Smart Agricultural Technology 11 (2025) 100879

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

A novel application with explainable machine learning (SHAP and LIME) to predict soil N, P, and K nutrient content in cabbage cultivation

Thilina Abekoon [a], Hirushan Sajindra [a], Namal Rathnayake [b], Imesh U. Ekanayake [c,d], Anuradha Jayakody [e], Upaka Rathnayake [f,*]

[a] Water Resources Management and Soft Computing Research Laboratory, Millennium City, Athurugiriya, 10150, Sri Lanka
[b] Department of Civil Engineering, Faculty of Engineering, The University of Tokyo, 1 Chome-1-1 Yayoi, Bunkyo City, Tokyo 113-8656, Japan
[c] Royal Melbourne Institute of Technology, Melbourne, Australia
[d] Ceylon Institute for Artificial Intelligence and Research (CIAIR) Limited, Colombo, Sri Lanka
[e] Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
[f] Department of Civil Engineering and Construction, Faculty of Engineering and Design, Atlantic Technological University, Sligo, F91 YW50, Ireland

ARTICLE INFO

Keywords:
Cabbage
Deep neural network
Explainable machine learning
Major soil nutrients
Plant growth characteristics

ABSTRACT

Cabbage (*Brassica oleracea* var. capitata) is commonly cultivated in high altitudes and features dense, tightly packed leaves. The Green Coronet variety is well-known for its robust growth and culinary versatility. Maximizing yield is crucial for food sustainability. It is essential to predict the soil's major nutrients (nitrogen, phosphorus, and potassium) to maximize the yield. Artificial intelligence is widely used for non-linear predictions with explainability. This research assessed the predictive capabilities of soil nitrogen, phosphorus, and potassium levels with explainable machine learning methods over an 85-day cabbage growth period. Experiments were conducted on cabbage plants grown in central hills of Sri Lanka. SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) were used to clarify the model's predictions. SHAP analysis showed that high feature values of the number of days and plant average leaf area negatively impacted for nutrient predictions, while high feature values of leaf count and plant height had a positive effect on the nutrient predictions. To validate the results, 15 greenhouse-grown cabbage plants at various growth stages were selected. The nitrogen, phosphorus, and potassium levels were measured and compared with the predicted values. These insights help refine predictive models and optimize agricultural practices. A user-friendly application was developed to improve the accessibility and interpretation of predictions. This tool is a user-friendly platform for end-users, enabling effective use of the model's predictive capabilities.

1. Introduction

Cabbage (*Brassica oleracea* var. capitata) is a biennial vegetable known for its dense heads. Belonging to the Brassicaceae family [1,2], it is a close relative of broccoli, cauliflower, and Brussels sprouts. Cultivated for centuries, cabbage is valued globally for its culinary versatility and nutritional value [3]. Cabbage is nutritionally rich, with low energy density, minimal fat, and essential micronutrients [4,5]. Cabbage is rich in micronutrients and contains a high amount of vitamin C (maximum 53.32 mg/100 g) for immune support and antioxidant power [6]. It contributes significantly to vitamin K for blood clotting and bone health, packs dietary fiber for gut health, and offers diverse B vitamins (folate, pantothenic acid, pyridoxine) and essential minerals (potassium, manganese) for a balanced diet [1,7]. There are several types of cabbage in the world. Napa cabbage (Pe-tsai group) features elongated, tender leaves ideal for wraps and kimchi production [8]. Savoy cabbage (Capitata group) boasts crinkled, loose leaves, often employed in soups and gratins due to their high water content [9]. Bok choy (Chinensis group) showcases thick stalks and dark green leaves, frequently utilized in stir-fries and broths [10]. Lastly, Brussels sprouts (Gemmifera group) form miniature cabbage-like heads on stalks, offering a unique flavor profile ideal for roasting and steaming [11].

Fertilization is crucial for successful cabbage cultivation [12]. Similar to other plants, cabbage requires adequate levels of primary macronutrients, nitrogen (N), phosphorus (P), and potassium (K) for optimal growth and development [1,13,14]. Traditionally, conventional

* Corresponding author.
E-mail address: upaka.rathnayake@atu.ie (U. Rathnayake).

https://doi.org/10.1016/j.atech.2025.100879
Received 24 January 2025; Received in revised form 3 March 2025; Accepted 4 March 2025
Available online 6 March 2025

cabbage cultivation relies on chemical fertilizers like urea, muriate, and triple superphosphate of potash to fulfill these nutritional needs [15]. Nevertheless, organic cultivation practices prioritize alternative sources of nutrients such as rice bran, poultry manure, cow urine, cow dung, and ash [16]. For both conventional and organic approaches, accurately assessing the existing nutrient content of the soil is crucial. Prediction models provide valuable insights for farmers, allowing for tailored fertilization strategies.

Artificial Intelligence (AI) driven precision agriculture is transforming farming practices [17]. AI algorithms, trained on expansive datasets encompassing weather [18], soil properties [1,18,19], historical yields [18,20], and remote sensing imagery [21,22,23], are now capable of performing sophisticated agricultural predictions. Previous studies have extensively examined heavy metal uptake in Chinese cabbage (*Brassica chinensis* L.) cultivated in soils contaminated with cadmium (Cd), copper (Cu), nickel (Ni), lead (Pb), and zinc (Zn). Research indicates that the bioavailability and subsequent accumulation of these metals in cabbage are influenced by soil properties, particularly pH levels. For instance, a study demonstrated that the uptake of Cd, Pb, and Zn by Chinese cabbage is significantly affected by soil pH, with lower pH levels enhancing metal solubility and uptake [24]. Additionally, the chemical forms of these metals in the soil play a crucial role in their absorption by plants. Advanced modeling techniques have been employed to improve predictions of metal uptake in Chinese cabbage. A notable approach involves integrating Geographic Information System data with deep learning algorithms to predict crop yields based on soil properties. For example, a modified Convolutional Neural Network-Recurrent Neural Network model was developed to predict cabbage yield by extracting soil characteristics from SoilGrids and soil suitability maps. This model effectively captures the complex relationships between soil attributes and crop performance [25]. Furthermore, deep neural networks have been utilized to predict soil nitrogen, phosphorus, and potassium contents using the plant properties of cabbage. By analyzing various plant growth parameters, these models can accurately estimate soil nutrient levels, thereby aiding in precision agriculture practices. Such predictive capabilities are essential for optimizing fertilization strategies and ensuring sustainable crop production [1]. However, transparency has been lost, and farmers cannot understand how the predictions occur and how the features affect the predictions [26,27]. Explainable AI facilitates the understanding of how predictions happen using explanations of the feature effects on the predictions. Explainable Artificial intelligence (XAI) makes AI models with decisions understandable to humans, such as Local Explanations, Global Explanations, Rule-based Explanations, and instance-based explanations [28]. Traditional black-box models, while achieving impressive accuracy, often leave farmers in the dark regarding the rationale behind their recommendations [28]. XAI techniques, including feature importance analysis, model introspection, and counterfactual reasoning, focus on these processes, revealing which variables in the vast agricultural data hold the most significant weight in impacting predictions [29,30]. Terminology such as black box, gray box, and white box is widely used to describe varying degrees of accessibility to the internal essence of a component in many research [31]. In contrast, a white box component is entirely transparent, revealing all aspects of its design to the user. On the other hand, intermediate levels of disclosure, referred to as grey box components, exist depending on the extent of available details. The black box environment within the context of AI is referred to as the challenge of the system's difficulty in providing a clear explanation for how it reached an answer, commonly known as "the black-box nature" [32]. This problem underscores the difficulty in comprehending the inner workings of AI systems, hindering users' ability to understand how the system arrives at specific answers. Therefore, the concept of XAI plays a crucial role in minimizing the black box concept problem commonly associated with complex machine learning models [27]. XAI offers transparency and interoperability of the machine learning approaches. Therefore, XAI seeks to expose the decision-making processes of complex machine learning (ML) models, making their outputs more understandable for human users. The significance of XAI becomes particularly pronounced in scenarios where the outcomes of AI systems have significant importance in various research domains including healthcare, agriculture, autonomous vehicle operations, etc. [33]. The key objective of XAI is to build trust among end users. Therefore, AI-generated predictions and decisions are more human-friendly [34]. This approach not only stimulates accountability and sustainable use of AI but also encourages broader adoption by ensuring that users can interpret and act upon AI outputs with confidence.

Cabbage is primarily considered an upcountry vegetable in Sri Lanka. During the 2023 Maha season, total cabbage production in the country reached 52,310.4 MT [35]. While cold climatic conditions generally favor cabbage cultivation, heat-tolerant varieties have been developed for successful growth in the dry zone [17,36]. The Department of Agriculture, Sri Lanka recommends specific cabbage cultivars for different climatic zones. Exotic and Hercules varieties thrive in the up-country wet zone due to their cold tolerance, while Exotic, AS Cross, and KY Cross offer broader adaptability for the mid-country. Recognizing market demands, breeders have developed hybrid varieties with desirable traits, such as those offered by Royal Sluis, Green Coronet, Golden Cross, Green 123, Tropicana, and GS Cabbage [1,37–39].

Sajindra et al. [1] have predicted the cabbage cultivation soil N, P, and K levels using traditional black-box models [1]. However, the explainability of the developed prediction model was not discussed. This research aims to develop an XAI model to predict soil nitrogen, phosphorus, and potassium levels in cabbage cultivation. By focusing on individual nutrient predictions, the study seeks to enhance model transparency and provide actionable insights for precision agriculture. This approach addresses the limitations of traditional black-box models by elucidating the complex relationships between cabbage plant characteristics and soil macronutrient content. The ultimate goal is to enable targeted fertilization strategies that optimize resource use, reduce environmental impact, and promote sustainable agricultural practices.

2. Materials and methods

2.1. Case study data set

The cabbage plants were systematically cultivated within carefully controlled greenhouse environments, adhering rigorously to established standards. The selected research sites comprised greenhouses situated in Marassana and Nuwara Eliya within the Central Province, as well as Welimada in the Uva Province of Sri Lanka (refer to Fig. 1). Additionally, each greenhouse in the mentioned locations was utilized to cultivate ten Green Coronet cabbage plants in individual soil pots for the study's purpose. The cultivation process adhered to standardized conditions and agricultural protocols intricately designed to optimize cabbage growth, ensuring a consistent and uniform environment. These protocols encompassed precise management of various parameters, including temperature, lighting, humidity, irrigation, ventilation, and soil pH, along with the thorough implementation of measures aimed at pest and disease control [1].

At weekly intervals (every 7 days), the concentrations of soil major nutrients, specifically nitrogen, phosphorus, and potassium, in the vicinity of the plant near the soil were quantified using an NPK conductivity sensor (JXBS3001-DLJ). This monitoring persisted over a span of 85 days [1].

The nutrient concentrations were quantified in milligrams per kilogram ($mgkg^{-1}$). Concurrently, various plant characteristics, including plant height, leaf count, and average leaf area, were systematically recorded at weekly intervals (every 7 days) over the entire 85-day observation period. The plant height, measured in centimeters (cm), and the average leaf area, measured in square centimeters (cm^2), for each selected plant, were determined using a flexible tape graduated in millimeters (mm) and a portable leaf area meter (LI-3000C) [40,41].

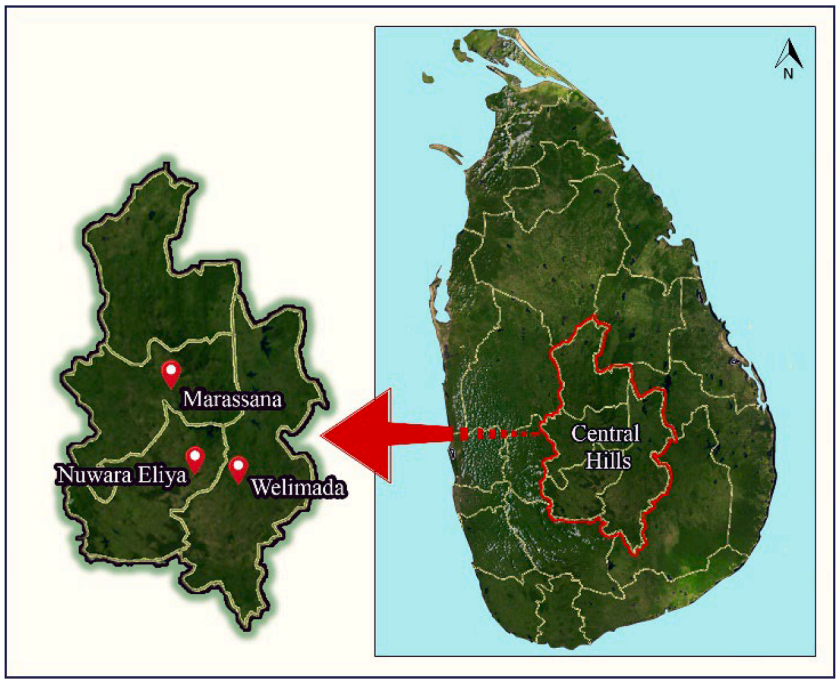

Fig. 1. Selected research sites of central hills in Sri Lanka.

