MEDIA KOMUNIKASI TEKNIK SIPIL
Jurnal Ilmu dan Terapan Bidang Teknik Sipil
Website : mkts.sipil.undip.ac.id
e-ISSN : 2549 - 6778
doi: 10.14710/mkts.v31i1.70716

An Artificial Intelligence-Based Model for Geopolymer Concrete Strength Prediction

*Riqi Radian Khasani[1], Ferry Hermawan[1], Akhmad Firdos Khoiril Khitam[2]
[1]Department of Civil Engineering, Diponegoro University, Indonesia
[2]Department of Civil and Construction Engineering, NTUST, Taiwan
*) riqi@live.undip.ac.id

Received: 30 January 2025 Revised: 26 May 2025 Accepted: 21 July 2025

Abstract

Geopolymer concrete (GPC) has emerged as a sustainable alternative to conventional concrete, offering reduced carbon emissions and enhanced mechanical properties. However, variability in compressive strength due to material composition poses challenges to its broader adoption. Traditional evaluation methods are often time-consuming and resource-intensive, necessitating the development of precise and efficient predictive tools. This study introduces the optimized least squares moment balanced machine with feature selection (OLSMBM-FS), an advanced AI-based model for accurately predicting GPC compressive strength. The model incorporates backpropagation neural networks (BPNN) for weight assignment, least squares support vector machines (LSSVM) for hyperplane optimization, and the optical microscope algorithm (OMA) for hyperparameter tuning. The study employs a systematic dataset, implementing normalization and feature selection techniques to improve the accuracy and efficiency of the model training process. The OLSMBM-FS was validated using 10-fold cross-validation and demonstrated superior performance compared to other machine learning models. It achieved the lowest RMSE (4.279), MAE (2.291), and MAPE (6.59%), alongside the highest R (0.901) and R^2 (0.813), confirming its robustness and predictive accuracy. These findings highlight the potential of OLSMBM-FS as a reliable tool for predicting GPC compressive strength, supporting its broader application in sustainable construction practices.

Keywords: *Geopolymer strength, machine learning, feature selection, compressive strength, sustainable construction*

Abstrak

Beton geopolimer (GPC) telah muncul sebagai alternatif berkelanjutan untuk beton konvensional, menawarkan pengurangan emisi karbon dan peningkatan sifat mekanis. Namun, variabilitas kekuatan tekan akibat komposisi material menjadi tantangan dalam penerapan yang lebih luas. Metode evaluasi tradisional sering kali memakan waktu dan sumber daya, sehingga diperlukan pengembangan alat prediksi yang presisi dan efisien. Studi ini memperkenalkan optimized least squares moment balanced machine with feature selection (OLSMBM-FS), sebuah model berbasis AI yang canggih untuk memprediksi kekuatan tekan GPC secara akurat. Model ini mengintegrasikan Backpropagation Neural Networks (BPNN) untuk penentuan bobot, least squares support vector machines (LSSVM) untuk optimasi hiperplane, dan optical microscope algorithm (OMA) untuk penyesuaian hiperparameter. Penelitian ini menggunakan dataset sistematis, menerapkan teknik normalisasi dan seleksi fitur untuk meningkatkan akurasi dan efisiensi proses pelatihan model. Validasi OLSMBM-FS dilakukan menggunakan 10-fold cross-validation dan menunjukkan kinerja yang unggul dibandingkan dengan model pembelajaran mesin lainnya. Model ini mencapai RMSE terendah (4.279), MAE (2.291), dan MAPE (6.59%), serta nilai R tertinggi (0.901) dan R^2 (0.813), yang mengonfirmasi kekuatan dan akurasi prediksinya. Temuan ini menyoroti potensi OLSMBM-FS sebagai alat yang andal untuk memprediksi kekuatan tekan GPC, mendukung penerapannya dalam praktik konstruksi berkelanjutan.

