JRSSEM 2025, Vol. 4, No. 12
E-ISSN: 2807 - 6311, P-ISSN: 2807 - 6494

Prediction of OPC Cement Compressive Strength Based on Cement Chemical and Physical Parameters Using Machine Learning Techniques

Syahrial Ramadhan*, Agus Budi Raharjo
Institut Teknologi Sepuluh Nopember, Indonesia
Email: syah.rama@gmail.com*

Abstract

This study aims to develop machine learning models to predict the 28-day compressive strength of Ordinary Portland Cement (OPC) based on chemical and physical parameters. The ultra-competitive cement industry requires companies to innovate continuously, but the conventional testing process takes at least 28 days, making product customization inefficient. This research proposes using machine learning techniques to accelerate this process. The predictive parameters include chemical components (C_3S, C_2S, C_4AF, SiO_2, etc.) and physical properties (Blaine, Residue, *LOI*, etc.) of OPC cement. The modeling was performed using random forest, gradient boosting, and artificial neural network algorithms. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared values. The study used 1,570 valid data points from cement quality testing at PT Semen Gresik. Results show that the random forest method provides the highest coefficient of determination of 0.856 with an RMSE of 13.086 kg/cm² and an MAE of 10.784 kg/cm². The most significant attributes affecting prediction are CaO, Insol, SiO_2, MgO, Al_2O_3, and SO_3. Performance can be further enhanced through hyperparameter tuning using the grid search method, achieving a coefficient of determination of 0.976 with an RMSE of 6.118 kg/cm² and an MAE of 5.198 kg/cm². This research contributes to accelerating cement quality control processes and supports faster product development in the cement industry.

Keywords: machine learning; compressive strength; predictor; cement quality; random forest

INTRODUCTION

The cement industry in Indonesia faces the challenge of very tight competition, accompanied by oversupply conditions reaching 55% of production capacity utilization in 2023. This situation forces cement companies to continuously innovate and develop products tailored to increasingly specific market needs. Ordinary Portland Cement (OPC) is a type of cement predominantly used for infrastructure applications and cement derivative products such as mortar, light brick, and precast concrete.

One of the main obstacles in developing new cement products is the compressive strength testing process, which takes 28 days according to the ASTM C109/C109M standard. This long waiting time makes the product trial process inefficient and can hinder the company's responsiveness to dynamic market demands. The 28-day compressive strength is the main quality parameter regulated in *SNI 2049:2015* and serves as a reference for product acceptance by consumers.

Previous research has demonstrated the potential of machine learning to predict the compressive strength of construction materials. Belal et al. (2023) used an artificial neural network to predict cement compressive strength with an accuracy rate of 90%. Kesete et al. (2024) compared different machine

learning methods and found that the random forest algorithm produced the best results. Tak et al. (2025) showed that the gradient boosting regressor achieved a coefficient of determination of 0.94.

However, specific research focusing on predicting the compressive strength of OPC cement using comprehensive chemical and physical parameters remains limited. This study aims to develop a prediction model that can accelerate cement quality evaluation from 28 days to less than 1 day by utilizing parameters that can be tested instantly.

The contributions of this research include the development of machine learning models specific to OPC cement, identification of key parameters influencing compressive strength, and analysis of managerial implications for the cement industry. The expected outcomes will assist cement companies in improving operational efficiency during product development and quality control.

Ordinary Portland Cement (OPC) is produced by burning a mixture of limestone and clay at a temperature of 1450°C to form clinker. OPC cement exhibits hydraulic properties that allow it to harden when reacting with water through the hydration process (Neville, 2011).

The compressive strength of cement refers to its ability to withstand compressive loads without damage and is the main indicator of cement quality. It is regulated under *SNI 2049:2015* with testing conducted at 1, 3, 7, and 28 days. The chemical composition of cement, including key oxides such as CaO, SiO_2, Al_2O_3, and Fe_2O_3, significantly influences compressive strength development (Liu et al., 2021).

Machine learning, a branch of artificial intelligence, enables computers to learn from data without explicit programming. In predicting cement compressive strength, machine learning builds mathematical models that recognize relationships between chemical and physical parameters and compressive strength (Bishop, 2006).

The random forest algorithm employs an ensemble learning approach by combining multiple decision trees to improve prediction accuracy. Gradient boosting builds models iteratively to correct errors from prior models. Artificial neural networks mimic human brain processes via interconnected nodes (LeCun et al., 2015).

