Journal of Experimental Biology and Agricultural Sciences, August - 2023; Volume – 11(4) page 720 – 735

Journal of Experimental Biology and Agricultural Sciences

http://www.jebas.org

ISSN No. 2320 – 8694

Acid Rain and Seed Germination: A Predictive Model Using ML-based CART Algorithm

Vasundhara Arora[1], Bikram Jit Singh[2], Navneet Bithel[3], Tapan Kumar Mukherjee[4], Sushil Kumar Upadhyay[5], Rippin Sehgal[5], Raj Singh[5*]

[1]Department of Botany C.C.S. University, Meerut, UP, India
[2]Department of Mechanical Engineering, MM Engineering College, Maharishi Markandeshwar (Deemed to be University) Mullana-Ambala, 133207, Haryana, India
[3]Department of Botany and Microbiology, Gurukul Kangri University, Haridwar -249404, Uttrakhand, India
[4]Department of Biotechnology, Amity University, Major Arterial Road, Action Area II-36, 37, 38, Rajarhat, New Town, Kolkata,West Bengal 700156, India
[5]Department of Bio-Sciences and Technology, Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, 133207, Haryana, India

Received – May 23, 2023; Revision – August 05, 2023; Accepted – August 18, 2023
Available Online – August 31, 2023

DOI: http://dx.doi.org/10.18006/2023.11(4).720.735

KEYWORDS

Simulated Acid Rain

Seed Germination

Brinjal

Cowpea

Decision Tree Algorithm

Predictive Model

ABSTRACT

The impact of acid rain on the germination of seeds is a significant concern in agricultural and environmental studies. Acid rain, characterized by elevated acidity levels due to pollutants like sulfur dioxide and nitrogen oxides, can adversely affect the germination process of various plant species. The objective of this study was to evaluate the impact of simulated acid rain (SAR) on the germination of Brinjal (*Solanum melongena* Linn.) and Cowpea (*Vigna unguiculata* ssp. *cylindrica* L. Walpers) crops. The experiments were conducted using eight plastic trays of approximately 25 cm. x 30 cm dimensions. Four trays were used for experiments with brinjal seeds (Set I), while the other four were used for cowpea seeds (Set II). One tray from each set was used as positive control and treated with normal pH 5.6, while the other three trays from each batch were treated with SAR solutions of pH 4.5, 3.5, and 2.5. Brinjal seed germination percentage and seed vigor were inferior to Cowpea seeds. The seeds treated with SAR (pH 4.5, 3.5, and 2.5) showed hindered seed germination. Furthermore, a more significant inhibitory effect was observed at lower pH values. The mean germination percentage of seeds was highest for standard SAR (pH 5.6) in the case of Brinjal seeds, while it was recorded lowest for Cowpea seeds. The results indicate that plants do not respond uniformly to SAR. To investigate the behavior of the simulated acid rain data, a Machine Learning-based Decision Tree Algorithm was employed to identify and optimize conditions. Cowpea was predicted to get 95% seed germination, whereas brinjal would only

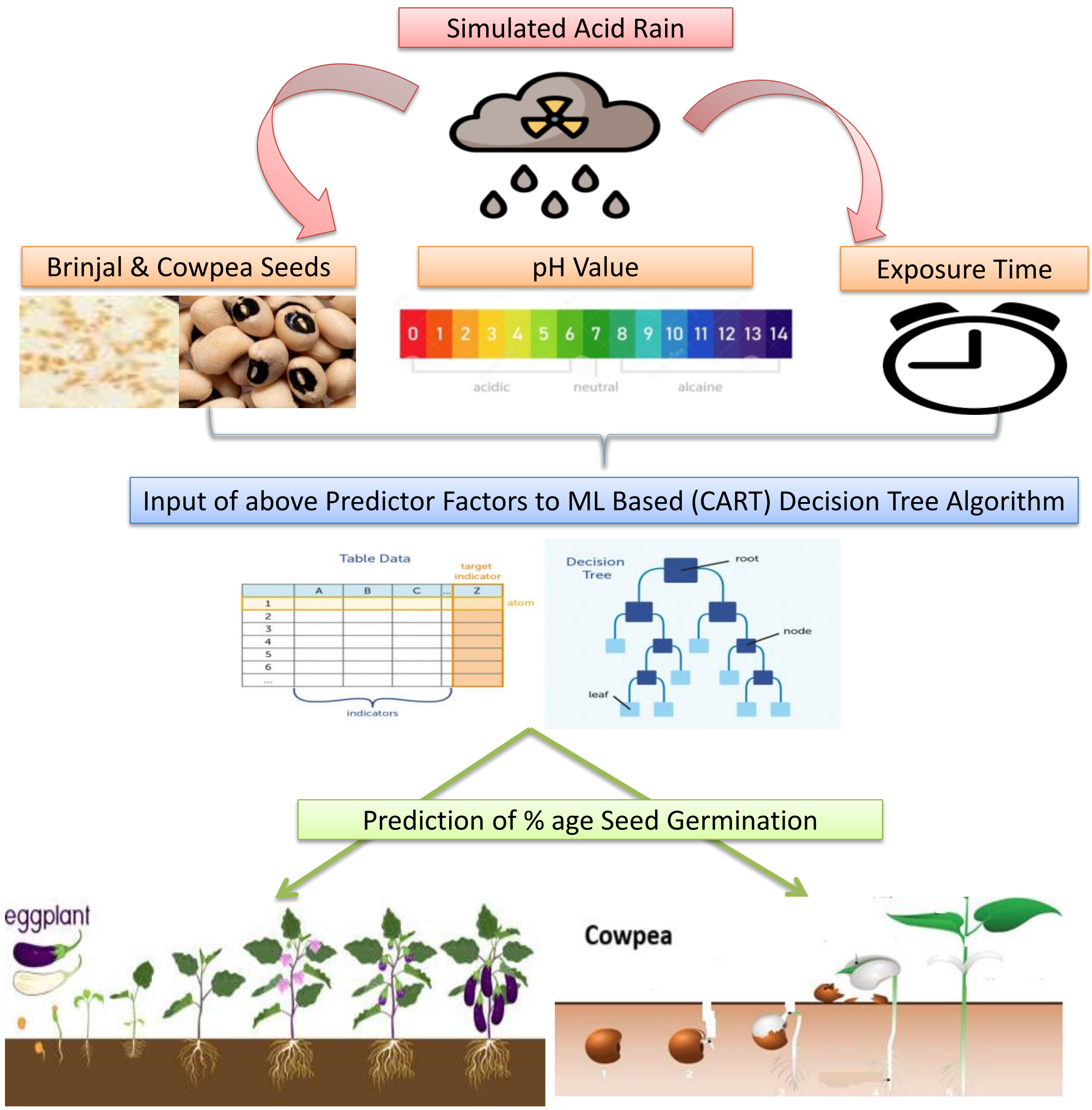
* Corresponding author
E-mail: dr.rajsingh09@gmail.com (Raj Singh)

Peer review under responsibility of Journal of Experimental Biology and Agricultural Sciences.

Production and Hosting by Horizon Publisher India [HPI] (http://www.horizonpublisherindia.in/).

be 64% in acid rain of pH value 5.05 for 36 hours. In conclusion, utilizing a Machine Learning-based CART algorithm has provided valuable insights into predicting the germination behavior of seeds under the influence of acid rain.

GRAPHICAL ABSTRACT



1 Introduction

Acid rain refers to precipitation that contains acidic compounds resulting from various sources. The term "acid deposition" is more accurate, including dry and wet deposition (Lee 1988; Burns et al. 2016). Pure rainwater has a pH of 5.6 as it contains dissolved carbon dioxide, but when the pH of rainwater falls below 5.6, it is referred to as acid rain (Park et al. 2015). In some cases, the pH of rain and fog water can drop as low as 2.2, as reported in Wheeling, West Virginia, and Pitlochry, Scotland (McCormick 2013). The increase in natural water and soil acidity is a global issue due to nitrogen and sulfur oxides resulting from atmospheric pollution

(Shi et al. 2021). The various gases react with water, oxygen, and other chemicals to produce different compounds causing acidity. Among these, H_2SO_4 primarily contributes to acidic precipitation (60-70%), followed by HNO_3 (30-40%) and HCl. Acid rain can also deteriorate monuments and building materials (Altae 2022).

While studying the impact of simulated acid rain on nutrient cycling and the ratios of Mg, Al, Ca, N, and P in tea plants in a subtropical plantation, Hu et al. (2019) found varied effects on plant growth. The levels of Mg and Ca decreased in the soil and within plant tissues with increasing acidity; in contrast, Al concentration increased. The study highlighted the complex interactions between acid rain and nutrient dynamics within the ecosystem and focused on the potential consequences on plant health and soil fertility. Depending on the acidity of precipitation and the plant species being studied, the leaf chlorophyll content was reduced by 6.71% per pH among deciduous species compared to evergreen species. This underscores the differential sensitivity of various plant types to changes in pH levels, emphasizing the need for tailored approaches to understanding and mitigating the impacts of acid rain on diverse ecosystems (Li et al. 2021; Du et al. 2017). Acid rain can cause various adverse effects, including large irregular lesions on leaf surfaces, hyperplasia, hypertrophy, and yield reduction (Arora et al. 2022). Ulrich et al. (1980) reported that changes in soil resulting from acid rain led to the death of fine roots. Some studies have shown the effects of acid rain on seed germination and seedling growth of different plant species, including Anitha and Ramanujam (1992), Verma and Prakash (1999), Tyagi (2006), Akinci and Akinci (2010), and Verma et al. (2010). A few studies have also reported the indirect effects of acid rain on plants through the soil microorganisms and abiotic environment in the soil rhizosphere. Acid deposition affects plant-associated microorganisms' distribution, composition, abundance, function, and activity. It can influence the dynamics of some substances in the soil in ways that may be detrimental to plants (Zhang et al. 2023).'

