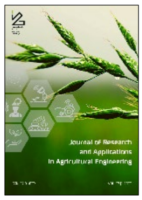
Article citation info:
Pieczarka K. S., Pentoś K., Cieniawska B., Mbah J. 2025. The potential of machine learning methods for the prediction of soil compaction. *Journal of Research and Applications in Agricultural Engineering* 70 (1): 5-11. https://doi.org/10.53502/jraae-203675

Łukasiewicz
Poznań Institute of Technology

Journal of Research and Applications in Agricultural Engineering

Journal website: https://tech-rol.eu/



The potential of machine learning methods for the prediction of soil compaction

Krzysztof S. Pieczarka[a] *
Katarzyna Pentoś[a]
Beata Cieniawska[a]
Jasper Mbah[a]

[a] Wrocław University of Environmental and Life Sciences, Wrocław, Poland

Article info
Received: 6 December 2024
Accepted: 12 March 2025
Published: 4 April 2025

Keywords
soil electrical conductivity
soil compaction
neural network

The practice of precision farming is contingent upon a comprehensive understanding of the spatial variability of a multitude of physical and chemical soil parameters. The acquisition of knowledge regarding soil parameters necessitates the undertaking of soil sampling and subsequent analysis, a process that is inherently labour-intensive and time-consuming. Consequently, precision farming employs the identification of homogeneous field regions through the utilisation of scanning techniques, with the objective of ascertaining soil electrical characteristics, including electrical conductivity and magnetic susceptibility. The objective of this study was to attempt to predict soil compaction based on selected electrical parameters. In order to predict compaction, machine learning methods, namely decision tree and support vector regression were employed. The highest R-value of 0.87 was obtained for the decision tree model and soil layer 0.1-0.2 m for the training set. For the test set, the highest R-value of 0.85 was obtained for soil layer 0.1-0.2 m and the support vector regression model, which also had the lowest MAPE error value of 11.31%. The prediction of soil compaction using electrical soil parameters based on machine learning methods represents a promising avenue of research.

DOI: 10.53502/jraae-203675

1. Introduction

Management zones are an effective approach to managing agricultural field variability and electrical conductivity (ECa), defined as a material's ability to conduct electric current is better known for representing fields based on nutrient needs [12, 10] and offers a cheaper approach for detecting soil variability patterns. The most useful tool in determining field variability that influences crop growth is Electromagnetic induction (EM). Its usefulness in precision agriculture is due to its depth of inquiry, longevity, and influence on crop growth parameters [1].

Soil ECa are directly related to soil texture, moisture, salinity, and nutrient concentration, and are inversely proportional to soil depth [7, 2]. Soil ECa, expressed in milli siemens is impacted by a combination of physicochemical properties such as soil texture and moisture contents, bulk density, soil pore size and distribution, and soil temperature [2]. Soil conductivity is either measured by the physical direct contact of a minimum of four electrodes with soil or by electromagnetic induction (EMI) that uses a transmitter coil to induce a field into the soil and a receiver coil to measure the response [4, 8].

Soil compaction refers to the reduction in soil volume caused by external influences that reduce soil

* Corresponding author: krzysztof.pieczarka@upwr.edu.pl

productivity and environmental quality [4, 5, 6]. Soil resistivity/compaction is a critical agricultural issue and is responsible for variable growth in agricultural areas [4]. Soil strength is the primary determinant of soil compaction and is influenced by various soil variables: bulk density, moisture, and soil texture [3, 11]. Soil compaction can be classified into topsoil and subsoil [11] and is found in all types of soils. Soil compaction occurs naturally, compression from machinery or stock trampling [5], and foot and vehicle traffic [11]. Soil compaction results in inadequate root systems and low yields affects trees and shrubs, reduces soil fertility, increases soil erosion, reduces soil porosity, and increases soil density and penetration resistance [12].

