Evaluation of Ensemble Machine Learning for Geospatial Prediction of Soil Iron in Croatia

Evaluacija kombinacije strojnog učenja za geoprostorno predviđanje sadržaja željeza u tlu u Hrvatskoj

Radočaj, D., Tuno, N., Mulahusić, A., Jurišić, M.

Poljoprivreda / Agriculture

ISSN: 1848-8080 (Online)
ISSN: 1330-7142 (Print)

https://doi.org/10.18047/poljo.29.2.7

Fakultet agrobiotehničkih znanosti Osijek, Poljoprivredni institut Osijek

Faculty of Agrobiotechnical Sciences Osijek, Agricultural Institute Osijek

ISSN 1330-7142
UDK=631.416.3:591.513-047.44
https://doi.org/10.18047/poljo.29.2.7

EVALUATION OF ENSEMBLE MACHINE LEARNING FOR GEOSPATIAL PREDICTION OF SOIL IRON IN CROATIA

Radočaj, D. [(1)]***, Tuno, N.*** [(2)]***, Mulahusić, A.*** [(2)]***, Jurišić, M.*** [(1)]

Original scientific paper
Izvorni znanstveni rad

SUMMARY

Soil fertility is pivotal for agricultural productivity, and iron (Fe) is a critical micronutrient essential for a successful crop development. This study investigates a potential of ensemble machine-learning methods in geospatial prediction of soil Fe in Croatia. Using a dataset of 686 soil samples, three individual machine-learning methods, including the extreme gradient boosting (XGB), support vector machine (SVM), and Cubist, as well as their ensemble, were evaluated for the soil Fe prediction. The ensemble method outperformed the individual models, exhibiting a higher prediction accuracy expressed by the coefficient of determination (R^2 = 0.578), with a lower root-mean-square error (RMSE = 0.837) and the mean absolute error (MAE = 0.550). The soil clay content emerged as the most influential predictor, followed by the sand content, pH values, and select bioclimatic variables. This study's results demonstrate the effectiveness of ensemble machine learning in an accurate prediction of soil Fe content and contribute to an informed decision-making in sustainable agricultural land-use planning and management. By including the complementary machine-learning methods into an ensemble with the representative environmental covariates, a geospatial prediction aids to a reliable comprehension of soil properties and their spatial variability.

Keywords: soil samples, extreme gradient boosting, support vector machine, cubist, land-use planning

INTRODUCTION

Micronutrients are essential for the insurance of an adequate crop development, and soil fertility is a key factor to the determination of an agricultural production. Since it is a crucial component of the enzymes involved in photosynthesis, respiration, and nitrogen fixation, iron (Fe), one of these micronutrients, has a special relevance, because it is engaged in crucial physiological processes within plants (Mondal and Bose, 2019). Although both excessive and inadequate Fe concentrations in the soil can have a negative impact on a crop growth, the availability and distribution of Fe in the soil can have a substantial influence on how it is absorbed by the crops. A Fe toxicity can cause a root damage, nutritional imbalances, and decreased nutrient absorption efficiency, whereas a Fe shortage can cause chlorosis, stunted growth, and decreased yield (Zaid et al., 2020) its threshold value in plants increases by diverse anthropogenic and natural sources, which results in the inhibition of plant growth and development. This inhibition is due to excess Fe availability, in the soil environment, leading to direct or indirect Fe toxicity. This toxicity, as well as the opposite, i.e., iron deficiency, results in disturbance of basic plant metabolism due to disruption in the rate of uptake and translocation of other essential and beneficial mineral nutrient elements. Since other key nutrients and excess Fe compete, in root rhizosphere(s. Traditionally, a difficult and time-consuming field collection and laboratory analysis of soil samples has been conducted to determine the soil Fe concentration. However, this method only offers a limited amount of information on the geographical distribution of Fe, making it difficult to locate the specific

(1) Dorijan Radočaj, Ph. D. (dradocaj@fazos.hr), Prof. Dr. Mladen Jurišić – Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia, (2) Assoc. Prof. Nedim Tuno, Prof. Dr. Admir Mulahusić – University of Sarajevo, Faculty of Civil Engineering, Patriotske lige 30, 71000 Sarajevo, Bosnia and Herzegovina

regions vulnerable to the Fe toxicity or insufficiency (Radočaj et al., 2021).

