PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol x(x): xx–xx
https://doi.org/10.14719/pst.2926

HORIZON
e-Publishing Group
HePG

RESEARCH ARTICLE

Optimization of a soil type prediction method based on the deep learning model and vegetation characteristics

Suwardi[1], Lilik Sutiarso[1]*, Herry Wirianata[2], Andri Prima Nugroho[1], Sukarman[3], Septa Primananda[3], Moch. Dasrial[3] & Badi Hariadi[3]*

[1]Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2]Department of Agrotechnology, Faculty of Agriculture, Institute Pertanian Stiper (Instiper), Yogyakarta, Indonesia
[3]Wilmar International Plantation, Central Kalimantan Region, Indonesia

*Email: lilik-soetiarso@ugm.ac.id

OPEN ACCESS

ARTICLE HISTORY
Received: 04 September 2023
Accepted: 25 December 2023

Available online
Version 1.0 : 31 December 2023

Check for updates

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE

Suwardi, Sutiarso L, Wirianata H, Nugroho A P, Sukarman, Primananda S, Dasrial M, Hariadi B. Optimization of a soil type prediction method based on the deep learning model and vegetation characteristics. Plant Science Today (Early Access). https://doi.org/10.14719/pst.2926

Abstract

The structure and composition of forest vegetation plays an important role in different ecosystem functions and services. This study aimed to identifying soil types based on vegetation characteristics using a deep learning model in the High Conservation Value (HCV) area of Central Kalimantan, spanning 632.04 hectares. The data on vegetation were collected using a combination method between line transect and quadratic plots were placed. The development of a deep learning model was based on the results of a vegetation survey and the processing of aerial photos using the Feature Classifier method. The results of applying a deep learning model could provide a relatively accurate and consistent prediction in identifying soil types (Entisols 62%, Spodosols 90%, Ultisols 90% accuracy). The composition of vegetation community in Ultisols was dominated of seedling and tree (closed canopy), meanwhile in Entisols and Spodosols was dominated of seedling and sapling (dominantly open canopy). Ultisols exhibited the highest species richness (57 species), followed by Spodosols (31 species) and Entisols (14 species). Ultisols, Entisols, and Spodosols displayed even species distribution(J' close to 1) without dominance of certain species (D < 0.5). The species diversity index was at a low to moderate level (H' < 3), while the species richness index remained at a very low level (D_mg > 3.5).

Keywords

Characteristics; deep learning; identification; spodosols; vegetation

Introduction

The structure and composition of forest vegetation play a pivotal role in ecosystem functions and services, such as carbon storage, nutrient cycling, and biodiversity conservation (1,2). Changes in forest structure and composition due to human activities, such as deforestation, can significantly impact forest ecology and the services it provides (3). Broadly, forest vegetation can be categorized into different layers based on its vertical arrangement (4). The uppermost layer, known as canopy, is composed of the tallest trees that creates a continuous leafy cover over the forest floor. The lower layer includes small trees, shrubs, and woody plants growing beneath the canopy, while the herbaceous layer consists of grasses, herbs, and other non-woody plants thriving on the forest floor (5,6). The horizontal stand structure delineates the distribution of individual species within their habitat (7).

The structure and composition of vegetation in the forest refer to the physical composition and variety of plant species within the forest ecosystem, as well as the continuity of these species (8). Forest vegetation structure can be categorized based on tree height and diameter, plant cover density, and vertical layers of vegetation (9,10). The composition of forest vegetation refers to the plant species present, emphasizing the need for inventory activities to consider Indonesian forest types and strata (11). Forests exhibit a rich diversity of tree, shrub, herbaceous, and grass species, with their presence influenced by a variety of factors, including soil type, climate, disturbance regimes, and human activities (12,13). Vegetation analysis is one of the ways to explore the composition of types, forms, and vegetation structures of plants (14).

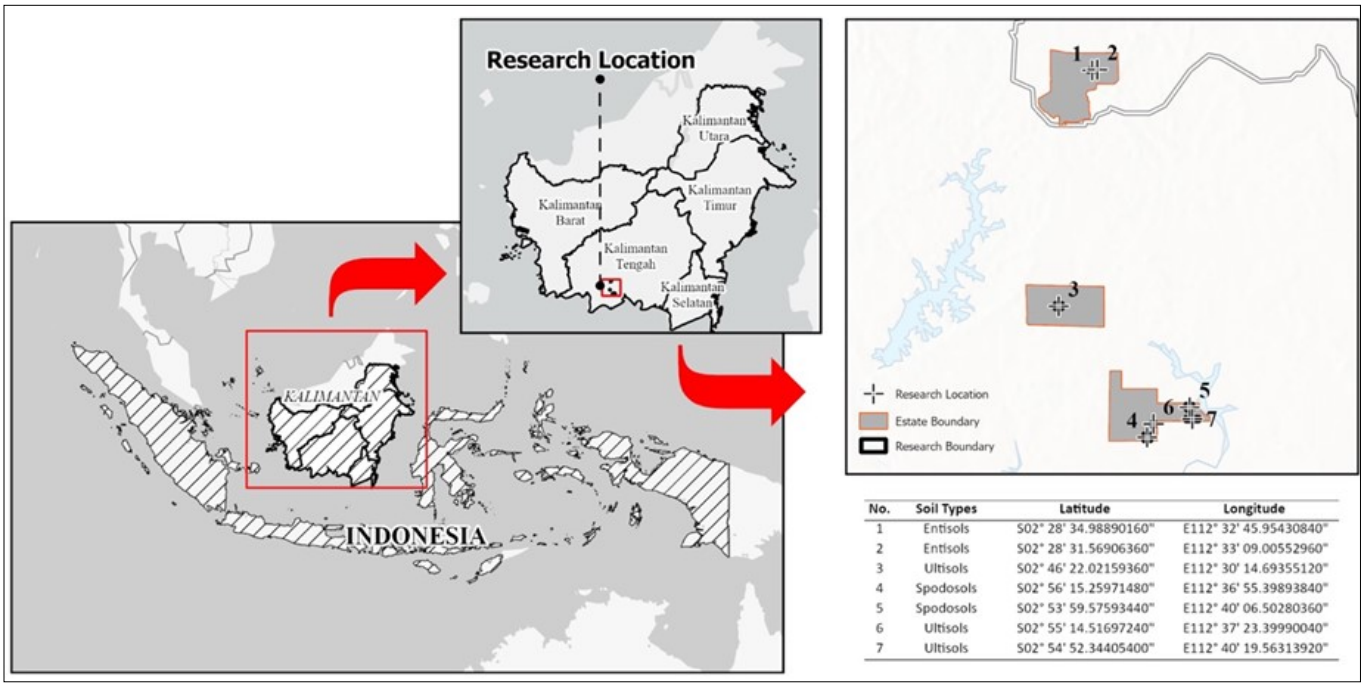
The structure and composition of vegetation may be directly influenced by soil types, as they are closely linked to the areas where the plants thrives (15,16). Additionally, the plant species that flourishes play a crucial role in determining the species of fauna that inhabit these ecosystems (17,18). In this study, the soil types used are: Entisols, Spodosols, and Ultisols. According to (19), Entisols are primarily sandy soils with a minimal soil development profile, lacking a distinct lower horizon. Spodosols are sandy mineral soils featuring a spodic layer <100 cm, marked by an acidic pH, and an accumulation of organic matter, Al, and/or lacking Fe oxide in the lower horizon (20,21). Ultisols, are clay mineral soils characterized by an acidic pH (lowest pH), low base saturation, an accumulation of clay in the lower horizon, and argillic characteristics (22,23).

Research into the structure and composition of vegetation typically seeks to enhance our understanding of fundamental ecology for the development of sustainable forest management (24,25). However, in this study, the characteristics of the vegetation were developed as a basis for identifying specific soil types. In addition to traditional methods, the incorporation of deep learning, specially through the Feature Classifier method, has proven effective in recognizing and classifying images for the identification of land cover and vegetation characteristics (26,27). The Feature Classifier is employed to categorize images into distinct groups (28,29). This study aims to identify soil types in Entisols, Spodosols, and Ultisols based on vegetation characteristics and deep learning models.

Materials and Methods

Study Area

This study was conducted in oil palm plantations located in Seruyan Regency and Kotawaringin Timur Regency, Central Kalimantan Province, within a HCV area covering 632.04 Ha (121.93 ha Entisols, 265.75 ha Spodosols, 244.36 ha Ultisols). The research spanned six months, from August 2022 to January 2023. The study site featured a flat to slightly undulating topography, with elevations ranging from 5 to 32 meters above sea level (m a.s.l). The area encompassed six soil orders: Oxisols, Ultisols, Inceptisols, Histosols, Spodosols, and Entisols. In this study, Ultisols were selected from the subgroups Aquic palaeudult and Typic paleaqult, Entisols from Albic Quartzipsamment, and Spodosols from Typic haplohumod (Fig.1.)



No.	Soil Types	Latitude	Longitude
1	Entisols	S02° 28' 34.98890160"	E112° 32' 45.95430840"
2	Entisols	S02° 28' 31.56906360"	E112° 33' 09.00552960"
3	Ultisols	S02° 46' 22.02159360"	E112° 30' 14.69355120"
4	Spodosols	S02° 56' 15.25971480"	E112° 36' 55.39893840"
5	Spodosols	S02° 53' 59.57593440"	E112° 40' 06.50280360"
6	Ultisols	S02° 55' 14.51697240"	E112° 37' 23.39990040"
7	Ultisols	S02° 54' 52.34405400"	E112° 40' 19.56313920"

Fig. 1. Location in oil palm plantations in Seruyan Regency and Kotawaringin Timur Regency, Central Kalimantan Province.

Procedures

Research tools and materials

The necessary tools and materials included a clinometer, compass, circumference gauge, 20 m plot thread, 10 m plot thread, 5 m plot thread, 2 m plot thread, label, GPS, camera, tape measure, wood plot, thermometer, hygrometer, soil tester, plastic bags, identification books, observation sheets, stationery, label paper, raffia rope, and alcohol 70%.

Data collection methods

The collected data encompassed information on the number of individuals in the seedling stage and the diameter at breast height (DBH) for each phase of rejuvenation or vegetation growth. The vegetation data collection method involved a combined approach using both line transects and quadratic plots (grid system). The size of each plot was adjusted to the observed vegetation level, with systematic placement carried out by establishing transects to the north, south, west and east. Observations of the undergrowth focused on non-woody plants, generally located above the forest floor. The observations of forest structure and composition involved categorizing the vegetation into different stages according to the following criteria: (Fig. 2).

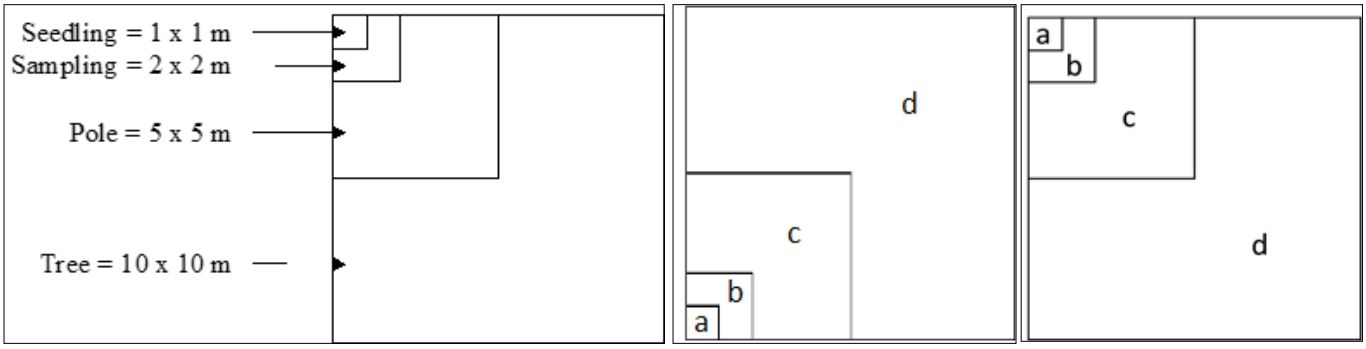


Fig. 2. Research sampling sketch.

