Applied Mathematics and Computational Intelligence
Volume 13, No. 1, Feb 2024 [136-155]

A new Soil Management System in Enhancing the Yield of Plantation Using IoT Technique and Machine Learning for Smart Pineapple Farming

Norhanna Amalin binti Che Ismail[1], Elmy Johana binti Mohamad[1,*]

[1]Department of Electronic Engineering, Faculty of Enelectric and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat Johor, Malaysia

*Corresponding author: elmy@uthm.edu.my

Received: 11 October 2023; Revised: 25 January 2024; Accepted: 31 January 2024; Published: 29 February 2024

ABSTRACT

The machine learning technique is studied to aid farmers in decision-making and analysing soil quality based on the nitrogen, phosphorus, and potassium NPK nutrients as the current soil in Malaysia experience degradation of soil organic that affect in production of the nutrient for the crops. The research aim is to study and analyse the Artificial Neural Network model in analysing the quality of soil based on the prediction of NPK level class, which the data collected from Smart Agri-Scan. Next objective is to evaluate the prediction and accuracy of the model. The ANN model is constructed in Neural Net Fitting App in MATLAB. A feedforward neural network is applied to the ANN model and trains it with two different training functions and a different number of neurons of hidden layers. The model with the smallest Mean Square Error is chosen for data analysis as it means the model has the best performance. From the prediction graph, the output of training and validation that corresponds to the prediction model is observed. The points of the output prediction close to the reference line are considered a good prediction model, which means it can analyse soil quality accurately. In future, the model might be able to do the analysis and decision directly at the monitoring platform based on the real-life prediction data.

Keywords: Agri-Scan, MATLAB, ANN model, Smart Agriculture, Machine Learning, Artificial Neural Network

1 INTRODUCTION

The demand for pineapple (Ananas Comosus vsrMD2), which has been classified as a high-value non-seasonal tropic fruit in Malaysia, is both positive and growing. According to statistics, peat soil is used to grow roughly 90% of pineapples, with mineral soil being used for the remaining 10% [1]. As of right now, the production of pineapple is still primarily dependent on experienced farmers in a variety of plantation-related areas, including the selection of the right soil for the pineapple, the application and distribution of fertilizers, and the influence of the weather. Even while farming experience is subjective and takes time to spread to additional farmers, it is still significant. Therefore, it is crucial to translate experience into scientific terms that enable resource optimization and weather forecasting.

The purpose of this project is to help establish a database for important elements affecting soil quality, nutrients, and pineapple production in Malaysia. In conjunction with that, an IoT device called Smart Agri-scan IoT-based Precision Agriculture will be used to focus on live scanning of environmental data such as temperature, moisture, PH, NPK, and other sorts. Farmers can use the "Plug & Sense" solar-powered system provided by the system by placing it in the field and receiving real-time data feeds on smartphones and tablets. It is accessible online thanks to cloud computing. This system occasionally allows for the study of numerous types of data using big data analytics. A predictive soil quality and nutrient report can be generated from the combined data, enabling the farmer to make informed decisions about soil fertility to increase productivity.

2 MATERIAL AND METHODS

2.1 Investigate of Machine Learning Methods for agricultural management

This study aims to show the steps on how to construct the Artificial Neural Network model for study the condition of soil based on the soil parameters from the Smart Agri-scan results. The flowchart illustrates how the whole project is conducted. Research on standardized soil parameters is required for classifying the soil parameter as their output. After collecting the data, Artificial Neural Network is constructed by studying the Artificial Neural Network Architecture. The model is trained with different training functions and numbers of hidden layer neurons until get the smallest error for predicting new data. Model performance is evaluated using Mean Square Error.

The flowchart in Figure 1 is the methodology step in evaluating soil conditions using Artificial Neural Network models.