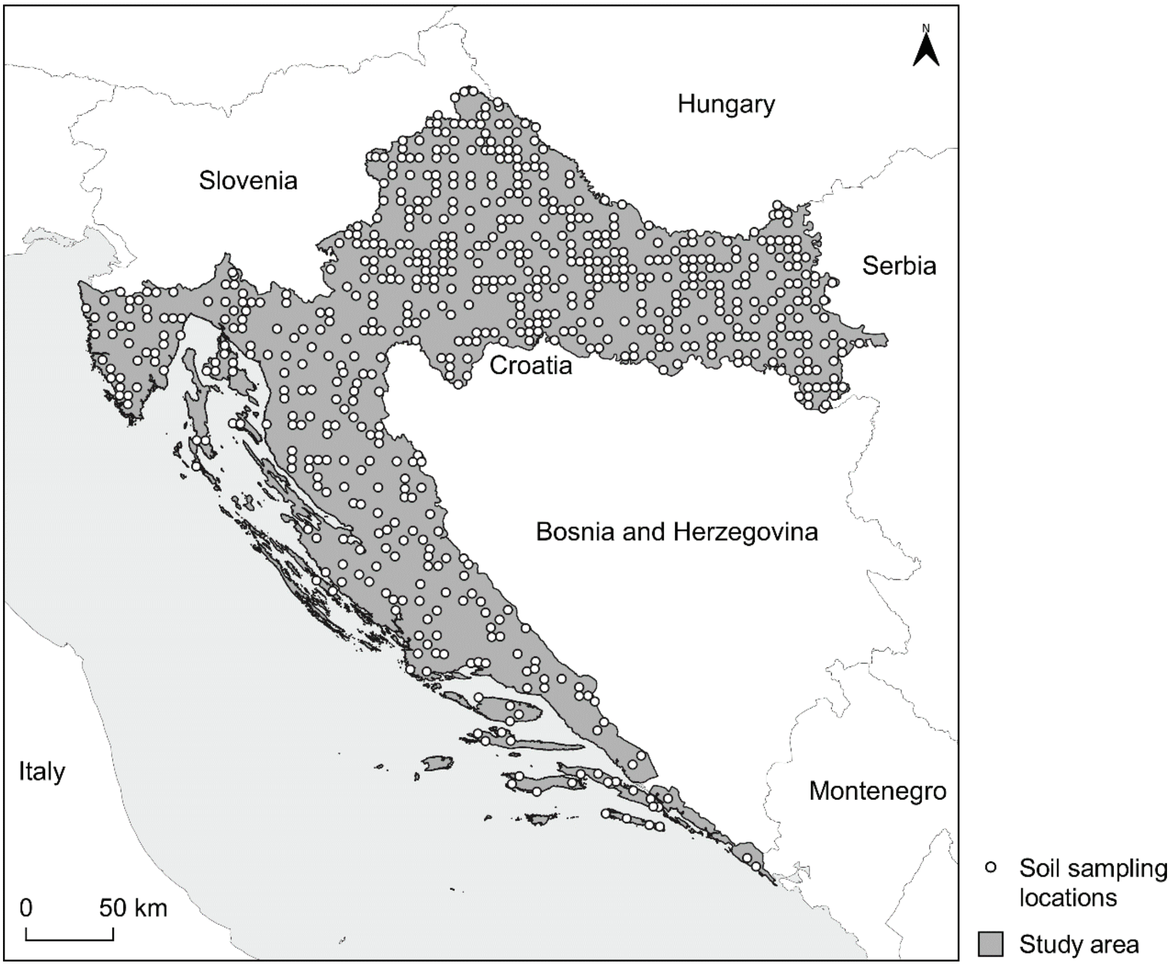
In order to assess the spatial variability of soil parameters, especially the Fe content, across the wide regions, geospatial soil prediction approaches have proven to be effective. A digital soil mapping offers the accurate and high-resolution soil Fe maps, aiding an informed decision-making in agricultural management (Jurišić et al., 2021). This is accomplished by combining the various sources of data, such as the soil samples, remote sensing, and geostatistical modeling. The precision, effectiveness, and dependability of forecasts pertaining to the soil properties have been greatly increased as a result of the application of machine-learning algorithms in geospatial prediction. Due to its capacity to integrate the capabilities of many machine-learning models, the ensemble machine-learning techniques are gaining interest recently (Hengl et al., 2017). This can be explained, as doing so improves prediction performance and lowers the inherent uncertainties associated with the individual models. When compared with the individual models, the ensemble machine-learning approach frequently outperforms them while minimizing the bias, boosting accuracy, and producing the more consistent predictions (Taghizadeh-Mehrjardi et al., 2020) that is, datasets in which all classes are approximately represented equally. Otherwise, the accuracy estimates may be unreliable and classes with only a few values are often misclassified or neglected. This is known as a class imbalance problem in machine learning and datasets that do not meet this criterion are referred to as imbalanced data. Most datasets of soil classes are, therefore, imbalanced data. One of our main objectives is to compare eight resampling strategies that have been developed to counteract the imbalanced data problem. We compared the performance of five of the most common ML algorithms with the resampling approaches. The highest increase in prediction accuracy was achieved with SMOTE (the synthetic minority oversampling technique). Utilizing the advantages of several distinct models while reducing their shortcomings and uncertainties, the ensemble approaches are highly successful. In geospatial soil prediction, where geographical heterogeneity can result in the complicated connections between the soil parameters and environmental variables, the ensemble approaches are essential for the improvement of an overall prediction accuracy and the reduction of overfitting (Radočaj et al., 2022) alternatives to conventional geostatistical methods for soil mapping are becoming increasingly more sophisticated. To provide a complete overview of their performance, this study performed cost–benefit analysis of four soil mapping methods based on five criteria: accuracy, processing time, robustness, scalability and applicability. The evaluated methods were ordinary kriging (OK).

The focus of this study was to evaluate the ensemble machine-learning approach in the geospatial prediction of soil Fe in Croatia. By combining the predictive capabilities of three individual machine-learning methods, a possibility to develop a more accurate and robust soil mapping method was evaluated. Through the integration of diverse algorithms, this research endeavors to provide more reliable soil Fe predictions, ensuring sustainable agricultural land-use planning and management.

