WARSAW UNIVERSITY OF TECHNOLOGY Index 351733
FACULTY OF CIVIL ENGINEERING
COMMITTEE FOR CIVIL AND WATER ENGINEERING
POLISH ACADEMY OF SCIENCES ISSN 1230-2945

DOI: 10.24425/ace.2026.157492

ARCHIVES OF CIVIL ENGINEERING

Vol. LXXII ISSUE 1 2026

pp. 565 –577



**Research paper**

# Could Machine Learning Algorithms affect the efficient management of anthropogenic materials?

**Justyna Dzięcioł[1], Roman Tracz[2], Jan Kowalski[3]**

**Abstract:** The construction industry is increasingly exploring alternatives to natural aggregates, driven by sustainability concerns and landfill waste reduction. Blast furnace slag, a byproduct of steel manufacturing, exemplifies this shift, serving as a substitute aggregate or concrete additive. This transition supports the circular economy principle, where yesterday's waste transforms into today's resources. Key to this practice is the precise determination of material parameters, which vary depending on their origin. Among these, the filtration coefficient is critical, affecting the performance of anthropogenic aggregates in construction and infrastructure. It indicates how well materials transmit water, a factor vital for structural integrity. Machine Learning (ML) presents a promising tool for estimating such parameters efficiently. This paper explores various ML techniques for predicting the filtration coefficient, comparing their effectiveness and examining the impact of the physical properties of aggregates on model accuracy. Through this approach, the paper aims to identify the most suitable methods for parameter estimation, which could enhance the durability and stability of constructions that utilize recycled materials. This research not only contributes to the field of civil engineering but also advances sustainable practices within the industry.

**Keywords:** coefficient of permeability, machine learning algorithms, blast furnace slag, sustainability, anthropogenic materials

[1]PhD., Eng., Warsaw University of Life Sciences, Institute of Civil Engineering, Nowoursynowska 166, 02-787 Warsaw, Poland, e-mail: justyna_dzieciol@sggw.edu.pl, ORCID: 0000-0002-2436-9748
[2]Prof., PhD., Eng., Warsaw University of Life S?ciences, Institute of Civil Engineering, Nowoursynowska 166, 02-787 Warsaw, Poland, e-mail: roman_tracz@sggw.edu.pl, ORCID: 0000-0001-6654-9870
[3]PhD., Eng., Warsaw University of Life Sciences, Institute of Civil Engineering, Nowoursynowska 166, 02-787 Warsaw, Poland, e-mail: jan_kowalski@sggw.edu.pl, ORCID: 0000-0002-8944-311X

# 1. Introduction

The development of the modern construction industry is increasingly driven by mitigating environmental impact and introducing innovative practices to ensure greater sustainability. This evolution has become a catalyst for changes in material procurement, notably the use of industrial by-products as alternative aggregates. About 85% of industrial waste ends up in landfills [1, 2]. Currently, the world faces a shortage of available landfill space and threats to environmental balance [3]. Therefore, it is crucial to efficiently utilize waste and by-products from industry. Among these is blast furnace slag, a by-product of iron and steel production, which is used as aggregate for road construction, soil stabilization, base layers, and flexible and semi-rigid pavements [4–6]. The adsorption properties of slag make it useful for phytoremediation applications and the removal of heavy metals from soil [7]. Many researchers have analyzed the use of steel slag as aggregate in designing asphalt concrete for road construction [8]. Effective recycling of construction waste requires assessing the permeability of aggregate, which is the ability to transmit fluids under pressure. The influx of waste from industry and building demolitions makes it impossible to determine these properties solely through laboratory or field methods. These materials are used in earthworks, including road embankments, where information on permeability is essential. Soil permeability depends on its grain size, structure, texture, and porosity. The coefficient of permeability, $k$ is used in predicting water flow through a porous medium [9]. The implementation of machine learning techniques opens new possibilities and increases the chances of precisely adapting slag properties to diverse construction requirements. Integrating technology in the assessment and utilization of slag not only enhances its application in civil engineering projects but also supports the industry's commitment to sustainability.