The cabbage plant height was measured as the vertical distance from the soil surface to the apex of the cabbage plant at each growth stage. The cabbage leaf area was assessed using a scanning head equipped with diodes and paired detectors, enabling precise measurement as the cabbage leaf passed through the scanning mechanism.

2.2. Data set evaluation

The dataset collected for the study on cabbage plant growth was devoid of missing values due to the systematic acquisition of relevant measurements for both plants and soil pots at consistent weekly intervals. Furthermore, a notable absence of complexity and a relatively low standard deviation were observed in the dataset (Table 1). The data distribution of the features is shown in Fig. 2.

According to the data distribution, the Day feature consistently showed a frequency value of 30, corresponding to the 30 plants selected for the study over the total 84 days. The Average Leaf Area, Plant Height, and Plant Leaf Count features exhibited an increase in frequency as the feature values increased. The distributions of the Day and Plant Leaf Count features have fewer values spread across the sample space.

For further clarification and to select a model for prediction, a pre-evaluation was conducted among Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Decision Tree, and Artificial Neural Network (ANN). Based on the results, ANN was selected for prediction in the model.

2.3. Overview of machine learning models employed in the study

In this study, SVR, KNN, Decision Tree, and ANN models were employed to predict soil nutrient levels due to their effectiveness in similar agricultural applications. SVR is adept at handling high-dimensional data and capturing complex relationships between variables, making it suitable for modeling soil nutrient dynamics. KNN offers simplicity and effectiveness in capturing local patterns within the data, which is beneficial for soil nutrient prediction. Decision Trees provide clear interpretability by illustrating decisions based on feature values, aiding in understanding the influence of different factors on soil nutrients. ANNs are capable of modeling complex, non-linear relationships, making them suitable for capturing intricate patterns in soil nutrient data. The feature scaling process was applied to the ANN, SVR, and KNN models. The 'StandardScaler' method was used for this purpose. However, for tree based models, such as Decision Trees, scaling was not required, as these models are insensitive to feature magnitudes. Therefore, they were applied directly without scaling. By leveraging these

Table 1
Descriptive data of factors used in the model.

Days	7	14	21	28	35	42	49	56	63	70	77	84
Mean height	0	5.03	8.1	10.83	15.17	16.27	19.00	20.1	22.13	23.37	23.37	23.37
Mean Average leaf area	0	5.57	53.07	93.9	148.63	192.53	261.63	309.37	354.6	378.77	381.73	382.97
Mean of leaf count	0	2.00	3.73	5.93	8.50	10.33	11.90	13.33	13.5	13.57	13.60	13.60
Mean N content ($mgkg^{-1}$)	43.75	42.27	40.82	35.9	32.47	30.91	25.11	23.82	21.37	20.17	19.99	19.81
Mean P content ($mgkg^{-1}$)	19.21	17.31	15.63	14.34	13.26	11.27	10.15	10.5	9.83	10.23	9.72	9.43
Mean K content ($mgkg^{-1}$)	51.55	45.35	44.32	43.01	38.88	32.75	30.45	28.42	27.64	26.41	26.65	26.45
SD of height	0	0.55	0.75	0.73	1.07	1.00	1.10	1.01	1.23	1.14	1.14	1.14
SD of Average leaf area	0	1.15	7.99	10.96	26.76	13.56	28.19	40.54	35.29	26.87	27.72	27.93
SD of leaf count	0	0	0.57	0.25	0.56	0.57	0.47	0.75	0.72	0.72	0.76	0.76
SD of N content ($mgkg^{-1}$)	2.00	2.01	1.91	1.84	1.67	1.59	1.33	1.15	0.87	0.80	0.63	0.60
SD of P content ($mgkg^{-1}$)	0.95	0.45	0.49	0.27	0.50	0.37	0.28	0.29	0.57	0.67	0.57	0.64
SD of K content ($mgkg^{-1}$)	0.73	1.08	1.57	0.82	1.24	0.76	0.84	0.77	0.93	0.63	0.61	0.59

SD denotes Standard Deviation.

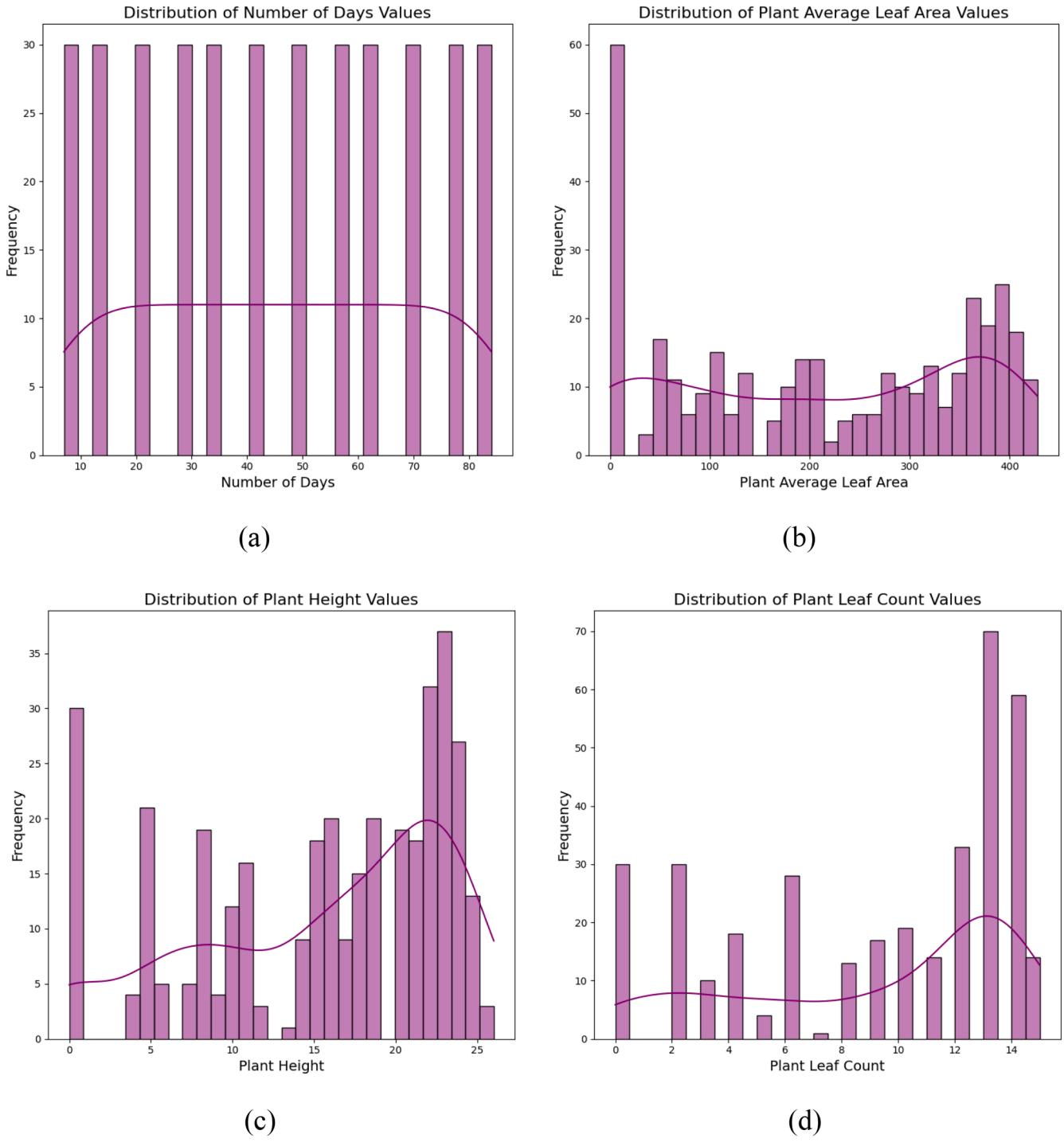

Fig. 2. Data distribution of the features (units: Plant Average Leaf Area (cm^2); Plant Height (cm)); The variables (a), (b), (c), and (d) represent the number of days, average leaf area, plant height, and leaf count, respectively, with each feature's deviation analyzed in relation to its frequency.

diverse methodologies, the study aimed to identify the most accurate and interpretable models for soil nutrient prediction, thereby enhancing precision agriculture practices.

2.3.1. *Support vector regression (SVR)*

SVR is a sophisticated and versatile machine learning model used primarily for regression tasks. It is an extension of the Support Vector Machine, a model traditionally used for classification [42]. SVR operates by finding a hyperplane that best fits the data while maintaining a margin of tolerance (epsilon) around the predicted values. This margin allows the model to be robust to outliers and noise, making SVR particularly useful in real-world scenarios where data may not be perfectly clean [43]. One of the key features of SVR is its ability to handle non-linear relationships through the use of kernel functions, such as radial basis function, polynomial, and sigmoid kernels. These kernels map the input data into a higher-dimensional space, allowing the model to capture complex patterns that linear models might miss [44]. However, SVR requires careful tuning of its hyperparameters, such as the regularization parameter, epsilon, and kernel parameters, to achieve optimal performance [45]. Additionally, while SVR is effective in high-dimensional spaces, it can be computationally expensive, especially for large datasets, due to the need to solve quadratic programming

problems during the training phase [46].

2.3.2. K-Nearest neighbors (KNN)

KNN is one of the simplest yet powerful machine learning models, widely used for both classification and regression tasks [47]. KNN is an instance-based, non-parametric learning algorithm, meaning it does not assume any specific form for the underlying data distribution. Instead, KNN operates by storing the entire training dataset and making predictions based on the similarity between the new data point and the existing points [48]. The similarity is typically measured using a distance metric such as Euclidean distance, though other metrics like Manhattan or Hamming distance can also be used depending on the nature of the data [49]. The key parameter in KNN is 'K,' the number of nearest neighbors considered when making a prediction. A smaller K makes the model sensitive to noise and outliers, while a larger K smooths the decision boundaries. Despite its simplicity and ease of implementation, KNN can become computationally intensive during the prediction phase, especially with large datasets, since it requires calculating distances to all training instances. Furthermore, KNN is highly sensitive to the scale of the features, necessitating proper data normalization or standardization before applying the model [50].

2.3.3. Decision Tree

Decision Trees are versatile machine learning models used for both classification and regression, known for their simplicity and interpretability. A Decision Tree model works by recursively splitting the dataset into subsets based on the most significant feature at each node, which is determined using criteria like Gini impurity, information gain, or variance reduction [51]. The tree grows by creating branches and leaves, where each branch represents a decision based on a feature, and each leaf node represents a final prediction [52]. One of the key strengths of Decision Trees is their interpretability; the model's decision-making process is transparent, allowing users to easily understand how the predictions are made. However, Decision Trees have a tendency to overfit, particularly with noisy or complex data, leading to poor generalization to new, unseen data. To mitigate overfitting, pruning techniques are often employed, which involve removing branches that add little predictive power to the model [53]. Despite their limitations, Decision Trees can handle both numerical and categorical data, making them a popular choice in various applications [12].