MATERIAL AND METHODS

A total of 686 soil samples collected on the agricultural land, forest land, and grasslands across the Republic of Croatia was used in the study (Picture 1). The soil sampling in the field was performed between April 2015 and October 2016 by the Croatian Geological Survey, Croatian Forest Research Institute, and by the former Agency for Agricultural Land of Croatia within the project entitled "*A Change in Soil Carbon Stocks and a Calculation of Total Nitrogen and Soil Organic Carbon Trends and C:N*" (Data Europa, 2021). A total soil Fe content was sampled in a 0–30 cm soil layer and determined according to the HRN ISO 11466 standard. The Shapiro-Wilk test statistic was used to determine a deviation of the input soil Fe data from a normal distribution. The p-value of the Shapiro-Wilk test was used as a measure of evidence against a null hypothesis, which stated that the data follows a normal distribution.

Three individual machine-learning methods with mutually different working prediction principles, including the extreme gradient boosting (XGB), support vector machine (SVM), and Cubist, as well as their ensemble, were evaluated for the sake of geospatial prediction of soil Fe. A machine-learning regression was performed in *R* v4.2.2.



Picture1. Study area and soil sampling locations.
Slika 1. Područje istraživanja i lokacije uzorkovanja tla.

The XGB is a widely used gradient-boosting algorithm that effectively computes the large datasets, capturing complex non-linear relationships and effectively handling the missing data (Ma et al., 2021). Its popularity stems from its ability to minimize the prediction errors while iteratively adding the weak learners and optimizing the objective functions. It also demonstrated a high prediction accuracy of soil parameters due to its capability to model the intricate soil-environment relationships effectively. The SVM belongs to the supervised machine learning, which calculates an optimal hyperplane that maximizes the margin between the different predictions (Heung et al., 2016)where a fitted model may then be used for prediction purposes on new data. Despite the growing number of machine-learning algorithms that have been developed, relatively few studies have provided a comparison of an array of different learners — typically, model comparison studies have been restricted to a comparison of only a few models. This study evaluates and compares a suite of 10 machine-learners as classification algorithms for the prediction of soil taxonomic units in the Lower Fraser Valley, British Columbia, Canada. A variety of machine-learners CART, CART with bagging, Random Forest, k-nearest neighbor, nearest shrunken centroid, artificial neural network, multinomial logistic regression, logistic model trees, and support vector machine. It has been effectively applied in soil prediction studies, particularly when dealing with the limited training data or with the non-linear relationships between the soil properties and environmental covariates. Cubist is a rule-based algorithm that generates the sets of the if-then rules to predict the continuous or categorical outcomes based on a combination of decision trees and linear models (Khaledian and Miller, 2020). Its advantages are interpretability and ability to reveal the complex interactions between the predictor variables, making it useful for calculating the soil-environment relationships. The ensemble machine-learning prediction was used by combining the XGB, the SVM, and Cubist based on a generalized linear model to produce a more accurate and robust prediction than the individual methods. The automated hyperparameter tuning was performed for the XGB, SVM, and Cubist tuning parameters listed in Table 1. to determine the best-performing parameters for the input soil Fe dataset.

Table 1. The tuned hyperparameters of the evaluated individual machine-learning methods.

Tablica 1. Podešeni hiperparametri evaluiranih pojedinačnih metoda strojnoga **učenja**.

Method / *Metoda*	Label / *Oznaka*	Description / *Opis*
XGB	nrounds	the number of boosting rounds or iterations, with each round adding a new decision tree
	max_depth	the maximum depth of each individual decision tree
	eta	learning rate which controls the step size at each iteration during the optimization process
	gamma	a regularization parameter which controls the minimum loss reduction required to make a further partition on a leaf node
	colsample_bytree	the fraction of features to be randomly sampled for building each decision tree
	subsample	the fraction of samples to be randomly sampled for each boosting round
SVM	sigma	the distance of a single training iteration reach
	C	the control of a trade-off between maximizing the margin and minimizing the classification error
Cubist	committees	the number of committees (or base models)
	neighbors	the number of nearest neighbors to consider when constructing the linear models

The accuracy assessment of the predicted soil Fe was performed applying the 10-fold cross-validation along with the three performance metrics, including the coefficient of determination (R^2), the root-mean-square error (RMSE), and the mean absolute error (MAE). The dataset was divided into ten subsets, and each subset was used as a validation set in turns, while the remaining nine subsets were utilized for training the predictive model. The R^2 provided insights into how accurately the predictions explained the variance in the observed soil Fe (Jurišić et al., 2020), while the RMSE and MAE metrics were calculated to determine the absolute prediction accuracy (Khaledian and Miller, 2020). The higher R^2 and the smaller RMSE and MAE values indicated a higher prediction accuracy of the evaluated machine-learning methods. The optimal hyperparameters for each method were selected based on the smallest RMSE during the tuning process.

