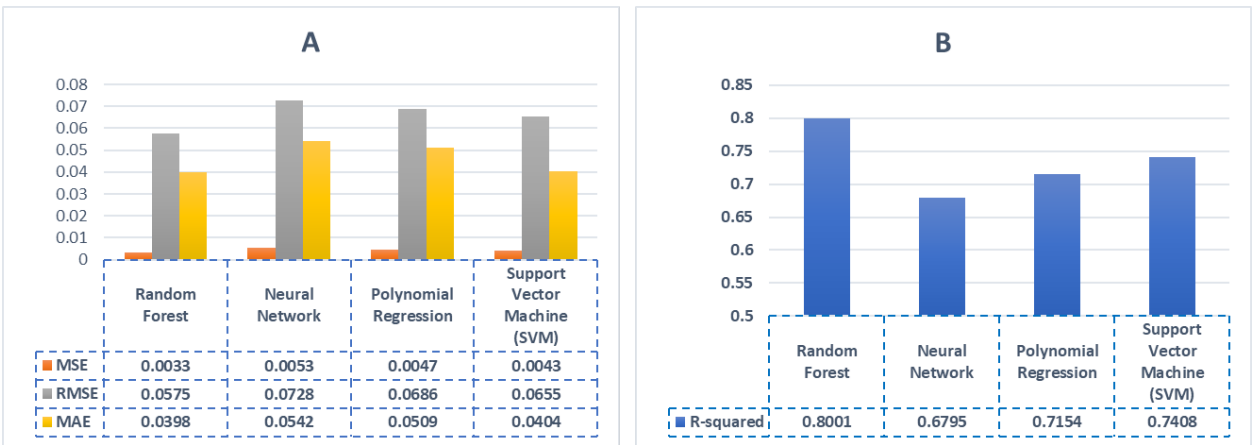

Figure 11. Model results comparison for NDVI :(A) Errors performance metrics, (B) R-squared performance metric

The results of the different NDVI trained models illustrated in Figure 11 are as follows:

Random Forest: The model has a coefficient of determination (R^2) of 0.8001, which means that it explains about 80.01% of the variance of the data. The mean square error (MSE) is 0.0033, indicating that the predictions are relatively close to the actual values. The RMSE is 0.0575 and the MAE is 0.0398.

Neural network: The model displays a coefficient of determination (R^2) of 0.6795, explaining about 67.95% of the variance of the data. The mean square error (MSE) is higher at 0.0053, indicating less accurate predictions compared to the Random Forest model. The RMSE is 0.0728 and the MAE is 0.0542.

Polynomial regression: The model has a coefficient of determination (R^2) of 0.7154, explaining about 71.54% of the variance of the data. The mean square error (MSE) is 0.0047, indicating relatively accurate predictions. The RMSE is 0.0686 and the MAE is 0.0509.

Support Vector Machine (SVM): The model displays a coefficient of determination (R^2) of 0.7408, explaining about 74.08% of the variance of the data. The mean square error (MSE) is 0.0043, indicating predictions relatively close to the actual values. The RMSE is 0.0655 and the MAE is 0.0404.

These results indicate that the Random Forest model has the best performance, with the highest coefficient of determination and the lowest errors. The Neural Network model shows the lowest performance, with a lower coefficient of determination and higher errors. The Polynomial Regression and Support Vector Machine (SVM) models fall somewhere in between in terms of performance.

NOTE: These four models can be employed for the purpose of interpolating missing values within a dataset that encompasses all three indices (NDWI, NDRE, NDVI). Additionally, they can be utilized to compute an index in situations where there exists a hardware limitation, such as a camera that is incapable of capturing all the necessary electromagnetic bands, or in the occurrence of a sensor malfunction.

