Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2025, 19(11), 59–75
https://doi.org/10.12913/22998624/208517
ISSN 2299-8624, License CC-BY 4.0

Received: 2025.04.30
Accepted: 2025.09.15
Published: 2025.10.01

Estimation of moisture content of cellular concrete with different apparent densities by time-domain reflectometry method using machine learning methods and regression analysis

Anna Futa[1], Magdalena Jastrzębska[1], Paweł Juszczyński[2], Anna Życzyńska[3], Henryk Sobczuk[4], Agata Zonik[4], Zbigniew Suchorab[4*]

[1] Department of Applied Mathematics, Faculty of Mathematics and Information Technology, Lublin University of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland
[2] Lublin University of Technology Doctoral School, Lublin University of Technology, Nadbystrzycka 38B, 20-618 Lublin, Poland
[3] Department of Conservation of Monuments, Faculty of Civil Engineering and Architecture, Lublin University of Technology, Nadbystrzycka 40, 20-618 Lublin, Poland
[4] Department of Water Supply and Wastewater Disposal, Faculty of Environmental Engineering and Energy, Lublin University of Technology, Nadbystrzycka 40B, 20-618 Lublin, Poland
* Corresponding author's e-mail: z.suchorab@pollub.pl

ABSTRACT

The moisture content of building materials, especially in porous media, such as cellular concrete, is a serious problem affecting the durability, quality of thermal insulation and safety of structures. Traditional methods of moisture assessment, based on gravimetric laboratory analyses, are time-consuming. The alternatives are the indirect techniques, such as time-domain reflectometry (TDR), which allow for quick measurements by analyzing the dielectric properties of materials. The aim of this work was to develop predictive models estimating moisture content of cellular concrete depending on its apparent density and other material parameters. The article used both classical regression models $\theta(\varepsilon)$, $\theta(\varepsilon,\rho)$ and artificial intelligence methods, including neural networks (NN), regression trees, support vector machines (SVM) and gaussian process regression (GPR) models. The $\theta(\varepsilon)$ model performed best at cellular concrete type 400 kg/m³ ($R^2 = 0.9433$), but its accuracy declined at higher densities. The universal $\theta(\varepsilon,\rho)$ model gives better results at 500 and 600 kg/m³ type cellular concretes, with an overall R^2 of 0.9340. However, AI models outperformed both of them. The GPR model achieved near-perfect predictions ($R^2 \approx 0.9999$, RMSE = 0.0021 – 0.0032 cm³/cm³), while SVM and NN also showed high accuracy ($R^2 = 0.9914–0.9960$) with significantly lower errors than deterministic regression models. Comparison of the effectiveness of these approaches allowed for the assessment of the accuracy of moisture prediction based on different data sources. The obtained results indicate the significant potential of AI application in monitoring the moisture content of building materials, offering more effective and precise diagnostic tools compared to traditional methods.

Keywords: moisture content, building materials, cellular concrete, time domain reflectometry, regression models, artificial intelligence.

INTRODUCTION

The presence of water in building materials is a fairly common phenomenon. This applies not only to old, historic buildings, but also to new facilities that may have problems with moisture due to improper execution of construction works. It is worth paying attention to this problem already at the design stage, because water in building partitions can lead to chemical and biological corrosion, mechanical damage to materials, a decrease in their insulating properties and cause significant financial losses [1]. The moisture content of walls and porous media is a result of the ability of these materials to absorb and store water. Porous media, such as walls, bricks, concrete, or soils, have a structure consisting of microscopic pores and cracks that can retain water [2, 3]. Water accumulates in them mainly due to the phenomenon of capillarity, which is when water penetrates narrow spaces (capillaries) under the influence of adhesion forces to the pore walls. These forces cause water molecules to climb up narrow spaces, even against the force of gravity. Water in porous materials can come from both atmospheric precipitation and moisture contained in the soil, which causes walls and other porous media to remain damp. This phenomenon is particularly visible in the buildings, where water from capillaries penetrates into the interior of the rooms, creating moisture-related problems such as mold or damage to building materials.

The moisture content of a porous medium, which includes most building materials, is expressed in various ways. One of the most commonly used parameters is the volumetric moisture content [4], also known as volumetric moisture content, expressed in [cm^3/cm^3] or [$\%_{vol}$]. According to the PN-EN ISO 12571:2013-12 standard [5], volumetric moisture content is the volume of water capable of evaporation related to the volume of dry material. The formula describing the volumetric moisture content is as follows [6]:

$$\theta_V = \frac{V_w}{V_{tot}} \tag{1}$$

where: θ_V – volumetric moisture content [cm^3/cm^3], V_w – volume of water able to evaporate [cm^3], V_{tot} – total sample volume [cm^3].

In addition to the indicators defining the moisture content of the material, the following parameters are also used to describe it: density, bulk density, tightness, porosity and maximum moisture content of the medium – water absorption by weight and volume [7]. These parameters result from the structure of the porous medium and are of high importance in assessing the level of moisture content of the building partition, constituting a reference point to which the measured moisture content is compared.

In connection with the problem of moisture in building materials, the development, elaboration and improvement of moisture detection techniques in building partitions play an important role. It should be noted that the most accurate determination of water content in partitions is possible owing to laboratory methods, which consist of material sampling, weighing and drying, which allows for a precise determination of the amount of water. However, these techniques are often impractical due to the need to drill the masonry and the long waiting time for the results. The alternatives are indirect techniques, which allow for quick measurements without damaging the structure of the partition and provide high precision.

One of the most important moisture detection techniques is the TDR method, namely the Time Domain Reflectometry. This is an electrical technique that involves measuring the dielectric parameters of a medium using reflectometry [8]. For many years, the technique has been used to measure moisture in soil media. The TDR technique works on the principle of sending electromagnetic pulses in the form of a wave through a sensor placed in the tested medium. When the wave encounters the transition between different sensor construction elements (e.g. beginning of the measuring element or its termination), part of the wave is reflected back to the sensor. The time that elapses from sending the pulse to its reflection allows for determining the dielectric parameters of the measured medium, which are closely related to its moisture.

The relationship between the dielectric parameters exhibited by moist porous media and the moisture of the media is most often presented in the form of physical and empirical models [9, 10]. The advantage of physical models is a certain independence from calibration tests. The disadvantages include a general, often complicated mathematical description. Another approach is to use empirical models based on laboratory measurements, which correlate the results of moisture measurements using the gravimetric method with dielectric permittivity. Among these, universal models and individual models can be distinguished, the former developed on the basis of various media, whereas the latter refer to a specific material, sensor or research procedure.

The most frequently cited empirical models used in the practical assessment of medium moisture include the Topp model [11], which takes the form of a third-degree polynomial:

$$\begin{gathered}\theta = -0.053 + 0.0292\varepsilon - \\ - 0.00055\varepsilon^2 + 0.0000043\varepsilon^3\end{gathered} \tag{2}$$

where: θ – volumetric water content in the tested porous medium [cm^3/cm^3], ε – dielectric permittivity of the medium measured by the TDR technique [-]. An alternative model that allowed for increased measurement accuracy was proposed in the work of Malicki et al. [12]:

$$\theta = \frac{\varepsilon^{0.5} - 0.819 - 0.168\rho - 0.159\rho^2}{7.17 + 1.18\rho} \quad (3)$$

where: ρ denotes the material density in a dry state [g/cm^3].

This model takes into account the density of the material in the dry state, which allowed for a more accurate representation of the moisture - dielectric permittivity relationship in materials characterized by different solid phase properties [13].

Regression analysis can be used to create empirical models for moisture assessment, using indirect detection methods as an alternative to traditional models. Regression is a statistical technique used to model the relationship between one dependent variable and one or more independent variables. The goal of regression is to predict the value of the dependent variable based on the values of the independent variables [14]. In the context of assessing the moisture content of materials, regression can be used to create the models that combine data from different measurement methods (e.g. gravimetric moisture content with dielectric permittivity) and predict the water content of media based on these variables. In this way, regression analysis provides a tool for developing more accurate and efficient moisture detection methods [15].

Artificial intelligence (AI) methods are becoming increasingly popular in the context of assessing the moisture content in building materials, offering a modern approach to monitoring and analyzing the moisture content [16, 17] The use of AI, including machine learning and deep learning, allows for the development of more precise and effective moisture detection tools, especially in comparison to traditional methods. One of the main applications of AI in this area is the analysis of measurement data from various sensors (e.g. moisture, temperature, dielectric permittivity). Machine learning algorithms, such as neural networks or decision trees, can be used to detect patterns in this data and predict the material moisture content based on the variables that are not always readily available in traditional measurement methods [18].