1) Seedling: A sapling with a height of less than 150 cm.
2) Sapling: A sapling with a height of more than 150 cm, but had a trunk diameter of less than 10 cm.
3) Pole: A young tree with a diameter ranging from 10 to 20 cm.
4) Tree: A woody plant with a diameter exceeding 20 cm.

In each observation station, a total of four quadratic transects were positioned. Each plant within the plot were labeled to facilitate data collection. Every individual present in each plot at the observed station was documented on the observation sheet. Subsequently, the number of individuals per species in each quadrat was calculated to determine the importance value of each species. Plant species were captured using a photo camera, and identification was performed using an identification book. Sampling Intensity (SI) for forest area < 1000 Ha was set at 10% of the area.

$$\text{Sampling Intensity (SI)} = \frac{n \times LPU}{LH}$$

n = Number of Plots/ PU

LPU = Plot Area/ PU

LH = Forest area inventoried

Data preparation with the feature classifier method

A deep learning model was developed based on the findings from a vegetation survey and the processing of aerial photos using the Feature Classifier method. The data utilized in this study consisted of plot data gathered during field surveys to determine the relative abundance and diversity of forest plant species within each of the Entisols, Spodosols and Ultisols soil types at the study site. The survey plot data were then employed to train a prediction model for soil types at the study site, which encompasses seven locations, each characterized by its specific soil type. The total number of sample plots used as training data amounted to 86 plots (Fig. 3).

The preparation of training data was executed using GIS software, resulting in generation of image data and label data for each image section (Fig. 4). A total of 1302 image chunks, each measuring 512x512 pixels, were employed in creating the deep learning model.

Data analysis

Relative abundance

Relative abundance represents the percentage composition of organisms of a specific type in relative to the total number of organisms in a given area, encompassing vegetation frequency, density, dominance, and the importance value index. Different populations within a community coexist in varying relative proportions. Vegetation frequency serves as an attribute that describes the probability of finding a species in a particular area. The probability is determined based on the specie's occurrence in a series of sample units. Frequency acts a measure of uniformity or distribution. The presence of a certain frequency type describes the distribution pattern of that type, indicating whether it is spread throughout an area or concentrated within a specific group.

$$\text{Absolute Frequency, } AF = \frac{\text{Area of } Plots\ in\ which\ a\ Species\ Occurs}{Total\ Area\ Sampled}$$

$$\text{Relative Frequency, } RF = \frac{Frequency\ of\ a\ Species}{Total\ Frequency\ of\ All\ Species} \times 100\%$$

Vegetation density is the percentage of a particular species of vegetation or plants that inhabit a certain area. This metric reveals the number of individual species within a defined unit area, providing a quantitative description of the species abundance at the observation site.

Fig. 3. Distribution of field survey plots used as training data for the deep learning model. The red dot is the location of the field survey plots. (**a**) and (**b**) show the distribution of samples for areas with a soil type of Entisols. (**c**), (**d**) and (**e**) show the distribution of samples for areas with a soil type of Ultisols. (**f**) and (**g**) show the distribution of samples for areas with a soil type of Spodosols.

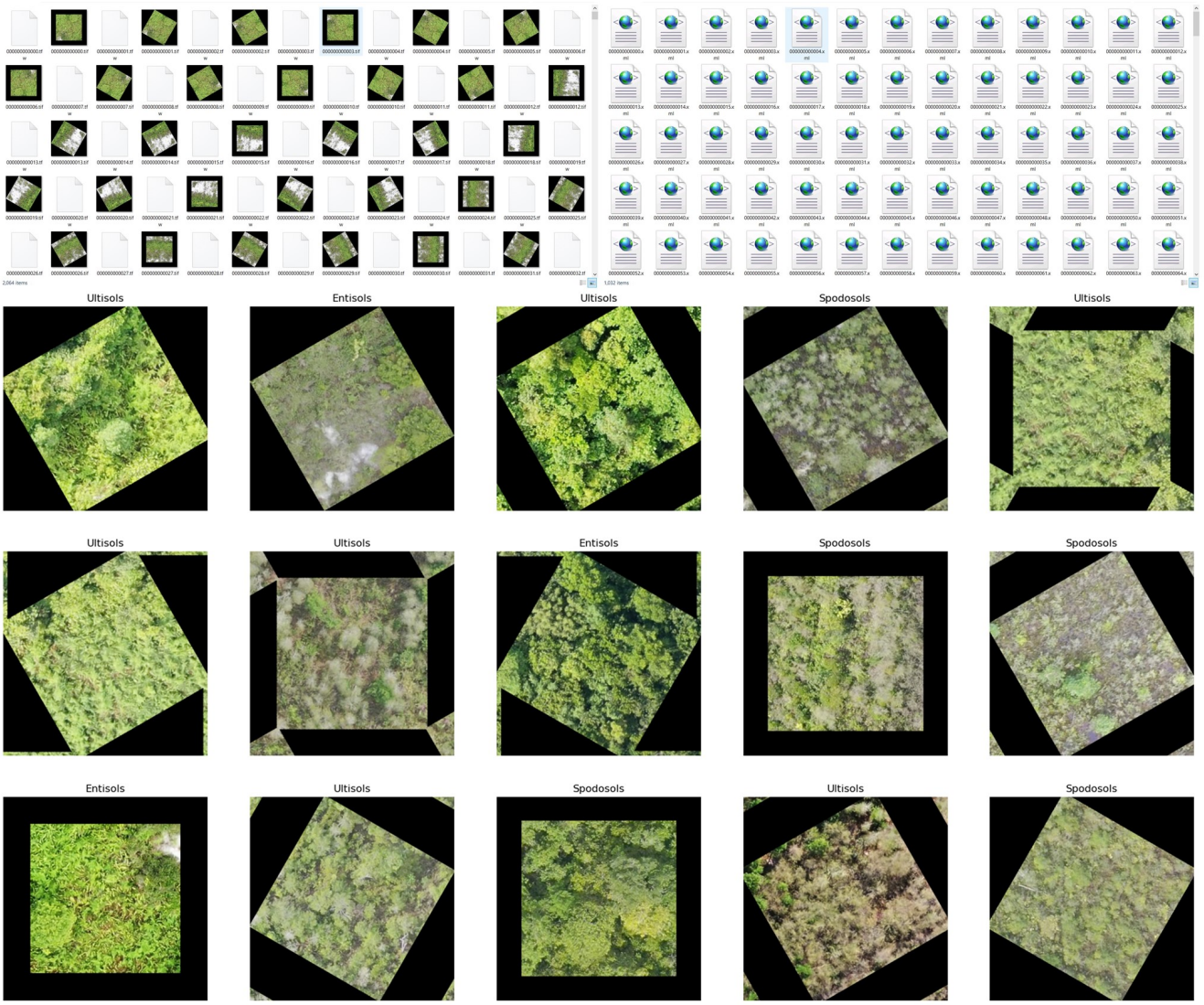


Fig. 4. Image chunks as data for the creation of deep learning model are provided with the label information of soil type.

Absolute Density, $$AD = \frac{Number\ of\ a\ Species}{Total\ Area\ Sampled}$$

Relative Density, $$RD = \frac{Density\ of\ a\ Species}{Total\ Density\ of\ All\ Species}\ x\ 100\%$$

Dominance can be interpreted as the dominance of one species over another. It can be in terms of space, light, and others. This dominance is often expressed through matrix like abundance and density, cover percentage and basal area (BA), volume, biomass, importance value index (IVI).

Absolute Dominance, $$D = \frac{Total\ Basal\ Area\ of\ a\ Species}{Total\ Area\ Sampled}$$

Relative Dominance, $$RDO = \frac{Dominance\ of\ a\ Species}{Total\ Dominance\ of\ All\ Species}\ x\ 100\%$$

Importance Value Index (IVI) is an analytical tool employed to assess the dominance of a species within a particular community. This index is calculated by summing the value of relative density, relative dominance, and relative frequency, with the maximum possible total being 300%. The higher the IVI value of a particular species indicates a greater degree of dominance within the community.

Importance Value Index (IVI) of Herbs/Seedling

$$= \text{Relative Density } (\%) + \text{Relative Frequency } (\%)$$

Importance Value Index (IVI) of Sapling, Pole & Tree

$$= \text{Relative Density } (\%) + \text{Relative Frequency } (\%) + \text{Relative Dominance } (\%)$$

Species diversity

Species diversity is defined as the number of different species in an ecosystem and their relative abundance.

$$H' = \sum_{i=1}^{S} (\rho i)\ (\ln \rho i)$$

H' = Shannon-Wiener diversity index

Pi = the proportion of the entire community made up of species

ni = number of individuals of a species *i*

N = total number of individuals

Two key components of measuring species diversity are species richness and species evenness. Species richness quantifies the number of genetically or functionally related groups of individuals. In most vegetation surveys, richness is expressed as the number of species, and is commonly referred to as species richness.

$$D_{mg} = \frac{S - 1}{Ln\ N}$$

D_{mg} = Margalef's index

S = total number of species

N = total number of individuals in the sample

Ln = natural logarithm

Species evenness, another component of diversity, is the proportion of species or functional groups present in a habitat. Greater evenness indicates a more proportionally distribution of species, signifying a balanced representation of various organisms. Conversely, a habitat with low evenness indicates that several species dominate that habitat, creating an uneven distribution where certain species hold greater influence.

$$J' = \frac{H'}{D_{max}}$$

J' = evenness value (between 0 – 1)

H' = Shannon-Wiener diversity index

D_{max} = maximum value of diversity index

Deep learning modeling with the feature classifier method

The deep learning model was developed using the Jupyter Notebook software, incorporating the Python libraries arcpy and arcgis.learn. The Object Classification Models, specially the Feature Classifier in the arcgis.learn library, were used to construct the deep learning model for the images within the provided training data. It is essential to note that to utilize the arcpy and arcgis.learn libraries, a valid license for ArcGIS Pro software with the Image Analyst extension is required.

The "prepare_data()" function serves to divide the training data into training and validation datasets, maintain a division ratio of 50:50.Subsequently, the "model.lr_find()" function is used to determine the optimal learning rate, which is established at 0.00069 for the model fitting process. The training of the model is executed using the "model.fit" function, iterating through 39 epochs until the process is automatically halted due to the model accuracy plateauing and showing minimal changes (Table 1). Upon inspecting the graph, it is evident that the formed model does not exhibit signs of overfitting (Fig. 5).

The evaluation of model performance through the Confusion Matrix method, applied to the validation data, reveals a high level of accuracy. The deep learning model effectively distinguishes images within the training data, categorizing them based on each soil type (Fig. 6).

Table 1. Train_Loss, Valid_Loss, Accuracy and Processing Time per Epoch values

EPOCH	TRAIN_LOSS	VALID_LOSS	ACCURACY	TIME
0	1.963028	1.301305	0.401163	0:55
1	1.838040	1.232017	0.445736	0:58
2	1.777831	1.109779	0.496124	0:54
3	1.623583	0.959240	0.569767	0:50
4	1.466226	0.917420	0.583333	0:52
…	…	…	…	…
34	0.158218	0.033353	0.988372	0:50
35	0.169254	0.018230	0.996124	0:50
36	0.170862	0.027969	0.988372	0:50
37	0.140715	0.016946	0.996124	0:50
38	0.117391	0.021682	0.990310	0:50

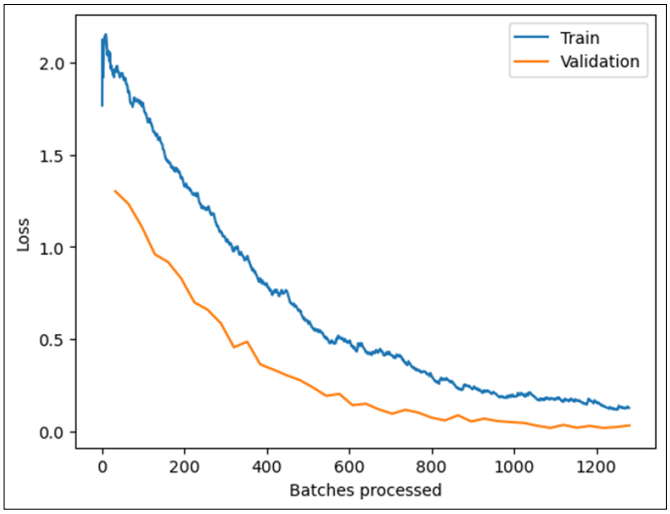


Fig. 5. Graph of Train_Loss and Validation_Loss during the creation of the deep learning model.