2.3.4. Artificial neural network (ANN)

ANNs are highly flexible and powerful models inspired by the structure and functioning of the human brain. ANNs consist of layers of interconnected neurons, where each neuron receives inputs, applies a transformation through an activation function, and passes the output to the neurons in the next layer. The architecture of an ANN typically includes an input layer, one or more hidden layers, and an output layer. The hidden layers are where the model learns to capture complex, non-linear relationships in the data [54]. ANNs are trained using a process called backpropagation, where the model adjusts the weights of the connections between neurons to minimize the error between predicted and actual values. This training process involves optimization techniques like gradient descent, which iteratively updates the weights to converge to a solution that best fits the data [55]. ANNs are incredibly versatile and can be applied to a wide range of tasks, including image recognition, natural language processing, and time-series forecasting [56,57,58].

2.4. Model development and training

A neural network model was developed in this study. The model comprises three hidden layers, excluding the input and output layers, with each hidden layer comprising one hundred nodes. Additionally, the input layer consists of four nodes, while the output layer consists of three nodes (Fig. 3). Eq. (1) represents the formulation of the relationship modeled in this study.

$$N_{content}, P_{content}, K_{content} = \phi \left(\begin{array}{c} Plant\ Height,\ Number\ of\ Leaves, \\ Average\ Leaf\ Area,\ Number\ of\ Days \end{array} \right) \quad (1)$$

The dataset was divided, with 80 % allocated for training the model and the remaining 20 % for testing. Additionally, 15 data points were separated to cover all stages of the plant for validation purposes [59].

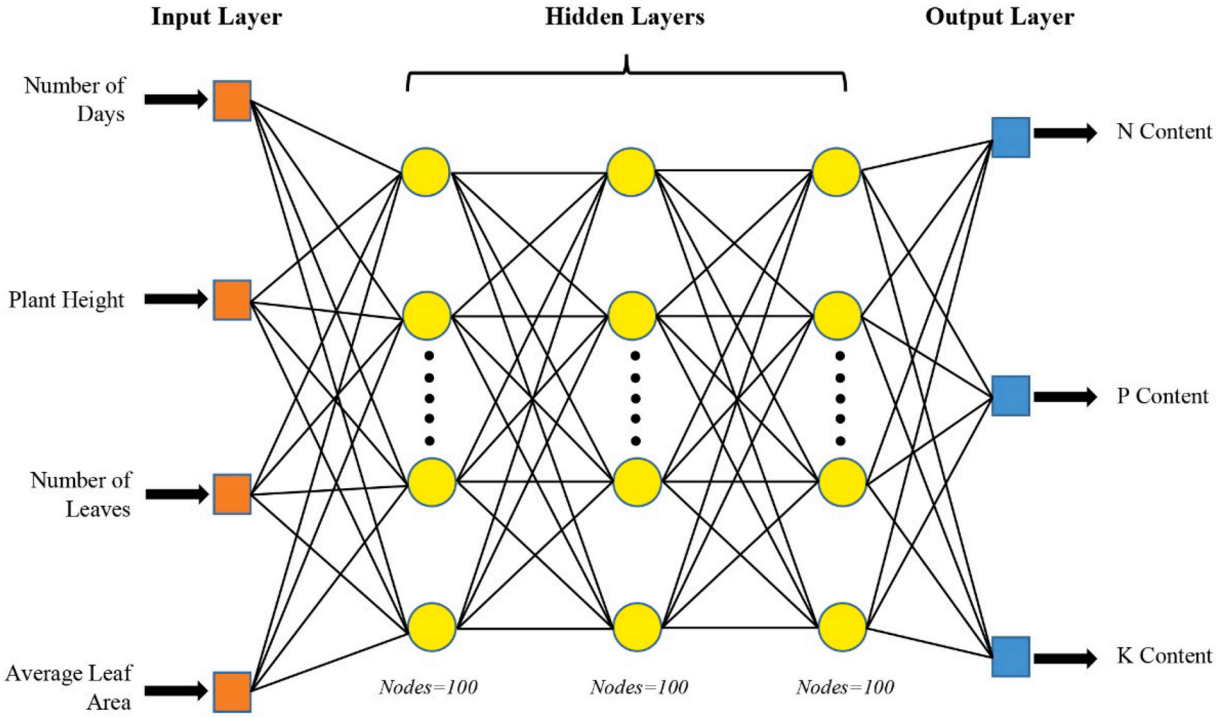

Fig. 3. Architecture of the ANN model.

Furthermore, the SHAP and LIME methods were employed to clarify the model's predictions [60]. The SHAP technique was utilized to comprehend the overall behavior of the model across the entire dataset, rather than focusing on individual predictions. Similarly, the LIME technique was utilized to explain the individual predictions made by the model in a locally interpretable manner. The entire computational framework was implemented using Python as the programming language. Google-Colab was used as the computational environment, providing the requisite resources for training and evaluation purposes [57].

Web Application Architecture and Data Flow - The tool was developed and deployed in a Google Colab environment, featuring a user-friendly front-end interface. Connectivity between the server and client sides was established using Ngrok, enabling the prediction of major soil nutrient content based on cabbage plant characteristics. The Ngrok is a flexible API gateway, that provides instant and secure connectivity for testing the web application, enabling remote access to the locally hosted model without requiring a complex setup.

The interface includes an HTML form with input fields corresponding to three key features derived from cabbage plants and weekly time intervals. When the user enters data and submits the form, the input is captured and sent via an HTTP POST request to the server-side. Here, the pre-trained machine learning model, stored in the cloud environment, processes the input and generates predictions.

Model Integration and Response Handling - Upon receiving the input data, the server-side loads the machine learning model as *.sav* file and feeds the input data into it for prediction. The model processes the input and returns the predicted values. These predictions are then encapsulated in the server's response and sent back to the client side. The HTML page updates to display the prediction results, providing real-time feedback to the user. This process, powered by Ngrok, services and uses HTTPS protocol to ensure secure and efficient communication between the client (HTML interface) and the server (model processing), all within the cloud environment.

2.5. Performance evaluation

In this model, we utilized 200 iterations to achieve a high level of accuracy, reaching 97 % for both training and testing. Fig. 4 illustrates the learning curve of the model, where the X-axis represents the number of iterations, and the Y-axis represents the loss during the training stage. The loss rapidly decreases in between iterations 0 to 24 during the training stage of the model as expected. Subsequently, the plot indicates minimal changes in loss during iterations 75 to 200. However, the model achieved good accuracy with a negligible difference between training and testing accuracy level under 200 iterations.

After completing the development of the Deep learning model, a user-friendly interface was developed. This application was designed with the aim of simplifying user interactions and enhancing overall usability, ensuring a seamless experience for users interacting with the model's functionalities.

2.6. Shapley additive explanations (SHAP)

The SHAP framework aims to explain a model's predictions by calculating the contribution of each feature to the predictions [61]. It utilizes Shapley Values, originally from game theory, to achieve this objective. As such, SHAP aligns with fundamental principles in game theory when determining feature importance. Moreover, SHAP offers insights at both local and global levels regarding an ML model's behavior [62,63,64]. Lundberg and Lee [61] have proposed employing the Shapley value to calculate feature importance. Additionally, Liang et al. [65] presented an intricate classification of explanation models, with SHAP being identified as a data-driven and perturbation-based technique. As a result, SHAP depends on input parameters and does not necessitate understanding the operation sequence of the ML model [62,63]. Lundberg and Lee [61] have introduced various versions of SHAP tailored to specific ML model categories, such as DeepSHAP, Kernel SHAP, LinearSHAP, and TreeSHAP [66]. In this study, Kernel SHAP is utilized to explain the deep learning model predictions. Moreover, this research utilizes keras, Scikit-learn (Sklearn), Matplotlib, NumPy, Pandas, and the SHAP libraries for implementation purposes.

Lundberg and Lee demonstrated that employing a weighted linear regression model as the local surrogate, along with an appropriate weighting kernel, allows the regression coefficients of the LIME surrogate to estimate the SHAP values [61]. The Shapley kernel used to derive SHAP values is as follows (refer to Eq. (2)).

$$\pi_{x'}(z') = \frac{(M-1)}{(M\ chose|z'|)|z'|(M-|z'|)} \tag{2}$$

M denotes the quantity of features, while $|z'|$ represents the number of non-zero features in the simplified input z'. The importance of Shapley values for XAI, can be described throughout all additive feature attribution methods [67]. Overall, the SHAP calculate the contribution of each feature (player) to the model prediction (game).

2.7. Local interpretable model agnostic explanations (LIME)

LIME is also an explanation method of XAI in local instances. Hence,

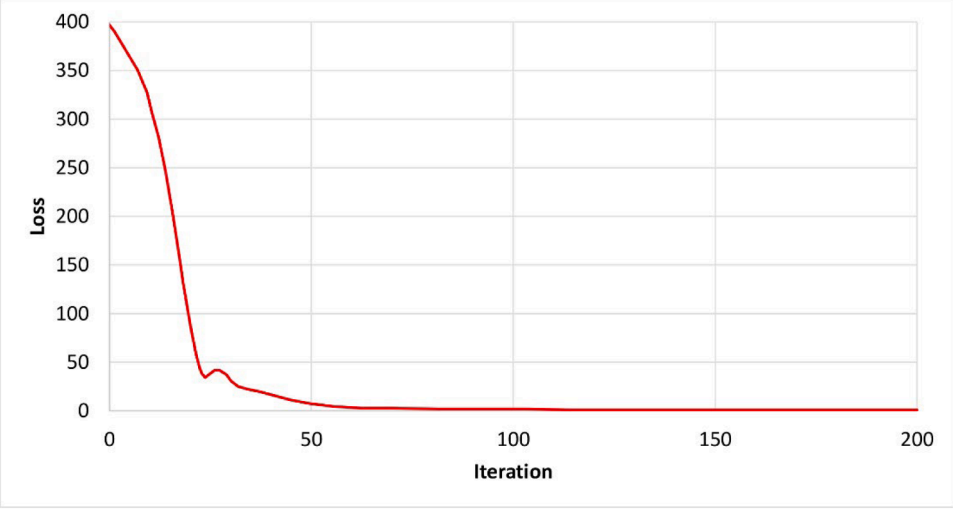

Fig. 4. Training stage learning curve of the model.

the LIME method is designed to provide reliable explanations for the predictions of any classifier or regression. It achieves this by locally approximating the model with an interpretable model. LIME is focused on providing an easily interpretable method with a strong emphasis on local reliability. Local reliability ensures that explanations for individual predictions accurately reflect how the model behaves in the immediate vicinity of the prediction being made [68]. LIME employs a structured approach to explain predictions made by the model. The process involves generating new samples and subsequently obtaining their predictions using the original model. These new samples are then weighed based on their proximity to the specific instance being explained. This methodical process allows LIME to provide reliable and interpretable explanations for individual predictions, enhancing the transparency and understanding of complex ML models [69].

By utilizing the output probabilities derived from a set of samples that encompass the relevant input area for clarification, LIME constructs a linear model. The weights of this substitute model are then utilized to evaluate the importance of input features. Moreover, LIME's model-agnostic nature allows it to be applied effectively to various machine-learning models without constraints [68,69].

2.8. Overall methodology

To provide a comprehensive understanding of the process, the overall methodology is illustrated in Fig. 5 as a block diagram. The following steps were carried out.

1. The web application is started on the cloud environment.
2. A tunnel to the private cloud environment is created through the specified port using the 'Ngrok' executable process.
3. The publicly accessible temporary URL is shared.
4. The received URL is shared with the end-users who need to access the web application.
5. The web URL is accessed by end users on the internet.