AI models are advanced computational systems based on artificial intelligence algorithms that can analyze data, recognize patterns, and make decisions based on the information provided. These models work on the principle of machine learning, which means that they can improve with the processing of more and more data. AI models can be divided into several main categories depending on the method of learning and the architecture used. The division of AI models is very wide, and each of them finds application in different fields. Supervised learning is ideal for classification and regression, unsupervised learning helps in the analysis of unknown patterns, and reinforcement learning allows the development of intelligent agents. Deep learning and generative models revolutionize image and text processing, whereas probabilistic and statistical models provide interpretability and accuracy in data analysis. Each of these categories of AI models has its unique features and applications; the selection of the right model depends on the problem to solve and the available computational resources. In this paper, the following AI models are presented: regression trees, GPR models, SVM models, and neural networks [19].

Regression trees are a special type of decision tree used to predict numerical values instead of classifying data [20]. Unlike classification trees, which assign data to specific categories, regression trees learn a function of the relationship between input variables and a continuous output value. The structure of a regression tree consists of decision nodes – defining the conditions for the partition, branches – representing possible decision paths, leaves (terminal nodes) – containing the predicted numerical values [21, 22]. A regression tree works by recursively splitting data into smaller groups in such a way as to minimize the prediction error. The partition into subsequent nodes is performed according to criteria such as: variance reduction – the tree looks for a partition that reduces the spread of values in the groups and mean squared error (MSE) – choosing a partition that minimizes the mean square error. Each leaf of the tree represents the average value of the samples contained in a given node. Simplicity of regression trees and intuitiveness make them widely used, but in practice they are often combined into more advanced ensemble models to increase their efficiency and stability of predictions [23].

GPR is a machine learning method that models the relationship between input and output data using a probabilistic approach. Instead of producing a single predicted value, it provides a distribution of possible outcomes, allowing the estimation of prediction uncertainty [24]. The process begins with training, where the input-output data is collected and assumed to come from an unknown function represented by a Gaussian process. A covariance function, or kernel, is chosen to define how input points are related, and a covariance matrix is built to capture these relationships. In the prediction stage, the model receives a new input and uses the Gaussian process and the existing covariance matrix to compute the predicted value along with its variance. This result includes both the expected output and a measure of the model's confidence in its prediction. Gaussian process regression is an advanced probabilistic regression method that allows not only for accurate prediction of values, but also for estimating the uncertainty of the prediction. It is particularly effective in modeling small data sets, where classic regression models may have difficulties. However, its high computational complexity makes it unsuitable for very large data sets, where methods such as random forests or boosting are more effective [25, 26].

Support vector machines (SVM) is a popular method used in machine learning, especially in classification tasks, although it can also be used for regression [23]. The main goal of SVM is to find a hypersurface (hyperplane) that best separates data into different classes. SVM is a supervised learning model that requires training data with assigned class labels. It works on the principle of maximizing the margin (distance) between classes, meaning that it looks for a hypersurface that maximizes the space between the closest data points from different classes, called support vectors. The Support Vector Machines algorithm works by finding a hypersurface (in multidimensional space) or line (in the case of two dimensions) that best separates different classes of data. The main steps of the SVM algorithm are hypersurface selection, margin maximization, kernel function usage, optimization and prediction [27]. Support Vector Machines are tools for solving classification and regression problems, especially effective in nonlinear cases and with large datasets with a high number of features. SVM is widely used in various fields, such as bioinformatics, image analysis, and pattern recognition. Although SVM can be computationally expensive and requires careful parameter selection, it remains one of the most effective algorithms in the field of classification [28]. Neural networks are a class of machine learning algorithms that are inspired by the human brain [29]. They are used to solve various problems, such as classification, regression, image recognition, text analysis, forecasting, and computer games [30]. Neural networks are the basis of deep learning, which has gained a lot of popularity in recent years. The algorithm of a neural network consists of passing data through layers of neurons, in which each layer processes information based on its weights and activation functions. The process starts with an input layer, which accepts input data, then the data passes through subsequent hidden layers, where it is processed, and finally reaches the output layer, which generates an output (e.g. prediction or classification). During training, the network compares its results with real values, calculates the error (the difference between the network output and the real value), and adjusts the weights using backpropagation of the error. The backpropagation algorithm calculates the error gradients and uses an optimization method (e.g. gradient descent) to change the weights so that the error decreases with each iteration. This process is repeated many times until the model achieves satisfactory accuracy. Neural networks are tools applied in the area of pattern recognition, image analysis, text and speech analysis. Although they have their drawbacks, such as high computational requirements and difficulties with interpretability, their capabilities in solving complex problems make them the foundation of modern artificial intelligence systems [31].

An example of utilizing AI models to detect porous material permittivity was presented in the article by Nimer et al. [9] The performance of the FFNN neural network model was compared with conventional approaches, including theoretical models (e.g., Silberstein, Birchak, and Looyenga) and empirical regression models. FFNN achieved correlation coefficients of 0.9942 (training), 0.9967 (validation), and 0.9977 (testing), significantly outperforming both the theoretical models (e.g., Silberstein: $R = 0.9163$, $MSE = 6.95$) and the best empirical model (full quadratic model: $R = 0.9773$, $MSE = 1.42$). These results indicate that machine learning can substantially enhance the accuracy of moisture and contamination assessments based on dielectric measurements.

AI methods, owing to their ability to recognize complex patterns in data, can significantly improve the accuracy and precision of estimation. In addition to artificial intelligence methods, traditional regression models were used in the article, such as second-degree polynomial regression of one and two variables. These models were used to extract the relationships between input variables (e.g. apparent density) and their significance for the prediction of moisture content of cellular concrete. The developed regression models were compared with the results obtained using artificial intelligence algorithms. This comparison allowed for an assessment of which of the methods gives better results in terms of precise moisture prediction and on which data the AI model is more effective. By using both artificial intelligence methods and classical regression models, it is possible to compare the different approaches of moisture forecasting and the assessment of their effectiveness in the context of different variables [32].

The aim of this article was to develop the predictive models that will allow for the estimation of moisture content of cellular concrete, taking into account the apparent density of this material, and to compare the effectiveness of different modeling approaches, including the use of artificial intelligence methods and classical regression models. The topic discussed in the article is of high importance, because it addresses a real and widespread problem in construction — the presence of moisture in building materials, which affects both new and historic structures. Moisture can lead to serious consequences, such as material degradation, mold growth, and reduced thermal insulation, directly impacting building durability and indoor comfort. The search for fast and accurate methods to detect moisture – such as TDR techniques and predictive modeling – offers practical solutions that are highly relevant for the construction industry. Moreover, the comparison between traditional regression and modern AI methods introduces an innovative and original aspect, highlighting the potential of machine learning in solving engineering problems. Previous approaches to calibrating the TDR method have included only conventional deterministic models. They mainly involved the relationship between the electrical permittivity and humidity of the tested media. However, the signals from TDR meters are very complex. A simple analysis consisting of determining the pulse propagation time and then the electrical permittivity causes a lot of information to be lost. The use of machine learning methods to determine the humidity of media allows for taking into account more information, owing to which the obtained measurements will be more accurate and the humidity estimation will be characterized by smaller errors.

MATERIALS AND METHODS

Measuring setup

Autoclaved cellular concrete samples with densities of 400, 500 and 600 kg/m³ from SOLBET Lubartów S.A. were utilized for the study (Lubartów, Poland). The materials were dried using a VO-500 laboratory dryer (Memmert, Germany). Weight measurements were taken using a WPT 6C/1 precision balance, manufactured by Radwag (Radom, Poland). Tests were carried out using a TDR meter "TDR / MUX / MPTS" from ETest (Lublin, Poland), allowing data recording. Calibration tests were performed for FP/mux probes from the same company, which were equipped with two metal rods of 100 mm length and 14 mm spacing. The test set prepared in this way allowed for accurate and comparative determination of the moisture content of different types of cellular concrete.

Materials

The test material was prepared in accordance with the established protocol. Each sample was meticulously trimmed to dimensions of approximately 50 × 50 × 120 mm, thereby ensuring optimal conditions for subsequent testing. The samples were subjected to a drying process for a duration of seven days, with the objective of achieving a constant weight. Subsequently, the TDR probes were installed into the prepared samples, and apparent permittivity was measured. Next, the samples were gradually moistened with water to achieve the state of saturation. During the whole calibration procedure, the probes were temporarily disconnected from the TDR multimeter, to measure and control their weight and moisture status. After that they were re-connected to the multimeter and apparent permittivity was read. This procedure was repeated for the samples of all densities used. Subsequent to the completion of the measurements, the mass data of the samples were converted to moisture content and the deterministic analysis was then performed. In parallel, machine learning algorithms were implemented to achieve optimal precision and discern the correlation between material parameters and their moisture content. To facilitate understanding of the experimental procedure, a block diagram is presented in Figure 1 which shows the particular stages of the experiment.