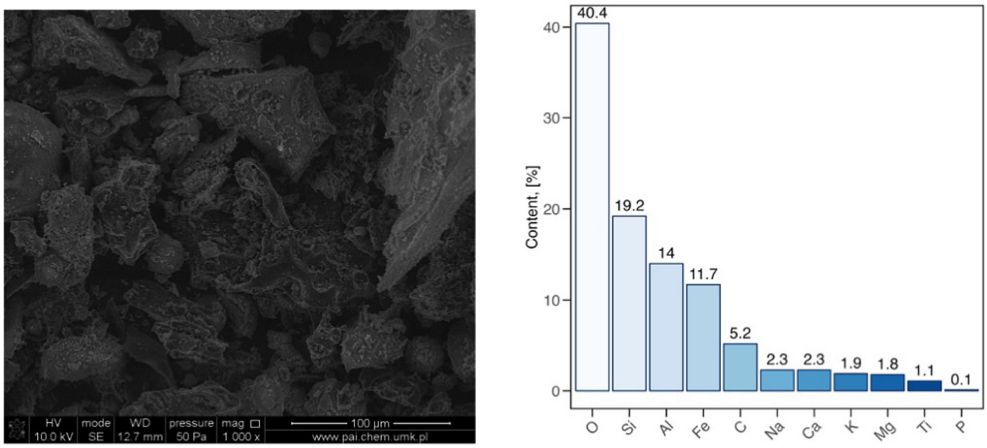
This study aims to present the predictive efficiency of machine learning algorithms applied to estimate the coefficient of permeability for anthropogenic aggregate – blast furnace slag. Considering the rapid development of machine learning techniques and their increasing selectivity regarding analyzed data, it is necessary to consider different types of algorithms to achieve the best predictive effect. Precise estimation of critical parameters used in construction, based on historical data analyzed by machine learning algorithms, can significantly influence the future development of this sector, reducing project design time and costs, thereby directly affecting the progress and margins of the entire investment.

# 2. Material and methodology

## 2.1. Material

Blast furnace slag, a byproduct of iron production in blast furnaces, involves the reduction of iron ore into molten iron while producing slag that floats on top. This slag can be cooled and processed in various ways to produce different types of materials useful in construction and other industries. Blast furnace slag can vary in appearance from a rough, rocky texture to a granular, sandy form, depending on the cooling process. The material typically has a lower

density than natural aggregates, making it lighter and beneficial in specific applications. It is relatively hard and stable, contributing to its utility as a construction material. Depending on the processing method, the particle size of blast furnace slag can range from coarse aggregates to very fine powders [10–12]. The chemical composition of the examined blast furnace slag includes silicon, aluminum, iron, and carbon, along with smaller amounts of other elements, such as magnesium (Figure 1b). The high content of silicon, aluminum, and calcium indicates that the blast furnace slag may be highly reactive, rendering it a suitable material for use as an additive in cement. In contrast, the presence of metals such as iron and titanium can influence the chemical stability and resistance of the slag to environmental conditions. Figure (1a) depicts an image of blast furnace slag obtained through scanning electron microscopy (SEM). The SEM image illustrates a complex microstructure, characterized by a porous and irregular structure. The images were captured at 1000x magnification, revealing particles of varying shapes, sizes, and jagged edges. The surface texture of these particles appears rough and complex, with numerous small protrusions and indentations.




(a) (b)

Fig. 1. The SEM photo of the test material (a), the chemical composition of the test material (b)

Table 1 presents the ranges of analyzed parameters from the database that provides input for estimating models using machine learning. The gradient ranges from 0.08 to 1.02, while particle sizes vary across percentiles, with the smallest particles ($d_5$) ranging from 0.60 mm to 0.70 mm and the largest particles ($d_{90}$) ranging from 16.00 mm to 19.50 mm. The volumetric density is relatively low, ranging from 0.80 g/cm$^3$ to 0.83 g/cm$^3$, indicating the slag's lightweight nature. Porosity values range from 0.65 to 0.66, and the index porosity varies from 1.83 to 1.93, reflecting a high degree of void spaces within the material. The coefficient of permeability ranges from 0.02569 m/s to 0.05120 m/s, indicating the possibility of using this material for applications requiring drainage and fluid movement.

Table 1. Summary of data ranges of analyzed parameters

| Description | Minimum | Maximum |
|---|---|---|
| Gradient ($i$), [–] | 0.08 | 1.02 |
| Particle size ($d_5$), [mm] | 0.60 | 0.70 |
| Particle size ($d_{10}$), [mm] | 0.98 | 1.00 |
| Particle size ($d_{17}$), [mm] | 2.00 | 7.30 |
| Particle size ($d_{20}$), [mm] | 3.20 | 8.50 |
| Particle size ($d_{30}$), [mm] | 9.30 | 9.90 |
| Particle size ($d_{50}$), [mm] | 11.10 | 14.00 |
| Particle size ($d_{60}$), [mm] | 12.00 | 14.00 |
| Particle size ($d_{90}$), [mm] | 16.00 | 19.50 |
| Volumetric density ( $\rho_d$), [g/cm3] | 0.80 | 0.83 |
| Porosity ($n$), [–] | 0.65 | 0.66 |
| Index porosity ($e$), [–] | 1.83 | 1.93 |
| Homogeneity index ($Cu$), [–] | 12.00 | 14.29 |
| Grain size curvature index ($Cc$), [–] | 7.14 | 7.26 |
| Coefficient of permeability ($i$), [m/s] | 0.02569 | 0.05120 |

## 2.2. Methodology

The permeability characteristics of BFS were tested using the constant head method. This technique is known for its simplicity and consistent testing conditions, and it is considered one of the most reliable methods for measuring permeability in non-cohesive soils [13]. This research employed a database comprising 192 original experiments, ensuring a highly representative sample that enhances the reliability and validity of the findings.

