Vol. 13, No. 4 (2024): 1182 - 1195 http://dx.doi.org/10.23960/jtep-l.v13i4.1182-1195

JURNAL TEKNIK PERTANIAN LAMPUNG

ISSN 2302-559X (print) / 2549-0818 (online)

Journal homepage : https://jurnal.fp.unila.ac.id/index.php/JTP

Prediction of Phenotypic Parameters of Sugarcane Plants Based on Multispectral Drone Imagery and Machine learning

Febri Hasskavendo[1], Mohamad Solahudin[1,✉], Supriyanto[1], Slamet Widodo[1]

[1] Department of Mechanical and Biosystem Engineering, Faculty of Agricultural Engineering and Technology, IPB University, Bogor, INDONESIA.

Article History:

Received : 05 April 2024
Revised : 09 June 2024
Accepted : 22 June 2024

Keywords:

Machine learning,
Multispectral drone imagery,
Phenotypic parameter,
Plant productivity,
Sugarcane.

Corresponding Author:
✉ mohamadso@apps.ipb.ac.id
(Mohamad Solahudin)

ABSTRACT

Measuring phenotypic parameters is important in evaluating the productivity of sugarcane. Existing manual measurements are considered less efficient, so a better alternative method is needed. This research aims to explore the potential of using multispectral drone imagery and machine learning to estimate phenotypic parameters of sugarcane plants that are efficient, accurate, inexpensive, and support sustainable agricultural practices. Spectrum data captured by drones, namely Green, Red, RedEdge and NIR are used as inputs to estimate phenotypic parameters including brix value, number of stands, stem diameter, and plant height. Based on the results of machine learning model development, the ANN algorithm model is most effective in predicting Brix Value with R^2 0.74 and RMSE 0.06 and number of stands with R^2 0.68 and RMSE 2.13. All models could not predict stem diameter and plant height well. The best model to predict plant height was obtained by RF algorithm with R^2 0.53 and RMSE 14.09. SVR algorithm was the best model to predict plant diameter with R^2 0.39. and RMSE 0.49. This indicates that the effectiveness of an algorithm depends on the specific parameter being predicted and there is no dominant algorithm for all phenotypic parameters.

1. INTRODUCTION

Sugarcane is one of the main plantation commodities in the Indonesian agribusiness sector. According to the report of the Central Statistics Agency (BPS, 2023), the sugarcane plantation area reached 490.01 thousand hectares, with its management divided into State Large Plantations (PBN), Private Large Plantations (PBS), and People's Plantations (PR) . In the context of the Indonesian economy, sugarcane is in fourth place in the list of strategic agricultural commodities, after rice, animal products, and oils and fats, which have a market contribution of 6.7 percent (Magfiroh *et al*., 2017). Over the past five years, sugar production in Indonesia has shown a positive trend with an annual increase of more than 3.5%, reaching 2.42 million tons (Pusdatin, 2022). The quality and quantity of sugarcane production are greatly influenced by various phenotypic parameters of the plant, such as sugar content in sugarcane plants (brix or pol value), number of stands, stem diameter, and plant height (de Oliveira *et al*., 2022). Conventionally, the measurement of these parameters is done manually. This process is not only time-consuming and labor-intensive, but also has the potential to damage plants (Wahyuni *et al*., 2023). In addition, the development of non-destructive methods for measuring phenotypic parameters of sugarcane plants is very important (Akbarian *et al*., 2022).

In the last decade, drone technology equipped with multispectral cameras has shown its potential as a non-destructive tool in precision agriculture (Candiago *et al*., 2015). The development of technology and access to cheap drones and remote sensing techniques have created opportunities for researchers to conduct studies on biomass estimation at harvest time and analysis of nitrogen content in leaves at a small plot scale (size 2m x 2m) through the use of UAV LIDAR together with multispectral image processing techniques (Shendryk *et al*., 2020). Other studies use drone-based remote sensing for detection plant disease. Solahudin *et al*. (2015), for example, developed an analytical approach to detect Gemini virus infection in chili plants (*Capsicum annuum* L.) using aerial photos and Bayes segmentation-based image processing techniques. This approach depends on input from three color channels (Red, Green, Blue) and four targeted segmentation objects.

Drones can be exploited to gather high-resolution imagery data from large areas in a short time (Narmilan *et al*., 2022). This technology, when combined with machine learning algorithms, is able to analyze image data for predicting phenotypic parameters with high accuracy (Vijayakumar *et al*., 2023). The use of this method reduces costs and increases the efficiency of agricultural management, as well as supports sustainable agricultural practices (Guo *et al*., 2020). Until recently, the research on crop yield prediction using multispectral cameras and machine learning (ML) have been developed rapidly. For example, (Costa *et al*., 2021) created a model based on machine learning to predict crop yields and related characteristics in wheat plants, which showed a high accuracy through the use of multispectral imagery. Kayad *et al*., (2019) applied ML techniques to project variations in the field and produced more reliable predictions compared to the vegetation index method when used in corn areas.

This study aims to predict phenotypic parameters of sugarcane plants based on multispectral imagery taken by drones and to determine which combination of machine learning algorithms is the most effective and efficient for developing prediction models. Field data acquisition was carried out by conducting direct observations in the field and using drones equipped with multispectral cameras. The obtained data were then processed using Qgis software and was analyzed using several machine learning algorithms such as Random Forest (RF), Artificial Neural Network (ANN), and Support Vector Regression (SVR) to predict phenotypic parameters of sugarcane plants including sugar content (Brix value), number of stands, stem diameter, and plant height. This analysis was expected to reveal the potential and limitations of each algorithm in the context of precision agriculture applications and provide recommendations on the best methods for implementation in the field. Through the combination of multispectral imagery technology and artificial intelligence, it is expected that innovative solutions are discovered in developing productive and sustainable sugarcane plantations.

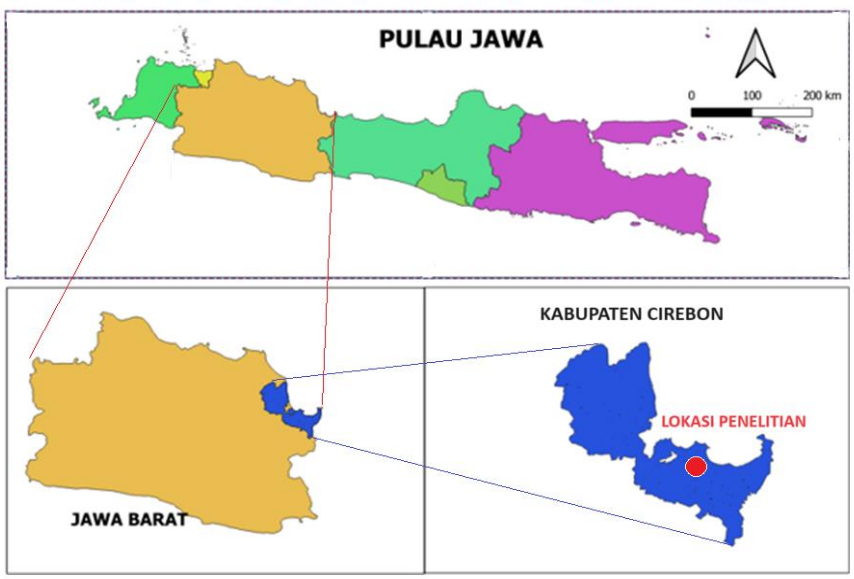


Figure 1. Location of research dataset collection

2. MATERIALS AND METHODS

The study was conducted from November 2023 to February 2024. Field sample data collection was located in the sugarcane plantation area of PT PG Rajawali II Sindang Laut, Sampih Palutungan Village, Susukanlebak District, Cirebon Regency, West Java (-6.852289° S, 108.603628° E). Details of the data collection area were described by polygons in Figure 1. Dataset processing and model development were carried out at the Bioinformatics Engineering Laboratory, Department of Mechanical and Biosystems Engineering, IPB University.

