



GEO-RADAR LNAPL MODEL BY ARTIFICIAL INTELLIGENT

Mimi Diana Ghazali^{*1,2}, Othman Zainon², Mohamad Adib Abdullah³, Faradina Marzukhi⁴

¹ School of Science Geomatics, University Teknologi MARA Perlis Branch, Campus of Arau, Arau, Perlis.
Email: mimidiana@uitm.edu.my

² Department of Geoinformation, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, UTM Skudai, Johor

Email: othmanz.kl@utm.my

³ AJ Surveyor and Engineering Sdn.Bhd

Email: m.adib731@gmail.com

⁴ School of Surveying Science and Geomatics, University Teknologi MARA, Perlis Branch, Campus of Arau.

Email: faradina@uitm.edu.my

* Corresponding Author

Article Info:

Article history:

Received date: 01.10.2021

Revised date: 01.11.2021

Accepted date: 20.11.2021

Published date: 01.12.2021

To cite this document:

Ghazali, M. D., Zainon, O., Abdullah, M. A., & Marzukhi, F. (2021). Geo-Radar LNAPL Model By Artificial Intelligent. *Journal of Information System and Technology Management*, 6 (24), 243-253.

DOI: 10.35631/JISTM.624023

This work is licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)



Abstract:

Ground Penetrating Radar (GPR) as non-destructive measurement of full-wave electromagnetic (EM) backscatter completely relies on dielectric permittivity. Interpreting GPR reflection configuration is a complex qualitative with positioning and depth determination would be misleading due to severe polarization and velocity mismatch in travelling-wave. As a result of these studies, a GPR signal segmentation algorithm model was developed to map and identify light non-aqueous liquid (LNAPL) contaminated in laterite soil utilizing dielectric permittivity prediction. Simultaneous registration of a Global Positioning System (GPS) signal was performed while acquiring georeferenced GPR data sets to pinpoint the appropriate location of the soil layers. In this way, georeferenced GPR dispersion was assessed and dielectric permittivity was retrieved by velocity extraction. Empirical model relationship was established by the higher-order regression. Calibration function used verification measurement with root mean square error (RMSE) and calibrated Performance Network Analyzer (PNA). Segmentation and classification using Support Vector Machine (SVM) classifier as Artificial Intelligence (AI) was executed using predicted dielectric permittivity to construct the GPR automated recognition model. The model was compared with actual data and logistic regression classification. The result shows both classification techniques have provided good quality with root mean square errors (RMSE),

which were 0.1391 and 0, respectively. The classification produces correct instances classified above 98% for SVM and 100% for logistic regression.

Keywords:

Ground Penetrating Radar, Support Vector Machine, LNAPL, Regression

Introduction

GPR is completely reliant on dielectric permittivity, which depends on variance moisture content. Dielectric permittivity is reflected in the velocity motion's effectiveness, which determines the accuracy of the actual depth. Thus, the presence of LNAPL in soil tomography will trigger the dielectric permittivity values (Punia et al., 2021). The response of that kind of LNAPL may affect the soil water content and the dielectric. The migration of LNAPL is controlled by structural characteristics, moisture content, LNAPL composition, temperature (Punia et al., 2021) and soil properties (Arun et al., 2020). Many studies have predicted dielectric permittivity via empirical relationship models from field-based correlation, apart from moisture content. Most of the empirical models are initiated by TDR, such as Topp's model (Topp et al., 1980) and Complex Refractive Index Model (CRIM)(Roth et al., 1992). However, these models are applied for soil with similar deposits. Besides, interpretation of GPR scattering images is intrinsically a challenge. EM wave propagation is a susceptible and qualitative feature limited to noisy data and it leads to misleading depth and positioning interpretation.

Consequently, an alternative method to reduce faults in interpreting GPR images is the automatic recognition model (ARM) which uses Artificial Intelligence (AI). AI is a computational technique that has progressed GPR forward from locating and testing to imaging. Many studies have been carried out to understand GPR signal behaviour by re-cluttering the GPR signal usage (AI) such as Artificial Neural Networks (Travassos et al., 2018), Convolutional Neural Networks, Machine learning etc.