There is a justifiable reason to propose an alternative method of determining soil compaction that is cheaper and easier and it may be possible to use scanning methods as an indirect method to estimate soil compaction. Soil compaction, electrical conductivity, and magnetic susceptibility correlate with soil texture, bulk density, moisture content, and organic matter content. The use of electrical soil parameters offers the advantage of quickly scanning large areas covering several hectares [13]. These techniques allow for fast, economical, and accurate assessment of soil texture differences, soil moisture levels, and soil salinity [14]. Nevertheless, there has been limited research focused on applying scanning methods to estimate soil compaction [15].

The aim of this study was to demonstrate that the appropriate application of machine learning methods can produce soil compaction prediction models based on soil electrical parameters and soil texture with sufficient accuracy to be useful in agricultural practice.

2. Materials and methods

2.1. Experimental data acquisition

In the spring of 2022, field tests were conducted on a plot of land with the registration number 021905_2.0007.22, spanning an area of 1.1 hectares and situated within the Świdnica district, Lower Silesia region, Poland (GNSS 50° 54' 55.027" N 16° 35' 35.994" E). Following the harvest of maize for grain during the 2020/2021 growing season, the crop residues were shredded using a mulcher and then cultivated with a disc harrow to a depth of 0.1 m. The maize was a forecrop for spring barley. Soil was classified as a silt loam, following the Polish Soil Classification System. The soil was not intentionally compacted and the variation in compaction was solely due to altitude and variable soil texture.

Parallel lines 10 m apart were drawn along the longest edge of the field. Measurements of electrical parameters and soil compaction were taken along these lines. The delineated lines were traversed at 10-metre intervals, resulting in 126 compaction measurement points throughout the plot. Soil compaction was assessed for soil layers ranging from 0 to 0.5 m using a cone penetrometer (Eijkelkamp Soil &Water, Giesbeek, The Netherlands) equipped with GPS. Field measurements were conducted using a cone with a base of 0.0001 m2 and an angle of 600, with a penetration speed of 0.03 m·s-1. The two scanners were utilised for measurements, namely Geonics EM38 scanner and Veris 3100 scanner.

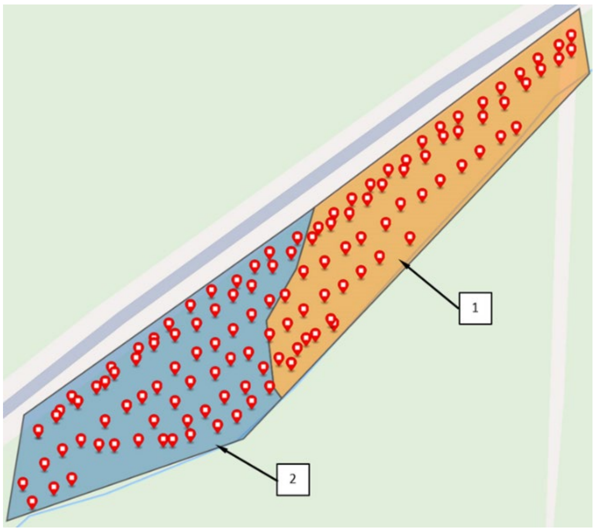


Fig. 1. The cadastral boundary of the surveyed field with the two-zone division and the soil compaction measurement points

The following parameters were collected: electrical conductivity measured using Geonics EM38 (ECa EM38), magnetic susceptibility measured using Geonics EM38 (MS EM38), electrical conductivity measured using the Veris 3100 (ECa Veris). Following the completion of the compaction and scanning measurements, the nearest corresponding point from the Geonics EM38 and Veris 3100 scanner was matched to each compaction measurement using GNSS coordinates, employing the least squares method. The division of the surveyed field into management zones numbered from 1 to 2 presented in figure 1 is based on soil electrical conductivity results from the Geonics EM38 scanner. The management zones were identified using the proprietary algorithm developed by the company carrying out the survey.

