

# Evaluation of Crop Yield from Biochar-Induced Soil Through Rough Set Approach

Sofia Nicole D. Andres\*, Louie Alfredrich E. Ducusin, Elyssa Vernadette R. Marcadejas, Beatriz A. Belmonte

Research Center for the Natural and Applied Sciences/Chemical Engineering Department, University of Santo Tomas, España Blvd., 1015 Manila, Philippines  
[sofianicole.andres.eng@ust.edu.ph](mailto:sofianicole.andres.eng@ust.edu.ph)

For years, biochar has gained increasing interest in the scientific community due to its significant potential to sequester carbon from the atmosphere and improve the physicochemical properties of soil which in turn increases crop yield. Multiple investigations on crop productivity using biochar-induced soil were done before but gave variable results. Different biochar properties, production methods, and application conditions have led to varied responses when applied to different soils, ranging from positive to neutral or even negative crop yield effects, necessitating the need to identify the most suitable combination of parameters to achieve the most favorable outcome. This study developed a model to maximize the beneficial effects of biochar in agricultural settings with the aid of rough set-based machine learning (RSML). Four if-then rules were accepted correlating the feedstock type, application rate, pyrolysis temperature, and soil type to the % change in crop yield. The coverage of Rules 1, 2, 3, and 4 in the training set are 19 %, 14 %, 11 %, and 6 % with an accuracy of 100 %. They also cover 13 %, 21 %, 14 %, and 4 % of the validation set at 100 % accuracy. The findings indicate that these condition attributes can have a notable impact on crop yield in biochar-induced soil. This study can also guide the agricultural sector in choosing the appropriate biochar parameters to improve soil quality and maximize crop productivity.

## 1. Introduction

In the pursuit of sustainable agriculture amidst escalating environmental concerns, integration of innovative agricultural practices is imperative. One promising and sustainable approach is the application of biochar to soil (Woolf et al., 2010). For years, there has been an increasing interest in investigating the potential of biochar in soil amendment and as a climate change mitigation tool. Biochar, a solid porous byproduct obtained from pyrolysis, the heating of biomass feedstock under a condition with limited or no oxygen (Brassard et al., 2019), has been widely known as a good addition to agricultural activity because of its auspicious effect on soil properties which in turn enhances crop yield (Novak et al., 2009). The production of biochar can also displace the use of fossil fuels by energy co-products (syngas and bio-oil). Most of the carbon in biochar is recalcitrant. Hence, the production of biochar from biomass feedstock via pyrolysis and inducing it to soil has great prospects for long-term withdrawal of CO<sub>2</sub> from the atmosphere as it is expected to sequester carbon for centuries (Brassard et al., 2016). This increase in soil carbon sequestration contributes to the improvement of soil quality due to the vital role of carbon in chemical, physical, and biological soil processes (Novak et al., 2009). Increased crop productivity is one prominent effect of biochar application to soil (Singh et al., 2022). Multiple investigations on crop productivity using biochar-induced soil have shown variable results due to the variation in biochar properties, soil conditions, and experimental conditions (Dai et al., 2020). The effect of the different combinations of the said parameters remains unclear. Such variations can be elucidated to find the optimum set of parameters and attain the most favorable outcome. The following process parameters highly influence the physicochemical properties of the biochar product: the type of biomass feedstock and the operating conditions during pyrolysis (Novak et al., 2009). Novak et al. (2009) theorized that the production process of

biochar could be engineered to generate designer biochars which consist of distinct physical and chemical characteristics suitable for the soil conditions where the biochar would be applied.

Data from previous studies focusing on the effects of biochar application on crop yield can be used to find trends, patterns, and correlations between the different parameters involved by using artificial intelligence. Artificial Intelligence (AI) is the ability of a computer derived from human intelligence, to develop advanced systems in order to keep up with the diversity and complexity of the data involved (Joiner, 2018). Machine Learning (ML) makes it possible to conduct large-scale observations, improve, and extend itself through the application of new knowledge (Woolf, 2009) as it uses computers to learn general concepts from the training sets, a subset used to generate rules (Hvidsten, 2013). ML is proven to be effective and more convenient especially with large data gathered ready for treatment. ML has been used in maximizing the capability of biochar by predicting its total yield and specific area (Hai et al., 2023) and in evaluating the removal of heavy metals by biochar (Liu et al., 2023) and from industrial wastewater (Dashti et al., 2023). Its application is not limited to this field, for it is widely used in numerous industries such as medicine, agriculture, food engineering, etc.