Prior to undertaking further analysis utilizing machine learning, a comprehensive data preparation process was undertaken to guarantee the efficacy and precision of the model. The data preparation process commenced with the collection and filtering of data to address any instances of missing values, which served as the foundation for constructing a machine learning model. It was of utmost importance to verify the data to eliminate any irrelevant or outdated information that could have had a detrimental effect on the model's performance. This entailed identifying errors within the database, standardizing the data collection and validation process, and ensuring data accuracy in accordance with predefined criteria. The subsequent steps included identifying missing values, ensuring the correct format of text and numerical data, and removing duplicate entries. Once the data was identified, standardized, and cleaned of unwanted and duplicate information, it was integrated into the database for further analysis. Any data record gaps were either addressed or removed if they could not be filled, ensuring a robust dataset for machine learning analysis.

Machine learning is a subset of artificial intelligence that involves developing algorithms that enable computers to learn from data and make decisions or estimates based on that data. The field has seen significant progress over the past few decades, driven by increases in computing power, advances in technology, and the availability of large datasets and innovative algorithmic techniques. Machine learning algorithms are broadly categorized into supervised, unsupervised, and reinforcement learning, each serving different purposes and applications.

Among the most well-known and widely used machine learning algorithms for estimation are the Gradient Boosting Machine (GBM), decision tree, random forest, neural network, and linear regression. These algorithms underpin many practical applications across various industries and are used in this study to estimate the coefficient of permeability, a crucial geotechnical parameter in the design of earth structures and road infrastructure. The following are brief characteristics of the algorithms used in the prediction process in this article.

The Gradient Boosting Machine is an ensemble learning technique that builds a series of decision trees sequentially, each correcting the errors of its predecessor by focusing on the residuals of the previous trees. Utilizing a gradient descent algorithm to minimize the loss function, GBM is effective for both classification and regression tasks. Introduced by Jerome H. Friedman in 2001 [14], it is celebrated for its high predictive accuracy and robustness against overfitting, particularly when combined with regularization and tree pruning techniques. However, GBM can be computationally demanding and requires meticulous parameter tuning, including the number of trees, tree depth, learning rate, and minimum samples required to split a node.

A Decision Tree is a non-parametric supervised learning algorithm used for classification and regression tasks. It divides the data into subsets based on the value of input features, creating a tree structure where each node represents a feature, each branch a decision rule, and each leaf an outcome [15].

Random Forest is an ensemble learning method that builds multiple decision trees during training and outputs the mode or mean prediction of these trees. Each tree is constructed from a random subset of the training data and features, enhancing the model's generalizability and robustness to overfitting. Introduced by Leo Breiman in 2001 [16], Random Forests are versatile and capable of handling large datasets with greater accuracy than individual decision trees. However, their complexity and longer training times, along with reduced interpretability compared to single decision trees, could be notable drawbacks. Key hyperparameters include the number of trees in the forest, maximum tree depth, minimum samples required to split a node, minimum samples required at a leaf node, and the number of features to consider when looking for the best split [17].

Neural networks, inspired by the human brain, excel at recognizing patterns and modeling complex relationships within data. There are numerous methods for configuring neural network nodes. The most straightforward approach is the feed-forward network, wherein signals progress in a single direction without looping back. The foundational model of such a network is the perceptron, established in the 1950s by Professor Frank [18]. These networks are adept at identifying non-linear dependencies and are particularly useful in fields such as image and speech recognition. However, they require significant amounts of data and computational power.

Their effectiveness largely depends on careful hyperparameter tuning and network architecture, such as the number of hidden layers, the number of neurons in each layer, learning rate, batch size, activation functions, and regularization coefficients. Techniques such as dropout and regularization are employed to prevent overfitting.

Linear Regression is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables. The method dates back to the early 19th century, with significant contributions from Adrien-Marie Legendre [19]. The model assumes a linear relationship, expressed by the equation (2.1):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon \tag{2.1}$$

where: $\beta$ – represents the coefficients, $\in$ – the error term.

Linear regression is easy to implement, interpret, and computationally efficient. However, it is limited to linear relationships and can be sensitive to outliers and multicollinearity among the predictors. Algorithmic model prediction and comprehensive analysis of estimation results and interpretation were performed using R Studio software.