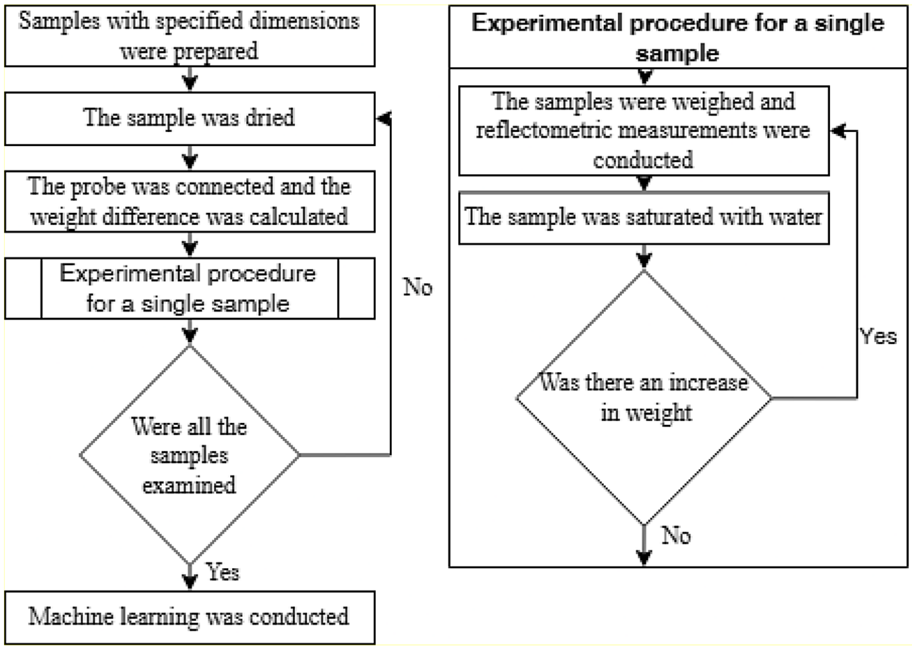

Figure 1. Block diagram of the experimental procedure

Description of regression analysis method

Regression is a data analysis method that examines the relationship between variables. A regression model allows predicting the value of one variable (dependent) based on other (independent) variables. A popular example is linear regression, which describes the relationship as a straight line. However, many physical phenomena have a non-linear nature, which in practice can often be expressed using a simple second-degree polynomial regression. Simple second-degree polynomial (quadratic) regression is a method of estimating the relationship between one independent variable and a dependent variable using a second-degree polynomial function. Contrary to classical linear regression, it allows modeling nonlinear relationships, which makes it a more flexible tool in data analysis [33]. The second-degree polynomial regression model takes the following form:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \tag{4}$$

where: y – dependent variable, x – independent variable, β_2, β_1, β_0 – structural model parameters, ϵ – random error.

This model can be interpreted as an extension of classical linear regression with an additional quadratic term, which allows for considering nonlinear effects in the data set. The β_2 coefficient determines the degree of curvature of the regression function – its sign indicates whether the function is convex ($\beta_2 > 0$) or concave ($\beta_2 < 0$). The β_1 coefficient represents the slope of the tangent to the regression line at x = 0, defining the effect of variable x on variable y. The β_0 coefficient is the intercept of the model, denoting the predicted value of y at $x = 0$.

In a situation where more than one independent variable is taken into account in the model, then a multiple regression model should be used. A frequently used model is a two-variable multiple regression model in the form of a second-degree polynomial:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \epsilon \tag{5}$$

where: y – dependent variable, x_1, x_2 – independent variables, β_0 – y-intercept, β_1, β_2 – linear effect coefficients, β_3, β_4 – quadratic effects coefficients, describing the nonlinear influences x_1 and x_2, ϵ – random error.

By taking into account quadratic terms, this model allows not only the prediction of the value of the dependent variable, but also the analysis of extreme points (maximum, minimum), which makes it particularly useful in process optimization [34]. The significance of the regression coefficients plays an important role in assessing the quality of the model and allows determining which variables have a real impact on the dependent variable. In second-degree polynomial regression, where both linear and quadratic terms are taken into account, significance analysis allows reducing the model by eliminating irrelevant terms, which reduces the risk of overfitting and improves the interpretability of the results. In classical regression analysis, the Student's t-test is used for each regression coefficient. This test examines the null hypothesis that a given independent variable has no significant impact on the dependent variable. If the null hypothesis is true, it means that the given coefficient can be removed from the model, because its impact is statistically insignificant. In turn, the alternative hypothesis is that this coefficient is different from zero, which means that the given variable has a significant impact on the dependent variable and should remain in the model [35]. To check whether the coefficient is significant, the value of the test statistic is calculated. The value of this statistic is then compared to the critical values of the Student t-distribution for the appropriate level of significance. Alternatively, the so-called p-value is calculated, which is the probability of obtaining such or more extreme values of the test statistic, provided that the null hypothesis is true. If $p < 0.05$, then the null hypothesis is rejected and the coefficient is considered to be statistically significant. This means that the variable in question does indeed affect the dependent variable. If $p > 0.05$, then there is no sufficient evidence to reject the null hypothesis, which suggests that the variable in question has no significant effect on the dependent variable and can be removed from the model. In summary, analyzing the significance of regression coefficients allows for a better understanding of the model and ensures that the variables considered really affect the dependent variable. Eliminating irrelevant parameters improves interpretation, reduces the risk of overfitting, and increases modeling efficiency [36].

Description of artificial intelligence models

The moisture content of cellular concrete was estimated using four artificial intelligence algorithms: GPR, SVM, Tree and neural network (NN). In the case of the GPR algorithm, Matern 5/2 GPR preset model with automatic kernel scale was used. In the case of the SVM algorithm, Linear SVM preset model was used with linear kernel function and automatic kernel scale. In the case of the Tree algorithm, Fine Tree preset model was applied with minimum leaf size equal 4 and finally for neural networks, Trilayered Neural Network was applied with the following sets: number of fully connected layers – 3, sizes of all layers were set to 10.

For teaching of the AI models the following amount of data was used: in the case of the aerated concrete 400 kg/m^3 the number of responses was equal 264, for concrete 500 kg/m^3 - 336 and 600 kg/m^3 - 360. The length of predictors for all densities was equal to 1023. The AI models were derived using Matlab software [37].

Regression and AI models fitting measures

An important aspect during the construction of regression models and AI models is the analysis of the degree of fit of these models to empirical data, which requires assessment using various fit measures. The most commonly used measures are root mean squared error (RMSE), mean absolute error (MAE) and determination coefficient (R^2).

The root mean squared error is the square root of the arithmetic mean of the squares of the differences between the predicted and observed values. RMSE measures the average distance, expressed in original units, between the empirical values of the explained variable and the theoretical values obtained on the basis of the model, cf. [38]. In other words, RMSE determines how much the empirical values of the explained variable deviate on average from the theoretical values calculated on the basis of the built model. The smaller the RMSE value, the closer the theoretical values are to the actual values, which indicates a better fit of the model.

The mean absolute error is another measure that assesses how much the observed values of the dependent variable differ, in absolute value, from the values predicted by the model. MAE is a metric that treats all errors equally, regardless of their size, meaning that it does not favor or discriminate against errors with a larger range [39]. Furthermore, when comparing MAE to RMSE, the RMSE value is usually larger, suggesting that there are errors with a larger range. Analyzing the difference between RMSE and MAE values can indicate the presence of large errors that have a significant impact on the model.

The determination coefficient indicates how well the predictor variables describe the variability of the explained variable. This coefficient determines what part (percentage) of the variance of the dependent variable is explained by the independent variables. The closer the R^2 coefficient value is to 100%, the better the regression model describes the behavior of the dependent variable being studied [40]. Adding insignificant predictor variables to the regression equation often increases the value of the R^2 coefficient. In order to eliminate this phenomenon, the so-called adjusted determination coefficient should be used.

RESULTS AND DISCUSSION

The moisture content of cellular concrete was estimated using different methods: regression models and artificial intelligence algorithms. The aim of this analysis is to determine which approach best predicts moisture content based on material properties such as permittivity (ε) and density (ρ).

