Applied Information System and Management (AISM)
Volume 8, (2) 2025, p. 297–304
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i2.46681

Cocoa Land Suitability Analysis Using ID3 Spatial Algorithm

Andi Nurkholis[1*], Ririn Wuri Rahayu[2], Andrey Ferriyan[3], Akfina Ni'mawati[4]

***Abstract*—Cocoa production in Indonesia encounters ongoing challenges due to declining plantation areas and suboptimal land utilization. This study applies the ID3 Spatial algorithm to evaluate land suitability for cocoa cultivation in Bogor Regency, West Java Province. The methodology integrates nine basic land characteristics, including elevation, drainage, relief, base saturation, cation exchange capacity, soil texture, soil pH, and mineral soil depth, derived from field surveys conducted by BBSDLP. Two classification models were developed and tested using spatial data preprocessing techniques. Model M1 was the baseline approach without constraints, while Model M2 incorporated a minimum planted area threshold of ≥1 ha. The results show notable performance differences between models. Model M2 achieved a reasonable accuracy of 87.27% compared to Model M1's 29.09%, with relief identified as the root node due to its higher gain value and reduced entropy. Classification results indicate that Bogor Regency's cocoa cultivation potential comprises 16,443 ha of S2 (moderately suitable) land and 231,018 ha of S3 (marginally suitable) land. The generated land suitability map may provide stakeholders with helpful guidance for identifying potential cultivation areas. The result suggests that artificial intelligence integration, specifically the ID3 spatial algorithm, could improve land suitability evaluation processes, potentially supporting more informed agricultural development decisions.**

***Index Terms*—Cocoa, ID3 spatial algorithm, land characteristics, spatial data.**

I. Introduction

Cocoa (*Theobroma cacao* L.) is a high-value crop that is an essential commodity in the global market, particularly in tropical and subtropical regions [1]. West Africa dominates global cocoa production, making it a cornerstone of the economies of producing countries, the livelihood of smallholder farmers, and the chocolate industry. In addition to its distinctive and delightful flavor, cocoa contains bioactive phytochemical compounds that offer various health benefits. These advantages have increasingly drawn global attention, positioning cocoa not only as a raw material for chocolate but also as a valuable commodity in the health and nutrition sectors [2].

Cocoa production in Indonesia faces significant challenges and has shown a declining trend. According to the Central Bureau of Statistics (BPS), there was a substantial surge in cocoa imports, with an increase of 119% monthly in January 2025. The total value of cocoa and its derivative imports rose sharply to US$304.41 million, up from US$140 million in December 2024 [3]. Amid this downward trend, Indonesia's cocoa production 2022 still reached 667.3 thousand tons, with more than half, approximately 385,981 tons, being exported. Ironically, despite being one of the world's largest cocoa producers, Indonesia still heavily imports raw cocoa beans to meet domestic industrial demand. This imbalance highlights that domestic production has yet to meet the needs of the processing industry, both in terms of quantity and quality. One of the main factors contributing to this situation is the continued decline in cocoa plantation area over the past decade, with an average annual decrease of 2.38%. In 2014, the total cocoa plantation area stood at 1.73 million hectares, but this figure has steadily decreased to just 1.39 million hectares by 2023 [4].

The decline in cocoa plantation area has not occurred without reason; one of the key drivers is the limited expansion of cultivation due to land characteristics less conducive to optimal development [5]. Every agricultural commodity, including cocoa, has specific land requirements for maximum yield. Without alignment between crop characteristics and land conditions, productivity is hindered. Therefore, land suitability mapping and analysis are crucial in cocoa cultivation [6]. The land's resource potential must be evaluated based on established standards to determine whether a particular area is suitable for cocoa farming. The Food and Agriculture Organization (FAO) classifies land suitability into four main categories: highly suitable (S1), moderately suitable (S2), marginally suitable (S3), and not suitable (N) [7]. Understanding this classification is essential for farmers and

Received: 28 May 2025; Revised: 01 July 2025; Accepted: 25 August 2025
*Corresponding author

[1]Andi Nurkholis, Department of Informatics, UPN Veteran Yogyakarta, Indonesia (e-mail: andinurkholis@upnyk.ac.id).
[2]Ririn Wuri Rahayu, Department of Geomatics Engineering, UPN Veteran Yogyakarta, Indonesia (e-mail: ririnwurirahayu@upnyk.ac.id).
[3]Andrey Ferriyan, Graduate School of Media and Governance, Keio University, Japan (e-mail: andrey@keio.jp).
[4]Akfina Ni'mawati, Department of Informatics, UPN Veteran Yogyakarta, Indonesia (e-mail: akfina.mhs@upnyk.ac.id).

stakeholders in designing more effective and sustainable cultivation strategies, enabling them to optimize yields even as plantation areas shrink.