Kata kunci: *Kekuatan geopolimer, pembelajaran mesin, pemilihan fitur, kekuatan tekan, konstruksi berkelanjutan*

Introduction

Geopolymer concrete (GPC) has gained recognition as a viable substitute for conventional concrete due to its potential to enhance sustainability and significantly reduce carbon emissions. Unlike traditional concrete that relies on Portland cement, GPC is produced using aluminosilicate-rich materials such as fly ash, metakaolin, or slag. These materials are combined with alkaline activators to form a strong and durable binder matrix (Tchadjie and Ekolu, 2018). This material surpasses the mechanical properties and durability of ordinary Portland cement concrete, while significantly lowering the carbon footprint associated with cement production. Additionally, GPC exhibits exceptional resistance to high temperatures and aggressive chemical environments, along with low shrinkage, which minimizes cracking risks and enhances long-term structural performance (Wong, 2022). Despite these advantages, GPC faces challenges, particularly the variability in its mechanical properties, which is heavily influenced by the composition of raw materials. Factors such as the type and proportion of aluminosilicate precursors, alkali solution concentration, and curing conditions significantly affected the compressive strength. Conventional evaluation methods are time-consuming, costly, and require extensive laboratory testing, highlighting the need for accurate and efficient predictive methods to facilitate broader adoption in sustainable construction (Alaneme et al., 2023).

Machine learning (ML) offers an innovative approach to predicting the mechanical properties of GPC. By leveraging large datasets and identifying complex patterns, ML models can provide accurate predictions and optimize material performance. Recent studies have demonstrated the potential of ML techniques in this area. For instance, Le et al. (2024) developed an ML model to predict the compressive strength of GPC using historical data on material composition (Le et al., 2024). Similarly, Rathnayaka et al. (2024) emphasized the importance of data quality in improving ML prediction reliability (Gad et al., 2024), while Ma et al. (2022) highlighted the need for a deeper understanding of input variables to improve practical applications (Ma et al., 2022). Despite the application of various ML models in compressive strength prediction, significant challenges remain, such as the lack of feature selection mechanisms, data quality variability, and advanced techniques to handle complex relationships within datasets (Rathnayaka et al., 2024). Incorporating feature selection can significantly improve model accuracy by focusing on the most influential variables. Furthermore, incomplete or unrepresentative datasets reduce prediction reliability, necessitating data pre-processing techniques, including normalization and feature selection, to ensure consistent and representative data for the ML model (Wang et al., 2024).

Recent advances in ML methodologies, particularly advanced and hybrid models, show significant potential to overcome these challenges. Predominantly used models include ANN, SVM, and Decision Tree (Amin et al., 2022)(Ahmad et al., 2022)(Ahmad et al., 2021). The least squares moment balance machine (LSMBM) model emerges as a promising and potent framework for addressing the complexities of GPC datasets. The LSMBM excels at capturing non-linear and higher-order relationships within diverse datasets for compressive strength prediction, outperforming traditional ML models in accuracy and generalization (Cheng and Khasani, 2024b). Its ability to handle imbalanced data and provide precise predictions makes it valuable for optimizing mix design and understanding the mechanical behavior of GPC.

The objective of this research is to introduced optimized least squares moment balanced machine with feature selection (OLSMBM-FS) model capable of accurately predicting the compressive strength of geopolymer concrete. The model is trained on a dataset compiled from multiple studies that incorporates various material compositions to enhance its generalizability. Comprehensive data pre-processing, including normalization and feature selection, ensures the quality and consistency of the dataset. Hyperparameter tuning is performed using the optical microscope algorithm (OMA), a metaheuristic optimization technique to optimize model performance. This research advances the field of sustainable construction materials by employing ML techniques to enhance predictive modeling and material optimization. The integration of feature selection and advanced optimization methods enhances the reliability of predictions and provides valuable insights into factors influencing GPC performance. This research advances the GPC field and demonstrates the broader applicability of ML in optimizing sustainable construction practices.