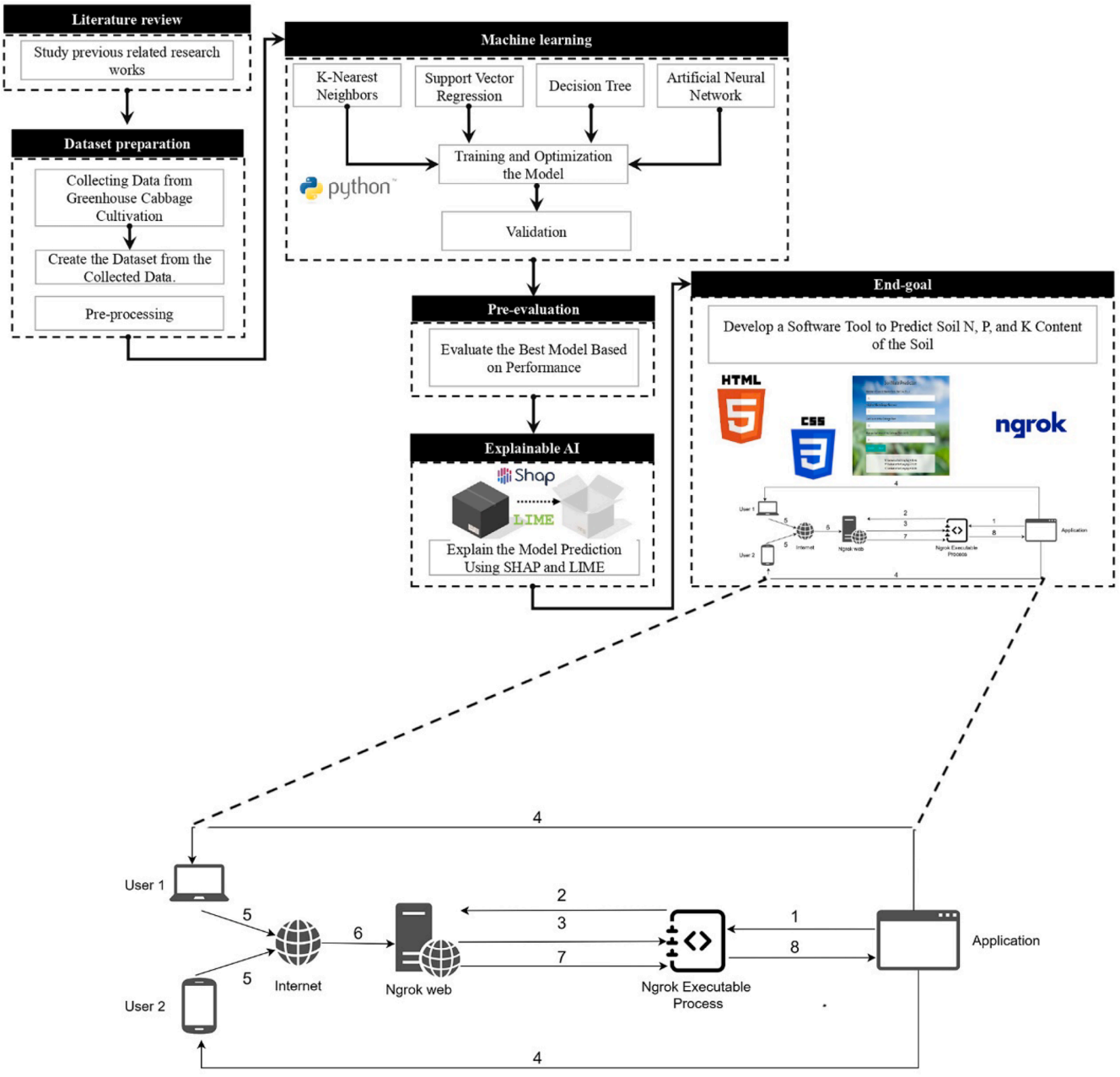

Fig. 5. Flow diagram of the study.

6. Once the request arrives at the 'Ngrok' web server (as the URL belongs to an ngrok.com subdomain, requests are first routed to the 'Ngrok' web server), it maps the URL to the corresponding application for tunneling.
7. The request is forwarded by Ngrok to the tunnel established in Step 2.
8. The forwarded request is connected to the application.

2.9. Model accuracy evaluation

The coefficient of determination, often denoted as R^2, is a statistical measure that assesses the proportion of the variance in the dependent variable (y) that is explained by the independent variable(s) (x) in a regression model. In other words, it quantifies the goodness of fit of the regression model to the observed data. The R^2 equation is shown in Eq. (3).

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2} \quad (3)$$

y_i denotes the actual y values, while $\hat{y}_i$ denotes predicted y values. $\overline{y}$ denotes the mean of the y values. SSR is a statistical term used to measure the amount of variation in the dependent variable (y) that is explained by the independent variable(s) (x) in a regression model. In other words, SSR quantifies how well the regression line fits the data points. SST is a statistical measure that represents the total variability in the dependent variable (y) without considering any explanatory variables. It quantifies the dispersion of the observed y values around their mean ($\bar{y}$), regardless of any relationship with the independent variable (s) [70].

The mean squared error (MSE) is a statistical measure that quantifies the average squared difference between the actual observed values of a variable and the corresponding predicted values produced by a model. In the context of regression analysis, MSE is commonly used to assess the accuracy of a predictive model (refer Eq. (4)) [1].

$$MSE = \frac{\sum_{i=1}^{N} [x_i - y_i]}{N} \quad (4)$$

Let x represent the observed values and y denote the predicted maximum value for a given observation, ranging from i to n. The parameter N signifies the total number of observations.

MBE (Eq. (5)) stands for Mean Bias Error, a statistical measure used to assess the accuracy of a predictive model by calculating the average bias or difference between the predicted and observed values. It helps in determining whether a model tends to overestimate or underestimate the actual values. A positive MBE indicates that the model, on average, overestimates the observed values, while a negative MBE suggests that the model, on average, underestimates them. This measure is particularly useful in evaluating model performance and understanding the direction of bias.

$$MBE = \frac{1}{n} \sum_{i=n}^{n} (P_i - O_i) \quad (5)$$

n represents the total number of observations or data points, P_i is the predicted value for the i th observation, and O_i is the observed (actual) value for the i th observation.

3. Results and discussion

3.1. Pre-evaluation and model predictions

The pre-evaluation results of the SVR, KNN, Decision Tree, and ANN models can be seen in Table 2. This section details how features such as average leaf area, plant height, number of leaves, and number of days function within these models, and presents their performance metrics—including R^2, RMSE, MSE, and MBE—across various models.

According to the results, comparatively lower R^2 values of 0.70 and 0.69 for training and testing, respectively, were recorded in KNN, while the highest values were observed in ANN, with 0.96 and 0.97 for training and testing, respectively. This indicates that the ANN model demonstrated superior regression performance compared to the other models. Furthermore, the results show that KNN recorded the highest MSE and RMSE values, with 14.69 and 3.83 for training, and 17.15 and 4.14 for testing, respectively. In contrast, ANN recorded the lowest MSE and RMSE values, with 1.82 and 1.35 for training, and 1.18 and 1.09 for testing, respectively. These findings highlight ANN as the most accurate model, outperforming the others in terms of predictive accuracy and error minimization. For further clarification of the model evaluations, Fig. 6 shows a comparison of the actual and predicted values in the training and testing graphs for each model. The ANN model demonstrates the best performance, with minimal deviations exceeding the 10 % and 20 % error lines, achieving high R^2 values for both training and testing (Fig. 6(d) and (h)). In both figures, all the actual versus predicted data points are clustered near the regression line, indicating a low error of the model.

According to the MBE values, the ANN model shows the closest training and testing values to zero, with 0.11 and −0.01, respectively. This indicates that the model has minimal overestimations and underestimations in both training and testing, suggesting low systematic bias and strong prediction accuracy compared to other models. When comparing R^2, RMSE, MSE, and MBE values with other models, the overall best model is ANN.

3.2. Explanation of model predictions using SHAP

Fig. 7 shows each feature importance for the prediction of soil N, P, and K content.

In the heat map, colors represent data values that indicate the impact of different features on predicting N, P, and K content. Features associated with higher values in the heat map have a greater effect on predicting the respective nutrients. According to Fig. 7, the Day feature had the highest impact on the prediction of soil N, and K, while the Plant Height feature had the lowest impact for the predictions. The least effect on the prediction was observed in the prediction of soil N content by the Plant Height feature, while the greatest effect was observed in the prediction of soil K content by the Day feature.

Fig. 8 represents a global dataset, with SHAP values displayed on the x-axis, indicating their impact on the model output. The y-axis represents the selected features in the model. As the feature values transition from low to high, the color gradient shifts from blue to red, denoting the change in feature values [71].

The levels of soil remaining N, P, and K significantly affect various features: height, leaf count, average leaf area, and days. This impact is

Table 2
Descriptive data of factors used in the model.

Index	Training				Testing			
	SVR	KNN	Decision Tree	ANN	SVR	KNN	Decision Tree	ANN
R^2	0.86	0.70	0.79	**0.96**	0.87	0.69	0.85	**0.97**
RMSE	2.89	3.83	3.15	**1.35**	3.03	4.14	2.97	**1.09**
MSE	8.37	14.69	9.90	**1.82**	9.19	17.15	8.80	**1.18**
MBE	−0.67	−1.13	−0.13	0.11	−0.89	−1.53	0.05	−0.01

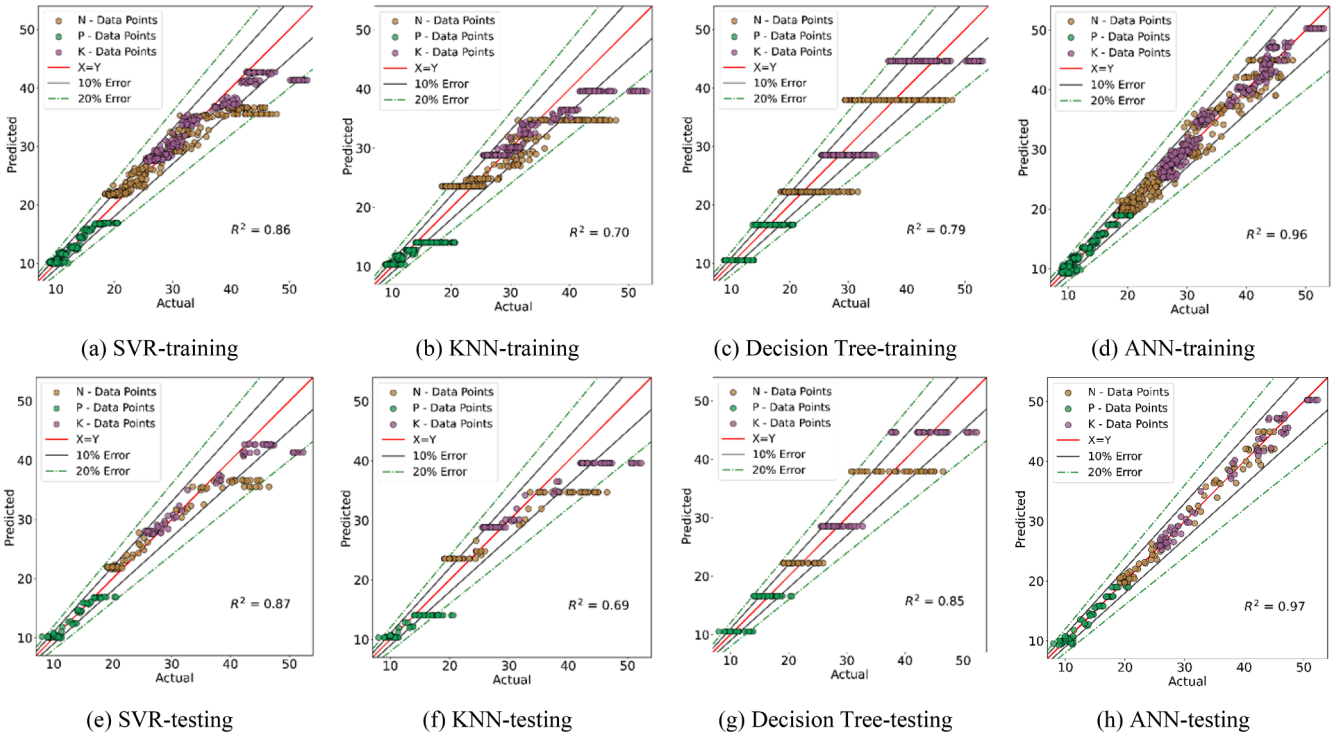

Fig. 6. Comparison of actual and predicted values with different Models.
(a) to (d) denote the training results of the models, while (e) to (h) denote the testing results of the models.