Some authors have examined the effect of pH and time on the germination of Brinjal and Cowpea seeds. For instance, Kumar et al. (2014) inferred that a pH range of 6-7 was optimal for germinating Brinjal seeds, while Sodhi and Singh (2013) discovered that an incubation period of 24-48 hours was ideal for germinating Cowpea seeds. In the present case, the algorithm's decision criteria for splitting nodes were adapted to access the particularities of brinjal and cowpea. It accounted for factors such as water pH levels and relevant time for each crop. CART can generate trees of varying complexity based on the data and relationships within each crop's ecosystem. The algorithm may discover unique patterns in brinjal and cowpea growth that aren't present in other crops, leading to different tree structures and depths. This enables the creation of models tailored to the intricacies of each crop, leading to more accurate predictions and actionable insights for optimizing their cultivation. Singh et al. (2018) and Singh and Sodhi (2015) developed a model for predicting Brinjal and Cowpea seed germination based on pH, time, and temperature, respectively. The incidence of acidic rain events has become more frequent in the Indo-Gangetic plains and Indian summer monsoon (Bisht et al. 2015; Majumdar et al. 2022). Industrialization of the National Capital Region in India and burning post-harvested crop residues have significantly contributed to this issue.

Therefore, this investigation evaluated the impact of simulated acid rain on the germination of Brinjal and Cowpea seeds, which are significant crops grown in Meerut, UP, India. This study emphasizes the significance of developing predictive models to forecast germination under different environmental conditions through CART algorithms, which are still to be explored. Gathered experimental data was analyzed appropriately through behavioral and predictive analysis for percentage germination. This study strategically regressed the simulated acid rain data for necessary behavioral investigation. It utilized a corresponding Machine Learning Decision Tree Algorithm to identify and optimize the conditions for properly germinating these plants. This analysis's findings can further help improve crop yield and productivity.

2 Materials and Methods

The seeds of Brinjal (*Solanum melongena* Linn.) and Cowpea (*Vigna unguiculata* L. Wapl.) were obtained from a National Seeds Corporation (NSC), Meerut, Uttar Pradesh, India. Pusa-1 and Gomati varieties of Brinjal and Cowpea were utilized, respectively.

2.1 Preparation of Plant Material

According to the Lee et al. (1980) methodology, approximately 25cm X 30cm eight plastic trays measuring were utilized for this study. The trays were filled with garden soil and kept moist for one week to maintain proper moisture levels. Four trays were designated for the Brinjal seed experiments (Set I), and the remaining four were used for the Cowpea seed experiments (Set II). Each tray of the experimental set I contained 20 seeds arranged in four rows of five seeds each. The seeds were spaced evenly apart. Before sowing, the seeds were sterilized with 0.1% $HgCl_2$ for two minutes and then rinsed thoroughly with distilled water to ensure good health. The trays in Set II were also planted analogously but with Cowpea seeds instead.

2.2 Preparation of Acid Solutions for SAR Experiments

A solution consisting of concentrated H_2SO_4 and HNO_3 (purchased from Merck) in a 7:3 (v/v) ratio was created to conduct the analysis. To obtain the aqueous solutions of predefined pH values (5.6, 4.5, 3.5, and 2.5), the concentrated stock solution was diluted with distilled water following the standard procedure of Lee et al.

(1980). The pH of the working acid solution for SAR experiments was then measured using a digital pH meter (EI-111).

2.3 Treatment of seeds with SAR

In each set, one tray was designated as the control, and the seeds were treated with distilled water, while the other four trays were treated with different pH values. For Set I (with Brinjal seeds), four trays were treated with predefined aquatic solutions of pH 5.6, 4.5, 3.5, and 2.5, respectively. A similar treatment was carried forward for Set II (with Cowpea seeds), respectively.

2.4 The Germination Response of Seeds to SAR

The seeds typically began to sprout within a day, and the emergence of a 1-2mm root was considered a successful germination. The number of germinated seeds was recorded daily until no further germination occurred. The seedlings were given treatments every alternate day, alternating between distilled water and the required solution. This procedure was conducted for six days. The germination percentage, mean germination frequency and seed vigor were calculated using the collected data.

$$\text{Germination percentage} = \frac{\text{Number of seed germinated}}{\text{Total no.of seeds}} \text{ X } 100$$

Mean germination frequency =

$$\frac{\text{Naximum no.of seed germinated}}{\text{Minimum period in which maximum germinatedis achivedv}} \text{ X } 100$$

It represents the rate of germination

The following formula determined seed vigour (an index of seed germination).

$$\text{Seed vigour} = \frac{\sum \text{Quotient of daily counts}}{\text{No.of days of germination}}$$

2.5 Data Analysis Techniques

The data obtained from the seed germination experiment was analyzed using various methods to gain meaningful insights. Multi-regression analysis was used to determine the effect of pH and time on the seed germination percentages for Brinjal and Cowpea plants. Main-effect plots were created to show the variation in germination percentage by considering one factor at a time. Regression statistics were generated to ensure the accuracy of the mode-fitment, and statistical equations were modelled to describe the overall germination behaviour. A heat map was used to illustrate necessary trends and effects.

A Machine Learning (ML) based Classification and Regression Trees (CART) algorithm was used to predict the germination percentage for predictive analysis. This algorithm was used to identify the optimal pH and time conditions for maximum seed germination of Brinjal and Cowpea plants. The CART algorithm selects a split point for each feature, using a metric such as Gini impurity or information gain to determine the best split. It can handle categorical and continuous input features, missing values, binary and multi-class classification tasks, and regression tasks. The algorithm's output was validated by plotting diagonal scatter plots for actual versus predicted values. Using these tools and techniques will provide a better understanding of the factors affecting seed germination behavior and help to determine optimal conditions for seed germination in agricultural practices.

3 Results

Compared to Cowpea seeds, Brinjal seeds' germination percentage and seed vigor were poorer. The use of SAR (pH 4.5) treatment had a significant inhibitory effect on the germination of both varieties of seeds. The treatment of seeds with higher pH SAR (pH 3.5 and 2.5) also reduced seed germination, with a more significant inhibitory effect observed with decreasing pH levels. In the case of brinjal, the mean germination percentage was highest with standard SAR treatment (pH 5.6) but was lowest for cowpea. A decrease in seed vigour was regularly observed for both plants as the pH decreased. As shown in Table 1, maximum seed vigour was observed in the control (pH 5.6) for both species.

3.1 Findings

The data collected was analyzed in two phases to evaluate the results comprehensively. The first phase focused on comprehending and normalizing the variations in germination related to the independent variables by conducting an appropriate "Behavioral Analysis." In the second phase, "Predictive Analysis" was performed using a Machine Learning (ML)-based Tree Algorithm to make necessary predictions about germination.

3.1.1 Phase-I: Behavioural Analysis

Both inferential and descriptive analyses are important for comprehensively understanding and analyzing data. Descriptive statistics can provide valuable insights into a dataset, such as the results of an experiment that examined the effect of pH, time, and their interaction on the germination percentage of Brinjal and Cowpea plants (Table 1). In this experiment, the predictors or independent variables were the pH levels and time duration, while the dependent variable or response was the germination percentage. The experiment used pH levels of 2.5, 3.5, 4.5, and 5.6 and time durations of 24, 48, 72, 96, 120, and 144 hours (Figure 1).

For brinjal, the germination percentage ranged from 0% to 30%, and for cowpea, it ranged from 10% to 100%. The data in the plot shows that as the pH level and time duration increased, the germination percentage also increased for both the selected plants. Cowpea had a germination percentage of 100% at pH 5.6 and a duration of 120 and 144 hours, while brinjal had a germination

Table 1 Effect of treatment with simulated acid rains (SAR) on germination percentage, mean germination frequency and seed vigour of Brinjal seeds and Cowpea seeds

Responses	Hours	pH							
		Brinjal seeds				Cowpea seeds			
		5.6*	4.5	3.5	2.5	5.6*	4.5	3.5	2.5
Seed Germination in Percentage (%)	24	5.00	Nil	Nil	Nil	40.00	25.00	20.00	10.00
	48	30.00	20.00	10.00	10.00	90.00	50.00	45.00	45.00
	72	30.00	20.00	10.00	10.00	90.00	70.00	70.00	70.00
	96	30.00	20.00	15.00	10.00	95.00	70.00	70.00	70.00
	120	30.00	20.00	15.00	10.00	100.00	70.00	65.00(R)	65.00(R)
	144	30.00	20.00	15.00	10.00	100.00	70.00	65.00(R)	65.00(R)
Mean Germination Frequency		25.00	8.33	3.12	4.16	16.66	19.44	19.44	19.44
Seed Vigor		5.16	3.33	2.16	1.66	15.50	11.83	11.16	11.83

* Treatment with a solution of pH 5.6 is regarded as a control in this experiment; R = Rotting of seedling, and 'Nil' represented no seed germination.