Research on land suitability has been extensively developed, with the matching method being one of the most commonly used evaluation approaches [8]. With the advancement of technology, land suitability assessments have continued to evolve, particularly through the integration of Artificial Intelligence (AI) [9]. AI enables the resolution of complex multi-index decision-making problems, allowing the analysis of diverse parameters to produce more accurate and comprehensive evaluation results.

Previous studies have used AI methods such as Analytical Hierarchy Processing (AHP) and Fuzzy to evaluate the suitability of wheat land [10]. In addition, one of the AI-based algorithms that can be utilized in land suitability evaluation is Iterative Dichotomiser 3 (ID3) spatial algorithm. The main advantage of the ID3 spatial algorithm over other research methods lies in its ability to analyze the correlation of spatial data for each attribute or factor [11]. In land suitability assessments, which are inherently based on geographic references, considering spatial relationships such as location, proximity, and orientation is crucial [12]. This algorithm ensures that these factors are optimally integrated into the analysis, resulting in more accurate and relevant evaluations. ID3 spatial algorithm has been successfully applied in various studies, including for garlic with an accuracy of 94.34% [13], rice with 96.67% accuracy [14], and oil palm with an accuracy of 98.18% [15]. Given its proven high performance, the ID3 spatial algorithm is proposed as an optimal approach for developing a land suitability model for cocoa cultivation.

This study aims to identify the fundamental problem—specifically, the inadequacy of technology in managing land suitability and its impact on domestic production capacity—and explore the application of the ID3 spatial algorithm as a solution to generate a suitability model that enhances precision in determining optimal and sustainable cocoa development areas. The algorithm is designed to develop a predictive model that classifies land suitability levels based on soil characteristics, adapted to the specific requirements of cocoa cultivation. The data used are derived from field surveys conducted by BBSDLP, encompassing relevant environmental parameters. The final output of this model is a cocoa land suitability map, which can serve as a strategic guide for stakeholders in identifying the most promising and sustainable cultivation areas.

II. Research Method

A. Study Area

This study focuses on Bogor Regency, a strategic region in West Java Province that plays a vital role in the agricultural sector. With Cibinong as its capital, the regency comprises 40 districts and covers 299,070 hectares (ha), making it one of the regions with vast and diverse agricultural land potential [16]. Land suitability mapping conducted by the Indonesian Center for Agricultural Land Resources Research and Development (BBSDLP) indicates that cocoa is one of the region's priority commodities, alongside rice, maize, soybeans, red chili, and elephant grass [17]. This highlights the importance of Bogor Regency as a key area for land suitability-based agricultural development. Given its substantial potential, the regency not only serves as a strong candidate to become a role model in land suitability evaluation but also holds a strategic position in promoting more productive, sustainable, and data-driven agriculture in Indonesia.

B. Research Data

This study utilized primary data obtained from BBSDLP through field surveys conducted across Bogor Regency in 2017. Additionally, elevation data was sourced from the United States Geological Survey (USGS) in raster DEM format to ensure comprehensive environmental profiling. The research dataset comprises nine attributes organized into two main categories: explanatory layers and target layers. The explanatory layers encompass eight soil characteristics and properties that serve as primary factors for evaluating land suitability for cocoa cultivation. These attributes include elevation, drainage, relief, base saturation, cation exchange capacity, soil texture, soil acidity, and mineral soil depth. The target layer represents cocoa land suitability data mapped based on BBSDLP's field survey results. This land suitability classification system consists of four main categories: highly suitable (S1), moderately suitable (S2), marginally suitable (S3), and not suitable (N). The details of the attributes, including descriptions, data formats, and data sources, are presented in Table 1.

The research adopts a mixed-methods quantitative design structured into five core stages: (1) spatial data acquisition from authoritative sources (BBSDLP and USGS), (2) spatial data preprocessing and harmonization using PostgreSQL/PostGIS and ArcMap, (3) ID3 spatial model construction with two scenarios (with and without land area constraint), (4) evaluation of model performance using confusion matrix metrics, and (5) visualization and validation of classification outputs against BBSDLP maps. Each stage is methodologically interdependent, ensuring the traceability and repeatability of the classification process in future applications. Figure 1 illustrates the overall research workflow.