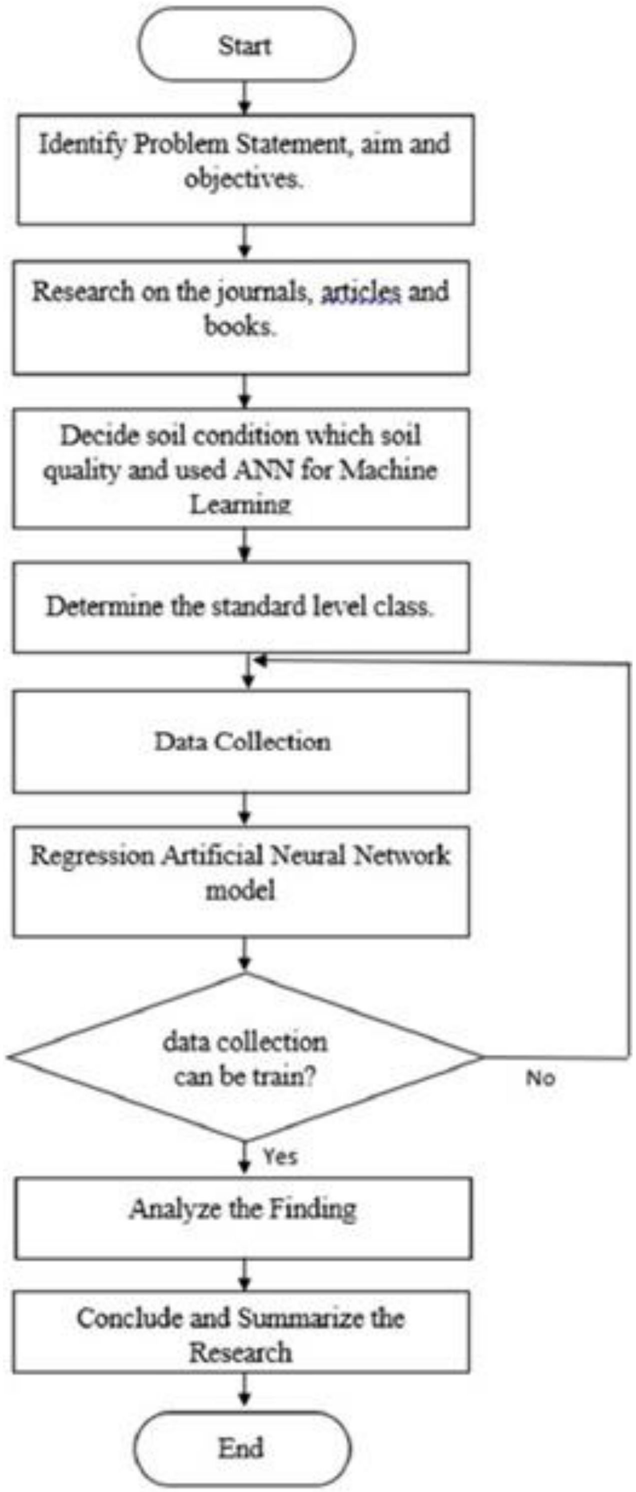

Figure 1: The flowchart of whole project

2.2 Standardize for the level class of NPK nutrient, Temperature and Moisture

The standardization of the NPK nutrients is required to indicate the target output, whether they are Very Low, Low, Normal, High, or Very High. This condition will be imported to the ANN model. The NPK nutrient level class tables are obtained by referring to Asean Guidelines on Soil and Nutrient Management [2].

Table 1: The standardized class level for Nitrogen, phosphorus and postassium

Class	N (mg/kg)
Very High	>3300
High	2600 – 3300
Moderate	1400 – 2600
Low	700- 1400
Very Low	<700

Class	P (mg/kg)
Very High	>20
High	15 - 19
Moderate	10 - 14
Low	5 - 9
Very Low	<5

Class	K (mg/kg)
Very High	>180
High	150 - 179
Moderate	90 - 149
Low	30 - 89
Very Low	<30

2.3 Data Collection: Smart Agri-Scan

Smart Agri-Scan is a soil monitoring device which in this study focused in measuring Nitrogen (mg/kg), Phosphorus (mg/kg) and Potassium (mg/kg). This system will display the trend of crops fertility and simultaneously give precision data time to time in the Arduino IOT Cloud Dashboard as shown in Figure 2

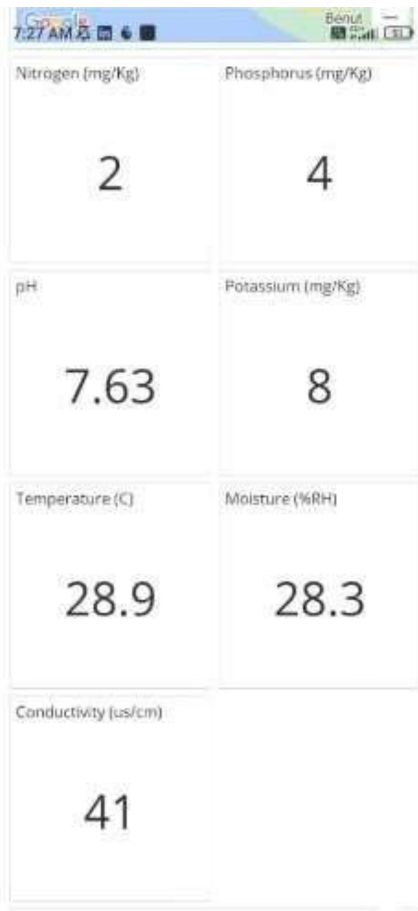

Figure 2: Arduino IoT Cloud Dashboard

In this study, the device is set up at plantation area in front of Tun Fatimah Residential College. The data collection is taken for two weeks for analysis purposes. The fertilizer is added to the soil as the NPK nutrient is considered too low and shortage of varies level class value for the training data in the ANN model. After filtering the data only 300 data is reconsidered for training and 100 new data for implementing in the model to predict class level for NPK nutrient and soil quality analysis.