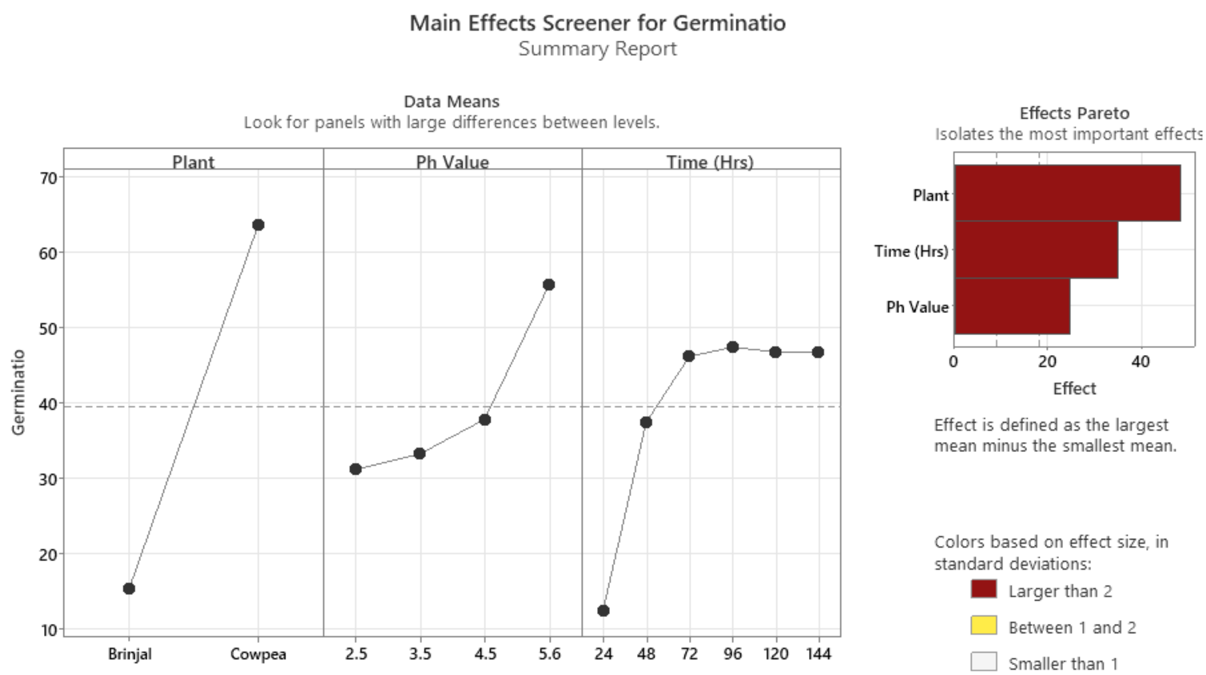


Figure 1 Main Effect Plots for Germination

percentage of only 30% under similar conditions. These data revealed that cowpea had a higher germination percentage than brinjal, which suggested that cowpea was more tolerant to changes in pH levels and time duration. Further, on the right side of the plot, the Pareto chart examines the relative impact of each predictive factor on response (i.e. germination percentage). A Huge difference (more than 40%) was observed due to Plant Type (Cowpea surpassed the brinjal), followed by Time duration (more than 30%) and ended with pH value (more than 20% difference in germination percentage). Overall, this plot delineates the importance of pH level and time duration on the germination percentage of plants (as observed from the sample data collected) and highlights the differences in germination behavior between different plant species.

Descriptive analysis has some limitations, such as it cannot be used to generalize findings beyond the collected dataset and does not provide a basis for making predictions or drawing inferences about

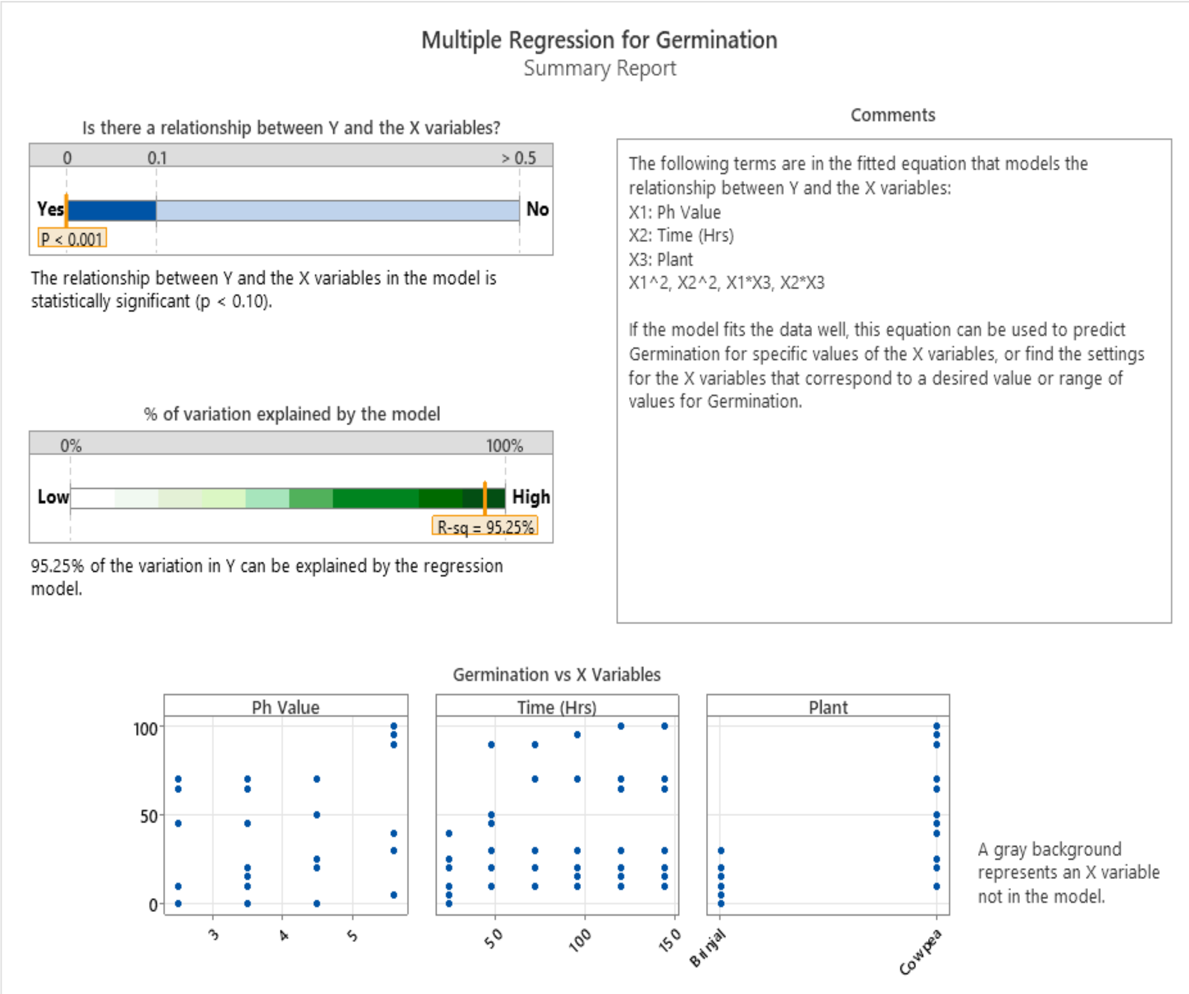


Figure 2 Summary Report for Percentage Germination

the population from which the data was collected. On the other hand, inferential statistics can help to determine the significance of differences or relationships observed in a dataset and can be used to make predictions and draw conclusions about a population. Cowpea was more sensitive to acid rain than brinjal, while percentage germination increased in both crops with increased pH value and time. However, it enhanced abruptly when pH varied from 4.5 to 5.6 and time raised to 72 hours, respectively.

This regression output from Minitab statistical software presents a regression model's coefficients and statistical information (Figure 2). The coefficient for pH Value is -21, which means that for every unit increase in the pH value, the predicted value of the response variable (percentage germination) decreases by 21 units, holding all other variables constant. This coefficient is statistically significant at the 0.05 significance level (p-value = 0.047). The coefficient for Time (Hrs.) is 1.086, which suggests that for every one-hour increase in time, the predicted value of the response variable increases by 1.086 units, holding all other variables constant. This coefficient is also statistically significant (p-value = 0). The regression model incorporates a categorical predictor variable named "Plant," encompassing two levels: Brinjal (used as the reference level) and Cowpea. The coefficient for the Cowpea level of the Plant variable is 48.33, which means that the predicted value of the response variable for cowpea is higher than brinjal by 48.33 units, holding all other variables constant.

