CHALLENGES OF EFFECTIVE PERMEABILITY COEFFICIENT ESTIMATION USING BOOSTING AND BAGGING TECHNIQUES FOR SELECTED ANTHROPOGENIC AGGREGATES

Justyna Dzięcioł, Warsaw University of Life Sciences, Institute of Civil Engineering, Assistant Prof., DSC, MSC, Eng., justyna_dzieciol@sggw.edu.pl
Wojciech Sas, Warsaw University of Life Sciences, Institute of Civil Engineering, Associate Prof. DSC (with habilitation), MSC, Eng., Head of Geotechnical Department, wojciech_sas@sggw.edu.pl

ABSTRACT
Reliable estimation of the permeability coefficient (k) for anthropogenic aggregates remains one of the critical research challenges in geotechnical and hydraulic engineering. Conventional empirical relationships, originally developed for natural sands, often fail to accurately reflect the hydraulic behaviour of these materials due to their distinctive particle morphology, elevated porosity, and heterogeneous grain-size distributions. In recent years, machine learning methodologies, particularly ensemble techniques such as bagging and boosting have shown considerable potential for improving predictive accuracy. Algorithms including Random Forest, AdaBoost, and XGBoost, combined with feature interpretation methods such as SHapley Additive exPlanations (SHAP), enable both robust predictions and consistent interpretations. This review synthesises current laboratory practices, empirical formulations, and advanced modelling approaches, critically assessing their limitations. The analysis highlights the need for standardisation of testing procedures, the development of large multi-source datasets, and the integration of hybrid frameworks based on data-driven approaches to support sustainable infrastructure design using machine learning algorithms and to evaluate their performance.

1. Introduction

In the European Union, construction and demolition waste accounts for more than one third of all waste, which makes this stream pivotal for the circular economy and resource security in the construction sector [1]. EU policy steers Member States toward high recovery levels through the Waste Framework Directive and its 2018 amendment, which consolidates targets and strengthens circular economy instruments [2,3]. In Poland, this direction was confirmed by the 2019 Roadmap for the Circular Economy that includes actions for civil engineering and construction [4,5]. The increased use of anthropogenic aggregates, such as recycled concrete aggregate and metallurgical slags, nevertheless exposes the challenge of reliably assessing hydraulic parameters that are essential for safe design of structural layers, drainage, and stability.

Determination of the permeability coefficient rests on Darcy's law and permeameter testing under constant or falling head, with laboratory procedures standardized by ASTM D2434 and ASTM D5084 [6,7]. The hydraulic performance, particularly the effectiveness of predicting the permeability coefficient is pivotal for drainage layers, bases and embankments, where inadequate permeability can compromise stability and serviceability [8–10]. Recycled concrete aggregate (RCA) and slag exhibit broader variability in gradation, angularity, porosity and surface chemistry than quarried aggregates, which complicates transferability of empirical soil formulae and motivates data-driven modelling [11]. Recent geotechnical reviews emphasize both the promise and the methodological pitfalls of machine learning (ML), including sampling bias, optimistic validation, and weak external generalization, issues that are amplified for heterogeneous recycled materials [12–14]. Within this context, field- and lab-scale evidence shows that RCA and steel-slag aggregates (SS) can deliver drainage capacity comparable to natural aggregates when properly processed and graded [15,16], yet prediction of permeability coefficient remains sensitive to fines content, threshold gradients and compaction energy [8,17]. Initial studies applying ML to permeability coefficient estimation for anthropogenic aggregates have yielded encouraging results. Dzięcioł et al. used a Random Forest model to predict the filtration coefficient of blast furnace slag (BFS)-based aggregates, achieving R^2 = 0.84-0.92 on validation data [18]. In another study, AdaBoost was used to estimate the permeability of RCA, attaining an R^2 = 0.89 on test samples (with even higher 0.95 on training), significantly better than a comparative Artificial Neural Network (ANN) which reached only R^2 = 0.64 under the same conditions [19]. These examples highlight the potential of boosting and bagging techniques to improve prediction accuracy for heterogeneous recycled materials. Nevertheless, significant challenges remain before such models can be broadly adopted in practice. One challenge is the limited size and representativeness of available datasets, models may overfit to lab-specific data and struggle to generalize to new sources of material. Another challenge is the interpretability of ML models: engineers need to trust and understand the predictions. Black-box models, no matter how accurate, may be met with skepticism if they cannot explain which material features drive the results. To address this, researchers have begun integrating eXplainable AI (XAI) techniques like SHapley Additive exPlanations (SHAP) to shed light on model decision-making [20–22]. This review paper aims to analyze the current state-of-the-art in ML-based permeability prediction for selected anthropogenic aggregates, focusing on ensemble methods (boosting and bagging) and their interpretability.