The equipment required in this study consisted of hardware for the acquisition of sugarcane plant dataset images and data sampling in the field, namely: Smartphone equipped with Qfield software, refractometer to calculate brix values, calipers to measure the diameter of sugarcane stems, 5 m measuring tape to measure the height of sugarcane and Parrot Bluegrass Drone equipped with a multispectral camera. Equipment for data processing included HP 240 G8 Notebook PC laptop that has been installed Python programming language. Furthermore, the software used for image acquisition using drones is FreeFlight Pro and Pix4Dcapture and Agisoft metashape while the software to support field dataset processing is QGis 3.28 Firenze and Google Colaboratory for data processing for the machine learning model used. The research workflow diagram is shown in the flow diagram in Figure 2.

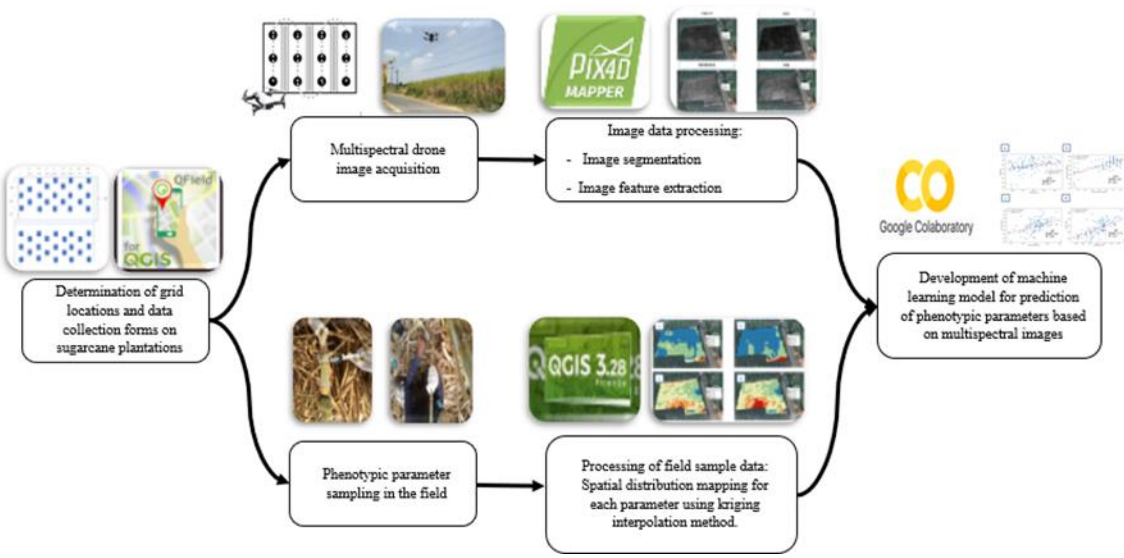


Figure 2. Research Workflow Diagram

2.1. Field Data Collection

2.1.1. Data Sample Collection

Field data sampling on sugarcane plants was performed using a 2×2 m grids to collect information on the condition of phenotypic parameters of sugarcane. The use of Qfield software facilitates field data collection with high efficiency, reducing registration time and data input processes into the database. It also capable to digitize in the field, edit geometry and attributes, and integration with GPS and Android device cameras. Qfield allows GIS project management from Qgis on Android devices, with simple configuration steps. Field data was recorded and digitized through three layers (points, lines, and polygons) with certain attributes. The attribute table connected to the three layers recorded the following information: plantation name (string), block (string), area (integer), variety (string), plant age (integer), stem diameter (decimal), plant height (integer), number of stands (integer), brix value (decimal). Projects was then synchronized using the Qfield Sync plugin or by copying the project file to a smartphone device.

2.1.2. Multispectral Drone Image Collection

Image acquisition was carried out by a Parrot Bluegrass multispectral drone. Considering the optimal height and no down draft effect to the plant object, the flight was kept above the plot at an altitude 50-70 m and average speed of 3-5 m/s. Images were taken automatically every 5 seconds (with a flight mission) without stopping the drone. The acquisition time between 11:30 and 13:30 is very good to avoid shadows formed by sunlight (Kerkech *et al*., 2020). The specifications of the drone and camera used in acquiring sugarcane plant images was detailed in Table 1.

Table 1. Specifications of Parrot Bluegrass Drone and Parrot Sequoia Camera

Specifications	Description
UAV	
Weight	1850 g
Size	50 × 44 × 12 cm
Flight altitude	50 - 70 m
Resolution	6.6 cm/pixel at 50 m
Flight range	Distance: max 2 km; Coverage: 30 hectares/battery
Battery	Lipo 6700 mAh
Average flight duration	25 minutes
Camera	
RGB Camera	Full HD 1080p video, 14 MP wide angle photo, 3-axis stabilization, 32 GB internal memory.
Multispektral Camera	Spectrum: Green (550 nm ± 20 nm), Red (660 nm ± 20 nm), RedEdge (735 nm ± 5 nm), near-infrared (790 nm ± 20 nm)

2.2. Data Processing

2.2.1. Field Sample Data Processing

After collecting field data using a smartphone with the Qfield application, the data was imported from the device and processed using a computer. Field sample data were then spatially mapped for each phenotypic parameter which was processed on a computer for statistical and graphical analysis. Data analysis conducted in this study utilized the Qgis programming module using kriging interpolation to analyze the phenotypic parameters.

The use of Geographic Information Systems (GIS) allows the collection of large and varied data, and allows for regular updates to initial information by adding the latest data (Nunes *et al*., 2009). Data analysis in this study used the Qgis programming module with the kriging interpolation method to analyze and provide estimates of an unsampled location point based on a weighted average of the surrounding data based on parameters of plant height, stem diameter, number of stands, and brix value in sugarcane plants.

The kriging interpolation model is specifically designed for spatial analysis. Spatial errors show spatial autocorrelation, while in linear models, errors are considered independent (Jurado-Expósito *et al*., 2021). This method is similar to multiple linear regression applied in a spatial context, with random variables as the regression variables and random variables at the desired points as the dependent variables (Schirrmann *et al*., 2017). Its level of accuracy is influenced by several factors, including: (1) the number and quality of samples at each point; (2) the distance between the samples and the points being evaluated; and (3) the spatial continuity that needs to be considered.

2.2.2. Multispectral Image Data Processing

Image analysis is a key component in developing a predictive model for sugarcane phenotypic parameters. The collected data are stored in *.TIF format. The use of Pix4D Mapper software was carried out to create an orthomosaic from aerial imagery and merge UAV images (Kou *et al*., 2022). The merged drone images produced aerial maps for each color band, namely red (R), green (G), blue (B), RedEdge (RE), and near-infrared (NIR). Plot segmentation on each image was carried out using QGIS 3.28 Firenze Software to increase the data resolution to 2 x 2 m cells. The Zonal Statistics tool was used to display the index values per grid based on the average pixels in the same area.