The model includes interaction terms between pH Value and Time (Hrs.). The coefficient for the pH Value*pH Value interaction term is 3.53, which means that the effect of pH Value on the response variable changes depending on the level of pH Value. This coefficient is statistically significant (p-value = 0.007). The coefficient for the Time (Hrs.)*Time (Hrs.) interaction is 0.005038, which means that the effect of Time (Hrs.) on the response variable changes depending on the level of Time (Hrs.). This coefficient is also statistically significant (p-value = 0). Overall, all the Xs and their squares substantially relate to the given Y (as p-value < 0.001). The R-squared (R-sq) value is

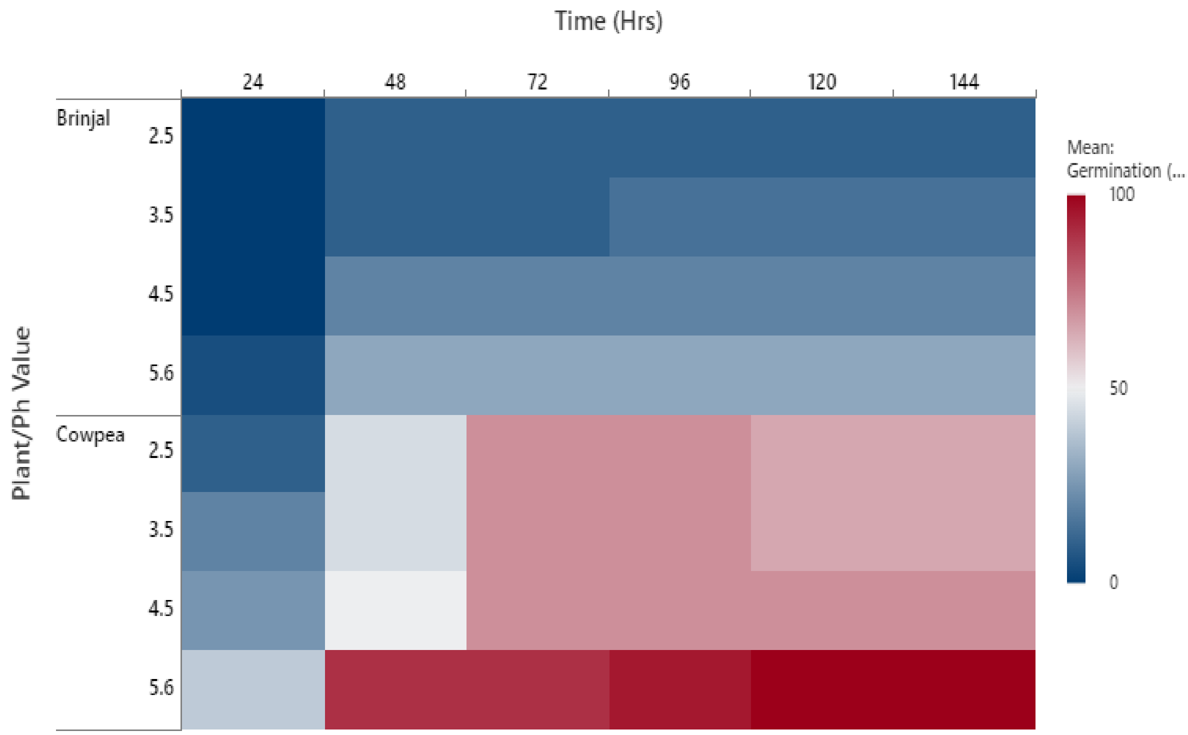


Figure 3 Heat Map for Percentage Germination

92.25%, indicating that the predictor variables can explain 92% of the variation in the response variable. In brief, the high R-squared and adjusted R-squared values indicate that the predictor variables in the model are good at explaining the variation in the response variable, and the model has an excellent fit to the data. Necessary variations of all the independent variables concerning the dependent one have been plotted in Figure 2.

For more precise insights, a heat map has been delineated between predictors (pH level and Duration) and response (Germination in percentage) with 95% confidence (Figure 3). These maps are graphical representations of data that use colors to indicate the magnitude of a variable.

According to Borner and Chen (2018), heat maps are a valuable tool for identifying patterns and trends in large data sets, as they allow for easy identification of areas of high or low response values. In a heat map, the intensity of the color corresponds to the value of the variable being represented. In a heat map of germination data, areas with a high percentage of germination would be represented by intense colors (red), while cooler areas would be represented by less intense colors (blue). According to Schiffer and Ha (2017), using color in heat maps can make it easier for users to identify patterns and trends in the data. Results presented in Figure 3 revealed that in cowpeas with a pH value of 5.6, maximum germination would be found in the time range from 48 hrs to 144 hrs, while the brinjal has only 50% germination at these predictor settings. The colour scale is used for percentage germination (shown on the left side of the map) for necessary interpretation. The decrease in pH will decrease germination capability in both plants. Time after around 72 hrs did not have much effect on brinjal, whereas Cow pea shows a continuous increase in germination up to 144 hrs. The two-dimensional map region is suitably divided among assorted colour shades (highlighting various levels of percentage germination) for corresponding predictor values.

To generalize, necessary statistical equations generated by Minitab at the end have been used (Figure 4). Two independent quadratic equations for predicting germination behaviour (w.r.t each plant) have been regressed with 95% confidence.

These equations have quantified the behavior of percentage germinations for each plant w.r.t. pH and time suitably and will provide needed information to practitioners or bio-scientists for future research. In the case of brinjal, percentage germination was directly proportional to pH value and its square, whereas cowpea showed inversely proportional relation with pH value but directly proportional to time and its quadratic term. The square of pH value also directly highlighted its relation with germination, which made it more sensitive than brinjal. Regression models allow us to understand the relationships between the predictor and response variables. This can provide insights into the factors that influence the response variable (Germination) and help identify areas for improvement.

X1: Ph Value X2: Time (Hrs) X3: Plant

Plant	Final Equations
Brinjal	Germination (%) = $9.6 - 22.97\, X1 + 0.956\, X2 + 3.534\, X1^2 - 0.005038\, X2^2$
Cowpea	Germination (%) = $20.1 - 18.98\, X1 + 1.215\, X2 + 3.534\, X1^2 - 0.005038\, X2^2$

Figure 4 Modelled Equations for Brinjal and Cowpea Plants

3.1.2 Phase-2: Predictive Analysis

Predictive analysis through machine learning (ML) based Tree Algorithms has become popular in various fields, including biosciences, biomaterials, finance, healthcare, and marketing (Qian et al. 2020). Tree Algorithms, such as Decision Trees and Random Forests, can handle categorical and continuous variables, making them highly versatile and effective for predicting outcomes (Yin et al. 2019). The ML-based Decision Tree Algorithm has been applied in the present case for more accurate and precise predictions for percentage germination. It will help predict germination percentage in plants exposed to environmental stressors such as acid rain. The model will utilize under-study predictors such as pH value, time of exposure and plant species to provide accurate predictions. Using such models can assist farmers and scientists in understanding the impacts of acid rain on plant growth and provide insights into potential mitigation strategies.

Table 2 provides information related to the Classification and Regression Trees (CART) Algorithm to predict the germination percentage of plants based on their pH value, time (hours), and type of plant. The CART® method uses node splitting to divide the data into smaller groups based on the independent variables. The splitting is done so that the variability within the groups is minimized. The method uses the least squared error to select the best split at each node. The breakup of that results in the lowest sum of squared errors is chosen. The analysis identifies the optimal tree that best fits the data.

Table 3 encapsulates the response (Percentage germination) for the given dataset. The percentage germination has a mean of 39.5%, a standard deviation of 30.35, and ranges from 0 to 100, respectively. The first quartile (Q1) is 11.25, the median is 30, and the third quartile (Q3) is 70 for germination. This table provides valuable information on the response variable's central tendency, variability, and range. The mean and median values give an idea of the typical value of the response variable, while the standard deviation indicates the degree of variation in the data. The quartile values provide information on the data's spread and help identify outliers.

In conclusion, this information can help understand the distribution of the percentage germination variable and select appropriate statistical methods for predicting the data suitably. Next, a line plot (Figure 5) has been drawn with R-squared values (in percentage) and the Number of Terminal Nodes (of a corresponding Tree Diagram).

The Minitab provides different tree structures for appropriate decision-making at different Number of Nodes, but the best fit will be the one having maximum R-squared value with minimum nodes (complexity). In the current study, the 4-Node Tree Diagram has been shortlisted for further predictions, as it will provide around 81.9% R-squared value with a less complex tree structure.

