Improving Concrete Mix Design Predictions with Machine Learning Algorithms

(Article begins on next page)

Improving Concrete Mix Design Predictions with Machine Learning Algorithms

Abstract—**The construction industry has a crucial impact on environmental pollution and on the phenomenon of climate change.**

One of the most impactful activities in this sector, is the concrete production. The concrete industry alone is a significant source of CO_2 emissions, with concrete being the most widely used construction material. Concrete is a material mainly composed of cement, water, aggregates and additives. Despite the percentage of cement in concrete mixtures is comprised between 10% and 15% by weight, this component contributes up to 90% of the associated greenhouse gas emissions.

The process involved in concrete production is complex and involve many uncertainties. To compensate for these unknowns and ensure structural safety, cement is frequently added in excess to improve mechanical performance. However, reducing the cement content in concrete formulations can have considerable environmental benefits. Therefore, enhancing the precision of predicting concrete's mechanical characteristics is essential for lowering cement consumption without compromising safety.

In this paper, the application of machine learning techniques to forecast the mechanical properties of concrete is presented. The main goal is to use the concrete mix design data to accurately predict the properties related to the resistance and the workability of concrete. To this extent, a dataset comprising roughly 1100 mix designs was utilized, combined with weather-related data from concrete batching plants. The study emphasizes regression and classification models to predict concrete properties, taking into account environmental and meteorological factors at the production site.

Index Terms—**Concrete, Mix Design, Machine Learning, Sustainability, Cost reduction, Artificial Intelligence**

I. INTRODUCTION

The construction industry is a crucial sector with both significant economic and environmental implications. It contributes approximately 10% to the global gross domestic product (GDP) [1], making it one of the largest industries worldwide. However, its rapid expansion, especially in developing regions, raises concerns about its long-term sustainability [2].

One of the most pressing issues associated with this sector is its environmental footprint. The industry is responsible for nearly 30% of global greenhouse gas emissions [3] and consumes approximately 50% of the world's raw materials and 40% of total energy resources [4]. As the demand for infrastructure and urban development grows, mitigating these environmental effects becomes increasingly urgent.

In recent years, the entire construction industry has been increasingly adopting optimization methods to improve the efficiency and sustainability of the entire construction process [5]. These efforts start with the optimization of material production, continue with the refinement of preliminary design stages [6], [7], and extend to the optimization of construction processes on-site [8], [9]. Additionally, advanced monitoring techniques, including non-destructive testing methods [10], [11], are being employed to prolong the useful life of structures [12], further reducing the environmental footprint.

Concrete production is among the most resource-intensive activities within the sector. Given that concrete is one of the most widely used construction materials, its impact is substantial. The mixture consists primarily of cement, water, and aggregates, with additional chemical components—such as superplasticizers, set retarders, and air entrainers—used to enhance performance characteristics.

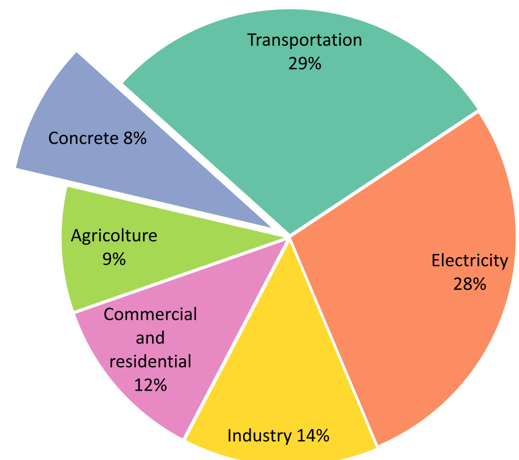
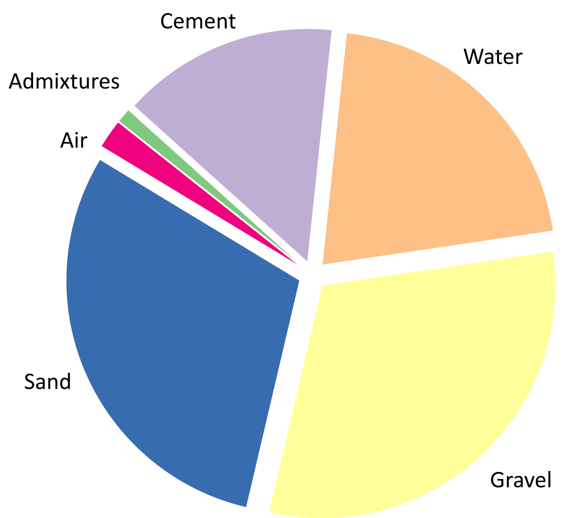