C. Data Preprocessing

Data preprocessing is a critical initial step in this study, aimed at ensuring the use of high-quality data [18], which is essential for producing an accurate model. Specifically, this stage focuses on the processing and integrating spatial data to optimize its use in the ID3 spatial classification [19]. In this research, data preprocessing was carried out in four main stages:

1) The first step involves integrating spatial and non-spatial data using the PostgreSQL Database Management System (DBMS). The data obtained from BBSDLP, which includes seven explanatory factors and cocoa land suitability data, consists of two formats: spatial data (vector) in shapefile format and non-spatial data in spreadsheet format. The spatial data refers to geometric information in polygons,

Applied Information System and Management (AISM)
Volume 8, (2) 2025, p. 297–304
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i2.46681

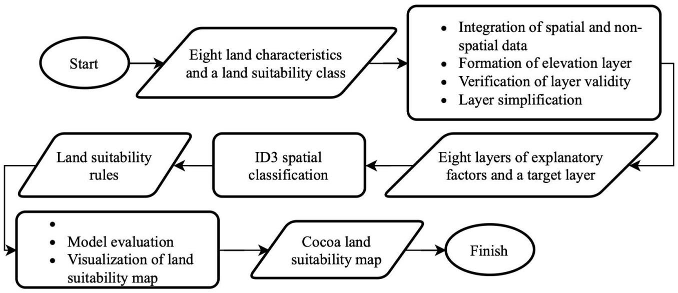

Fig. 1. Research stages.

while the non-spatial data contains attributes that can be linked through land map units (Satuan Peta Tanah, SPT). In this study, one SPT may encompass multiple polygons.

2) The next stage involves creating the elevation layer, which will be used as the eighth attribute in the classification process. The elevation data from the United States Geological Survey (USGS) is in raster format, whereas the ID3 spatial algorithm operates on vector data. Therefore, data conversion is necessary using the Digital Elevation Model (DEM), a three-dimensional representation of the Earth's surface based on elevation data. This process is carried out using ArcMap version 10.3.

3) The third step is to ensure that all layers used in this study are geometrically valid. Both the explanatory layers and the target layer are in polygon format, which is a common geometry type in spatial data. Therefore, a validity verification process is required to prevent self-intersection, a condition where a polygon intersects with itself. Polygons exhibiting self-intersection are considered non-compliant with OpenGIS standards and cannot be used in spatial relationship operations within ID3 spatial classification. If any invalid polygons are detected, they are corrected by trimming the affected small sections using ArcMap version 10.3.

4) The final stage of data preprocessing is simplifying the layer names and the attributes within each layer to facilitate easier use during the classification process. This simplification is carried out using spatial queries in the PostgreSQL DBMS. The expected outcome of this stage is that all layers are stored within a single database in tabular format, making them ready for use in ID3 spatial.

D. ID3 Spatial Classification

Classification analyzes data within a database to discover patterns or land suitability rules (S1, S2, S3, and N). One effective classification method for handling noisy data, which is a significant challenge in spatial data processing, is the ID3 spatial algorithm [20]. This technique is an enhancement of the ID3 algorithm and is often referred to as a spatial decision tree. ID3 spatial operates based on the decision tree concept, where each node represents an attribute, branches indicate attribute values, and leaves refer to a specific class [13]. More specifically, the spatial decision tree has three main characteristics: (1) each internal node functions as a decision point based on spatial data, (2) each branch represents the outcome of a test, and (3) each leaf represents a land suitability class. With this structure, ID3 spatial becomes a powerful tool for analyzing and classifying spatial data more accurately and efficiently.

The first step in building an ID3 spatial classification model is to perform a spatial relationship operation between the explanatory layer and the target layer, which in this study is in polygon format. The spatial relationship between layers enables the calculation of quantitative values such as the distance between points or the intersection area between two polygons. In this study, the spatial measure used is the intersection area between the explanatory and target layers, which serves as the basis for the classification process. Figure 2 shows the sequential steps involved in ID3 spatial classification. For instance, consider a layer L_i (explanatory layer) and a layer L_j (target layer) where $i \neq j$. For each feature r_i, where $i = 1,2, \dots, p$ and p represents the total number of layers in the set L_i, and $j = 1,2, \dots, q$, with q being the total number of layers in L_j

Table 1.
Research data

Attributes	Description	Format	Source
Elevation	Land height above sea level based on digital elevation model (DEM)	Raster	USGS
Drainage	The rate per location of water into the soil against air aeration in the soil	Vector	BBSDLP
Relief	Land slope is measured in %	Vector	BBSDLP
Base saturation	The number of bases (NH4OAc) present in 100g of soil sample	Vector	BBSDLP
Cation exchange capacity	Cation exchange capacity of clay fraction	Vector	BBSDLP
Soil texture	Terms in the distribution of fine soil particles with a size of <2 mm	Vector	BBSDLP
Soil pH	Soil pH value in the land	Vector	BBSDLP
Soil mineral depth	The depth of minerals in the soil layer	Vector	BBSDLP
Lan suitability class of cocoa	The level of suitability of a land area for cocoa cultivation is classified into three classes, namely S2 (moderately suitable), S3 (marginally suitable), and N (not suitable).	Vector	BBSDLP