RESULTS AND DISCUSSION

The input soil Fe dataset had a median value of 3.440 mg kg^{-1}, with the coefficient of variation amounting to 0.354, indicating a moderate variability. The minimum value was 0.000 mg kg^{-1} in several samples in the dataset, while the maximum value was 7.175 mg kg^{-1}. The test statistic of the Shapiro-Wilk normality test resulted in 0.9366, which indicated some deviation from a normal distribution. Moreover, the p-value of 2.519e^{-16} confirmed that the input soil Fe dataset does not follow a normal distribution, as the p-value was significantly lesser than the selected significance level of 0.05, rejecting the null hypothesis. A relationship between the soil Fe and the major soil properties is presented in Table 2, based on the results of the Kruskal-Wallis test for a soil type and linear regression for the soil pH (H_2O), clay, silt, and sand content (Data Europa, 2021). Their results indicate that there were statistically significant differences of soil Fe values per soil type class, as well as that the soil Fe was strongly affected by the soil texture components, having a notable relationship with soil pH and having a weaker significance level.

Table 2. A relationship between the soil Fe and major soil properties based on the Kruskal-Wallis test (soil type) and linear regression (soil pH, clay, silt, and sand content)

Tablica 2. Odnos između sadržaja Fe u tlu s važnijim svojstvima tla temeljenim na Kruskal-Wallisovu testu (vrsta tla) i linearnoj regresiji (pH tla, udio gline, praha i pijeska)

Soil property / *Svojstvo tla*	Test statistics / *Statistička vrijednost testa*	Probability (*p*) / *Vjerojatnost (p)*	Significance level / *Razina signifikantnosti*
soil type	187.13	< 0.001	***
pH	2.31	0.021	*
clay	13.04	< 0.001	***
silt	–3.79	< 0.001	***
sand	–7.43	< 0.001	***

Statistical significance levels: '***' 0.001, '**' 0.01, '*' 0.05.

The tuned XGB model had a relatively shallow tree structure (max_depth = 1), which reduced the overfitting (Fig. 2). The learning rate (eta) was moderate (0.3), and no regularization was applied (gamma = 0). The feature sampling was set to 80% (colsample_bytree = 0.8), while all samples were used for training (subsample = 1). The sigma value in the SVM did not change from the initial value of 0.017, while the C value of 1 determined the decision boundary as smooth, allowing some mispredictions but reducing a possibility of overfitting (Fig. 2). The Cubist created 20 separate regression trees (models) and utilized their combined predictions for the production of the final one (committees = 20), using the zero neighbors for the creation of linear models, without incorporating the information from the nearby data points. In comparison to the previous studies, the hyperparameter tuning for all evaluated methods focused on the reduction of overfitting and on the optimization of generalization performance, its disadvantage being a slightly lower prediction accuracy (Beucher et al., 2022; Kavzoglu and Teke, 2022).

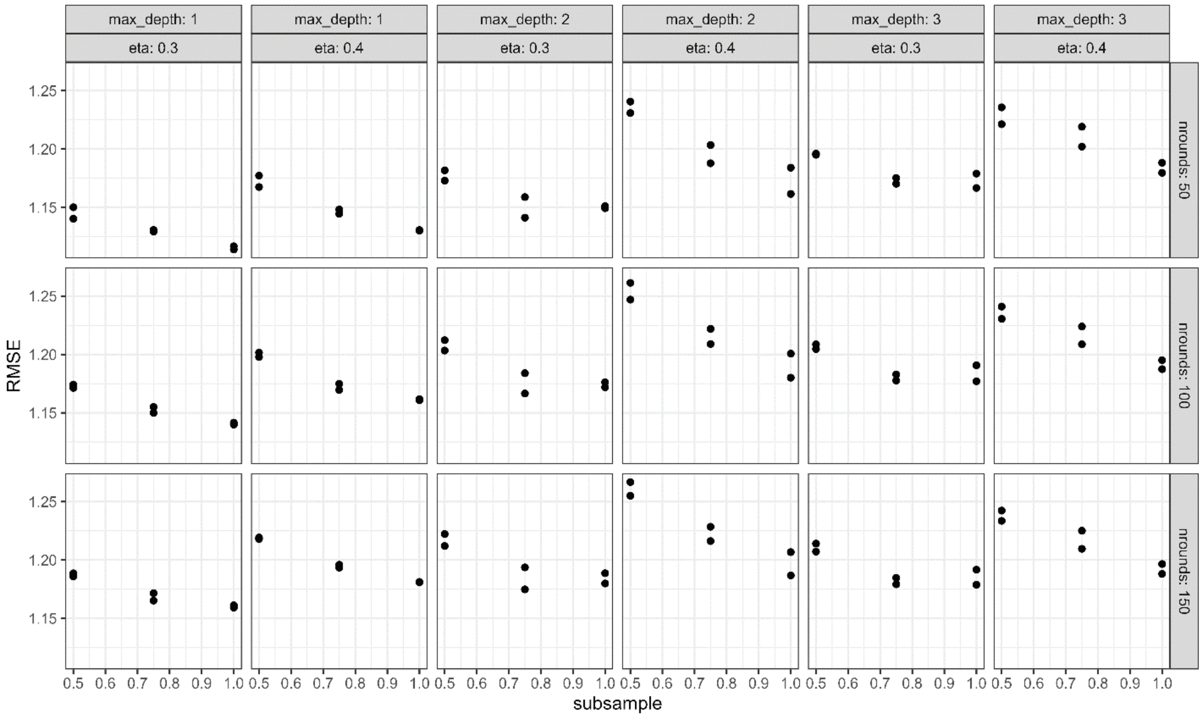


Figure 1. The combination of evaluated tuning hyperparameters for the XGB.