Fig. 1: Contribution of Concrete to CO_2 Emissions Compared to Other Sectors

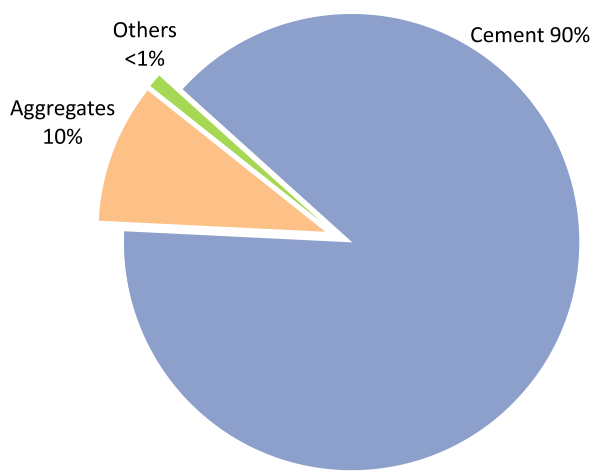
Cement, a key ingredient in concrete, plays a particularly significant role in environmental concerns. Its manufacturing process alone accounts for roughly 8-10% of worldwide CO_2 emissions [13], [14] (see Figure 1). Consequently, the industry's contribution to climate change cannot be overlooked, and solutions aimed at improving material efficiency and reducing emissions are becoming a priority.

In response to these challenges, initiatives such as the European Green Deal have set ambitious sustainability targets, including a 55% reduction in CO_2 emissions by 2030 (compared to 1990 levels) and achieving carbon neutrality by 2050. These measures emphasize the need for innovation within the construction sector to balance economic growth with environmental responsibility.

Despite making up only 10% to 15% of the total mass in a concrete mixture (Figure 2a), cement is the primary contributor to the greenhouse gas emissions generated during concrete production. Studies estimate that its use is responsible for as much as 90% of these emissions (Figure 2b), underscoring its substantial environmental footprint relative to its proportion in the mix.



(a) Concrete mix-design



(b) Concrete mix-design CO_2 emissions

Fig. 2: Graphical representation of the proportion of key components in a concrete mix (a) and their relative share of CO_2 emissions generated during production (b).

Optimizing cement usage in concrete production is crucial for reducing global CO_2 emissions [15]. However, because cement content directly influences mechanical properties, lowering its proportion introduces a risk of compromising structural strength. The complexity of concrete production stems from the variability in raw materials [16], [17], diverse manufacturing processes, and fluctuating environmental conditions, all of which make predicting final material properties particularly challenging [18].

To compensate for these uncertainties, an excess amount of cement is often added to the mix as a precaution. While this ensures compliance with safety and performance standards, it also results in unnecessary material consumption and associated environmental impacts. Given that concrete properties can change throughout their lifespan, various monitoring techniques have been developed to assess their structural performance over time [10], [19].

A key challenge in production lies in the storage of aggregates, which are typically kept in open stockpiles where exposure to environmental factors leads to variations in moisture content. Although modern plants utilize sensors to measure surface moisture, these readings do not always reflect internal humidity levels, which can significantly differ from surface measurements. This discrepancy introduces inconsistencies in water content, ultimately affecting the quality and mechanical properties of the final mix.

This research explores the potential of machine learning techniques to enhance predictive accuracy in concrete mix design. By applying regression and classification algorithms, the goal is to develop more reliable estimates of concrete's mechanical characteristics [20]. Unlike traditional approaches, this study introduces meteorological data as an additional input, allowing for better modeling of the hidden moisture within aggregates. The environmental parameters used in this research were sourced from the publicly available database of ARPAL (Agenzia Regionale per la Protezione dell'Ambiente Ligure).