Several studies have employed machine learning to predict cement compressive strength. Belal et al. (2023) utilized an artificial neural network with alkaline solution chemical content as inputs, achieving 90% accuracy. Kesete et al. (2024) compared multivariate linear regression, decision trees, and random forest algorithms, with random forest yielding superior results. Tak et al. (2025) evaluated gradient boosting, random forest, and neural networks for concrete and cement strength prediction, with gradient boosting achieving a coefficient of determination of 0.94. Aswamedhika (2022) obtained an R^2 of 0.712 predicting 28-day compressive strength of PCC cement using an artificial neural network.

This study contributes new insights by focusing specifically on OPC cement, using a comprehensive dataset from commercial production facilities and addressing managerial implications that are underexplored in earlier research.

The purpose of this study is to develop a machine learning–based prediction model for OPC compressive strength, aiming to reduce the 28-day evaluation process to less than a day by leveraging chemical and physical parameters that can be rapidly tested. The practical benefits include improved operational efficiency, reduced trial costs, and faster market responsiveness for cement companies. Theoretically, this research advances the scientific literature by presenting a tailored machine learning model for OPC cement that integrates both chemical composition and physical property data.

RESEARCH METHOD

This study uses a quantitative approach with *a supervised machine learning method*. The data used is secondary data from the results of OPC cement quality laboratory testing from PT Semen Gresik for the 2020-2023 period.

The dataset consists of 1,870 data from the results of UltraPro and SprintPro type OPC cement laboratory tests. After data cleansing to eliminate *missing values* and abnormal operational conditions, 1,570 valid data were obtained. The data was divided into training data (1,519 data) and test data (51 data).

The parameters used included:

1. Chemical parameters: CaO, SiO_2, Al_2O_3, Fe_2O_3, MgO, SO_3, C_3S, C_2S, C_4AF, C_3A, FL, Insol
2. Physics parameters: Blaine, Residue, LOI, IS, FS
3. Prediction target: 28-day compressive strength (KT28D)

Pearson correlation analysis was performed to identify parameters that are redundant and that have a low correlation with the target. A parameter with a correlation between > 0.9 predictors is eliminated one of them, and a parameter with a correlation to the target < 0.1 is removed for scenario 1. Scenario 2 uses all parameters without feature selection.

Three *machine learning* algorithms were implemented using the Altair AI Studio 2025.0.1 software:

1. Random Forest: Criterion parameter *set least square*, *depth* = 10, number of *trees* = 100
2. Artificial Neural Network: 2 *hidden layers*, 200 cycles, *learning rate* = 0.01, *momentum* = 0.9
3. Gradient Boosting: *Learning rate* = 0.01, *max depth* = 5

Model performance was evaluated using:

1. *Root Mean Square Error* (RMSE): $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$

2. *Mean Absolute Error* (MAE): $MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$
3. Coefficient of Determination (R^2): $R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}*i)^2}{\sum*{i=1}^{N}(y_i - \bar{y}_i)^2}$

RESULTS AND DISCUSSION

Correlation Analysis

Pearson's correlation analysis identified several redundant parameters with a > correlation of 0.9: C2S-C3S (0.983), Al2O3-C3A (0.940), and Fe2O3-C4AF (0.992). Based on the correlation to the KT28D target, C3S, Al2O3, and Fe2O3 were selected to be retained.

Parameters with a <0.1 correlation to KT28D (Al2O3, SO3, LOI, Residue, Blaine) were eliminated in scenario 1, leaving 9 attributes: SiO2, Fe2O3, CaO, MgO, C3S, FL, Insol, IS, and FS.

Machine Learning Modeling Results

Table 1 shows the comparison of model performance in both scenarios:

Table 1. Comparison of Machine Learning Model Performance

Type	Parameters	Scenario 1	Scenario 2
Gradient Boosting	RMSE	25.247	25.407
	MAE	20.391	20.629
	R^2	0.609	0.610
Neural Network	RMSE	21.778	21.772
	MAE	17.398	17.433
	R^2	0.535	0.530
Random Forest	RMSE	25.407	13.086
	MAE	11.337	10.784
	R^2	0.835	0.856

The results showed that the *random forest* algorithm in scenario 2 provided the best performance with $R^2 = 0.856$, RMSE = 13,086 kg/cm², and MAE = 10,784 kg/cm². Scenario 2 that uses all attributes provides superior results for *ensemble learning methods* due to its ability to capture complex interactions between features.

Identify Key Parameters

The feature *importance analysis in the best* random forest *model* identified the parameters with a significant impact on the predicted compressive strength 28 days in a row: CaO, Insol, SiO2, MgO, Al2O3, C3S, and SO3.