Subsequently, the previously developed deep learning model was employed to categorize soil types within the research area. Before applying the model, it was necessary to create a grid overlay on the aerial photographs of the study site, with each plot measuring 20 x 20 meters. Consequently, a total of 11,203 plots were established across the entire research area. Precision, recall and f1-score values were computed through the calculation of true & false positives, as well as true & false negatives (Fig. 7).

True Negative (TN) occurs when a plot that should have a soil type of non-Entisols or non-Spodosols, or non-Ultisols is predicted as a soil type of non-Entisols, or non-Spodosols, or non-Ultisols, respectively.

True Positive (TP) occurs when a plot that should have a soil type of Entisols, or Spodosols, or Ultisols is predicted as a soil type of Entisols, or Spodosols, or Ultisols, respectively.

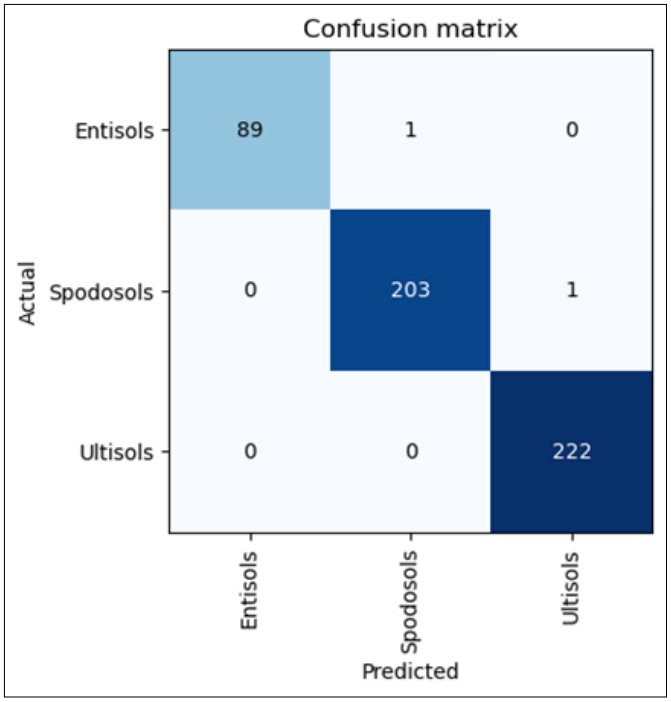


Fig. 6. The Confusion Matrix of performance test of the deep learning model against validation data.

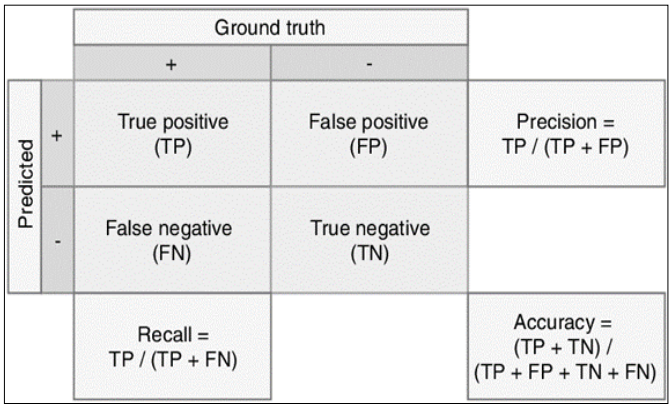


Fig. 7. The Matrix of TP, FP, FN, TN, Precision, Recall dan Accuracy.

False Negative (FN) (type-2 error) occurs when a plot that should have a soil type of Entisols, or Spodosols, or Ultisols is predicted as a soil type of non-Entisols, or non -Spodosols, or non-Ultisols, respectively.

False Positive (FP) (type-1 error) occurs when a plot that should have a soil type of non-Entisols, or non-Spodosols, or non-Ultisols is predicted as a soil type of Entisols, or Spodosols, or Ultisols, respectively.

Precision, recall and f1-score values are calculated using the formula below :

precision = TP / (TP + FP) .....(1)

recall = TP / (TP + FN).....(2)

f1-score = 2*(recall*precision) / (recall + precision).....(3)

accuracy = (TP + TN) / (TP + FP + TN + FN).....(4)

Precision incorporates the False Positive (FP) variable, and recall involves the False Negative (FN) variable. Consequently, a smaller False Positive (FP) results in a higher precision value, while a smaller False Negative (FN) leads to greater recall. The f1-score falls within a range of 0 – 1, with value closer to 1 indicating superior precision and recall values for the formed deep learning model formed. Meanwhile, accuracy describes how accurate the deep learning model's proficiency in accurately classifying soil types within the research area.

Results and Discussion

Number of Vegetation Species and Individuals in Entisols, Spodosols, and Ultisols

Examining Fig. 3, it is evident that all vegetation levels (seedling, sapling, pole, tree) in Ultisols exhibited a higher number of species compared to Entisols and Spodosols. Entisols had the lowest number of species, particularly at the tree vegetation level. The number of species observed at the sapling vegetation level surpassed those at the seedling, pole, and tree levels. Interestingly, both Spodosols and Ultisols exhibited the same number of species at the pole and tree vegetation level.

The abundance of species does not always correlate with the overall number of individuals in a given area, as evident in the comparison between Fig. 8 and 9. Despite Ultisols having a higher number of species (seedling and sapling vegetation level) compared to Spodosols, the latter exhibited a greater total number of individual. At the pole and tree vegetation levels in Ultisols, the total number of individuals exceeded those in Spodosols and Entisols. Within the studied vegetation community, Entisols consistently displayed the lowest number of individuals and the least presence at the tree vegetation level. In compared to Spodosols and Ultisols, Entisols had

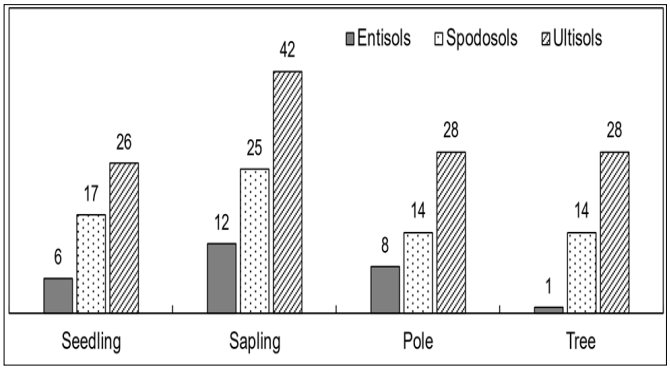


Fig. 8. Number of species in the vegetation community of Entisols, Spodosols and Ultisols.

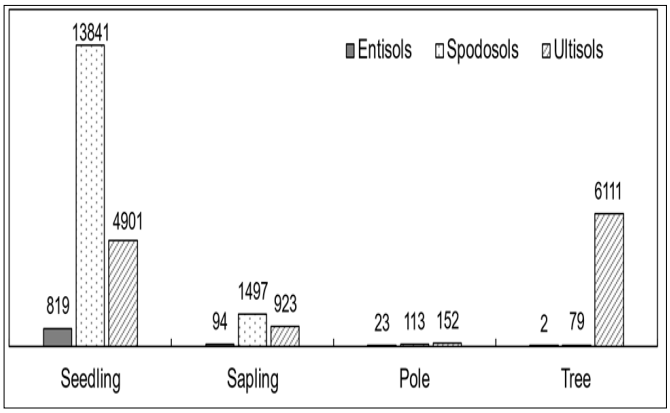


Fig. 9. Total individuals in the vegetation community of Entisols, Spodosols and Ultisols.

the fewest species, totaling 14. Spodosols, on the other hand, harboured 31 species across the seedling, sapling, pole, and tree vegetation levels, while Ultisols boasted the highest species diversity, with 57 identified species.

At the seedling level, Ultisols boasted 26 species, totaling 4,901 individual plants. Spodosols, with 17 species, exhibited the highest total number of individual plants at 13,841. Entisols had the least number of species (6 species) and the lowest total number of individual plants, amounting to 819. Moving to sapling level, Ultisols led with the highest number of species at 41 species (923 plants). However, Spodosols, despite having 25 species, recorded the highest total number of individual at 1,497. Entisols, with 12 species, had the lowest total number of individual plants, totaling 94. At the pole level, Ultisols excelled with the highest number of species (29 species) and total individuals (152 plants), surpassing Spodosols (14 species, 113 plants) and Entisols (8 species, 23 plants). Remarkably, Entisols exhibited only one species with a mere 2 individual plants, markedly lower than Ultisols with 28 species and 6,111 individual plants, and Spodosols with 14 species and 79 individual plants

Based on the provided data, Ultisols exhibited the highest number of species at the seedling, sapling, pole, and tree levels, consistent with the total number of individual plants at the pole and tree levels. However, at the seedling and sapling levels, Spodosols recorded the highest number of individuals despite having a lower number of species. Spodosols hosted only 4 plant species at the study site: *Barringtonia sp*, *Calophyllum hosei*, *Ochanostachys sp*, and *Polyalthia glauca*. In Entisols, 4 plant species were identified, namely *Acacia mangium*, *Alstonia scholaris*, *Cocos nucifera*, and *Syzygium sp2*. Ultisols, on the other hand, harbored a unique set of 29 plant species at the study site, including *Aglaia sp*, *Aquilaria malacensis*, *Campnosperma coriaceum*, *Compassia excelsa*, *Cratoxylum sp*., *Cryptocarya nitensyang* and other species outlined in Table 2.