Image segmentation and classification are applications in the machine learning model such as Support Vector Machine (SVM) in computer vision which is constructed by training data. The training data is labelled by a number of classes for each pixel from a set of pixels. While in image classification, the classifier determines the non-linear boundaries generated from the training data by separating classes in the feature space.

The study aims is to carry out a detailed examination of the efficacy of SVM in modelling LNAPL contamination based on GPR images, with the hypothesis that SVM would provide a better understanding for GPR image interpretation.

Methodology***Study Area***

This study used experimental design is to which formulated an empirical model for petrophysical parameters for Terap Red soil in the most extensive rice-producing area of Peninsular Malaysia. Terap Red soil samples have been were collected from the Harum Manis

cultivation area surrounding Perlis, Malaysia (6.2659N, 100.1648E). This soil has its own unique characteristics that have been is able to retain water for a long time (Ghazali et al., 2020).

Empirical Relationship Model

Such A relationship that describes dielectric permittivity from moisture content, $\epsilon_r(\theta_v)$ was has been established from four (4) regression analyses through Pearson technique: (i) simple linear, (ii) logarithmic, (iii) second-order polynomial and (iv) third-order polynomial. Empirical relationship appears to be the most accurate model to describe correlated θ_v and ϵ_r . The model that suggests higher-order regression by the Pearson technique is depicted in the following relationship:

$$\epsilon_r = \alpha + \beta_1\theta_v + \beta_2\theta_v^2 + \beta_3\theta_v^3 \quad (1)$$

where ϵ_r is real relative permittivity element, while β denotes constant-coefficient for each predictor variable of θ_v (moisture content) and α (constant of intercept in $\epsilon_r(\theta_v)$ plot) to represent the value of ϵ_r at $\theta_v = 0$.

Pearson's technique has been selected as available data which is a parametric form obtained from the normal distribution histogram as illustrated in Figure 1.

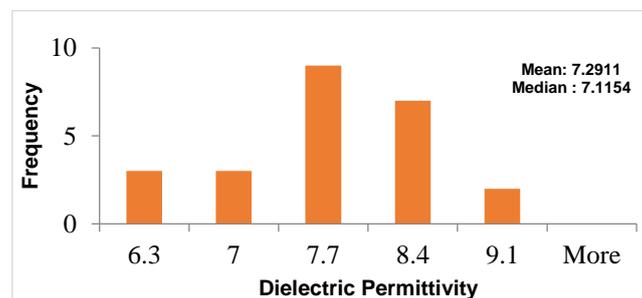


Figure 1 Normal Distribution Of Terap Red Soil

The Pearson's correlation and linear regression have been continued after the null hypothesis has been was rejected in the t-statistic test. Pearson's correlation coefficient (r or R) was also calculated to retrieve the degree of relationship between the dielectric permittivity and moisture content indicated in the soil. The coefficient correlation is calculated with the following equation (CY Piaw, 2006);

$$r^2 = 1 - \frac{\sum (Y - Y_p)^2 \div N}{\sum (Y - \bar{Y})^2 \div N} \quad (2)$$

where r^2 is the variance, Y is the dependent variable from regression and Y_p is the predicted value of Y from linear regression and \bar{Y} the mean of the dependent variable.

The correlation coefficient and linear regression model significance were evaluated using the value F-value using the ANOVA test. The F-value was also intended to test the null hypothesis by comparing the variance used by Amador-Muñoz et al. (2020). The level of significance (α) for the F-value test was set at $p < 0.05$ (95% confident level). This significant $p < 0.05$ was used by a number of researchers in their statistical studies, such as Leewis et al. (2013), Radziemska & Fronczyk (2015) and Saint-Laurent & Arsenault-Boucher, (2020).

Accuracy Verification

To quantify the accuracy of best-fitting relationships, i) standard error and (ii) root mean square error (RMSE) have been used to exhibit a high degree of predicted dielectric permittivity, as prescribed by C. H. Roth et al. (1992) and Szyplowska et al., (2018). RMSE is determined using,

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\epsilon_{\text{predicted}} - \epsilon_{\text{measured}})^2}{N}} \quad (3)$$

where N is the sample number, $\epsilon_{\text{measured}}$ is the relative permittivity obtained from GPR measurement, and $\epsilon_{\text{predict}}$ is the relative permittivity obtained from the predicted data. The standard regression error is the average distance between the observed values and the regression line.