Deterministic models

At first, two regression models for estimating the moisture content of cellular concrete were analyzed. Deterministic models were developed based on the values of electrical permittivity read with FP/mts probes and the moisture values determined gravimetrically in the laboratory. The $\theta(\varepsilon)$ model was based solely on the apparent permittivity (ε), for each density separately. Model $\theta(\varepsilon,\rho)$, on the other hand, is an ensemble model that combines the permittivity readings from all samples with different densities. Table 1 presents the regression models developed to estimate the moisture content of cellular concrete based on dielectric permittivity and, in one case, apparent density. For each density level (400, 500, and 600 kg/m³), separate models were developed using only dielectric permittivity as the predictor. These models show that the relationship between permittivity and moisture content is nonlinear and varies depending on the density of the material, which means that each density class requires an individual calibration for accurate moisture estimation. Additionally, a universal model was developed using the data from all densities combined. This model includes both dielectric permittivity and density as variables, aiming to generalize the prediction across different material types. However, despite the inclusion of density, earlier statistical analysis indicates that its influence on moisture estimation may not be statistically significant. This suggests that dielectric permittivity remains the dominant factor in predicting moisture content, while the role of density may be limited or dependent on the specific context.

The performance of the obtained models is also presented graphically. Figure 2 shows the relationship between the values obtained using the TDR method for three density levels and those predicted by the single-variable regression model $\theta(\varepsilon)$. In each subplot, the predicted values (yellow) follow the overall trend of the actual values (blue), which indicates that the model captures the general behavior of the data. However, some discrepancies between predicted and true values are visible, especially in the regions with rapid moisture content changes. This suggests that while the model is effective, there may be certain areas where its predictive accuracy decreases, potentially due to variability in material structure or measurement noise.

Figure 3 presents the estimated values along with the fit lines for this model. The closer the points lie to the diagonal (perfect prediction line), the better the model performance. For all three densities, the points generally align along the diagonal, confirming that the $\theta(\varepsilon)$ model provides reasonably accurate predictions. The 600 kg/m³ model appears to have the tightest clustering around the diagonal, indicating the highest accuracy among the three. The 400 kg/m³ model shows greater scatter, suggesting slightly lower prediction accuracy for lower-density concrete.

Table 1 also includes the universal model of two variables $\theta(\varepsilon,\rho)$. This model additionally includes the density (ρ) and is a universal model, applied for all densities simultaneously. Figure 4 presents the relationship between the values obtained using the TDR method together for all density levels and the values predicted by the $\theta(\varepsilon,\rho)$ universal regression model (a) and the

Table 1. Regression model equations

Density [kg/m³]	Regression model
400	$\theta(\varepsilon) = -0.1009\cdot\varepsilon^2+3.6079\cdot\varepsilon-2.3334$
500	$\theta(\varepsilon) = -0.1114\cdot\varepsilon^2+3.9804\cdot\varepsilon-2.9716$
600	$\theta(\varepsilon) = -0.1018\cdot\varepsilon^2+3.9148\cdot\varepsilon-1.8776$
all	$\theta(\varepsilon,\rho) = -0.09574\cdot\varepsilon^2+3.693\cdot\varepsilon+0.00003591\cdot\rho^2-0.02321\cdot\rho+0.2936$

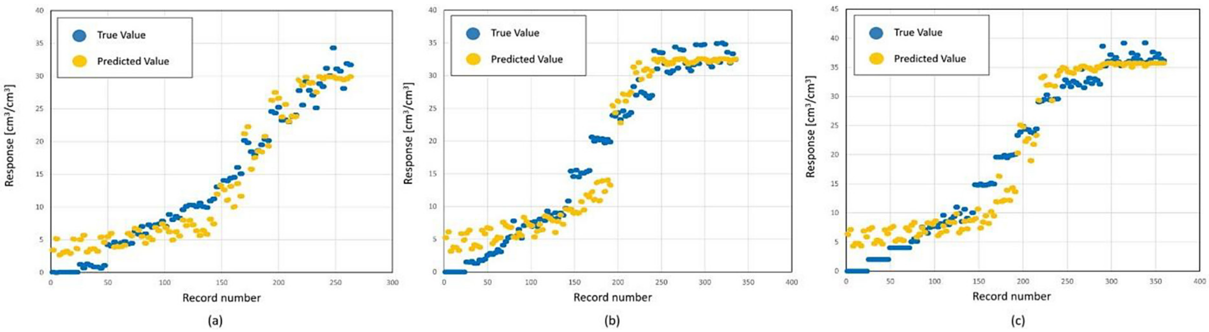

Figure 2. Relationships between values obtained using the TDR method and those predicted by the *θ(ε)* model for: (a) density 400 kg/m^3, (b) density 500 kg/m^3 and (c) density 600 kg/m^3

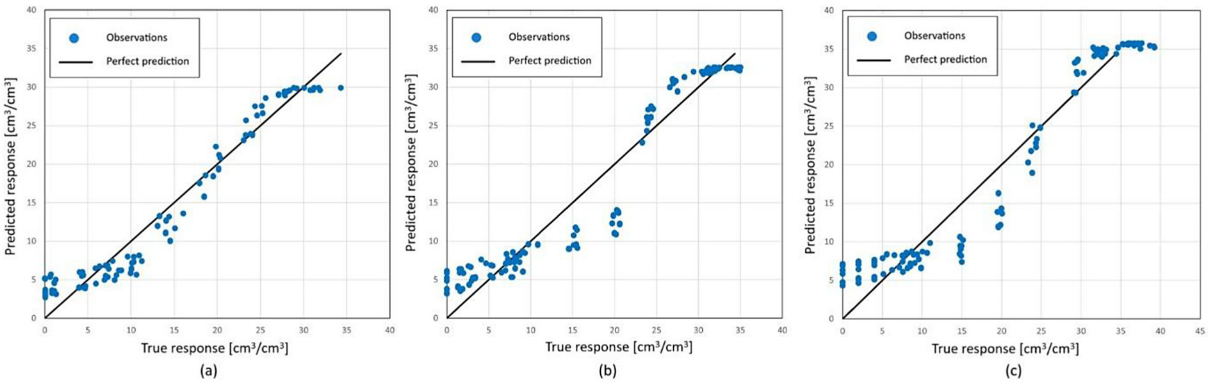

Figure 3. Estimated values along with the *θ(ε)* model fit line for: (a) density 400 kg/m^3, (b) density 500 kg/m^3 and (c) density 600 kg/m^3

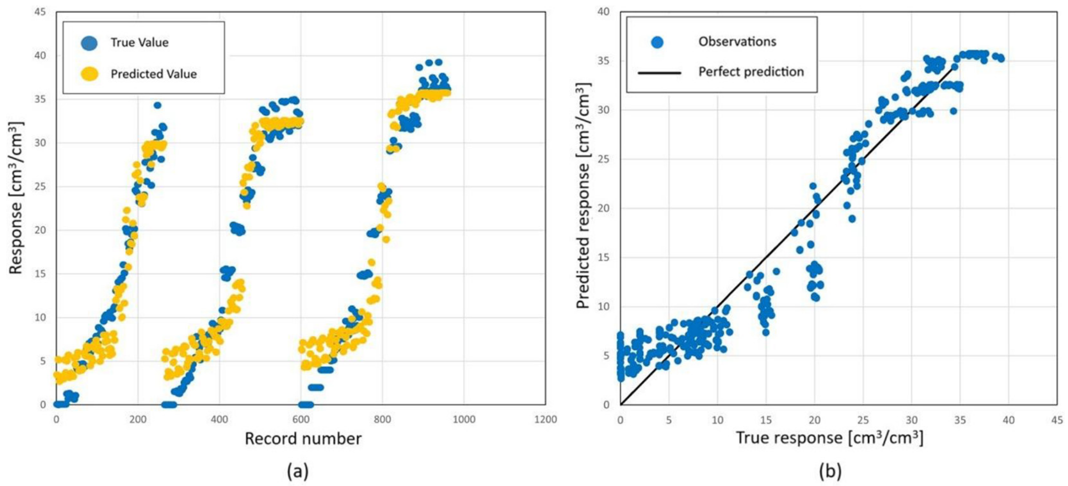

Figure 4. (a) Relationships between the values obtained using the TDR method and predicted values by the *θ(ε, ρ)* model and (b) the estimated values together with the fit line of the *θ(ε, ρ)* model

estimated values together with the fit line (b). On the left (a), predicted values (blue) closely follow the true values (yellow), indicating good model performance, though some deviations appear at lower and mid moisture levels. On the right (b), the predicted values aligning well along the line of perfect prediction confirm this, showing high accuracy. On the basis of Table 2, the parameter *ε* is consistently statistically significant across all models (400, 500, 600, and the universal model). All associated p-values are less than 2e^{-16}, indicating a very strong significance level (denoted