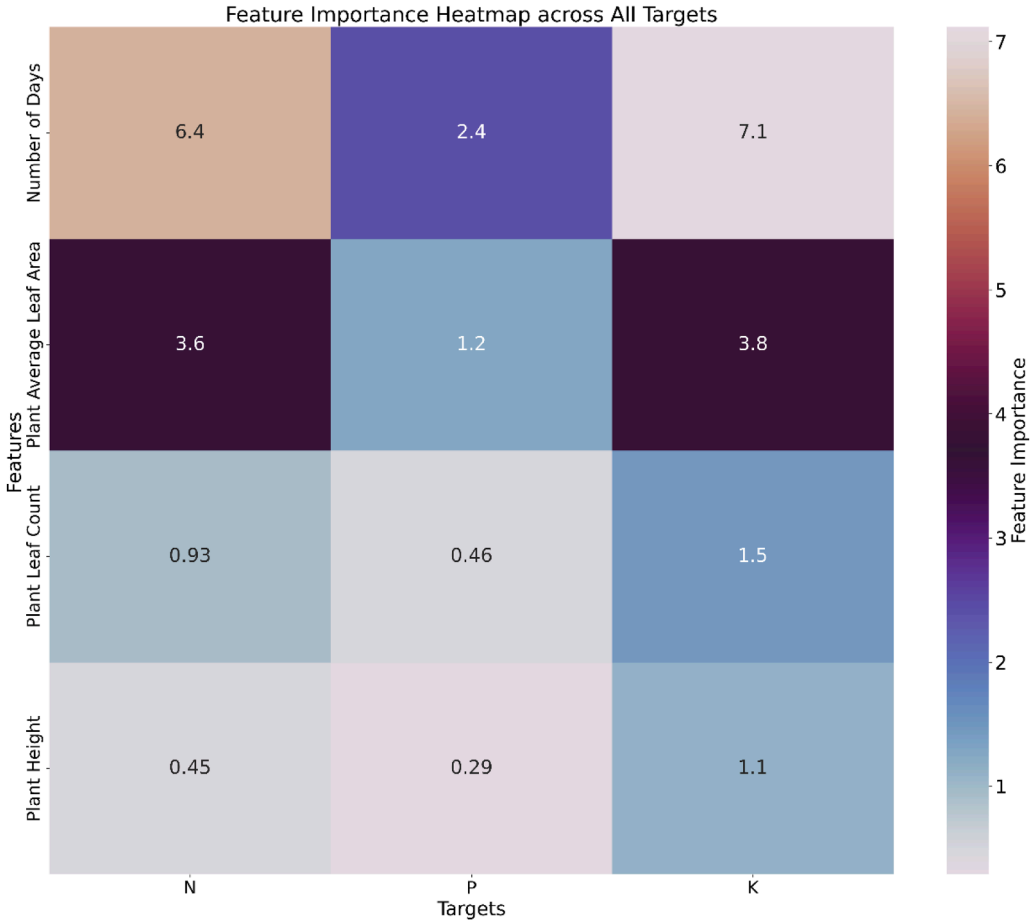

Fig. 7. Each feature importance for the prediction.

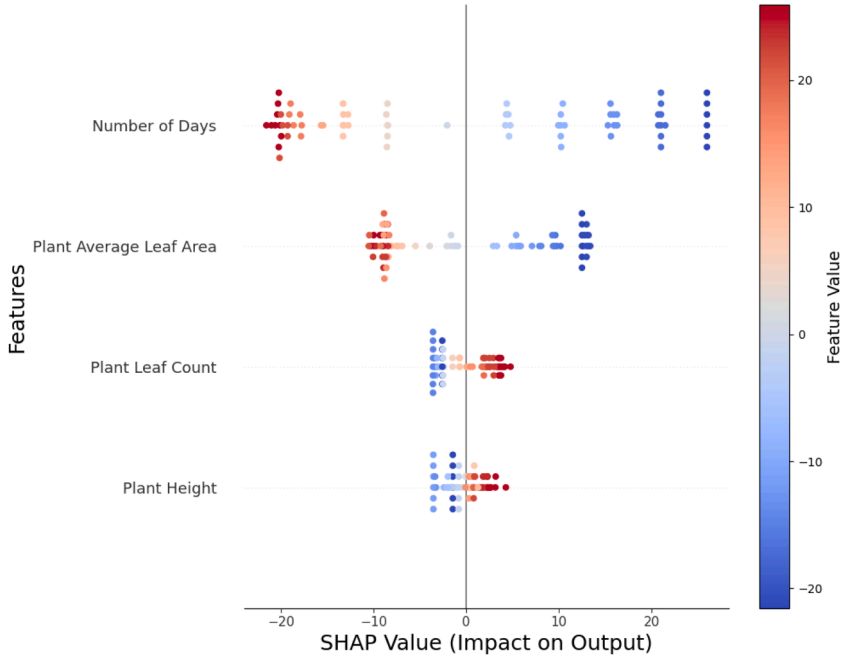

Fig. 8. SHAP values for the different features used in the model.

discernible through the distribution range of feature values with the SHAP values. The analysis of SHAP values reveals notable insights into the dynamic relationship between various factors and the model's predictive accuracy. Firstly, as the number of days features increases, it affects negatively on the model predictions of soil nutrient content, as evidenced by the presence of negative SHAP values. Similarly, when the average leaf area feature increases, the model predictions of soil nutrients content show a negative impact with negative SHAP values.

When increasing plant leaf count feature, it affects to positively to the model prediction of soil nutrient showing positive SHAP values. When the plant height feature increases it shows improved predictions of soil nutrients by showing positive SHAP values. This finding suggests that as the number of leaves and plant height increase over time, the model benefits from more robust and distinguishable features, enhancing its ability to make accurate predictions. Overall, these observations underscore the intricate interaction among temporal factors, such as cultivation duration and evolving plant characteristics, and their influence on the model's predictive performance, emphasizing the significance of considering these dynamics in agricultural modeling and analysis.

3.3. Explanation of model predictions using LIME

When considering instances, LIME was used in six specific instances [72], with two instances each categorized as high, middle, and low. These instances corresponded to different stages of plant growth, including the near-harvest stage of cabbage, the middle growth phase, and the initial growth phase of cabbage.

In the 1st high instance, Plant Height and Plant Leaf Count recorded positive contribution values for the prediction of N, P, and K, with respective values of 1.3, 0.68, 1.92 and 2.45, 1.22, 3.78, indicating a positive impact on the prediction. Conversely, Average Leaf Area and Day features recorded negative contribution values for the prediction of N, P, and K, with respective values of −5.84, −1.88, −6.05 and −10.21, −3.52, −10.80, indicating a negative impact on the prediction. In the 2nd high instance, Plant Height and Plant Leaf Count recorded positive contribution values for the prediction of N, P, and K, with respective values of 1.25, 0.74, 1.65 and 1.20, 0.43, 1.57, indicating a positive impact on the prediction. Conversely, Average Leaf Area and Day features recorded negative contribution values for the prediction of N, P, and K, with respective values of −5.84, −1.88, −6.05 and −10.21, −3.52, −10.80, indicating a negative impact on the prediction.

Based on the analysis of two high instances, as depicted in Figs. 9 (a–c) and 9 (d–f), it was observed that the soil K content prediction is positively influenced by the features of leaf count and height, while it is negatively affected by the features of days and average leaf area. Similarly, for the prediction of soil N content, leaf count and height exhibit a positive impact, whereas days and average leaf area demonstrate a negative impact on the prediction, as seen in the aforementioned instances. Additionally, for soil P content prediction, height and leaf count also show a positive impact, while days and average leaf area exhibit a negative impact. These observations suggest that certain plant characteristics, such as leaf count and height, contribute positively to the prediction of soil nutrient content, while other factors, such as the number of days and average leaf area, may have a detrimental effect. The lowest contribution value (−10.80) Day feature consistently exhibited the greatest negative impact on the prediction in both instances, while the highest positive contribution value (3.78) on the prediction was observed with the leaf count feature in predicting K content of soil in 1st high instance (Fig. 9(c)). The highest positive contribution value (1.65) on the prediction in the 2nd high instance was observed with the Plant Height feature in predicting K content of soil.

In the 1st medium instance, Average leaf area recorded positive contribution values for the prediction of N, P, and K, with respective values of 1.5, 0.4, 1.69, indicating a positive impact on the prediction. Conversely, the Day, Plant Leaf Count, and Plant Height features recorded negative contribution values for the prediction of N, P, and K, with respective values of −0.1, −0.17, −0.08; −1.22, −0.52, −1.78; and −1.27, −0.71, −1.72, indicating a negative impact on the prediction. In the 2nd high instance, Plant Leaf Count recorded positive contribution values for the prediction of N, P, and K, with respective values of 1.04, 1.03, 1.94, indicating a positive impact on the prediction. Conversely, the Day, Plant Height, and Average leaf area features recorded negative contribution values for the prediction of N, P, and K, with respective values of −0.32, −0.35, −0.30; −1.34, −0.73, −1.81; and −4.21, −1.57, −5.10, indicating a negative impact on the prediction.

Based on the analysis of two medium instances for predicting the N content of the soil, the Average Leaf Area feature shows a positive effect

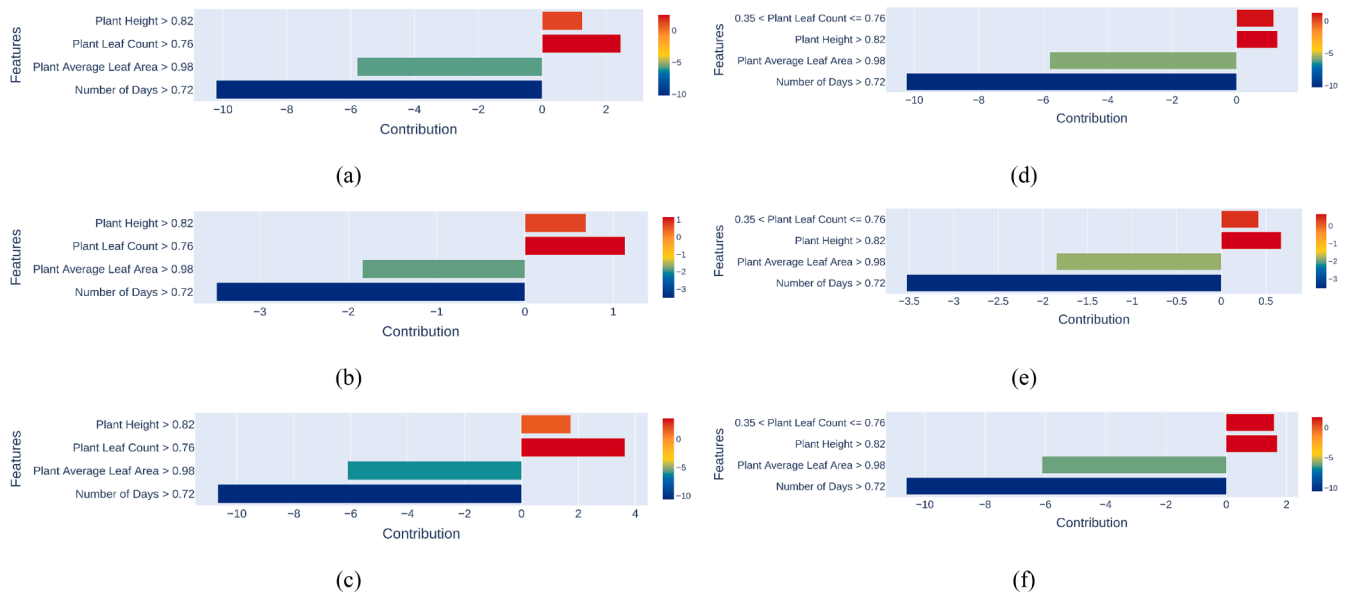

Fig. 9. LIME graphs for selected two high instances: (a), (b), and (c) represent the contribution of features for predicting soil N, P, and K, respectively, in the 1st high instance; (d), (e), and (f) represent the contribution of features for predicting soil N, P, and K, respectively, in the 2nd high instance of the testing data.

on the model's prediction in the 1st medium instance (Fig. 10a), while it shows a negative effect in the 2nd medium instance (Fig. 10d). Conversely, Plant Leaf Count shows a negative effect on the model's prediction in the 1st medium instance (Fig. 10a), while it shows a positive effect in the 2nd medium instance (Fig. 10d). In both instances, the Day and Plant Height features negatively affect the model in predicting soil K content. When predicting the P content of the soil, similar to the effect on N, the Average Leaf Area feature shows a positive effect on the model's prediction in the 1st medium instance (Fig. 10b), while it shows a negative effect in the 2nd medium instance (Fig. 10e). Conversely, Plant Leaf Count shows a negative effect on the model's prediction in the 1st medium instance (Fig. 10b), while it shows a positive effect in the 2nd medium instance (Fig. 10e). In both instances, the Day and Plant Height features negatively affect the model in predicting soil P content. When predicting the K content of the soil, similar to the effect on N and P, the Average Leaf Area feature shows a positive effect on the model's prediction in the 1st medium instance (Fig. 10c), while it shows a negative effect in the 2nd medium instance (Fig. 10f). Conversely, Plant Leaf Count shows a negative effect on the model's prediction in the 1st medium instance (Fig. 10c), while it shows a positive effect in the 2nd medium instance (Fig. 10f). In both instances, the Day and Plant Height features negatively affect the model in predicting soil K content. The greatest positive and negative contribution values in the 1st instance were 1.69 and −1.78, respectively, while in the 2nd instance, they were 1.94 and −5.10, respectively, in the prediction of soil K content.