Thus, in order to gain a functional value for the dielectric permittivity of Terap Red soil, the calibration value was using used an Agilent Performance Network Analyzer E8562B (PNA E8562B), as shown in Figure 2(a). The dielectric permittivity analyses were taken with an open-ended coaxial probe method, as applied by Šarlah et al. (2019). The PNA calculated the reflection coefficient using the backscatter response of electrical networks whenever the antenna was exposed to soil or any material, as shown in Figure 2(b).

Furthermore, prior to extracting the GPR data, the processed B-Scan GPR data were evaluated using Signal to Noise Ratio (SNR) and Normalize Root Mean Square Error (RMSE). Signal to noise to the ratio (SNR) is the calculated ratio of a received signal that compares the intended signal level towards the background noise level (Ciampoli et al., 2019). SNR (in decibel, dB) shall be specified as (Ai et al., 2018):

$$SNR = 10 \log_{10} \frac{\sum_k^N r(k)^2}{\sum_k^N [f(k) - r(k)]^2} \quad (4)$$

where $f(k)$ is the signal or amplitude of the raw data containing noise at a given scale level k to N , $r(k)$ is the de-noised signal or amplitude and N is the length of the signal or might be the number of samples composing the signals. The effective value for quantifying the strength of the signals in the relevant SNR should be greater than 15dB. The increased SNR value provides a higher GPR dataset interpretability (Ciampoli et al., 2019).



Figure 2 The Measurement Of Dielectric Permittivity (A) PNA E8562B And (B) Open-Ended Coaxial Probe Method

While, NRMSE is a subset of RMSE that takes into account the variances of two groups of different scales. In GPR data processing, the NRSME represents the amplitude of the original data between corresponding de-noised or processed amplitude, which is defined in the following terms (Baili et al., 2009):

$$NRMSE = \sqrt{\frac{[f(k) - r(k)]^2}{[f(k) - \mu(k)]^2}} \quad (5)$$

where, $\mu(k)$ is the mean value of the GPR signal. The NRMSE illustrates how the de-noised amplitude of the GPR signal is close to that of the original signal.

Classification Model Process

Support vector machine which is a method based on statistical learning theory was used for image segmentation with an AI to classify areas potentially containing object reflections in Weka software developed by the University of Waikato. SVMs were operated by optimizing a line that best divides the data into two classes using only data instances from the training dataset. The training data set was formed due to the semantic pixel-by-pixel labelling task, which required classifying each pixel of an image into several categories.

The attributes of training datasets extracted from B-Scan GPR processed data were segmented as follows: (i) X coordinates (ii) Y coordinates (iii) depth, (iv) dielectric permittivity and (v) class: Contaminated & Non-contaminated. The dataset's coordinates were related to GPS measurements in the Rectified Skew Orthomorphic Projection-Geodetic Datum of Malaysia 2000 (RSO-GDM2000) coordinate system via the GPR data georeferenced. The dielectric permittivity was calculated using an empirical relationship model established in this research study. The findings of SVM were compared to those of the old standard statistical classification method, regression logistics (LR).

Many studies have compared LR and SVM for the purpose of determining the efficacy of new machine learning approaches, SVMs such as Musa (2013) and Kalantar et al. (2017). LR describes the relationship between categorical variables and a number of dependent factors, which can be categorical, continuous or binary variables. Independent variables in the LR could be labelled as 0 and 1, denoting the contaminated and clean areas, respectively, without requiring normal distribution criteria.

Results And Discussion

Empirical Relationship Model between Dielectric Permittivity and Moisture Content

Overall, the empirical relationship model for GPR data with SNR(dB)= 15.2438 and NRSME=0.2145 in each form of regression appears to be a good fit. All regressions have surpassed the significant level with a strong positive correlation as indicated by the R squared value of less than 0.9. However, the third-order polynomial tends to adhere more closely to the GPR permittivity measurement obtained by most other studies. A reasonable fit, $R^2 = 0.9892$, implies 98.92% was found for 800Mhz data in 24 hours of measurement for the Terap Red soil, as seen in Figure 3. The exceptional agreement is indicated by the third-order polynomial model with a positively strong correlation exhibits is that it was the best fitting for predictive dielectric permittivity, ϵ_r , of Terap Red soil contaminated by diesel. As shown in Table 1, The accuracy of the best-fit regression is acceptable with a standard error of 0.076212.