The Optimal Tree Diagram for percentage germination has been delineated through Minitab, as it is a helpful tool for decision-making since it allows the visualization of the decision tree model and helps identify the critical variables and their relationships to the outcome variable (Figure 6). The diagram provides a clear and intuitive representation of the decision-making process and helps identify the optimal path (based on the available data). The decision

Table 2 4 Node CART Regression Settings for Percentage Germination

Node splitting	Least squared error
Optimal tree	Within 1 standard error of maximum R-squared
Model validation	10-fold cross-validation
Rows used	48

Table 3 Response Information

Mean	StDev	Minimum	Q1	Median	Q3	Maximum
39.5833	30.3496	0	11.25	30	70	100

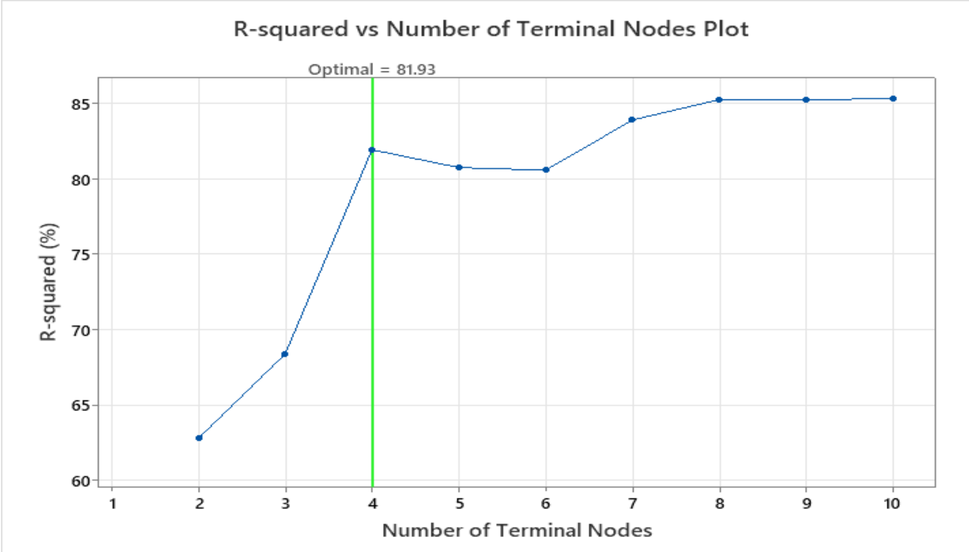


Figure 5 Line Plot to Select Optimal Decision Tree

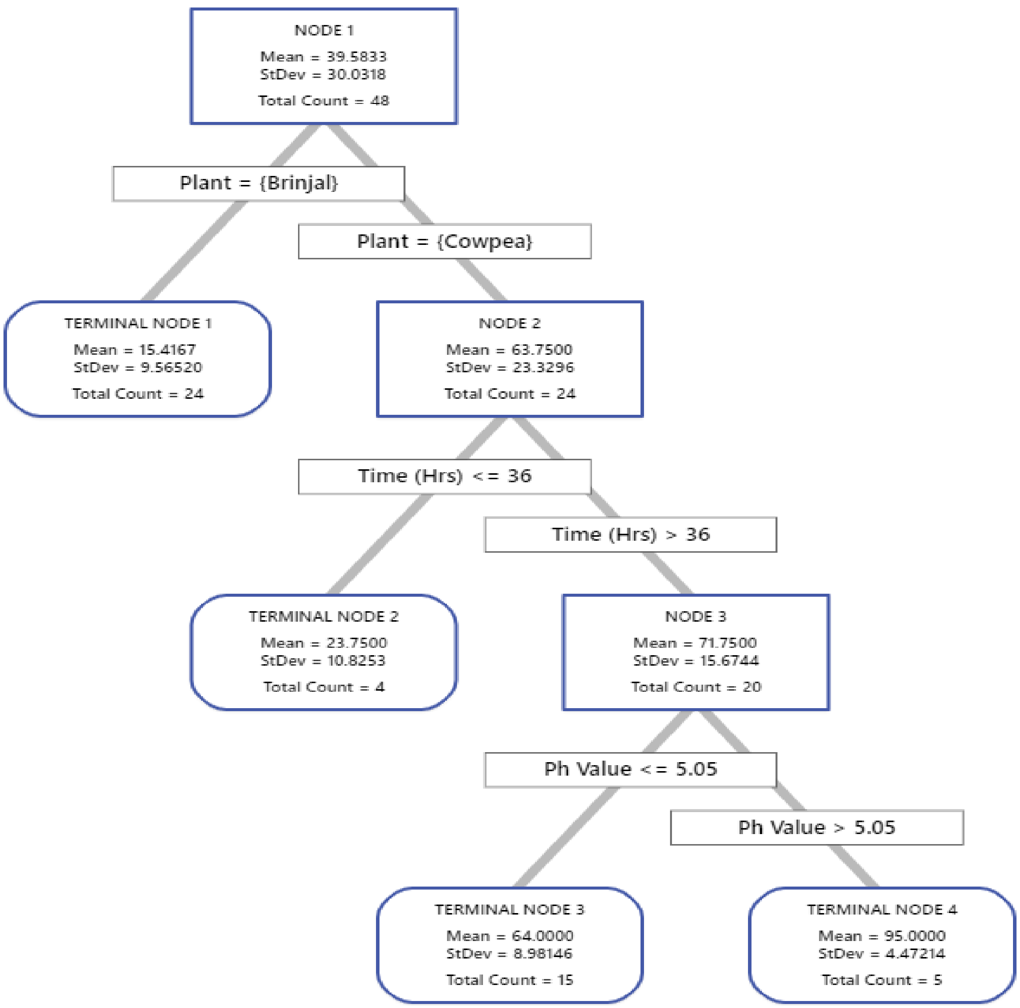


Figure 6 Optimal Decision Tree

Table 4 Model Summary While Training and Testing

Total predictors	3	
Important predictors	3	
Number of terminal nodes	4	
Minimum terminal node size	4	
Statistics	**Training**	**Test**
R-squared	90.82%	81.93%
Root mean squared error (RMSE)	9.0997	12.7670
Mean squared error (MSE)	82.8038	162.9955
Mean absolute deviation (MAD)	7.3056	8.8456
Mean absolute percent error (MAPE)	0.3030	0.3352

tree algorithm is a machine learning technique that recursively partitions data based on feature values, creating a tree-like structure to make decisions or predictions. Several studies have demonstrated the usefulness of Optimal Tree Diagrams in decision-making (De Ste Croix et al. 2016). The optimal tree diagram aims to predict the percentage germination of Brinjal and Cowpea plants under different conditions of pH value and time (in hours) using a dataset of 48 observations. The tree starts at Node-1, which represents the mean and standard deviation of the response variable (percentage germination) for the entire dataset. The mean percentage germination for the dataset is 39.58%, and the standard deviation is 30.03. The tree splits into two branches based on the type of plant, i.e. Brinjal and Cowpea.

Terminal Node-1 represents the mean and standard deviation of the response variable for the subset of observations that correspond to Brinjal plants. The mean percentage germination for Brinjal plants is 15.41, and the standard deviation is 9.56. Node-2 represents the mean and standard deviation of the response variable for the subset of observations that correspond to Cowpea plants. The mean percentage germination for Cowpea plants is also 15.41, and the standard deviation is 9.56. Further, The tree splits the Cowpea branch into two branches based on the time (in hours) variable. Terminal Node-2 represents the mean and standard deviation of the response variable for the subset of observations corresponding to Cowpea plants exposed to less than or equal to 36 hours of simulated acid rain. The mean percentage germination for these plants is 23.75, and the standard deviation is 10.82. The total count of observations in this terminal node is 4. Node-3 represents the mean and standard deviation of the response variable for the subset of observations corresponding to Cowpea plants exposed to more than 36 hours of simulated acid rain. The mean percentage germination for these plants is 71.7, and the standard deviation is 15.67. The total count of observations in this node is 20.

Finally, Node-3 is split into two terminal nodes based on the pH value variable. Terminal Node-3 represents the mean and standard deviation of the response variable for the subset of observations corresponding to Cowpea plants that were exposed to more than 36 hours of simulated acid rain and had a pH value of less than or equal to 5.05. The mean percentage germination for these plants is 64, and the standard deviation is 8.98. The total count of observations in this terminal node is 15. Terminal Node-4 represents the mean and standard deviation of the response variable for the subset of observations corresponding to Cowpea plants exposed to more than 36 hours of simulated acid rain and with a pH value of more than 5.05. The mean percentage germination for cowpeas is 95, and the standard deviation is 4.472. The total count of observations in this terminal node is 5. This decision tree can help predict Brinjal and Cowpea plants' percentage germination under different pH values and time conditions. It can also be used to identify which factors are most important in predicting germination, as the tree splits are based on the importance of these variables.

Further, Table 4 quoted the model summary for a decision tree model with three important predictors and four terminal nodes. The model has been evaluated on both training and test data. The R-squared value for the training set is 90.82%, indicating that the model explains a substantial portion of the variance in the data. The R-squared value for the test set is slightly lower at 81.93%, indicating that the model may have some degree of overfitting. The root mean squared error (RMSE) is 9.0997 for the training set and 12.7670 for the test set. The lower RMSE value for the training set suggests that the model fits the training data better than the test data.