In the 1st low instance, Average leaf area and day features recorded positive contribution values for the prediction of N, P, and K, with respective values of 9.07, 3.24, 9.16 and 13, 4.1, 13.72, indicating a positive impact on the prediction. In predicting soil N content, the Plant Height feature shows a minimal positive impact, resulting in a contribution value of 0.01. Conversely, Plant Height and Plant Leaf Count features recorded negative contribution values for the prediction of P,

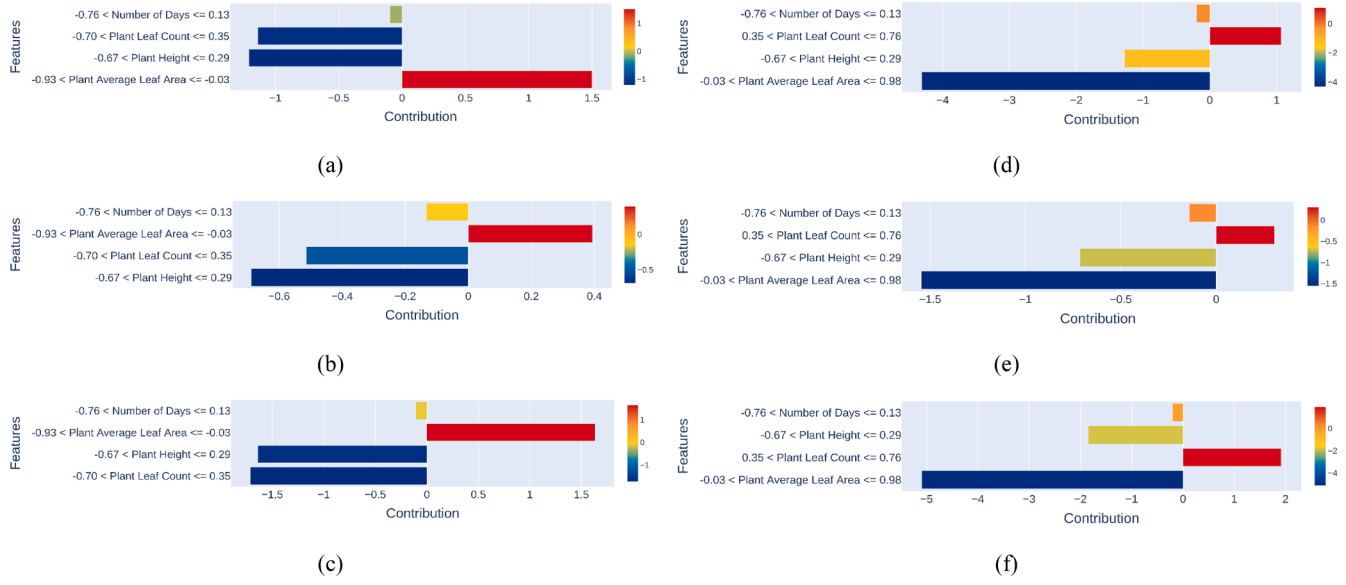

Fig. 10. LIME graphs for selected two medium instances: (a), (b), and (c) represent the contribution of features for predicting soil N, P, and K, respectively, in the 1st medium instance; (d), (e), and (f) represent the contribution of features for predicting soil N, P, and K, respectively, in the 2nd medium instance of the testing data.

and K, with respective values of −0.20, −0.30 and −0.91, −3.96, indicating a negative impact on the prediction. In predicting soil N content, the Plant Leaf Count shows a negative impact, resulting in a contribution value of −2.45. In the 2nd low instance, Average leaf area and day features recorded positive contribution values for the prediction of N, P, and K, with respective values of 1.45, 0.48, 1.76 and 13.11, 4.66, 13.71, indicating a positive impact on the prediction. In predicting soil N content, the Plant Height feature shows a minimal positive impact, resulting in a contribution value of 0.08. Conversely, Plant Height and Plant Leaf Count features recorded negative contribution values for the prediction of P, and K, with respective values of −0.03, −0.50 and −0.93, −3.77, indicating a negative impact on the prediction. In predicting soil N content, the Plant Leaf Count shows a negative impact, resulting in a contribution value of −2.70.

Based on the analysis of two low instances for predicting the N content of the soil, the Average Leaf Area, Plant Height, and Day features show a positive effect on the model's prediction, resulting in positive contribution values, while the Plant Leaf Count feature shows a negative impact on the model's prediction, resulting in negative contribution values (Fig. 11(a) and 11(d). For predicting the P content of the soil in two low instances, the Average Leaf Area and Day features show a positive effect on the model's prediction, resulting in positive contribution values, while the Plant Leaf Count and Plant Height features show a negative impact on the model's prediction, resulting in negative contribution values (Fig. 11(b) and 11(e)). Similar to the effects observed in P content prediction, for predicting the K content of the soil in two low instances, the Average Leaf Area and Day features show a positive effect on the model's prediction, resulting in positive contribution values, while the Plant Leaf Count and Plant Height features show a negative impact, resulting in negative contribution values (Fig. 11(b) and 11(e)). The greatest positive and negative contribution values in the 1st instance were 13.72 and −3.96, respectively, while in the 2nd instance, they were 13.7 and −3.77, respectively, in the prediction of soil K content.

3.4. 'SoilNutriPredictor' web application

With these results, a web application named "SoilNutriPredictor" has been developed with a user-friendly interface (refer to Fig. 12) to facilitate users' tasks. Users can feed four input parameters including the number of days on a weekly basis, the height of the cabbage plant in cm, the leaf count of the cabbage plant, and the average leaf area of the cabbage plant in cm^2. These inputs undergo a scaling process before being transferred to the developed deep-learning model. The model then predicts the N, P, and K content of the soil based on the characteristics of the plant, providing outputs in units of $mgkg^{-1}$. In addition, a reset button was created to clear the entered input values and prepare for new input.

Hence, the user-friendly application that can be used by the farmers showcases a greater novelty. This enhances the sustainability concept of food and agriculture which led to the reach of Sustainable Development Goal 2 – "Zero Hunger". Agriculture and food sustainability is an appealing topic among many communities in an increasing population phase. Therefore, the developments presented in this research would be highly important for understanding sustainable fertilizer levels to reach maximum benefits. In addition, optimal fertilizer levels in the farmlands minimize surface water runoff pollution. This is an added benefit from the developed model.

Mobile applications are more popular among modern society including farmers due to their ease of use and the evolution of the mobile phone industry. Hence a mobile application can be developed for this purpose, and future endeavors can be expanded to automatically detect all input features such as the number of days, height, leaf count, and average leaf area of the plant using computer vision techniques. Therefore, predicted outputs can be obtained by farmers and agriculture specialists in real-time situations without consuming much time. This would be the future of the Internet of Things (IoT) in agriculture.

Validation was conducted using newly grown, randomly selected greenhouse-grown cabbage plants at different stages. To validate the developed tool, 15 points of real-time data were used. After prediction, the differences between the actual and predicted data were observed (Fig. 13). The predicted values showed minimal differences from the real-time data, indicating that the model performed accurately with minimal errors.

4. Conclusions

Predictions for cabbage and other vegetables have been attempted in previous research using various machine learning techniques. However, a significant gap in the literature has been identified due to the lack of clear explanations regarding how these predictions are generated. The introduction of explainable artificial intelligence aimed to understand

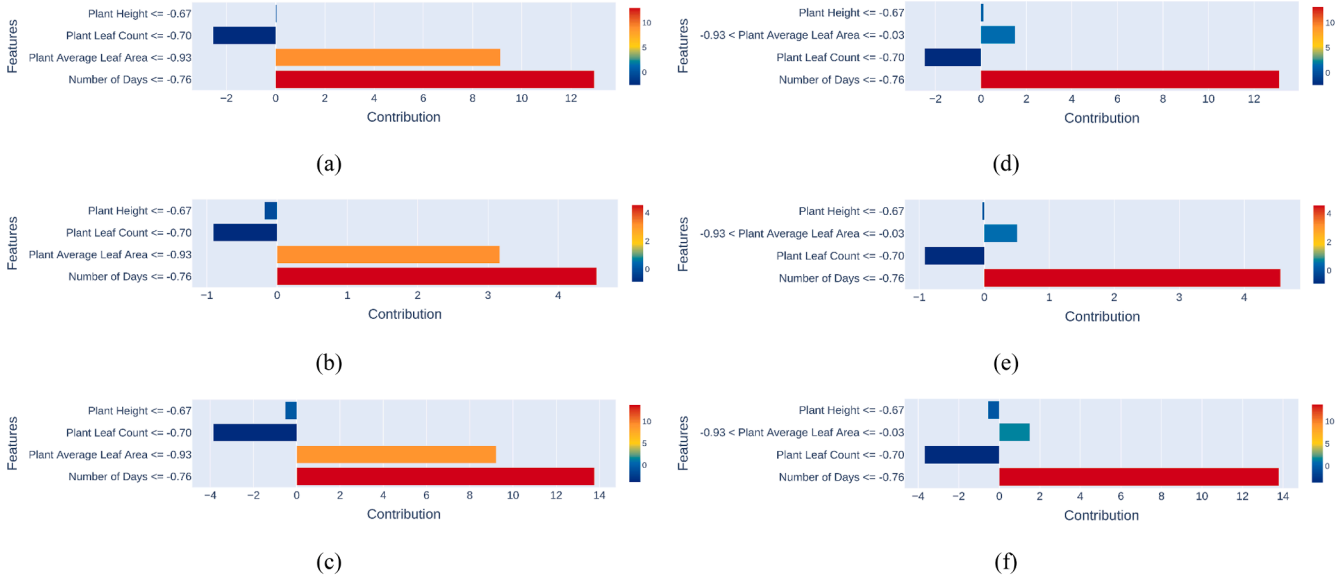

Fig. 11. LIME graphs for selected two low instances: (a), (b), and (c) represent the contribution of features for predicting soil N, P, and K, respectively, in the 1st low instance; (d), (e), and (f) represent the contribution of features for predicting soil N, P, and K, respectively, in the 2nd low instance of the testing data.