In accordance with the results from the Veris 3100 scanner, a soil coefficient (SCoef) was calculated using the formula specified in equation 1. This value was then incorporated into the modelling of soil compactness.

$$SCoef = \frac{Clay\%}{10} + \frac{Silt\%}{100} + \frac{Sand\%}{1000} \qquad (1)$$

2.1. Machine learning models

In this research the two machine learning algorithms were used, namely Decision Tree (DT) and Support Vector Regression (SVR). For the development of models the skelarn.tree and skelarn.skelarn.svm packages in Python 3.12 were employed. DT is a representation of a set of rules for making decisions based on features or attributes of a given dataset. The construction of a DT entails the incremental partitioning of the training dataset, with the objective of reducing the variance of the explanatory variable within each subset. DTs offer a straightforward means of comprehension and provide a graphical representation of their internal logic. However, decision trees are not without limitations, including the potential for overfitting and sensitivity to alterations in the data. These issues can be addressed through techniques such as pruning and setting the maximum depth of the tree. The generation of decision trees necessitates the adjustment of a number of hyper-parameters, including: splitter – strategy for choosing the split at each node, the maximum depth of the tree, the minimum number of samples required to be in a leaf node, the maximum number of leaf nodes in the tree, and the function used to measure the quality of a split. SVR is a powerful algorithm used for predicting continuous values in machine learning tasks. Its primary objective is to determine a function that closely estimates the target variable while maintaining a maximum margin between the predicted and actual values. The "support vectors" are data points lying closest to the regression line, defining the margin and significantly influencing the final model. SVR's effectiveness relies on fine-tuning hyperparameters such as the regularization parameter C (which balances error tolerance and model complexity) and kernel parameters (used for transforming data into higher-dimensional spaces for nonlinear problems). The ε-insensitive loss function used in SVR permits small deviations within a defined margin (ε), where such deviations are not penalized. Errors outside this range are penalized more stringently to ensure accurate predictions. In this research the RBF kernel with various γ values was employed. All hyper-parameters of DT and SVR model development were fine-tuned using a grid search method (separately for each soil depth). The ranges of hyper-parameters were as follows:

1. Decision Tree (DT): splitter [best, random], the maximum depth of the tree [None, 3, 5, 7, 9], the minimum number of samples required to be in a leaf node [1, 2, 4, 6], the maximum number of leaf nodes in the tree [None, 10, 20, 30], and the function used to measure the quality of a split [mse, friedman_mse, mae].
2. Support Vector Regression (SVR): C – from 100 to 1000, γ = [0.001, 0.01, 0.1, 0.2, 0.3].

Independent models were constructed for the measurement of soil compaction in the following soil layers: 0.1–0.2 m, 0.2–0.3 mm and 0.3–0.4 m. The following parameters were employed as independent variables in the models: electrical conductivity measured using Geonics EM38, magnetic susceptibility measured using Geonics EM38, electrical conductivity measured using the Veris 3100, soil coefficient. The 126 data sets acquired from the field experiment were randomly split 80:20 into a training set and a test set. The data were normalized.

The three metrics described below were used to evaluate the performance of the models.

The mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_t - Y_p}{Y_t}\right| \qquad (2)$$

The correlation coefficient (R) between the target and predicted values:

$$R = \frac{\sum(Y_t-\bar{Y}_t)(Y_p-\bar{Y}_p)}{\sqrt{\sum(Y_t-\bar{Y}_t)^2\sum\left(Y_p-\bar{Y}_p\right)^2}} \qquad (3)$$

The root-mean-square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_p - Y_t)^2} \quad (4)$$

where Y_t is the absolute target value, $\overline{Y}_t$ is the mean target value Y_p is the absolute predicted value, $\overline{Y}_p$ is the mean predicted value, and n is the amount of vectors in a data set.

3. Results and their analysis

Figure 2 illustrates the electrical parameters of the soil, as measured with the Geonics EM38 scanner, alongside the soil compaction, as measured with a cone penetrometer, divided into soil layers of 0.1–0.2 m, 0.2–0.3 m, and 0.3–0.4 m.

Upon examination of the soil electrical conductivity map presented in Figure 2, it becomes evident that two distinct areas can be identified within the surveyed field, thereby corroborating the delineation of the two management zones illustrated in Figure 1. Upon analysis of the soil compaction values of the individual layers, it becomes evident that there is minimal discrepancy between them. Notably, these variations occur in specific regions of the field, exhibiting a discernible pattern when comparing the soil compaction maps with the electrical conductivity map. The analogous variation in soil compaction across the layers and electrical conductivity may serve as a foundation for the assertion that selected electrical parameters of the soil can be utilized to predict its corresponding mechanical parameters.