Table 2. Significance of regression coefficients

Model	Parameter coefficient at	p-value
400	ε^2	$<2e^{-16}$ (***)
	ε	$<2e^{-16}$ (***)
	y-intercept	$7.6e^{-9}$ (***)
500	ε^2	$<2e^{-16}$ (***)
	ε	$<2e^{-16}$ (***)
	y-intercept	$7.43e^{-8}$ (***)
600	ε^2	$<2e^{-16}$ (***)
	ε	$<2e^{-16}$ (***)
	y-intercept	0.000108 (***)
Universal	ε^2	$<2e^{-16}$ (***)
	ε	$<2e^{-16}$ (***)
	ρ^2	0.102
	ρ	0.294
	y-intercept	0.957

by ***), suggesting that permittivity has a robust and reliable influence on the model outcomes. In contrast, parameter ρ is not statistically significant in the universal model. The p-values for ρ^2 (0.102), ρ (0.294), and the y-intercept (0.957) are all greater than the standard significance threshold of 0.05. This indicates that density does not have a statistically significant effect in this context. According to Malicki [12], including density and permittivity in the model to estimate the moisture content of porous media improves the model fit. However, in this paper, the statistical analysis showed no significance of density as an independent variable, which suggests that in the case of lightweight porous materials such as cellular concrete, apparent permittivity is a sufficient parameter to assess the moisture content.

AI models

The AI models generated using artificial intelligence such as GPR, SVM, Tree (Decision Tree) and NN (Neural Networks) were also analyzed. Figures 5, 6, 7 present the relationship between the values read using the TDR method and those predicted using AI models for densities of 400, 500 and 600 kg/m³, respectively. The GPR and Tree models consistently demonstrated the highest prediction accuracy, with data points closely aligned with the actual values and minimal scatter. The SVM model showed larger prediction errors and deviations, particularly at higher moisture content levels. Neural networks captured the overall trend well but exhibited greater variability and some local inaccuracies. An increase in concrete density did not significantly affect the performance of the top models, confirming their robustness and generalizability. In summary, GPR and Tree emerged as the most reliable models for predicting the moisture content of aerated concrete based on dielectric permittivity measurements across different densities.

Figures 8, 9, 10 show the estimated values with the fit line for these models. In all three cases, GPR models show the highest prediction accuracy, with data points closely aligned along the ideal prediction line. Neural networks, particularly those with multiple layers, also perform well, especially in predicting moisture content and density. In contrast, Tree models show significant scatter and poorer accuracy, particularly for higher values. Linear SVMs perform moderately but are generally less accurate than GPR and neural networks. Overall, GPR models prove to be the most effective in capturing complex relationships between input data and material properties. These results suggest that advanced nonlinear models like GPR and deep neural networks are well-suited for modeling the behavior of cellular concrete.

The quality assessment of deterministic and AI models

In order to assess the quality of fit of regression and AI models to real data, three standard measures of error and accuracy were used: RMSE, mean absolute error (MAE) and the determination coefficient (R^2). The values of individual

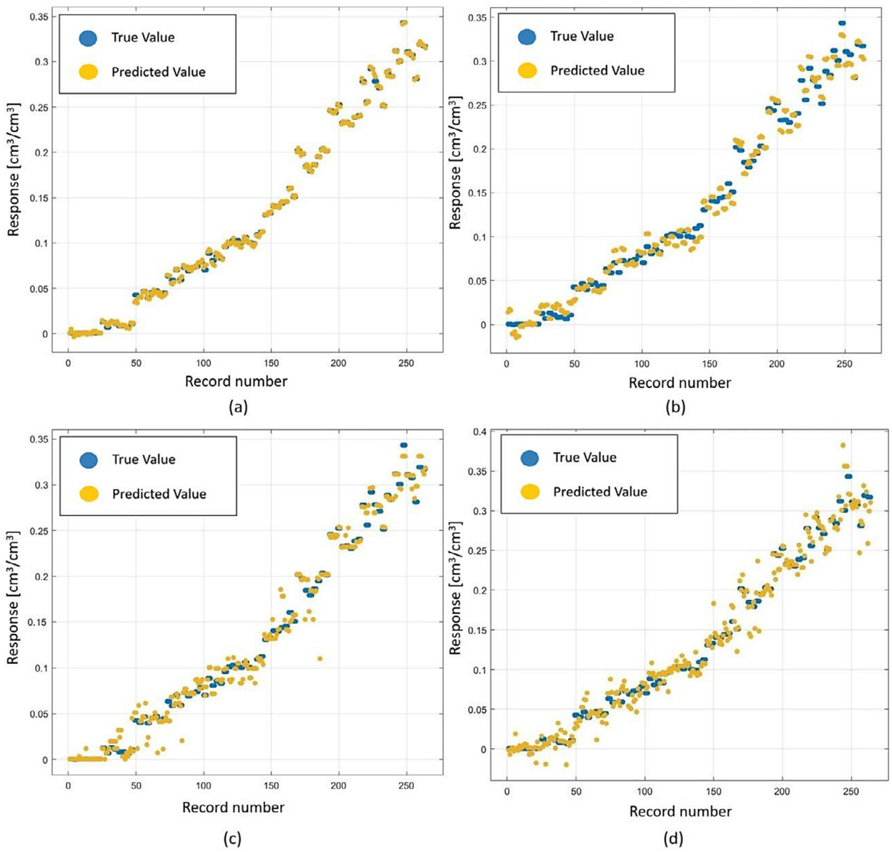

Figure 5. Relationships between values obtained using the TDR method and predicted values for a density of 400 kg/m^3 by AI models: (a) GPR, (b) SVM, (c) Tree, (d) NN

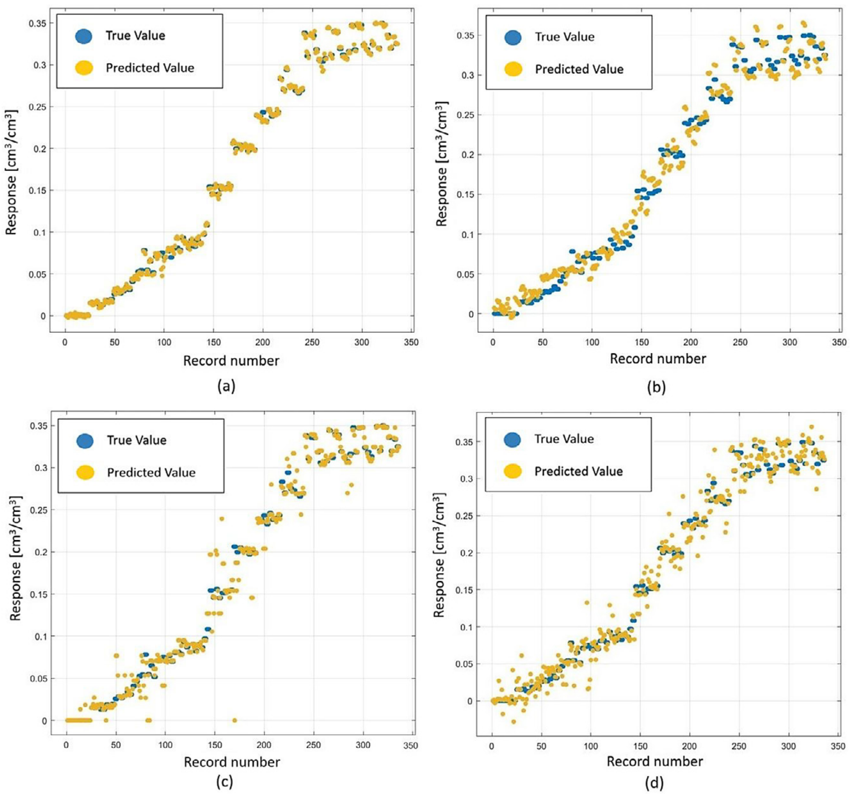

Figure 6. Relationships between values obtained using the TDR method and predicted values for a density of 500 kg/m^3 by AI models: (a) GPR, (b) SVM, (c) Tree, (d) NN