Figure 3: The portable Smart Agri-Scan device

2.4 Artificial Neural Network Model

The Artificial Neural Network modelling is constructed using MATLAB's Neural Net Fitting app under Deep learning toolbox [3]. This Neural Net Fitting app defines the fitting problem. In this project, each soil parameter, Nitrogen, Potassium, Phosphorus, is the input and the level class based on the referred standardized is the model output [4]. Based on Figure 4, the flowchart shows the steps of training the ANN Neural Network modelling, which starts with defining the input and output parameters. Next is determined the Artificial Neural Network Architecture as shown Figure 4 [5]. The model is trained until get the acceptable error. After that, if the validate ANN model performance is acceptable, the model is ready for prediction but if the validate ANN model performance is not acceptable, required to start back with reconstruct the ANN Architecture.

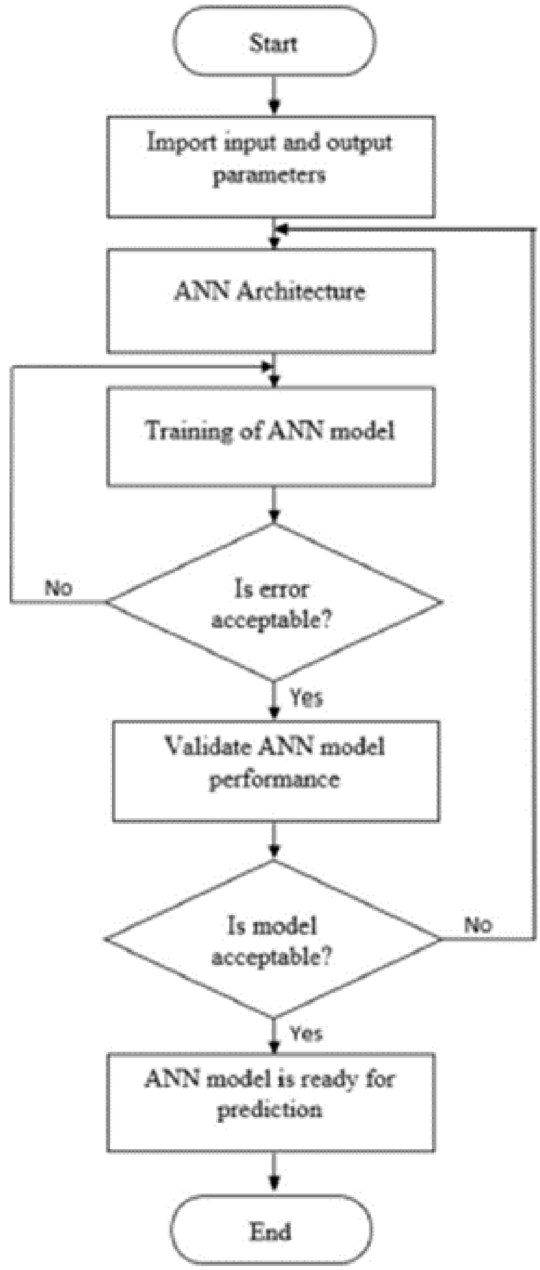

Figure 4: The flowchart is graphical step of constructing ANN modelling [6]

2.5 Artificial Neural Network Architecture

The Architecture of Artificial Neural Network is a feedforward neural network with three layers which input, sigmoid hidden neurons with ten neurons and a linear output neuron as shown in Figure 3.6. In addition, in order to avoid over-fitting and enhance the efficiency of fitting, Cross-Validation is used. Based on [7], Cross-Validation is a statistical method for evaluating and comparing learning algorithms by dividing data into two segments. one segment is used to train a model and the other segment is used to validate the model. The model will learn the correlation between input and target output with setup division of data for Training and Validation is 70:30.

Levenberg-Marquardt training Algorithm is used to train the neural network by updating weight and bias values based on its optimization. The Levenberg-Marquardt is the fastest training algorithm, but it required a lot of memory compared to others training algorithms. Scaled Conjugate Gradient back propagation is a fast convergence training function with low memory requirements. This algorithm also applied as of the collected data is 300 which might require a lot of space [8]. Two different numbers of neurons of hidden layer which is 10 and 20 is performed in training the model to identify the best model performance. Also, a coding for determining the error of both training and validation is constructed as shown in Figure 3.5. The errors are observed and compared, which if the gap error of the error is huge, the model might overfit. Overfitting occurs when ANN model produces accurate prediction for training data bit not for new data.