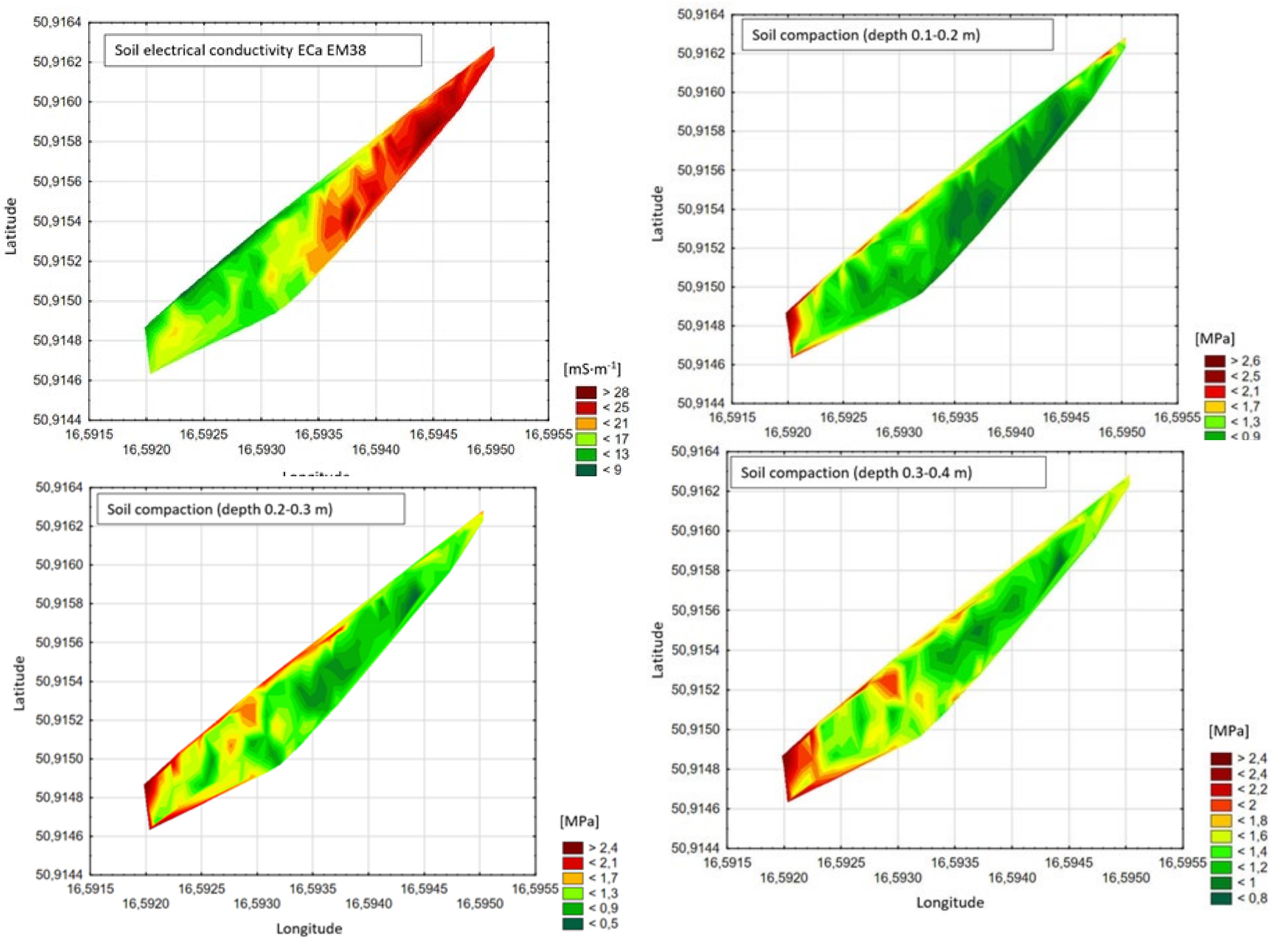


Fig. 2. The cadastral boundary of the surveyed field with the two-zone division and the soil compaction measurement points