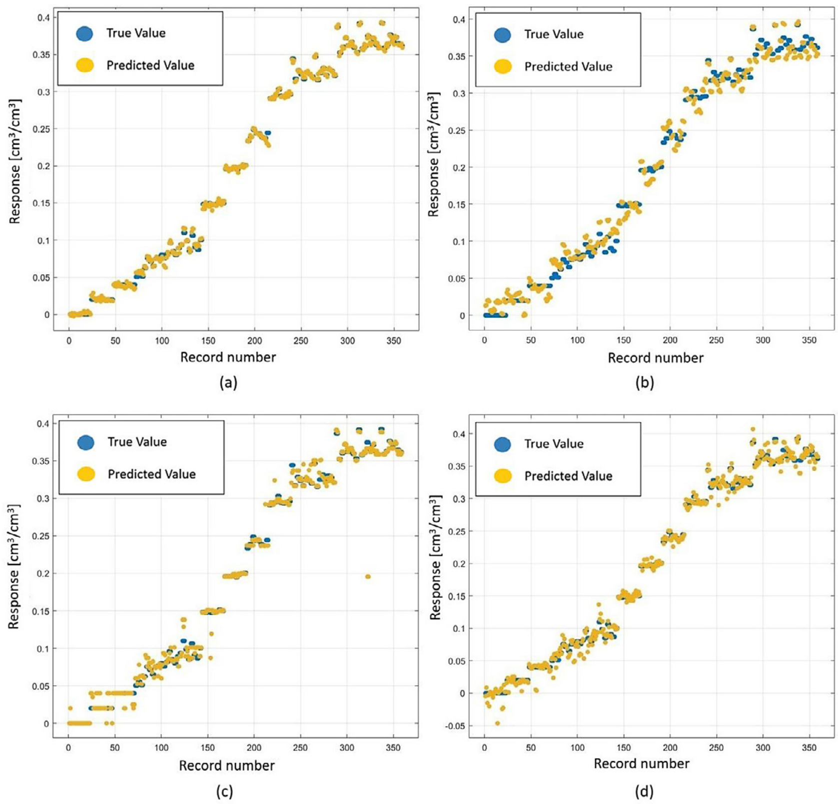

Figure 7. Relationships between values obtained using the TDR method and predicted values for a density of 600 kg/m^3 by AI models: (a) GPR, (b) SVM, (c) Tree, (d) NN

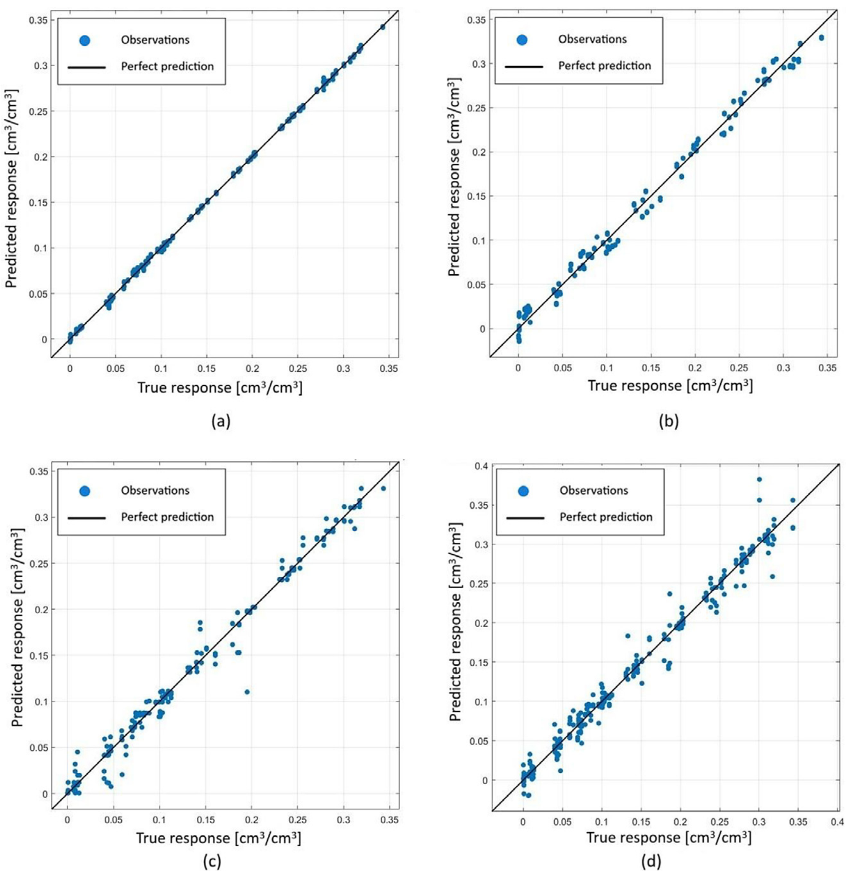

Figure 8. Estimated values with fit line for density 400 kg/m^3 of AI models: (a) GPR, (b) SVM, (c) Tree, (d) NN

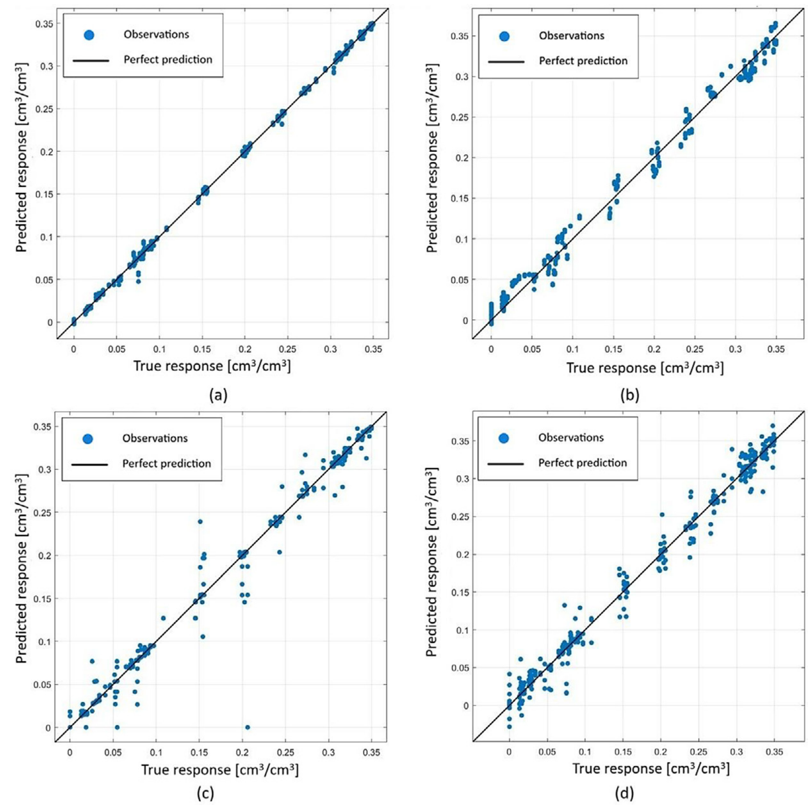

Figure 9. Estimated values with fit line for density 500 kg/m^3 of AI models: (a) GPR, (b) SVM, (c) Tree, (d) NN

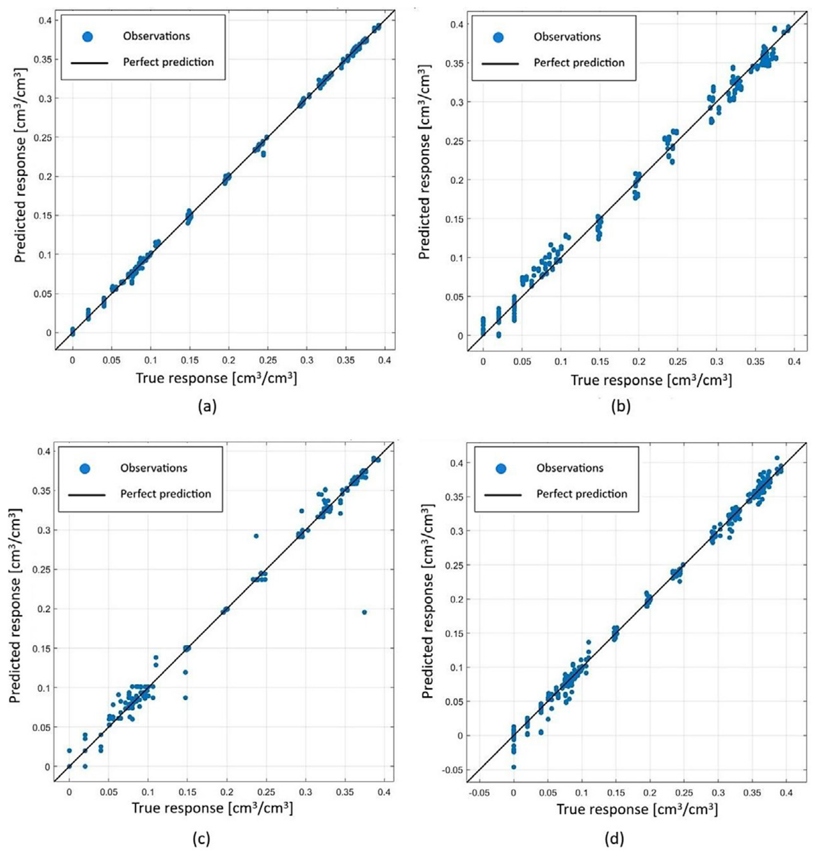

Figure 10. Estimated values with fit line for density 600 kg/m^3 of AI models: (a) GPR, (b) SVM, (c) Tree, (d) NN