Methodology

Model structure

The OLSMBM-FS incorporates a combination of advanced techniques aimed at improving regression accuracy. In this model, BPNN is employed to generate individualized weights for each data instance, indicating their respective significance within the dataset, as described in Eq. (1). The initial predictions from the BPNN algorithm are used to determine the weight given to each data point (G_k) as defined in Eq. (2). Where Yk denotes the actual target value and $\widehat{Y}k$ represents the predicted output. This inverse formulation ensures that samples with higher prediction errors, indicating lower reliability, are assigned smaller weights, whereas samples with lower prediction errors, reflecting higher reliability, are assigned larger weights. These weights are integrated with the least squares support vector machines (LSSVM) principles to create optimal hyperplane moments. The model's objective function aims to minimize moment imbalance, thereby ensuring optimal conditions for effective regression analysis, as demonstrated in Eq. (3). Where, the input and output variables are denoted by x_k and y_k, respectively. The weight vector, represented as v, is coupled with the regularization parameter λ, which controls the trade-off between model complexity and generalization ability. The term μ corresponds to the moments associated with each data point, while εk denotes the error term. The dataset comprises a total of m data points. To improve the model's predictive performance, the OMA is utilized for optimal tuning of the model parameters (Cheng and Khasani, 2024c).

$$D = (x_1, y_1, G_1), (x_2, y_2, G_2), \dots, (x_k, y_k, G_k)\} \in \mathbb{R}^n \quad (1)$$

$$G_k = \left(\frac{Y_k - \widehat{Y}_k}{Y_k}\right)^{-1} \quad (2)$$

$$\text{Minimize } f(\mathrm{v}, \varepsilon) = \frac{1}{2}||\mathrm{v}||^2 \lambda \frac{1}{2} \textstyle\sum_{k=1}^{m} \mu \quad (3)$$

Where: $\mu = G_k {\varepsilon_k}^2$

Subject to $y_k = \mathrm{v}.\varphi(x_k) + b + \varepsilon_k$

OMA is a new metaheuristic algorithm inspired by the dual-stage magnification process of compound optical microscopes, specifically simulating the behavior of objective lens magnification (global search) and eyepiece magnification (local search) to locate and refine optimal solutions within a given search space (Cheng and Sholeh, 2023). OMA inspired by the dual-stage magnification process of compound optical microscopes. It specifically simulates the behavior of objective lens magnification (global search) and eyepiece magnification (local search) to identify and refine optimal solutions within a given search space. In the initial stage, known as the objective lens magnification phase, OMA conducts a broad global search to extensively explore the solution space. The position of the candidate solution Mi is updated in relation to the current best solution Mbest, as demonstrated in Eq (4).

In the subsequent stage, the eyepiece magnification phase, OMA fine-tunes the solution locally. A neighboring solution Mj is chosen, and the magnification space is defined based on the fitness comparison between Mi and Mj, as illustrated in Eq (5). A new solution is then generated using Eq (6). This iterative process of alternating between the two magnification phases allows OMA to effectively balance exploration and exploitation. This optimization process enables the OLSMBM-FS model to deliver accurate and reliable predictions, demonstrating its robustness and adaptability to handle complex datasets. The architecture of the LSMBM is shown in Figure 1.

$$M_{\text{inew}} = M_i + mr \times 1.40 \times M_{\text{best}} \quad (4)$$

$$space = \begin{cases} M_j - M_i \ \ if \ \ f(M_i) \geq f(M_j) \\ M_i - M_j \ \ if \ \ f(M_i) < f(M_j) \end{cases} \quad (5)$$

$$M_{\text{inew}} = M_i + mr \times 0.55 \times \text{space} \quad (6)$$

Model Adaption

A systematic framework was developed to implement the OLSMBM-FS model for predicting the compressive strength of GPC, organized into five essential stages. The flowchart in Figure 2 provides a detailed overview of the key steps involved in adapting the OLSMBM-FS model. In Step 1, involves constructing a comprehensive geopolymer concrete dataset by collecting and organizing historical data from various sources.