4.3. Clustering

In this section, we implemented an advanced clustering analysis to transform the vegetation indices data into actionable insights for precision agriculture. The clustering approach serves multiple practical purposes: identifying zones requiring immediate irrigation intervention, detecting areas of nutrient deficiency, and highlighting regions of optimal crop health. This segmentation enables targeted resource allocation and precise intervention strategies. Using the K-means algorithm, complemented by the "Elbow" method, we determined the optimal number of clusters for our dataset. The Elbow technique, selected for its empirical robustness and implementation simplicity, plots the explained variations against cluster numbers to identify the ideal segmentation point (Figure 12).

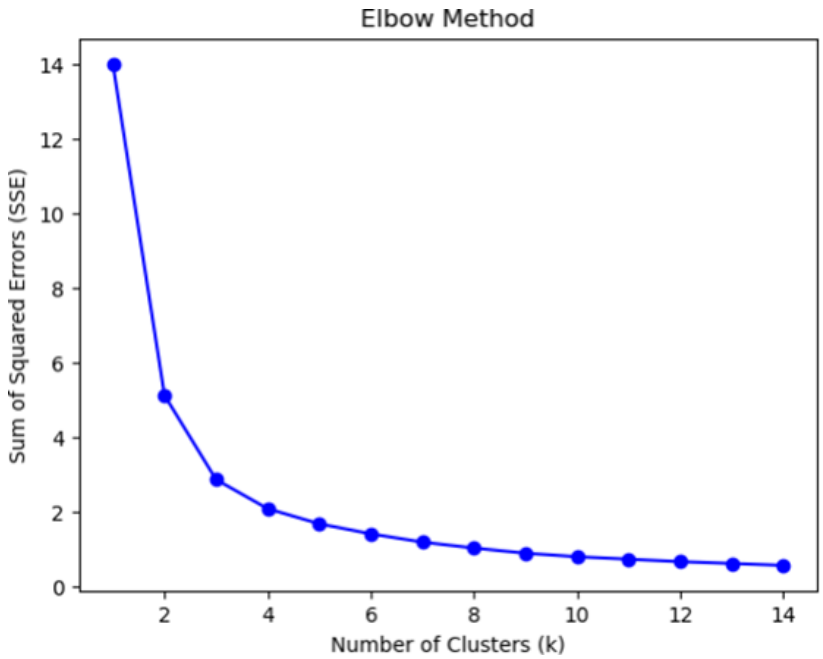

Figure 12. Elbow Curve for K means Clustering

We computed the sum of squared errors within-cluster (SSEWC) to evaluate clustering quality, running the K-means algorithm 14 times with incrementally increasing cluster numbers. The analysis revealed three distinct clusters (Figure 13), each representing specific agricultural management zones:

1. Cluster ”High”: characterized by optimal values across all three vegetation indices, indicating zones with adequate water content, healthy chlorophyll levels, and robust vegetation density. These areas serve as benchmarks for ideal crop conditions.

2. Cluster ”Medium”: represents transitional zones showing early signs of stress but not yet critical. These areas warrant preventive measures and increased monitoring to prevent deterioration.

3. Cluster ”Low”: Identifies critical zones requiring immediate intervention, characterized by poor values across all indices. These areas likely suffer from significant water stress, nutrient deficiencies, or other growth-limiting factors.