Grafikon 1. Kombinacija evaluiranih hiperparametara podešavanja za XGB.

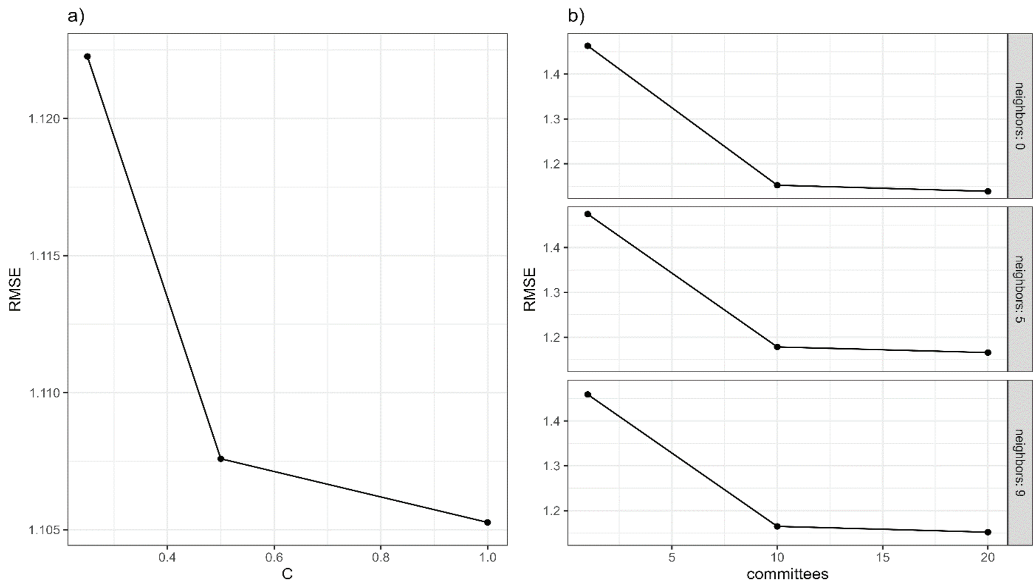


Figure 2. The combination of evaluated tuning hyperparameters for: a) the SVM and b) the Cubist.

Grafikon 2. Kombinacija evaluiranih hiperparametara podešavanja za: a) SVM i b) Cubist.

Among the four evaluated machine-learning models, the ensemble model stood out as the most effective and accurate predictor (Table 3, Fig. 4). The ensemble methods combined the predictions from several individual models, reducing the risk of overfitting and compensating for weaknesses in any single model. By averaging, or weighting, the predictions of diverse models, the ensemble model achieved a more robust and accurate prediction, as demonstrated by its higher R^2 value if compared with the other models. The lower

RMSE indicates that the ensemble model's predicted values are, on average, closer to the actual ones, while the smaller MAE suggests fewer significant deviations from the ground truth. These results are crucial in the digital soil mapping, where accurate predictions lead to a more accurate geospatial representation and informed decision-making. The prediction accuracy of the XGB and the SVM suggested their ability to capture a reasonable portion of the variance in the target variable. The SVM, being a powerful classifier, often performed well in the high-dimensional spaces based on a previous research, while the XGB, a gradient-boosting algorithm, excelled in handling the complex relationships between the features (Zeng et al., 2023). Despite their relative advantages, both models exhibited the higher RMSE and MAE values when compared with the ensemble model, indicating that they were not as accurate in predicting the concentration of soil Fe. The Cubist, a rule-based model, provided interpretability by generating the rules for decision-making, but it made more substantial prediction errors, as reflected in its higher RMSE and MAE values. However, the accuracy of machine-learning methods can vary across different datasets, soil types, and geographic regions (Khanal et al., 2018) high quality soil and crop yield maps. Integration of remotely sensed data and machine learning algorithms offers cost-and time-effective approach for spatial prediction of soil properties and crop yield compared to conventional approaches. The objectives of this study were to: (i. Hence, rigorous testing on diverse datasets can ensure that the selected method remains reliable and accurate across various study areas.