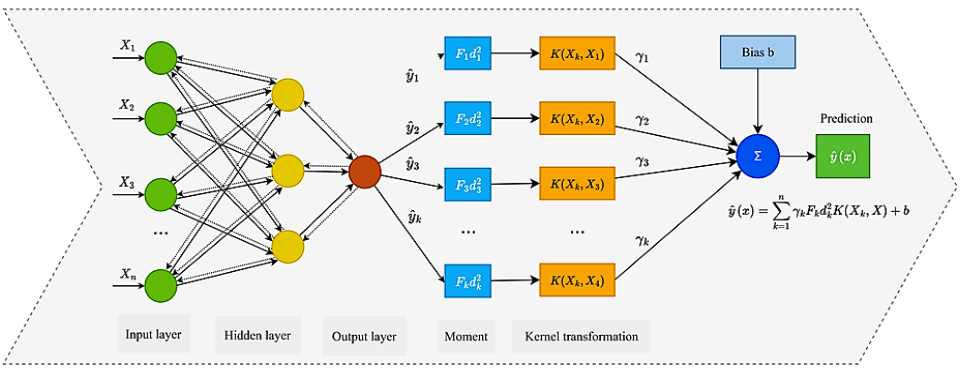


Figure 1. Model architecture

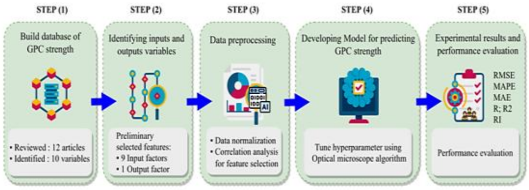


Figure 2. Model framework

The dataset used in this study effectively represents the correlation between the input variables and the compressive strength of GPC, serving as a solid foundation for predictive modeling. In Step 2, the input and output variables were identified to establish the foundation for the predictive modeling process. Critical input features, including material proportions and activator concentrations, were selected based on their known impact on compressive strength of GPC. Step 3 involved data pre-processing procedures to prepare the dataset for effective model development and analysis.

Normalization techniques were applied to the input variables to maintain consistency and allow for fair comparison across different data sources. Feature selection procedures were then implemented to isolate and retain the most relevant variables, contributing to improved model accuracy and computational efficiency. Step 4 involved constructing the predictive model using OLSMBM-FS. The model was trained on the processed dataset to effectively learn the relationship between the selected input and output variables. Advanced optimization techniques were applied to refine the model parameters, aiming to improve its overall predictive performance. In Step 5, The predictive capability of the model was evaluated by analyzing its performance using a variety of metrics to assess its accuracy and reliability. The evaluation results confirmed that the OLSMBM-FS model was capable of accurately predicting the compressive strength of GPC, highlighting its potential as a reliable tool for forecasting material performance.

Data relevant to the compressive strength of GPC were collected and organized for analysis. The dataset was evaluated to identify the critical features and variables that influence the strength of the GPC. During the data pre-processing phase, normalization was applied. To ensure uniformity across variables with different scales and units, all numerical features were normalized to a common scale, typically within the range of 0 to 1. To enhance the robustness and generalization capability of the model, a 10-fold cross-validation method was utilized. This technique involves partitioning the dataset into 10 equal segments. During each iteration, the model is trained using nine segments of the data and tested on the one remaining segment. This procedure is repeated ten times, allowing each segment to function as the validation set exactly once. The performance metrics from all iterations are then averaged to provide a comprehensive assessment of the model's effectiveness. Figure 3 illustrates the structure and workflow of the 10-fold cross-validation method.

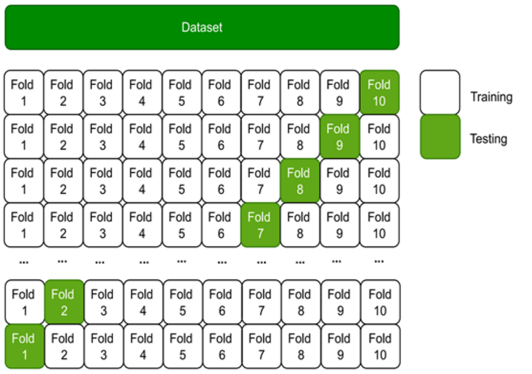


Figure 3. 10-fold cross-validation

Figure 4. illustrates the framework for the OLSMBM-FS model, designed to predict the compressive strength of GPC. The process begins with acquiring a geopolymer concrete dataset, followed by data pre-processing, which includes feature selection and normalization. This ensures consistency and enhances the quality of input data for model training and testing. The dataset is then divided into a training set and a testing set using 10-fold cross-validation strategy to ensure robust evaluation and generalization. The training set is used to develop the OLSMBM-FS model. The BPNN was implemented using default parameter settings in MATLAB, including a single hidden layer with a number of neurons equal to the number of input features, a tansig activation function for the hidden layer, a purelin function for the output layer, and the Levenberg–Marquardt training algorithm.