Vegetation Characteristics Based on Vegetation Community Level

The overall vegetation community in Entisols at the study site was predominantly characterized by the seedling level. In Entisols, no species emerged as dominant in the community (D value < 0.5), except at the tree vegetation

Table 2. The names of plants found in each type of soil

No	Species	Ultisols (Ult)				Spodosols (Spo)				Entisols (Ent)				Existence of Species		
		Se	Sa	Po	Tr	S	P	T	Po	S	P	T	Po	Ult	Spo	Ent
1	*Acacia mangium*										√			-	-	√
2	*Aglaia sp*	√	√											√	-	-
3	*Alstonia scholaris*											√		-	-	√
4	*Alstonia sp1*	√	√	√	√	√		√	√					√	√	-
5	*Aporusa sp*		√				√							√	√	-
6	*Aquilaria malacensis*		√											√	-	-
7	*Baccaurea sp*		√	√			√							√	√	-
8	*Baeckea frutescens*	√	√			√	√			√	√	√		√	√	√
9	*Barringtonia sp*						√							-	√	-
10	*Calophyllum hosei*					√	√	√	√					-	√	-
11	*Calophyllum sclerophyllum*				√									√	-	-
12	*Calophyllum sp.*				√	√	√	√						√	√	-
13	*Campnosperma coriaceum*			√										√	-	-
14	*Cocos nucifera*									√	√			-	-	√
15	*Combretocarpus rotundatus*		√				√	√	√					√	√	-
16	*Compassia excelsa*				√									√	-	-
17	*Cratoxylum arborescens*	√	√	√	√	√	√	√	√		√	√		√	√	√
18	*Cratoxylum sp.*				√									√	-	-
19	*Cryptocarya nitens*			√	√									√	-	-
20	*Dillenia eximia*		√											√	-	-
21	*Diospyros sp*		√											√	-	-
22	*Diospyros sp2*			√	√									√	-	-
23	*Eurycoma longifolia*		√						√		√			√	√	√
24	*Eusideroxylon zwageri*		√	√	√									√	-	-
25	*Ficus sp.*				√		√					√		√	√	√
26	*Flacourtia rukam*	√	√	√	√		√							√	√	-
27	*Ganua motleyana*			√										√	-	-
28	*Garcinea sp*		√											√	-	-
29	*Gordonia sp*	√	√	√	√	√	√		√					√	√	-
30	*Gymnacranthera paniculata*	√	√			√								√	√	-
31	*Lea indica*	√	√											√	-	-
32	*Lithocarpus beccarianus*	√	√	√	√									√	-	-
33	*Macaranga gigantea*	√	√	√	√	√	√	√			√			√	√	√
34	*Madhuca motleyana*		√	√										√	-	-
35	*Mangifera magnifica*	√	√				√							√	√	-
36	*Melaleuca leucadendron*	√	√	√	√	√	√	√	√					√	√	-
37	*Mezzetia parviflora*	√	√		√									√	-	-
38	*Nephelium lappaceum*	√	√	√										√	-	-
39	*Nothaphoebe umbelliflora*			√										√	-	-
40	*Ochanostachys sp*								√					-	√	-

41	*Polyalthia glauca*					√	√							-	√	-
42	*Ptemandra coerulescens*	√	√	√	√	√	√	√						√	√	-
43	*Ptemandra cordata*		√											√	-	-
44	*Quasia borneensis*		√	√		√	√	√						√	√	-
45	*Quercus sp.*		√	√	√									√	-	-
46	*Rhodomyrtus tomentosa*	√	√				√			√	√			√	√	√
47	*Rhothmannia sp.*		√											√	-	-
48	*Sandoricum beccarianum*			√	√				√					√	√	-
49	*Schima wallichii*	√	√	√	√									√	-	-
50	*Shorea balangeran*				√	√	√	√	√					√	√	-
51	*Shorea laevifolia*				√									√	-	-
52	*Shorea smithiana*	√	√	√	√									√	-	-
53	*Stemonurus capitata*	√	√											√	-	-
54	*Syxygium sp1*		√	√			√							√	√	-
55	*Syzygium sp2*										√	√		-	-	√
56	*Syzygium sp3*	√							√	√	√	√		√	√	√
57	*Syzygium tawahense*	√	√	√	√	√	√	√	√	√	√	√		√	√	√
58	*Syzygium tendens*	√	√	√	√	√	√	√	√					√	√	-
59	*Syzygium zeylanicum*	√	√			√	√			√	√			√	√	√
60	*Tetramerista glabra*		√		√									√	-	-
61	*Tristaniopsis Merguensis*	√	√	√		√	√	√			√	√	√	√	√	√
62	*Tristaniopsis whiteana*	√												√	-	-
63	*Vatica sp.*	√	√	√	√		√	√	√					√	√	-
64	*Vitex pinnata*		√	√	√									√	-	-
65	*Xanthophyllum sp*		√											√	-	-
		26	42	28	28	17	25	14	14	6	12	8	1	57	31	14

Note : Se = *Seedling* Sa= *Sapling* Po = *Pole* Tr = Tree

Fig. 10. Vegetation structure and composition in Entisols (**left**), Spodosols (**middle**) and Ultisols (**right**).

level where only 1 species was identified in the sample plot. The vegetation community of Entisols exhibited very low level of species richness (D_{mg})at the seedling, sapling, pole, and tree levels, all scoring below 3.5 (Table 3). Species diversity (H') was moderate at the seedling,sapling, pole, and tree levels, while at the seedling and tree levels, it was classified as low in Entisols, with an index value of H' < 1.5. The species abundance/species evenness (J') values in Entisols was close to 1, indicating an absence of species dominance in the overall vegetation community, except at the tree level where only 1 tree data point was collected across all samples at the location.

The overall vegetation community in Spodosols at the study site was predominantly characterized by the seedling level. In Spodosols, no species emerged as dominant in the community, with a D value less than 0.5. The vegetation community in Spodosols exhibited a low level of species richness (D_{mg}) at the seedling, sapling, pole, and tree levels, all scoring below 3.5. At all levels of the vegetation community in Spodosols, species diversity (H') was moderate, with an index value of (H') less than 1.5. The species abundance/evenness (J') value in Spodosols was close to 1, indicating an absence of species dominance in the overall vegetation community (Table 3).

Similarly, the overall vegetation community in Ultisols at the study site was dominated by the seedling

Table 3. The comparison of species richness index, dominance index, species diversity index and species evenness index in Entisols, Spodosols, and Ultisols

Soil Type	Index in Entisols				Index in Spodosols				Index in Ultisols			
	D_{mg}	D	H'	J'	D_{mg}	D	H'	J'	D_{mg}	D	H'	J'
Seedling	0.75	0.22	1.43	0.80	1.68	0.13	2.02	0.71	2.94	0.12	1.84	0.57
Sapling	2.42	0.11	2.13	0.86	3.28	0.13	2.26	0.70	6.01	0.09	2.70	0.72
Pole	2.23	0.16	1.87	0.90	2.75	0.18	1.94	0.74	5.37	0.10	2.63	0.79
Tree	0.00	1.00	0.00	0.00	2.98	0.17	1.96	0.74	5.50	0.18	2.25	0.68

and tree levels. In Ultisols, no species emerged as dominant in the community, with a D value less than 0.5. The vegetation community in Ultisols exhibited a low level of species richness (D_{mg}) at the seedling level, with an index value below 3.5, but had a high level of species richness from the sapling to tree vegetation levels, with an index value of 5 or greater. At all levels of the vegetation community in Ultisols, species diversity (H') was moderate, with an index value of H' less than 1.5. The species abundance/evenness (J') value in Ultisols was close to 1, indicating an absence of species dominance in the overall vegetation community (Table 3).

Based on Table 3, the highest species richness index was observed in Ultisols, followed by Spodosols and Entisols. Although Entisols exhibited a higher dominance index due to the presence of a dominant plant, overall, there were no dominant species in Entisols, Spodosols, and Ultisols (D < 0.5). The species diversity index (H') in Ultisols was higher compared to Spodosols and Entisols. The value of species abundance/species evenness (J') in Entisols, Spodosols, and Ultisols was close to 1, indicating an absence of species dominance in the vegetation community as a whole. Aerial photos revealed differences in color and density in the vegetation of Entisols, Spodosols, and Ultisols, as illustrated in Fig. 11. The results suggest that Ultisols in the study site were dominated by the seedling and tree levels, forming a closed canopy. In contrast, Entisols and Spodosols were dominated by the seedling and sapling levels, resulting in a more open canopy.

Vegetation Characteristics Based on Vegetation Structure

Based on the vegetation structure at each rejuvenation level on Entisols land, the highest Importance Value Index (IVI) value indicates the dominant vegetation species. The IVI can be interpreted based on its value, where a higher

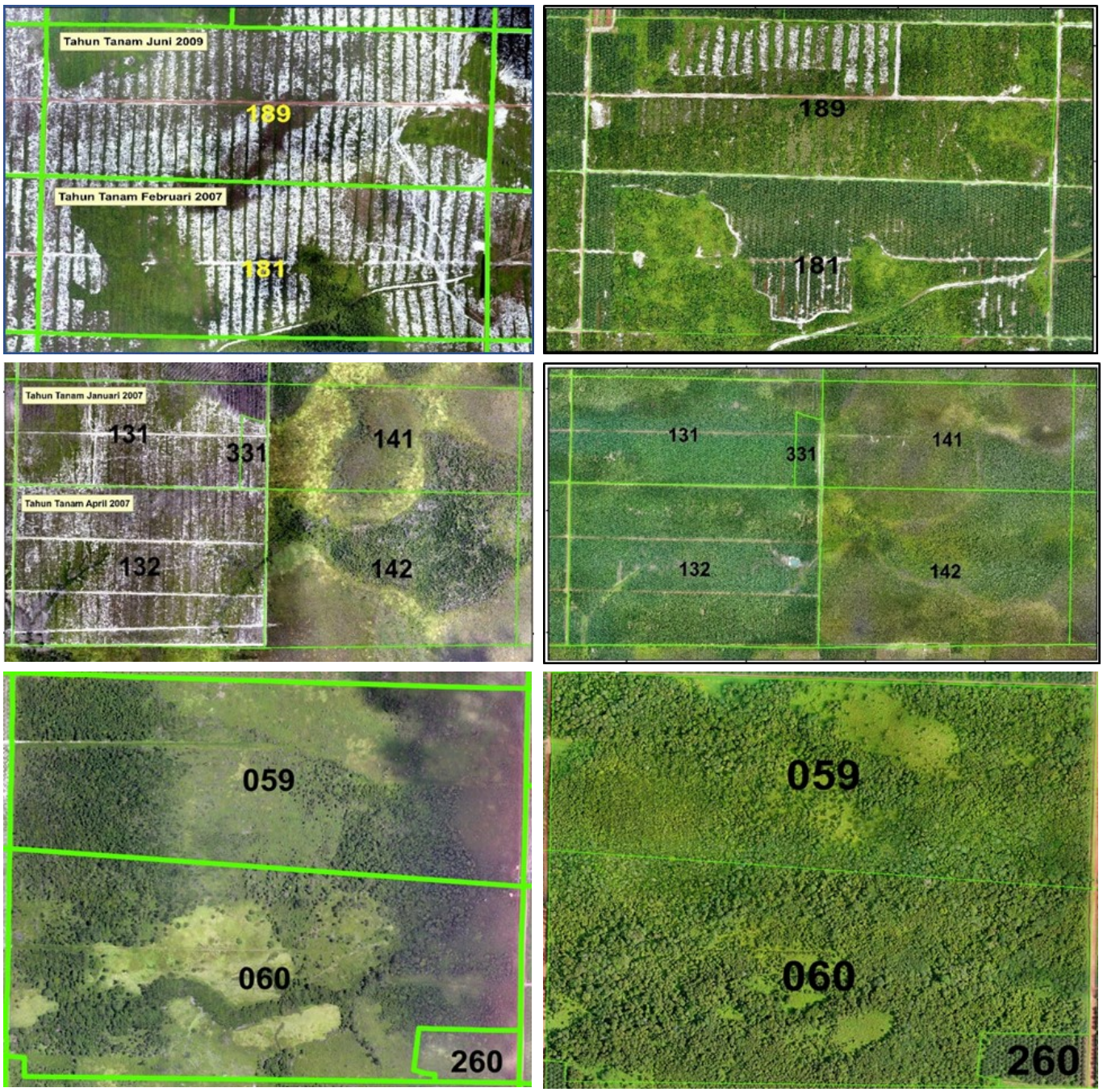


Fig. 11. The vegetation development of Entisols (**top**), Spodosols (**middle**) and Ultisols (**bottom**) as seen from the aerial photos from 2007 to 2021.

value signifies a greater level of dominance in the community, and vice versa. Table 4 below presents the three most dominant vegetation species in each vegetation level stratum at the study site, along with the individual potential estimate per hectare. At the seedling level, dominance was observed for *Syzygium tawahense*, *Syzygium sp*, and *Rhodomyrtus tomentosa*. Each vegetation level exhibited different dominant species, as illustrated in Table 4, reflecting variations in species composition in the sapling (*Baeckea frutescens*, *Rhodomyrtus tomentosa*, *Syzygium zeylanicum*), pole (*Tristaniopsis Merguensis*, *Syzygium sp*, *Baeckea frutescens*), and tree (*Tristaniopsis Merguensis*) stages.