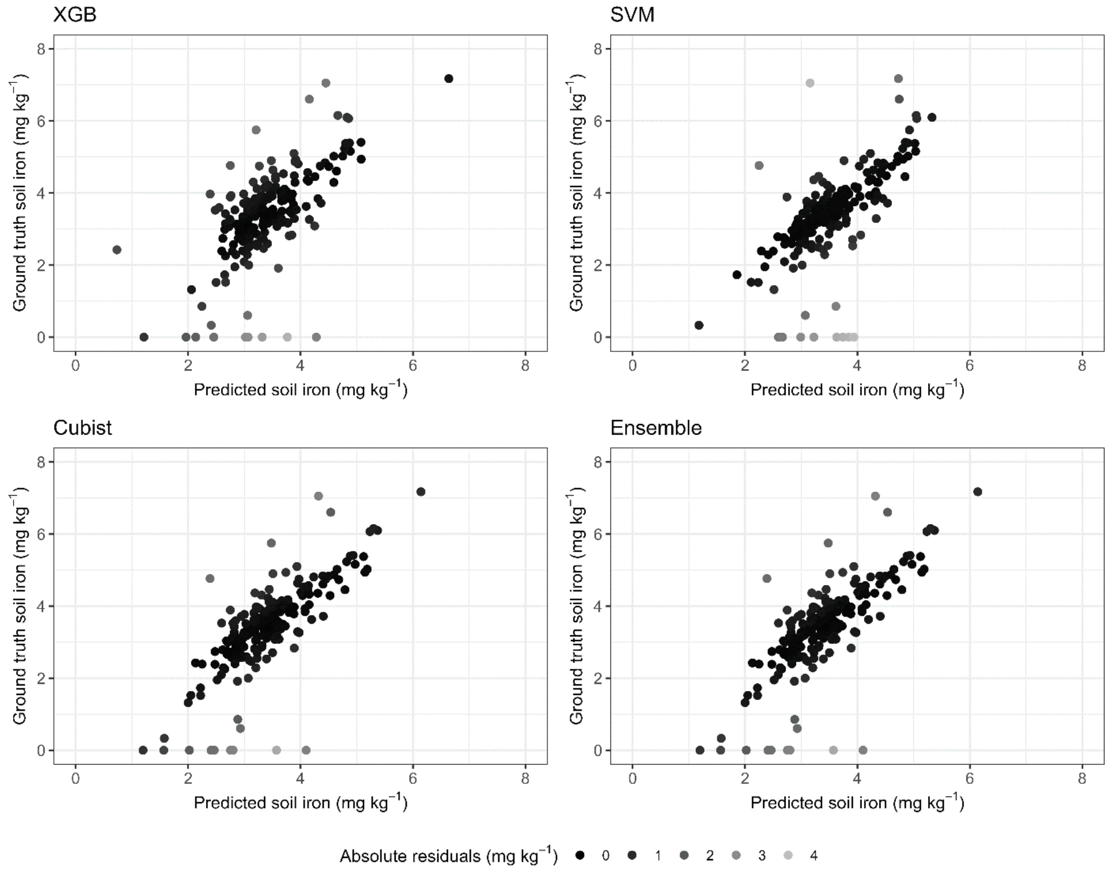


Figure 3. The scatterplots of machine-learning predictions according to the ground truth data.
Grafikon 3. Dijagrami raspršenosti predviđanja strojnoga učenja prema podatcima uzorkovanja tla.

Table 3. Accuracy assessment of the evaluated machine-learning methods for the prediction of soil Fe.
Tablica 3. Procjena točnosti evaluiranih metoda strojnoga učenja za predviđanje Fe u tlu.

Method *Metoda*	R^2	RMSE	MAE
XGB	0.425	0.953	0.661
SVM	0.429	0.962	0.557
Cubist	0.366	0.997	0.656
Ensemble	0.578	0.837	0.550

The variable importance metrics obtained from the machine-learning predictions for the input environmental covariates are presented on Figure 4. The variable importance metrics provide the insights into the significance of each variable in the prediction of target outcome, and the results can be instrumental in making the data-driven decisions and optimizing the processes in digital soil mapping. The soil clay content emerged as a key predictor, with the highest importance value among all machine learning methods. Kaur et al (2018) noted that the comprehension of this relationship can be crucial for agricultural practices, as soil composition influences crop growth, water retention, and nutrient availability. The soil sand content and pH values were also among the most impactful environmental covariates resulting in an ensemble machine-learning prediction, followed by the median annual surface reflectance in the 2016 shortwave infrared band. Bioclimatic variables bio04 (temperature seasonality) and bio12 (annual precipitation) also exhibited relatively high importance values across the models, which play a vital role in modeling the ecological habitats and influencing the species' distributions in general (Cervellini et al., 2021). Despite the noted relationship between the soil Fe and the soil type, pH and the soil texture components based on the Kruskal-Wallis test, and linear regression, the variable importance calculation ensured an exact quantification of their impact on the prediction accuracy, which could not be distinguished based on their respective test statistics.