The mean squared error (MSE) is 82.8038 for the training set and 162.9955 for the test set. This indicates that the model has a higher error level on the test set than the training set. The mean absolute deviation (MAD) is 7.3056 for the training set and 8.8456 for the test set. The mean absolute percent error (MAPE) is 0.3030 for the training set and 0.3352 for the test set. These values measure the model's accuracy in predicting the response variable. In brief, the

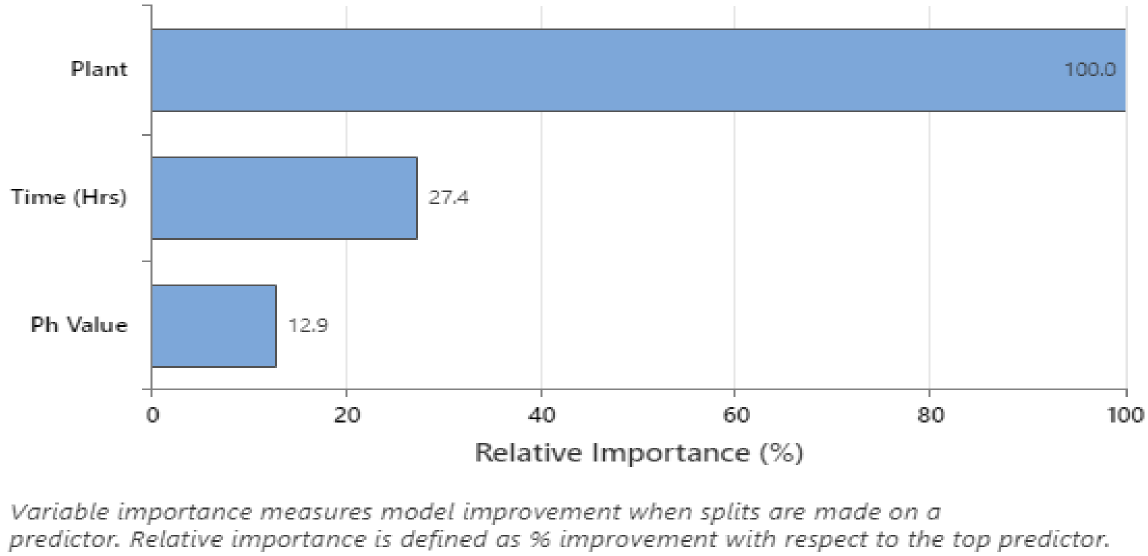


Figure 7 Relative Significance of Predictors

model seems to perform well on the training set but may have some degree of overfitting (which can be ignored). The performance on the test set is slightly lower, indicating that the model may not generalize well to new data, but it is insufficient to add any substantial error in the final prediction.

Additionally, the relative importance of predictors is determined based on their contribution to the model's predictive accuracy. It helps to identify the most important predictors that significantly impact the response variable and can be used to improve the model's accuracy. The horizontal bar graph illustrated in Figure 7 uncovered that the top predictor is 'Plant' with a relative importance of 100%, which means that it has the highest impact on the response variable compared to other predictors.

The relative importance of 'pH value' is 12.9%, and the relative importance of 'Time (Hrs.)' is 27.4% respectively. This suggests that both 'pH Value' and 'Time (Hrs.)' also have some influence on the response variable **(**percentage germination), but they are not as significant as the 'Plant' predictor.

In the CART (Classification and Regression Tree) algorithm, the scatter plots of response fits versus actual values are used to evaluate the performance of the model (Figure 8). These plots

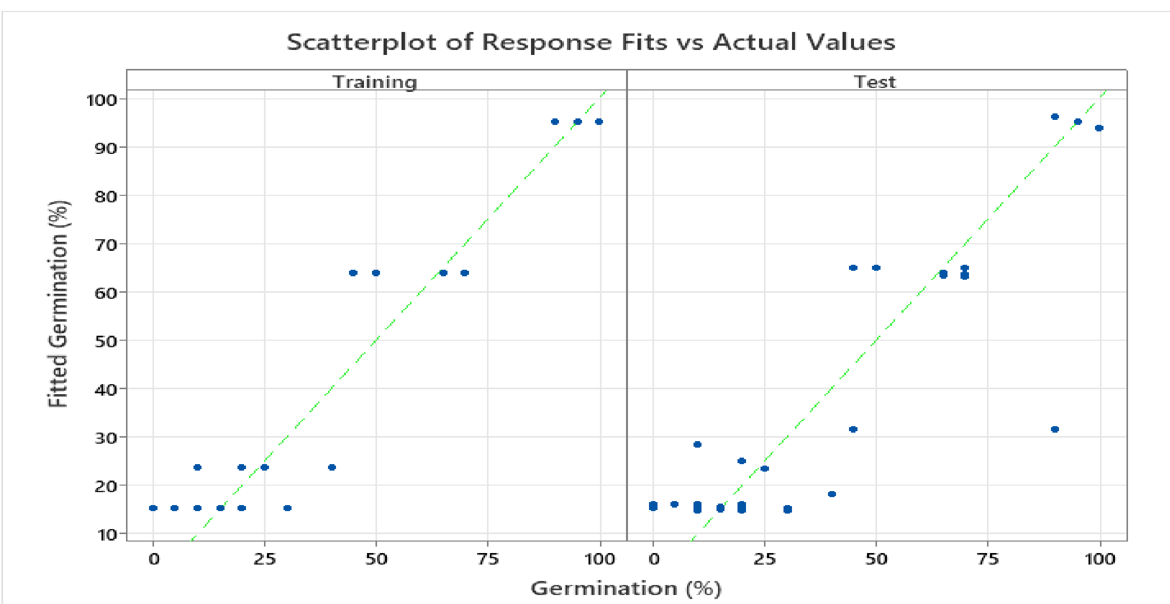


Figure 8 Scatter Plots for Testing and Training

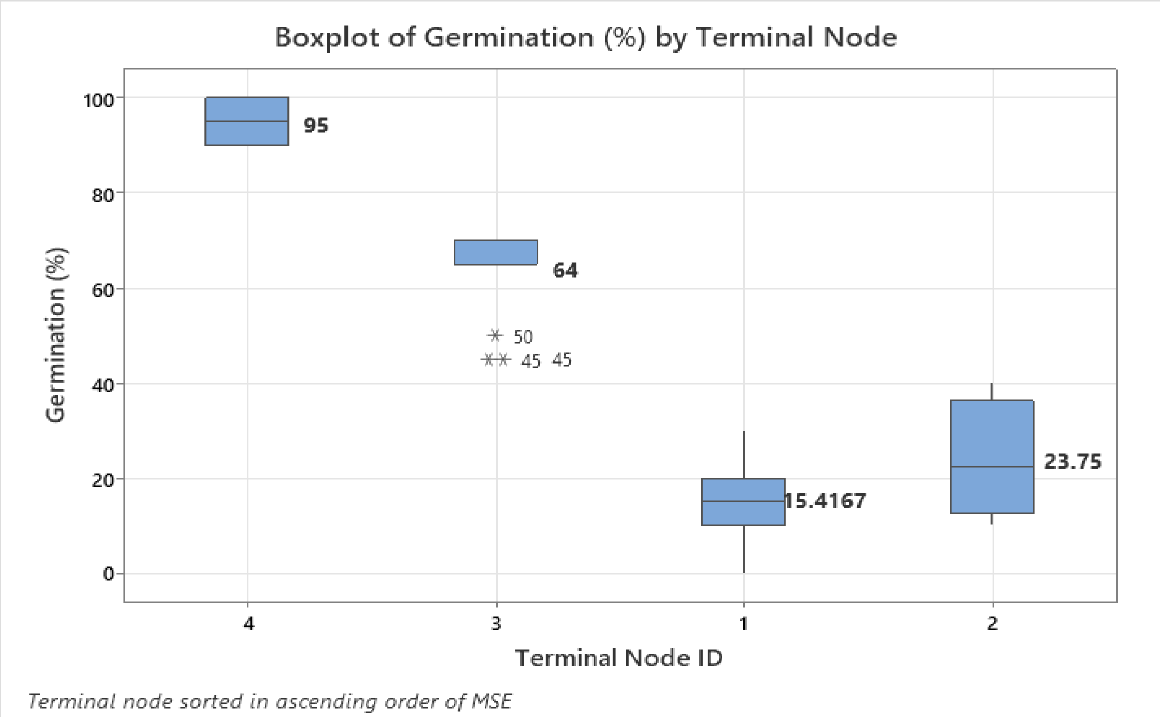


Figure 9 Boxplot for Percentage Germination w.r.t. Nodes

pointed out the relationship between the predicted response (% age germination) values and the actual response values, while model building (testing) and actual execution (training), respectively. In a scatter plot, the predicted response values are plotted on the x-axis, and the actual response values are plotted on the y-axis. In the present scenario, the points on the scatter plot fall on or around the diagonal line, which indicates a sufficient match between the predicted and actual response values.

The scatter plot allows us to visually examine the pattern of errors and identify any systematic deviations or outliers. These scatter plots are a valuable tool for evaluating the accuracy of the CART model and identifying areas for further improvement.

