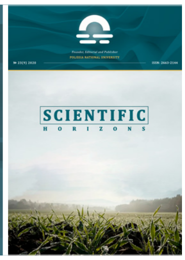
SCIENTIFIC HORIZONS

Journal homepage: https://sciencehorizon.com.ua

Scientific Horizons, 28(8), 165-178

UDC 633.18:631.4:519.876.5(594) DOI: 10.48077/scihor8.2025.165



Machine learning-based prediction of rice yield from rain feeding in monsoonal tropical area of Java, Indonesia

Azhari Rizal*

Master

Politeknik LPP Yogyakarta

55222, No. 1A Jl. LPP Balapan, Yogyakarta, Indonesia

https://orcid.org/0009-0006-7607-8746

Abdul Aziz

Master

LPP Agro Nusantara

55222, No. 01 LPP Str., Yogyakarta, Indonesia

http://orcid.org/0000-0002-9688-3180

Komariah

Doctor of Applied Science

Universitas Sebelas Maret

57126, No. 36 Ir. Sutami Str., Surakarta, Indonesia

https://orcid.org/0000-0001-7704-0754

Muhammad Rizky Romadhon

Master

Lambung Mangkurat University

70714, KM.36 Jenderal A. Yani Str., Banjarbaru, Indonesia

http://orcid.org/0000-0002-7929-769X

Muhamad Mustangin

Master

Politeknik LPP Yogyakarta

55222, No. 1A Jl. LPP Balapan, Yogyakarta, Indonesia

https://orcid.org/0009-0001-5999-4519

Article's History:

Received: 6.03.2025

Revised: 6.08.2025

Accepted: 27.08.2025

Abstract. Rainfed rice cultivation in monsoonal tropical areas of Java, Indonesia, is challenged by nutrient deficiencies, unpredictable rainfall amounts, and limited agricultural investment, leading to fluctuating yields. The purpose of this study was to develop a precise rice yield prediction model using machine learning tailored to specific toposequences in Central Java. A combination of survey-based field and laboratory methods was employed, integrating climate, soil, socio-economic, and

Suggested Citation:

Rizal, A., Aziz, A., Komariah, Romadhon, M., R., & Mustangin, M. (2025). Machine learning-based prediction of rice yield from rain feeding in monsoonal tropical area of Java, Indonesia. *Scientific Horizons*, 28(8), 165-178. doi: 10.48077/scihor8.2025.165.

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

**Corresponding author*

land management variables from 87 targeted sampling points. Machine learning analysis using Bayesian Neural Networks (BNN) demonstrated moderate accuracy with $R2 = 0.840$ and $RMSE = 0.442$ overall, but accuracy improved significantly when models were adjusted to elevation-specific categories, achieving R2 values up to 0.999. Lowland paddy field predictions were most influenced by available phosphorus (P), while rainfall, gender, education, and seed variety were key factors in medium-altitude zones; slope, available P, gender, and cropping patterns were dominant in highland areas. Pareto analysis supported the identification of these key yield determinants in each toposequence. The integration of BNN and Pareto approaches enabled the creation of a high-precision, location-specific yield prediction model. This work demonstrated that tailoring machine learning models to elevation-based agroecological zones enhances their performance and practical application. The findings are particularly valuable for agricultural stakeholders including policymakers, extension services, and farmers, who can leverage these predictive insights to optimise rainfed rice management practices and improve productivity under variable climatic conditions

Keywords: agricultural sustainability; Bayesian Neural Network (BNN); food security; Pareto analysis; precision agriculture

INTRODUCTION

Rainfed rice production in monsoonal tropical environments continues to be a crucial element of food systems in many developing regions. These systems are typified by seasonal rainfall patterns, limited irrigation infrastructure, nutrient-depleted soils, and socio-economic challenges. The productivity of rainfed rice fields is constrained by environmental variability, soil degradation, and the inability of smallholder farmers to adopt efficient management strategies. These challenges result in fluctuating and often suboptimal yields, which limit long-term food security and economic resilience in rural areas. As global demand for rice continues to rise and climate patterns shift unpredictably, the development of accurate, site-specific predictive models for rice yields is vital to inform agricultural planning and optimise input allocation in such high-risk production systems.

Machine learning (ML) has gained popularity in agricultural modelling due to its ability to manage complex nonlinear interactions between diverse input data. J. Zhang *et al*. (2023) developed a hybrid ML model combining climatic and soil factors for smallholder farms in Southeast Asia and reported a significant increase in prediction accuracy when socio-environmental interactions were considered. Analogously, N. Din *et al*. (2024) employed convolutional neural networks for rice yield forecasting across South Asia, demonstrating that spatial-temporal data integration markedly enhanced model robustness. P. Velmurugan *et al*. (2023) introduced a fuzzy enumeration technique in big data environments for agricultural forecasting and emphasised the predictive advantage of including farmer demographics and cropping patterns. D. Paudel *et al*. (2021) proposed large-scale ML frameworks for European agricultural systems, highlighting the scalability of data-driven models while cautioning that prediction accuracy declines without site-specific calibration. S. Liu *et al*. (2017) compared various ML methods and concluded that Bayesian Neural Networks (BNN) outperform conventional regression models in handling uncertainty and sparse datasets widespread in rainfed regions.

J. Qian *et al*. (2025) examined the effects of declining rainfall suitability on paddy production in tropical monsoon climates, asserting the need for adaptive modelling that accounts for regional climate vulnerabilities. In a pixel-scale study, S. Jeong *et al*. (2022) applied deep learning models to rice yield prediction using satellite imagery, concluding that predictive power increased substantially when incorporating slope, elevation, and localised agroecological variables. J. Pant *et al*. (2021) emphasised the significance of non-agronomic inputs, such as education level and access to improved seed varieties, in boosting prediction accuracy in rainfed rice fields. T. Tu *et al*. (2023) further demonstrated that in highland paddy systems, land slope and contour significantly influenced yields, and concluded that predictive models must be tailored to local topographic and management realities. Although these studies have contributed to advancements in ML-based crop yield forecasting, gaps persist in understanding how location-specific topographical characteristics, particularly in tropical rainfed zones such as Central Java interact with agronomic and socio-economic variables to affect model performance. Prior research has tended to focus either on climate or soil characteristics alone, without a holistic integration of diverse influencing factors within a localised, elevation-specific context.

The purpose of this study was to construct a rice yield prediction model for rainfed agricultural systems in Central Java using a Bayesian Neural Network (BNN) integrated with Pareto analysis, with the aim of identifying key determinants of yield across multiple mountain toposequences classified by elevation.