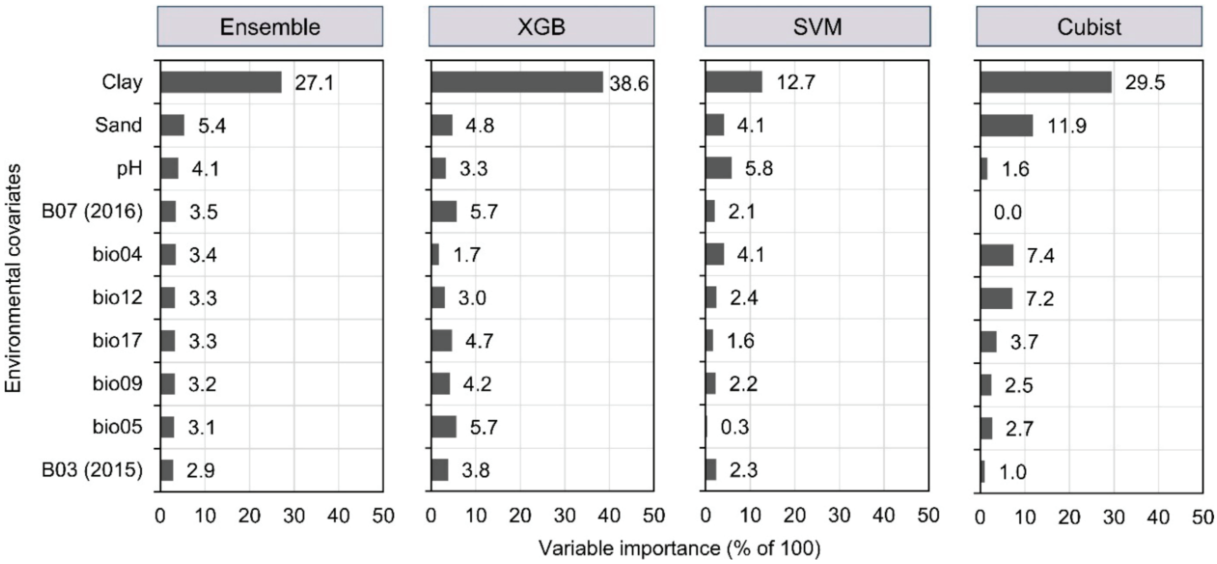


Figure 4. The top variable importance values per machine learning method.
Grafikon 4. Najviše vrijednosti utjecaja varijabla prema metodi strojnoga učenja.

CONCLUSION

In comparison with the individual methods, a superior prediction accuracy of ensemble machine learning provided the upgrade of the conventional approach for soil Fe prediction and indicated similar advantages in a digital soil mapping in general. The ensemble machine learning showcased a superior prediction accuracy, as evidenced by its lower RMSE value amounting to 0.837 and its MAE amounting to 0.550. These smaller error values indicate that the predicted soil values were closer to the actually observed ones, signifying a higher precision and accuracy in the ensemble machine-learning predictions. This aspect is particularly important in soil-related applications, where accurate predictions can lead to better decision-making in various fields, such as agriculture, environmental management, and land-use planning. By combining the strengths of the XGB, SVM, and Cubist, this research contributes to the development of a powerful ensemble model that can enhance soil property predictions and aid in making the informed decisions for sustainable land use and environmental management. The capacity of ensemble machine learning to provide precise and reliable predictions can, therefore, significantly contribute to the achievement of geospatial zoning pertaining to the areas having an excessive and insufficient Fe soil content. While this study provides a strong evidence of the ensemble machine-learning superiority, further research is warranted to explore the additional aspects of ensemble approach, including various soil properties. Investigating different ensemble techniques, such as bagging and boosting, and exploring the impact of varying ensemble sizes on a predictive accuracy could provide additional insights. Identifying the specific types of base models and feature engineering strategies that complement each other within the ensemble model could further enhance its performance. However, it is essential to consider the other factors, such as computational complexity and interpretability, when selecting the most suitable model for specific soil-related applications. Further research and validation of independent datasets may also be necessary to confirm the robustness and generalizability of ensemble machine-learning predictions.