Table 3. The fit measures of deterministic and AI models

Model	400			500			600		
	RMSE [cm³/cm³]	MAE [cm³/cm³]	R^2	RMSE [cm³/cm³]	MAE [cm³/cm³]	R^2	RMSE [cm³/cm³]	MAE [cm³/cm³]	R^2
$\theta(\varepsilon)$	0.0244	0.0203	0.9433	0.0332	0.0255	0.9280	0.0339	0.0265	0.9363
GPR	0.0021	0.0015	0.9996	0.0037	0.0023	0.9991	0.0032	0.0021	0.9995
SVM	0.0095	0.0082	0.9914	0.0127	0.0109	0.9895	0.0123	0.0100	0.9917
Tree	0.0115	0.0067	0.9875	0.0189	0.0081	0.9771	0.0161	0.0056	0.9857
NN	0.0140	0.0087	0.9816	0.0148	0.0097	0.9858	0.0085	0.0056	0.9960

Table 4. The fit measures of universal regression model

Model	RMSE [cm³/cm³]	MAE [cm³/cm³]	R^2
$\theta(\varepsilon,\rho)$	0.0321	0.0253	0.9340

measures of model fit are presented in Tables 3 and 4. The $\theta(\varepsilon)$ model performed best for a density of 400 kg/m³, where it had the smallest errors and the highest coefficient of determination value of 0.9433. This means that in this case, the model predicted moisture very well. However, for higher densities (500 and 600 kg/m³) the quality of prediction deteriorated significantly – RMSE increased and R^2 dropped to about 0.9280 - 0.9363. The universal $\theta(\varepsilon,\rho)$ model achieved a better fit than the $\theta(\varepsilon)$ model for densities of 500 and 600, but not for 400. The overall R^2 coefficient was 0.9340, which means better stability, but still not equal to the AI models.

The AI models were much more accurate in predicting moisture. The GPR model was found to be the best model. It achieved almost perfect fit with R^2 values close to 0.9999. Its errors were extremely low (RMSE at the level of 0.0021 - 0.0032 cm³/cm³), which means that they allowed the error values to be reduced almost tenfold compared to deterministic models. This type of model almost perfectly predicts moisture based on permittivity. The SVM and NN models also achieved very good results. The R^2 coefficients of determination at the level of 0.9914–0.9960 mean that these models predict moisture very well. The RMSE are smaller than those in regression models by a factor of 5. This means that they are a much better alternative to classical regression. The weakest of the AI models turned out to be the Tree model. Although the model achieved better results than deterministic regression models, it was significantly worse than GPR, SVM and NN. The RMSE and R^2 values of this model suggest that it makes larger errors and has lower accuracy. It is still better than regression models, but it is not the best choice.

In summary, the AI models are the most accurate at estimating the moisture content of cellular concrete – especially GPR, which fits the data almost perfectly. If the goal is maximum accuracy, this is the model to use. Regression models have limited effectiveness – although they give decent results, their accuracy is much lower than that of AI models. Including density (ρ) improves the prediction but does not solve all problems – even the universal model $\theta(\varepsilon,\rho)$ still has larger errors than AI models. If model simplicity is key, regression can be used, but AI provides much higher precision. If accuracy is a priority, GPR should be chosen. If simplicity and ease of interpretation are more important, the regression model $\theta(\varepsilon,\rho)$ can be a compromise, but one has to reckon with a larger error.

When comparing the obtained measurement errors with those indicated by other authors, it should be noted that they are comparable to or smaller than those obtained by other authors, even when it comes to deterministic models. Many basic articles on this method emphasized that these RMSE errors range from 0.01 to 0.066 cm³/cm³ depending on the characteristics of the tested material (Ju et al. [41]), Roth et al. [42] established the RMSE of their research at the level of 0.08–0.037 cm³/cm³, Malicki [12] with two-parameter model, taking into account the effect of density, estimated the RMSE for soils at the level of 0.03 cm³/cm³. Byun et al. [43] at the level of 0.04–0.05 cm³/cm³. More recent studies have also explored the use of predictive models for estimating moisture content. Paśnikowska et al. reported a RMSE of 0.07 cm^3/cm^3 [44] and Kojima et al. [45] between 0.066 cm^3/cm^3 and 0.220 cm^3/cm^3 using conventional models. However, it is important to note that these experiments were conducted using a non-invasive probe on silicate material. Similarly, the authors themselves in their previous studies obtained: RMSE values for TDR sensors of various types at the level of 0.024 to 0.032 cm^3/cm^3 [46] or 0.013 cm^3/cm^3 [47].

The use of machine learning methods improves the accuracy of moisture estimation using the TDR method several times, which was confirmed by Wan et al. [48], who observed similar capabilities of machine learning algorithms in relation to the studies using TDR equipment or Méndez-Patiño et al. [49], who observed a fourfold decrease in MSE errors compared to data analyzed using classical methods. Similar relationships were observed by the authors of this study in their works using artificial intelligence, e.g. Mikušová et al. [50] in which the use of SVM algorithms allowed for a reduction of RMSE errors by approximately a factor of 3, while GPR algorithms reduced moisture estimation errors about ten-fold. In another research Paśnikowska et al. [44] confirmed that machine learning models enabled prediction of silicate brick moisture with RMSE errors equal to 0.015 cm^3/cm^3 (GPR) and 0.025 cm^3/cm^3 (SVM).

CONCLUSIONS

The most important conclusions resulting from the conducted research include the following observations:

1. Taking into account the effect of density on moisture estimation in the case of aerated concrete does not significantly improve the quality of the estimation. This is confirmed by the significance levels of the p-value of the estimators responsible these parameters of the model. Unlike the Malicki model, which was developed for soils, in the case of aerated concrete, this is due to the small variation in the value of this parameter.
2. The use of machine learning methods improves the quality of moisture prediction of aerated concrete several times.
3. Of all the machine learning algorithms, the most effective are GPR models, which are characterized by approximately ten-fold smaller RMSE error values comparing to the deterministic methods.
4. SVM and tree algorithms are characterized by worse predictive capabilities than GPR algorithms, but they reduce RMSE errors by more than two-fold compared to traditional deterministic models based on single- and two-factor regression.

Acknowledgements

This work was financially supported within the authors' research of particular scientific units under subvention for a Scientific Disciplines program.

REFERENCES

1. Suchorab Z., Barnat-Hunek D., Smarzewski P., Pavlík Z., Černý R. Free of volatile organic compounds protection against moisture in building materials. Ecol Chem Eng S. 2014;21(3):401–411. https://doi.org/10.2478/eces-2014-0029
2. Elsadany S.M., Fayed M.N., Sorour T.M., Anwar A.M., Nasr N.E. Response reduction factor for structures with significant irregularities on different soil stratum. Civil Eng. J. 2024;10(3):757–78. https://doi.org/10.28991/CEJ-2024-010-03-07
3. Koestoer R.H., Ligayanti T., Kartohardjono S., Susanto H. Down-streaming small-scale green ammonia to nitrogen-phosphorus fertilizer tablets for rural communities. Emerg Sci J. 2024;8(2):625–43. https://doi.org/10.28991/ESJ-2024-08-02-016
4. Wang Y., Zhao J., Zhang Z., Li Y. A comprehensive evaluation method for sponge city construction based on the analytic network process and the entropy weight method. Sustainability. 2020;12(19):7855. https://doi.org/10.3390/su12197855
5. PN-EN ISO 12571:2013-12. Cieplno-wilgotnościowe właściwości użytkowe materiałów i wyrobów budowlanych – Określanie właściwości sorpcyjnych.
6. O'Connor K.M., Dowding C.H. GeoMeasurements by pulsing TDR cables and probes. CRC Press; 2020. https://doi.org/10.1201/9781003067726
7. Stefańczyk B. editor. Budownictwo ogólne. Tom 1. Materiały i wyroby budowlane. Warszawa: Arkady; 2007.
8. Wiendl A., Yan G., Scheuermann A., Fillibeck J., Cudmani R. An experimental study on applying spatial TDR to determine bentonite suspension penetration. Measurement. 2025;242(Pt E):116310. https://doi.org/10.1016/j.measurement.2024.116310
9. Nimer H., Ismail R., Rawashdeh A., Al-Mattarneh H., Khodier M., Hatamleh R., Abuaddous M.