Image segmentation is the process of separating certain parts of an image that have certain characteristics or objects. A Region of Interest (ROI) is an area in an image that is relevant and interesting for a particular analysis. Image segmentation methods include Thresholding and color-based segmentation using a specific color space (such as RGB, HSV, LAB) or extraction of color features such as hue, saturation, or intensity. After image segmentation and ROI determination, the next step is to isolate the relevant regions for further analysis, such as object recognition, measurement, or image processing. Image feature extraction involves extracting important information from an image for pattern recognition, classification, or image analysis. Multispectral imagery is used in remote sensing and image analysis to extract features from various color bands such as Green, Red, RedEdge, and NIR.

2.3. Machine Learning Model Development

Machine learning (ML) is a branch of artificial intelligence that focuses on developing computer systems that can learn and adapt to data without the need for explicit programming. In machine learning, computers use algorithms and statistical models to recognize patterns in data and make decisions or predictions based on those patterns. Mathematically, machine learning processes input in the form of a tuple (X, Y) where X is an independent feature and Y is a dependent or target variable. Supervised learning in machine learning is divided into two general problems, namely regression and classification (Cedric *et al*., 2022). A machine learning (ML) model for predicting sugarcane phenotypic parameters was developed by combining GIS data parameters (brix value, number of stands, diameter, and height) generated from the kriging interpolation method with the results of image feature extraction, namely: Green, Red, RedEdge, NIR. In this study, three machine learning models are used, as described in the following.

2.3.1. Random Forest (RF)

Random Forest is a technique often used to predict crop yields and is able to handle high data dimensions, is resistant to overfitting, and can explore important input variables (Munir *et al*., 2023). This Random Forest regression algorithm consists of decision trees that learn simultaneously and independently, with prediction results that combine information from each tree (Hanif & Gunawan, 2022).

2.3.2. Support Vector Regression (SVR)

Support Vector Regression (SVR) is a technique in regression analysis in machine learning (Khanal *et al*., 2018). The goal is to develop a hyperplane or a group of hyperplanes in an N-dimensional space, where N is the number of variables in the dataset. SVR tries to identify the hyperplane that best fits the highest number of points in the regression (Akbarian *et al*., 2023). This algorithm is popular in crop yield prediction because it is able to identify non-linearities in the data, resulting in a more reliable and accurate model. SVR has advantages in strong generalization, high prediction accuracy, and easy implementation (Haryadi *et al*., 2022).

2.3.3. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a powerful non-linear regression approach in machine learning to predict crop yields (Kumar *et al*., 2015). Overfitting can occur in non-linear models such as neural networks. In this study, a non-linear Auto-Regressive network with exogenous inputs was used using Bayesian backpropagation regularization as a training function to avoid overfitting. ANN has been shown to be the most reliable and efficient in predicting multivariable data and non-linear data series (Ellafi *et al*., 2021). The entire data processing process for the ANN model was carried out in Google Colaboratory software.

The performance of the three ML models was evaluated using two indicators of prediction accuracy, namely the coefficient of determination (R^2) and the root mean squared error (RMSE). R^2 can be interpreted as the fraction of variance explained by a given predictor as seen in Equation (1). RMSE in Equation (2) was used to estimate the measurement error. The greater the error associated with the associated error, the greater the RMSE value.

$$R^2 = \frac{\sum_{i=1}^{n}(yi-\hat{y}i)^2}{\sum_{i=1}^{n}(yi-\bar{y})^2} \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi-\hat{y}i)^2} \tag{2}$$

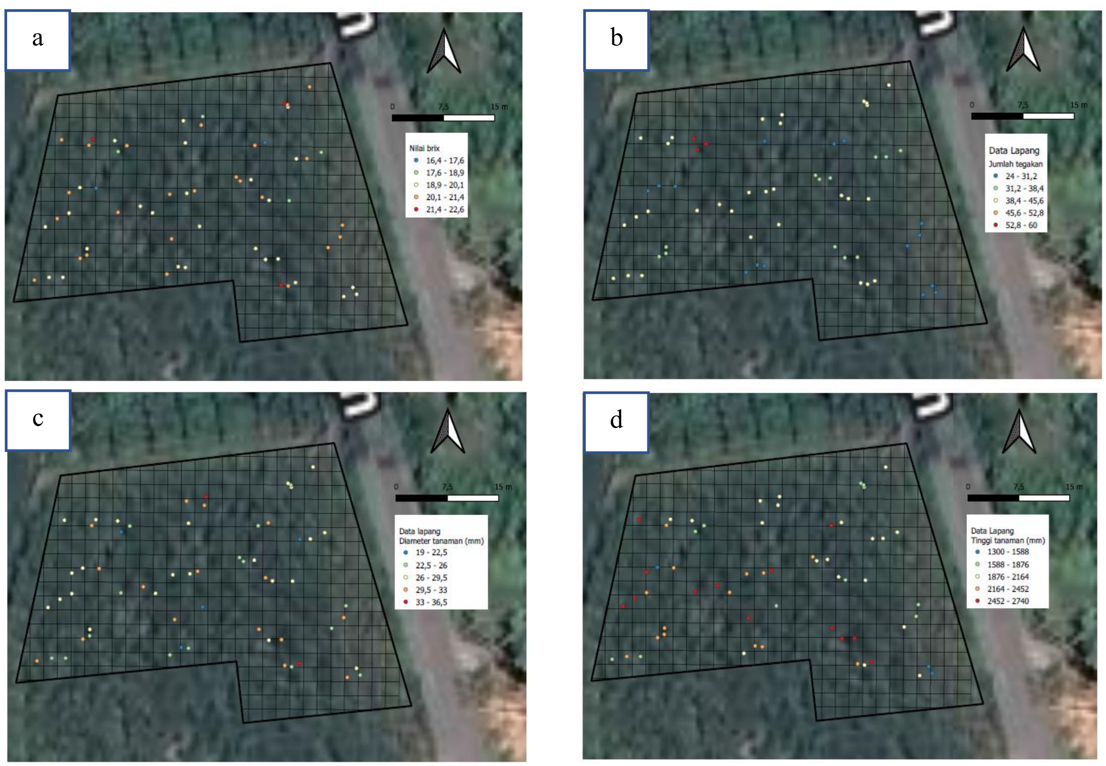
where $\hat{y}i$ is the predicted yield (t/ha), yi is the actual yield (t/ha), $(\hat{y})$ is the average value of the actual yield, and n is the number of iterations.

The models were assessed in terms of the accuracy of predicting phenotypic parameters by comparing the yield values obtained from the ML model with the potential values observed from field measurements. Each ML technique was performed through several iterations. Iterations by dividing the data based on each parameter into training data, training and testing, so that the R^2 and RMSE values indicate the average of all iterations and the accuracy of the model.