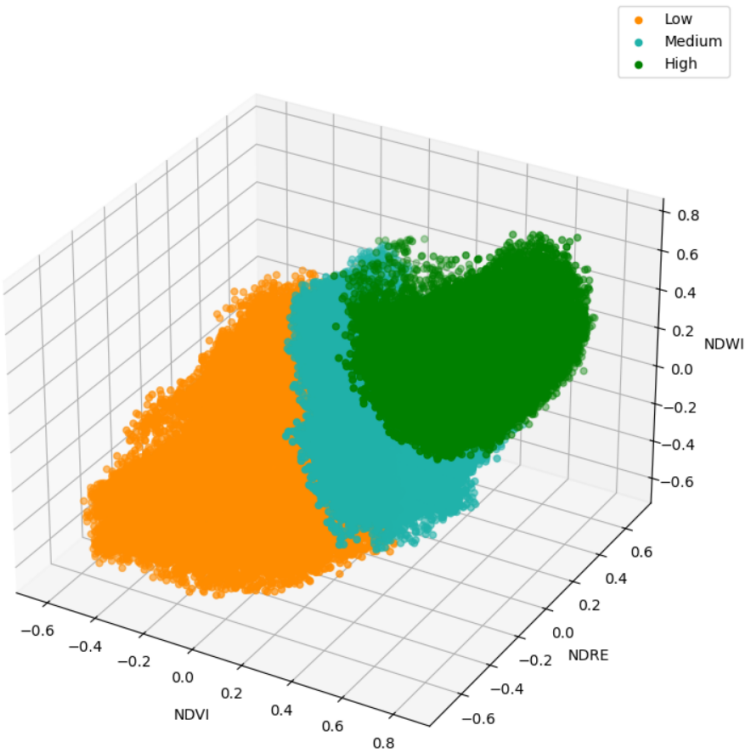

Figure 13. Cluster Visualization

This clustering framework serves as a practical decision-support tool that empowers farmers with data-driven insights for precision agriculture management. The system enables farmers to prioritize irrigation schedules by identifying zones with varying degrees of water stress severity, allowing for more efficient water resource allocation. Additionally, it facilitates targeted fertilizer applications by precisely mapping nutrient-deficient zones, reducing waste and optimizing input costs. The framework also supports strategic resource allocation by helping farmers focus their interventions on critical areas that require immediate attention. Through temporal cluster analysis, farmers can monitor crop health trends over time, identifying patterns and seasonal variations that inform long-term management strategies. Furthermore, the system functions as an early warning mechanism for crop stress detection, enabling proactive interventions before issues become severe and potentially compromise yield. The clustering

results can be directly visualized on crop imagery (Figure 14), providing an intuitive interface for field management decisions. While this model serves as a valuable reference for predicting vegetation states and guiding agricultural interventions, results should be interpreted alongside other agronomic data and field observations for comprehensive decision-making. This clustering approach transforms complex multispectral data into practical management zones, enabling precision agriculture practices that optimize resource use while maximizing crop health and productivity.

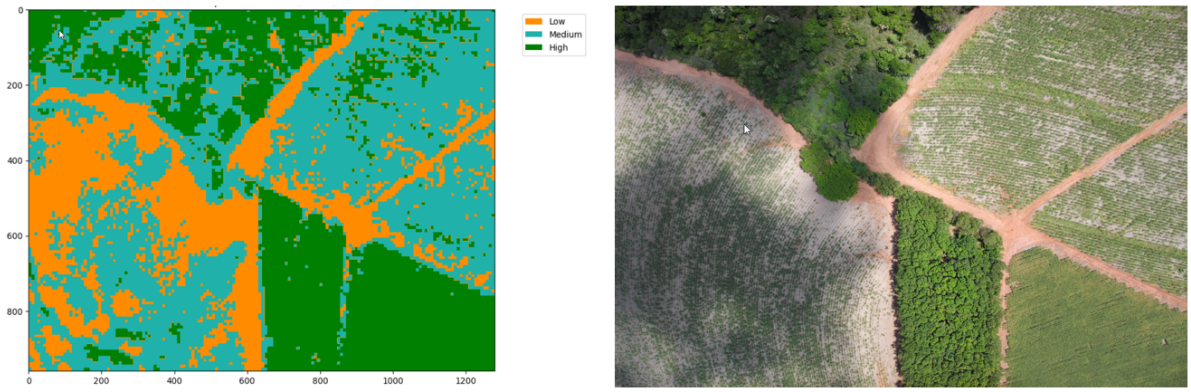

Figure 14. Clustering result applied to crop image :(A) Cluster map visualization, (B) Original Field RGB