By incorporating external climatic variables into the predictive framework, this methodology aims to refine mix design precision. Ultimately, the objective is to reduce cement consumption while maintaining structural integrity, leading to both economic and environmental benefits in concrete production.

II. Integrating Environmental Data into Concrete Property Prediction

This study investigates the use of machine learning techniques to estimate the mechanical properties of concrete, integrating real-world industrial mix designs with meteorological data from batching plants. The research is based on a dataset of approximately 1,100 samples, which includes key variables such as cement type, water and cement content per cubic meter, the presence of admixtures, and theoretical density.

To facilitate supervised learning, each data point is labeled with measured mechanical properties, specifically compressive strength at 28 days (R_{ck}) and consistency class. Additionally, the dataset records the date of production and the location of the cement manufacturing plant, allowing for correlation with environmental conditions at the time of mixing.

One of the key challenges in concrete production is the influence of weather conditions on raw materials, particularly aggregates stored in open-air environments. Variations in moisture content due to rainfall and temperature fluctuations can significantly affect the water-to-cement ratio, ultimately influencing the final mechanical properties of the concrete. To address this problem, the study incorporates meteorological data, such as cumulative rainfall and average temperature in the days preceding production, into predictive models.

The meteorological data used in this investigation were obtained from the publicly available records of ARPAL (Agenzia Regionale per la Protezione dell'Ambiente Ligure). Although the data set does not specify the precise location of the concrete batching plants, industry knowledge suggests that cement is typically used within a 100 km radius of its production site. Since the cement analyzed in this study was produced in Genoa, Italy, the possible geographical area for the corresponding batching plants was estimated (see Figure 3).

By integrating environmental parameters into machine learning models, this research aims to improve the accuracy of concrete property predictions, ultimately contributing to more efficient mix designs and reducing unnecessary cement consumption.

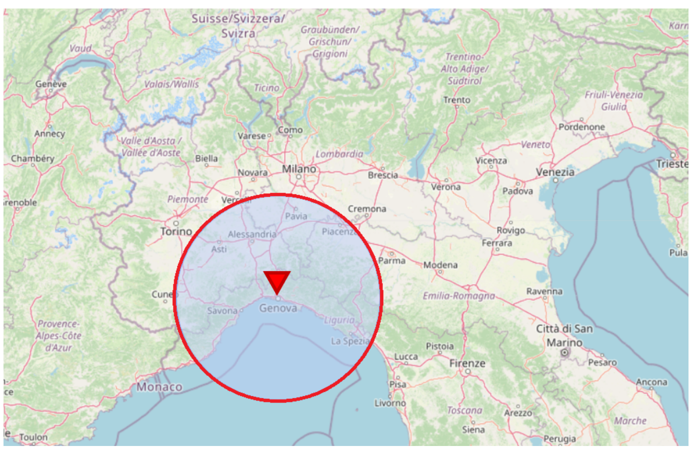


Fig. 3: Possible location area for the batching plant.

Taking into account the broad geographical area illustrated in Figure 3 and the diverse climatic conditions of the region, meteorological data from Genoa (Italy) were selected as a reference point for this study.

To enhance predictive accuracy, environmental variables such as cumulative rainfall and average temperature recorded in the days preceding concrete production were integrated into the training dataset.

This study explores the effectiveness of various machine learning algorithms in estimating concrete's mechanical properties. The models were trained using historical weather data that corresponded to the location of the batching plant at the time of production.

The analysis focuses specifically on forecasting two essential concrete characteristics: compressive strength and consistency class.

A. Prediction of Concrete Workability through Machine Learning

Concrete workability refers to how easily fresh concrete can be mixed, transported, placed, and compacted without segregation or excessive effort. It is a crucial factor in construction, as it influences both the ease of handling and the overall quality of the finished structure. If the concrete is too stiff, it becomes difficult to place and compact properly, increasing the risk of defects like honeycombing and voids. Conversely, if it is too fluid, segregation and excessive bleeding may occur, negatively impacting its mechanical properties.