3. RESULTS AND DISCUSSION

3.1. Field Data Collection Results

Data collection was carried out on sugarcane plantations that had rows of plants with a length of 8 meters. The location of sample data collection was determined based on the rows of plants with a grid size of 2 mx 2 m. In each grid, parameters were calculated including stem diameter, plant height, brix value and number of stands on sugarcane plants with multiples of 10 in the row of plants. Based on the analysis of field data from 62 samples taken from sugarcane plantations using a 2 mx 2 m grid, the average brix value of the measured samples was 20.06°Bx with minimum and maximum variations of 16.4 Bx and 21.4°Bx, respectively. This brix value indicates a fairly high concentration of sucrose in sugarcane, which indicates good sugar yield potential. The number of stands on sugarcane plantations has an average of 37 stems per sampling area (2 mx 2 m), with the lowest and highest variations between 24 stems and 60 stems. This indicates a varied plant density throughout the plantation area. The average stem diameter of sugarcane plants was 27.59 mm. The observed stem diameter range was between 19 mm and 36.5 mm, indicating consistency in the physical growth of the plant. The average plant height was 2110 mm, with a height variation of 1300 mm to 2740 mm. This variation reflects the differences in growth phases among plants in the same population. The field sampling map can be seen in Figure 3 and the description of the field data of 62 samples with a 2 mx 2 m grid as in Table 2.



a. Brix Value (°Bx)	b. Number of stands	c. Stem diameter (mm)	d. Plant height (mm)
16,4 - 17,6	24 - 31,2	19 - 22,5	1300 - 1588
17,6 - 18,9	31,2 - 38,4	22,5 - 26	1588 - 1876
18,9 - 20,1	38,4 - 45,6	26 - 29,5	1876 - 2164
20,1 - 21,4	45,6 - 52,8	29,5 - 33	2164 - 2452
21,4 - 22,6	52,8 - 60	33 - 36,5	2452 - 2740

Figure 3. Sample collection map of 62 samples in the sugarcane plantation: (a) Brix value, (b) Number of stalks, (c) Stem diameter, (d) Plant height.

Table 2. Field data description of 62 samples with a 2 m x 2 m grid.

Description	Average	Std	Min	Max
Brix Value (°Bx)	20.06	1.03	16,4	21.4
Number of Stalks	37	8	24	60
Diameter (mm)	27.59	3.45	19	36.5
Height (mm)	2110	34.81	1300	2740

3.2. Data Processing Results

3.2.1. Sample Data Processing Results

In this study, the sample data was interpolated using kriging to estimate the location of an unsampled point based on the weighted average of the surrounding data. The flowchart of the field sample data processing process can be seen in Figure 4. For each phenotypic parameter, a variogram analysis was performed to measure the variability between points (experimental variogram) and lines (theoretical variogram). The experimental variogram fitting to the variogram model was performed using the parameter with the smallest RMSE value selected as the best.

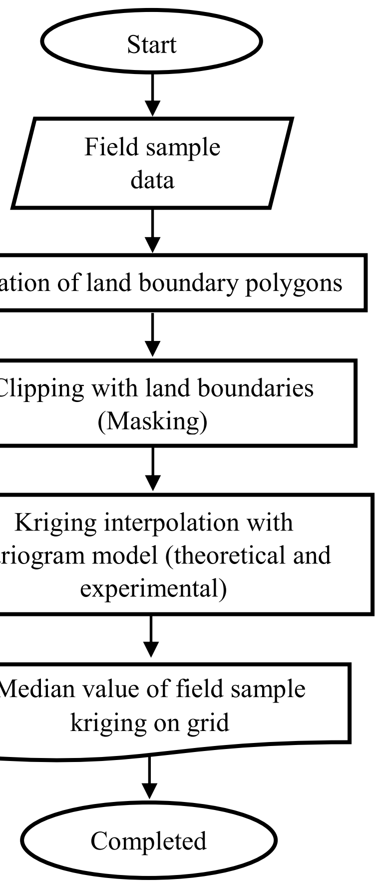


Figure 4. Flowchart of the sample data processing

Figure 5 shows the best variogram model based on the results of fitting phenotypic parameters. Figure 5(a) shows the experimental variogram on the brix value parameter using the adjust linear to sill model with the smallest RMSE with a value of 0.073. Figure 5(b) shows the experimental variogram on the number of stands parameter using the adjust exponential model with the smallest RMSE with a value of 297.944. Figure 5(c) shows the experimental vario gram on the plant diameter parameter using the adjust Spherical model with the smallest RMSE with a value of 158.62.

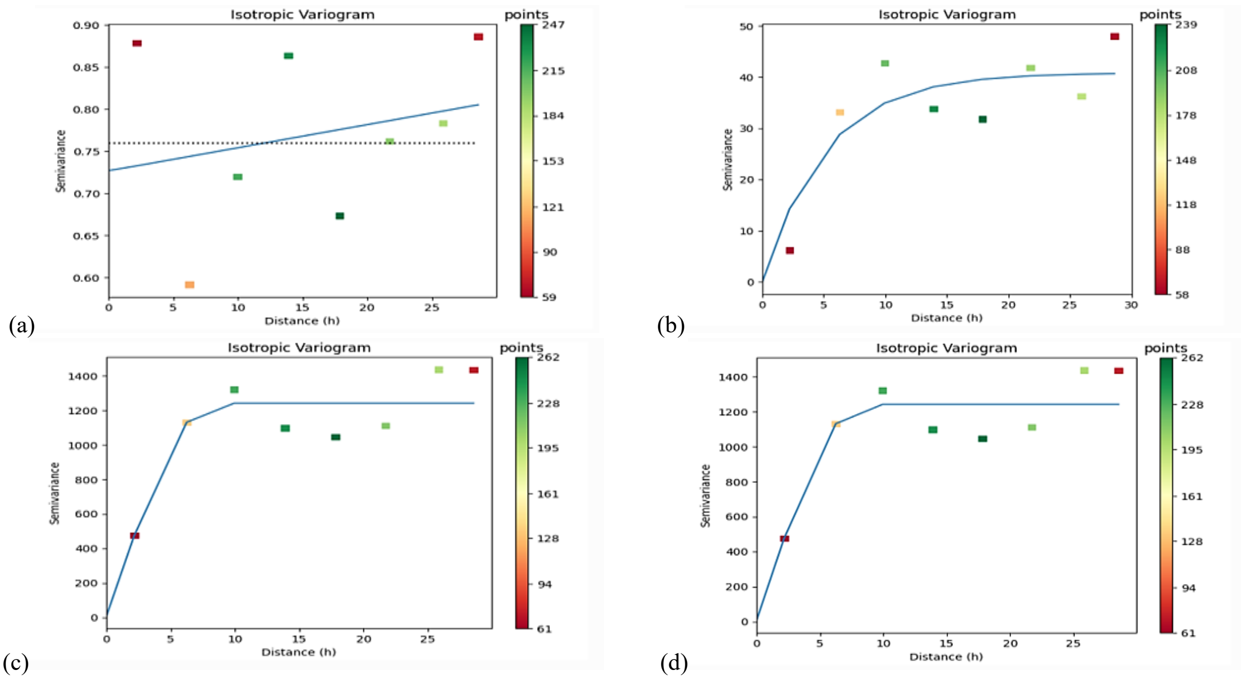
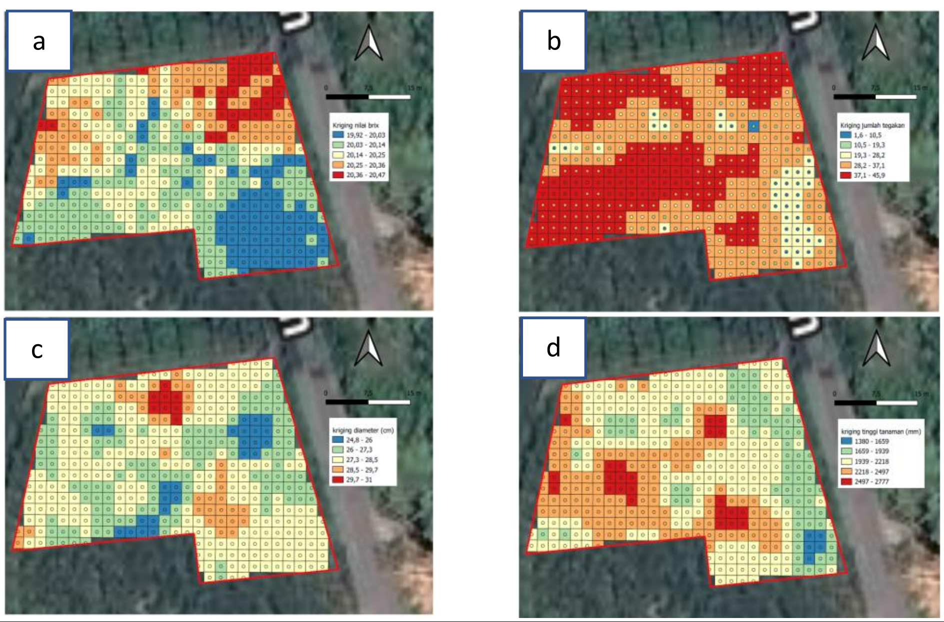