MATERIALS AND METHODS

Study area. Six mountain toposequences were the subject of the study in Central Java, Indonesia, including the eastern slopes of Mount Slamet (7°12’19.0”S – 7°17’07.1”S & 109°17’37.2”E – 109°26’11.4”E), southwest of Mount Sindoro (7°22’25.8”S – 7°29’13.1”S & 109°47’21.4”E – 109°53’34.3”E), south of Mount Sumbing (7°26’09.6”S – 7°37’16.0”S & 110°02’59.0”E – 110°08’04.3”E), northwest of Mount Merbabu (7°20’48.0”S – 7°26’05.2”S & 110°13’02.5”E – 110°20’38.7”E), southwest of Mount Merapi (7°37’08.0”S – 7°39’34.7”S & 110°18’12.0”E – 110°20’08.0”E), and south of Mount Lawu (7°46’6.3”S – 7°53’19.2”S & 111°9’27.3”E – 111°13’34.9”E) (Fig. 1). The gradient of the land topography was considered during the sampling procedure. A total of 57 locations were sampled with a gradient of 0-8%, 21 locations with a gradient of 8-15%, and 9 locations with a gradient of 15-25%. Additionally, topography was also considered for the analysis. There are 34 locations with an altitude of under 400 m above sea level, 33 locations with an altitude of 400-700 m above sea level, and 20 locations with an altitude of over 700 m above sea level. The survey used soil samples to identify their composition, as well as interviews to learn about land management practices and farmers’ socio-economic circumstances. The average yield from rainfed rice crops at the research site was 2.12 tonnes ha^{-1}.

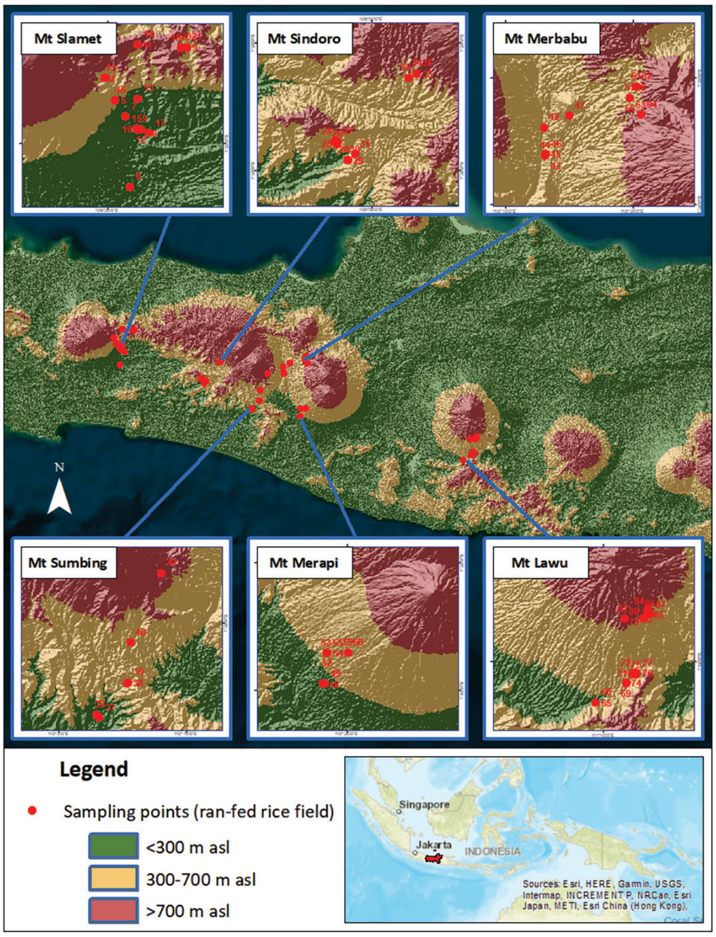


***Figure 1**. Research Site Sampling*

***Source:** ESRI (2022)*

Characteristics of the soil. The growth of plants depends on the characteristics of the soil, which also significantly influence agricultural yields. According to field observations and the outcomes of laboratory analyses of soil properties, this study employed data on soil characteristics in the field (Table 1). Soil types in the study area ranged between entisols, inceptisols, alfisols, ultisols, and andisols, with a slope of 0-25%. Generally, the soil pH at the study site ranged within 5-6, with organic C ranging within 0.39-7.74%. Soil sampling points were determined according to the toposequence, with 15-33 sampling points at each toposequence. Soil pH was measured with a pH stick (Jackson, 1973), and organic C was measured according to the Walkley and Black method (Walkley & Black, 1934). Total N was determined using the Kjeldahl method (Bremner & Mulvaney, 1982), and available P was determined using the Olsen method (soil pH > 5.5) (Olsen *et al.*, 1954). The method of ammonium acetate extraction was employed to analyse available K in soil (Thomas, 1982).

***Table 1**. Research parameters of soil characteristics at rainfed rice fields on location-specific toposequences*

Parameters (units)	Lowland paddy fields	Midland paddy fields	Highland paddy fields	Total
pH	5–6	5–6	5–6	5–6
Organic C (%)	0.39–6.03	0.67–4.02	0.96–7.74	0.39–7.74
Total N (%)	0.08–1.63	0.17–1.21	0.17–1.40	0.08–1.63
Available P (ppm)	0.29–23.74	1.04–18.16	1.59–25.78	1.04–25.78
Available K (meq 100g^{-1})	0.25–0.66	0.20–0.77	0.23–0.81	0.20–0.81
Type of soil	Entisols, Inceptisols, Alfisols, Ultisols and Andisols	Entisols, Inceptisols, Alfisols and Andisols	Entisols, Inceptisols, and Andisols	Entisols, Inceptisols, Alfisols, Ultisols and Andisols
Slope (%)	0-25	0-25	0-25	0-25

***Source:** compiled based on the analysis of soil characteristics based on (BRMP) Indonesian Agricultural Assembly and Modernization Agency (2005) and data on plant yield recap based on a social survey carried out by researchers at the research site*

Climate characteristics. The model incorporated climate data to estimate the amount of water available and usable by plants. In the yield prediction model, climate factors significantly affect crop productivity, with rainfall and air temperature as the primary determinants of yield outcomes. This predictive model considered two key variables: rainfall and evapotranspiration. The analysis relied on rainfall and air temperature data collected from climatology stations at the research site between 2014 and 2023, as presented in Table 2. Evapotranspiration was calculated using the Thornthwaite method (Moeletsi *et al.*, 2013).

***Table 2**. Climate characteristics at the research site*

Parameters (units)	Lowland paddy fields	Midland paddy fields	Highland paddy fields	Total
Rainfall (mm year^{-1})	1,845-4,874	1,845-4,000	2,246-3,954	1,845-4,874
Evapotranspiration (mm day^{-1})	1.90-5.81	1.90-5.81	1.90-5.27	1.90-5.81

***Source:** compiled based on the analysis of climate characteristics based on climatology stations at the research site (2014-2023)*

Farmers' socio-economic land management. Data were gathered through firsthand interviews with farmers who manage their land. A total of 87 respondents were included in the study, of whom 75 were male and 12 were female. The majority of respondents were aged between 51 and 60 years. The distribution of farmers' education was very diverse, ranging from uneducated to a master's degree. Most of the farmers still use conventional methods of cultivation. The farmer's tools and techniques affect the yields. To create a site-specific prediction model, this data presented an overview of farmers' socio-economic circumstances and unique land management techniques. However, this study used specific data on farmers' socio-economic circumstances (Table 3) and land management practices (Table 4) to obtain a site-specific prediction model that can support precision farming. Typically, the development of predictive models only uses climate and soil data as the input model.