The box plot in Figure 9 illustrates the mean germination percentage and the maximum and minimum feasible germination percentages for four different nodes. The nodes represent various combinations of the independent variables, essential germination predictors (Figure 6). Node-1 has a mean germination percentage of 15.42%, less than the minimum feasible germination percentage of 30%. This suggests that the conditions represented by Node-1 are not favorable for germination. Similarly, Node-2 has a mean germination percentage of 23.75%, within the feasible range of 10-40%. This suggests that the conditions represented by Node-2 are favorable for germination but not optimal. Similarly, Node-3 has a mean germination percentage of 64%, within the feasible range of 45-70%. This suggests that the conditions represented by Node-3 are quite favorable for germination. Lastly, Node-4 has a mean germination percentage of 95%, close to the maximum possible germination percentage of 100%. This suggests that the conditions represented by Node-4 are optimal for germination.

In concisely, the analysis of these nodes suggests that the independent variables used in this study significantly impact germination and that specific combinations of these variables are more favorable for germination than others.

Finally, the study results provide predictive data on the germination of Brinjal and Cowpea plants under different pH values, time intervals, node numbers and actual germination percentages. For Brinjal plants, the pH values range from 2.5 to 5.6, the time intervals range from 24 to 144 hours, and the node number is constant at 1. The actual germination percentages range from 0.0% to 30.0%, while the predicted percentage is constant at 15.4%. Cowpea plants' pH values also range from 2.5 to 5.6, the time intervals range from 24 to 144 hours, and the node number ranges from 2 to 4. The actual germination percentages range from 10.0% to 100.0%, while the predicted germination percentage is constant at 95.0% for node number 4 and 64.0% for node numbers 2 and 3. Overall, the table suggests that pH values and time intervals substantially impact the germination of Brinjal and Cowpea plants. However, the predicted germination percentage is constant across different pH values, time intervals, and node numbers, suggesting that other factors may play a role in determining the actual germination percentage.

4 Discussion

The present study aimed to investigate the effects of pH and time on the Brinjal and Cowpea plant's germination percentage. The study utilized both behavioral analyses by regression and predictive analysis by ML-based Decision Tree Algorithm to understand the relationship between the variables and their impact on germination. The behavioral analysis showed that pH and time significantly impacted the germination percentage of both Brinjal and Cowpea. The regression analysis showed that the R-squared values for both plants were significant, indicating that the model could explain considerable variation in the data. The results indicated that pH positively affected the germination percentage of both plants, with a pH range of 5-7 being the most favorable for germination (Smith et al. 2018). Moreover, the time variable also positively impacted the germination percentage of both plants. The results showed that the germination percentage increased with time up to a certain limit, after which it decreased. This was agreed with previous studies that reported time's effect on seed germination (Kaur et al. 2016).

In the present study, pH 5.6 was taken as the reference point. Differing viewpoints exist concerning the pH value selected as the reference (control) for investigating the impact of SAR on plants. While Sequeira (1982) and Charlson and Rhode (1982) express reservations about the suitability of pH 5.6 as the baseline reference, several researchers in previous studies have adopted pH 5.6 as the established control point. Like any other pollutant, SAR also may affect plants at any stage of development. Since seed germination is the first step in the life cycle of a seed plant, it was decided to study the effects of acid precipitation on the seed germination of Brinjal and Cowpea.

There are conflicting reports as far as the effect of simulated acid rain on the germination behavior of crop seeds is concerned. Various studies have reported a positive impact of acidity on the vigour and viability of crop seeds. In this concern, Baldwin (1934) reported that the seeds of *Picea rubra* germinate better in acidic conditions. Similarly, Verma and Prakash (1999) recorded stimulation in the germination of Cowpea seeds at pH 4.5 (though a decrease was observed at lower pH 3.5 and 2.5). Goubitz et al. (2003) and Perez-Fernandez et al. (2006) believed that acidic pH negatively affects seed germination in *Pinus alpensis*. The results of the present study revealed that the germination percentage of seeds of both the plants (Brinjal and Cowpea) was much lesser in trays treated with SAR pH 4.5 than those with SAR pH 5.6.

In the present study, rotting was observed in some Cowpea seeds germinating in solutions at pH 3.5 and 2.5. It is possible because tissue hardening does not occur much at the seedling stage. Yuan et al. (2011) observed that simulated acid rain of pH 2.5 seriously decreased the maize seed germination. The seedlings with softer tissues are more prone to adverse effects of acidic solutions. However, this issue requires further analysis and intensive investigation. The inhibitory action of SAR on the seed germination process has been attributed to the presence of bio-toxic radicals, i.e. sulfite and bisulfate, which alter the seed water interaction necessary for ongoing enzyme activity. Acid rain is an abiotic hazard that affects the growth and development of endogenous hormones, photosynthesis, antioxidant defence and molecular mechanisms under elevated acid rain (Debnath et al. 2023).

The impact of simulated acid precipitation on seed germination of Brinjal and Cowpea has sparked conflicting findings within the scientific community. While some studies indicate detrimental effects on germination behavior, others suggest limited or positive effects. This disparity can be attributed to factors like Acid Rain Composition, Soil characteristics and Crop variability. Acid rain is a complex mixture of acidic components. The specific chemical composition, including sulfur dioxide (SO_2) and nitrogen oxides (NOx), along with other pollutants, can vary widely based on geographical location and industrial activities (Wang et al., 2020). This variability can lead to different effects on seed germination. Similarly, other crops respond differently to environmental stressors. Brinjal and cowpeas may exhibit varying sensitivities to acid rain due to differences in genetics, physiological traits, and inherent tolerance levels (Bhattacharya 2022). Soil pH, composition, and nutrient levels can mediate the effects of acid rain on seeds. Altered soil conditions can either amplify or mitigate the impact of acidity on germination (Yadav et al. 2020). The predictive analysis was conducted using an ML-based Decision Tree Algorithm to predict the germination percentage of Brinjal and Cowpea plants based on pH and time variables. The results showed that the Decision Tree Algorithm could predict the germination percentage of both plants with a high degree of accuracy. These findings align with previous studies that have reported the effectiveness of Decision Tree Algorithms in predicting plant growth (Zhang et al. 2019). Like any other pollutant, SAR may also affect plants at any stage of plant development; it was decided to study the effects of acid precipitation on seed germination development. The seed germination is the first step in the life cycle of a seed of Brinjal and Cowpea. The germination percentage of seeds of both the plants (Brinjal and Cowpea) was much lesser in trays treated with SAR pH 4.5 than those with SAR pH 5.6. The inhibitory effect was quite marked. The lower the SAR pH, the more seed germination inhibition. It can be concluded that all the plants do not respond to SAR uniformly. The present study observed rotting in some Cowpea seeds germinating in pH 3.5 and 4.5 solutions. The seedling also suffered rotting at pH 3.5 and 2.5. It is possible because tissue hardening does not occur much at the seedling stage. The seedlings with softer tissues are more prone to adverse effects of acidic solutions. However, these issues require further analysis and intensive investigation.

The present study explored the effect of pH and time on the germination of Brinjal and Cowpea plants. The results showed that both pH and time significantly impacted the germination of these plants. Specifically, a pH of 7 was optimal for Brinjal and Cowpea seeds germination, while the germination rate decreased at extreme pH values. Moreover, the germination rate increased with time, up to a certain point, and then plateaued. Furthermore, regression analysis was employed to determine the relationship between the dependent (percentage germination) and independent variables (pH and time). The analysis revealed a significant linear relationship between pH and percentage germination, indicating that the germination rate increased as pH approached the optimal value of 7. Additionally, the predictive analysis using the decision tree algorithm demonstrated that pH was the most important variable in predicting the germination rate of these plants.

Conclusions

These findings have important implications for agriculture in optimizing the germination of Brinjal and Cowpea plants. Farmers can maximize the yield and quality of these crops by controlling the pH and time of germination. Additionally, machine learning algorithms can help predict the germination rate of these plants under different conditions, thereby providing valuable insights for plant breeders and researchers. The present study has highlighted the significance of pH and time in the germination percentage of Brinjal and Cowpea plants. The study utilized behavioral and predictive analysis to understand the relationship between the variables and their impact on germination. Farmers and agricultural researchers can use the findings to optimize the growth conditions of these important vegetable and fodder crop plants. Further research could explore the effect of other variables, such as temperature and light, on the germination of these plants. Overall, the present study contributes to our understanding of the factors that affect the germination of Brinjal and Cowpea plants and provides a foundation for future research in this area.

Acknowledgements

We are thankful to the Department of Botany CCS University, Meerut, UP, India and the Department of Botany and Microbiology, Gurukul Kangri University, Haridwar, Uttarakhand, India, for providing all facilities required while performing this experiment. We are also thankful to the Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, Haryana, India, for data analysis and presentation.

Conflict of Interest

The authors declare no conflict of interest.

Data availability

The data can be supplied as a supplementary file.

Funding Information

Self-funded.

References

Akinci, I. E., & Akinci, S. (2010). Effect of chromium toxicity on gormination and early seedling crowth in melon (*Cucumis melo* L.). *African Journal of Biotechnology*, *9*, 4589- 4594.

Altae, M. (2022). Study of the effect of acid rain prepared in the laboratory on building materials in Salah Al-Din/Iraq. *British Journal of Global Ecology and Sustainable Development*, *3*, 18-31.