Figure 5. Best kriging interpolation variogram model based on phenotypic parameters: a) Brix value b) Number of stands c) Stem diameter d) Plant height.



a. Kriging Brix Value (°Bx)				b. Kriging number of stands				c. Kriging diameter (mm)				d. Kriging height (cm)			
Color	Range	# Grid	%	Color	Range	# Grid	%	Color	Range	# Grid	%	Color	Range	# Grid	%
Blue	19.92-20.03	79	19.65	Blue	1.6-10.5	1	0.25	Blue	24.8-26.0	25	6.22	Blue	130.0-165.9	5	1.24
Green	20.03-20.14	119	29.60	Green	10.5-19.3	0	0.00	Green	26,0-27.3	118	29.35	Green	165.9-193.9	74	18.41
Yellow	20.14-20.25	103	25.62	Yellow	19.3-28.2	38	9.45	Yellow	27.3-28.5	217	53.98	Yellow	193.9-221.8	181	45.02
Orange	20.25-20.36	70	17.41	Orange	28.2-31.1	180	44.78	Orange	28.5-29.7	34	8.46	Orange	221.8-249.7	119	29.60
Red	20.36-20.47	31	7.71	Red	37.1-45.9	183	45.52	Red	29.7-31.0	8	1.99	Red	249.7-277.7	23	5.72

Figure 6. Spatial distribution map of kriging interpolation results for phenotypic parameters: (a) Brix value, (b) Number of stands, (c) Plant diameter, and (d) Plant height.

While Figure 5(d) shows the experimental variogram on the plant height parameter using the adjust Spherical model with the smallest RMSE with a value of 21.35.

The estimation results of the kriging interpolation method using the best variogram model according to the smallest range value of the selected variogram were then carried out by the point kriging method with an exponential variogram. Figure 5 shows the spatial distribution of phenotypic parameters estimated by the kriging interpolation method using the best variogram model.

Figure 6 (a) is a spatial distribution map of brix values, showing the dominant distribution of brix values with a range of values 20.03 - 20.14 (green color) and occupying approximately 29.60% of the total area of the study area. Figure 6 (b) is a spatial distribution map of the number of stands, showing the dominant number of stands with a value of 37.1-45.9 (red color) and occupying more than 45.52% of the total area of the study area. Figure 6 (c) is a spatial distribution map of plant diameter, showing the dominant plant diameter with a value of 27.3-28.5 (yellow color) and occupying more than 53.98% of the total area of the study area. Figure 6 (d) is a spatial distribution map of plant height, showing the dominant plant height with a value of 193.9-221.8 (yellow color) and occupying more than 45.02% of the total area of the study area.

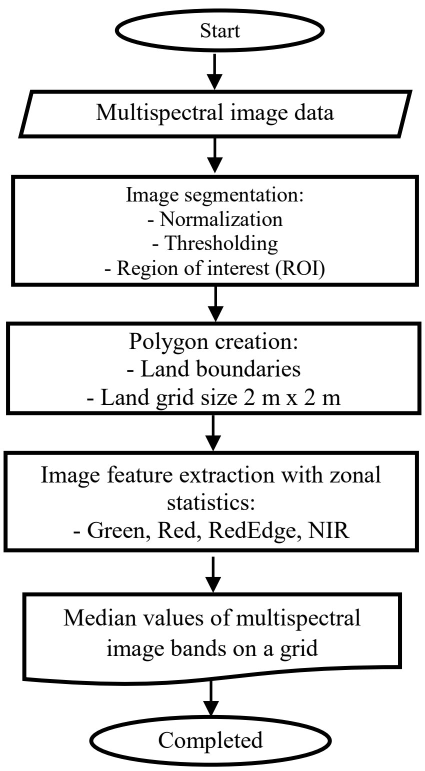
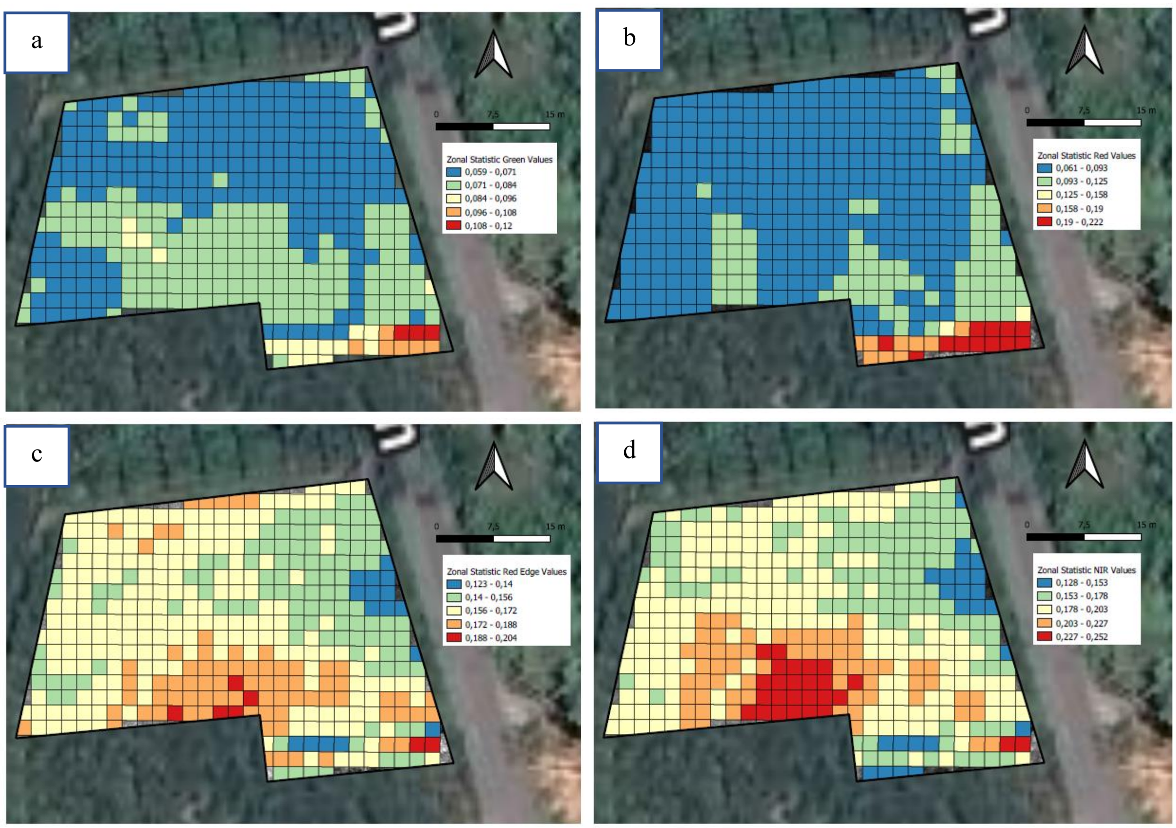