REFERENCES

1. Beucher, A., Rasmussen, C. B., Moeslund, T. B., & Greve, M. H. (2022). Interpretation of Convolutional Neural Networks for Acid Sulfate Soil Classification. *Frontiers in Environmental Science*, *9*. https://www.frontiersin.org/articles/10.3389/fenvs.2021.809995
2. Cervellini, M., Di Musciano, M., Zannini, P., Fattorini, S., Jiménez-Alfaro, B., Agrillo, E., Attorre, F., Angelini, P., Beierkuhnlein, C., Casella, L., Field, R., Fischer, J.-C., Genovesi, P., Hoffmann, S., Irl, S. D. H., Nascimbene, J., Rocchini, D., Steinbauer, M., Vetaas, O. R., & Chiarucci, A. (2021). Diversity of European habitat types is correlated with geography more than climate and human pressure. *Ecology and Evolution*, *11*(24), 18111–18124. https://doi.org/10.1002/ece3.8409
3. Data Europa. (2021). *Changes in soil carbon stocks and calculation of trends in total nitrogen and organic carbon in soil and C: N ratios.* https://data.europa.eu/data/datasets/zaliha-ugljika-u-tlu-izracun-trendova-ukupnog-dusika-i-organskog-ugljika-te-odnosa-c-n?locale=en
4. Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748
5. Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, *265*, 62–77. https://doi.org/10.1016/j.geoderma.2015.11.014
6. Jurišić, M., Radočaj, D., Krčmar, S., Plaščak, I., & Gašparović, M. (2020). Geostatistical Analysis of Soil C/N Deficiency and Its Effect on Agricultural Land Management of Major Crops in Eastern Croatia. *Agronomy*, *10*(12), Article 12. https://doi.org/10.3390/agronomy10121996
7. Jurišić, M., Radočaj, D., Plaščak, I., & Rapčan, I. (2021). A Comparison of Precise Fertilization Prescription Rates to a Conventional Approach Based on the Open Source Gis Software. *Poljoprivreda*, *27*(1), 52–59. https://doi.org/10.18047/poljo.27.1.7
8. Kaur, S., Kaur, R., & Chauhan, B. S. (2018). Understanding crop-weed-fertilizer-water interactions and their implications for weed management in agricultural systems. *Crop Protection*, *103*, 65–72. https://doi.org/10.1016/j.cropro.2017.09.011
9. Kavzoglu, T., & Teke, A. (2022). Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (XGBoost). *Bulletin of Engineering Geology and the Environment*, *81*(5), 201. https://doi.org/10.1007/s10064-022-02708-w
10. Khaledian, Y., & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, *81*, 401–418. https://doi.org/10.1016/j.apm.2019.12.016
11. Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., & Shearer, S. (2018). Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture*, *153*, 213–225. https://doi.org/10.1016/j.compag.2018.07.016
12. Ma, G., Ding, J., Han, L., Zhang, Z., & Ran, S. (2021). Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. *Regional Sustainability*, *2*(2), 177–188. https://doi.org/10.1016/j.regsus.2021.06.001
13. Mondal, S., & Bose, B. (2019). Impact of micronutrient seed priming on germination, growth, development, nutritional status and yield aspects of plants. *Journal of Plant Nutrition*, *42*(19), 2577–2599. https://doi.org/10.1080/01904167.2019.1655032
14. Radočaj, D., Jug, I., Vukadinović, V., Jurišić, M., & Gašparović, M. (2021). The Effect of Soil Sampling Density and Spatial Autocorrelation on Interpolation Accuracy of Chemical Soil Properties in Arable Cropland. *Agronomy*, *11*(12), Article 12. https://doi.org/10.3390/agronomy11122430
15. Radočaj, D., Jurišić, M., Antonić, O., Šiljeg, A., Cukrov, N., Rapčan, I., Plaščak, I., & Gašparović, M. (2022). A Multiscale Cost–Benefit Analysis of Digital Soil Mapping Methods for Sustainable Land Management. *Sustainability*, *14*(19), Article 19. https://doi.org/10.3390/su141912170
16. Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari, K., Behrens, T., Jamshidi, M., Davatgar, N., Toomanian, N., & Scholten, T. (2020). Synthetic resampling strategies and machine learning for digital soil mapping in Iran. *European Journal of Soil Science*, *71*(3), 352–368. https://doi.org/10.1111/ejss.12893
17. Zaid, A., Ahmad, B., Jaleel, H., Wani, S. H., & Hasanuzzaman, M. (2020). A Critical Review on Iron Toxicity and Tolerance in Plants: Role of Exogenous Phytoprotectants. In T. Aftab & K. R. Hakeem (Eds.), *Plant Micronutrients: Deficiency and Toxicity Management* (pp. 83–99). Springer International Publishing. https://doi.org/10.1007/978-3-030-49856-6_4
18. Zeng, T., Wu, L., Peduto, D., Glade, T., Hayakawa, Y. S., & Yin, K. (2023). Ensemble learning framework for landslide susceptibility mapping: Different basic classifier and ensemble strategy. *Geoscience Frontiers*, *14*(6), 101645. https://doi.org/10.1016/j.gsf.2023.101645

EVALUACIJA KOMBINACIJE STROJNOGA UČENJA ZA GEOPROSTORNO PREDVIĐANJE SADRŽAJA ŽELJEZA U TLU U HRVATSKOJ

SAŽETAK

Plodnost tla ključna je za produktivnost poljoprivredne proizvodnje, a željezo (Fe) je ključni mikroelement nužan za uspješan razvoj usjeva. Ova studija istražuje potencijal kombinacije metoda strojnoga učenja u geoprostornome predviđanju Fe u tlu u Hrvatskoj. Korištenjem skupa podataka od 686 uzoraka tla, tri pojedinačne metode strojnoga učenja, uključujući **extreme gradient boosting** ***(XGB),*** **support vector machine** ***(SVM) i*** **Cubist*****, kao i njihova kombinacija, evaluirani su za predviđanje Fe u tlu. Metoda kombinacije nadmašila je pojedinačne modele, pokazujući veću točnost predviđanja izraženu koeficijentom determinacije (R^2 = 0,578), s nižom srednjom kvadratnom pogreškom (RMSE = 0,837) i srednjom apsolutnom pogreškom (MAE = 0,550). Sadržaj gline u tlu pokazao se najutjecajnijim prediktorom, a slijede ga sadržaj pijeska, pH vrijednosti i odabrane bioklimatske varijable. Rezultati ove studije pokazuju učinkovitost kombinacije strojnoga učenja u točnome predviđanju sadržaja Fe u tlu, čime doprinose informiranomu donošenju odluka u planiranju i upravljanju održivim poljoprivrednim zemljištem. Uključivanjem komplementarnih metoda strojnoga učenja u skup s reprezentativnim ekološkim kovarijatama, geoprostorno predviđanje pomaže pri pouzdanomu razumijevanju svojstava tla i njihove prostorne varijabilnosti.***

Ključne riječi: uzorci tla, **extreme gradient boosting*****,*** **support vector machine*****,*** **Cubist*****, planiranje korištenja zemljišta***

(Received on July 24, 2023; accepted on November 23, 2023 – *Primljeno 24. srpnja 2023.; prihvaćeno 23. studenoga 2023*.)