4.4. User Interface

Building an application to simplify certain tasks is of crucial importance in our modern society. A well-designed and user-friendly application can save users time, improve productivity and reduce human error. By providing the right technology solution, this application can help farmers optimize their processes. It can enable real-time monitoring of crops. By automating complex tasks and providing valuable data, such an application can contribute to environmental sustainability, economic efficiency, thus strengthening the agricultural sector and responding to our world's growing food production challenges. An application Figure 18 is developed with a user-friendly interface for multispectral image processing. The user can easily import images corresponding to four different bands: Green, Red, NIR and Red edge. A tabbed view is available at the bottom of the interface, allowing the user to view their location. It is important to note that the near-infrared band (NIR) plays an essential role in the calculation of the required indices (NDVI, NDWI, NDRE). Therefore, this band cannot be ignored or omitted when importing images. If the user is unable to import an image for the NIR band, the application will provide tips to remind them to import it. However, if a different band from the NIR band is missing due to material limitations or other issues, the application has a management mechanism to fill in the gaps. We have integrated four pre-trained interpolation algorithms that can estimate missing values for a specific band based on the values of other available bands. This allows the application to calculate the required indices even in the presence of missing data, except for the NIR band, which must always be present.

Image (A) in Figure 15 shows the paths to the 4 imported band images, as well as their visualization in the dedicated panel. After clicking "Calculate", the indices will be calculated, after the calculations are completed, the values of the indices are displayed in the application interface (B), allowing the farmer to visualize and analyze the results obtained. This provides valuable information on the characteristics of the areas studied, such as vegetation, water and reflective elements, which can be used for agricultural, environmental and other applications. This allows them to make clear decisions, identify areas of interest and better understand the environmental or agricultural parameters around them. The clear and accurate display of indices in the application interface facilitates the interpretation of results and allows users to use this information to make relevant and informed decisions. Interpolation algorithms were adopted to

Figure 15. Application User Interface : (A) Image upload, (B) Indices calculation, (C) Ignoring the Green Band, (D) Indices prediction

estimate the missing values. The two images (C) and (D) showed how our application handles the process of interpolating missing values for a specific band. When the user imports the images, the application checks whether all the necessary bands are present. If a band is missing, our interpolation system comes into play to estimate the missing values. As an example, the Green band was not imported, but it was interpolated. When the user imports the images, our application detects the absence of the Green band (C). Thanks to our trained prediction model, we are able to fill this gap by estimating the missing values of NDWI despite the absence of the Green band (D). When the NDVI, NDRE and NDWI indices are calculated, our application offers a functionality to display them graphically (Figure 16). This feature allows users to visualize and analyze spatial variations in indices on imported images. It is important to note that the visualization of indices depends on the availability of data for each specific index. If a band is missing, the calculations associated with this index cannot be performed and, therefore, results cannot be displayed.

5. Conclusion and outlook

The present research explored the integration of advanced machine learning techniques with multispectral analysis for precision irrigation management. Our work primarily focused on developing predictive systems for crucial vegetation indices like NDVI, NDRE, and NDWI, harnessing various