***Table 3**. Research parameters of farmers' socio-economic characteristics at rainfed rice fields on location-specific toposequence*

Parameters (units)	Lowland paddy fields	Midland paddy fields	Highland paddy fields	Total
Farmer's age (year)	20-70	31-70	31-70	20-70
Gender	Male and female	Male and female	Male and female	Male and female
Farming experience (years)	1-30	6-30	6-30	1-30
Education	No education–bachelor	No education – senior high school	No education–junior high school	No education–bachelor
Farmer's side job	No side job, labourer, civil servant	No side job, labourer, civil servant, trader	No side job, labourer, civil servant	No side job, labourer, civil servant, trader
Family involvement	Only the head of the family, several family members, and all family members	Only the head of the family, several family members, and all family members	Only the head of the family, several family members	Only the head of the family, several family members, and all family members

***Source:** developed by the authors of this study based on questionnaire data collection*

Table 4. Research parameters on farmers' land management characteristics at rainfed rice fields on location-specific toposequence				
Parameters (units)	**Lowland paddy fields**	**Midland paddy fields**	**Highland paddy fields**	**Total**
Seed variety	IR64, IR32, ciherang, Brengkele, Liwu, Mlati, Inpari 43, Mapan, Mekong and Sugal	IR64, IR32, ciherang, bengawan, mentik wangi, umbul-umbul, and barito	IR64, ciherang, bengawan and cepogo	IR64, IR32, ciherang, Brengkele, bengawan, Liwu, Mlati, Inpari 43, HT, umbul-umbul, Mapan, Mekong, Sugal, barito and cepogo
Cropping method	Jajar legowo and conventional	Jajar legowo and conventional	Jajar legowo and conventional	Jajar legowo and conventional
Cropping technique	Semi-organic and conventional	Semi-organic and conventional	Semi-organic and conventional	Semi-organic and conventional
Cropping pattern	Paddy-paddy-paddy, paddy-paddy-secondary crops, paddy-secondary crops-secondary crops, paddy-paddy-fallow, paddy-paddy-horticulture,	Paddy-paddy-paddy, paddy-paddy-secondary crops, paddy-secondary crops-secondary crops, paddy-paddy-fallow, paddy-secondary crops-fallow, paddy-paddy-horticulture, paddy- horticulture-horticulture	Paddy-paddy-paddy, Paddy-paddy-secondary crops, paddy-secondary crops-secondary crops, paddy-paddy-fallow, paddy-secondary crops-fallow	Paddy-paddy-paddy, paddy-paddy-secondary crops, paddy-secondary crops-secondary crops, paddy-paddy-fallow, paddy-secondary crops-fallow, paddy-paddy-horticulture, paddy- horticulture-horticulture
Fertiliser type	Compost, manure, urea, ZA, TS, Phonska, SP36, TSP, KCl, NPK, dolomite, TSP	Compost, manure, urea, ZA, TS, Phonska, SP36, TSP, KCl, NPK	Compost, manure, urea, ZA, TS, Phonska, SP36, TSP, KCl, NPK	Compost, manure, urea, ZA, TS, Phonska, SP36, TSP, KCl, NPK, dolomite, TSP
Pest disease control	No pest disease control, preventive, chemical, mechanical	No pest disease control, chemical, mechanical	No pest disease control, preventive, chemical, mechanical	No pest disease control, preventive, chemical, mechanical
Yield (tonnes/ha)	0.62-4.34	0.8-4	0.91-3.57	0.62-4.34

***Source:** developed by the authors of this study based on questionnaire data collection*

Machine learning. The BNN analysis results used 87 data samples, with 22 input parameters including 4 indicators: soil, climate, farmers' socio-economic data, and land management. The data was operated and tested using different numbers of neurons to get the most effective results and approach the factual results. The framework of the analysis conducted in this study is presented in Figure 2. The data was subsequently analysed using Pareto to optimise and identify the model determinants.

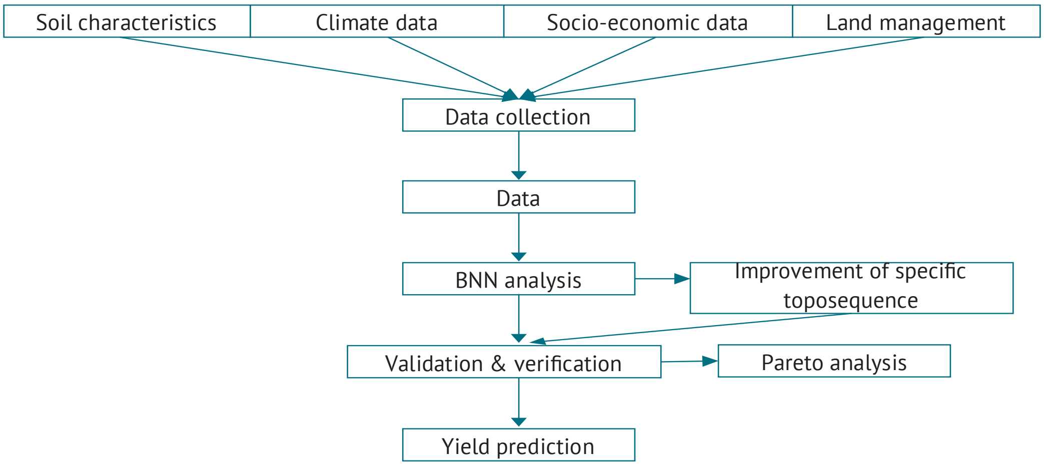


***Figure 2**. Framework of yield estimation using machine learning (Bayesian and Pareto approach)*
***Source:** developed by the authors of this study*

Six mountainous toposequences were used to create the forecast model for rainfed rice yields, using local and field data to produce a precise model. The BNN prediction model was made based on specific altitude characteristics to increase accuracy at particular places.

Model accuracy analysis. The capability and reliability of the predictive model were assessed using accuracy analysis, which was also employed to evaluate the model's performance. During the study, the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE) were calculated. The following equation was used for this evaluation:

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

$$\mathrm{MAPE} = \sum_{i=1}^{n}\left|\frac{y_i - y}{y_i}\right| \times 100\%, \tag{2}$$

where y_i is the prediction result; y is the average factual yield; n is the amount of data sample.

Measurement with root mean square error (RMSE) involved the analysis to measure the magnitude of error in the prediction results. The smaller the RMSE value, the more accurate the prediction result. Meanwhile, measurement with mean absolute percentage error (MAPE) involved statistical measurement to see the accuracy of estimates in forecasting methods.