Artificial intelligence using FFNN models for computing soil complex permittivity and diesel pollution content. Civil Eng. J. 2024;10(9):3053–69. https://doi.org/10.28991/CEJ-2024-010-09-018.

10. Černý R. Time-domain reflectometry method and its application for measuring moisture content in porous materials: A review. Measurement. 2009;42(3):329–336. https://doi.org/10.1016/j.measurement.2008.08.011
11. Topp G.C., Davis J.L., Annan A.P. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. Water Resour Res. 1980;16(3):574–582. https://doi.org/10.1029/WR016i003p00574
12. Malicki M.A., Plagge R., Roth C.H. Improving the calibration of dielectric TDR soil moisture determination taking into account the solid soil. Eur J Soil Sci. 1996;47(3):357–366. https://doi.org/10.1111/j.1365-2389.1996.tb01409.x
13. Skierucha W., Wilczek A., Alokhina O. Calibration of a TDR probe for low soil water content measurements. Sens Actuators A Phys. 2008;147(2):544–552. https://doi.org/10.1016/j.sna.2008.06.015
14. Klyve D. Regression to the mean. Central Washington University, Ellensburg; 2019.
15. Olive D.J. Multivariate models. In: Linear regression. Springer International Publishing; 2017. 213–242. https://doi.org/10.1007/978-3-319-55252-1_6
16. Frenzel D., Blaschke O., Franzen C., Brand F., Haas F., Troi A., Drese KS. Quantification of moisture in masonry via AI-evaluated broadband radar reflectometry. Heritage. 2023;6(7):5030–50. https://doi.org/10.3390/heritage6070266
17. Dafico L.C.M., Barreira E., Almeida R.M.S.F., Vicente R. Machine learning models applied to moisture assessment in building materials. Constr Build Mater. 2023;405:133330. https://doi.org/10.1016/j.conbuildmat.2023.133330
18. Russell S., Norvig P. Artificial intelligence: a modern approach. 4th ed. Pearson; 2021.
19. Poole D., Mackworth A. Artificial intelligence: foundations of computational agents. 2nd ed. Cambridge University Press, Cambridge, UK; 2017.
20. Zaki A.M., Zayed M.E., Bargal M.H.S., Saif A.G.H., Chen H., Rehman S., Alhems L.M., Nour El-deen E.S.H. Environmental and energy performance analyses of HVAC systems in office buildings using boosted ensembled regression trees: machine learning strategy for energy saving of air conditioning and lighting facilities. Process Saf Environ Prot. 2025;198:107214. https://doi.org/10.1016/j.psep.2025.107214
21. James G., Witten D., Hastie T., Tibshirani R. An introduction to statistical learning. 2nd ed. Springer, New York, USA; 2021. https://doi.org/10.1007/978-1-0716-1418-1
22. Souza V.F.C., Cicalese F., Laber E.S., Molinaro M. Decision trees with short explainable rules. Theor Comput Sci. 2025;1047:115344. https://doi.org/10.1016/j.tcs.2025.115344
23. Altork Y. Comparative analysis of machine learning models for wind speed forecasting: support vector machines, fine tree, and linear regression approaches. Int J Thermofluids. 2025;27:101217. https://doi.org/10.1016/j.ijft.2025.101217
24. Souza e Silva A., Gonçalves A.C., de Abreu W.V., da Silva A.C., Martinez A.S. Power distribution reconstruction from in-core detector measurements using gaussian process regression. Ann Nucl Energy. 2025;202:111581. https://doi.org/10.1016/j.anucene.2025.111581
25. Rasmussen C.E., Williams C.K.I. Gaussian processes for machine learning. MIT Press, Cambridge, MA, USA; 2006.
26. Chen S., Kong D., Chen D. Gaussian process regression for enhancing the positional accuracy of collaborative robot. Measurement. 2025;256(Pt A):118142. https://doi.org/10.1016/j.measurement.2025.118142
27. Zhou Y., Chen W., Zou H. A link prediction algorithm based on support vector machine. Physica A. 2025;673:130674. https://doi.org/10.1016/j.physa.2025.130674
28. Cristianini N., Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge, UK; 2000.
29. Zhang K., Hao W., Yu X., Shao T. An interpretable image classification model combining a fuzzy neural network with a variational autoencoder inspired by the human brain. Inf Sci. 2024;661:119885. https://doi.org/10.1016/j.ins.2023.119885
30. Gdoura A., Bernhard S. Regression and classification of Windkessel parameters from non-invasive cardiovascular quantities using a fully connected neural network. Inform Med Unlocked. 2025;53:101614. https://doi.org/10.1016/j.imu.2025.101614
31. Nielsen M.A. Neural networks and deep learning: a textbook. Determination Press, USA; 2015.
32. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, Cambridge, MA, USA; 2016.
33. Rawlings J.O., Pantula S.G., Dickey D.A. Applied regression analysis: a research tool. Springer, New York, USA; 1998.
34. Rencher A.C. Methods of multivariate analysis. John Wiley & Sons, New York, USA; 2002.
35. Keith T.Z. Multiple regression and beyond. Journal of Educational Measurement, New York, USA; 2015.
36. Ramsey J.B. Tests for specification errors in classical

linear least-squares regression analysis. Blackwell Publishing for the Royal Statistical Society, Michigan, USA; 1969.

37. MATLAB – MathWorks. [Internet]. Natick, MA: The MathWorks, Inc.; [cited 2025 Jun 20]. Available from: https://www.mathworks.com/products/matlab.html
38. Chai T., Draxler R.R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geosci Model Dev. 2014;7(3):1247–1257. https://doi.org/10.5194/gmd-7-1247-2014
39. James G. An introduction to statistical learning. Springer International Publishing, New York, USA; 2023.
40. Weisberg S. Applied linear regression. John Wiley & Sons, New Jersey, USA; 2014.
41. Ju Z., Liu X., Ren T., Hu C. Measuring soil water content with time domain reflectometry: An improved calibration considering soil bulk density. Soil Sci. 2010;175(10):469–473. https://doi.org/10.1097/SS.0b013e3181f55aa3
42. Roth K., Schulin R., Flühler H., Attinger W. Calibration of time domain reflectometry for water content measurement using a composite dielectric approach. Water Resour Res. 1990;26(10):2267–2273. https://doi.org/10.1029/WR026i010p02267
43. Byun Y.H., Hong W.T., Yoon H.K. Characterization of cementation factor of unconsolidated granular materials through time domain reflectometry with variable saturated conditions. Materials. 2019;12(8):1340. https://doi.org/10.3390/ma12081340
44. Paśnikowska-Łukaszuk M., Szulżyk-Cieplak J., Wlazło M., Zubrzycki J., Łazuka E., Urzędowski A., Suchorab Z. The use of 3D printing filaments to build moisture sensors in porous materials. Materials. 2025;18(1):115. https://doi.org/10.3390/ma18010115
45. Kojima Y., Okumura K., Aoki S., Noborio K., Kamiya K., Horton R. A four-parameter-based thermo-TDR approach to estimate water and NAPL contents of soil liquid. Geoderma. 2023;429:116263. https://doi.org/10.1016/j.geoderma.2022.116263
46. Suchorab Z., Majerek D., Kočí V., Černý R. Time domain reflectometry flat sensor for non-invasive monitoring of moisture changes in building materials. Measurement. 2020;165:108091. https://doi.org/10.1016/j.measurement.2020.108091
47. Suchorab Z., Widomski M.K., Łagód G., Barnat-Hunek D., Majerek D. A noninvasive TDR sensor to measure the moisture content of rigid porous materials. Sensors. 2018;18(11):3935. https://doi.org/10.3390/s18113935
48. Wan H., Hongwei Q., Shang S. Estimating soil water and salt contents from field measurements with time domain reflectometry using machine learning algorithms. Agric Water Manag. 2023;285:108364. https://doi.org/10.1016/j.agwat.2023.108364
49. Méndez-Patiño A., Gutierrez-Gnecchi J.A., Reyes-Archundia E., Tellez Anguiano A.D.C. Evaluation of feedforward artificial neural networks (ANN) to adjust soil moisture estimates derived from time domain reflectometry (TDR) measurements using soil temperature and gravimetric data. Int J New Technol Res. 2017;3(12).
50. Mikušová D., Suchorab Z., Paśnikowska-Łukaszuk M., Zaburko J., Trnik A. Applying the machine learning method to improve calibration quality of time domain reflectometry measuring technique. Adv Sci Technol Res J. 2024;18(3):270–279. https://doi.org/10.12913/22998624/187007