Table 1. Parameters of models

Dependent variable	Model	Training data			Test data		
		R	RMSE	MAPE	R	RMSE	MAPE
Soil compaction (depth 0.1-0.2 m)	DT	0.87	0.18	13.31	0.70	0.21	16.91
Soil compaction (depth 0.1-0.2 m)	SVR	0.86	0.19	14.24	0.85	0.16	11.31
Soil compaction (depth 0.2-0.3 m)	DT	0.80	0.22	13.28	0.77	0.22	11.62
Soil compaction (depth 0.2-0.3 m)	SVR	0.83	0.20	11.68	0.76	0.24	13.78
Soil compaction (depth 0.3-0.4 m)	DT	0.73	0.20	9.62	0.73	0.21	12.01
Soil compaction (depth 0.3-0.4 m)	SVR	0.77	0.119	8.26	0.72	0.21	11.75

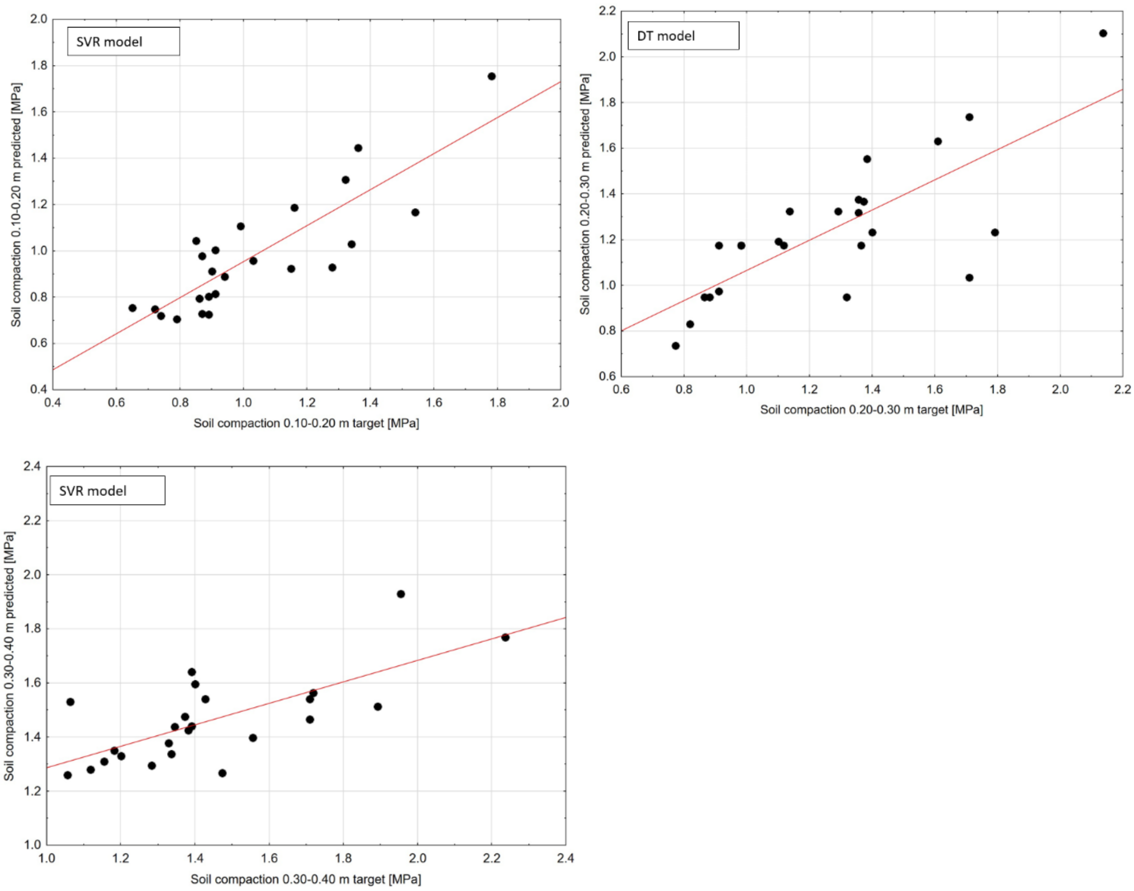


Fig. 3. Predicted values versus target values of soil compaction [MPa] measured at different depths for the best models