RESULTS

Machine learning. The trial-and-error method frequently starts with choosing the ideal weights, the number of hidden layers, and the neurons in the hidden layers (Fig. 3). The backpropagation algorithm's control settings often depend on the procedure of trial and error. The operation is analogous to the human brain. Numerous linked neurons translate input into new, useful information. Regardless of the magnitude of the input data being processed, NN systems can be trained on massive volumes of data rapidly and effectively. Additionally, NN creates patterns and trends for analysis that are unachievable for humans and ensures the extraction of pertinent data from extremely complicated data sets.

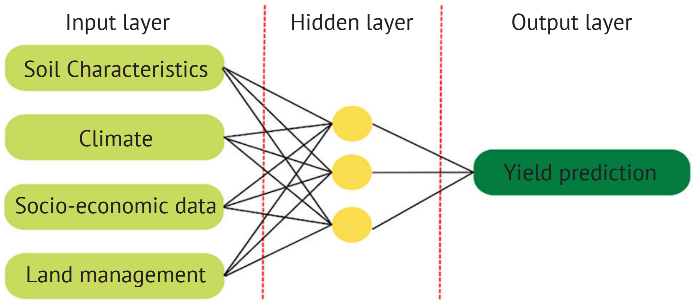


***Figure 3**. Schematic of prediction yield using BNN*

***Source:** developed by the authors of this study*

The study employed an efficient model comprising three neurons, based on the findings of the test. The test involved ten different input neurons (Fig. 3). Table 5 presents trial and error testing using ten different neurons, with the lowest RMSE of 0.442 and the most outstanding R^2 score of 0.840.

***Table 5**. Comparison of model accuracy and the number of neurons employed*

Toposequence	Parameters	Number of neurons									
		1	2	3	4	5	6	7	8	9	10
Total	RMSE	0.701	0.607	0.442*	0.696	0.704	0.704	0.704	0.705	0.704	0.704
	R^2	0.519	0.638	0.840*	0.525	0.513	0.512	0.512	0.511	0.512	0.512

***Note:** RMSE – smaller values are better; R^2 – greater values are better; * – best values*

***Source:** developed by the authors of this study*

Step 1: BNN analysis. The accuracy of yield projections at certain places was greater than the overall yield. The findings of this study revealed that predictions made using location-specific altitude data produced satisfactory outcomes, with the three models' coefficient of determination over 0.9. The ideal number of neurons for the models at the three toposequences varied, with highland paddy fields requiring 8 neurons, midland paddy fields – 1 neuron, and lowland paddy fields – 3 neurons. The results of the model accuracy comparison are presented in Figure 4.

As demonstrated in Figure 4, the three revisions of the toposequence-based model demonstrated an increase in performance, with the RMSE and R^2 values showing extremely satisfactory results when compared to when the model was built using total. Pearson values in all four models were more significant than 0.9, with the midland paddy fields prediction model having the greatest value. The total model value was 0.8396 based on the coefficient of determination, whereas the value of the three model modifications was above 0.99 or remarkably close to the factual value. Analogously to the RMSE and MAPE values, the evaluation value of the error of this model was low in the lowland paddy fields, midland paddy fields, and highland paddy fields models when compared to the total model with RMSE (0.4421) and MAPE (21.95%). This demonstrates that when the prediction model was modified based on the accuracy test value, the Pearson and R^2 values were close to 1, while the RMSE and MAPE values were close to 0.

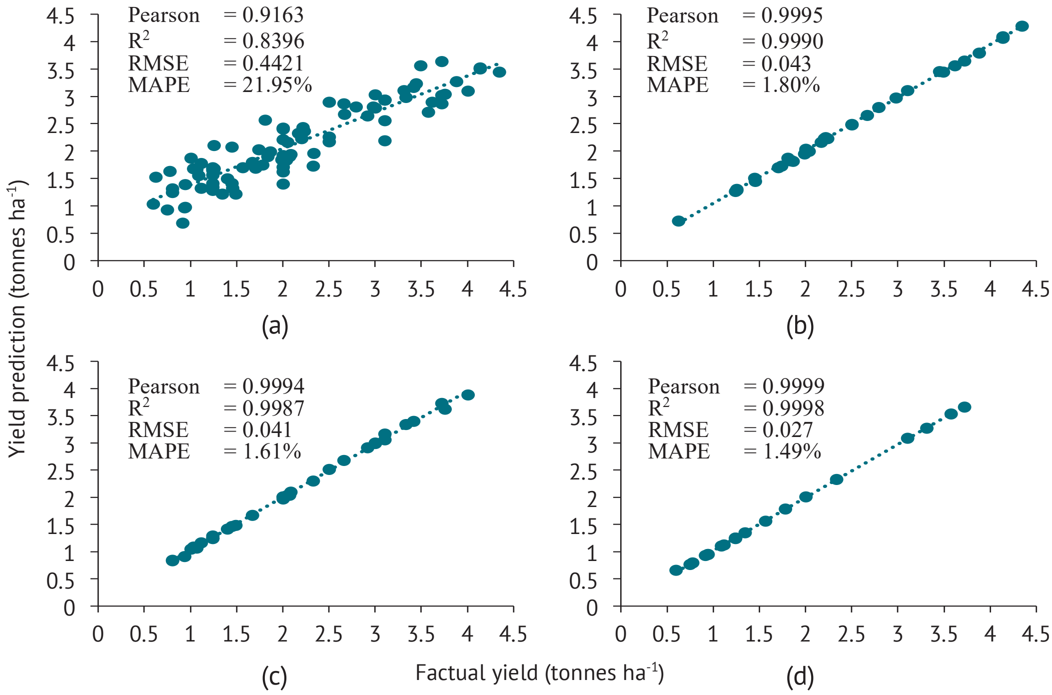


***Figure 4**. Location-specific crop yield prediction accuracy (a) Total; (b) Lowland paddy fields; (c) Midland paddy fields; (d) Highland paddy fields*

***Source:** developed by the authors of this study*

Table 6 depicts trial-and-error testing with ten different neurons, with the variation in toposequence indicating a different number of neurons used to construct a prediction model. According to M.K.A. Kadir *et al.* (2015), testing neurons using the trial-and-error approach is critical since this method can identify a model with the ideal composition of neurons and parameters. According to O. Bazrafshan *et al.* (2022), in machine learning models, the conditions in the hidden layer heavily influence the outcomes of the prediction models formed; the results of modifications and tests in Tables 5 and 6 demonstrate that the number of neurons in a different hidden layer produces different performance.