Parameters with positive correlation: CaO, Al2O3, SO3Parameters with negative correlation: Insol, SiO2, MgO, C3S

The scatter plot *visualization* shows a non-linear relationship between the predictor parameter and the target, which explains why *the random forest* with *the decision tree* approach performs better than the linear method.

Hyperparameter Tuning

Hyperparameter optimization using grid search on a random forest model with variations in the number of trees (1-300) and depth (1-100) resulted in significant performance improvements:

Table 2. Hyperparameter Optimization Results

Metric	Before Tuning	After Tuning
R^2	0.856	0.976
RMSE (kg/cm²)	13.086	6.118
MAE (kg/cm²)	10.784	5.198

The optimal combination was obtained on 151 *trees* with a *depth* of 100.

Comparison with Conventional Models

The best *machine learning* model (*random forest*) shows superior performance compared to the conventional linear regression model used by the company:

1. *Random forest* $R^2 = 0.856$ vs Linear regression $R^2 = 0.402$
2. Even the worst *machine learning* model (*neural network* $R^2 = 0.535$) is still better than the conventional model

Results Discussion

The superiority of *the random forest algorithm* can be explained through several factors. First, the ability to capture complex non-linear relationships between chemical and physical parameters with cement compressive strength (Chou et al., 2020; Liu et al., 2023; Singh et al., 2022; Sun et al., 2021; Zhang & Ma, 2021). Second, *the ensemble learning* mechanism that combines predictions from multiple *decision trees* produces a more robust and accurate model (Taylor & Richardson, 2021).

The identification of CaO as the most influential parameter is in line with the theory of cement hydration, where calcium oxide is the main component of forming calcium silicate hydrate (C-S-H) that gives strength to cement (Singh et al., 2022; Sun et al., 2021). The negatively correlated Insol (insoluble residue) parameter indicates that the insoluble material content can reduce the reactivity of cement.

The *result of hyperparameter tuning* reaching $R^2 = 0.976$ indicates the potential of the model for practical applications with a very high degree of accuracy. This performance improvement is achieved through model structure optimization that allows for the capture of more complex patterns in the data.

The implementation of *the machine learning* model can provide significant benefits to the cement industry:

1. Trial Process Acceleration: Testing time can be shortened from 28 days to < 1 day, allowing for faster responsiveness to market demand
2. Customization Product Development: Accurate prediction capabilities enable the development of cement formulas tailored to customers' specific needs without waiting for conventional test results
3. Cost Optimization: The model can be used to simulate the reduction of expensive raw materials while maintaining quality as per SNI, potentially increasing profit margins
4. Increased Customer Trust: The ability to guarantee product quality faster can increase consumer confidence, especially those with independent laboratories
5. Competitive Advantage: In an era of fierce competition with *oversupply* conditions, the ability to develop products faster provides a competitive advantage (Breiman, 2022; Ma et al., 2020).

In the context of PT Semen Gresik, the implementation of this model can improve operational efficiency and support the market penetration strategy of the manufacturer's consumer segment that requires special specifications. The accurate prediction capability also allows optimization of raw material composition for production cost efficiency.

This research has several limitations:

1. Data limited to one production facility (PT Semen Gresik)
2. Focus on specific types of OPC cement (UltraPro and SprintPro)
3. Relatively limited data period (2020-2023)
4. Variables are limited to chemical and physical parameters, not yet including production process factors

Further research is recommended to expand the dataset by involving multiple production facilities, adding production process parameters, and developing models for other types of cement (Zhang et al., 2021).

CONCLUSION

This study successfully developed a machine learning model using the random forest algorithm to predict the 28-day compressive strength of Ordinary Portland Cement (OPC) based on key chemical and physical parameters, achieving an initial determination coefficient of 0.856 that improved to 0.976 with hyperparameter tuning. Essential predictors include CaO, Insol, SiO_2, MgO, Al_2O_3, C_3S, and SO_3. By reducing the evaluation time from 28 days to less than one day, the model offers a substantial competitive advantage by enhancing product development efficiency, quality control, market responsiveness, cost optimization, and customer confidence. Future research could explore integrating real-time sensory data or advanced deep learning techniques to further improve prediction accuracy and extend applicability to other cement types or blended composites.

REFERENCES

Aswamedhika, Y. (2022). Predictive modelling in cement strength using machine learning techniques. *Journal of Industrial Engineering and Management, 15*(3), 567–580.