The model suggests higher-order regression with an outperformed test for goodness of fit as depicted in the following relationship:

$$\varepsilon_r = -9.2058 + 4.58583\theta_v - 0.4518\theta_v^2 + 0.0156\theta_v^3 \quad (6)$$

The significance of coefficient value is determined by using t-statistics test, which gives hypothesis $H(0): \alpha = 0$ against $H(a): \alpha \neq 0$, and $H(0): \beta = 0$ against $H(a): \beta \neq 0$. When this p (value of best-fit model is compared to the significance level, as shown in Table 2, it is less than 0.05, implying a 95 percent confidence level. A significant coefficient, P-value of regression was obtained 0.00578, 6.3391E-05, 7.4258E-05 and 3.8848E-05. As a result, the null hypothesis which was associated with the four (4) coefficients was rejected, and the model was considered highly significant. Throughout this best-fit empirical relationship model, the standard error of the regression coefficient was estimated at 2.9796, 0.9108, 0.0909, and 0.0029, respectively. The yield model tends to agree with the results stated by Topp (1980), Steelman & Endres (2011), Bello. Y. Idi (2013) and Patriarca et al. (2013), and Karim et al. (2018) that reported that the volumetric water content of soil has an indirect and a significant relationship with the dielectric permittivity value.

Frequent quantitative estimations of $RMSE_{\text{predicted}}$ values estimated for the best-fit empirical relationship model, with values as low as 9.57456E-14. The relative difference in the predicted dielectric permittivity values was calculated for best-fit relationship models between GPR and calibrated VNA measurements which was less than 0.60%.

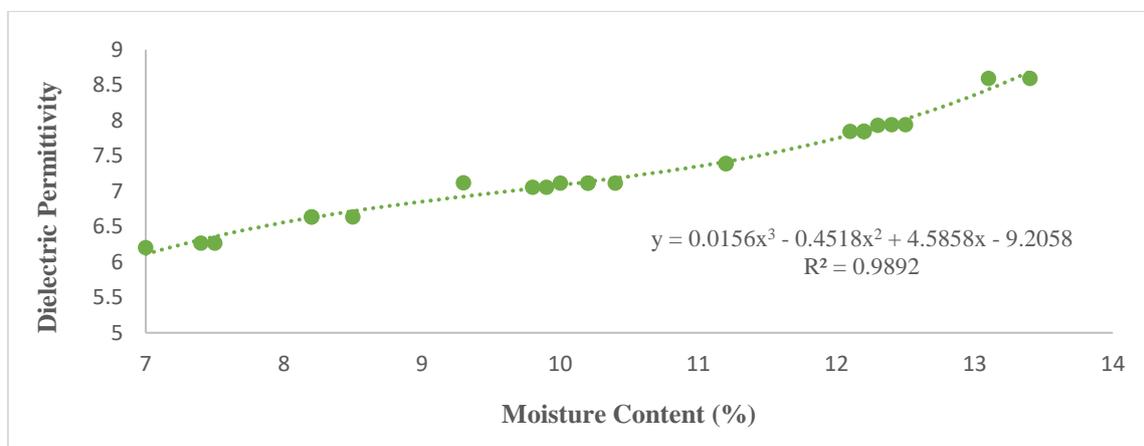


Figure 3 Line Of A Fit Plot Of The Empirical Relationship Model For Third-Order Polynomial Regression

Table 1 A Summary Of Empirical Relationship Model Outputs And ANOVA Analysis

Type of Regression	R ²	Df		F value	Significant F	Standard Error
		Regression	Residual			
Linear Regression	0.9599	1	23	526.6478	7.385E-17	0.1399
Logarithmic	0.9352	1	23	317.7474	1.4555E-14	0.1778
2nd-order polynomial	0.9743	2	22	397.7726	2.0301E-17	0.1147
3rd-order polynomial	0.9892	3	21	609.7075	8.0671E-20	0.0762