Figure 7. Flowchart of the multispectral drone image data processing

3.2.2. Multispectral Drone Image Data

In this study, multispectral drone imagery data was processed using zonal statistics tools from Qgis 3.28.0 Firenze software to analyze phenotypic parameters in the form of brix values, number of stands, plant diameter and plant height including the average, smallest value, largest value and standard deviation with a grid size of 2×2 m. The flow diagram of the multispectral drone imagery data processing process can be seen in Figure 7. Multispectral imagery from drones processed using Pix4D is in the form of a single band image with gray color. Observation of each spectrum (Green, Red, RedEdge and NIR) will be easier if the gray color is transformed into pseudocolor or full color as shown in Figure 8, making it easier to analyze the image visually based on color classification.

Figure 8(a) is a spatial distribution map of Green values, showing the dominant distribution of Green values with a range of values 0.059-0.071 (blue color) and occupying approximately 53.23% of the total area of the study area. Figure 8(b) is a spatial distribution map of Red values, showing the dominant Red values with values 0.061-0.093 (blue color) and occupying more than 74.13% of the total area of the study area. Figure 8(c) is a spatial distribution map of Rededge values, showing the dominant Rededge with values 0.156-0.172 (yellow color) and occupying more than 48.76% of the total area of the study area. Figure 8(d) is a spatial distribution map of NIR values, showing the dominant NIR with values 0.178-0.203 (yellow color) and occupying more than 45.52% of the total area of the study area. The multispectral statistical zonal description of 402 samples with a 2 m x 2 m grid can be seen in Table 3.



a. Green

Color	Range	# Grid	%
Blue	0.059-0.071	214	53.23
Green	0.071-.084	161	40.05
Yellow	0.084-0.096	17	4.23
Orange	0.096-0.108	6	1.49
Red	0.108-0.120	4	1.00

b. Red

Color	Range	# Grid	%
Blue	0.061-0.093	298	74.13
Green	0.093-0.125	80	19.90
Yellow	0.125-0.158	2	0.50
Orange	0.158-0.190	9	2.24
Red	0.190-0.222	13	3.23

c. RedEdge

Color	Range	# Grid	%
Blue	0.123-0.140	15	3.73
Green	0.140-0.156	108	26.87
Yellow	0.156-0.172	196	48.76
Orange	0.172-0.188	75	18.66
Red	0.188-0.204	8	1.99

d. NIR

Color	Range	# Grid	%
Blue	0.128-0.153	23	5.72
Green	0.153-0.178	107	26.62
Yellow	0.178-0.203	183	45.52
Orange	0.203-0.227	62	15.42
Red	0.277-0.252	27	6.72

Figure 8. Zonal spatial distribution map of multispectral image statistics.

Table 3. Multispectral statistical zonal description of 402 samples with a 2 m x 2 m grid

Description	Average	Std	Min	Max
Green	0.071	0.005	0.059	0.012
Red	0.084	0.009	0.061	0.222
Rededge	0.162	0.011	0.123	0.204
Nir	0.188	0.022	0.128	0.252

3.3. Results of Machine Learning Model Development

Prediction of phenotypic parameters of sugarcane plants was developed by combining the results of multispectral image feature extraction, namely: Green, Red, RedEdge, NIR with data parameters (brix value, number of stands, stem diameter, plant height) generated from the kriging interpolation method into three machine learning models, namely Random Forest (RF), Artificial Neural Network (ANN), and Support Vector Regression (SVR). The three machine learning model parameters were adjusted by testing different combinations of model hyperparameters.

Hyperparameter tuning on Random Forest (RF) for testing data is done to find the optimal configuration explaining the dataset and produces the best-fit model. In this tuning process, we examine six key parameters. These parameters include the number of decision trees (n estimator), the maximum depth of the tree (max depth), the minimum number of samples required to split an internal node (min sample split), the minimum number of samples per leaf, the number of features (max feature) and max leaf nodes. Hyperparameter values are taken randomly with a certain number of iterations. The best hyperparameters are determined based on the best coefficient of determination (R^2).

Hyperparameters for the Artificial Neural Network (ANN) model on the test data are implemented to identify the best configuration that can accurately map the dataset and create the most effective model. During this optimization process, five main parameters are analyzed. These parameters include the number of units in the hidden layer, activation function, weights in the network (solver), learning rate, number of iterations. Hyperparameter values are determined through random selection in a predetermined number of iterations. The most effective hyperparameters are identified based on the highest coefficient of determination (R^2) value.

For the Support Vector Regression (SVR) model, six unique kernels were developed and their specific parameters, including Scale, Epsilon, and Gamma, were adjusted for each combination. The SVR model that showed the highest effectiveness was determined by the best coefficient of determination R^2 value. Detailed results of hyperparameter adjustments for each algorithm model are described in Table 4.

Table 4. Hyperparameters for test data

Algorithm	Parameter	Value
Random Forest	N estimator	20
	Max depth	None
	Min samples split	5
	Min_weight_fraction leaf	0.0
	Max features	Auto
	Mx_leaf_nodes	None
Artficial Neural Network	Hidden layer	8
	Activation	Relu
	Solver	Lbfgs
	Learning rate	Adaptive
	Max_iter	20000
Support Vector Regression	C	1.0
	Epsilon	0.1
	Kernel	Rbf
	Gamma	Auto

Machine learning methods with predetermined hyperparameters, implemented through cross-validation techniques, divide the dataset into ten separate parts. In each round, the method examines the results of multispectral image feature extraction, namely: Green, Red, RedEdge, NIR as input parameters to estimate phenotypic parameters, namely: brix value, number of stands, stem diameter, plant height. The effectiveness of the predictive model is then measured using the coefficient of determination (R^2) and root mean squared error (RMSE), the details of which are presented in Table 5. It shows that the Random Forest (RF) algorithm excels in predicting plant height with RMSE 14.09 and R^2 0.53. On the other hand, Artificial Neural Network (ANN) proved to be the most efficient in predicting brix value with RMSE 0.06 and R^2 0.74 and number of stands with RMSE 2.13 and R^2 0.68. The results obtained from using the ANN algorithm are more suitable for rapid assessment because of its efficiency in predicting important parameters such as brix value and number of stands with high accuracy. Meanwhile, Support Vector Regression (SVR) showed the best performance in predicting plant diameter with RMSE 0.49 and R^2 0.39. This indicates that the effectiveness of an algorithm is highly dependent on the specific parameters being predicted, with no algorithm being absolutely dominant in all aspects. The suitability of predictions and actual algorithm development for predicting phenotypic parameters in the field can be seen in Figure 9.