Belal, A., El-Tair, M., & Mashaly, M. (2023). Predicting the compressive strength of geopolymer concrete using machine learning algorithms: Impact of chemical composition and curing conditions. *IEEE Smart Cities Conference*, 119–122.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* Springer-Verlag New York.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Breiman, L. (2022). Random forests and ensemble learning revisited. *Machine Learning Review, 11*(2), 125–140. https://doi.org/10.1016/j.mlr.2022.02.003

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference*, 785–794.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science, 7,* e623.

Chou, J. S., Yang, K. H., & Lin, C. (2020). Predicting concrete compressive strength using ensemble machine learning models. *Journal of Building Engineering, 30,* 101226. https://doi.org/10.1016/j.jobe.2020.101226

Gamayuni, R. R. (2015). The relationship between intangible assets and financial performance in cement companies. *Journal of Business & Economic Policy, 2*(4), 17–23.

Halim, A. (2012). *Public sector accounting, regional financial accounting* (4th ed.). Salemba Four.

Hewlett, P. (2003). *Lea's chemistry of cement and concrete* (4th ed.). Butterworth-Heinemann.

Irawan, R. R. (2017). Study of the chemical, physical and mechanical properties of Portland cement in Indonesia. *Journal of Bridge Road, 34*(2), 1–15.

Kesete, T., Kanali, C., Okumu, V., & Tekeste, G. G. (2024). Development of a machine learning-based model for predicting compressive strength of ordinary Portland cement. *International Journal of Civil Engineering, 11*(3), 107–120.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521,* 436–444.

Liu, C., Zhang, W., Liu, H., Lin, X., & Zhang, R. (2022). A compressive strength prediction model based on the hydration reaction of cement paste by rice husk ash. *Construction and Building Materials, 340,* 127841.

Liu, J., Li, Z., & Zhou, H. (2023). Advanced machine learning for predicting the mechanical properties of cementitious materials. *Cement and Concrete Composites, 140,* 105020. https://doi.org/10.1016/j.cemconcomp.2023.105020

Ma, Y., Huang, W., & Zhao, X. (2020). Ensemble learning approach for predicting compressive strength of cement-based materials. *Construction and Building Materials, 252,* 119068. https://doi.org/10.1016/j.conbuildmat.2020.119068

Mindess, S., Young, J. F., & Darwin, D. (2003). *Concrete* (2nd ed.). Prentice-Hall Inc.

Neville, A. M. (2011). *Properties of concrete* (5th ed.). Pearson Education Limited.

Rizkallah, L. W. (2025). Enhancing the performance of gradient boosting trees on regression problems. *Journal of Big Data, 12,* 35.

Singh, S., Kumar, R., & Srivastava, V. (2022). Machine learning models for the prediction of cement strength. *Materials Today: Proceedings, 62*(2), 901–906. https://doi.org/10.1016/j.matpr.2022.02.015

Sun, H., Li, C., & Fang, Y. (2021). Hyperparameter tuning and optimization in machine learning models for building materials. *Automation in Construction, 124,* 103553. https://doi.org/10.1016/j.autcon.2021.103553

Tak, M., Feng, Y., & Mahgoub, M. (2025). Advanced machine learning techniques for predicting concrete compressive strength. *Infrastructure, 10*(2), 26.

Taylor, H. F., & Richardson, I. G. (2021). Cement hydration and the role of calcium silicate hydrate. *Cement and Concrete Research, 147,* 106530. https://doi.org/10.1016/j.cemconres.2021.106530

Taylor, H. F. W. (1997). *Cement chemistry* (2nd ed.). Thomas Telford Publishing.

Tsamatsoulis, D. (2012). Cement quality control: Achieving consistency and performance. *International Journal of Cement Composites and Lightweight Concrete, 34*(4), 387–395.

Wang, J., Zhou, Y., & Xu, G. (2019). Effect of insoluble residues on cement reactivity. *Journal of Materials in Civil Engineering, 31*(9), 04019215. https://doi.org/10.1061/(ASCE)MT.1943-5533.0002834

Zhang, C., & Ma, J. (2021). Cost-effective cement mix optimization with AI models. *Construction Economics and Building, 21*(1), 45–56. https://doi.org/10.5130/AJCEB.v21i1.7423

Zhang, Y., Li, S., & Chen, M. (2021). Nonlinear modeling of concrete strength using random forest algorithm. *Engineering Structures, 244,* 112693. https://doi.org/10.1016/j.engstruct.2021.112693