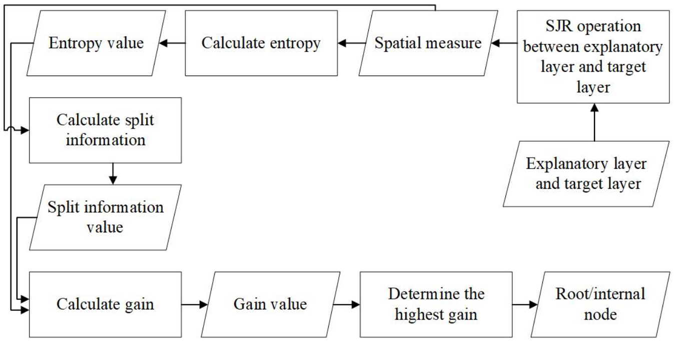

Fig. 2. Sequential steps involved in ID3 spatial classifications.

(which in this study is just one), the spatial measure for r_i, denoted as $SpatMes(r_i)$, can be calculated using a formula developed in previous research based on the Spatial Join Relation (SJR), as formulated in (1) [15].

$$SpatMes(r) = f\left(SpatMes(l_{i1} \cap l_{j1}), ..., SpatMes(l_{im} \cap l_{jn})\right) \quad (1)$$

The spatial measure is calculated using the function f, where f is the summation function, m represents the number of polygons in the explanatory layer, and n is the number of polygons in the target layer. The spatial relationship between L_i and L_j produces a new layer R, which is then further processed to obtain the combined SJR relationship of all features in L_i and L_j. In this case, r is a feature in R associated with feature p in L_i and feature q in L_j, as formulated in (2) [15].

$$SJR = \{(p, SpatMes(r), q)\} \quad (2)$$

After the $SpatMes$ is generated, the next step is to calculate entropy, which reflects the level of uncertainty in the information within the data. If the target layer S has l different class attributes $(C_1, C_2, ..., C_l)$, the entropy value of S indicates the amount of information required to classify the entire dataset, as formulated in (3) [15].

$$H(S) = -\sum_{i=1}^{l} \frac{SpatMes(S_{ci})}{SpatMes(S)} log_2 \frac{SpatMes(S_{ci})}{SpatMes(S)} \quad (3)$$

After obtaining the entropy values from the sample data, gain calculations are performed to assess the effectiveness of an attribute in the classification process. In this calculation, the number of target classes for a particular attribute is represented by $SpatMes(L(v_j, S))$, the total number of target classes is defined by $SpatMes(S)$, and the entropy value of the attribute is denoted by $H(L(v_j, S))$. This process utilizes split information, which ensures the optimal selection of characteristics in constructing the decision tree structure, as formulated in (4) [15].

$$H(S|L) = \sum_{j=1}^{q} \frac{SpatMes(L(v_j,S))}{SpatMes(S)} H(L(v_j, S)) \quad (4)$$

The final step is to calculate an attribute's gain. In this context, the entropy value is represented as $H(S)$, while the split information is denoted by$H(S|L)$. The gain is calculated using (5) [15].

$$Gain\ (L) = H(S) - H(S|L) \quad (5)$$

The attribute with the highest gain is selected as the root node in the ID3 spatial algorithm. The subsequent nodes in the tree are progressively filled with attributes with lower gain values until the final nodes (leaf nodes) represent the classification results. The ID3 spatial algorithm will terminate when one of the following two conditions is met:

1) *There is only one explanatory layer in the set L.*
   In this case, the algorithm assigns the leaf node based on the majority class obtained from SJR for both the best layer and the explanatory layer.
2) *The SJR for the best layer and the explanatory layer.*
   It contains only one common class c, which results in the algorithm labeling the leaf node with class c.

After implementing spatial ID3 classification modeling, an assessment was conducted to measure the effectiveness of the classification model in processing test data from Bogor Regency by evaluating its accuracy to ensure the reliability of the generated model. The evaluation metrics applied are based on previous studies that provide comprehensive guidelines regarding evaluation techniques in classification [21]. The analysis results were obtained through calculations using the confusion matrix formula outlined in (6), which serves as the reference for evaluating the model's performance.

$$Accuracy = \frac{tp+tn}{tp+fn+fp+tn} \quad (6)$$

where,
tp (true positive) = positive data that are correctly classified
tn (true negative) = positive data that are incorrectly classified
fp (false positive) = negative data that are correctly classified
fn (false negative) = negative data that are incorrectly classified

E. Cocoa Land Suitability Visualization

This phase converts the model's classification output into a more interpretable and comprehensible cocoa land suitability mapping. The generated classification regulations are implemented on soil characteristic data in Bogor Regency, creating a definitive spatial representation of areas with optimal potential for cocoa cultivation. The visualization process is executed using ArcMap 10.3, which facilitates high-resolution spatial mapping. This map serves not merely as an instrument for researchers and policymakers, but also as a strategic reference for more efficient and sustainable land management.