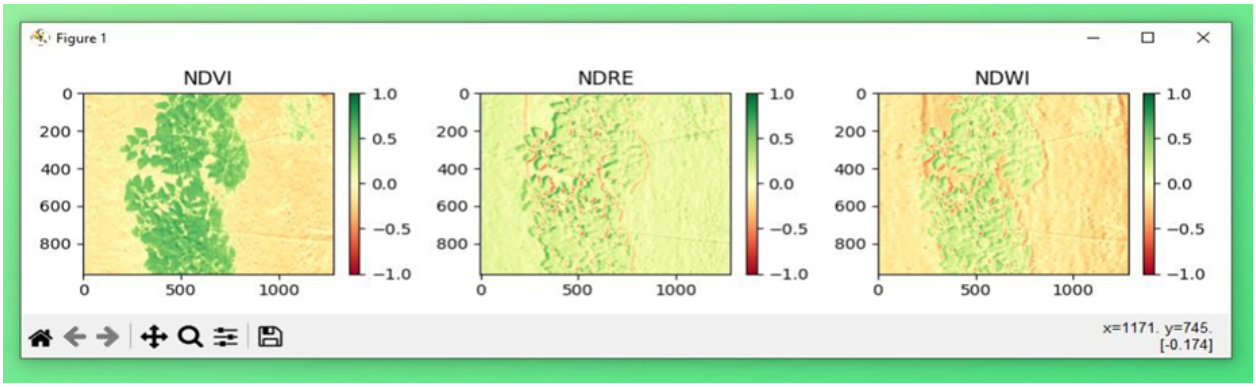

Figure 16. Calculated indices visualization

machine learning models. The Support Vector Machine algorithm, for instance, showed favorable scores with a Mean Squared Error (MSE) of 0.0006 and 0.0012 for NDRE and NDWI, respectively. However, the Random Forest model outperformed with a lower MSE of 0.0033 for NDVI prediction. We also detailed the creation of a graphical user interface that significantly eases the process of index calculation and interpretation. Also as additional work, a clustering model was trained to find relation patterns between indices thus classifying the crop areas according to their characteristics. This study marks a significant stride in precision agriculture, particularly in optimizing irrigation practices through intelligent data analysis.

Beyond the current achievements, our future research directions will focus on expanding the geographical and temporal scope of our analysis across diverse regions, multiple crop types, and different growing seasons. This expansion will strengthen the model's scalability and provide deeper insights into seasonal vegetation patterns. Technical advancements will incorporate additional vegetation indices for more comprehensive crop health assessment, enhanced feature analysis techniques, and improved visualization methods to better explain relationships between spectral bands and indices. The integration of IoT technologies and real-time data analytics represents a key development area, enabling automated decision-making systems that respond dynamically to environmental conditions. Cloud and edge computing adoption will address the challenges of processing large-scale multispectral data, while system integration improvements will create seamless data collection and analysis pipelines, including early warning systems for crop stress detection. These advancements aim to create a more comprehensive, scalable, and practical solution for precision agriculture. Through continued research in these areas, we envision a system that optimizes irrigation management and democratizes access to advanced agricultural technology for farmers worldwide, ultimately contributing to more sustainable and efficient farming practices through the practical application of machine learning and multispectral analysis.

REFERENCES

[1] S. Adhikary, B. Biswas, M. K. Naskar, B. Mukherjee, A. P. Singh, K. Atta, S. Adhikary, B. Biswas, M. K. Naskar, B. Mukherjee, A. P. Singh, and K. Atta, "Remote Sensing for Agricultural Applications," in *Arid Environment - Perspectives, Challenges and Management*, IntechOpen, Sept. 2022.

[2] S. Candiago, F. Remondino, M. De Giglio, M. Dubbini, and M. Gattelli, "Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images," *Remote Sensing*, vol. 7, pp. 4026–4047, Apr. 2015.

[3] Y.-S. Lee, S. Lee, and H.-S. Jung, "Mapping Forest Vertical Structure in Gong-ju, Korea Using Sentinel-2 Satellite Images and Artificial Neural Networks," *Applied Sciences*, 2020.

[4] A. Saddik, R. Latif, M. Elhoseny, and A. El Ouardi, "Real-time evaluation of different indexes in precision agriculture using a heterogeneous embedded system," *Sustainable Computing: Informatics and Systems*, vol. 30, p. 100506, June 2021.

[5] K. Lasko, "Gap Filling Cloudy Sentinel-2 NDVI and NDWI Pixels with Multi-Frequency Denoised C-Band and L-Band Synthetic Aperture Radar (SAR), Texture, and Shallow Learning Techniques," *Remote Sensing*, 2022.

[6] V. Nasiri, A. A. Darvishsefat, H. Arefi, V. C. Griess, S. M. M. Sadeghi, and S. A. Borz, "Modeling Forest Canopy Cover: A Synergistic Use of Sentinel-2, Aerial Photogrammetry Data, and Machine Learning," *Remote Sensing*, 2022.