Huge data sets are often indiscernible and difficult to characterize precisely. A mathematical approach to decision-making that can deal with such problems is the rough set theory (RST). RST is a useful tool for decision support systems that deal with vagueness, uncertainty, and imprecision of data involved in the decision process (Pawlak, 1997). This study pioneers the use of a rough set approach with the aid of machine learning to develop a rule-based model to forecast the influence of several factors on crop yield after subjecting the soil to biochar application. If-then rules were generated and interpreted in this study, which can be readily applied in practical situations. The findings of this study can guide the agriculture sector in choosing the appropriate biochar parameters to improve soil quality, maximize crop productivity, and optimize the soil's climate change mitigation potential.

## 2. Methodology

In developing the rule-based model, six steps were followed (Belmonte et al., 2023), as shown in Figure 1. The dataset used is the supplementary data on the influence of biochar applications on soil's physicochemical properties, ranging from 2012 to 2021, obtained from the meta-analysis study of Singh et al. (2022). The meta-analysis accepted a total of 59 studies that provided comparisons between the control (no biochar) and amendments (treated w/ biochar) groups. The effects of various condition attributes such as feedstock type, application rate (t/ha), pyrolysis temperature (°C), and soil type on % change in crop yield (decision attribute) were evaluated.

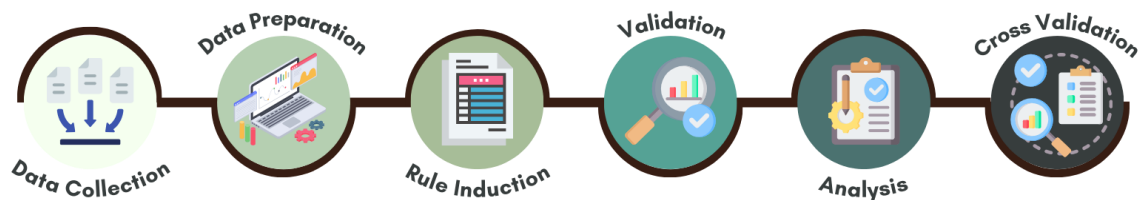


Figure 1. Flow diagram of the research methodology

A rough set-based machine learning (RSML), which uses Pawlak's rough set theory and Boolean reasoning for inducing the rules are the framework implemented in this study (Hvidsten, 2013). The raw data collected were pre-processed and discretized using the ROSETTA software. The discretized data include feedstock type, application rate (t/ha), pyrolysis temperature (°C), soil type, and % change in crop yield, and were represented with corresponding number-coded values, as shown in Table 1. From the data obtained, biochar-induced soils have varied responses, ranging from positive (177.586 %) to neutral (0.000 %) or even negative (27.607 %) crop yield effects.

Rule induction was done using Johnson's algorithm, a fast algorithm that uses a greedy search to find one reduct and produces rules in the ROSETTA software (Hvidsten, 2013). From the supplementary data, a total of 122 objects were acquired, and these were randomly divided into two subsets: 85 objects (70 %) were used for the training set, a subset for generating the rule-based models by learning the patterns present in the data; and 37 objects (30 %) were used for the validation set, a subset for validating the performance of the trained rule-based models. The training sets are then reduced to minimal sets of condition attributes called reducts that are sufficient for describing the decision attribute (Gue et al., 2021).

After the generation of the rule-based models, four if-then rules that exhibited high performance based on the percentage coverage were accepted and validated using the validation set. The coverage factor,  $cov_S(\Phi, \Psi)$ , represented by Eq(1), is a measure of how well the decision attribute ( $\Psi$ ) covers all the necessary condition attributes ( $\Phi$ ). A high coverage factor means that all or most of the condition attributes were considered, while a very low coverage factor means that there are still parameter/s not being considered in the study (Pawlak, 2002). On the other hand, the certainty factor,  $cer_S(\Phi, \Psi)$ , also known as accuracy factor, represented by Eq(2), is a measure of how strongly the combination of condition attributes ( $\Phi$ ) suits and supports the decision attribute ( $\Psi$ ). A high certainty factor means that there is a strong relationship between the condition attributes and the decision attribute, while a very low certainty factor means a little to no relationship at all (Pawlak, 2002). That is why in the ROSETTA System, the value for the % Accuracy for each generated rule is always 100 % unless the then-part contains several decisions (Hvidsten, 2013).