III. RESULT

A. Preprocessing Data Result

All explanatory layers along with the target layer are stored in a database within the PostgreSQL DBMS. For ease of data processing, the layer names and attributes of each layer were

Table 2.
Layers in PostgreSQL

Layer name	Amount of polygons	Class
Elevation (msal)	1,209	<100, 100-600, 601-700, 701-1600, 1601-1750, 1751-2000, >2000
Drainage*	57	Swift, good, slightly hamper, hamper
Relief (%)	187	Flat (0), slightly flat (1−3), slightly slope (4−8), slope (9−15), slightly steep (16−25), steep (26−40), very steep (>40)
Base saturation (%)	53	Low (20−35), medium (36−60), high (61−80), very high (>80)
Cation exchange capacity (cmol)	65	Low (5−16), medium (17−24), high (24−40), very high (>40)
Soil texture*	94	Very smooth, smooth, slightly smooth, medium, slightly rude, rude
Soil pH (°)	76	Acid (4.5−5.5), slightly acid (5.6−6.5), neutral (6.6−7.5)
Soil mineral depth (cm)	90	Very shallow (<25), shallow (25−50), medium (51-75), deep (76−100), very deep (>100)
Land suitability class of cocoa*	238	S2 (moderately suitable), S3 (marginally suitable), and N (not suitable).

*)Attributes do not have numeric values, but only class values

changed to shorter codes. The list of names and the number of polygons for each layer shown in Table 2.

B. Variation of Model Classification

The spatial ID3 classification technique was implemented using the Python programming language to generate two distinct models, aimed at deriving optimal rules based on spatial area condition approaches previously applied in earlier research [15]. The first model (M1) represents the original baseline model without any conditional constraints applied. The second model (M2) incorporates additional conditional limitations by introducing a minimum planted area threshold of ≥ 1 hectare within the SJR module. This spatial area constraint approach follows established morphological analysis principles where rules are modeled as networks of finite state transducers, ensuring systematic processing. The 1-hectare minimum threshold is established based on agricultural research indicating that the smallest viable plantation area for any commodity is 1 hectare [22]. Consequently, areas below this threshold are considered to have negligible or minimal impact in land suitability evaluation processes.

C. Model Evaluation

Both model outputs underwent validation testing using actual data sourced from BBSDLP. The detailed assessment findings are displayed in Table 3. The two tested models demonstrated significant differences in accuracy, as shown in Table 3. The first model (M1), implemented without constraints, achieved only 29.09% accuracy from 55 SPT data.

Table 3.
Model evaluation

Model	Root node	Amount of rules	Attributes not included	Accuracy
M1	Soil texture	62	-	87.27%
M2	Relief	53	Drainage	29.09%

In contrast, the second model (M2), which applied a minimum planted area constraint of ≥ 1 hectare, dramatically improved to 87.27% accuracy. This enhancement may be attributed to using information ratios, such as entropy and gain, in the model's decision-making process. By considering more relevant area parameters, M2 was able to generate superior class separation, thereby improving overall accuracy. The confusion matrix results of M2 on 55 SPT test data can be seen in Table 4.

Furthermore, M1's root node was soil texture, while M2 utilized relief as its root node. The selection of relief as the root node in M2 is explained by its higher gain value and lower entropy. Low entropy indicates that the relief attribute produces reduced uncertainty levels in data classification, enabling more apparent inter-class separation. This demonstrates that relief contains more relevant information, making it more effective in supporting land suitability analysis. Consequently, appropriate attribute selection is crucial for enhancing model performance in spatial analysis.

M1 generated 62 rules regarding rule quantity, while M2 produced 53. Although M1 had more rules, the soil texture-based root node tended to create broader and potentially less specific rules. Conversely, M2, with its relief-based root node, produced more focused and relevant rules, despite their

Table 4.
Confusion matrix of M2

SPT=55		Actual data			
		S2	S3	N	Unclassified
Predictive data	S2	6			0
	S3	1	32	4	0
	N		1	10	0
	Unclassifed			1	0

smaller quantity. This indicates that the M2 approach is more efficient in filtering the most significant attributes. Additionally, applying area constraint conditions assists the model in optimizing separation and improving the quality of information obtained from the data. The optimization method used in this study is known as the pruning technique, which has been previously utilized in an earlier study [23].

Regarding attribute involvement, M1 incorporated all attributes: elevation, drainage, relief, base saturation, cation exchange capacity, soil texture, soil acidity, and soil mineral depth. Meanwhile, M2 excluded the drainage attribute, which demonstrated less significant gain and entropy values in the model. The low entropy in M2 indicates that the selected attributes possess superior capabilities in reducing classification uncertainty, demonstrating more apparent inter-class separation. By focusing on more relevant attributes, M2 emphasizes the importance of attribute selection in spatial models to achieve optimal results.