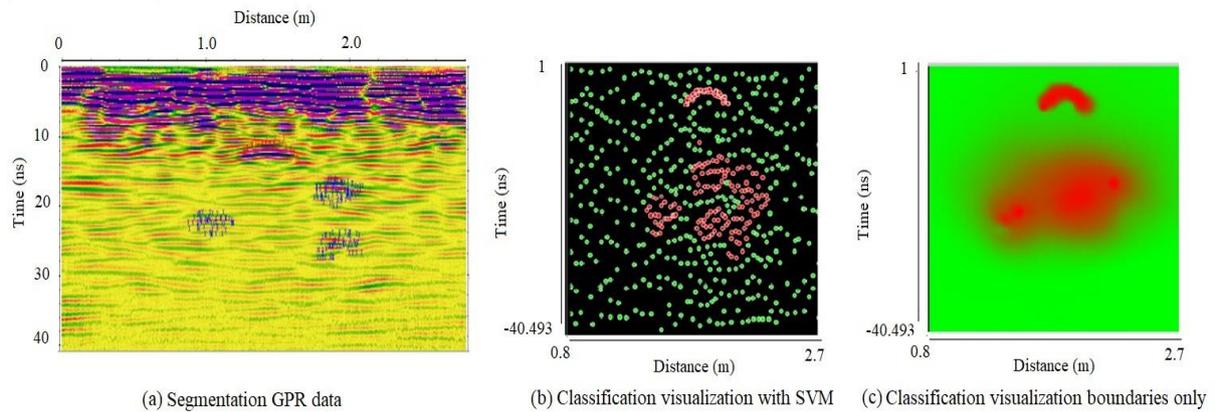
Table 2 A Summary Of Best-Fit Empirical Relationship Model By The Third-Order Polynomial Coefficient Analysis

	Coefficients	Standard Error	t Stat	P-value
Intercept	-9.2058	2.9796	-3.0896	0.0057
Moisture Content(%)(θ)	4.5858	0.9108	5.0352	6.3391E-05
Moisture Content (%)(θ) ²	-0.4518	0.0909	-4.9663	7.4258E-05
Moisture Content (%)(θ) ³	0.0156	0.0029	5.2493	3.8848E-05

LNAPL Automated Recognition Model

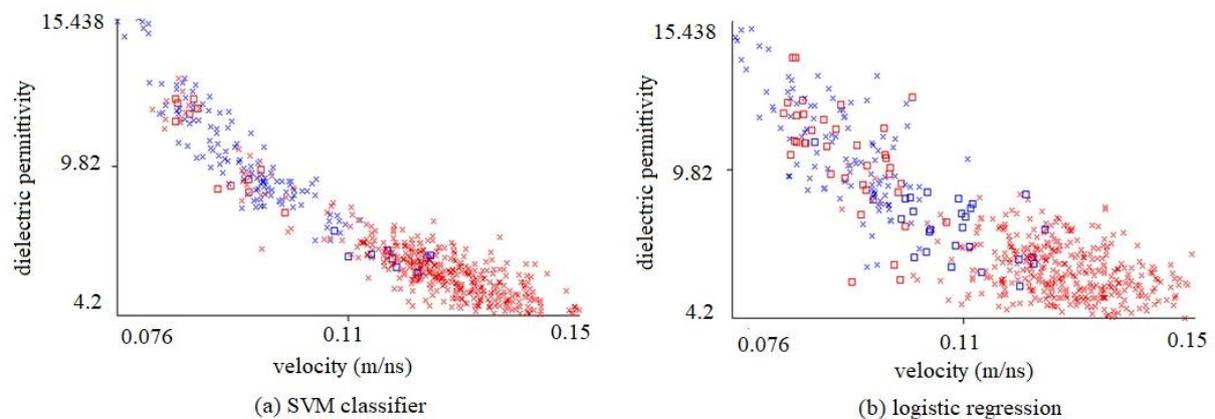
Image segmentation with an AI to classify classifies areas that are potentially containing object reflections. The aim is to map all of the different detected areas on the GPR B-Scan image that are perceived as contaminated zones. A semantic pixel-wise labelling task sets each pixel of an image to belongs to some classes (dielectric permittivity, depth, time, material (clean and contaminated)). Figure 4 shows the visual of the classification boundary of LNAPL using SVM via dielectric permittivity prediction and training data.