## 3. Results

In the research discussed in the article, cross-validation was used as the sample division method, recognized as a key tool in machine learning for assessing a model's ability to generalize to new data. The sample for each fold was randomly selected. This validation uses a parameter called "$i$", which specifies the number of splits into which the dataset is divided [20]. In practice, cross-validation involves using a limited data set to evaluate the model's efficiency in predictions on previously unseen data. This method is valued for its simplicity and clarity, which helps in creating more reliable models through iterative learning and testing phases. The effectiveness of the model is assessed by averaging the results from each iteration of $i$-fold cross-validation [21]. These results are further verified through error analysis, allowing for a more accurate assessment of individual models. The method of calculating estimation errors and verifying the convergence of models with empirical data is presented below:

– Mean Square Error (*MSE*) (3.1): Represents the average squared difference between the original and predicted values in the dataset. It measures the variance of the residuals.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.1}$$

where: $y_i$ – actual value, $\hat{y}_i$ – predicted value, $n$ – number of observations

– Root Mean Square Error (*RMSE*) (3.2): The square root of the Mean Square Error. It measures the standard deviation of the residuals.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.2}$$

where: $y_i$ – actual value, $\hat{y}_i$ – predicted value, $n$ – number of observations.

– Mean Absolute Error (*MAE*) (3.3): Represents the average absolute difference between the actual and predicted values in the dataset. It measures the average of the residuals.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}| \tag{3.3}$$

where: $y_i$ – actual value, $\hat{y}_i$ – predicted value, $n$ – number of observations.

– Coefficient of Determination ($R^2$) (3.4): Represents the proportion of the variance in the dependent variable that is explained by the linear regression model. It is a unitless measure that ranges from 0 to 1, where 1 indicates a perfect fit of the model to the data and 0 indicates no fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{3.4}$$

where: $y_i$ – actual value, $\hat{y}_i$ – predicted value, $\overline{y}$ – mean of actual values, $n$ – number of observations.

Using these metrics together, one can get a holistic view of a model's predictive accuracy, reliability, and overall performance in capturing patterns in the data. *MSE* and *RMSE* are useful for understanding the variability and magnitude of errors, especially for larger discrepancies. *MAE* provides a robust and interpretable measure of the average error magnitude, which is less affected by outliers. $R^2$ helps in assessing the explanatory power of a model and comparing different models. This multi-faceted approach ensures that the evaluation covers various aspects of model performance, leading to more informed decisions on model selection and improvement. The results of prediction using the GBM, Decision Tree, Random Forest, Neural Network, and Linear Regression algorithms are presented in Table 2.

Table 2. Prediction results using machine learning algorithms

| **Model** | ***MSE*** | ***RMSE*** | ***MAE*** | $\boldsymbol{R^2}$ |
|---|---|---|---|---|
| GBM | 1.76E–06 | 0.0013 | 0.0010 | 0.96 |
| Decision Tree | 3.62E–06 | 0.0019 | 0.0016 | 0.91 |
| Random Forest | 1.32E–05 | 0.0036 | 0.0028 | 0.90 |
| Neural Network | 6.08E–06 | 0.0025 | 0.0019 | 0.85 |
| Linear Regression | 7.14E–06 | 0.0027 | 0.0021 | 0.83 |

The algorithm with the best predictive performance was the model obtained for the GBM algorithm, which obtained an $R^2$ of 0.96. The model that performed the worst with the estimation of the filter coefficient was the model that was created using linear regression here the coefficient of determination was 0.83.

Figure 2 presents scatter plots showing the relationship between observed values and predicted values for five different machine learning models: Gradient Boosting Machine (GBM), Neural Network, Random Forest, Linear Regression, and Decision Tree.