The Importance Value Index (IVI) on Spodosols reveals different dominant species compared to Entisols and Ultisols. Table 5 presents the three most dominant vegetation species in each stratum of the vegetation level in Spodosols at the study site, along with the individual potential estimate per hectare. Each vegetation level exhibited distinct IVI values for seedling (*Syzygium zeylanicum*, *Cratoxylum arborescens*, *Tristaniopsis merguensis*), sapling (*Cratoxylum arborescens*, *Tristaniopsis merguensis*, *Melaleuca Leucadendron*), pole (*Cratoxylum arborescens*, *Combretocarpus rotundatus*, *Syzygium tawahense*), and tree (*Shorea balangeran*, *Cratoxylum arborescens*, *Combretocarpus rotundatus*).

Table 6 illustrates the Importance Value Index (IVI) in Ultisols, presenting the three most dominant vegetation species in each stratum of the vegetation level at the study site, along with the individual potential estimate per hectare. Distinct IVI values were observed for seedling (*Baeckea frutescens*, *Syzygium tendens*, *Cratoxylum arborescens*), sapling (*Melaleuca leucadendron*, *Cratoxylum arborescens*, *Syzygium tawahense*), pole (*Schima wallichii*, *Melaleuca Leucadendron*, *Syzygium tawahense*), and tree (*Schima wallichii*, *Syzygium tawahense*, *Macaranga gigantea*). Notably, at the seedling vegetation level, there was species similarity between the IVI of Ultisols and Spodosols, specifically *Cratoxylum arborescens*. Conversely, there was no species similarity in Entisols. The similarity in species between the IVI of Ultisols and Spodosols was also observed at the sapling (*Melaleuca leucadendron*) and pole (*Syzygium tawahense*) levels. Additionally, a species similarity between the IVI of Entisols and Ultisols was found at the sapling level, namely Cratoxylum arborescens. However, there was no species similarity at the tree vegetation level in Entisols, Spodosols, or Ultisols.

Several dominant plant species were identified to expedite the identification of soil types, specifically Ultisols, Entisols, and Spodosols, across various vegetation levels, including seedling, sapling, pole, and tree. To ascertain more precise dominant plant species,

Table 4. Importance Value Index (IVI) for the Entisols land

Vegetation	Species	Vernacular	Individual Potential (ha)	IVI (%)
	Syzygium tawahense	*Ubar merah*	46,250	56.14
Seedling	*Syzygium sp3*	*Galam tikus*	44,375	54.68
	Rhodomyrtus tomentosa	*Karamunting*	19,219	45.02
	Baeckea frutescens	*Ujung atap*	250	51.30
Sapling	*Rhodomyrtus tomentosa*	*Karamunting*	750	50.91
	Syzygium zeylanicum	*Nasi nasi*	275	35.28
	Tristaniopsis merguensis	*Pelawan*	38	67.71
Pole	*Syzygium sp2*	*Galam tikus*	31	58.71
	Baeckea frutescens	*Ujung atap*	25	56.92
Tree	*Tristaniopsis merguensis*	*Pelawan*	3	300

Table 5. Importance Value Index (IVI) for the Spodosols land

Vegetation	Species	Vernacular	Individual Potential (ha)	IVI (%)
	Syzygium zeylanicum	*Nasi nasi*	252,500	43.94
Seedling	*Cratoxylum arborescens*	*Geronggang*	203,676	35.67
	Tristaniopsis merguensis	*Pelawan*	183,015	32.77
	Cratoxylum arborescens	*Geronggang*	4,859	75.21
Sapling	*Tristaniopsis merguensis*	*Pelawan*	3,329	48.97
	Melaleuca leucadendron	*Galam*	2,059	35.19
	Cratoxylum arborescens	*Geronggang*	124	102.88
Pole	*Combretocarpus rotundatus*	*Tumih*	65	53.26
	Syzygium tawahense	*Ubar merah*	41	37.22
	Shorea balangeran	*Belangeran*	17	87.95
Tree	*Cratoxylum arborescens*	*Geronggang*	16	72.60
	Combretocarpus rotundatus	*Tumih*	8	31.75

Table 6. Importance Value Index (IVI) for the Ultisols land

Vegetation	Species	Vernacular	Individual Potential (ha)	IVI (%)
	Baeckea frutescens	Ujung atap	156,351	48.35
Seedling	*Syzygium tendens*	Ubar putih	50,135	37.87
	Cratoxylum arborescens	Geronggang	48,649	22.65
	Melaleuca leucadendron	Galam	2,486	57.16
Sapling	*Cratoxylum arborescens*	Geronggang	1,632	44.93
	Syzygium tawahense	Ubar merah	681	24.34
	Schima wallichii	Gandawari	132	81.41
Pole	*Melaleuca leucadendron*	Galam	38	23.88
	Syzygium tawahense	Ubar merah	27	20.71
	Schima wallichii	Gandawari	45	121.88
Tree	*Syzygium tawahense*	Ubar merah	5	20.15
	Macaranga gigantea	Mahang	3	14.54

the researchers employed an approach based on the Importance Value Index (IVI) results at different vegetation levels, providing a comprehensive understanding. IVI effectively elucidates the density, dominance, and frequency of plant species within a given location. In Ultisols, characterized by a higher value of species diversity (H'), determining the dominant plant species became challenging due to the abundance of diverse species. As per Table 6, several dominant plants in Ultisols were identified, encompassing *Baeckea frutescens*, *Melaleuca Leucadendron*, and *Schima wallichii*.

In Entisols, the predominant plant species identified at the study site is *Baeckea frutescens*, particularly at the sapling and pole levels. Conversely, in Spodosols, the prevailing plant species is *Cratoxylum arborescens*, found across the seedling, sapling, pole, and tree levels. Both *Baeckea frutescens* and *Cratoxylum arborescens* exhibit characteristics that make them well-suited for sandy soils. These adaptations contribute to their ability to thrive in less fertile and well-drained soil conditions.One shared characteristic is their extensive and deep-rooted systems, allowing efficient absorption of available water and nutrients from sandy soils. This adaptability enables these plants to endure conditions with lower fertility. Additionally, both *Baeckea frutescens* and *Cratoxylum arborescens* demonstrate drought tolerance, surviving in areas with limited water availability. The small leaves of these plants reduce water evaporation, while their robust root systems facilitate moisture absorption in sandy soils, making them suitable for regions with low rainfall or during dry seasons. A study by (30) indicated that the native vegetation of Spodosols primarily consists of gelam plants (*Melaleuca leucodendron*), with *Fibrimstylis sp* and Lalang grass (*Imperata cylindrica*) being common secondary vegetation. Another study by Hartati (31) added that Spodosols with disturbed vegetation mainly comprise pioneer species like Lalang (Imperata cylindrica) and ferns (Pteridophytasp).

As illustrated in Fig. 12, *Baeckea frutescens* exhibits a leaf morphology that is narrower compared to *Cratoxylum arborescens*. This distinction is a key factor contributing to the dominant growth of *Baeckea frutescens* on Entisols. The narrower leaves indicate a higher tolerance to evaporation and water scarcity (32), which aligns well with the lower water availability in Entisols compared to Spodosols (33).While the water availability in Spodosols is relatively better than in Entisols, cultivating Spodosols requires more intensive treatment efforts. This is attributed to the presence of a spodic layer in Spodosols (20). The spodic horizon in Spodosols has a limited capacity to retain nutrients, resulting in the degradation of organic matter and the loss of nutrient availability from fertilizers through percolation water (34–36). Despite these challenges, many planters, including both PBN and PBS, persist in cultivating oil palm plantations on Spodosols. Current technology for managing oil palm on Spodosols focuses on aspects such as improving the growing media, breaking hardpan and mounding, enhancing the microclimate, and implementing proper fertilization (37).

In various Entisols landscapes, the potential for dominant plant species to thrive can vary. Nevertheless, plants that typically grow on sandy soils generally share morphological characteristics similar to *Baeckea frutescens*. Plants adapted to less fertile and water-deficient conditions, such as sandy soils, typically exhibit distinctive morphological features (38). These include a deep and well-developed root system that extends into the deeper layers of sandy soils to access water sources (39–41). Furthermore, these plants demonstrate stability and optimize water absorption under environmental stress through morphological adaptations, involving changes in the structure and composition of root tissue (42,43). Adaptation mechanisms, such as increasing the number of lateral roots and root hairs, are observed in plants thriving in water-scarce environments (32,44). Li (45) also highlighted that certain plants on sandy soils develop adventitious roots to cope with water shortages and enhance water absorption.

Ultisols exhibit a higher species diversity value (H'), posing a challenge in identifying a specific dominant plant

species. However, to distinguish between Entisols and

Fig. 12. Dominant species in the Entisols land vegetation (**a**). *Baeckea frutescens* , Spodosols (**b**). *Cratoxylum arborescens* and Ultisols (**c**) *Schima walichi*.

Spodosols, reference can be made to the values of species diversity (H') and species richness (D_mg) in the vegetation's structure and composition. This is attributed to the fertility of Ultisols, contributing to a more abundant diversity and richness of plant species (46). Ultisols possess physical and chemical properties conducive to plant growth and diversity, including higher organic matter and nutrient content, resulting in elevated productivity and plant diversity (47,48). Additionally, regional and topographical factors influence plant dominance in Ultisols, as observed in studies correlating plant growth requirements with the distribution of plant species in specific areas (8). Recent research by Ott (3) and Xiang (49) has emphasized the impact of topography on habitat diversity and quality, extending to fauna as well.

Each region exhibits a distinct structure and composition of vegetation, and these variations are generally influenced by soil characteristics (physical, chemical, biological), climate, topography, and the presence of invasive species that interact with the soil, contributing to unique vegetation features for each soil type (50–53). These discrepancies in the structure and composition of vegetation can serve as indicative signs or general characteristics for the rapid identification of soil types. Differences resulting from variations in vegetation structure and composition in different areas and climates may include species diversity, vegetation height and density, species composition, water availability, and resilience to drought or excess water (54–56).

Vegetation Characteristics Based on Feature Classifier Method

The GIS software utilized the Classify Objects Using Deep Learning function along with the pre-trained deep learning model for predicting soil types in each plot. The software analyzed the image or spatial data of the plots employing the developed algorithm. The outcomes of the soil type prediction were visually displayed on the map (Fig. 13), offering crucial information for users to discern and comprehend the soil characteristics in each plot. Leveraging GIS and deep learning technology rendered this process efficient and accurate. The deep learning model could discern patterns and distinctive features from past soil data, facilitating the classification of soil types with a high degree of accuracy. This information proved valuable in various contexts, enhancing decision-making and resource utilization efficiency based on the unique soil characteristics in each plot.