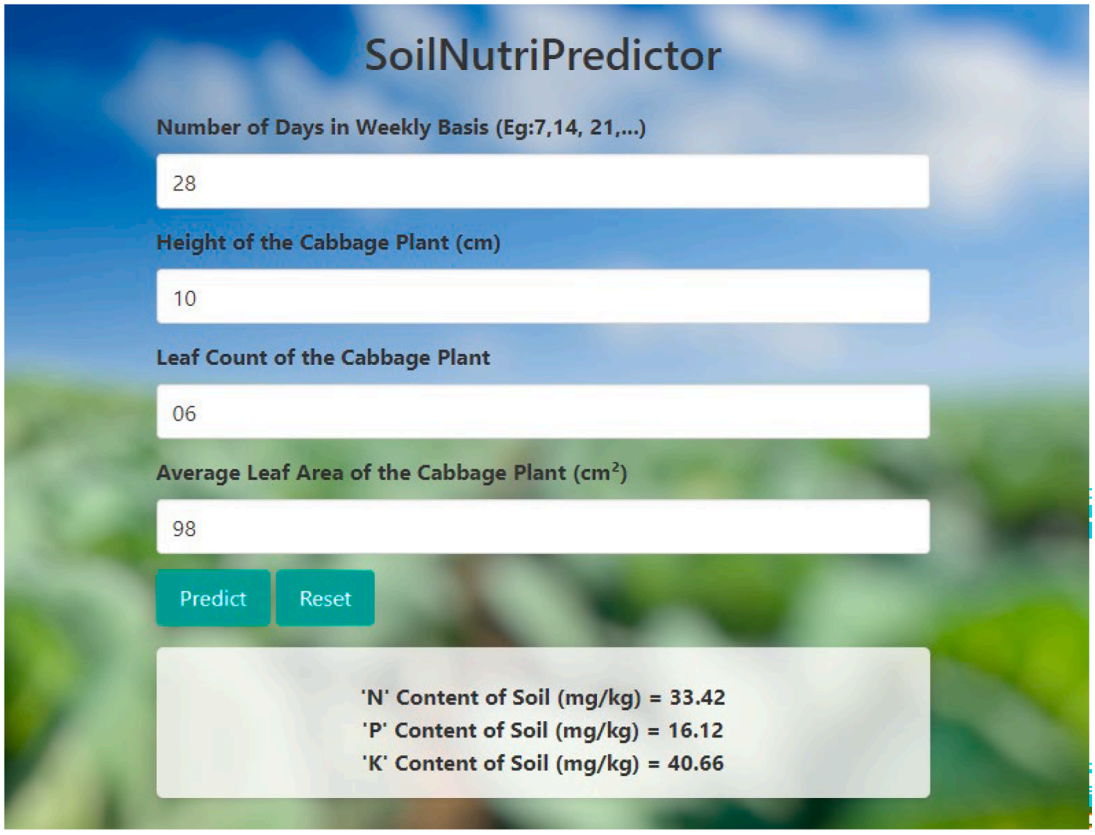

Fig. 12. 'SoilNutriPredictor' web application interface.

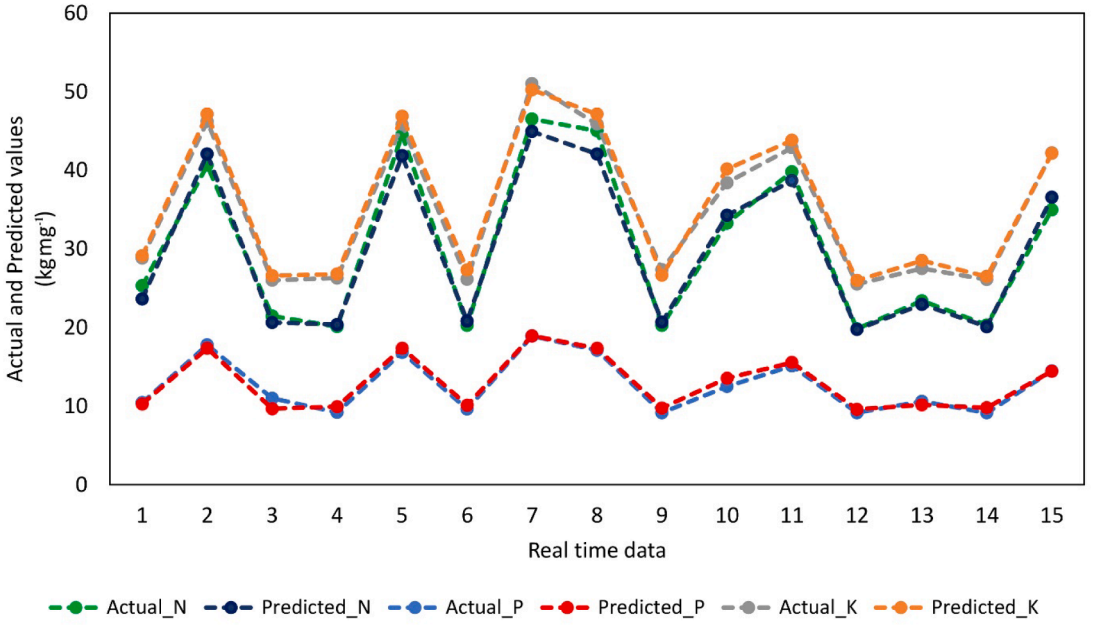

Fig. 13. Validation plot of the developed tool based on real-time data.

the model predictions. In this study, the prediction and elucidation of soil N, P, and K levels alongside cabbage plant growth characteristics were demonstrated. To achieve this, SHAP and LIME techniques were employed to provide insights into both the global dataset and specific instances, respectively. In the initial pre-evaluation of models, ANN was identified as the most suitable model for prediction, as it recorded lower MSE (Training=1.82, Testing=1.18) and higher R^2 (Training=0.96, Testing=0.97) values compared to SVR, KNN, and Decision Tree models. In the SHAP, certain features such as day and average leaf area were observed to exhibit a negative impact on the model's predictions of soil N, P, and K content, while others like the number of leaves and plant height showed a positive impact on the model. Notably, the impact of these features across different instances was highlighted by LIME analysis, offering valuable insights into the prediction process. To enhance

user accessibility and usability, a user-friendly interface application was developed. This application serves as a tool for individuals with varying levels of technical expertise, facilitating their utilization of the predictive model. By bridging the gap between complex machine learning algorithms and end-users, our research contributes to the democratization of predictive analytics in agriculture and beyond.

Integrating the predictive model with IoT technologies can facilitate real-time soil nutrient monitoring. By deploying IoT-based soil sensors capable of measuring parameters such as moisture content, pH levels, temperature, and nutrient concentrations, data can be collected continuously and transmitted wirelessly to a central system. Advanced algorithms can then analyze this data to provide actionable insights, enabling timely interventions and optimizing fertilization strategies.

The fusion of machine learning models with IoT devices represents a significant advancement in precision agriculture. This integration allows for the development of intelligent systems that not only monitor soil health in real-time but also predict future nutrient requirements based on continuous data analysis. Such systems can lead to more efficient resource utilization, reduced environmental impact, and enhanced crop yields.

CRediT authorship contribution statement

Thilina Abekoon: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Hirushan Sajindra:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Namal Rathnayake:** Writing – original draft, Software, Formal analysis. **Imesh U. Ekanayake:** Supervision, Software. **Anuradha Jayakody:** Resources. **Upaka Rathnayake:** Writing – review & editing, Validation, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research received no external funding.

Data availability

Data used in this paper may available on request for research purposes.

References

[1] H. Sajindra, T. Abekoon, J.A.D.C.A. Jayakody, U. Rathnayake, A novel deep learning model to predict the soil nutrient levels (N, P, and K) in cabbage cultivation, Smart Agric. Technol. 7 (2024) 100395, https://doi.org/10.1016/j.atech.2023.100395.

[2] S.A. Saganuwan, Samuel A Junior, A.S. Tsekohol, N. Wanmi, Acute toxicity study, haematological, biochemical and histopathological effects of brassica oleracea var. Capitata, allium porrum and cucurbita maxima in Rattus norvegicus, RPS Pharm. Pharmacol. Rep. 2 (2023), https://doi.org/10.1093/rpsppr/rqad025.

[3] A. Zayed, M. Sheashea, I a.A Kassem, M.A. Farag, Red and white cabbages: an updated comparative review of bioactives, extraction methods, processing practices, and health benefits, Crit. Rev. Food Sci. Nutr. 63 (2022) 7025–7042, https://doi.org/10.1080/10408398.2022.2040416.

[4] J.J. Uuh-Narvaez, M.R. Segura-Campos, Cabbage (Brassica oleracea var. capitata): a food with functional properties aimed to type 2 diabetes prevention and management, J. Food Sci. 86 (2021) 4775–4798, https://doi.org/10.1111/1750-3841.15939.

[5] Y. Xu, Y. Xiao, C. Lagnika, et al., A comparative evaluation of nutritional properties, antioxidant capacity and physical characteristics of cabbage (Brassica oleracea var. Capitate var L.) subjected to different drying methods, Food Chem. 309 (2020) 124935, https://doi.org/10.1016/j.foodchem.2019.06.002.

[6] S. Rana, K.S. Thakur, R.K. Bhardwaj, et al., Effect of biofertilizers and micronutrients on growth and quality attributes of cabbage (Brassica oleracea var. capitata L.), Int. J. Chem. Stud. 8 (2020) 1656–1660, https://doi.org/10.22271/chemi.2020.v8.i1x.8501.

[7] N. Moreb, A. Murphy, S. Jaiswal, A.K. Jaiswal, Cabbage. Elsevier Ebooks, 2020, pp. 33–54.

[8] Y.-H. Cho, S.-M. Hong, C.-H. Kim, Isolation and characterization of lactic acid bacteria from Kimchi, Korean traditional fermented food to apply into fermented dairy products, Korean J. Food Sci. Anim. Resour. 33 (2013) 75–82, https://doi.org/10.5851/kosfa.2013.33.1.75.

[9] A. Marasek-Ciolakowska, G. Soika, W. Warabieda, et al., Investigation on the relationship between morphological and anatomical characteristic of savoy cabbage and kale leaves and infestation by cabbage whitefly (Aleyrodes proletella L.), Agronomy 11 (2021) 275, https://doi.org/10.3390/agronomy11020275.

[10] K.F. Ling, The Food of Asia - featuring authentic recipes from Chefs (pdf), PDF Room (2002). https://pdfroom.com/books/the-food-of-asia-featuring-authentic-recipes-from-master-chefs-kong-foong-ling/bXgPXjWAgev. Accessed 3 January 2024.

[11] S.Z. Viña, D.F. Olivera, C.M. Marani, et al., Quality of Brussels sprouts (Brassica oleracea L. gemmifera DC) as affected by blanching method, J. Food Eng. 80 (2007) 218–225, https://doi.org/10.1016/j.jfoodeng.2006.02.049.

[12] P. Chohura, E. Kołota, Suitability of some nitrogen fertilizers for the cultivation of early cabbage, J. Elementol. (2014), https://doi.org/10.5601/jelem.2014.19.3.704.

[13] R.B. Chhetri, Effect of organic nutrients on the yield of cabbage, Bhutan J. Nat. Resour. Dev. 1 (2014) 32–36, https://doi.org/10.17102/cnr.2014.05.

[14] K.S. Karthika, P.S. Philip, S. Neenu, Brassicaceae Plants Response and Tolerance to Nutrient Deficiencies, Springer eBooks, 2020, pp. 337–362.

[15] S. Shahena, M. Rajan, V. Chandran, L. Mathew, Conventional Methods of Fertilizer Release, Elsevier eBooks, 2021, pp. 1–24.

[16] E.I. Moyin-Jesu, Use of different organic fertilizers on soil fertility improvement, growth and head yield parameters of cabbage (Brassica oleraceae L), Int. J. Recycl. Organic Waste Agric. 4 (2015) 291–298, https://doi.org/10.1007/s40093-015-0108-0.

[17] E.M.B.M. Karunathilake, A.T. Le, S. Heo, et al., The path to smart farming: innovations and opportunities in precision agriculture, Agriculture 13 (2023) 1593, https://doi.org/10.3390/agriculture13081593.

[18] M.M. Uttsha, A.K.M.N. Haque, T.T. Banna, et al., Enhancing agricultural automation through weather invariant soil parameter prediction using machine learning, Heliyon 10 (2024) e28626, https://doi.org/10.1016/j.heliyon.2024.e28626.

[19] W.Z. Taffese, K.A. Abegaz, Artificial intelligence for prediction of physical and mechanical properties of stabilized soil for affordable housing, Appl. Sci. 11 (2021) 7503, https://doi.org/10.3390/app11167503.

[20] A. Morales, F.J. Villalobos, Using machine learning for crop yield prediction in the past or the future, Front. Plant Sci. 14 (2023), https://doi.org/10.3389/fpls.2023.1128388.

[21] E. Elbasi, C. Zaki, A.E. Topcu, et al., Crop prediction model using machine learning algorithms, Appl. Sci. 13 (2023) 9288, https://doi.org/10.3390/app13169288.