To assess workability, concrete is classified based on its consistency, which is typically determined through standardized tests. One of the most widely used methods is the slump test, which provides a straightforward measure of the material's flowability. This test involves filling a truncated cone mold (also known as an Abrams cone) with fresh concrete in three layers. Each layer is compacted with a tamping rod to remove trapped air and ensure uniformity. Once the mold is completely filled, it is carefully lifted vertically, allowing the concrete to settle naturally under its own weight. The difference between the initial height of the mold and the final height of the slumped concrete, measured in millimeters, indicates its consistency. Based on this measurement, the concrete is classified into different consistency classes (see Table I).

A higher slump value means the concrete is more fluid and easier to work with, making it suitable for applications with dense reinforcement or complex formworks. In contrast, a lower slump indicates a stiffer mix, which may be more appropriate for pavements or large-volume pours where excessive flow is undesirable. Selecting the right consistency is essential to achieving durable and high-quality concrete structures.

TABLE I: Definition of consistency class for concrete.

Consistency Class	Slump [mm]
S1	10-40
S2	50-90
S3	100-150
S4	160-200
SR	>200

In this study, machine learning techniques were applied to classify concrete consistency based on mix properties. Various classification algorithms, including decision trees, logistic regression, support vector machines, neural networks, and ensemble methods, were tested and compared in terms of precision. By leveraging meteorological and material data, these models aim to improve the prediction of concrete behavior, optimize the design of the mix and reduce unnecessary material use.

B. Prediction of Concrete Resistance through Machine Learning

This study aims to predict the compressive strength of concrete 28 days after casting, an essential parameter influencing both the final strength and durability of a structure.

Given its critical role in structural design and verification, accurately estimating compressive strength is vital to ensuring the integrity of a structure and minimizing potential risks.

The compressive strength, represented as R_c, is typically determined by conducting compression tests on cubic concrete samples with 15 cm side lengths. However, even within a single batch of concrete, test results can exhibit variability. To address this, a statistical approach is used to construct a distribution curve that reflects the variability in concrete compressive strength. From this curve, the characteristic compressive strength, R_{ck}, is defined as the 95th percentile of the strength's normal distribution [21].

In order to predict the compressive strength, the study employs various machine learning-based regression models, including linear regression, support vector machines, regression trees, Gaussian process regression, kernel approximation regression, ensemble trees, and neural networks. To evaluate the performance of these models, a set of classical regression metrics are used, such as the Root Mean Square Error (RMSE), Maximum Absolute Error (MAE), and the Coefficient of Determination (R^2). This comparative analysis allows for a thorough assessment of the most effective model for estimating concrete's compressive strength.

III. Results and Comparative Analysis

A. Description of the Dataset

The dataset utilized in this study, which was compiled specifically for this research, includes several key parameters for each concrete mix design. These parameters cover a range of variables, such as the type of cement used (either type II or type III), the amount of cement in kilograms, and the cement's 28-day strength, categorized by class. It also records the volume of water added to the mix, measured in liters, the maximum diameter of the aggregates in millimeters, and the quantity of any additives incorporated into the mixture.

In addition to these, the dataset provides information about the consistency class and the 28-day compressive strength of the concrete after it has set, which are treated as the target variables to be predicted. The neural network models are trained using this dataset, where the known parameters of each mix design serve as the input features, and the goal is to predict both the consistency class and the compressive strength of the resulting concrete.

Figure 4 illustrates the discrepancy between the measured compressive strength and the required compressive strength for a sample of 1100 concrete mixes. In the graph, a positive difference (ΔR_{ck}) indicates that the produced concrete has a compressive strength greater than the required value.

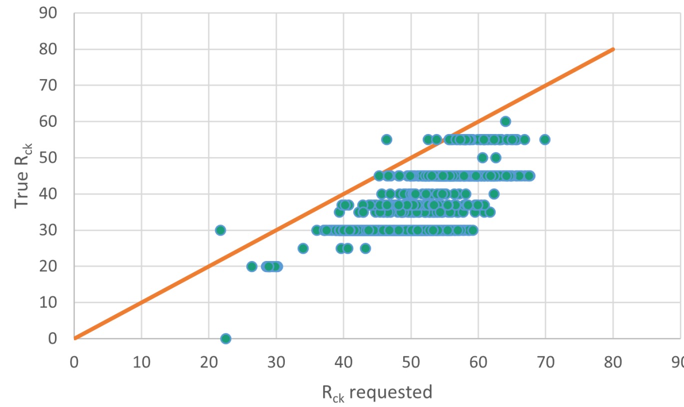


Fig. 4: Comparison between the measured and required compressive strength for a sample of 1100 concrete mixes. A positive ΔR_{ck} value indicates that the measured compressive strength exceeds the required value, while a negative ΔR_{ck} reflects the opposite.