Table 1 presents the parameters of the DT and SVR models. With regard to the training set, the DT model for a soil layer of 0.1–0.2 m yielded the highest R-value of 0.87, while the SVR model for the same soil layer exhibited a slightly lower value of 0.86. The lowest R-value of 0.73 was observed for the DT model for a soil layer of 0.3–0.4 m. In the training set, the lowest MAPE error values were observed for depth 0.3–0.4 m for the SVR model (8.26%) and for the DT model (9.62%). With regard to the test set, the highest R-value of 0.85 was obtained for the SVR model for a soil layer 0.1–0.2 m. This model also exhibited the lowest MAPE error value of 11.31%. The hyper-parameters of the model were as follows: C = 100, and γ=0.001. In the case of the DT model and depth 0.1–0.2 m, the highest MAPE error value of 16.91% was observed, accompanied by the lowest R-value of 0.7. The differences in the quality of the models obtained for soil compaction measured at different depths are not great, but it can be seen that the models for depths of 0.3–0.4 m are of slightly lower quality. This may be due to the fact that the ECa measurements with the Veris 3100 scanner were taken for a shallow (0–0.3 m) soil depth. This means that this parameter better illustrates changes in soil compaction for soil layers closer to the surface.

4. Discussion

The use of soil electrical parameters to assess soil compaction or other parameters that are its key indicators has been presented in the scientific literature. Predictive models have been developed using both statistical and machine learning methods. Pathirana et al [15] used geophysical data to predict soil bulk density, a key indicator of soil compaction. They used random forest regression and obtained a model accuracy assessed by a coefficient of determination of R2>0.8. An unsupervised machine learning algorithm was successfully used by Romero-Ruiz et al [16] to identify field clusters associated with soil compaction values. The XGBoost model developed by Driba et al. was reported to be an accurate tool for predicting the spatial variability of soil parameters based on ECa values [17]. EM38-mk scanner data as input parameters to ML models were used by Zeyliger et al. [18] to produce accurate (R2>0.6) cartographic models of gravimetric soil moisture.

5. Conclusions

The challenges facing agricultural crop production are significant, with climate change and market competition representing two of the most pressing issues. Farmers are compelled to enhance yield while curtailing production costs in order to optimise profitability. An understanding of the prevailing climatic conditions in the field, the abundance of soil nutrients and the use of informed chemical crop protection are of significant importance in achieving maximum yields. The market offers a range of technical solutions that assist farmers in monitoring various soil and climate parameters. One parameter that has a significant impact on plant growth and yield is soil compaction, as evidenced by the findings of numerous researchers. Excessive soil compaction impairs the formation of the root system, thereby reducing the uptake of nutrients and ultimately leading to a decline in yield. However, measuring soil compaction with a cone penetrometer is a laborious and time-consuming process, making it challenging to implement on large-scale farms. The study concluded that the knowledge of soil electrical parameters, such as electrical conductivity and magnetic susceptibility, can be employed to predict soil compaction through the utilisation of machine learning techniques. The results of the analyses indicated that the SVR and DT models analysed in this study have the potential to predict soil compaction in different soil layers. However, the analysis did not provide a definitive answer regarding the superiority of either model for predicting compaction. The combination of soil scanning and machine learning methods represents a promising approach for a prediction of soil compaction in a rapid and cost-effective manner. Knowing the spatial distribution of soil compaction makes it possible to plan tillage at variable depths, resulting in reduced fuel consumption, which has a positive impact on the economic and environmental aspects of farming. It is also possible to plan soil loosening treatments according to the depth of the root system of the crops being grown. Quick and inexpensive prediction of soil compaction may lead to more farmers taking such measurements. Using our results on a large scale will require additional trials on soils of different textures.