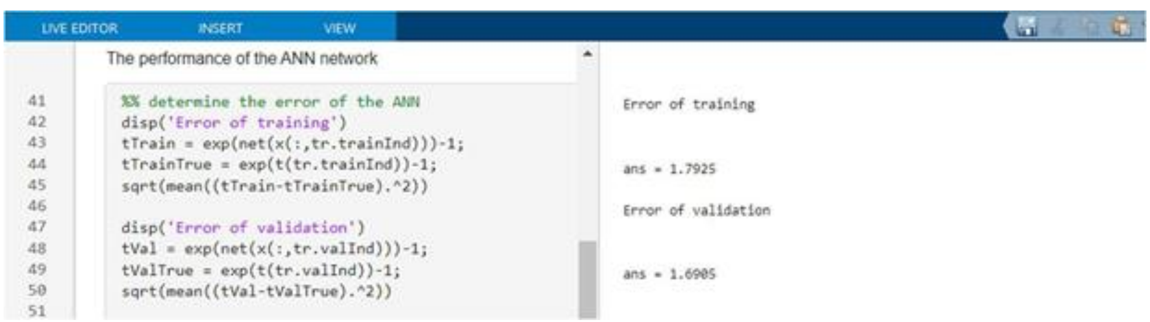

Figure 5: The percentage error of training and validation ANN network

Applied Mathematics and Computational Intelligence
Volume 13, No. 1, Feb 2024 [136-155]

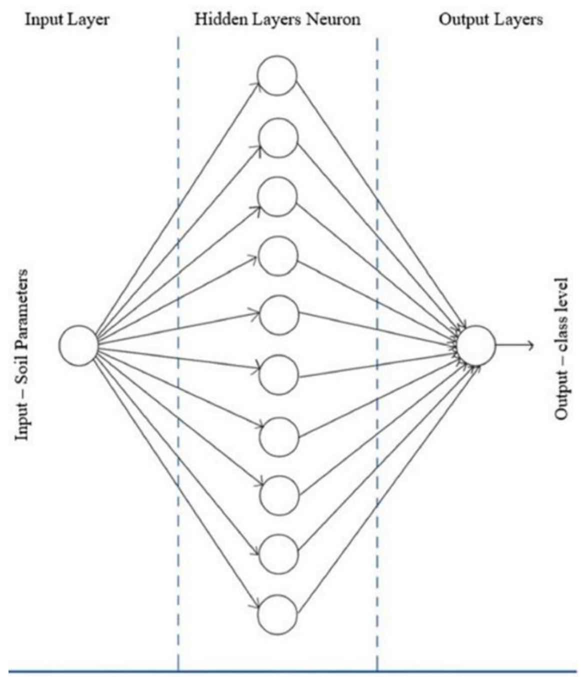

Figure 6: The feed-forward neural network with three layers [9]

2.6 Implementation of data collection for training in ANN model

The collected data is trained using Net Fitting app in MATLAB. The network will be trained with Levenberg- Marquardt back propagation algorithm (trainlm), and because of the collected data is greater than 100, Scaled Conjugate Gradient back propagation (trainscg) is used [9]. Both training functions are trained with two different neurons numbers of hidden layers which is 10 and 20 to identify the best performance of ANN model. The 210 data is trained as the division of data for training is set up as 70%. Since, the training data is selected randomly by the algorithm for every data training. Hence the mean square error of the model in each training will always be different as the random data is taken from the whole data set and it is not fixed [10]. The model with smallest mean square error, MSE is selected.

The table of ANN model cases shows as following:

Table 2: The ANN Training Model

No.	Training Function	No. of Neuron
Model 1	TRAINLM	10
Model 2	TRAINLM	20
Model 3	TRAINSCG	10
Model 4	TRAINSCG	20

2.7 Validation of Data Set

The training of the data is completed after the error of the training is accepted or below the tolerance levels. The division of data for validation is 30% which out of 300 data, 90 data are used for validation.

2.8 Prediction of NPK Nutrient

Prediction of new NPK nutrients of ANN model is tested once finish the data training until get the best performance of training.