Table 5.
Cocoa land suitability area

Class	Total area (ha)	
	BBSDLP	Model M2
S2, Moderately Suitable	18,849	16,443
S3, Marginally Suitable	217,392	231,018
N, Not Suitable	61,838	49,253
Unclassified	0	1,363

Overall, M2 demonstrated significant superiority compared to M1, both in terms of accuracy and efficiency. M2 successfully utilized more relevant attributes, producing more apparent class separation despite having fewer rules. This approach emphasizes the importance of attribute selection in spatial analysis, making M2 a more optimal model for improving information quality and classification accuracy. The following are five of 53 rules generated by M2:

1) *IF Relief is Slope AND Soil Texture is Rude THEN Class is S3, Marginally Suitable.*
2) *IF Relief is Slope AND Soil Texture is Slightly Smooth THEN Class is S3, Marginally Suitable.*
3) *IF Relief is Very Steep AND Soil Texture is Slightly Smooth THEN Class is N, Not Suitable.*
4) *IF Relief is Flat THEN Class is S3, Marginally Suitable.*
5) *IF Relief is Slightly Flat AND Base Saturation is very high THEN Class is S2, Moderately Suitable.*

Subsequently, M2 rules were applied to the land characteristic data of Bogor Regency to generate visualizations of cocoa land suitability predictions using ArcMap 10.3. This process was also conducted to display comparisons between the actual BBSDLP version of the cocoa land suitability maps and the M2 prediction results. The visualization comparison of cocoa land suitability maps is shown in Fig. 3.

Based on Figure 3, differences in classification were identified between the modeling results map in Bogor Regency and the original BBSDLP data. This disparity is visible through the comparison between Figure 3(b), which shows the modeling results with red polygon lines, and Figure 3(a), which presents the actual data from BBSDLP. Several areas marked with red polygon lines in the model results indicate classification errors, where regions that should have been categorized as S2 according to the BBSDLP version were instead classified as S3 by the model, and vice versa. This miscategorization is a consequence of the rules generated by the model in differentiating between these two suitability classes. Area assessment was subsequently conducted for each land suitability class utilizing the ST_Area function in PostgreSQL DBMS to present comprehensive guidance for cocoa land suitability. The calculation output for the cocoa land use area in Bogor Regency is presented in Table 5.

Table 5 demonstrates area discrepancies in cocoa land suitability categories between the modeling output and BBSDLP data. These discrepancies include increased area in the S3 category and decreases in the N category. This variation results from the achieved accuracy level of 87.27%, indicating that not all data has been correctly categorized. According to Table 5, most cocoa land suitability categories in Bogor