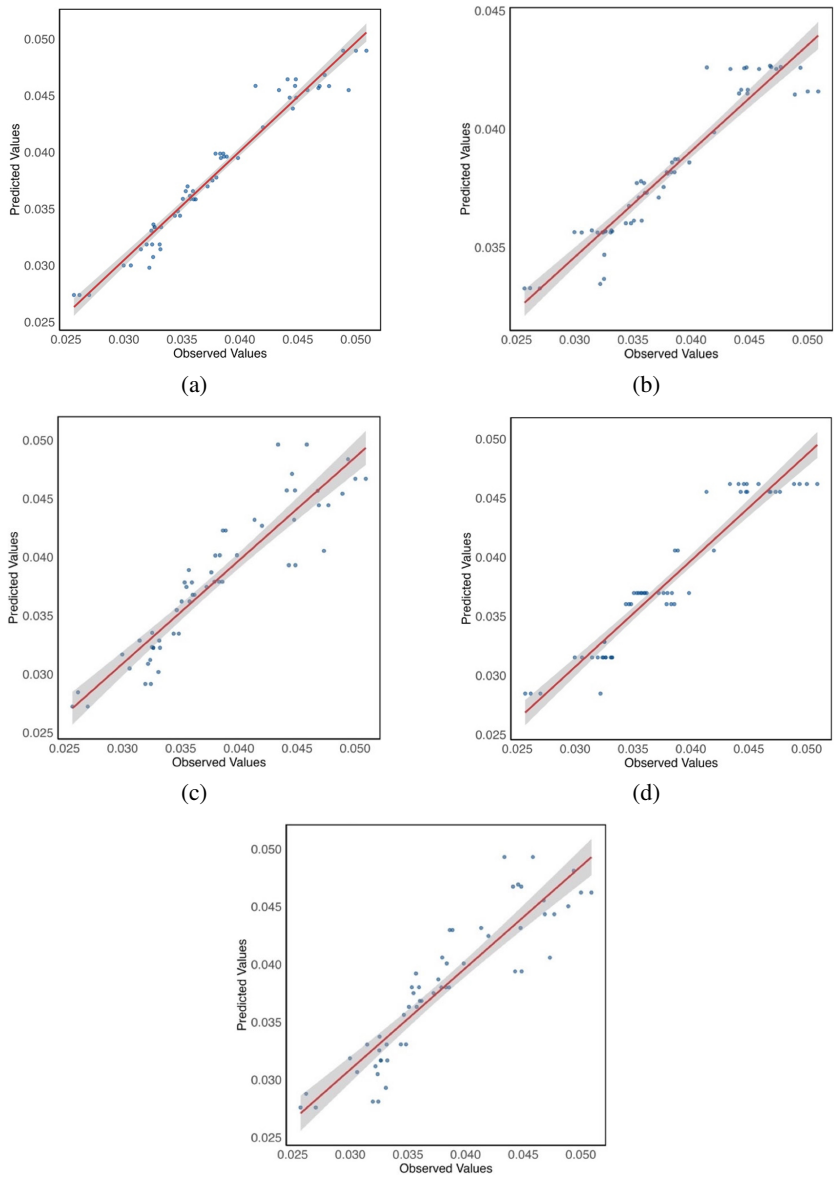
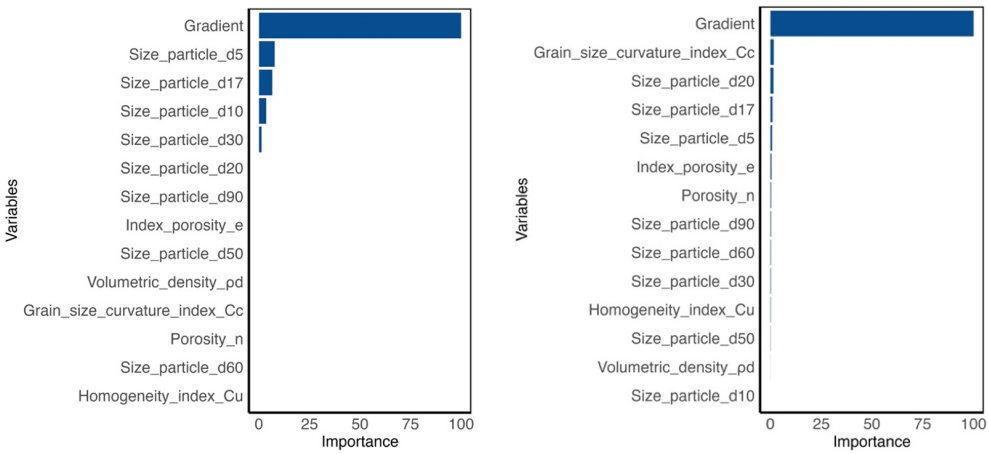



Fig. 2. Graphical presentation of prediction results versus observation results for individual algorithms: (a) GBM; (b) Random Forest; (c) Neural Network; (d) Decision Tree; (e) Linear Regression

Each graph includes a regression line (red) and a shaded confidence interval to help understand the accuracy and reliability of the predictions made by each model. The graph for GBM shows a strong linear relationship between observed and predicted values, as indicated by the closely matched points along the regression line. This suggests that the GBM model is highly accurate, effectively capturing the underlying pattern in the data. The narrow confidence interval further indicates high confidence in the predictions. For what it's worth, the linear regression plot shows a linear trend, but with a noticeable scatter around the regression lines, especially at the extremes. Linear regression captures the overall trend in the data but has more significant prediction errors compared to ensemble methods such as GBM and Random Forest. The confidence interval is wider, indicating less confidence in the predictions. Other algorithms like neural networks and Random Forest show good performance with moderate prediction confidence. Decision Tree has the lowest certainty of outcome as evidenced by the arrangement of points on the graph, indicating potential problems with over-fitting or lack thereof.

Interpretive graphs for various machine learning models, such as neural networks, decision trees, gradient boosting machines, linear regression, and random forests, can be efficiently generated using the R programming language. By using packages such as 'ggplot2' for visualization and 'caret' for modeling, it is possible to understand which variables have the greatest impact on model predictions. The 'varImp' function from the caret package is an example of a tool to assess the importance of variables, which is extremely useful in the feature selection process.

The graphs in Figure 3 illustrate the importance of various functions for different machine learning models. The importance of the functions indicates the extent to which each function contributes to the prediction performed by the model.