2.8.1 Graph: Actual data corresponding to the predicted data

The prediction from the ANN will be visualized in a scatter plot of the actual data which prediction of model and the output is predicted data which is the plotted point of training and validation data set prediction. The scatter plot is an effective graph that presents the comparison of output model and model prediction. The graph can determine the accuracy of the model. If the output model is plotted at or close to the reference line corresponds to the predicted value, the model has high accuracy. If the output prediction point is far from the reference line, it means that the model is not well predicted. Figure 7 shows the sample of the scatter graph of the testing and validation corresponding to the prediction model.

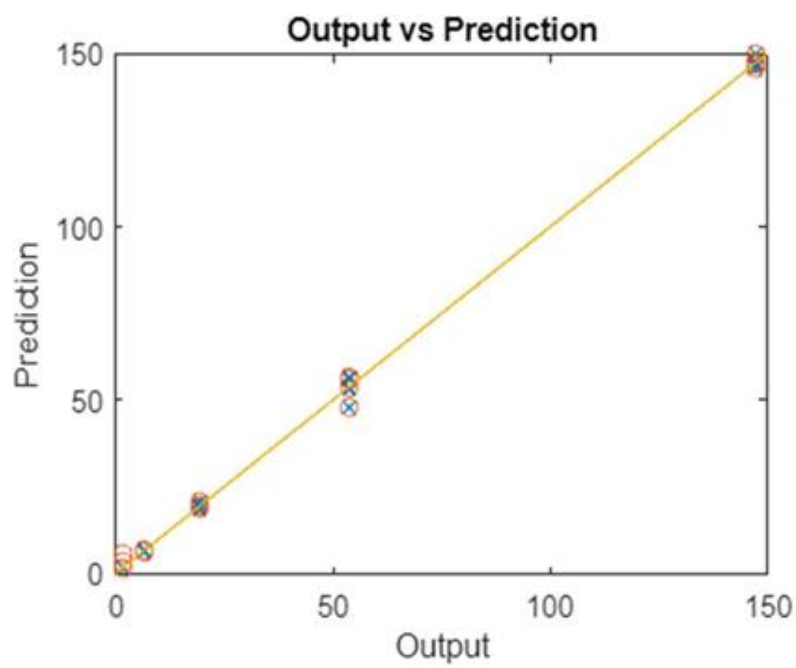

Figure 7: The (x) is the training set and (o) is the validation set

2.8.2 Coding: Predict the NPK Nutrient Class Level

Besides that, a simple training script is generated by using a command-line functionality of the Neural Net Fitting toolbox. The coding is modified to customize the network training for predicting or optimizing the model and run it at the MATLAB Live Editor. Next, import a set of new input in the model as shown in line 87 in Figure 8 The level class of Very High, High, Normal, Low and Very Low is converted into scale form, 5, 4, 3, 2, and 1 for MALAB easy to read or interpreted the class level but it will convert back to the level class for observatory or farmer understand the data better. The model's performance and prediction are displayed in the output display in Live Editor window. Then the outcomes are compared and analysed with the expected value to assess the accuracy of the machine learning model. The soil quality is evaluated by analysing the predicted class level NPK nutrient.[11].

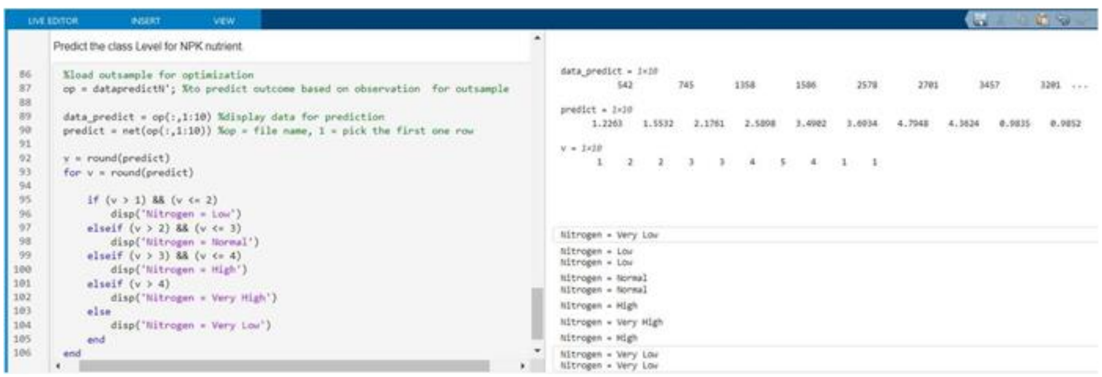

Figure 8: Predict the class level of Nitrogen

2.9 Mean Square Error and Regression value of the ANN model

In this research, the error performance validation is by evaluating the mean square error, MSE. Besides, the Regression, R-value is analyzed as it indicates the consequential relationship between the input with its class level. The input is determined to be correlated to the target if the R value is close to one. Mean Square Error (MSE) Mean Square Error is a standard deviation that evaluates the difference of the predicted deviation outcome with the actual value. MSE is to evaluate the mean squared difference between the observed value and the predicted values which if MSE obtained is zero, then the model has no error. The MSE is increased corresponding to the error happened on the model [12]. Formula for Mean Square Error (MSE).