B. Predictive Models for Consistency Class Classification

The results for the best-performing classification models used to predict the consistency class of the concrete samples are shown in Table II. These models are highly efficient, both in terms of memory consumption and computational load, allowing for the execution of thousands of predictions per second. Some of the models even achieve an accuracy rate greater than 94%, showcasing their effectiveness in reliably predicting the consistency class of concrete.

Figure 5 presents the outcomes of the best classification model. It is an ensemble tree that was trained considering the cumulative rainfall referred to one day before the concrete production. This model records the best performance with an accuracy of 94.5%. This model is based on an ensemble of decision trees, trained using the complete set of concrete mix data. Interestingly, it relies solely on the total rainfall from the day preceding the pouring as the environmental input.

Figure 5 presents the outcomes of classification model 6.13, which records the best performance with an accuracy of 94.5%. This model is based on an ensemble of decision trees, trained using the complete set of concrete mix data. Interestingly, it relies solely on the total rainfall from the day preceding the pouring as the environmental input.

C. Regression Models for Predicting Compressive Strength

Table III displays the performance of various machine learning models used to forecast the compressive strength of concrete. The models listed in the table include Ensemble Trees, Decision Trees, Gaussian Process Regression, Support Vector Machines, and Neural Networks. These models were selected as the most effective for the regression task of estimating concrete compressive strength. The table emphasizes the strong performance of these models, especially when assessed using metrics such as RMSE, MAE, and R^2. Importantly, the top-performing models achieved an RMSE lower than 4 MPa and an MAE under 2.78 MPa, demonstrating their precision in predicting compressive strength.

The ensemble of decision trees emerged as the most successful model for predicting compressive strength in this study. This model was developed using a comprehensive set of mix design data, along with an additional layer of information provided by the consistency class predicted by the classifi-

TABLE II: Comparison of the best Machine Learning-based Classification models

Model Type	Environmental Conditions	Size [kB]	Speed [obs/sec]	Accuracy (Validation)
Ensemble Tree	Rain 1d	647	6400	94.5%
Ensemble Tree	Rain 1d, 3d, 5d	852	7100	94.3%
Ensemble Tree	Temp 1d, 3d, 5d	977	3600	94.1%
Naive Bayes	Rain 1d	440	1900	85.3%
Neural Network	Rain 1d	12	47000	79.7%
SVM	Rain 1d	112	15000	79.5%
Neural Network	Rain 1d, 3d, 5d	13	64000	79.5%

TABLE III: Comparison of Top Machine Learning Regression Models

Model Name	Environmental Conditions	Size [kB]	RMSE	MAE
Ensemble Tree 3.15	Rain 1d, Temp 1d	567	3.623	2.78
Ensemble Tree 2.15	Rain 1d, 3d, 5d, Temp 1d, 3d, 5d	1013	3.681	2.84
Decision Tree 9	Rain 1d, Temp 1d	212	3.747	2.86
Gaussian Process 7	Rain 1d, Temp 1d	200	3.833	2.84
SVM 2.12	Rain 1d, Temp 1d, 3d, 5d	195	3.959	2.98
Neural Network 3.24	Rain 1d, Temp 1d	14	4.164	3.00

cation models discussed earlier. As shown in Figure 6(a), the response plot for this top-performing Ensemble Tree model highlights its effectiveness in estimating concrete compressive strength. Although the model demonstrates impressive accuracy, there remains potential for further refinement and optimization. In Figure 6(b), the predicted compressive strength values are compared to the actual measured values, which reveals that the model generally provides accurate predictions. Most of the data points are situated near the diagonal, indicating a close match between the predicted and observed values. However, a few discrepancies can still be observed, suggesting that some improvements in model performance could be made with additional fine-tuning or more diverse training data.