LNAPL boundary plume is defined by red and green for the clean area. Distribution classification indicates that the training data has a fairly significant consistency, which can be seen from the small number of red squares representing the data misinterpretation with correct instances classified (CIC) close to 100%, as seen in Figure 5. The difference in CIC is found to be negligible for SVM and LR, just less than 2%, as shown in Table 3. The LNAPL plume classification accuracy exhibits a good performance with RMSE value which obtains less than 0.2 for both classification techniques, SVM and LR. Similar accuracy which is obtained by Liu et al. (2020) to assesses the rebar positioning. Thus, according to Kappa Statistics, there is no significant difference between SVM and LR, and both have performed well with a distinct value of just 0.0577, as obtained by Kalantar et al. (2017).



(a) Segmentation GPR data (b) Classification visualization with SVM (c) Classification visualization boundaries only

Figure 4 Classifier Boundary Visualization for Terap Red soil



(a) SVM classifier (b) logistic regression

Figure 5 Classification Distribution Graph For SVM And LR

Table 3 Classification Accuracy Output

Type of Classification	SVM	Logistic Regression
Correct Instances Classified	98.0655%	100 %
Kappa Statistic	0.9423	1
Mean Absolute Error	0.0239	0
RMSE	0.1391	0
Relative Square Error	5.6432 %	0.0003%

Conclusion

SVM classifier model is modelled via dielectric permittivity prediction which is developed by higher-order regression. It has shown the ideal tools for the understanding of GPR data. The correct instance classified for SVM is achieved at 98.0655% with RMSE 0.1391. However, to facilitate the optimum model of classification, data training needs to be multiplied.

Acknowledgement

The researchers would like to express Dr Azremi, Dean of the School of Computer and Communication Engineering at Universiti Malaysia Perlis, for supplying the VNA. Furthermore, the author would also like to express gratitude to Universiti Teknologi MARA, Perlis, to provide funds, namely the Student Trust Fund (TAPA).

References

- Ai, Q., Liu, Q., Meng, W., & Xie, S. Q. (2018). Neuromuscular Signal Acquisition and Processing. In *Advanced Rehabilitative Technology*. doi.org/10.1016/b978-0-12-814597-5.00003-5
- Amador-Muñoz, O., Martínez-Domínguez, Y. M., Gómez-Arroyo, S., & Peralta, O. (2020). Current situation of polycyclic aromatic hydrocarbons (PAH) in PM_{2.5} in a receptor site in Mexico City and estimation of carcinogenic PAH by combining non-real-time and real-time measurement techniques. *Science of the Total Environment*, 703, 134526. doi.org/10.1016/j.scitotenv.2019.134526
- Arun, S., Kumar, R. M., Ruppa, J., Mukhopadhyay, M., Ilango, K., & Chakraborty, P. (2020). Occurrence, sources and risk assessment of fluoroquinolones in dumpsite soil and sewage sludge from Chennai, India. *Environmental Toxicology and Pharmacology*, 79(January), 103410. doi.org/10.1016/j.etap.2020.103410
- Baili, J., Lahouar, S., Hergli, M., Al-Qadi, I. L., & Besbes, K. (2009). GPR signal de-noising by discrete wavelet transform. *NDT and E International*, 42(8), 696–703. doi.org/10.1016/j.ndteint.2009.06.003
- Bello, Y. Idi. (2013). Mapping and modelling petrophysical biogenic gas concentration of Pontian Peatland, Southwest Malaysia with ground-penetrating radar.
- Ciampoli, L. B., Tosti, F., Economou, N., & Benedetto, F. (2019). Signal processing of GPR data for road surveys. *Geosciences (Switzerland)*, 9(2). doi.org/10.3390/geosciences9020096
- Piaw, C. Y. (2006). *Kaedah Penyelidikan buku 2*. Kuala Lumpur: McGraw-Hill (Malaysia) Sdn. Bhd.
- Ghazali, M. D., Zainon, O., Idris, K. M., Nor, S., Zainon, A., Karim, M. N. A., Anshah, S. A., & Talib, N. A. (2020). The Assessment of Relative Permittivity on Diesel Vapour in the Moisture Content of Terap Red Soil by Ground Penetrating Radar. 15–17. doi.org/10.1177/1178622120930661
- Kalantar, B., Pradhan, B., Naghibi, S. A., Motevalli, A., & Mansor, S. (2017). Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). 5705(December). doi.org/10.1080/19475705.2017.1407368
- Karim, N. I. A., Kamaruddin, S. A., & Hasan, R. C. (2018). Modeling of petrophysical relationship of Soil Water Content estimation at peat lands. *International Journal of Integrated Engineering*, 10(7), 177–187. doi.org/10.30880/ijie.2018.10.07.017
- Liu, H., Lin, C., Cui, J., Fan, L., Xie, X., & Spencer, B. F. (2020). Detection and localization of rebar in concrete by deep learning using ground penetrating radar. *Automation in Construction*, 118(December 2019), 103279. doi.org/10.1016/j.autcon.2020.103279