$$MSE = \frac{1}{n}\sum(y - Y)^{z}$$

$y = is\ actual\ of\ y$

$Y = is\ a\ prediction\ value\ of\ y$

$n = number\ of\ observation$

3 RESULT AND DISCUSSION

3.1 Result of Artificial Neural Network Modelling

Table 3 shows the MSE results of each model training for the Nitrogen content in the soil. For training, it is the four conditions to determine the best performance model.

Table 3: Neural Network MSE results of Nitrogen

No.	Training Function	No. of Neuron of hidden layer	Mean Square Error (MSE)
Model 1	TRAINLM	10	0.0170
Model 2	TRAINLM	20	**0.0054**
Model 3	TRAINSCG	10	0.0256
Model 4	TRAINSCG	20	0.0187

Table 4 shows the MSE results of each model training for the Phosphorus content in the soil with same conditions as nitrogen in order to determine the best performance model.

Table 4: Neural Network MSE results of Phosphorus

No.	Training Function	No. of Neuron of hidden layer	Mean Square Error (MSE)
Model 1	TRAINLM	10	0.00440
Model 2	TRAINLM	20	0.00740
Model 3	TRAINSCG	10	0.02340
Model 4	TRAINSCG	20	0.00247

Table 5 shows the MSE results of each model training for the Potassium content in the soil with the same conditions as nitrogen and phosphorus nutrient.

Table 5: Neural Network MSE results of Potassium

No.	Training Function	No. of Neuron of hidden layer	Mean Square Error (MSE)
Model 1	TRAINLM	10	0.01790
Model 2	TRAINLM	20	**0.00534**
Model 3	TRAINSCG	10	0.03270
Model 4	TRAINSCG	20	0.02850

3.2 Implementation of new set data in the ANN model

Based on the selected model for the NPK, collected data from a mango plantation is trained In the ANN models. Table 6, Table 7, and Table 8 show the prediction of NPK level class for the soil quality assessment.

Table 6: The prediction result of nitrogen data

Class	N(mg/kg)	Num. of Sample
Very High	>3300	15
High	2600 - 3300	8
Normal	1400 - 2600	17
Low	700 - 1400	8
Very Low	<700	52

Table 7: The prediction result of phosphorus data

Class	P(mg/kg)	Num. of Sample
Very High	>20	50
High	15 - 19	12
Normal	10 - 14	11
Low	5 - 9	18
Very Low	<5	9

Table 8: The prediction result of phosphorus data

Class	K(mg/kg)	Num. of Sample
Very High	>180	5
High	150 - 179	3
Normal	90 - 149	6
Low	30 - 89	34
Very Low	<30	52

4 CONCLUSION

This paper proposes a novel smart farming system utilizing advanced Artificial Intelligence techniques to monitor and control critical factors influencing the growth of MD2 pineapples. With the increasing global population, ensuring an adequate food supply is of utmost importance, and this system presents a significant step forward in enhancing agricultural production at the farm and plantation level. The effectiveness of this system hinges on harnessing and analysing data and insights from the environment. To facilitate this, web-based dashboards will be developed, enabling farmers to visualize and analyse the collected data. Through these dashboards, farmers can observe the trends and variability in the data, empowering them to take proactive measures for future fruit management. By leveraging the capabilities of this smart farming system, we can improve the efficiency and productivity of pineapple cultivation, contributing to a sustainable and thriving agricultural sector. As the data kept training, discovered that with the increasing number of neurons in an ANN model, the MSE value is decreased. Training function Levenberg-Marquardt backpropagation is the best function training. Scaled Conjugate Gradient backpropagation is performed well for a wide variety of problems for networks with a large number of weights. The training function Levenberg-Marquardt backpropagation algorithm with 20 neurons of the hidden layer is the best-predicted model for Nitrogen and Potassium with the lowest MSE of 0.00540 and 0.00534. The regression plot for both models is 0.99859 and 0.99856. The training function Scaled Conjugate Gradient backpropagation with 20 neurons of the hidden layer is the best-predicted model for the Phosphorus model. The MSE and Regression values of the model is 0.00247 and 0.99838. The new data that implemented in these models, proved that the model can improve decision-making, which helps farmers make better decisions in managing the crops, especially for crops that require NPK nutrients to grow, such as pineapple, mango, and papaya trees.