$$cov_S(\Phi, \Psi) = \frac{card(\|\Phi \wedge \Psi\|_S)}{card(\|\Psi\|_S)} \quad (1)$$

$$cer_S(\Phi, \Psi) = \frac{card(\|\Phi \wedge \Psi\|_S)}{card(\|\Phi\|_S)} \quad (2)$$

The mechanistic plausibility of the accepted four top-performing if-then rules was further assessed and proven using various supporting biochar literature. Lastly, a 10-fold cross-validation was done to prevent overfitting. Overfitting occurs when the rule-based model performs well on the training dataset but poorly on new or unseen dataset (Hvidsten, 2013).

Table 1: Discretization of the attributes

Feedstock Type		Application Rate (t/ha)
1	herbaceous	1 [1, 3)
2	lignocellulosic	2 [3, 13)
3	wood	3 [13, 60)
Soil Type		Pyrolysis Temperature (°C)
1	{loamy, sandy loam}	1 [200, 325)
2	{loamy silt clay}	2 [325, 463)
3	{loam, sandy, silt loam}	3 [463, 800)
4	{loamy sand, silty clay}	
5	{sandy clay loam, silty clay loam, sandy loam}	
		% Change in Crop Yield
		1 [-27.607, 11.642)
		2 [11.642, 177.586)

### 3. Results and Discussion

This section discusses how the rules were generated and evaluated to determine the percentage change in crop yield after subjecting the soil to biochar application. Pertinent studies in the biochar literature that corroborate the validity of the accepted rules are further explained.

Table 2 is a confusion matrix, which is one of the common metrics for evaluating the model performance. It comprises columns with the predicted percentage change in crop yield values generated by the rule-based model, and rows with the actual values. In total, the model accurately classified 75.29 % of the training dataset. Notably, among the 122 objects in the dataset, 85 were utilized as the training set, with 64 of them being correctly classified as highlighted in green, which are the reducts that are sufficient for describing the two discretized ranges of the decision attribute (Hvidsten, 2013). This means that 44 objects in the dataset indicate the number of correct predictions for the % change in crop yield ranging from -27.607 % to 11.642 %, while the 20 objects are for the % change in crop yield range of 11.642 % to 177.586 %.

Table 2: Overall performance of the rules generated from the training set

		Predicted		
		1	2	
Actual	1	44	4	Overall Accuracy 75.29 %
	2	17	20	

The rules with the highest performance in terms of both percentage coverage and percentage accuracy from the training set (refer to Table 4) were accepted and given in Table 3. Rule 1 specifies that if the application rate ranges from 3 to 13 t/ha, the pyrolysis temperature is 463 °C – 800 °C, and the soil type is sandy clay loam, silty clay loam, or sandy loam, then the % change in crop yield ranges from -27.607 % to 11.642 %. Comparably, Rule 4 states that if the application rate is 13 t/ha – 60 t/ha, the pyrolysis temperature is 463 °C – 800 °C, and the soil type is loamy sand or silty clay, then the % change in crop yield ranges from -27.607 % to 11.642 %. While it is acceptable that all these attributes (application rate, pyrolysis temperature, and soil type) can affect crop productivity, it can be deduced from these rules that the pyrolysis temperature is the most influential factor that can affect crop productivity among the condition attributes in these rules. This further implies that the biochar produced at a pyrolytic temperature equal to or greater than 463 °C will result in a % change in crop yield that is less than or equal to 11.642 %. The study by Adekiya et al. (2019) showed that wood-based biochars produced at a pyrolysis temperature of 580 °C with an application rate of 25 t/ha resulted in 0 %, 4 %, and 4.55 % changes in crop yield from differing crop yield controls. The paper of Tammeorg et al. (2014) revealed that wood-based biochars produced at a pyrolysis temperature of 600 °C with an application rate varying from 5, 10, 20, and 30 t/ha resulted in negative % changes in crop yield ranging from -2.96 % to -20 %. The temperature during pyrolysis has a significant influence on the physicochemical properties of the biochar (Ding et al., 2014), which in turn affects plant growth when applied to soil. A linear increase in biochar pH was observed with increasing pyrolytic temperature (Li et al., 2019). Application of biochar to soil can further increase soil pH. Soil pH is determined as one of the key factors influencing plant growth (Dai et al., 2017) since the availability of nutrients for plants is pH-dependent (Purakayastha et al., 2019). The study of Jeffery et al. (2011) revealed that the increase in soil pH was found to be positively correlated to the increased amount of plant productivity. However, it is also important to note that optimal pH levels should be maintained as soils with high pH can hinder plant growth due to toxic concentrations of mineral elements and/or low nutrient availability (Läuchli, 2012). Purakayastha et al. (2019) denoted that biochars produced at a lower pyrolysis temperature could offer greater soil nutrient availabilities than that prepared at a higher pyrolysis temperature. In the training set, the coverage achieved for Rule 1 and Rule 4 are 19 % and 6 % with 100 % accuracy.