Figure 9(a) shows the results of the brix value prediction with the best ANN algorithm with R^2 0.74 and RMSE 0.06. Figure 9(b) shows the results of the number of stands prediction with the ANN algorithm as the best with R^2 0.68 and RMSE 2.13. Figure 9(c) shows the results of the plant diameter prediction with the SVR algorithm as the best with R^2 0.39 and RMSE 0.49. Figure 9(d) shows the results of the plant height prediction with the RF algorithm as the best with R2 0.53 and RMSE 14.09.

Table 5. Results of algorithm development to predict field parameters

Phenotypic parameter	Algorithm					
	RF		ANN		SVR	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
Brix value	0.67	0.64	0.06	0.74	0.06	0.61
Number of stalks	2.15	0.60	2.13	0.68	2.51	0.48
Stem diameter	0.51	0.33	0.56	0.17	0.49	0.39
Plant height	14.09	0.53	15.74	0.42	16.86	0.34

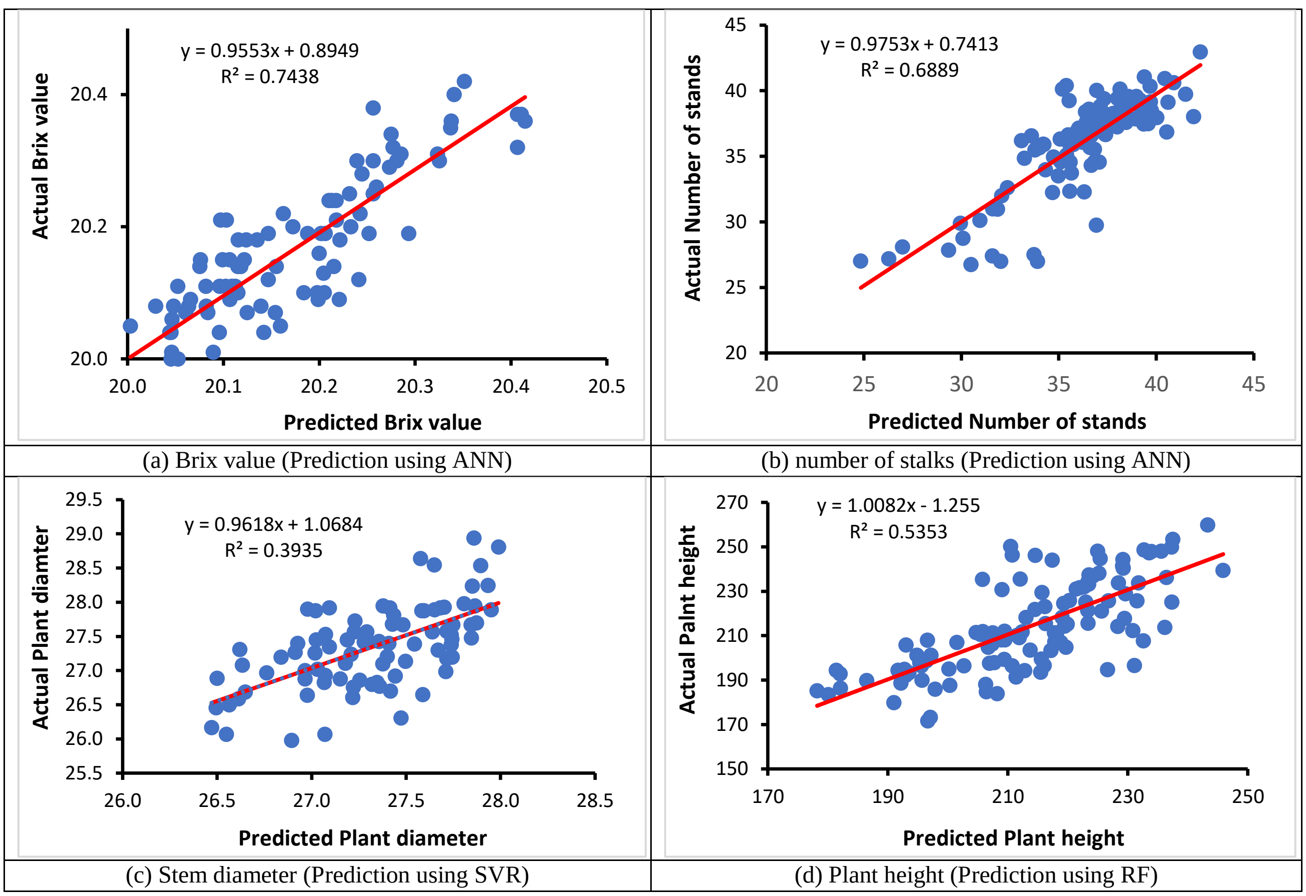


Figure 9. The fit between predictions and actual algorithm development for predicting phenotypic parameters in the field: (a) Brix value, (b) Number of stands, (c) Stem diameter, and (d) Plant height.

4. CONCLUSION

The development of non-destructive methods to measure phenotypic parameters of sugarcane plants is urgently needed because the current manual process is inefficient and potentially damaging to plants. The integration of drones with multispectral imagery and machine learning algorithms promises the ability to produce accurate phenotypic parameter predictions, reduce costs, and support sustainable agricultural practices. In this study, the implementation of field data sampling on sugarcane plants used a grid size of 2 mx 2 m with 62 samples to provide detailed information on the condition of phenotypic parameters in sugarcane plants. Then the data was analyzed using the Qgis programming module using kriging interpolation to analyze each phenotypic parameter. Multispectral imagery was taken using the Parrot Bluegrass drone , then the drone images that had been combined produced aerial maps for each color band, namely red (R), green (G), RedEdge (RE), and near-infrared (NIR) and were segmented using the zonal statistics tool to display the index value per grid based on the average pixel in the same area. Data from the multispectral spectrum captured by the drone was used as input to estimate phenotypic parameters such as brix value, number of stands, stem diameter, and plant height. Based on the results of the machine learning model development , the ANN algorithm model is the most efficient in predicting brix values with RMSE 0.06 and R^2 0.74 and the number of stands with RMSE 2.13 and R^2 0.68. The RF algorithm model stands out in predicting plant height with RMSE 14.09 and R^2 0.53. The SVR algorithm model shows the best performance in predicting plant diameter with RMSE 0.49 and R^2 0.39. This indicates that the effectiveness of an algorithm is highly dependent on the specific parameters being predicted, with no algorithm being absolutely dominant in all aspects.