ACKNOWLEDGEMENT

This research was supported by Ministry oh Higher Education (MOHE), through Fundamental Research Grant Scheme (GPPS). We also want to thank Universiti Tun Hussein Onn Malaysia for the Research Grant through Multidisciplinary Research (Vot H476). Sincere gratitude to Faculty of Electrical Engineering for the endless support.

REFERENCES

[1] A. M. Nazri and D. E. Pebrian, "Analysis of Energy Consumption in Pineapple Cultivation in Malaysia: A Case Study", *Pertanika J. Sci. & Technol*. vol. 25, no. 1, pp. 17 – 28, 2017.

[2] M. K. Doula and A. Sarris, "Soil Environment," *Environment and Development*, pp. 213– 286, 2016.

[3] MATLAB, Version R2022b. Natick, Massachusetts: MathWorks Inc, 2022.

[4] Neural net fitting. Fit Data with a Shallow Neural Network - MATLAB & Simulink-MathWorks India. from https://in.mathworks.com/help/deeplearning/gs/fit-data-with- a-neural- network.html.

[5] Z. Bai et al., "Effects of agricultural management practices on soil quality: A review of long-term experiments for Europe and China," *Agriculture, Ecosystems & Environment*, vol. 265, pp. 1–7, Oct. 2018, doi: https://doi.org/10.1016/j.agee.2018.05.028.

[6] M. S. Suchithra and M. L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters," *Information Processing in Agriculture*, May 2019.

[7] C. W. MohdNoor, R. Mamat and A. N. Ahmed, "Comparative study of artificial neural network and mathematical model on marine diesel engine performance prediction," *International Journal of Innovative Computing Information and Control*, vol. 14, pp. 959-969, 2018.

[8] S. Mishra, Bibhu Prasad Ganthia, A. Sridharan, Perumal Rajakumar, D. Padmapriya, and S. Kaliappan, "Optimization of Load Forecasting in Smartgrid using Artificial Neural Network based NFTOOL and NNTOOL," vol. 2161, no. 1, pp. 012068–012068, Jan. 2022.

[9] Neural net fitting. Fit Data with a Shallow Neural Network - MATLAB & Simulink. (n.d.), from https://www.mathworks.com/help/deeplearning/gs/fitdata-with-a-neural-network.html.

[10] S. Motia and S. Reddy, "Exploration of machine learning methods for prediction and assessment of soil properties for agricultural soil management: a quantitative evaluation," *Journal of Physics: Conference Series*, vol. 1950, no. 1, p. 012037, Aug. 2021.

[11] Mansi Phute, Jyoti Madake, and Shripad Bhatlawande, "Precision Farming Based Soil Fertility Assessment Techniques," Sep. 2021,

[12] Mean squared error in machine learning, Naukri.com.Available at: https://www.naukri.com/learning/articles/mean-squared-error-in-machinelearning/.

[13] A. M. Nazri and D. E. Pebrian, "Analysis of Energy Consumption in Pineapple Cultivation in Malaysia: A Case Study", *Pertanika J. Sci. & Technol*. vol. 25, no. 1, pp. 17 – 28, 2017.

[14] MIPB. Pineapple info. Malaysian Pineapple Industry Board. Retrieved August 20, from www.mpib.gov.my, 2012.

[15] K. E. Danso, K. O. Ayeh, V. Oduro, S. Amiteye and Amoatey, “Effect of 6- Benzylaminopurine and Naphthalene Acetic Acid on in vitro production of MD2 pineapple planting materials”, *World Applied*, 2008.

[16] M. A. Akbar, B. Karmarka and Roy, “Callus induction and high frequency plant regeneration of pineapple (Ananas comosus L. Merr.)”, *Plant Tissue Culture*, vol. 13, no. 2, pp. 109–116, 2003.

[17] A. Ahmadi Thalip and T. Pei Sin and N, Casey, “The MD2 'Super Sweet' pineapple (Ananas comosus)”, *UTAR Agriculture Science Journal*, 2015.

[18] K. Jaji, N. Man, and N. M. Nawi, “Factors affecting pineapple market supply in Johor, Malaysia,” *Int. Food Res. J.*, 2018.

[19] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, “An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges,” *IEEE Internet Things J.*, 2018.

[20] A. De Mauro, M. Greco and M. Grimaldi, "A formal definition of Big Data based on its essential features", *Library Review*, vol. 65, no. 3, pp. 122-135, 2016.

[21] M. I. F. Musaddin, F. P. Rahman, S. N. A. Rahman, M. S. M. Ali, A. Amron, and M. S. Azizan, “Development of dashboard to identify rice germplasm in MARDI,” *Econ. Technol. Manag. Rev.*, vol. 13, pp. 111–120, 2018.