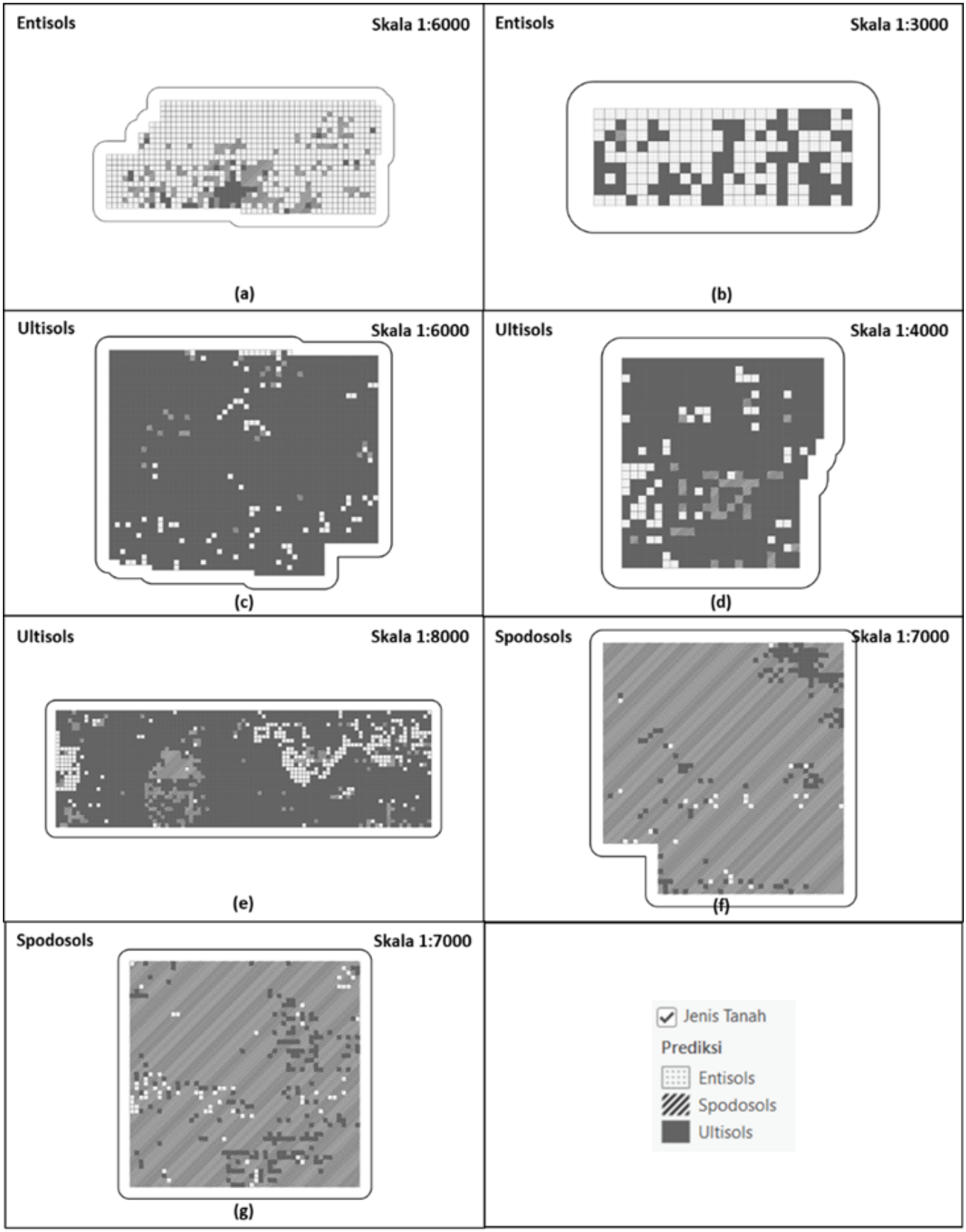


Fig. 13. The results of the prediction of soil types in plots at the study site.

Table 2 presents a comparison between the number of plots with actual soil types and the predicted soil types. In Entisols, depicted in Fig. 13 (a) and (b), the deep learning model exhibited an accuracy of 75% in predicting the total number of plots designated as Entisols. Notably, there was a percentage error of 14% where Entisols were incorrectly predicted as Spodosols and a 12% error where Entisols were predicted as Ultisols. Moving to Spodosols, illustrated in Fig. 13 (f) and (g), the deep learning model demonstrated a commendable accuracy of 91% in predicting the total plots assigned to Spodosols. The prediction errors were minimal, with a 2% error where

Spodosols were predicted as Entisols and a 7% error where Spodosols were predicted as Ultisols. Regarding Ultisols, as shown in Fig. 13 (c), (d), and (e), the deep learning model achieved an accuracy of 86% in predicting the total plots categorized as Ultisols. However, there were errors, with 8% of Ultisols predicted as Entisols and 6% predicted as Spodosols. This analysis underscores that the deep learning model yielded generally satisfactory results in predicting soil types for each plot, albeit with a level of error that requires attention.

Moreover, to assess the performance of the employed deep learning model, the confusion matrix method was applied using the scikit-learn Python Library. The process was executed via a Jupyter notebook, as depicted in Fig. 14. This method facilitated the presentation of the results of the model's performance evaluation through a graphical representation of the confusion matrix, offering a visual depiction of the model's capability to accurately classify soil types. Additionally, the evaluation outcomes were presented in tabular format, encompassing precision, recall, and f1-score metrics for each predicted soil type. The results also included accuracy values for the deep learning model, macro average, and balanced average based on weighted average, as illustrated in Table 7. This comprehensive evaluation furnished detailed insights into the performance and reliability of the deep learning model in soil type prediction, serving as a crucial reference for further assessment and refinement of the model.

The precision values for the deep learning model in predicting the soil types of Entisols, Spodosols, and Ultisols were 0.62, 0.9, and 0.9, respectively. The deep learning model exhibited accurate predictions relative to the total positive predictions, achieving an accuracy rate of 62% for Entisols, 90% for Spodosols, and 90% for Ultisols. In this context, the precision value served as a crucial indicator for evaluating the model's ability to classify soil types with a high degree of accuracy. A higher precision value reflected the model's enhanced capacity to minimize errors in positive predictions.

The recall values for the deep learning model in predicting the soil types of Entisols, Spodosols, and Ultisols were 0.75, 0.91, and 0.86, respectively. This signifies that the deep learning model successfully identified and classified soils correctly, relative to the total negative estimates, achieving a success rate of 75% for Entisols, 91% for Spodosols, and 86% for Ultisols in the study area. In this context, the recall value played a crucial role in evaluating the model's proficiency in accurately identifying and classifying positive soils. A higher recall value indicated the model's improved ability to detect and correctly classify the intended soil type.

Table 7. Performance evaluation values of the deep learning model

	Precision	Recall	F1-Score	Support
Entisols	0.62	0.75	0.68	1170
Spodosols	0.90	0.91	0.90	4740
Ultisols	0.90	0.86	0.88	5293
accuracy			0.87	11203
macro avg	0.81	0.84	0.82	11203
weighted avg	0.87	0.87	0.87	11203

When evaluating the performance of the deep learning model, precision and recall values, along with f1-score and accuracy values, were considered. The f1-score value illustrated the harmony between precision and recall, offering insight into the overall quality of the model's predictions. The evaluation revealed f1-scores of 68% for Entisols, 90% for Spodosols, and 88% for Ultisols. The deep learning model demonstrated a good balance between precision and overall predictions, performing well in predicting the soil types of Spodosols and Ultisols. Furthermore, accuracy served as a measure of the overall model performance, with a value of 87% for predicting the three soil types. This indicated a high success rate in classifying the correct soil type overall. Despite achieving good f1-score and accuracy values, it is essential to emphasize continuous evaluation and correction of the model's performance, addressing any inaccuracies encountered in predicting soil types. Several factors contribute to differences in the confusion matrix results for identifying soil types in this study, including:

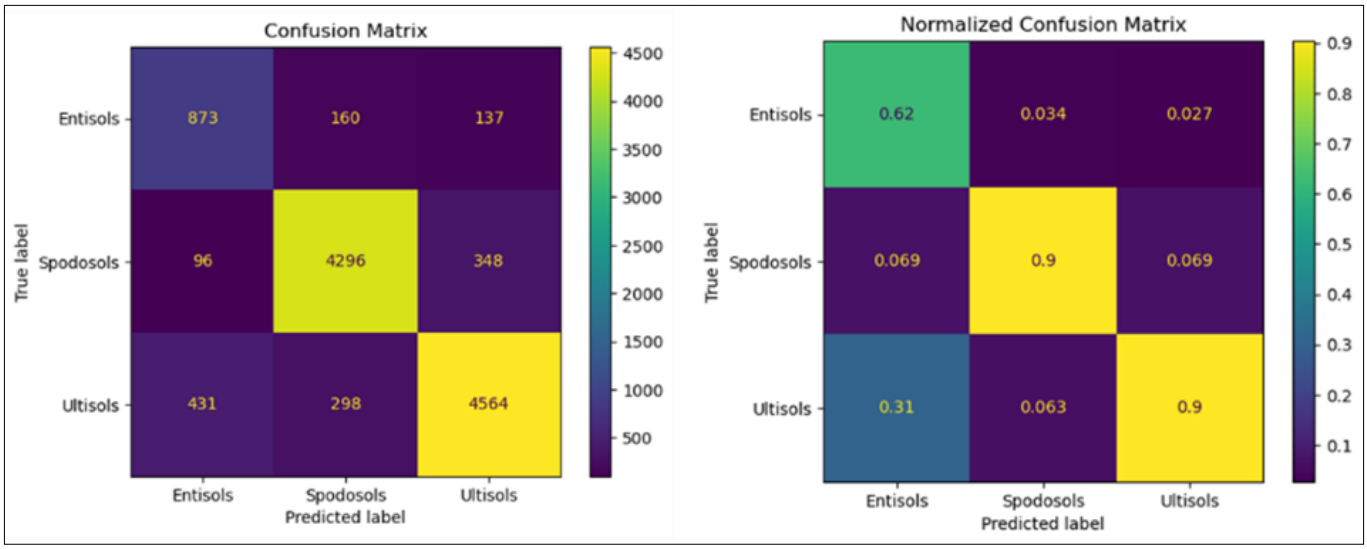


Fig. 14. The left side of the confusion matrix graph shows the number of plots, while the right side shows the precision value of the prediction results.

Factors of the vegetation community composition

The composition of the vegetation community can significantly impact the outcomes of the confusion matrix in soil type identification, particularly for Ultisols, Entisols, and Spodosols. In Ultisols, the prevalence of high levels of seedling and tree vegetation (closed canopy) tends to dominate the overall vegetation composition. This dominance has the potential to influence the confusion matrix results, enhancing the model's capacity to accurately classify Ultisols. The closed canopy formed by these trees imparts distinct characteristics to the Ultisols, facilitating the deep learning model in recognizing specific patterns and features associated with this soil type.

Conversely, in Entisols and Spodosols, the soil type exhibited a vegetation composition dominated by seedling and sapling stages, characterized by a predominantly open canopy. The prevalence of seedlings and saplings imparts distinct features to the soil types of Entisols and Spodosols. The deep learning model is inclined to identify and classify these soil types by considering the characteristics associated with the vegetation composition, primarily characterized by seedling and sapling stages. Consequently, the confusion matrix for Spodosols may reflect a high level of accuracy in classifying this particular soil type.

The lower accuracy observed in Entisols compared to Spodosols may be attributed to variations in vegetation composition. Entisols exhibited a lower tree vegetation level, whereas the seedling and sapling stages had a higher prevalence. This imbalance could pose challenges for the deep learning model in discerning and classifying numerous small plants. Given the dominance of small plants in Entisols, the deep learning model might encounter difficulties in distinguishing and recognizing distinctive features associated with the Entisols soil type. In contrast, Spodosols, with its distinct vegetation composition, facilitated easier identification and classification by the deep learning model. The diverse characteristics of the vegetation composition in each soil type offer specific cues that deep learning models can leverage to enhance the accuracy of soil type identification. A detailed understanding of vegetation patterns and canopy structures associated with specific soil types contributes to refining the performance and precision of the confusion matrix results in soil type identification using the deep learning model.

Factors of species diversity

The greater diversity of soil types in Ultisols provides an advantage in utilizing deep learning models for soil type identification. Ultisols, with 57 species, exhibit higher diversity compared to Spodosols (31 species) and Entisols (14 species). The model, trained on a more extensive range of soil types, can learn and recognize diverse patterns and characteristics, leading to improved accuracy in classifying soil types. This is evident in the 90% accuracy achieved for Ultisols. Spodosols, characterized by lower soil type diversity, may encounter challenges in classifying soil types that are infrequently represented in the training data. Although Spodosols still maintain a relatively high accuracy of 90%, the limited variety of soil types can impact the model's ability to recognize and classify those not well-represented in the training data. Entisols, with the lowest soil type diversity, may face difficulties classifying rarely appearing soil types in the training data, contributing to the lower accuracy of 62%. The restricted number of soil type variations can influence the model's capacity to recognize and classify types that were not well-represented in the training data.