2. Methods Review and Current State of Knowledge

Estimating the permeability coefficient for anthropogenic aggregates is a multi-faceted challenge requiring a combination of laboratory testing, empirical modelling, and data-driven prediction. Traditional methods, such as constant head and falling head permeability tests, have been adapted for recycled and industrial by-product aggregates, often revealing increased variability in results compared to natural aggregates [23–25]. Laboratory procedures for RCA, BFS, SS, Fly Ash (FA), and Bottom Ash (BA) must account for factors such as grain shape angularity, higher porosity, and potential contaminant leaching [26,27]. Empirical formulas, including Hazen's, Kozeny-Carman, and Slichter's equations, have been historically applied to granular soils [28–30]. However, their applicability to anthropogenic materials is limited due to deviations in grain size distribution, fines content, and surface roughness. For example, RCA often exhibits higher water absorption and irregular particle morphology, which increase tortuosity and reduce the accuracy of empirical correlations [31,32].

In recent years, Machine Learning (ML) approaches have gained prominence. Among them, boosting and bagging stand out for their ability to integrate heterogeneous predictors and minimize bias-variance trade-offs [33,34]. AdaBoost algorithms (boosting) sequentially train weak learners, typically shallow decision trees by emphasizing previously mispredicted cases, thereby improving prediction accuracy for complex datasets [35,36]. Random Forests algorithms (bagging), build multiple decision trees on bootstrapped datasets and aggregate their predictions, effectively reducing overfitting [33,37,38].

Recent applications in geotechnics include permeability estimation for RCA using AdaBoost, yielding high training accuracy (R^2 = 0.959) and good generalization performance (R^2 = 0.886) compared to ANN-based models [19,39]. Similar studies for BFS and FA have reported that tree-based ensembles outperform linear regression and support vector machines when grain-size heterogeneity is high.

Interpretability frameworks have also emerged as essential complements to predictive modelling. Tools such as Shapley Additive Explanations (SHAP) allow quantification of the influence of individual features, such as hydraulic gradient, bulk density, and d_{50} particle size on predicted permeability values [10,18]. This transparency is particularly important for engineering acceptance and regulatory compliance.

Beyond AdaBoost and Random Forest, advanced data-driven techniques are being explored for permeability estimation. For instance, integrating fractal-based grain size descriptors into a Random Forest model was shown to improve prediction accuracy by about 62% compared to the Hazen equation for broadly graded sands [40]. Support Vector Regression models, especially with Bayesian optimization, have also demonstrated the ability to capture non-linear effects in extreme gradation cases (e.g. very high uniformity coefficients) better than traditional empirical methods. Nonetheless, modern ensemble and deep-learning approaches typically achieve higher accuracy on complex datasets. Recent studies have designed physics-informed neural networks and incorporated transfer learning with XGBoost to leverage prior knowledge, successfully handling high-dimensional feature interactions even when data are limited [40]. While these cutting-edge methods push the performance frontier, their complexity and data requirements have so far limited their practical deployment in routine permeability assessments.

3. Future Challenges

Future research on the estimation of the permeability coefficient for anthropogenic aggregates using boosting and bagging techniques will need to address several critical gaps in knowledge, methodology, and implementation. The first challenge involves the systematic integration of multi-scale data, including grain-scale imaging, micro-computed tomography, and pore-network modeling, with ensemble learning algorithms. Although machine learning models such as Random Forest, Gradient Boosting, and AdaBoost have demonstrated strong predictive capacity for recycled concrete aggregate, blast furnace slag, and fly ash mixtures [18,26], their input features often lack direct descriptors of pore connectivity and tortuosity, which limits their capacity to generalize across materials with different hydraulic behavior [41]. A second future challenge is the development of transferable models capable of operating under diverse climatic, geochemical, and compaction conditions. The majority of current datasets are regional in scope, which means that trained models may not accurately predict permeability for aggregates sourced from different production facilities or processed under alternative quality control protocols. Cross-regional calibration and the creation of large, open-access, quality-controlled datasets would be necessary to achieve generalizable solutions. This will require collaboration between industrial producers, research laboratories, and governmental agencies to standardize testing procedures and data-sharing protocols.