It is important to note that not all discrepancies have the same implications. If the predicted compressive strength is higher than the actual value, the prediction is less accurate but still leans in favor of safety, as it would not undermine structural integrity. On the other hand, when the predicted strength is lower than the actual value, it introduces a potential safety concern, as the structure may be underestimated in terms of its durability. In such cases, improvements in prediction accuracy are crucial, or alternatively, a safety factor may be applied to the results to account for this discrepancy. The magnitude of such a safety factor requires a detailed and parametric analysis of the results, which the authors plan to conduct in future developments of this work.

IV. Conclusions

This study explored the application of machine learning algorithms to predict the mechanical properties of concrete, with a specific focus on the consistency class and compressive strength at 28 days. By integrating a dataset that includes real-world mix designs and environmental data from concrete batching plants, we were able to develop models that exhibit strong accuracy and reliability in predicting these key properties.

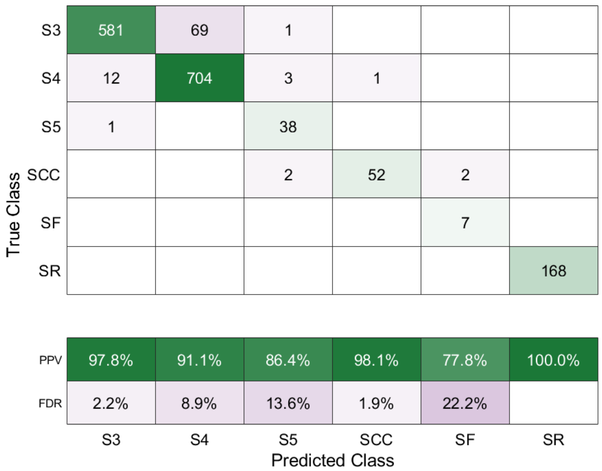
The results demonstrate that machine learning techniques, particularly ensemble methods, show great potential in predicting both the consistency class and compressive strength of concrete. The models developed in this study were able to accurately forecast these properties using mix design data as input. For predicting the consistency class, the best-performing neural network architectures achieved around 80% accuracy. In contrast, the compressive strength predictions, while promising, still require further refinement to improve performance, even though they already compare favorably to traditional human-based methods.

One important aspect to highlight is the presence of uncertainty in the data. The experimental nature of the dataset, with around 1100 samples, introduces variability, especially due to external factors like weather conditions, which can affect the water content in the mix. However, despite these challenges, the models demonstrated notable accuracy, and further exploration of the underlying uncertainties in the data will be essential for future improvements.

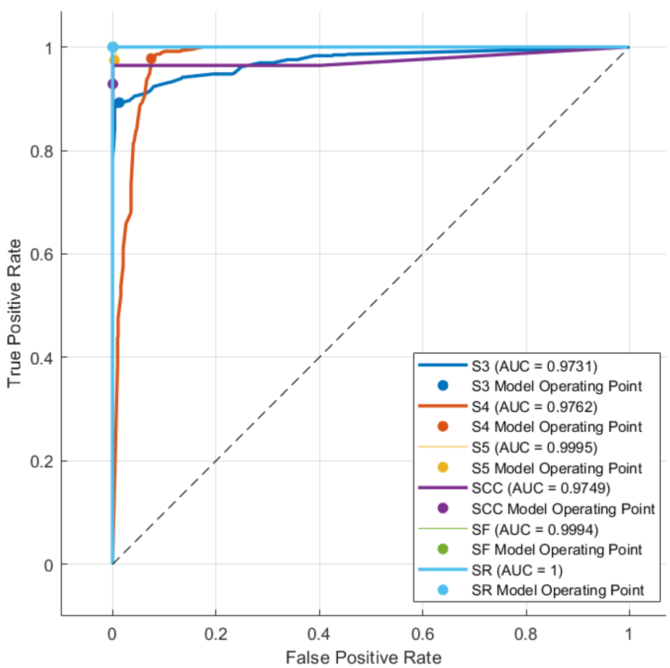
Concrete, being the most widely used construction material globally, plays a crucial role in the building industry due to its versatility, durability, and cost-effectiveness. However, its production contributes significantly to global CO_2 emissions, approximately 7-8%. Therefore, reducing the environmental impact of concrete is critical. By enhancing the accuracy of predictions linking the initial mix design to the final mechanical properties of the concrete, this study contributes to optimizing the design process and minimizing unnecessary material use.