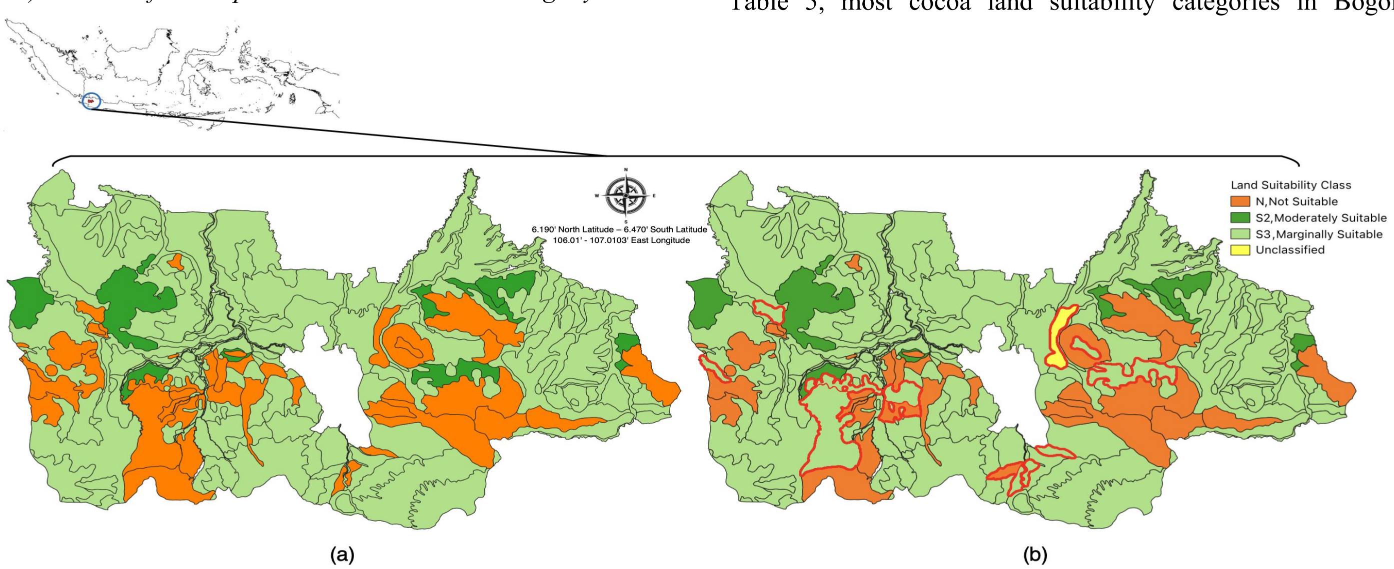

Fig. 3. Cocoa land suitability maps

Regency consist of S2 and S3 classes. Based on these conditions, establishing cocoa cultivation development areas in Bogor Regency can prioritize S2 and S3 class regions. Furthermore, according to FAO guidelines, land suitability categories can be optimized through land quality improvement [7], thereby enabling S2 class transformation into S1. Land quality enhancement can be implemented by modifying parameter values for growth requirements. For example, soil pH that is initially acidic can be modified to slightly acidic to meet S1 category requirements through nutrient addition [5], and similarly for other parameters. This approach provides a practical pathway for upgrading land suitability classifications through targeted agricultural interventions and soil management practices.

IV. Conclusion

ID3 spatial algorithm implementation for cocoa land suitability evaluation in Bogor Regency, Indonesia, demonstrates effectiveness through a comprehensive analysis of nine critical soil attributes: elevation, drainage, relief, base saturation, cation exchange capacity, soil texture, soil pH, and mineral soil depth. Two distinct models were developed and rigorously tested to optimize classification performance. Model M1 served as the baseline approach, while Model M2 incorporated a strategic minimum land area constraint of ≥1 hectare. This methodological enhancement yielded quite convincing results - M2 achieved an impressive 87.27% accuracy, representing a dramatic improvement over M1's modest 29.09% performance. Relief emerged as the dominant predictor in the superior M2 model, selected as the root node due to its exceptional gain value and minimal entropy. This attribute selection enabled superior class separation and significantly reduced classification uncertainty. Spatial analysis results identified Bogor Regency's cocoa cultivation potential across distinct suitability zones: 16,443 hectares classified as S2 (moderately suitable) and 231,018 hectares as S3 (marginally suitable). The study concludes that integrating Artificial Intelligence, specifically the ID3 spatial algorithm, enhances the precision of land suitability evaluations, providing valuable insights for sustainable and data-driven agricultural development. Future research could explore comparative analyses with other land suitability assessment methods to further validate ID3's advantages and enhance its applicability in various agricultural contexts.

Acknowledgment

Thank you to the Institute for Research and Community Service of Universitas Pembangunan Nasional "Veteran" Yogyakarta for supporting through research grant number 368/UN62.21/PG.00.00/2025.

References

[1] D. K. Terzungwe, S. Chimuka, and S. Gospel, "The economic importance of cocoa industry in countries of the gulf of guinea," *The Scientific Heritage*, vol. 150, pp. 64–74.

[2] J. E. Kongor, M. Owusu, and C. Oduro-Yeboah, "Cocoa production in the 2020s: Challenges and solutions," *CABI Agriculture and Bioscience*, vol. 5, Art. no. 102, 2024, doi: 10.1186/s43170-024-00310-6.

[3] Badan Pusat Statistik, "Berita Resmi Statistik (17 Februari 2025)," 2025. [Online]. Available: https://www.bps.go.id/id/pressrelease/2025/02/17/2409/ekspor-januari-2025-mencapai-us-21-45-miliar--turun-8-56persen-dibandingkan-dengan-desember-20242--impor-januari-2025-senilai-us-18-00-miliar--turun-15-18persen-dibandingkan-dengan-desember-2024.html (Accessed: May 2025).

[4] Pusat Data dan Sistem Informasi Pertanian, *Buku Outlook Komoditas Perkebunan Kakao*, 2023rd ed. Sekretariat Jenderal - Kementerian Pertanian, 2023.

[5] D. Djaenudin, M. H., S. H., and A. Hidayat, *Petunjuk Teknis Evaluasi Lahan untuk Komoditas Pertanian*, 2nd ed. Bogor (ID): Badan Penelitian dan Pengembangan Pertanian, 2011.

[6] A. Nurkholis, I. S. Sitanggang, Annisa, and Sobir, "Spatial decision tree model for garlic land suitability evaluation," *International Journal of Artificial intelligence (IJ-AI)*, vol. 10, no. 3, pp. 666–675, 2021, doi: 10.11591/ijai.v10.i3.pp666-675.

[7] FAO, *A framework for land evaluation*, 1st ed. Rome (IT): Food and Agriculture Organization of The United Nations, 1976. doi: 10.1007/BF01203810.