*Table 3: Accepted rules for the crop yield response of biochar-induced soil*

Top Performing Rules	
Rule 1	IF Application Rate (t/ha) = 2 AND Pyrolysis Temperature (°C) = 3 AND Soil Type = 5 THEN % Change in Crop Yield = 1
Rule 2	IF Feedstock Type = 1 AND Pyrolysis Temperature (°C) = 2 THEN % Change in Crop Yield = 2
Rule 3	IF Feedstock Type = 1 AND Application Rate (t/ha) = 3 THEN % Change in Crop Yield = 2
Rule 4	IF Application Rate (t/ha) = 3 AND Pyrolysis Temperature (°C) = 3 AND Soil Type = 4 THEN % Change in Crop Yield = 1

Rule 2 states that if the feedstock is herbaceous, and the pyrolysis temperature is within the range of 325 °C to 463 °C, then the % change in crop yield is 11.642 % - 177.586 %. Similarly, Rule 3 specifies that if the biochar feedstock is herbaceous, and the application rate is 13 t/ha – 60 t/ha, then the % change in crop yield is 11.642 % - 177.586 %. These results imply that herbaceous-based biochar has positive response and can significantly increase crop yield when applied to soil. From the meta-analysis of Singh et al. (2022), among the other feedstock types, herbaceous-based biochar had resulted to greatest significant increase in crop yield. It was followed by the lignocellulosic-based and wood-based biochars. Herbaceous-based biochar has greater amounts of nutrients compared to biochar produced from other feedstocks (Singh et al, 2022). Specifically compared to wood-derived biochar, Latini et al. (2019) reported that biochar made from wheat straw, which is herbaceous, has higher ash %, cation exchange capacity (CEC), contents of N, P, Ca, and Mg, pH, and wt. % of C, Na, and K, which are nutrients necessary to facilitate crop growth (Tandzi and Mutengwa, 2020). The impact of biochar properties on crop yields appeared to be most prominently influenced by the temperature during the pyrolysis process (Ye et. al., 2019). Biochars produced at pyrolytic temperature less than or equal to 400 °C resulted to largest increase in crop yield (Ye et. al., 2019). It is also important to note that these biochars were mostly produced from cereal residues, a type of herbaceous feedstock. On the other hand, biochars produced at temperatures greater than 500 °C induced a smaller effect on crop yield. This is consistent with the findings of Li et al. (2019) and Han et al. (2023), wherein the pyrolysis temperature range of 400 – 500 °C proved to be the most beneficial in enhancing crop yields, while biochars produced at elevated temperatures, particularly exceeding 600 °C, resulted in a decline in crop productivity. Biochars prepared at high pyrolytic temperatures is capable of tightly holding water and dissolved minerals. Restricting the accessibility of essential water and minerals that the crop needs for their development (Li et al., 2019). It was found that the effects of

pyrolysis temperature on crop yield were similar to the effects on N retention, which may be hypothesized that crop growth is closely connected to the availability of N in soil (Li et al., 2019) which is a nutrient that can be found in greater amounts in herbaceous biochars compared to biochars produced from other feedstocks. In the training set, the coverage achieved for Rule 2 and Rule 3 are 14 % and 11 % with 100 % accuracy.

The accepted if-then rules for the crop yield response of biochar-induced soil underwent validation to evaluate their performance. Shown in Table 4 are the percentage coverage and percentage accuracy of the rules in the validation dataset where all rules performed well as reflected on their coverages. Rule 2 has the highest coverage of 21 %, followed by Rule 3 with 14 %, Rule 1 with 13 %, and Rule 4 with 4 %.

To prevent overfitting and gain further insight into the performance of the rules, 10-fold cross-validation was also done. The dataset underwent 10 random divisions into the training set and validation set, each with 90 % of 122 data points utilized as the training set and 10 % allocated to the cross-validation set.