***Table 6**. Comparison of the number of effective neurons at each toposequence*

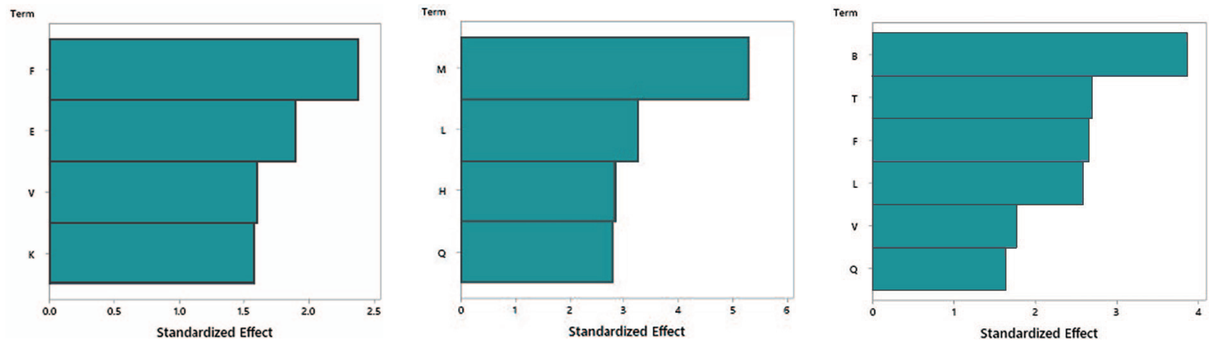
Toposequence	Parameters	Number of neurons									
		1	2	3	4	5	6	7	8	9	10
lowland paddy fields	RMSE	0.946	0.946	0.043*	0.946	0.946	0.946	0.946	0.946	0.946	0.946
	R^2	0.580	0.587	0.999*	0.580	0.585	0.588	0.588	0.588	0.576	0.588
midland paddy fields	RMSE	0.041*	0.089	0.093	0.090	0.095	0.105	0.104	0.098	0.104	0.107
	R^2	0.999*	0.994	0.993	0.993	0.992	0.990	0.991	0.992	0.991	0.990
highland paddy fields	RMSE	0.032	0.034	0.028	0.034	0.034	0.027	0.033	0.026*	0.029	0.034
	R^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000*	1.000	1.000

***Note:** RMSE – smaller values are better; R^2 greater values are better; * – best values*

***Source:** developed by the author of this study*

The results of the analysis by location classification are presented in Table 6 and Figure 4, demonstrating that the correlation values were highly significant and closely related to the analysis of input data in specific locations. This proves that presenting data based on specific altitude zones simplifies decision-making for users on issues related to agriculture. They show the same significance of correlation between lowland, midland, and highland rice fields. In this case, this was factored in when calculating the amount of input data that needed to be provided during the cultivation phase to obtain the best yields.

Step 2: Pareto analysis. According to the findings of this study, the rice harvest prediction model based on particular heights produces greater accuracy. According to the Pareto analysis's findings, each elevation criterion had a unique effect standard, as presented in Figure 5. The available P parameter was the factor that had the most significant influence according to the Pareto analysis results at a lowland paddy field. The farmer's education determines the prediction results for midland paddy fields, whereas the slope factor is the critical element for highland paddy fields.



(a) (b) (c)

A	Type of soil	G	Available K	M	Education	R	Cropping method
B	Slope	H	Rainfall	N	Farming experience	S	Cropping technique
C	pH	I	Elevation	O	Farmer side job	T	Cropping pattern
D	Organic C	J	Evapotranspiration	P	Family involvement	U	Fertiliser type
E	Total N	K	Farmer's age	Q	Seed variety	V	Pest disease control
F	Available P	L	Gender				

***Figure 5**. Pareto analysis at each toposequence: (a) Lowland paddy fields; (b) Midland paddy fields; (c) Highland paddy fields*

***Source:** developed by the authors of this study*

The essential parameters – Available P, Total N, Pest disease control, and Farmer age, – presented in Figure 5, are what the lowland paddy fields prediction model depended on. However, only Available P had a p-value < 0.05 (Table 7). In the lowland paddy fields, primarily Available P, the data distribution varied, dominated by levels of Available P (< 4 ppm) with 41.2% (Table 8). Available P in lowland paddy fields, which had the greatest average yields, was at levels of > 15 ppm with 2.97 tonnes ha^{-1}. This shows that elevated levels of Available P can increase rice yields in rainfed rice fields. Based on research by S. Supriyadi *et al.* (2022), available P is closely related to the soil quality index, especially in paddy fields. This shows that Available P in lowland paddy fields is one of the determining factors for this toposequence.

***Table 7**. The p-value of each parameter classified by a specific toposequence*

Toposequence	Parameter	p-value
Lowland paddy fields	Available P	0.024
Midland paddy fields	Rainfall	0.008
	Gender	0.003
	Education	0.000
	Seed variety	0.009
Highland paddy fields	Slope	0.002
	Available P	0.020
	Gender	0.023
	Cropping pattern	0.018

***Source:** developed by the authors of this study*

Table 8. Prediction determinants of the rainfed lowland rice harvest for each toposesequence				
Toposequence	**Parameter**	**Parameter class**	**Percentage (%)**	**Yield (ton ha^{-1})**
Lowland paddy fields	Available P (ppm)	< 4	41.2	2.23
		4-8	14.7	2.17
		8-10	14.7	2.36
		10-15	5.9	2.33
		> 15	23.5	2.97
Midland paddy fields	Rainfall (mm year^{-1})	<1500	0.0	0.00
		1,500-2,000	15.2	2.79
		2,000-2,500	39.4	2.44
		2,500-3,000	9.1	1.39
		>3,000	36.4	1.64
	Farmers' gender	Male	84.8	2.19
		Female	15.2	1.65
	Farmers' education	No education	3.0	1.24
		Elementary school	66.7	2.39
		Junior high school	24.2	1.36
		Senior high school	6.1	2.36
		Bachelor's degree	0.0	0.00
	Seed variety	Barito	3.0	3.10
		Ciherang	27.3	2.38
		IR32	6.1	1.32
		IR64	57.6	1.96
		Mentik wangi	3.0	3.00
		Umbul-umbul	3.0	2.05
Highland paddy fields	Slope (%)	0–8	60.0	3.57
		8–15	35.0	1.93
		15–25	5.0	1.37
		25–40	0.0	0.00
		>40	0.0	0.00
	Available-P (ppm)	<4	55.0	1.29
		4–8	10.0	2.11
		8–10	20.0	2.02
		10–15	5.0	2.00
		>15	10.0	2.54
	Farmers' gender	Male	90.0	2.31
		Female	10.0	1.61
	Cropping pattern	paddy-paddy-paddy	10.0	2.33
		paddy-paddy-fallow	15.0	2.68
		paddy-paddy-secondary crop	35.0	1.63
		paddy-secondary crop-secondary crop	35.0	1.17
		Paddy-secondary crop-fallow	5.0	1.24

Source: *developed by the authors of this study*

In contrast, the midland paddy fields model used four parameters: rainfall, education level, gender, and seed type (Fig. 5b), with all parameters having a p-value < 0.05 (Table 7). Based on climatic characteristics, rainfall is one of the determining factors in rice cultivation in midland paddy fields. Based on the land suitability guidelines for food crop cultivation, rainfed rice is optimally cultivated with rainfall conditions of 1,500-2,000 mm year^{-1} (Wahyunto *et al.*, 2016). This condition is under the distribution of analytical data in this study, where land with rainfall of 1,500-2,000 mm year^{-1} had the greatest average yield of 2.79 tonnes ha^{-1}. Meanwhile, land with rainfall above 2,500 mm year^{-1} had an average yield of 1.64 tonnes ha^{-1}. The use of seed varieties in the paddy fields medium also varied. In total, 6 types of varieties were cultivated in the medium of paddy fields. IR64 dominated the existing seed varieties by as much as 57.6%, while the greatest average yield was for the Barito variety, which had 3.10 tonnes ha^{-1}.