(a) (b)

Fig. 3. [cont.]

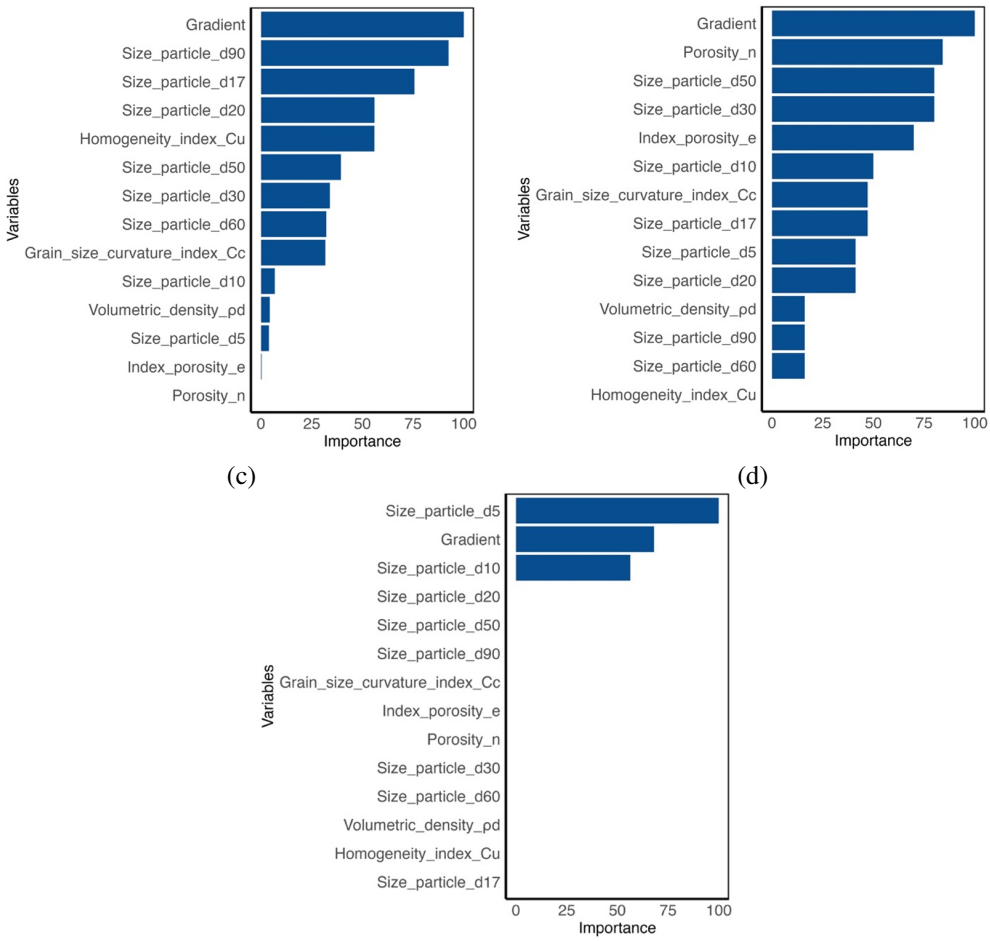



Fig. 3. Interpretive graphs for machine learning models: : (a) GBM; (b) Random Forest; (c) Neural Network; (d) Decision Tree; (e) Linear Regression

In the GBM models, the gradient variable stands out as particularly significant, influencing the model's prediction. This underscores its pivotal role in shaping the model's responses, especially when paired with specific particle size metrics such as size particle $d_5$, $d_{17}$, and $d_{10}$. The gradient, by capturing key trends and directional changes in the data, allows the GBM model to adjust its predictions more accurately, reflecting nuances in the dataset. At the same time, variable importance analysis showed that the gradient is consistently the most important variable in all models, highlighting its universal importance for the accuracy of various machine learning methods. In models such as neural networks, the gradient along with particle sizes (particle size $d_{90}$, $d_{17}$, $d_{20}$) plays a key role. Similarly, in decision trees and GBM models, Gradient in combination with other variables like porosity or size particle $d_5$ accounts for the predictive power of the model.

For linear regression, particle size (size particle $d_5$) and gradient were the variables with the greatest impact on results. For random forests, again the gradient along with grain size curvature index and particle size $d_{20}$ were the most influential.

Using the interpretive functions available in *R* language packages to analyze and interpret the importance of variables is key to optimizing models and understanding which features most affect their performance. Such information is especially relevant for further modeling based on other structures and materials, which will improve the accuracy and efficiency of machine learning models.