Anitha, P. C., & Ramanujam, M. P. (1992). Impact of simulated acid rain on germination and seedling growth of groundnut. *Advances in Plant Sciences*, *5*, 180- 186.

Arora, V., Singh, B. J., Bithel, N., Malik, N., Upadhyay, S. K., & Singh, R. (2022). Effect of simulated acid rain on plant growth behaviour of *Solanum melongena* Linn. and *Vigna unguiculata* ssp *cylindrica* (L.) Walp. *Environment, Development and Sustainability*, 1-29. https://doi.org/10.1007/s10668-022-02726-4

Baldwin, H. I. (1934). Germination of red spruce. *Plant Physiololgy*, *9*, 491-532.

Bhattacharya, A. (2022). Effect of low-temperature stress on germination, growth, and phenology of plants: A review. In A. Bhattacharya (eds) *Physiological processes in plants under low temperature stress*, (pp. 1-106), Singapore Springer. DOI: https://doi.org/10.1007/978-981-16-9037-2_1.

Bisht, D. S., Tiwari, S., Srivastava, A. K., Singh, J. V., Singh, B. P., & Srivastava, M. K. (2015). High concentration of acidic species in rainwater at Varanasi in the Indo-Gangetic Plains, India. *Natural Hazards*, *75*(3), 2985-3003.

Borner, K., & Chen, C. (2018). Visualization. In A. Pertti, B. Leonard, & B. Julia B (Eds), the SAGE handbook of social research methods (pp. 303-326). Sage Publications. DOI: https://doi.org/10.4135/9781446212165.

Burns, D. A., Aherne, J., Gay, D. A., & Lehmann, C. (2016). Acid rain and its environmental effects: Recent scientific advances. *Atmospheric Environment*, *146*, 1-4.

Charlson, R. J., & Rhode, H. (1982). Factors controlling the acidity of natural rainwater. *Nature*, *295*, 683-685.

Debnath, B., Akhi, M. Z., Rob, M. M., Sikder, A., Rahman, M. M., Islam, M. S., & Ahammed, G. J. (2023). Role of Phytohormones in

Plant Responses to Acid Rain. In *Plant Hormones and Climate Change* (pp. 95-124). Singapore: Springer Nature.

De Ste Croix, M., Armstrong, N., Welsman, J., & Sharpe, P. (2016). Longitudinal changes in physical activity levels in children and adolescents. *Sports Medicine*, *46*(11), 1493-1518.

Du, E., Dong, D., Zeng, X., Sun, Z., Jiang, X., & de Vries, W. (2017). Direct effect of acid rain on leaf chlorophyll content of terrestrial plants in China. *Science of the Total Environment*, *605*, 764-769.

Goubitz, S., Werger, M. J. A., & Neeman, G. (2003). Germination response to five related factors of Seeds from non-serotinous and serotinous cones. *Plant Ecology*, *169*, 196-204.

Hu, X. F., Wu, A. Q., Wang, F. C., & Chen, F. S. (2019). The effects of simulated acid rain on internal nutrient cycling and the ratios of Mg, Al, Ca, N, and P in tea plants of a subtropical plantation. *Environmental monitoring and assessment*, *191*(2), 1-14.

Kaur, G., Asthir, B., & Bains, N. S. (2016). Influence of temperature and light on seed germination of wild marigold (*Tagetes minuta* L.). *International Journal of Applied and Pure Science and Agriculture*, *2*(8), 58-64.

Kumar, A., Singh, R. K., & Singh, P. (2014). Effect of pH and temperature on germination and growth of Brinjal (*Solanum melongena* L.) seedlings. *International Journal of Agriculture, Environment and Biotechnology*, *7*(3), 551-556.

Lee, J. (1988). Acid rain'. *Biological Sciences Review*, *1*, 15-18.

Lee, J. J., Neely, G. E., Perrigan, S. C., & Grothaus, L. C. (1980). Effects of simulated sulfuric acid rain on yield, growth, and foliar injury of several crops. United States. https://doi.org/10.2172/6560198.

Li, Y., Li, T., Zhao, D., Wang, Z., & Liao, Y. (2021). Different tillage practices change assembly, composition, and co-occurrence patterns of wheat rhizosphere diazotrophs. *Science of the Total Environment*, *767*, 144252.

Majumdar, A., Samanta, D., & Das, R. (2022). Chemical Characteristics and Trends of Indian Summer Monsoon Rainfall: A Review. *Aerosoland Air Quality Research*, *22*, 220019. https://doi.org/10.4209/aaqr.220019

McCormick, J. (2013). Acid Earth: the global threat of acid pollution. Routledge.

Park, S. M., Seo, B. K., Lee, G., Kahng, S. H., & Jang, Y. W. (2015). Chemical composition of water-soluble inorganic species in precipitation at Shihwa Basin, Korea. *Atmosphere*, *6*(6), 732-750.

Perez Fernandez, M. A., Calvo-Magro, E., Montanero Fernandez, J. & Oyala-velasco, J. A. (2006). Seed germination in response to chemicals: effect of nitrogen and pH in the media. *Journal of Environmental Biology*, *27*, 13-20.

Qian, G., Yang, N., Ma, A. H. Y., Wang, L., Li, G., Chen, X., & Chen, X. (2020). COVID-19 transmission within a family cluster by presymptomatic carriers in China. *Clinical Infectious Disease, 71*(15), 861-862.

Schiffer, M., & Ha, U. H. (2017). Heat maps. Wiley Statsref: Statistics Reference Online. doi: 10.1002/ 9781118445112.stat06596

Sequeira, R. (1982). Acid rain: an assessment based on acid-base considerations. *JAPCA*, *32*, 241-245.

Shi, Z., Zhang, J., Xiao, Z., Lu, T., Ren, X., & Wei, H. (2021). Effects of acid rain on plant growth: A meta-analysis. *Journal of Environmental Management*, *297*, 113213.

Singh, S., & Sodhi, Y. S. (2015). Predictive models for germination and emergence of cowpea. *Journal of Agricultural Science*, *7*(9), 105-117.

Singh, S., Singh, R. K., & Singh, P. (2018). Predictive model for germination of Brinjal (*Solanum melongena* L.) seeds. *International Journal of Agriculture, Environment and Biotechnology*, *11*(3), 547-551.

Smith, S. E., Fretz, T. A., & Gugino, B. K. (2018). Effects of soil pH on the germination and emergence of vegetable crops. *Journal of Horticulture, Environment, and Biotechnology*, *59*(2), 215-223.

Sodhi, Y. S., & Singh, S. (2013). Germination and emergence of cowpea as influenced by temperature and time of incubation. *International Journal of Seed Spices*, *3*(1), 67-71.

Tyagi, K. (2006). Acid rain pollution studies on some leguminous plante. Ph.D. thesis submitted to the CCS University, Meerut, UP, India.

Ulrich, B., Mayer, R., & Khanna, P. K. (1980). Chemical changes due to acid precipitation in a loess derived soil in Central Europe. *Soil Science*, *130*, 193-199.

Verma, A., Tewari, A., & Azami, A. (2010). An impact of simulated acid rain of different pH levels on some major vegetable plants in India. *Report and Opinion*, *2*, 38-40.

Verma, S.P., & Prakash, G. (1999). Simulated acid rain injury to germination and seedling growth of *Vigna unguiculata* L. Walpers. *Acta Botanica Indica*, *27*, 53-59.

Wang, M. H. S., Wang, L. K., & Shammas, N. K. (2020). Glossary of acid rain management and environmental protection. In Y.S., Hung, L.K., Wang, N.K., Shammas (eds.) *Handbook of environment and waste management: Acid rain and greenhouse gas pollution control* (pp. 719-749), World Scientific.

Yadav, D. S., Jaiswal, B., Gautam, M., & Agrawal, M. (2020). Soil acidification and its impact on plants. In P. Singh, S.K. Singh, S.M. Prasad (eds) *Plant responses to soil pollution* (pp. 1-26), Singapore Springer. https://doi.org/10.1007/978-981-15-4964-9_1.

Yin, M., Malkhi, D., Reiter, M. K., Gueta, G. G., & Abraham, I. (2019, July). HotStuff: BFT consensus with linearity and responsiveness. In *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing* (pp. 347-356).

Yuan, Z. Z., Zeng, S. & Zhou, Y. Y. (2011). Effects of simulated acid rain to seed germination and seedling growth in maize. *Journal of Shanxi Agricultural Sciences*, *39*(11), 1161-1164.

Zhang, W., Liu, X., Cai, D., & Zhang, Y. (2019). A decision tree algorithm for predicting plant growth in a complex environment. *Journal of Intelligent & Fuzzy Systems*, *37*(3), 3575-3582.

Zhang, Y., Li, J., Tan, J., Li, W., Singh, B.P., Yang, X., Bolan, N., Chen, X., Xu, S., Bao, Y. & Lv, D. (2023). An overview of the direct and indirect effects of acid rain on plants: Relationships among acid rain, soil, microorganisms, and plants. *Science of The Total Environment*, *873*, 162388.