[22] H. Sajindra, T. Abekoon, E.M. Wimalasiri, et al., An artificial neural network for predicting groundnut yield using climatic data, AgriEngineering 5 (2023) 1713–1736, https://doi.org/10.3390/agriengineering5040106.

[23] L. Zhang, L. Zhang, Artificial Intelligence for Remote sensing Data analysis: a review of challenges and opportunities, IEEE Geosci. Remote Sens. Magazine 10 (2022) 270–294, https://doi.org/10.1109/mgrs.2022.3145854.

[24] S. Ai, R. Guo, B. Liu, L. Ren, et al., A field study on the dynamic uptake and transfer of heavy metals in Chinese cabbage and radish in weak alkaline soils, Environ. Sci. Pollut. Res. Int. 23 (2016) 20719–20727, https://doi.org/10.1007/s11356-016-7277-x.

[25] Y.O. Choi, J. Lee, J.H. Sim, t al, Development of a modified model for predicting cabbage yield based on soil properties using GIS, J. Korean Soc. Survey. Geodesy Photogr. Cartogr. 40 (2022) 449–456, https://doi.org/10.7848/ksgpc.2022.40.5.449.

[26] I.U. Ekanayake, D.P.P. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). Case Studies in Construction Materials 16:e01059, 2022, https://doi.org/10.1016/j.cscm.2022.e01059.

[27] V. Hassija, V. Chamola, A. Mahapatra, et al., Interpreting black-box models: a review on explainable artificial intelligence, Cognit. Comput. 16 (2023) 45–74, https://doi.org/10.1007/s12559-023-10179-8.

[28] E.M. Kenny, C. Ford, M. Quinn, M.T. Keane, Explaining black-box classifiers using post-hoc explanations-by-example: the effect of explanations and error-rates in XAI user studies, Artif. Intell. 294 (2021) 103459, https://doi.org/10.1016/j.artint.2021.103459.

[29] M.L. Guindo, M.H. Kabir, R. Chen, et al., Chemometric approach based on explainable AI for rapid assessment of macronutrients in different organic fertilizers using fusion spectra, Molecules 28 (2023) 799, https://doi.org/10.3390/molecules28020799.

[30] D. Gwiazdowska, K. Juś, J. Jasnowska-Małecka, K. Kluczyńska, The impact of polyphenols on bifidobacterium growth, Acta Biochim. Pol. 62 (2015) 895–901, https://doi.org/10.18388/abp.2015_1154.

[31] R.R. Suman, R. Mall, S. Sukumaran, M. Satpathy, Extracting state models for Black-Box software components, J. Object Technol. 9 (2010) 79, https://doi.org/10.5381/jot.2010.9.3.a3.

[32] A. Adadi, M. Berrada, Peeking inside the Black-box: a survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018) 52138–52160, https://doi.org/10.1109/access.2018.2870052.

[33] U. Ehsan, Q.V. Liao, M. Muller, et al., Expanding explainability: towards social transparency in AI systems, CHI Conf. Human Factor Comput. Syst. (2021).
[34] A. Tocchetti, M. Brambilla, The role of human knowledge in explainable AI, Data 7 (2022) 93, https://doi.org/10.3390/data7070093.
[35] In: department of Census and Statistics. http://www.statistics.gov.lk/. Accessed 4 Jan 2024.
[36] L. Rajapaksha, D.C. Gunathilake, S. Pathirana, T. Fernando, Reducing post-harvest losses in fruits and vegetables for ensuring food security – Case of Sri Lanka, MOJ Food Process. Technol. 9 (2021) 7–16, https://doi.org/10.15406/mojfpt.2021.09.00255.
[37] W.A.H. Champa, K.B. Palipane, W.A.P. Weerakkody, M.D. Fernando, Maturity indices for harvesting of cabbage (Brassica oleracear L.) variety green coronet, Trop. Agric. Res. 19 (2007) 254–264.
[38] W a.P Weerakkody, S.M.M.R. Mawalagedera, Recent Developments in Vegetable Production Technologies in Sri Lanka, Springer eBooks, 2020, pp. 189–214.
[39] J.P. Marasinghe, S.H.P.P. Karunarane, S.N. Surendran, et al., Insecticide resistance, resistance mechanisms, and phylogeny of three myzus persicae populations in cabbage from three agroclimatic zones of Sri Lanka, Trop. Agric. Res. 34 (2023) 80–93, https://doi.org/10.4038/tar.v34i2.8621.
[40] O.E. Apolo-Apolo, M. Pérez-Ruiz, J. Martínez-Guanter, G. Egea, A mixed data-based deep neural network to estimate leaf area index in wheat breeding trials, Agronomy 10 (2020) 175, https://doi.org/10.3390/agronomy10020175.
[41] M. Yeshitila, Non-destructive prediction models for estimation of leaf area for most commonly grown vegetable crops in Ethiopia, Sci. J. Appl. Math. Stat. 4 (2016) 202, https://doi.org/10.11648/j.sjams.20160405.13.
[42] M. Sharifzadeh, A. Sikinioti-Lock, N. Shah, Machine-learning methods for integrated renewable power generation: a comparative study of artificial neural networks, support vector regression, and Gaussian Process regression, Renew. Sustain. Energy Rev. 108 (2019) 513–538, https://doi.org/10.1016/j.rser.2019.03.040.
[43] A. Ayodeji, Y.-K. Liu, SVR optimization with soft computing algorithms for incipient SGTR diagnosis, Ann. Nucl. Energy 121 (2018) 89–100, https://doi.org/10.1016/j.anucene.2018.07.011.
[44] H. Shafizadeh-Moghadam, A. Tayyebi, M. Ahmadlou, et al., Integration of genetic algorithm and multiple kernel support vector regression for modeling urban growth, Comput. Environ. Urban Syst. 65 (2017) 28–40, https://doi.org/10.1016/j.compenvurbsys.2017.04.011.
[45] N.J.-T. Jeng, Hybrid approach of selecting hyperparameters of support vector machine for regression, IEEE Trans. Syst. Man. Cybern. 36 (2006) 699–709, https://doi.org/10.1109/tsmcb.2005.861067.
[46] Z. Zhang, G. Gao, Y. Tian, J. Yue, Two-phase multi-kernel LP-SVR for feature sparsification and forecasting, Neurocomputing 214 (2016) 594–606, https://doi.org/10.1016/j.neucom.2016.06.049.
[47] M. Suyal, P. Goyal, A review on analysis of K-nearest neighbor classification machine learning algorithms based on supervised learning, Int. J. Eng. Trends Technol. 70 (2022) 43–48, https://doi.org/10.14445/22315381/ijett-v70i7p205.
[48] M. Steinbach, P.N. Tan, kNN: k-nearest neighbors. The Top Ten Algorithms in Data Mining, Chapman and HallCRC, 2009, pp. 165–176.
[49] H.A.A. Alfeilat, A.B.A. Hassanat, O. Lasassmeh, et al., Effects of distance measure choice on K-nearest neighbor classifier performance: a review, Big Data 7 (2019) 221–248, https://doi.org/10.1089/big.2018.0175.
[50] S.B. Imandoust, M. Bolandraftar, Application of k-nearest neighbor (knn) approach for predicting economic events. theoretical background, Int. J. Eng. Res. Appl. 3 (2013) 605–610.
[51] C. Strobl, A.-L. Boulesteix, T. Augustin, Unbiased split selection for classification trees based on the Gini Index, Comput. Stat. Data Anal. 52 (2007) 483–501, https://doi.org/10.1016/j.csda.2006.12.030.
[52] B. De Ville, Decision trees, Wiley Interdiscip. Rev. Comput. Stat. 5 (2013) 448–455, https://doi.org/10.1002/wics.1278.
[53] F. Esposito, D. Malerba, G. Semeraro, J. Kay, A comparative analysis of methods for pruning decision trees, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 476–493, https://doi.org/10.1109/34.589207.
[54] S. Agatonovic-Kustrin, R. Beresford, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research, J. Pharm. Biomed. Anal. 22 (2000) 717–727, https://doi.org/10.1016/s0731-7085(99)00272-1.
[55] A. Azadeh, A. Maghsoudi, S. Sohrabkhani, An integrated artificial neural networks approach for predicting global radiation, Energy Convers. Manage. 50 (2009) 1497–1505, https://doi.org/10.1016/j.enconman.2009.02.019.
[56] T. Abekoon, B.L.S.K. Buthpitiya, H. Sajindra, et al., A comprehensive review to evaluate the synergy of intelligent food packaging with modern food technology and artificial intelligence field, Discover Sustain. 5 (2024), https://doi.org/10.1007/s43621-024-00371-7.
[57] T. Abekoon, H. Sajindra, J. Jayakody, et al., Image processing techniques to identify tomato quality under market conditions, Smart Agric. Technol. 7 (2024) 100433, https://doi.org/10.1016/j.atech.2024.100433.
[58] I.I. Baskin, D. Winkler, I.V. Tetko, A renaissance of neural networks in drug discovery, Expert Opin. Drug Discov. 11 (2016) 785–795, https://doi.org/10.1080/17460441.2016.1201262.
[59] B. Pradhan, S. Lee, Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling, Environ. Modell. Softw. 25 (2010) 747–759, https://doi.org/10.1016/j.envsoft.2009.10.016.
[60] MdM Hasan, Understanding model predictions: a comparative analysis of SHAP and LIME on various ML algorithms, J. Sci. Technol. Res. 5 (2024) 17–26, https://doi.org/10.59738/jstr.v5i1.23(17-26).eaqr5800.
[61] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, Adv. Neural. Inf. Process. Syst. 30 (2017).
[62] P. Meddage, I. Ekanayake, U.S. Perera, Azamathulla HMd, M.A. Md Said, U. Rathnayake, Interpretation of machine-learning-based (black-box) wind pressure predictions for low-rise gable-roofed buildings using Shapley additive explanations (SHAP), Buildings 12 (2022) 734, https://doi.org/10.3390/buildings12060734.
[63] A.B. Parsa, A. Movahedi, H. Taghipour, et al., Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis, Accident Anal. Prevent. 136 (2020) 105405, https://doi.org/10.1016/j.aap.2019.105405.
[64] M. Yap, R.L. Johnston, H. Foley, S. MacDonald, O. Kondrashova, et al., Verifying explainability of a deep learning tissue classifier trained on RNA-seq data, Sci. Rep. 11 (2021) 1–2.
[65] Y. Liang, S. Li, C. Yan, et al., Explaining the black-box model: a survey of local interpretation methods for deep neural networks, Neurocomputing 419 (2021) 168–182, https://doi.org/10.1016/j.neucom.2020.08.011.
[66] C. Madhushani, K. Dananjaya, I.U. Ekanayake, et al., Modeling streamflow in non-gauged watersheds with sparse data considering physiographic, dynamic climate, and anthropogenic factors using explainable soft computing techniques, J. Hydrol. 631 (2024) 130846, https://doi.org/10.1016/j.jhydrol.2024.130846.
[67] A. Mamalakis, E.A. Barnes, I. Ebert-Uphoff, Carefully choose the baseline: lessons learned from applying XAI attribution methods for regression tasks in geoscience, Artif. Intell. Earth Syst. 2 (2023), https://doi.org/10.1175/aies-D-22-0058.1.
[68] H.T.T. Nguyen, H.Q. Cao, K.V.T. Nguyen, N.D.K. Pham, Evaluation of explainable artificial intelligence: SHAP, LIME, and CAM, Proc. FPT AI Conf. (2021) 1–6.
[69] M.T. Ribeiro, S. Singh, C. Guestrin, why should I trust you?, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, https://doi.org/10.1145/2939672.2939778.
[70] R. Wieland, T. Lakes, C. Nendel, Using Shapley additive explanations to interpret extreme gradient boosting predictions of grassland degradation in Xilingol, China Geosci. Model Dev. 14 (2021) 1493–1510, https://doi.org/10.5194/gmd-14-1493-2021.
[71] D. Zhang, A coefficient of determination for generalized linear models, Am. Stat. 71 (2017) 310–316, https://doi.org/10.1080/00031305.2016.1256839.
[72] M.R. Zafar, N. Khan, Deterministic local interpretable model-agnostic explanations for stable explainability, Mach. Learn. Know. Extraction 3 (2021) 525–541, https://doi.org/10.3390/make3030027.