References

1. Brevik EC., Fenton T.E., Lazari A.: Soil electrical conductivity as a function of soil water content and implications for soil mapping. Precision Agriculture, 2006, 7(6): 393-404.
2. Che Othaman N.N., Md Isa M.N., Ismail R.C., Ahmad M.I., Hui C.K.: Factors that affect soil electrical conductivity (EC) based system for smart farming application. The 2nd International Conference on Applied Photonics and Electronics 2019, 2020, 2203(1): 020055-1- 020055-6.
3. Friedman SP.: Soil properties influencing apparent electrical conductivity: a review. Computers and Electronics in Agriculture, 2005, 46(1): 45-70.
4. Hemmat A., Adamchuk VI.: Sensor systems for measuring soil compaction: Review and analysis. Computers and Electronics in Agriculture, 2008, 63(2): 89-103.
5. Hiura T., Okada H., Terada C., Nakamura M., Kaneko N.: Effects of soil compaction on above-and belowground interactions during the early stage of forest development. Urban Forestry & Urban Greening, 2024, 102: 128565.
6. Hoorman J., Moraes J.C., Reeder R.: The Biology of Soil Compaction. Agriculture and Natural Resources, 2009.
7. Lipiec J., Usowicz B.: Temporal variability in water content, temperature, and electrical conductivity of sandy soil: Responses to soil-improving cropping systems (SICS). 2024, Soil and Tillage Research, 224: 106166.
8. Mat Su AS., Adamchuk VI.: Temporal and operation-induced instability of apparent soil electrical conductivity measurements, Frontiers in Soil Science, 2023, 3: 1137731.
9. Nawaz M., Bourrié G., Trolard, F.: Soil compaction impact and modeling. A review. Agronomy for Sustainable Development, 2013, 33: 291-309.

10. Pentoś K., Mbah J., Pieczarka K., Niedbała G., Wojciechowski T.: Evaluation of Multiple Linear Regression and Machine Learning Approaches to Predict Soil Compaction and Shear Stress Based on Electrical Parameters. Applied Sciences, 2022, 12(17): 8791.
11. Tekin AB., Yalçin H.: Design and development of a front-mounted on-the-go soil strength profile sensor. Turkish Journal of Agriculture and Forestry, 2019, 46(2): 151-163.
12. Vrindts E., Mouazen A.M., Reyniers M., Maertens K., Maleki M.R., Ramon H., De Baerdemaeker J.: Management Zones based on Correlation between Soil Compaction Yield and Crop Data. Biosystems Engineering, 2005, 92(4): 419-428.
13. Brogi C., Huisman J., Herbst M., Weihermüller L, Klosterhalfen A., Montzka C., Reichenau T. G., Vereecken H.: Simulation of spatial variability in crop leaf area index and yield using agroecosystem modeling and geophysics-based quantitative soil information. Vadose Zone Journal, 2020, 19(1):e20009.
14. De Carlo L., Vivaldi G.A., Caputo M.C.: Electromagnetic induction measurements for investigating soil salinization caused by saline reclaimed water. Atmosphere, 2022, 13(1): 73.
15. Pathirana S., Lambot S., Krishnapillai M., Cheema M., Smeaton C., Galagedara L.: Integrated ground-penetrating radar and electromagnetic induction offer a non-destructive approach to predict soil bulk density in boreal podzolic soil. Geoderma, 2024, 450: 117028.
16. Romero-Ruiz A., O'Leary D., Daly E., Tuohy P., Milne A., Coleman K., Whitmore A. P.: An agrogeophysical modelling framework for the detection of soil compaction spatial variability due to grazing using field-scale electromagnetic induction data. Soil Use and Management, 2024, 40: e13039.
17. Driba D.L., Emmanuel E.D., Doro K.O.: Predicting wetland soil properties using machine learning, geophysics, and soil measurement data. Journal of Soils and Sediments, 2024, 24: 2398–2415.
18. Zeyliger A., Chinilin A., Ermolaeva O.: Spatial Interpolation of Gravimetric Soil Moisture Using EM38-mk Induction and Ensemble Machine Learning (Case Study from Dry Steppe Zone in Volgograd Region). Sensors, 2022, 22: 6153.