# 4. Conclusions

The results presented in the article provide clear evidence of the effectiveness of various machine learning models, particularly highlighting the superior performance of the Gradient Boosting Machine (GBM) and the key role of cross-validation in model evaluation. Here are the exact conclusions drawn from the study:

- The GBM model showed exceptional performance with an $R^2$ value of 0.96, significantly higher than other models tested. This high value indicates an excellent fit to the data and superior predictive accuracy, demonstrating GBM effectiveness in complex data environments.
- The study used MSE, RMSE, MAE, and $R^2$ to provide a comprehensive overview of the model's performance. It is worth noting that the GBM model recorded the lowest MSE of 1.76E-06 and RMSE of 0.0013, indicating minimal prediction errors compared to other models such as decision tree, random forest, neural network, and linear regression.
- Graphical analyses of observed values versus model-estimated values confirmed previous findings from the error analysis. The scatter plot for GBM showed a strong linear relationship between observed and predicted values and a narrow confidence interval, indicating high confidence in the model.
- The Gradient variable consistently showed the highest variable significance in all models, confirming its fundamental impact on predictive performance. This underscores the importance of feature selection in training and optimizing models. At the same time, it is important to note the significance of the impact of finer material fractions, such as size particle $d_5$, $d_{10}$, and $d_{17}$.
- The use of *i*-fold cross-validation allowed the analysis to be carried out equally for all tested machine learning models. This method consistently supports the reliability and accuracy of model evaluations by using multiple subsets of data.

Thus, what with the question posed in the title "Could Machine Learning Algorithms affect the efficient management of anthropogenic materials?", it is certainly so, it is still important to develop the techniques and explore their capabilities in terms of increasingly large databases. Based on the data from this study, future research should be focused on applying these models to different data sets or predictive scenarios, which would help verify the adaptability and robustness of these methods. In particular, a wider range of materials, especially anthropogenic materials, from different sources should be considered.

## References

[1] B. Bhardwaj and P. Kumar, "Waste foundry sand in concrete: A review", *Construction and Building Materials*, vol. 156, pp. 661–674, 2017, doi: 10.1016/j.conbuildmat.2017.09.010.

[2] M.K. Dash, S.K. Patro, and A.K. Rath, "Sustainable use of industrial-waste as partial replacement of fine aggregate for preparation of concrete – A review", *International Journal of Sustainable Built Environment*, vol. 5, no. 2, pp. 484–516, 2016, doi: 10.1016/j.ijsbe.2016.04.006.

[3] S.K. Kirthika, M. Surya, and S.K. Singh, "Effect of clay in alternative fine aggregates on performance of concrete", *Construction and Building Materials*, vol. 228, art. no. 116811, 2019, doi: 10.1016/j.conbuildmat.2019.116811.

[4] J. Dzięcioł and W. Sas, "Perspective on the Application of Machine Learning Algorithms for Flow Parameter Estimation in Recycled Concrete Aggregate", *Materials*, vol. 16, no. 4, art. no. 1500, 2023, doi: 10.3390/MA16041500.

[5] I. Yuksel, "Blast-furnace slag", in *Waste and Supplementary Cementitious Materials in Concrete: Characterisation, Properties and Applications*, R. Siddique and P. Cachim, Eds. Woodhead Publishing, 2018, pp. 361–415, doi: 10.1016/B978-0-08-102156-9.00012-2.

[6] W. Sas, et al., "Geotechnical and environmental assessment of Blast Furnace Slag for engineering applications", *Materials*, vol. 14, no. 20, art. no. 6029, 2021, doi: 10.3390/ma14206029.

[7] J. Dzięcioł and M. Radziemska, "Blast Furnace Slag, Post-Industrial Waste or Valuable Building Materials with Remediation Potential?", *Minerals*, vol. 12, no. 4, art. no. 478, 2022, doi: 10.3390/min12040478.

[8] L. Rafalski, *Road Foundations*. Warsaw: Road and Bridge Research Institute, 2007 (in Polish).

[9] Z. Pazdro, *Basic hydrogeology*, 3rd ed. Warsaw: Geological Publishers, 983 (in Polish).

[10] G.J. Osborne, "Durability of Portland blast-furnace slag cement concrete", *Cement and Concrete Composites*, vol. 21, no. 1, pp. 11–21, 1999, doi: 10.1016/S0958-9465(98)00032-8.

[11] G. Habert, J.B. d'Espinose de Lacaillerie, and N. Roussel, "An environmental evaluation of geopolymer based concrete production: Reviewing current research trends", *Journal of Cleaner Production*, vol. 19, no. 11, pp. 1229–1238, 2011, doi: 10.1016/j.jclepro.2011.03.012.

[12] Q. Li, "Properties of one-part CaO-Na2CO3 activated ground granulated blast furnace slag", *Archives of Civil Engineering*, vol. 70, no. 2, pp. 461–476, 2024, doi: 10.24425/ace.2024.149875.

[13] W. Sas and J. Dzięcioł, "Determination of the filtration rate for anthropogenic soil from the recycled concrete aggregate by analytical methods", *Scienti?c Review – Engineering and Environmental Sciences*, vol. 27, no. 2, pp. 236–248, 2018, doi: 10.22630/PNIKS.2018.27.2.23 (in Polish).