[8] R. Rahmawaty, S. Frastika, A. Rauf, R. Batubara, and F. S. Harahap, "Land suitability assessment for Lansium domesticum cultivation on agroforestry land using matching method and geographic information system," *Biodiversitas Journal of Biological Diversity*, vol. 21, no. 8, Jul. 2020, doi: 10.13057/biodiv/d210835.

[9] L. Qu, Y. Shao, and L. Zhang, "Land suitability evaluation method based on GIS technology," in *2nd International Conference on Agro-Geoinformatics: Information for Sustainable Agriculture, Agro-Geoinformatics*, 2013, pp. 7–12. doi: 10.1109/Argo-Geoinformatics.2013.6621869.

[10] O. M. Kılıc, K. Ersayın, H. Gunal, A. Khalofah, and M. S. Alsubeie, "Combination of fuzzy-AHP and GIS techniques in land suitability assessment for wheat (Triticum aestivum) cultivation," *Saudi J Biol Sci*, vol. 29, no. 4, pp. 2634–2644, 2022.

[11] A. Nurkholis and I. S. Sitanggang, "A spatial analysis of soybean land suitability using spatial decision tree algorithm," in *Sixth International Symposium on LAPAN-IPB Satellite*, T. D. Pham, K. D. Kanniah, K. Arai, G. J. P. Perez, Y. Setiawan, L. B. Prasetyo, and Y. Murayama, Eds., SPIE, Dec. 2019, p. 65. doi: 10.1117/12.2541555.

[12] S. Rinzivillo and F. Turini, "Classification in geographical information systems," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3202, pp. 374–385, 2004, doi: 10.1007/978-3-540-30116-5_35.

[13] A. Nurkholis, I. S. Sitanggang, Annisa, and Sobir, "Spatial decision tree model for garlic land suitability evaluation," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 666–675, 2021, doi: 10.11591/ijai.v10.i3.pp666-675.

[14] A. Nurkholis, M. Muhaqiqin, and T. Susanto, "Algoritme spatial decision tree untuk evaluasi kesesuaian lahan padi sawah irigasi," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 5, pp. 978–987, 2020, doi: 10.29207/resti.v4i5.2476.

[15] A. Nurkholis and I. S. Sitanggang, "Optimalisasi model prediksi kesesuaian lahan kelapa sawit menggunakan algoritme pohon keputusan spasial," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 3, pp. 192–200, 2020, doi: 10.14710/jtsiskom.2020.13657.

[16] Balai Besar Penelitian dan Pengembangan Sumberdaya Lahan Pertanian (BBSDLP), *Atlas Peta Tanah Semi Detail Skala 1:50.000, Kabupaten Bogor, Provinsi Jawa Barat*. Bogor: Badan Penelitian dan Pengembangan Pertanian - Kementerian Pertanian, 2016.

[17] BBSDLP, *Atlas Peta Kesesuaian Lahan Dan Arahan Komoditas Pertanian Pertanian, Kabupaten Bogor, Provinsi Jawa Barat, Skala 1:50.000*, 2nd ed. Bogor (ID): Badan Penelitian dan Pengembangan Pertanian, Kementerian Pertanian, 2016.
[18] A. Nurkholis, S. Styawati, and A. Suhartanto, "Firefly algorithm for SVM multi-class optimization on soybean land suitability analysis," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 2, pp. 592–597, 2024.
[19] A. Nurkholis, M. Muhaqiqin, and T. Susanto, "Land suitability analysis for upland rice based on soil and weather characteristics using spatial ID3," *JUITA: Jurnal Informatika*, vol. 8, no. 2, pp. 235–244, 2020, doi: 10.30595/juita.v8i2.8311.
[20] I. S. Sitanggang, R. Yaakob, N. Mustapha, and A. A. B. Nuruddin, "An extended ID3 decision tree algorithm for spatial data," in *IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services*, IEEE, Jun. 2011, pp. 48–53. doi: 10.1109/ICSDM.2011.5969003.
[21] A. Nurkholis, S. Styawati, S. Alim, H. Saputra, and A. Ferriyan, "Sentiment analysis of COVID-19 booster vaccines on twitter using multi-class support vector machine," *Applied Information System and Management (AISM)*, vol. 8, no. 1, pp. 29–36, 2025.
[22] Pusat Data dan Sistem Informasi, "Produksi, luas panen dan produktivitas sayuran di Indonesia," *Kementerian Pertanian Republik Indonesia*. Accessed: Aug. 05, 2020. [Online]. Available: https://aplikasi2.pertanian.go.id/bdsp
[23] I. S. Sitanggang, R. Yaakob, N. Mustapha, and A. A. N., "A decision tree based on spatial relationships for predicting hotspots in peatlands," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 12, no. 2, pp. 511–518, Jun. 2014, doi: 10.12928/TELKOMNIKA.v12i2.2036.