[7] S. Khaki and L. Wang, "Crop Yield Prediction Using Deep Neural Networks," *Frontiers in Plant Science*, vol. 10, 2019.

[8] J. Padarian, B. Minasny, and A. B. McBratney, "Using deep learning for digital soil mapping," *SOIL*, vol. 5, pp. 79–89, Feb. 2019.

[9] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review," *Sensors*, vol. 18, p. 2674, Aug. 2018.

[10] I. Sa, M. Popović, R. Khanna, Z. Chen, P. Lottes, F. Liebisch, J. Nieto, C. Stachniss, A. Walter, and R. Siegwart, "WeedMap: A Large-Scale Semantic Weed Mapping Framework Using Aerial Multispectral Imaging and Deep Neural Network for Precision Farming," *Remote Sensing*, vol. 10, p. 1423, Sept. 2018.

[11] P. Feng, B. Wang, D. L. Liu, C. Waters, and Q. Yu, "Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia," *Agricultural and Forest Meteorology*, vol. 275, pp. 100–113, Sept. 2019.

[12] A. Moussaid, S. El Fkihi, Y. Zennayi, O. Lahlou, I. Kassou, F. Bourzeix, L. El Mansouri, and Y. Imani, "Machine Learning Applied to Tree Crop Yield Prediction Using Field Data and Satellite Imagery: A Case Study in a Citrus Orchard," *Informatics*, vol. 9, p. 80, Dec. 2022.

[13] H. Auernhammer, "Precision farming — the environmental challenge," *Computers and Electronics in Agriculture*, vol. 30, pp. 31–43, Feb. 2001.

[14] B. Koch, R. Khosla, W. M. Frasier, D. G. Westfall, and D. Inman, "Economic Feasibility of Variable-Rate Nitrogen Application Utilizing Site-Specific Management Zones," *Agronomy Journal*, vol. 96, no. 6, pp. 1572–1580, 2004.

[15] B. Robertson, C. Bednarz, and C. Burmester, "Growth and development: first 60 days," vol. 13, no. 2, 2007.

[16] R. Grisso, M. Alley, W. Thomason, D. Holshouser, and O. Roberson, "Precision farming tools: Variable-rate application," *Precision, Geospatial, & Sensor Technologies*, pp. 442–505, Jan. 2011.

[17] R. Bongiovanni and J. Lowenberg-Deboer, "Precision Agriculture and Sustainability," *Precision Agriculture*, vol. 5, pp. 359–387, Aug. 2004.

[18] S. Fountas, S. Blackmore, D. Ess, S. Hawkins, G. Blumhoff, J. Lowenberg-Deboer, and C. G. Sorensen, "Farmer Experience with Precision Agriculture in Denmark and the US Eastern Corn Belt," *Precision Agriculture*, vol. 6, pp. 121–141, Apr. 2005.

[19] D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps," *Biosystems Engineering*, vol. 114, pp. 358–371, Apr. 2013.

[20] B. Abderrazak, D. Morin, F. Bonn, and A. Huete, "A review of vegetation indices," *Remote Sensing Reviews*, vol. 13, pp. 95–120, Jan. 1996.

[21] J. Xue and B. Su, "Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications," *Journal of Sensors*, vol. 2017, p. e1353691, May 2017.

[22] I. Becker-Reshef, J. Chris, S. Mark, E. Vermote, T. Compton, A. Anyamba, S. Jen, P. Ed, M. Ed, S. Jeff, H. Matthew, P. Kyle, B. Charon, W. Derrick, C. Reynolds, and D. Bradley, "Monitoring Global Croplands with Coarse Resolution Earth Observations: The Global Agriculture Monitoring (GLAM) Project," *Remote Sensing*, vol. 2, June 2010.