[14] J.H. Friedman, "Greedy function approximation: A gradient boosting machine", *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.

[15] J.R. Quinlan, "Induction of Decision Trees", *Machine Learning*, vol. 1, no. 1, 1986, doi: 10.1023/A:1022643204877.

[16] L. Breiman, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[17] J. Dzięcioł and W. Sas, "Estimation of the coefficient of permeability as an example of the application of the Random Forest algorithm in Civil Engineering", *Archives of Civil Engineering*, vol. 70, no. 2, pp. 119–134, 2024, doi: 10.24425/ace.2024.149854 .

[18] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958, doi: 10.1037/h0042519.

[19] A.M. Legendre, *Nouvelles méthodes pour la détermination des orbites des comètes*. 1806.

[20] M.W. Browne, "Cross-Validation Methods", *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 108–132, 2000, doi: 10.1006/JMPS.1999.1279.

[21] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation", *Statistics and Computing*, vol. 21, no. 2, pp. 137–146, 2011, doi: 10.1007/S11222-009-9153-8.

# Czy algorytmy uczenia maszynowego mogą wpłynąć na efektywne zarządzanie materiałami antropogenicznymi?

**Słowa kluczowe:** współczynnik filtracji, algorytmy uczenia maszynowego, żużel wielkopiecowy, zrównoważony rozwój, materiały antropogeniczne

## Streszczenie:

Poszukiwanie alternatyw dla kruszyw naturalnych w branży budowlanej doprowadziło do przewartościowania podejścia do odpadów budowlanych i poprzemysłowych, które są coraz częściej wykorzystywane w inżynierii lądowej i wodnej. Przykładem takiego materiału jest żużel wielkopiecowy, produkt uboczny produkcji stali, który może być stosowany jako kruszywo lub dodatek do betonu. Wykorzystanie materiałów antropogenicznych jako substytutów naturalnych kruszyw może pomóc zmniejszyć ilość odpadów na składowiskach. Ważne jest jednak dokładne określenie parametrów tych materiałów, które mogą się różnić i zależeć od pochodzenia lub genezy materiału. Biorąc pod uwagę, że wczorajsze odpady stają się dziś surowcami, zrównoważone zarządzanie nimi stanowi kluczowy element gospodarki o obiegu zamkniętym. Poszukiwanie rozwiązań pozwalających na ich jak najkorzystniejsze wykorzystanie staje się kolejnym z wyzwań. Czy algorytmy uczenia maszynowego mogą okazać się jednym z możliwych rozwiązań? Na przykładzie procesu estymacji przedstawiono potencjał wykorzystania algorytmów uczenia maszynowego do oszacowania współczynnika filtracji. Współczynnik filtracji jest kluczowym parametrem dla prawidłowego zastosowania kruszyw antropogenicznych w budownictwie i infrastrukturze drogowej, określając zdolność kruszyw do przepuszczania wody, co ma kluczowe znaczenie dla projektowania. Prawidłowe oszacowanie tego współczynnika może mieć znaczący wpływ na trwałość i stabilność konstrukcji. Celem niniejszego artykułu było zbadanie różnych metod szacowania współczynnika filtracji pochodzących z różnych technik uczenia maszynowego oraz wskazanie kierunku dalszych badań w tym obszarze. Przeanalizowano wyniki uzyskane z modeli Gradient Boosting Machine (GBM), drzewa decyzyjnego, lasu losowego, sieci neuronowej i regresji liniowej. Model GBM osiągnął najwyższą dokładność predykcyjną z wartością współczynnika determinacji $R^2$ wynoszącą 0.96, podczas gdy model regresji liniowej miał najniższą wartość $R^2$ wynoszącą 0.83. GBM osiągnąłrównież najniższe wartości *MSE* (1.76E-06) i *RMSE* (0.0013), co oznacza minimalne błędy predykcji. Zastosowanie k-krotnej walidacji krzyżowej zapewniło rzetelną ocenę wydajności modeli, zwiększając wiarygodność wyników. Badanie wykazało, że algorytmy uczenia maszynowego, mogą znacząco wpłynąć na efektywne zarządzanie materiałami antropogenicznymi. Przyszłe badania powinny skupić się na testowaniu modeli na różnych zestawach danych i różnych materiałach, aby zweryfikować ich adaptacyjność. Badania zaprezentowane w artykule potwierdziły potencjałtych technik w oszacowaniu kluczowych parametrów geotechnicznych. W związku z pytaniem postawionym w tytule "Czy algorytmy uczenia maszynowego mogą wpłynąć na efektywne zarządzanie materiałami antropogenicznymi?", z pewnością tak, nadal ważne jest jednak dopracowanie technik, procedur i zbadanie ich możliwości w odniesieniu do coraz większych baz danych.

Received: 2024-09-30, Revised: 2024-10-21