REFERENCES

Akbarian, S., Jamnani, M.R., Xu, C., Wang, W., & Lim, S. (2023). Plot level sugarcane yield estimation by machine learning on multispectral images: A case study of Bundaberg, Australia. *Information Processing in Agriculture*, *xxxx*. https://doi.org/10.1016/j.inpa.2023.06.004

Akbarian, S., Xu, C., Wang, W., Ginns, S., & Lim, S. (2022). Sugarcane yields prediction at the row level using a novel cross-validation approach to multi-year multispectral images. *Computers and Electronics in Agriculture*, ***198*****,** 107024. https://doi.org/10.1016/j.compag.2022.107024

Badan Pusat Statistik. (2023). *Statisitik Tebu Indonesia 2022* (Direktorat Statistik Tanaman Pangan (ed.); Vol. 13). BPS RI/BPS-Statistics Indonesia.

Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., & Gattelli, M. (2015). Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sensing*, ***7***(4), 4026–4047. https://doi.org/10.3390/rs70404026

Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M., & Kimpolo, C. L. M. (2022). Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agricultural Technology*, ***2***. https://doi.org/10.1016/j.atech.2022.100049

Costa, L., Ampatzidis, Y., Rohla, C., Maness, N., Cheary, B., & Zhang, L. (2021). Measuring pecan nut growth utilizing machine vision and deep learning for the better understanding of the fruit growth curve. *Computers and Electronics in Agriculture*, ***181***, 105964. https://doi.org/10.1016/j.compag.2020.105964

de Oliveira, R.P., Barbosa Júnior, M.R., Pinto, A.A., Oliveira, J.L.P., Zerbato, C., & Furlani, C.E.A. (2022). Predicting Sugarcane Biometric Parameters by UAV Multispectral Images and Machine Learning. *Agronomy*, ***12***(9). https://doi.org/10.3390/agronomy12091992

Ellafi, M.A., Deeks, L.K., & Simmons, R.W. (2021). Application of artificial neural networks to the design of subsurface drainage systems in Libyan agricultural projects. *Journal of Hydrology: Regional Studies*, ***35***, 100832. https://doi.org/10.1016/j.ejrh.2021.100832

Guo, Y., Wang, H., Wu, Z., Wang, S., Sun, H., Senthilnath, J., Wang, J., Bryant, C.R., & Fu, Y. (2020). Modified red blue vegetation index for chlorophyll estimation and yield prediction of maize from visible images captured by uav. *Sensors (Switzerland)*, ***20***(18), 1–16. https://doi.org/10.3390/s20185055

Hanif, W.H., & Gunawan, F.E. (2022). Random forest regression to predict catalyst deactivation in industrial catalytic process. *TEM Journal*, ***11***(1), 104–110. https://doi.org/10.18421/TEM111-12

Haryadi, D., Hakim, A.R., Atmaja, D.M.U., & Yutia, S.N. (2022). Implementation of support vector regression for polkadot cryptocurrency price prediction. *International Journal on Informatics Visualization*, ***6***(1–2), 201–207. https://doi.org/10.30630/joiv.6.1-2.945

Jurado-Expósito, M., Torres-Sánchez, J., López-Granados, F., & Jiménez-Brenes, F.M. (2021). Monitoring the spatial variability of knapweed (*Centaurea diluta* aiton) in wheat crops using geostatistics and uav imagery: Probability maps for risk assessment in site-specific control. *Agronomy*, ***11***(5). https://doi.org/10.3390/agronomy11050880

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). https://doi.org/10.3390/rs11232873

Kerkech, M., Hafiane, A., & Canals, R. (2020). Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Computers and Electronics in Agriculture*, ***174***. https://doi.org/10.1016/j.compag.2020.105446

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., & Shearer, S. (2018). Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture*, ***153***, 213–225. https://doi.org/10.1016/j.compag.2018.07.016

Kou, J., Duan, L., Yin, C., Ma, L., Chen, X., Gao, P., & Lv, X. (2022). Predicting leaf nitrogen content in cotton with UAV RGB images. *Sustainability (Switzerland)*, ***14***(15). https://doi.org/10.3390/su14159259

Kumar, R., Singh, M.P., Kumar, P., & Singh, J.P. (2015). Crop selection method to maximize crop yield rate using machine learning technique. *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, ICSTM 2015 - Proceedings*, 138–145. https://doi.org/10.1109/ICSTM.2015.7225403

Magfiroh, I.S., Setyawati, I.K., & Wibowo, R. (2017). Mutu tebu industri gula di Indonesia. *Prosiding Seminar Nasional Pembangunan Pertanian II*, 94–100.

Munir, S., Seminar, K.B., Sudradjat, Sukoco, H., & Buono, A. (2023). The use of random forest regression for estimating leaf nitrogen content of oil palm based on Sentinel 1-A Imagery. *Information (Switzerland)*, ***14***(1). https://doi.org/10.3390/info14010010

Narmilan, A., Gonzalez, F., Salgadoe, A.S.A., Kumarasiri, U.W.L.M., Weerasinghe, H.A.S., & Kulasekara, B.R. (2022). Predicting canopy chlorophyll content in sugarcane crops using machine learning algorithms and spectral vegetation indices derived from UAV multispectral imagery. *Remote Sensing*, ***14***(5).

Nunes, M., Ferreira, Ó., Schaefer, M., Clifton, J., Baily, B., Moura, D., & Loureiro, C. (2009). Hazard assessment in rock cliffs at Central Algarve (Portugal): A tool for coastal management. *Ocean and Coastal Management*, ***52***(10), 506–515. https://doi.org/10.1016/j.ocecoaman.2009.08.004

Pusdatin. (2022). Outlook Komoditas perkebunan tebu. In A.A. Susanti & R.K. Putra (Eds.), *Pusat Data dan Sistem informasi Petanian* (Tahun 2022, Issue 1). Pusat Data dan Sistem inofrmasi Pertanian.Sekretariat Jendral-Kementrian Pertanian.

Schirrmann, M., Hamdorf, A., Giebel, A., Gleiniger, F., Pflanz, M., & Dammer, K.H. (2017). Regression kriging for improving crop height models fusing ultra-sonic sensing with UAV imagery. *Remote Sensing*, ***9***(7), 1–18. https://doi.org/10.3390/rs9070665

Shendryk, Y., Sofonia, J., Garrard, R., Rist, Y., Skocaj, D., & Thorburn, P. (2020). Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using UAV LiDAR and multispectral imaging. *International Journal of Applied Earth Observation and Geoinformation*, ***92***, 102177. https://doi.org/10.1016/j.jag.2020.102177

Solahudin, M., Pramudya, B., Liyantono, Supriyanto, & Manaf, R. (2015). Gemini virus attack analysis in field of chili (*Capsicum Annuum* L.) using aerial photography and bayesian segmentation method. *Procedia Environmental Sciences*, ***24***, 254–257. https://doi.org/10.1016/j.proenv.2015.03.033

Vijayakumar, V., Ampatzidis, Y., & Costa, L. (2023). Tree-level citrus yield prediction utilizing ground and aerial machine vision and machine learning. *Smart Agricultural Technology*, ***3***, 100077. https://doi.org/10.1016/j.atech.2022.100077

Wahyuni, A.P., Astuti, I.S., & Purwanto, P. (2023). Estimasi fase pertumbuhan dan produktivitas tebu menggunakan citra Sentinel 2 di Kecamatan Dampit, Kabupaten Malang. *Jurnal Integrasi dan Harmoni Inovatif Ilmu-Ilmu Sosial (JIHI3S)*, ***3***(2), 104–122. https://doi.org/10.17977/um063v3i2p104-122