Apart from climatic factors and seed varieties, the midland paddy fields also had a significant correlation with the socio-economic aspects of farmers in the midland paddy fields, especially gender and education. Farmers managing rainfed lowland rice fields were dominated by men, reaching 84.8%, while women accounted for 15.2%. Additionally, the yields on land managed by men had greater yields, with 2.19 tonnes ha^{-1}, against those managed by women, with 1.65 tonnes ha^{-1}. Whereas in education, the distribution of data on midland paddy fields varied widely, starting from no education (3%), elementary school (66.7%), junior high school (24.2%), and senior high school (6.1%). The average yield based on the highest education classification in the elementary school education class was 2.39 tonnes ha^{-1} and senior high school 2.36 tonnes ha^{-1}. In comparison, no education parameter resulted in the lowest indicators, with an average yield of 1.24 tonnes ha^{-1}.

The highland paddy fields based on Pareto analysis had six key parameters, namely Slope, cropping pattern, Available P, Gender, Pest disease control, and Seed variables that influenced the prediction model, with the four parameters having a p-value < 0.05. According to the findings of the p-value test in Table 7, like lowland paddy fields, Available P in highland paddy fields had a significant correlation with crop yields. Available P yields are directly proportional to yields, as presented in Table 8. Available P levels < 4 ppm had the lowest yields with 1.29 tonnes ha^{-1}, while Available P levels > 15 ppm were the greatest with 2.54 tonnes ha^{-1}. Gender also had a significant correlation with rice yields, as established for midland paddy fields. In the highland paddy fields, male farmers accounted for 90%, with an average harvest of 2.31 tonnes ha^{-1}, while only 10% were women with an average harvest of 1.61 tonnes ha^{-1}.

Slope was one of the predictive determinants with the most striking difference in each class on highland paddy fields (Table 8). The analysis results revealed that a slope of 0-8% dominated the area with 60% and had the greatest average yield, which reached 3.57 tonnes ha^{-1}, while the distribution of data which had a slope class of 8-15% with an average harvest of 1.93 tonnes ha^{-1}, and 15-25% with an average harvest of only 1.37 tonnes ha^{-1}. Slope is one of the critical points in every crop cultivation, including rice. According to N. Takeda *et al.* (2019), cultivation of rice on steep slopes poses significant challenges in the management of irrigation and groundwater resources. At the same time, according to T. Tu *et al.* (2023), rice cultivation in the highlands with steep slopes has many obstacles, such as management, difficulty in mechanisation, and distribution of fertiliser inputs.

DISCUSSION

The tests conducted in this study, which applied four principal input parameters and three neurons in the hidden layer, yielded the most effective results among all the network configurations assessed. As demonstrated by L. Mou *et al.* (2022), Bayesian Neural Networks (BNN) were capable of simultaneously producing predicted outcomes and measurement uncertainties, thereby indicating the deviation of input data from the training data distribution. This approach, as highlighted by S.K. Bal *et al.* (2022), effectively mitigated the negative effects caused by out-of-distribution (OOD) data and prevented erroneous predictions, even with a single model parameter. The development of the BNN model in the present study was based on the specific toposequence of the farmer's land to enhance the accuracy of machine learning analysis. According to D. Paudel *et al.* (2021) regional and location-specific input data were shown to support the construction of reliable predictive models. The model's accuracy along defined toposequences corresponded closely with observed values, illustrating the high effectiveness of the predictive modelling approach. T. Hengl *et al.* (2018) recognised modern machine learning methods as powerful, data-driven tools for predicting soil properties and landscape variability. Moreover, A.M. J.-C. Wadoux *et al.* (2023) found that the integration of high-resolution, site-specific data – including environmental and management variables – substantially reduced prediction uncertainty and enhanced model robustness. This finding underscored the significance of tailoring modelling strategies to various elevation zones within a toposequence, as each zone was influenced by distinct factors.

It was therefore clear that individual elevation categories required separate modelling frameworks to account for their specific determinants. T. Van Klompenburg *et al.* (2020) demonstrated that yield prediction models incorporating site-specific factors markedly reduced errors. Furthermore, according to D. Paudel *et al.* (2022), such models provided stakeholders with improved insights and decision-making tools for policy development and strategic planning in agriculture. The prediction model developed using elevation categories followed an analogous structure, beginning with yield values and identifying key production determinants. According to Z.C. He *et al.* (2022), yield prediction methods generally required site-specific measurements due to diverse local conditions such as soil composition, climate variability, socio-cultural norms, and farmers' management practices. S. Jeong *et al.* (2022) observed that large-scale forecasts at the district or city level tended to be less sensitive to rice yield variations, owing to the compounding effects of climate change, soil heterogeneity, and inconsistent land management. Conversely, X. Yue *et al.* (2022) highlighted the significance of landscape-level predictions, as they provided deeper insights into crop responses under specific environmental and agronomic practices.

The Bayesian estimation approach employed in the present study contributed to the broader modelling efforts aimed at explaining long-term trends in Indonesia's agricultural productivity, as previously conceptualised in the literature. Specifically, it added nuance to existing research stock models by incorporating the complexity of Java's toposequences. Moreover, the current study extended this framework by introducing uncertainty analysis into the process of forming new research capital stocks from agricultural research investments. Although the unpredictable nature of individual research outcomes was already recognised within Indonesia's agricultural science system, this study offered empirical evidence to substantiate that claim. This uncertainty, as demonstrated in the findings, underscored the necessity of maintaining a diverse an d adaptable research portfolio. Supporting varied productivity-enhancing innovations and promoting flexible, incentive-based policies in agricultural science and technology were recommended strategies to address this challenge effectively. Understanding how neural networks learned was also considered essential. According to standard practice, several visualisation techniques were employed, including intermediate activation visualisation, convolution filter mapping, heat maps, and saliency diagrams. These methods, as shown in the study, illustrated how the model internally represented visual environments and how its feature extraction evolved across neural layers – enhancing compositional awareness, class discrimination, and yield prediction performance. Furthermore, visualisation tools helped in identifying model weaknesses and guiding model refinement.