This work represents a pioneering approach to applying machine learning to concrete production. The models developed in this study offer a promising foundation for future research, potentially leading to optimized concrete mix designs that not only improve construction efficiency but also reduce CO_2+ emissions globally. As more data becomes available, more sophisticated machine learning models, such as 1D convolutional neural networks (CNNs), will be explored to further enhance prediction accuracy.

In conclusion, this study demonstrates the significant potential of machine learning in revolutionizing the concrete industry by improving the predictability of its properties, which is a crucial step toward more sustainable construction

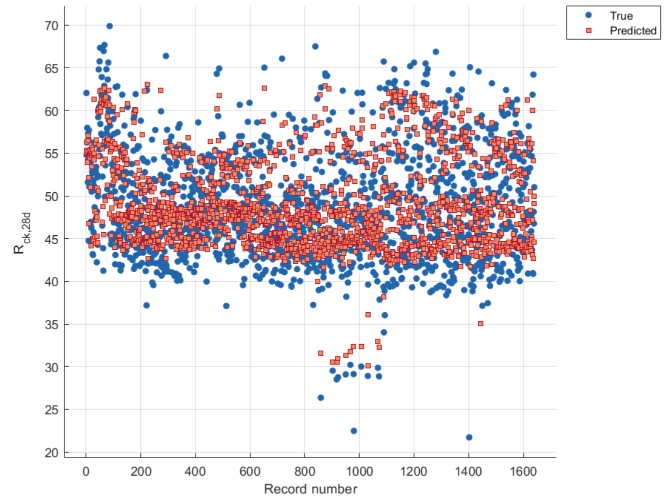


(a) Ensemble tree Confusion Matrix

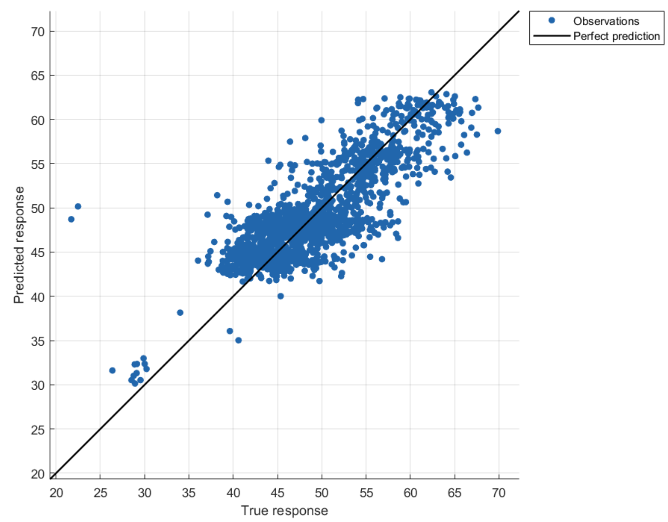


(b) Ensemble tree ROC Curve

Fig. 5: Confusion Matrix (a) and ROC Curves (b) corresponding to the top-performing classification model.



(a) Regression Response Plot



(b) Predicted vs Actual

Fig. 6: (a) Regression response plot for the best-performing Ensemble Tree model, showing the model's predicted compressive strength against the actual values. (b) Predicted vs. Actual plot for the same model, highlighting the relationship between predicted and actual compressive strength values. The closer the data points are to the diagonal, the better the model's prediction accuracy.

practices.

References

[1] E. Commission, "European commission: Internal market, industry, entrepreneurship and smes," 2024, accessed: (15/04/2024). [Online]. Available: https://single-market-economy.ec.europa.eu/sectors/construction_en

[2] N. D. Lagaros, "The environmental and economic impact of structural optimization," *Structural and Multidisciplinary Optimization*, vol. 58, no. 4, pp. 1751–1768, 2018.

[3] J. Min, G. Yan, A. M. Abed, S. Elattar, M. Amine Khadimallah, A. Jan, and H. Elhosiny Ali, "The effect of carbon dioxide emissions on the building energy efficiency," *Fuel*, vol. 326, p. 124842, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016236122016854

[4] M. M. Klufallah, M. F. Nuruddin, M. F. Khamidi, and N. Jamaludin, "Assessment of carbon emission reduction for buildings projects in malaysia-a comparative analysis," in *E3S web of conferences*, vol. 3. EDP Sciences, 2014, p. 01016. [Online]. Available: https://doi.org/10.1051/e3sconf/20140301016

[5] M. Arashpour, R. Wakefield, N. Blismas, and J. Minas, "Optimization of process integration and multi-skilled resource utilization in off-site construction," *Automation in Construction*, vol. 50, pp. 72–80, 2015.