The diversity of soil types within the training data significantly influences the accuracy of soil type identification in the confusion matrix. A higher diversity of plant species in the training data enhances the deep learning model's ability to recognize and classify soil types with greater accuracy. The diversity of plant species plays a crucial role in enriching the dataset with a variety of features and patterns. A broader representation of plant species in the training data provides the deep learning model with more opportunities to learn and comprehend the distinct characteristics associated with each soil type. This comprehensive understanding allows the model to identify soil types more effectively, leading to higher accuracy results in the confusion matrix. Therefore, ensuring that the training data encompasses a wide range of plant species is essential for the model to develop a thorough understanding of the relationships between soil types and existing vegetation communities.

In addition to the factors mentioned above, several crucial elements influence the outcomes of the confusion matrix evaluation in soil type identification. The quality and representativeness of the training data are pivotal in determining the model's predictive accuracy (57,58). If the training data lack diversity in soil types, the model might struggle to classify rarely occurring types or those inadequately represented (59). Moreover, the choice of an algorithm or model that aligns with the soil data's characteristics significantly impacts prediction results (60).The quality and representativeness of the test data also emerge as critical factors, as test data failing to capture variations in soil types can lead to prediction errors (61,62). Human-related factors, including sampling or labeling errors, sampling angle, aerial photo resolution, and data quality, exert a notable influence on prediction outcomes (63,64). Additionally, establishing the threshold for classifying soil types becomes a crucial determinant affecting prediction results and the configuration of the confusion matrix (65). Comprehensive understanding of these factors underscores the importance of meticulous and ongoing evaluation of the confusion matrix to optimize model performance and ensure accurate decision-making in soil type classification applications.

Conclusion

The analysis of vegetation characteristics, incorporating both the examination of vegetation structure and composition and the application of the deep learning model, yielded relatively accurate and consistent

predictions in identifying Entisols, Spodosols, and Ultisols soil types in the study area. The deep learning model exhibited high accuracy in predicting total positive outcomes, with precision values for Entisols (62%), Spodosols (90%), and Ultisols (90%). Furthermore, the model demonstrated its capacity to accurately identify soil types relative to the total negative estimates, achieving recall values of Entisols (75%), Spodosols (91%), and Ultisols (86%). The deep learning model showcased commendable performance in soil type classification, achieving f1-score values of Entisols (68%), Spodosols (90%), and Ultisols (88%). It is worth noting that variations in the composition of the vegetation community and the diversity of plant species can contribute to differences in the confusion matrix results when identifying soil types in Entisols, Spodosols, and Ultisols.

The vegetation community composition in Ultisols was characterized by a prevalence of seedlings and trees, forming a closed canopy. In contrast, the composition of the vegetation community in Entisols and Spodosols was dominated by seedlings and saplings, indicating a predominantly open canopy. Ultisols exhibited superior fertility and suitability for plant growth compared to sandy soils (Spodosols and Entisols), as evidenced by the highest number of species in Ultisols (57 species), followed by Spodosols (31 species) and Entisols (14 species). Ultisols boasted the highest number of species at the seedling level (26 species, 4,901 plants) and sapling level (43 species, 923 plants). However, the overall number of individuals was lower than Spodosols, which had fewer species at the seedling level (17 species, 13,841 plants) and sapling level (25 species, 1,497 plants). The distribution of species in Ultisols, Entisols, and Spodosols was even (J' close to 1), indicating the absence of dominance by specific species (D < 0.5). The species diversity index was at a low to moderate level (H' < 3), and the species richness index was at a very low level (D_{mg} > 3.5).The dominant plants in Entisols (*Baeckea frutescens*) and Spodosols (*Cratoxylum arborescens*) exhibited similar morphologies, characterized by narrow leaves. However, *C. arborescens* had wider leaves. Ultisols displayed a more abundant species diversity value (H') due to its higher fertility level, making it challenging to identify a specific dominant plant species. Dominance was more influenced by regional factors and topography, directly impacting the conditions for plant growth.

Acknowledgements

This project becomes possible because of the financial support of PT. Rimba Harapan Sakti, PT. Sarana Titian Permata, PT. Kerry Sawit Indonesia, and PT. Mustika Sembuluh. I would like to acknowledge Mr. Lo Koon Wai (Plantation Head Kalimantan-Wilmar International Plantation) for providing the necessary resources, facilities, and conducive environment for research. Special thanks also go to the study participants for sharing experiences crucial to the understanding of the research subject, and Mr. Hairul Fatah, Dedy Safriyanto, Fajar Hariadi, Dasrial Efendi and others for the technical support that enhances the rigor of the work. This research would not have been possible without the collaborative efforts of everyone mentioned above. Thank you all for your invaluable contributions.

Authors contributions

Conceptualization, SW, LS, HW, APN, and S; methodology, SW, LS, HW, APN, S, and BD; validation, SW, S, and BD; formal analysis, SW, S, MD, and BD; investigation, SW, LS, HW, APN, S, SP, MD, BD; resources, SW, S, SP, MD, and BD; data curation, SW, LS, APN, S, and BD; writing—original draft preparation, SW, S, and BD; writing—review and editing, SW, LS, HW, APN, S, and BD; supervision, S, SP, MD, and BD; project administration, S; funding acquisition, SW, and S. All authors have read and agreed to the published version of the manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None.

References

1. Zhang D, Peng Y, Li F, Yang G, Wang J, Yu J *et al*. Changes in above-/below-ground biodiversity and plant functional composition mediate soil respiration response to nitrogen input. Funct Ecol. 2021 May 1;35(5):1171-82. https://doi.org/10.1111/1365-2435.137832
2. Rambey R, Susilowati A, Rangkuti AB, Onrizal O, Desrita, Ardi R *et al*. Plant diversity, structure and composition of vegetation around barumun watershed, North Sumatra, Indonesia. Biodiversitas. 2021 Aug 1;22(8):3250-56. https://doi.org/10.13057/biodiv/d2208193
3. Ott RF. How lithology impacts global topography, vegetation and animal biodiversity: A global-scale analysis of mountainous regions. Geophys Res Lett [Internet]. 2020;47(20):1-11. Available from: https://doi.org/10.3929/ethz-b-000449983
4. Cruz ACR, Corrêa N de M, Murakami MM da S, Amorim T de A, Nunes-Freitas AF, Sylvestre L da S. Importance of the vertical gradient in the variation of epiphyte community structure in the Brazilian Atlantic forest. Flora. 2022 Oct 1;295:152137. https://doi.org/10.1016/j.flora.2022.152137
5. Fahey RT, Atkins JW, Gough CM, Hardiman BS, Nave LE, Tallant JM *et al*. Defining a spectrum of integrative trait-based vegetation canopy structural types. Ecology Letters. Blackwell Publishing Ltd. 2019;22:2049-59. https://doi.org/10.1111/ele.13388
6. Zhao X, Feng Y, Xu K, Cao M, Hu S, Yang Q *et al*. Canopy structure: An intermediate factor regulating grassland diversity-function relationships under human disturbances. Fundamental Research. 2022 Mar 1;3:179-87. https://doi.org/10.1016/j.fmre.2022.10.007
7. Walter JA, Stovall AEL, Atkins JW. Vegetation structural complexity and biodiversity in the great smoky mountains. Ecosphere. 2021 Mar 1;12(3). https://doi.org/10.1002/ecs2.3390
8. Anderle M, Paniccia C, Brambilla M, Hilpold A, Volani S, Tasser E *et al*. The contribution of landscape features, climate and topography in shaping taxonomical and functional diversity of

avian communities and a heterogeneous alpine region. Oecologia. 2022 Jul 1;199:499-512. https://doi.org/10.1007/s00442-022-05134-7

9. Chen J, Jin S, Du P. Roles of horizontal and vertical tree canopy structure in mitigating daytime and nighttime urban heat island effects. International Journal of Applied Earth Observation and Geoinformation. 2020 Jul 1;89:1-11. https://doi.org/10.1016/j.jag.2020.102060
10. Verduchi V. How complexity in vegetation structure and distance from the ground-level habitat influence spontaneous plant diversity on green-roofs. 2021 Dec.
11. Hamka H, Hapid A, Maiwa A. Analisis vegetasi di kawasan lindung desa betania kabupaten poso. Jurnal Pendidikan MIPA. 2022 Sep 16;12(3):808-13. https://doi.org/10.37630/jpm.v12i3.688
12. Singh V, Shukla S, Singh A. The principal factors responsible for biodiversity loss. Open J Plant Sci [Internet]. 2021;6(1):11-14. Available from: https://dx.doi.org/10.17352/jps.000026
13. Gizachew GT. Spatial-temporal and factors influencing the distribution of biodiversity: A review. Scientific Reports in Life Sciences. 2021;2(4):1-19.
14. Damptey FG, Birkhofer K, Menor IO, de la Riva EG. The functional structure of tropical plant communities and soil properties enhance ecosystem functioning and multifunctionality in different ecosystems in Ghana. Forests. 2022 Feb 1;13(2). https://doi.org/10.3390/f13020297
15. Javed A, Ali E, Binte Afzal K, Osman A, Riaz DrS. Soil fertility: Factors affecting soil fertility and biodiversity responsible for soil fertility. International Journal of Plant, Animal and Environmental Sciences. 2022;12(01). https://doi.org/10.26502/ijpaes.202129
16. Montagna M, Brunetti M, Spada A, Cussigh A, Alali S, Cremonesi P *et al*. Factors affecting soil invertebrate biodiversity in agroecosystems of the po plain area (Italy). ARPHA Conference Abstracts. 2022 Sep 29;5. https://doi.org/10.3897/aca.5.e95808
17. Kooch Y, Noghre N. The effect of shrubland and grassland vegetation types on soil fauna and flora activities in a mountainous semi-arid landscape of Iran. Science of The Total Environment. 2020 Feb 10;703:135497. https://doi.org/10.1016/j.scitotenv.2019.135497
18. Lu P, Tan Y, Dai N, Di M, Weng X, Zhan Y *et al*. Composition and structure of soil fauna communities and their relationships with environmental factors in copper mine waste rock after re-vegetation. Glob Ecol Conserv. 2021 Dec 1;32:e01889. https://doi.org/10.1016/j.gecco.2021.e01889
19. Chaparro MAE, Moralejo M del P, Böhnel HN, Acebal SG. Iron oxide mineralogy in Mollisols, Aridisols and Entisols from Southwestern Pampean region (Argentina) by environmental magnetism approach. Catena (Amst). 2020 Jul 1;190:104534. https://doi.org/10.1016/j.catena.2020.104534
20. Suwardi, Sutiarso L, Wirianata H, Nugroho AP, Sukarman, Primananda S. Substantial changes in physical and chemical properties of spodosols soil by hardpan breaking and mounding in oil palm plantation. In: Proceedings of the International Conference on Sustainable Environment, Agriculture and Tourism (ICOSEAT 2022). Atlantis Press. 2023. https://doi.org/10.2991/978-94-6463-086-2_20
21. Junior AJ de S, Camêlo D de L, Arruda DL de, Souza Junior VS de, Rocha AT da, Corrêa MM. Spodosol formation on sandy ruins in a semi-arid climate in the Catimbau National Park, Northeast Brazil. Catena (Amst). 2023 Aug 1;229:107226. https://doi.org/10.1016/j.catena.2023.107226
22. Shi Y xiao xiao, Cui J qi, Zhang F, Li K wei, Jiang J, Xu R kou. Effects of soil pH and organic carbon content on *in vitro* Cr bioaccessibility in Ultisol, Alfisol, and Inceptisol. Chemosphere. 2023 Sep 1;336:139274. https://doi.org/10.1016/j.chemosphere.2023.139274
23. Xu P, Liu Y, Zhu J, Shi L, Fu Q, Chen J *et al*. Influence mechanisms of long-term fertilizations on the mineralization of organic matter in Ultisol. Soil Tillage Res. 2020 Jul 1;201:104594. https://doi.org/10.1016/j.still.2020.104594
24. Tamang B, Hedman C, Haines F, Stone D, Andreu M. Upland forest community composition and structure by ecoregion in 73 Florida state parks – Insights for ongoing management. For Ecol Manage [Internet]. 2023 Oct 1 [cited 2023 Jul 16];545:121237. https://doi.org/10.1016/j.foreco.2023.121237
25. Jamison EAK, D'Amato AW, Dodds KJ. Describing a landscape mosaic: Forest structure and composition across community types and management regimes in inland northeastern pitch pine barrens. For Ecol Manage. 2023 May 15;536:120859. https://doi.org/10.1016/j.foreco.2023.120859
26. Morgan GR, Wang C, Li Z, Schill SR, Morgan DR. Deep learning of high-resolution aerial imagery for coastal marsh change detection: A comparative study. ISPRS Int J Geoinf. 2022 Feb 1;11(2):1-21. https://doi.org/10.3390/ijgi11020100
27. Jiang G, Zheng Q. Remote sensing recognition and classification of forest vegetation based on image feature depth learning. Mobile Information Systems. 2022;2022:1-11. https://doi.org/10.1155/2022/9548552
28. Liu M, Fu B, Xie S, He H, Lan F, Li Y *et al*. Comparison of multi-source satellite images for classifying marsh vegetation using DeepLabV3 Plus deep learning algorithm. Ecol Indic. 2021 Jun 1;125:107562. https://doi.org/10.1016/j.ecolind.2021.107562
29. Fu B, Zuo P, Liu M, Lan G, He H, Lao Z *et al*. Classifying vegetation communities karst wetland synergistic use of image fusion and object-based machine learning algorithm with Jilin-1 and UAV multispectral images. Ecol Indic. 2022 Jul 1;140:108989. https://doi.org/10.1016/j.ecolind.2022.108989
30. Paramananthan S. Managing marginal soils for sustainable growth of oil palms in the tropics. Journal of Oil Palm and the Environment. 2013 Jan;4(1):1-16. https://doi.org/10.5366/jope.2013.1
31. Hartati W, Arifin J, Sudarmadji T, Ruhiyat D. Spodosols of east kalimantan: Land cover disturbances induced degradation of soil properties. In: Proceedings of the Joint Symposium on Tropical Studies (JSTS-19). Atlantis Press BV. 2021; p. 403-09. https://doi.org/10.2991/absr.k.210408.066
32. Meng F, Zhang T, Yin D. The effects of soil drought stress on growth characteristics, root system and tissue anatomy of *Pinus sylvestris* var. *mongolica*. PeerJ. 2023 Jan 9;11. https://doi.org/10.7717/peerj.14578
33. Sukarman, Saidy AR, Rusmayadi G, Adriani DE, Primananda S, Suwardi *et al*. Effect of water deficit of Ultisols, Entisols, Spodosols and Histosols on oil palm productivity in Central Kalimantan. Sains Tanah. 2022 Dec 1;19(2):180-91. https://doi.org/10.20961/stjssa.v19i2.65455
34. Nottingham AC, Thompson JA, Wood F, Edwards PJ, Strager MP. Mapping pedomemory of spodic morphology using a species distribution model. Geoderma. 2019 Oct 15;352:330-41. https://doi.org/10.1016/j.geoderma.2017.10.044
35. Schaetzl RJ, Kasmerchak C, Samonil P, Baish C, Hadden M, Rothstein D. Acidification and weathering associated with deep tongues in sandy Spodosols, Michigan, USA. Geoderma Regional. 2020 Dec 1;23:e00332. https://doi.org/10.1016/j.geodrs.2020.e00332
36. Souza Junior AJ de, Camêlo D de L, Arruda DL de, Souza Junior VS de, Rocha AT da, Corrêa MM. Spodosol formation on sandy ruins in a semi-arid climate in the Catimbau National Park, Northeast Brazil. Catena (Amst). 2023 Aug 1;229:107226. https://doi.org/10.1016/j.catena.2023.107226