The Bayesian approach also enabled the identification of critical limiting factors, which could inform policy recommendations for agricultural communities and governmental bodies aiming to strengthen and sustain the agrarian sector in pursuit of food security. Finally, W.S. Dewi *et al*. (2022) highlighted that agriculture is a cornerstone of global resource utilisation, consuming approximately 70% of surface and groundwater and 30% of global energy output, emphasising its centrality in food and energy supply systems. The creation of this framework for evaluating the performance of farming systems could make a critical contribution to the current study by presenting a systematic performance appraisal model for the distribution system using a Bayesian neural network. Previously, the development of such frameworks focused on increasing demand-oriented management of on-farm agricultural systems. Notably, the paradigm is limited to agricultural distribution systems as a critical element of off-farm agrarian management. The presented framework can be used as a decision-support tool to prioritise modernisation solutions for the farming distribution system. It also aids the yielding process and governance in determining the technical implications of building these systems and the environmental consequences.

CONCLUSIONS

The presented study achieved high accuracy in predicting rainfed rice yields in the tropical monsoon climate of central Java, Indonesia, by combining machine learning methods, specifically Bayesian neural networks (BNN), and Pareto analysis. Models adapted to elevation toposequences showed a significant advantage over general approaches, demonstrating a coefficient of determination (R^2) of up to 0.999, while the general model only reached 0.8396. The error indicators (RMSE = 0.4421; MAPE = 21.95%) were low, while Pearson's correlation coefficients exceeded 0.90, confirming the high reliability of the models. The study identified specific factors that determine yield in different agroecological conditions. In lowland and mountainous areas, the most significant factor was the content of available phosphorus in the soil: at P levels > 15 ppm, the yield reached 2.97 tonnes ha^{-1} in lowlands and 2.54 tonnes ha^{-1} in highlands. In mid-mountain conditions, climatic (precipitation 1,500-2,000 mm/year – 2.79 tonnes ha^{-1}) and socio-economic factors played a decisive role: level of education, gender of the farmer, and choice of seed variety. It was found that male farmers, as well as those with at least a basic education, achieved greater yields.

Furthermore, the slope indicator proved to be critical for mountainous areas: plots with a slope of 0-8% had the greatest average yield (3.57 tonnes ha^{-1}), while on slopes of 15-25% it decreased to 1.37 tonnes ha^{-1}. This confirmed scientific observations regarding the limitations in mechanisation, irrigation, and fertiliser application on steep slopes. The findings confirmed the effectiveness of integrating environmental, socio-economic, and technological parameters into precision agriculture models. The high sensitivity of models to local conditions allows optimising the agrotechnical solutions, increasing resource efficiency, and supporting for sustainable agricultural development in conditions of climate instability. Prospects for further research include conducting long-term field observations involving dynamic agroclimatic changes, expanding the socio-economic database of farms, developing integrated resource management models based on remote sensing of the Earth, and adapting artificial intelligence to decision-making systems in precision farming.

ACKNOWLEDGMENTS

None.

FUNDING

This study received no funding.

CONFLICT OF INTEREST

The authors of this study declare no conflict of interest.

REFERENCES

[1] Bal, S.K., Sandeep, V.M., Kumar, P.V., Rao, A.V.M.S., Pramod, V.P., Manikandan, N., Rao, C.S., Singh, N.P., & Bhaskar, S. (2022). Assessing impact of dry spells on the principal rainfed crops in major dryland regions of India. *Agricultural and Forest Meteorology*, 313, article number 108768. doi: 10.1016/j.agrformet.2021.108768.

[2] Bazrafshan, O., Ehteram, M., Moshizi, Z.G., & Jamshidi, S. (2022). Evaluation and uncertainty assessment of wheat yield prediction by multilayer perceptron model with bayesian and copula bayesian approaches. *Agricultural Water Management*, 273, article number 107881. doi: 10.1016/J.AGWAT.2022.107881.
[3] Bremner, J.M., & Mulvaney, C.S. (1982). Nitrogen – total. In A.L. Page (Ed.), *Methods of soil analysis: Part 2. Chemical and microbiological properties* (595-624). Madison, WI: American Society of Agronomy and Soil Science Society of America. doi: 10.2134/agronmonogr9.2.2ed.c31.
[4] Dewi, W.S., Romadhon, M.R., Amalina, D.D., & Aziz, A. (2022). Paddy soil quality assessment to sustaining food security. *IOP Conference Series: Earth and Environmental Science,* 1107, article number 012051. doi: 10.1088/1755-1315/1107/1/012051.
[5] Din, N. M. U., Assad, A., Dar, R. A., Rasool, M., Sabha, S. U., Majeed, T., Islam, Z. U., Gulzar, W., & Yaseen, A. (2024). RiceNet: A deep convolutional neural network approach for classification of rice varieties. *Expert Systems with Applications*, *235*, 121214. https://doi.org/10.1016/j.eswa.2023.121214
[6] ESRI. (2022). GIS software for mapping and spatial analytics. Retrieved from https://www.esri.com/en-us/home.
[7] He, Z.C., Huo, S.L., Li, E., Cheng, H.T., & Zhang, L.M. (2022). Data-driven approach to characterize and optimize properties of carbon fiber non-woven composite materials. *Composite Structures*, 297, article number 115961. doi: 10.1016/j.compstruct.2022.115961.
[8] Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 9, article number e11088.
[9] Indonesian Agricultural Assembly and Modernization Agency. (2005). *Chemical analysis of soil, plants, water, and fertilizers*. Retrieved from https://repository.pertanian.go.id/server/api/core/bitstreams/77f52e6b-6a13-48bc-96d1-d6a35025d793/content.
[10] Jackson, M.L. (1973). *Soil chemical analysis*. Prentice Hall of India.
[11] Jeong, S., Ko, J., & Yeom, J.M. (2022). Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea. *Science of The Total Environment*, 802, article number 149726. doi: 10.1016/j.scitotenv.2021.149726.
[12] Kadir, M.K.A., Ayob, M.Z., & Miniappan, N. (2015). Wheat yield prediction: Artificial neural network based approach. In *2014 4th International conference on engineering technology and technopreneuship (ICE2T)* (pp. 161-165). Kuala Lumpur: IEEE. doi: 10.1109/ICE2T.2014.7006239.
[13] Liu, S., Wang, X., Liu, M., & Zhu, J. (2017). Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics*, 1(1), 48-56. doi: 10.1016/j.visinf.2017.01.006.
[14] Moeletsi, M.E., Walker, S., & Hamandawana, H. (2013). Comparison of the Hargreaves and Samani equation and the Thornthwaite equation for estimating dekadal evapotranspiration in the Free State Province, South Africa. *Physics and Chemistry of the Earth*, 66, 4-15. doi: 10.1016/j.pce.2013.08.003.
[15] Mou, L., Liang, L., Gao, Z., & Wang, X. (2022). A multi-scale anomaly detection framework for retinal OCT images based on the Bayesian neural network. *Biomedical Signal Processing and Control*, 75, article number 103619. doi: 10.1016/j.bspc.2022.103619.
[16] Olsen, S.R., Cole, C.V., Watanabe, F.S., & Dean, L.A. (1954). *Estimation of available phosphorus in soils by extraction with sodium bicarbonate* (USDA Circular No. 939). Washington: U.S. Government Printing Office.
[17] Pant, J., Pant, R.P., Singh, M.K., Singh, D.P., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*, 46, 10922-10926. doi: 10.1016/j.matpr.2021.01.948.
[18] Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I.N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187, article number 103016. doi: 10.1016/j.agsy.2020.103016.
[19] Paudel, D., Boogaard, H., de Wit, A., van der Velde, M., Claverie, M., Nisini, L., Janssen, S., Osinga, S., & Athanasiadis, I.N. (2022). Machine learning for regional crop yield forecasting in Europe. *Field Crops Research*, 276, article number 108377. doi: 10.1016/j.fcr.2021.108377.
[20] Qian, J., Zhao, H., Wang, X., Wang, T., Feng, Z., Cao, C., Li, X., & Zhang, A. (2025). Declining suitability for conversion of drylands to paddy fields in Northeast China: Impact of future climate and socio-economic changes. *Geography and Sustainability*, 6(1), article number 100199. doi: 10.1016/j.geosus.2024.05.004.
[21] Supriyadi, S., Ustiatik, R., Mukti, B., Minardi, S., Widijanto, H., & Sakti, M.B.G. (2022). Soil quality status under Hazton's paddy farming: A case study in Banyumas Regency, Indonesia. *SAINS TANAH – Journal of Soil Science and Agroclimatology*, 19(2), 123-131. doi: 10.20961/STJSSA.V19I2.58375.
[22] Takeda, N., López-Galvis, L., Pineda, D., Castilla, A., Takahashi, T., Fukuda, S., & Okada, K. (2019). Evaluation of water dynamics of contour-levee irrigation system in sloped rice fields in Colombia. *Agricultural Water Management*, 217, 107-118. doi: 10.1016/J.AGWAT.2019.02.032.