[23] N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. Tucker, and N. C. Stenseth, "Using the satellite-derived NDVI to assess ecological responses to environmental change," *Trends in ecology & evolution*, vol. 20, pp. 503–10, Oct. 2005.

[24] J. U. H. Eitel, D. S. Long, P. E. Gessler, and A. M. S. Smith, "Using in-situ measurements to evaluate the new RapidEye™ satellite series for prediction of wheat nitrogen status," *International Journal of Remote Sensing*, vol. 28, pp. 4183–4190, Sept. 2007.

[25] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sensing of Environment*, vol. 25, pp. 295–309, Aug. 1988.

[26] H. Q. Liu and A. Huete, "A feedback based modification of the NDVI to minimize canopy background and atmospheric noise," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 457–465, Mar. 1995.

[27] R. Fensholt and I. Sandholt, "Derivation of a Shortwave Infrared Water Stress Index From MODIS Near- and Shortwave Infrared Data in a Semiarid Environment," *Remote Sensing of Environment*, vol. 87, pp. 111–121, Sept. 2003.

[28] E. R. Hunt and B. N. Rock, "Detection of changes in leaf water content using Near- and Middle-Infrared reflectances," *Remote Sensing of Environment*, vol. 30, pp. 43–54, Oct. 1989.

[29] S. Ma, Y. Zhou, P. H. Gowda, J. Dong, G. Zhang, V. G. Kakani, P. Wagle, L. Chen, K. C. Flynn, and W. Jiang, "Application of the water-related spectral reflectance indices: A review," *Ecological Indicators*, vol. 98, pp. 68–79, Mar. 2019.

[30] C. Zhang and J. Kovacs, "The application of small unmanned aerial systems for precision agriculture: A review," *Precision Agriculture*, vol. 13, Dec. 2012.

[31] Y. Shi, J. A. Thomasson, S. C. Murray, N. A. Pugh, W. L. Rooney, S. Shafian, N. Rajan, G. Rouze, C. L. S. Morgan, H. L. Neely, A. Rana, M. V. Bagavathiannan, J. Henrickson, E. Bowden, J. Valasek, J. Olsenholler, M. P. Bishop, R. Sheridan, E. B. Putman, S. Popescu, T. Burks, D. Cope, A. Ibrahim, B. F. McCutchen, D. D. Baltensperger, R. V. A. Jr, M. Vidrine, and C. Yang, "Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research," *PLOS ONE*, vol. 11, p. e0159781, July 2016.

[32] D. P. Roy, M. A. Wulder, T. R. Loveland, W. C.e., R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy, T. A. Scambos, C. B. Schaaf, J. R. Schott, Y. Sheng, E. F. Vermote, A. S. Belward, R. Bindschadler, W. B. Cohen, F. Gao, J. D. Hipple, P. Hostert, J. Huntington, C. O. Justice, A. Kilic, V. Kovalskyy, Z. P. Lee, L. Lymburner, J. G. Masek, J. McCorkel, Y. Shuai, R. Trezza, J. Vogelmann, R. H. Wynne, and Z. Zhu, "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sensing of Environment*, vol. 145, pp. 154–172, Apr. 2014.

[33] "Corn field mapped with eBee Ag in Brazil," Feb. 2021.

[34] A. Couturier and M. A. Akhloufi, "A Review on Deep Learning for UAV Absolute Visual Localization," *Drones*, vol. 8, p. 622, Nov. 2024. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[35] J. C. Nowack, L. K. Atencia-Payares, A. M. Tarquis, and M. Gomez-del Campo, "Application of Unmanned Aerial Vehicle (UAV) Sensing for Water Status Estimation in Vineyards under Different Pruning Strategies," *Plants*, vol. 13, p. 1350, Jan. 2024. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[36] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, Jan. 2012.

[37] C. Li and B. Niu, "Design of smart agriculture based on big data and Internet of things," *International Journal of Distributed Sensor Networks*, vol. 16, p. 1550147720917065, May 2020.