37. Suwardi. Manajemen tanah spodosols melalui sistem pecah hardpan dan mounding untuk meningkatkan produksi tanaman kelapa sawit. Instiper Yogyakarta. Instiper. 2021.
38. Ali Z, Merrium S, Habib-ur-Rahman M, Hakeem S, Saddique MAB, Sher MA. Wetting mechanism and morphological adaptation; leaf rolling enhancing atmospheric water acquisition in wheat crop—A review. Environmental Science and Pollution Research. 2022 May 1;29(21):30967-85. https://doi.org/10.1007/s11356-022-18846-3
39. Cai G, Carminati A, Abdalla M, Ahmed MA. Soil textures rather than root hairs dominate water uptake and soil-plant hydraulics under drought. Plant Physiol. 2021 Oct 1;187:858-72. https://doi.org/10.1093/plphys/kiab271
40. Shoaib M, Banerjee BP, Hayden M, Kant S. Roots' drought adaptive traits in crop improvement. Plants. MDPI. 2022;11:1-20. https://doi.org/10.3390/plants11172256
41. Dong X, Zhang Z, Wang S, Shen Z, Cheng X, Lv X *et al*. Soil properties, root morphology and physiological responses to cotton stalk biochar addition in two continuous cropping cotton field soils from Xinjiang, China. PeerJ. 2022 Feb 16;10. https://doi.org/10.7717/peerj.12928
42. Phillips AJ, Govedich FR, Moser WE. Leeches in the extreme: Morphological, physiological, and behavioral adaptations to inhospitable habitats. Int J Parasitol Parasites Wildl. 2020 Aug 1;12:318-25. https://doi.org/10.1016/j.ijppaw.2020.09.003
43. Li WT, Liang N, Zhan J, Wang H, Zhang P. Morphological and anatomical characteristics of eelgrass *Zostera marina* L. at two distinct environments of Shandong Peninsula, China: An implication of adaptation strategy of seagrasses. Aquat Bot. 2023 May 1;186:103612. https://doi.org/10.1016/j.aquabot.2022.103612
44. Lynch JP, Strock CF, Schneider HM, Sidhu JS, Ajmera I, Galindo-Castañeda T *et al*. Root anatomy and soil resource capture. Plant and Soil. Springer Science and Business Media Deutschland GmbH. 2021;466:21-63. https://doi.org/10.1007/s11104-021-05010-y
45. Li X, Dang X, Gao Y, Meng Z, Chen X, Wang Y. Response mechanisms of adventitious root architectural characteristics of *Nitraria tangutorum* shrubs to soil nutrients in Nabkha. Plants. 2022 Dec 1;11(23):1-17. https://doi.org/10.3390/plants11233218
46. Sellan G, Thompson J, Majalap N, Brearley FQ. Soil characteristics influence species composition and forest structure differentially among tree size classes in a Bornean Heath forest. Plant Soil. 2019;1-39. https://doi.org/10.1007/s11104-019-04000-5
47. Kome GK, Enang RK, Tabi FO, Yerima BPK. Influence of clay minerals on some soil fertility attributes: A review. Open Journal of Soil Science. 2019;09:155-88. https://doi.org/10.4236/ojss.2019.99010
48. Reichert JM, Morales B, Lima EM, de Bastos F, Morales CAS, de Araújo EF. Soil morphological, physical and chemical properties affecting *Eucalyptus* spp. productivity on Entisols and Ultisols. Soil Tillage Res. 2023 Feb 1;226:105563. https://doi.org/10.1016/j.still.2022.105563
49. Xiang Q, Kan A, Yu X, Liu F, Huang H, Li W *et al*. Assessment of topographic effect on habitat quality in mountainous area using InVEST model. Land (Basel). 2023 Jan 1;12(1):1-17. https://doi.org/10.3390/land12010186
50. Shang R, Li S, Huang X, Liu W, Lang X, Su J. Effects of soil properties and plant diversity on soil microbial community composition and diversity during secondary succession. Forests. 2021 Jun 1;12(805):1-12. https://doi.org/10.3390/f12060805
51. Kumar U, Saqib HSA, Islam W, Prashant P, Patel N, Chen W *et al*. Landscape composition and soil physical–chemical properties drive the assemblages of bacteria and fungi in conventional vegetable fields. Microorganisms. 2022 Jun 1;10(1202):1-18. https://doi.org/10.3390/microorganisms10061202
52. Zaldívar-Cruz B, Pérez-Ceballos R, Zaldívar-Jiménez A, Canales-Delgadillo J, Endañu-Huerta E, Flores AB *et al*. Structural and diversity changes in coastal dunes from the Mexican Caribbean: The case of the invasive Australian pine (*Casuarina equisetifolia*). Management of Biological Invasions. 2022 Mar 1;13(1):131-46. https://doi.org/10.3391/mbi.2022.13.1.08
53. de Francesco MC, Tozzi FP, Buffa G, Fantinato E, Innangi M, Stanisci A. Identifying critical thresholds in the impacts of invasive alien plants and dune paths on native coastal dune vegetation. Land (Basel). 2023 Jan 1;12(135):1-16. https://doi.org/10.3390/land12010135
54. Debouk H, Emeterio LS, Marí T, Canals RM, Sebastià MT. Plant functional diversity, climate and grazer type regulate soil activity in natural grasslands. Agronomy. 2020 Sep 1;10(1291):1-20. https://doi.org/10.3390/agronomy10091291
55. Diatta SBD, Tall LN, Ndour YB, Sembene M, Assigbetsé K. Composition and diversity of soil bacterial communities along an environmental gradient in the Sudano-Sahelian region of Senegal. Open Journal of Soil Science. 2020;10(02):58-89. https://doi.org/10.4236/ojss.2020.102004
56. Yaseen M, Fan G, Zhou X, Long W, Feng G. Plant diversity and soil nutrients in a tropical coastal secondary forest: Association ordination and sampling year differences. Forests. 2022 Mar 1;13(376):1-10. https://doi.org/10.3390/f13030376
57. Liu Y, Yin X, Yue G, Zheng Z, Jiang J, He Q *et al*. Blind omnidirectional image quality assessment with representative features and viewport oriented statistical features. J Vis Commun Image Represent. 2023 Mar 1;91:103770. https://doi.org/10.1016/j.jvcir.2023.103770
58. Stampoulis D, Damavandi HG, Boscovic D, Sabo J. Using satellite remote sensing and machine learning techniques towards precipitation prediction and vegetation classification. Journal of Environmental Informatics. 2021;37(1):1-15.
59. Saputro IW, Sari BW. Uji performa algoritma naïve bayes untuk prediksi masa studi mahasiswa. Citec Journal. 2019;6(1):1-11. https://doi.org/10.24076/citec.2019v6i1.178
60. Piaser E, Villa P. Evaluating capabilities of machine learning algorithms for aquatic vegetation classification in temperate wetlands using multi-temporal sentinel-2 data. International Journal of Applied Earth Observation and Geoinformation. 2023 Mar 1;117:1-12. https://doi.org/10.1016/j.jag.2023.103202
61. Yang P. Exploring the interrelated effects of soil background, canopy structure and sun-observer geometry on canopy photochemical reflectance index. Remote Sens Environ. 2022 Sep 15;279:113133. https://doi.org/10.1016/j.rse.2022.113133
62. Atemkeng CC, Tapimo R, Tonnang EHZ, Tchinda R. Inverse radiative transfer problem for soil properties retrieval from bidirectional reflectance measurements. Results in Optics. 2023 May 1;11:100409. https://doi.org/10.1016/j.rio.2023.100409
63. Bouguettaya A, Zarzour H, Kechida A, Taberkit AM. Deep learning techniques to classify agricultural crops through UAV imagery: A review. Neural Computing and Applications. Springer Science and Business Media Deutschland GmbH. 2022;34:9511-36. https://doi.org/10.1007/s00521-022-07104-9
64. Ramli NE, Yahya ZR, Said NA. Confusion matrix as performance measure for corner detectors. Journal of Advanced Research in Applied Sciences and Engineering Technology. 2022;29(1):256-65. https://doi.org/10.37934/araset.29.1.256265
65. Asfour M, Menon C, Jiang X. Feature–classifier pairing compatibility for sEMG signals in hand gesture recognition under joint effects of processing procedures. Bioengineering. 2022 Nov 1;9(11):1-18.