- Musa, A. B. (2013). Comparative study on classification performance between support vector machine and logistic regression. 13–24. doi.org/10.1007/s13042-012-0068-x
- Patriarca, C., Tosti, F., Velds, C., Benedetto, A., Lambot, S., & Slob, E. (2013). Frequency-dependent electric properties of homogeneous multi-phase lossy media in the ground-penetrating radar frequency range. *Journal of Applied Geophysics*, 97, 81–88. doi.org/10.1016/j.jappgeo.2013.05.003
- Punia, P., Bharti, M. K., Chalia, S., Dhar, R., Ravelo, B., Thakur, P., & Thakur, A. (2021). Recent advances in synthesis, characterization and applications of nano-particles for contaminated water treatment-A review. *Ceramics International*, 47(2),1526–1550. doi.org/10.1016/j.ceramint.2020.09.050
- Radziemska, M., & Fronczyk, J. (2015). Level and contamination assessment of soil along an expressway in an ecologically valuable area in Central Poland. *International Journal of Environmental Research and Public Health*, 12(10),13372–13387. doi.org/10.3390/ijerph121013372
- Roth, C. H., Malicki, M. A., & Plagge, R. (1992). Empirical evaluation of the relationship between soil dielectric constant and volumetric water content as the basis for calibrating soil moisture measurements by TDR. *Journal of Soil Science*, 43(1), 1–13.
- Saint-Laurent, D., & Arsenault-Boucher, L. (2020). Properties of alluvial and non-alluvial soils in fragmented mixed deciduous forest patches in southern Québec, Canada. *Catena*, 184(October 2019), 104254. doi.org/10.1016/j.catena.2019.104254
- Šarlah, N., Podobnikar, T., Mongus, D., Ambrožič, T., & Mušič, B. (2019). Kinematic GPR-TPS model for infrastructure asset identification with high 3D georeference accuracy developed in a real urban test field. *Remote Sensing*,11(12). //doi.org/10.3390/rs11121457
- Steelman, C. M., & Endres, A. L. (2011). Comparison of petrophysical relationships for soil moisture estimation using GPR ground waves. *Vadose Zone Journal*, 10(1), 270–285. doi.org/10.2136/vzj2010.0040
- Szypłowska, A., Szerement, J., Lewandowski, A., Kafarski, M., Wilczek, A., & Skierucha, W. (2018). Impact of Soil Salinity on the Relation between Soil Moisture and Dielectric Permittivity. *2018 12th International Conference on Electromagnetic Wave Interaction with Water and Moist Substances, ISEMA2018*,1–3. doi.org/10.1109/ISEMA.2018.8442298
- Topp, G. C., Davis, J. L., & Annan, A. P. (1980). Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resources Research*,16(3),574–582. doi.org/10.1029/WR016i003p00574
- Travassos, X. L., Avila, S. L., & Ida, N. (2018). Artificial Neural Networks and Machine Learning techniques applied to Ground Penetrating Radar: A review. *Applied Computing and Informatics*, 1–7. http://doi.org/10.1016/j.aci.2018.10.001