The repeatedly occurring top-performing rules identified through cross-validation are tabulated in Table 4, indicating their frequency of appearance, the highest percentage coverage, and the highest percentage accuracy. Rule 1 was present in 8 trials, Rules 2 and 4 in 9 trials, and Rule 3 in 10 trials. This implies that the four top-performing rules were evaluated multiple times, showing good performance not only on the training set (70-30 split), but also on the new trained dataset (90-10 split). Lastly, this result indicates that the performance of the four top-performing rules is satisfactory during the first validation through cross-validation. This is an implication that the methodology implemented using the software ROSETTA yields acceptable outcomes.

*Table 4: Performance of the top-performing rules*

Rules	Training Dataset		Validation Dataset		10-Fold Cross Validation		
	% Coverage	% Accuracy	% Coverage	% Accuracy	Frequency of Occurrences	Highest % Coverage	Highest % Accuracy
1	19	100	13	100	8	42	100
2	14	100	21	100	9	100	100
3	11	100	14	100	10	33	100
4	6	100	4	100	9	25	100

#### 4. Conclusion

Rough-set theory was employed with the aid of machine learning to develop a model to forecast the influence of different factors on % change in crop yield after subjecting the soil to biochar application. The model consists of if-then rules correlating the feedstock type, application rate, pyrolysis temperature, and soil type to the % change in crop yield. The performance of the rules was further evaluated for mechanistic plausibility. The findings indicate that these condition attributes can have a notable impact on crop yield in biochar-induced soil. A 10-fold cross-validation was done to ensure that the generated rule-based model performed well not only in the training set but also in the new trained dataset of the cross-validation. The results exhibit good performance from the first validation until the cross-validation. Hence, the findings of this study can guide the agricultural sector in choosing the appropriate biochar parameters to improve soil quality, maximize crop productivity, and optimize the soil's climate change mitigation potential. The rule-based model developed in this study can guide the production and application of biochar suitable for improving soil properties, which translates to increased crop productivity. Further research can consider additional parameters not included in this study and utilize larger datasets from other meta-analyses to enhance the prediction accuracy of the rough set-based model while adapting the same methodology. These parameters might include the type of crop to be planted, the location of biochar application, and the initial conditions of the soil's physicochemical properties.

#### Acknowledgement

The authors gratefully acknowledge the financial support from the Department of Science and Technology – Science Education Institute (DOST-SEI), Philippines and the Research Center for the Natural and Applied Sciences (RCNAS) of the University of Santo Tomas, Philippines.

#### References

- Adekiya A.O., Agbede T.M., Aboyeji C.M., Dunsin O., Simeon V.T., 2019, Effects of biochar and poultry manure on soil characteristics and the yield of Radish. *Scientia Horticulturae*, 243, 457–463.
- Belmonte B.A., Aquino T.M.B., Atis R.C.R., Marron A.L.L., Gue I.H.V., Aviso K.B., Tan R.R., 2023, A Rough Set-based Model for Evaluating the Stability of Biochar Based on Accelerated Aging Data. *Chemical Engineering Transactions*, 106, 85–90.