The third area of future research concerns the physical interpretability of ensemble predictions. While explainable AI tools such as SHAP have allowed engineers to rank the relative importance of features, these rankings still require validation against physical models and controlled experiments [10]. Integrating physical constraints directly into model training, for example through physics-informed machine learning frameworks, may help ensure that predictions remain consistent with known hydraulic principles, thereby improving the acceptability of these methods in regulatory and design contexts. Another emerging challenge is the computational efficiency of high-complexity models. The environmental and operational benefits of rapid permeability estimation can only be realized if models can be deployed in situ via portable devices or integrated into real-time monitoring systems. This will require research into lightweight ensemble architectures and hybrid approaches that combine simplified physical equations with machine learning surrogates to reduce computational cost without sacrificing accuracy [42].

Finally, future research must address the sustainability implications of using anthropogenic aggregates. Although the reuse of blast furnace slag, steel slag, and recycled concrete aggregate can significantly reduce the environmental footprint of construction [43,44], the hydraulic performance of these materials under long-term environmental exposure remains insufficiently documented. Long-term field trials, coupled with periodic permeability measurements and model updates, will be essential to ensure that design assumptions remain valid over the service life of geotechnical and hydraulic structures.

4. Conclusions

The review of current methodologies for permeability coefficient estimation in anthropogenic aggregates demonstrates that ensemble learning methods, particularly bagging and boosting techniques, offer a substantial improvement in predictive accuracy compared to traditional empirical and analytical approaches. This superiority is most evident when models are trained on datasets incorporating both granulometric descriptors and state variables such as compaction energy, hydraulic gradient, and density, which together capture the complex and nonlinear nature of flow in recycled and industrial aggregates. Applications to blast furnace slag, steel slag, recycled concrete aggregate, fly ash, and bottom ash confirm that these materials require modelling frameworks capable of addressing heterogeneity in particle shape, pore structure, and mineral composition. Ensemble learners have been shown to handle this variability more robustly than single algorithms, and post-hoc interpretability tools such as SHAP can provide physically coherent explanations that support engineering acceptance. Nevertheless, the adoption of ensemble learning in permeability prediction is constrained by data scarcity, lack of standardization, and limited transferability across regions and production processes. Addressing these issues will require coordinated efforts to develop large, open, quality-controlled datasets, to integrate microstructural descriptors into modelling, and to validate interpretability outputs against physical theory and experiments.

Future work should also focus on creating computationally efficient ensemble models that can be embedded in portable testing devices and real-time monitoring systems, ensuring that advances in model accuracy translate into practical engineering tools. Finally, the long-term hydraulic performance and environmental durability of anthropogenic aggregates under field conditions should be monitored and modelled to ensure that their sustainable use in geotechnical and hydraulic engineering.

Reference:

1. European Commission Construction and Demolition Waste Available online: https://environment.ec.europa.eu/topics/waste-and-recycling/construction-and-demolition-waste_en (accessed on 13 August 2025).
2. European Parliament & Council Directive 2008/98/EC on Waste (Waste Framework Directive) Available online: https://eur-lex.europa.eu/eli/dir/2008/98/oj/eng (accessed on 13 August 2025).
3. European Commission Waste Framework Directive Available online: https://environment.ec.europa.eu/topics/waste-and-recycling/waste-framework-directive_en (accessed on 13 August 2025).
4. Eurostat Waste Statistics Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Waste_statistics (accessed on 13 August 2025).
5. Ministry of Development (Poland) Road Map towards the Transition to Circular Economy Available online: https://www.gov.pl/attachment/e9c53c81-61f7-4ccd-b918-af40cd57cb7c (accessed on 13 August 2025).
6. ASTM D2434-22 Test Method for Permeability of Granular Soils (Constant Head). ASTM International 2022.
7. ASTM D5084-16a Test Methods for Measurement of Hydraulic Conductivity of Saturated Porous Materials Using a Flexible Wall Permeameter. ASTM International 2016.
8. Głuchowski, A.; Sas, W.; Dzięcioł, J.; Soból, E.; Szymański, A. Permeability and Leaching Properties of Recycled Concrete Aggregate as an Emerging Material in Civil Engineering. Applied Sciences 2018, 9, 81, doi:10.3390/app9010081.
9. Oikonomou, N.D. Recycled Concrete Aggregates. Cem Concr Compos 2005, 27, 315–318, doi:10.1016/j.cemconcomp.2004.02.020.
10. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. Adv Neural Inf Process Syst 2017, 2017-Decem, 4766–4775, doi:10.48550/arxiv.1705.07874.
11. Sas, W.; Dzięcioł, J.; Głuchowski, A. Estimation of Recycled Concrete Aggregate's Water Permeability Coefficient as Earth Construction Material with the Application of an Analytical Method. Materials 2019, 12, 2920, doi:10.3390/ma12182920.
12. Zhao, T.; Shen, F.; Xu, L. Review and Comparison of Machine Learning Methods in Developing Optimal Models for Predicting Geotechnical Properties with Consideration of Feature Selection. Soils and Foundations 2024, 64, 101523, doi:10.1016/j.sandf.2024.101523.
13. Bozorgzadeh, N.; Feng, Y. Evaluation Structures for Machine Learning Models in Geotechnical Engineering. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 2024, 18, 52–59, doi:10.1080/17499518.2024.2313485.
14. Shao, W.; Yue, W.; Zhang, Y.; Zhou, T.; Zhang, Y.; Dang, Y.; Wang, H.; Feng, X.; Chao, Z. The Application of Machine Learning Techniques in Geotechnical Engineering: A Review and Comparison. Mathematics 2023, 11, 3976, doi:10.3390/math11183976.
15. Roque, A.J.; da Silva, P.F.; de Almeida, R.P.M. Recycling of Crushed Concrete and Steel Slag in Drainage Structures of Geotechnical Works and Road Pavements. J Mater Cycles Waste Manag 2022, 24, 2385–2400, doi:10.1007/s10163-022-01486-7.
16. McCulloch, T.; Kang, D.; Shamet, R.; Lee, S.J.; Nam, B.H. Long-Term Performance of Recycled Concrete Aggregate for Subsurface Drainage. Journal of Performance of Constructed Facilities 2017, 31, doi:10.1061/(ASCE)CF.1943-5509.0000994.

17. Nguyen, T.L.; Nguyen, V.T.; Nguyen, H.G.; Matsuno, A.; Sakanakura, H.; Kawamoto, K. Mechanical and Hydraulic Properties of Recycled Concrete Aggregates Mixed with Clay Brick Aggregates and Particle Breakage Characteristics for Unbound Road Base and Subbase Materials in Vietnam. Sustainability 2022, 14, 4854, doi:10.3390/su14084854.
18. Dzięcioł, J.; Sas, W. Estimation of the Coefficient of Permeability as an Example of the Application of the Random Forest Algorithm in Civil Engineering. Archives of Civil Engineering 2024, 119–134, doi:10.24425/ace.2024.149854.
19. Dzięcioł, J.; Sas, W. Estimation of Permeability Coefficient Using Machine Learning Algorithms on the Example of AdaBoost and Artificial Neural Network Algorithms. In Geotechnical Engineering Challenges to Meet Current and Emerging Needs of Society; CRC Press: London, 2024; pp. 1572–1577.
20. Dzięcioł, J.; Sas, W. Perspective on the Application of Machine Learning Algorithms for Flow Parameter Estimation in Recycled Concrete Aggregate. Materials 2023, Vol. 16, Page 1500 2023, 16, 1500, doi:10.3390/MA16041500.
21. Hart, S. Shapley Value. Game Theory 1989, 210–216, doi:10.1007/978-1-349-20181-5_25.
22. Rozemberczki, B.; Watson, L.; Bayer, P.; Yang, H.T.; Kiss, O.; Nilsson, S.; Sarkar, R. The Shapley Value in Machine Learning. IJCAI International Joint Conference on Artificial Intelligence 2022, 5572–5579, doi:10.48550/arxiv.2202.05594.
23. Head, K.H.; Epps, R. Manual of Soil Laboratory Testing, Volume 2. Permeability, Shear Strength and Compressibility Tests. Manual of soil laboratory testing, volume 2. Permeability, shear strength and compressibility tests. 1994, doi:10.1016/0016-7061(95)90001-2.
24. Elhakim, A.F. Estimation of Soil Permeability. Alexandria Engineering Journal 2016, 55, 2631–2638, doi:10.1016/j.aej.2016.07.034.
25. Tran, V.Q. Predicting and Investigating the Permeability Coefficient of Soil with Aided Single Machine Learning Algorithm. Complexity 2022, 2022, doi:10.1155/2022/8089428.
26. Kumar, N.; Prakash, S.; Ghani, S.; Gupta, M.; Saharan, S. Data-Driven Machine Learning Approaches for Predicting Permeability and Corrosion Risk in Hybrid Concrete Incorporating Blast Furnace Slag and Fly Ash. Asian Journal of Civil Engineering 2024, 25, 3263–3275, doi:10.1007/s42107-023-00977-8.
27. Yu, F.; Chu, W.; Zhang, R.; Gao, Z.; Yang, Y. Predicting the Permeability and Compressive Strength of Pervious Concrete Using a Stacking Ensemble Machine Learning Approach. Sci Rep 2025, 15, 22506, doi:10.1038/s41598-025-08479-0.
28. Chapuis, R. Predicting the Saturated Hydraulic Conductivity of Sand and Gravel Using Effective Diameter and Void Ratio. Can Geotech J. 2004, 41, 787–795.
29. Slichter, S. Theoretical Investigation of the Motion of Ground Water. USGS 19th Annual Report, part II, 1899, 301–380.
30. Hazen A Some Physical Properties of Sand and Gravel, with Special Reference to Their Use in Filtration. 24th Annual Report, State Board of Health, Commonwealth of Massachusetts, Boston 1892, 539–556.
31. Poon, C.S.; Chan, D. Feasible Use of Recycled Concrete Aggregates and Crushed Clay Brick as Unbound Road Sub-Base. Constr Build Mater 2006, 20, 578–585, doi:10.1016/j.conbuildmat.2005.01.045.
32. Danish, A.; Mosaberpanah, M.A. A Review on Recycled Concrete Aggregates (RCA) Characteristics to Promote RCA Utilization in Developing Sustainable Recycled Aggregate Concrete (RAC). European Journal of Environmental and Civil Engineering 2022, 26, 6505–6539, doi:10.1080/19648189.2021.1946721.
33. Breiman, L. Random Forests. Mach Learn 2001, 45, 5–32, doi:10.1023/A:1010933404324.
34. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J Comput Syst Sci 1997, 55, 119–139, doi:10.1006/jcss.1997.1504.
35. Zhou, Z.-H.; Yu, Y. AdaBoost. The Top Ten Algorithms in Data Mining 2009, 141–164.
36. Solomatine, D.P.; Shrestha, D.L. AdaBoost.RT: A Boosting Algorithm for Regression Problems. In Proceedings of the IEEE International Conference on Neural Networks - Conference Proceedings; 2004; Vol. 2, pp. 1163–1168.

37. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random Forests. In Ensemble Machine Learning: Methods and Applications; Springer, Boston, MA: Boston, MA, 2012; pp. 157–175 ISBN 9781441993267.
38. Genuer, R.; Poggi, J.-M.; Tuleau, C. Random Forests: Some Methodological Insights. 2008, doi:10.48550/arxiv.0811.3619.
39. Kumar, A.; Ahuja, R. Prediction of Permeability of Amended Soil Using Ensembled Artificial Intelligence Models. AI in Civil Engineering 2025, 4, 9, doi:10.1007/s43503-025-00052-y.
40. Wang, J.; Ding, H.; Guan, L.; Wang, Y. Prediction on Permeability Coefficient of Continuously Graded Coarse-Grained Soils: A Data-Driven Machine Learning Method. Applied Sciences 2025, 15, 5248, doi:10.3390/app15105248.
41. Li, H.; Yang, J.; Yu, X.; Zhang, Y.; Zhang, L. Permeability Prediction of Pervious Concrete Based on Mix Proportions and Pore Characteristics. Constr Build Mater 2023, 395, 132247, doi:10.1016/j.conbuildmat.2023.132247.
42. Wang, J.; François, B.; Lambert, P. Equations for Hydraulic Conductivity Estimation from Particle Size Distribution: A Dimensional Analysis. Water Resour Res 2017, 53, 8127–8134, doi:10.1002/2017WR020888.
43. Islam, A.; Alengaram, U.J.; Jumaat, M.Z.; Bashar, I.I.; Kabir, S.M.A. Engineering Properties and Carbon Footprint of Ground Granulated Blast-Furnace Slag-Palm Oil Fuel Ash-Based Structural Geopolymer Concrete. Constr Build Mater 2015, doi:10.1016/j.conbuildmat.2015.10.026.
44. Pomponi, F.; Moncaster, A. Circular Economy for the Built Environment: A Research Framework. J Clean Prod 2017, 143, 710–718, doi:10.1016/J.JCLEPRO.2016.12.055.