[23] Thomas, G.W. (1982). Exchangeable cations. In A.L. Page (Ed.), *Methods of soil analysis: Part 2. Chemical and microbiological properties* (pp. 159-165). Madison, WI: American Society of Agronomy and Soil Science Society of America. doi: 10.2134/agronmonogr9.2.2ed.c9.
[24] Tu, T., *et al.* (2023). Methods and experiments for collecting information and constructing models of bottom-layer contours in paddy fields. *Computers and Electronics in Agriculture*, 207, article number 107719. doi:10.1016/j.compag.2023.107719.
[25] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, article number 105709. doi: 10.1016/j.compag.2020.105709.
[26] Velmurugan, P., Kannagi, A., & Varsha, M. (2023). Superior fuzzy enumeration crop prediction algorithm for big data agriculture applications. *Materials Today: Proceedings*, 81(2), 112-117. doi: 10.1016/j.matpr.2021.02.578.
[27] Wadoux, A.M.J.-C., Minasny, B., & McBratney, A.B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, article number 103359. doi: 10.1016/j.earscirev.2020.103359.
[28] Wahyunto, *et al.* (2016). *Technical guidelines for land suitability assessment for strategic agricultural commodities at a semi-detailed level on a scale of 1:50,000*. Retrieved from https://repository.pertanian.go.id/handle/123456789/20342.
[29] Walkley, A., & Black, I.A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, 37(1), 29-38. doi: 10.1097/00010694-193401000-00003.
[30] Yue, X., Li, Z., Li, H., Wang, F., & Jin, S. (2022). Multi-temporal variations in surface albedo on Urumqi glacier No.1 in Tien Shan, under arid and semi-arid environment. *Remote Sensing*, 14(4), article number 808. doi: 10.3390/rs14040808.
[31] Zhang, J., Jing, X., Song, X., Zhang, T., Duan, W., & Su, J. (2023). Hyperspectral estimation of wheat stripe rust using fractional order differential equations and Gaussian process methods. *Computers and Electronics in Agriculture*, 206, article number 107671. doi: 10.1016/j.compag.2023.107671.

Прогнозування урожайності рису на богарних землях у мусонній тропічній зоні Яви, Індонезія з використанням машинного навчання

Ажарі Різал
Магістр
Політехнічний інститут ЛПП Джок'якарта
55222, вул. ЛПП Балапан, 1А, м. Джок'якарта, Індонезія
https://orcid.org/0009-0006-7607-8746

Абдул Азіз
Магістр
ЛПП Агро Нусантара
55222, вул. ЛПП, 01, м. Джок'якарта, Індонезія
http://orcid.org/0000-0002-9688-3180

Комаріах
Доктор прикладних наук
Університет Себелас Марет
57126, вул. Ір. Сутабі, 36, м. Суракарта, Індонезія
https://orcid.org/0000-0001-7704-0754

Мухаммад Різкі Ромадхон
Магістр
Університет Ламбунг Мангкурат
70714, вул. Дженерал А. Яні, КМ.36, м. Банджарбару, Індонезія
http://orcid.org/0000-0002-7929-769X

Мухамад Мустангін
Магістр
Політехнічний інститут ЛПП Джок'якарта
55222, вул. ЛПП Балапан, 1А, м. Джок'якарта, Індонезія
https://orcid.org/0009-0001-5999-4519

Анотація. Вирощування рису на богарних землях у тропічних районах Яви, Індонезія, що знаходяться під впливом мусонних дощів, ускладнюється через нестачу поживних речовин, непередбачуваність опадів та обмежені інвестиції в сільське господарство, що призводить до коливань врожайності. Метою цього дослідження було розроблення точної моделі прогнозування врожайності рису з використанням машинного навчання, адаптованого до конкретних топологічних послідовностей у Центральній Яві. Застосовано комплекс польових та лабораторних методів, заснованих на опитуваннях, з урахуванням кліматичних, ґрунтових, соціально-економічних та агроменеджментних змінних із 87 цільових точок відбору проб. Аналіз за допомогою баєсівських нейронних мереж (BNN) показав помірну точність (R^2 = 0,840; RMSE = 0,442), однак точність суттєво підвищувалась при адаптації моделей до категорій за висотою, досягаючи R^2 до 0,999. Для низинних полів найбільший вплив мала наявність доступного фосфору (P); у зонах середньої висоти ключовими чинниками були кількість опадів, стать, освіта та сорт насіння; для високогірних районів – ухил, доступний фосфор, стать і сівозміна. Аналіз Парето підтвердив ідентифікацію цих основних детермінант урожайності у кожному топосліді. Інтеграція методів BNN і Парето дозволила створити високоточну, локалізовану модель прогнозування. Робота довела, що адаптація моделей машинного навчання до агроекологічних зон за висотою покращує їхню ефективність і практичне застосування. Результати є особливо цінними для аграрних стейкґолдерів – зокрема для політиків, дорадчих служб і фермерів, які можуть використати прогностичні дані для оптимізації управління богарним рисівництвом в умовах змінного клімату

Ключові слова: аграрна стійкість; баєсівська нейронна мережа (BNN); продовольча безпека; аналіз Парето; точне землеробство