- Brassard P., Godbout S., Raghavan V., 2016, Soil biochar amendment as a climate change mitigation tool: Key parameters and mechanisms involved. *Journal of Environmental Management*, 181, 484–497.
- Brassard P., Godbout S., Levesque V., Palacios J.H., Raghavan V., Ahmed A., Hogue R., Jeanne T., Verma M., 2019, Biochar for soil amendment. *Char and Carbon Materials Derived from Biomass*, 109–146.
- Dai Z., Zhang X., Tang C., Muhammad N., Wu J., Brookes P.C., Xu, J., 2017, Potential role of biochars in decreasing soil acidification - A critical review. *Science of the Total Environment*, 581–582, 601–611.
- Dashti A., Raji M., Harami H.R., Zhou J.L., Asghari M., 2023, Biochar performance evaluation for heavy metals removal from industrial wastewater based on machine learning: Application for environmental protection. *Separation and Purification Technology*, 312, 123399.
- Ding W., Dong X., Ime I.M., Gao B., Ma L.Q., 2014, Pyrolytic temperatures impact lead sorption mechanisms by bagasse biochars. *Chemosphere*, 105, 68–74.
- Gue I.H.V., Lopez N.S., Chiu A., Ubando A.T., Tan R.R., 2021, Rough Set-based Model of Waste Management Systems towards Circular City Economies. *Chemical Engineering Transactions*, 89, 133–138.
- Hai A., Bharath G., Abdul Patah M.F., Daud W.M., Rambabu K., Show P., Banat F., 2023, Machine learning models for the prediction of total yield and specific surface area of biochar derived from agricultural biomass by pyrolysis. *Environmental Technology & Innovation*, 30, 103071.
- Han M., Zhang J., Zhang L., Wang Z., 2023, Effect of biochar addition on crop yield, water and nitrogen use efficiency: A meta-analysis. *Journal of Cleaner Production*, 420, 138425.
- Hvidsten T.R., 2013, A tutorial-based guide to the ROSETTA system: A Rough Set Toolkit for Analysis of Data, <[trhvidsten.com/docs/ROSETTATutorial.pdf](http://trhvidsten.com/docs/ROSETTATutorial.pdf)>, accessed 25.06.2023.
- Jeffery S., Verheijen F.G.A., van der Velde M., Bastos A.C., 2011, A quantitative review of the effects of biochar application to soils on crop productivity using meta-analysis. *Agriculture, Ecosystem & Environment*, 144, 175–187.
- Joiner I.A., 2018, Artificial Intelligence: AI is Nearby. *Emerging Library Technologies*, 1–22.
- Latini A., Bacci G., Teodoro M., Gattia D.M., Bevivino A., Trakal L., 2019, The Impact of Soil-Applied Biochars From Different Vegetal Feedstocks on Durum Wheat Plant Performance and Rhizospheric Bacterial Microbiota in Low Metal-Contaminated Soil. *Front Microbiol*, 10, 2694.
- Läuchli A. Grattan S.R., 2012, Soil pH extremes. *Plant Stress Physiol*, 194–209.
- Li S., Harris S., Anandhi A., Chen G., 2019, Predicting biochar properties and functions based on feedstock and pyrolysis temperature: A review and data syntheses. *Journal of Cleaner Production*, 215, 890–902.
- Liu J., Xu Z., Zhang W., 2023, Unraveling the role of Fe in As (III & V) removal by biochar via machine learning exploration. *Separation and Purification Technology*, 311, 123245.
- Novak J.M., Lima I., Xing B., Gaskin J.W., Steiner C., Das K.C., Ahmedna M., Rehrh D., Watts D.W., Busscher W.J., Schomberg H., 2009, Characterization of Designer Biochar Produced at Different Temperatures and Their Effects on a Loamy Sand. *Annals of Environmental Science*, 3, 195–206.
- Pawlak Z., 1997, Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, 99(1), 48–57.
- Pawlak Z., 2002, Rough sets and intelligent data analysis. *Information Sciences*, 147(1-4), 1–12.
- Purakayastha T., Bera T., Bhaduri D., Sarkar B., Mandal S., Wade P., Kumari S., Biswas S., Menon M., Pathak H., Tsang D., 2019, A review on biochar modulated soil condition improvements and nutrient dynamics concerning crop yields: Pathways to climate change mitigation and global food security. *Chemosphere*, 227, 345–365.
- Singh H., Northup B.K., Rice C.W., Vara Prasad P.V., 2022, Biochar applications influence soil physical and chemical properties, microbial diversity, and crop productivity: a meta-analysis. *Biochar*, 4, 8.
- Tammeorg P., Sijojoki A., Mäkelä P., Stoddard F.L., Alakukku L., Helenius J., 2014, Short-term effects of biochar on soil properties and wheat yield formation with meat bone meal and inorganic fertiliser on a boreal loamy sand. *Agriculture, Ecosystems & Environment*, 191, 108–116.
- Tandzi N.L., Mutengwa S.C., 2020, Factors Affecting Yield of Crops. *Agronomy - Climate Change & Food Security*, 1–16.
- Woolf B.P., 2009, Machine Learning. *Building Intelligent Interactive Tutors*, 221, 223–297.
- Woolf D., Amonette J.E., Street-Perrott F.A., Lehmann J., Joseph S., 2010, Sustainable biochar to mitigate global climate change. *Nature Communications*, 1, 56.
- Ye L., Camps-Arbestain M., Shen Q., Lehmann J., Singh B., Sabir M., 2019, Biochar effects on crop yields with and without fertilizer: A meta-analysis of field studies using separate controls. *Soil Use and Management*, 36(1), 2–18.