[6] J. Melchiorre, A. Manuello, F. Marmo, S. Adriaenssens, and G. Marano, "Differential formulation and numerical solution for elastic arches with

variable curvature and tapered cross-sections," *European Journal of Mechanics-A/Solids*, vol. 97, p. 104757, 2023.

[7] A. M. Bertetto, J. Melchiorre, and G. C. Marano, "Improved multi-body rope approach for free-form gridshell structures using equal-length element strategy," *Automation in Construction*, vol. 161, p. 105340, 2024.

[8] J. Melchiorre, S. Soutiropoulos, A. Manuello Bertetto, G. C. Marano, and F. Marmo, "Grid-shell multi-step structural optimization with improved multi-body rope approach and multi-objective genetic algorithm," in *Italian Workshop on Shell and Spatial Structures*. Springer, 2023, pp. 62–72.

[9] J. Melchiorre, A. Manuello Bertetto, S. Adriaenssens, and G. C. Marano, "Form-finding and metaheuristic multiobjective optimization methodology for sustainable gridshells with reduced construction complexity and waste," *Available at SSRN 5073825*, 2025.

[10] J. Melchiorre, L. D'Amato, F. Agostini, and A. M. Rizzo, "Acoustic emission onset time detection for structural monitoring with u-net neural network architecture," *Developments in the Built Environment*, vol. 18, p. 100449, 2024.

[11] J. Melchiorre, A. Manuello Bertetto, M. M. Rosso, and G. C. Marano, "Acoustic emission and artificial intelligence procedure for crack source localization," *Sensors*, vol. 23, no. 2, p. 693, 2023.

[12] J. Melchiorre, A. M. Bertetto, S. Invernizzi, G. C. Marano, and F. Bazzucchi, "Dakar mosque gridshell: Exploring the benefits of the improved multibody rope approach through postbuckling analysis," in *Structures*, vol. 61. Elsevier, 2024, p. 106059.

[13] C. Fetting, "The european green deal," *ESDN Report, December*, 2020.

[14] P. J. Monteiro, S. A. Miller, and A. Horvath, "Towards sustainable concrete," *Nature materials*, vol. 16, no. 7, pp. 698–699, 2017.

[15] I. York and I. Europe, "Concrete needs to lose its colossal carbon footprint," *Nature*, vol. 597, no. 7878, pp. 593–594, 2021.

[16] G. Bertagnoli, F. Tondolo, F. Canonico, C. Anerdi, L. Buzzi, and M. Giuseppe, "Calcium sulfoaluminate based concrete - mechanical characterization," *IOP Conference Series: Materials Science and Engineering*, vol. 471, p. 032006, 2018.

[17] S. Sgobba, G. C. Marano, M. Borsa, and M. Molfetta, "Use of rubber particles from recycled tires as concrete aggregate for engineering applications," in *2nd International conference on sustainable construction materials and technologies*, 2010.

[18] A. Fiore, G. C. Marano, P. Monaco, and A. Morbi, "Preliminary experimental study on the effects of surface-applied photocatalytic products on the durability of reinforced concrete," *Construction and Building Materials*, vol. 48, pp. 137–143, 2013.

[19] J. Melchiorre, M. M. Rosso, R. Cucuzza, E. D'Alto, A. Manuello, and G. C. Marano, "Deep acoustic emission detection trained on seismic signals," in *Applications of Artificial Intelligence and Neural Systems to Data Science*. Springer, 2023, pp. 83–92.

[20] P. Ziolkowski and M. Niedostatkiewicz, "Machine learning techniques in concrete mix design," *Materials*, vol. 12, no. 8, p. 1256, 2019.

[21] S. L. Matthews, "fib model code 2020: Enabling advances for new and existing concrete structures," *Structural Concrete*, vol. 24, no. 4, pp. 4341–4351, 2023.