[22] A. Suk Oh, “Smart urban farming service model with IoT based open platform,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 1, p. 320, Oct. 2020, doi: 10.11591/ijeecs.v20.i1.pp320-328.

[23] T. T. Gurmessa, “A Big Data Analytics Framework in Climate Smart Agriculture,” *Computer Engineering and Intelligent Systems*, vol. 10, no. 6, pp. 1–6, 2019, doi: 10.7176/ceis/10-6-01.

[24] Y. Shekhar, E. Dagur, S. Mishra, R. J. Tom, M. Veeramanikandan, and S. Sankaranarayanan, “Intelligent IoT based automated Irrigation system,” *International Journal of Applied Engineering Research*, vol. 12, no. 18, pp. 7306–7320, 2017.

[25] Essays UK, “The Advantages of Information Visualization Information Technology Essay,” 2016. [Online]. Available: https://www.ukessays.com/essays/information- technology/the-advantages-of-information-visualization-information-technology-essay.php.

[26] R. S. King, Cluster analysis and data mining, Boston, MA, 2015.

[27] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms. USA, Kluwer Academic Publishers, 1981.

APPENDIX A: COLLECTED DATA OF NITROGEN, PHOSPHORUS AND POTASSIUM

No.	Nitrogen	No.	Nitrogen	No.	Nitrogen
1	0	35	46	69	2245
2	1	36	50	70	2299
3	2	37	55	71	2304
4	2	38	64	72	2306
5	2	39	88	73	2555
6	3	40	106	74	2590
7	3	41	110	75	2618
8	4	42	205	76	2648
9	5	43	272	77	2689
10	5	44	323	78	2832
11	6	45	374	79	3038
12	6	46	415	80	3041
13	7	47	461	81	3072
14	9	48	463	82	3073
15	10	49	670	83	3094
16	10	50	678	84	3140
17	11	51	695	85	3271
18	13	52	717	86	3361
19	15	53	753	87	3429
20	16	54	849	88	3549
21	17	55	927	89	3574
22	17	56	1037	90	3625
23	17	57	1110	91	3726
24	18	58	1147	92	3773
25	18	59	1378	93	3860
26	18	60	1420	94	3906
27	21	61	1497	95	4280
28	26	62	1804	96	4282
29	32	63	1907	97	4370
30	34	64	1999	98	4415
31	38	65	2000	99	4416
32	40	66	2148	100	8224
33	42	67	2154		
34	45	68	2188		

No.	Phosphorus	No.	Phosphorus	No.	Phosphorus
1	1	35	14	69	24
2	1	36	14	70	24
3	1	37	14	71	25
4	2	38	14	72	25
5	2	39	15	73	26
6	3	40	15	74	27
7	3	41	16	75	28
8	4	42	16	76	28
9	4	43	17	77	28
10	5	44	17	78	28
11	6	45	17	79	28
12	6	46	18	80	30
13	6	47	18	81	30
14	7	48	19	82	30
15	7	49	19	83	30
16	7	50	19	84	31
17	7	51	20	85	32
18	8	52	20	86	32
19	8	53	21	87	33
20	8	54	22	88	34
21	8	55	22	89	34
22	8	56	22	90	35
23	8	57	22	91	35
24	8	58	22	92	35
25	9	59	22	93	36
26	9	60	23	94	52
27	9	61	23	95	55
28	10	62	23	96	57
29	11	63	23	97	63
30	11	64	23	98	64
31	12	65	24	99	77
32	12	66	24	100	124
33	12	67	24		
34	13	68	24		

No.	Potassium	No.	Potassium	No.	Potassium
1	0	35	23	69	38
2	1	36	23	70	39
3	2	37	23	71	39
4	2	38	24	72	40
5	2	39	24	73	41
6	3	40	24	74	42
7	3	41	24	75	42
8	4	42	25	76	43
9	5	43	26	77	43
10	5	44	27	78	44
11	6	45	27	79	44
12	6	46	28	80	45
13	7	47	28	81	45
14	9	48	29	82	47
15	10	49	29	83	48
16	10	50	29	84	57
17	11	51	30	85	77
18	13	52	30	86	88
19	15	53	31	87	98
20	16	54	31	88	100
21	17	55	32	89	106
22	17	56	33	90	110
23	17	57	33	91	123
24	18	58	34	92	134
25	18	59	34	93	151
26	18	60	34	94	163
27	21	61	34	95	167
28	26	62	34	96	190
29	32	63	34	97	225
30	34	64	35	98	236
31	38	65	35	99	255
32	40	66	35	100	310
33	42	67	36		
34	45	68	36		