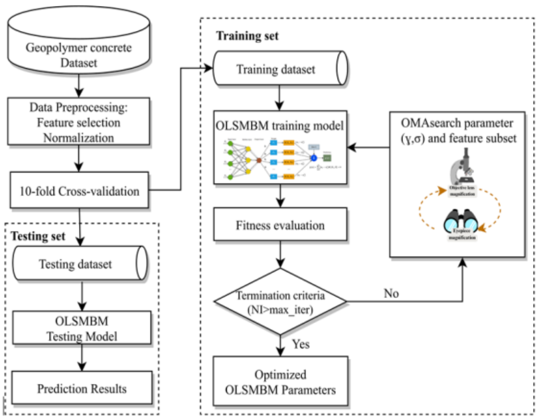


Figure 4. Model adaptation

The OLSMBM-FS training model undergoes an iterative process driven by the OMA algorithm, which searches for the optimal combination of parameters (γ and σ). This search involves fitness evaluation based on the RMSE, continuing iteratively until the maximum termination criteria are achieved. RMSE provides a smooth and continuous error surface, which is advantageous for metaheuristic optimization. Furthermore, RMSE is widely used in the machine learning literature and benchmark studies (Cheng and Khasani, 2024c). Once optimization is complete, the resulting OLSMBM-FS parameters are applied to the test dataset. The OLSMBM-FS testing model then generates prediction results, evaluated using performance metrics to confirm the accuracy and reliability of the model. This framework highlights the integration of advanced machine learning and metaheuristic optimization to enhance the predictive performance of the model.

Performance evaluation

A comprehensive evaluation was carried out to examine the predictive accuracy and overall performance of the machine learning models. This assessment involved the use of multiple performance indicators, including Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), the coefficient of determination (R^2), and the correlation coefficient (R). Furthermore, the Reference Index (RI) was employed to provide a unified performance score by equally weighting the contributions of all five individual metrics (Cheng and Khasani, 2024a). To compute RI, the values of each metric were normalized and the normalized scores were averaged to produce a single index. RI ranges from 0 to 1, with a value of 1 indicating optimal performance across all metrics. This comprehensive approach enables an overall comparison of the effectiveness of the algorithm by capturing both individual and aggregate performance. The detailed mathematical formulations for each performance metric and RI calculation are presented in Table 1.

Results and Discussion

Data collection and pre-processing

The data for this study was compiled by integrating datasets from various sources, including Chindaprasirt and Chalee (Chindaprasirt and Chalee, 2014), Kusbiantoro et al. (Kusbiantoro et al., 2012), Diaz-Loya et al. (Diaz-Loya et al., 2011), Kupwade and Erez (Kupwade-Patil and Allouche, 2013), Lavanya and Jegan (Lavanya and Jegan, 2015), Pane et al. (Pane et al., 2018), Topark-Ngarm et al. (Topark-Ngarm et al., 2015), Nuaklong et al. (Nuaklong et al., 2018), Embong et al. (Embong et al., 2016), Muthadhi and Dhivya (Muthadhi and Dhivya, 2017), Phoo-ngernkham et al. (Phoo-ngernkham et al., 2018), and Mehta and Siddique (Mehta and Siddique, 2017). The dataset comprises nine input and a single output variable, as illustrated in Figure 5.

Figure 5. Factors influencing GPC compressive strength

This study employed correlation analysis to identify relevant variables for predicting the compressive strength of GPC. The correlation matrix in Figure 6 quantifies the linear relationships between the input and target variables. The correlation coefficients (r) and p-values were assessed to ensure statistical significance, retaining variables that contributed significantly to the predictive model. The results of the correlation analysis for nine influencing factors are presented in Table 2. Of these factors, eight met the significance threshold ($p < 0.05$) and were selected as input variables, while one was excluded. Factor 4 exhibited the highest positive correlation, while factor 2 showed a strong negative correlation, both were statistically significant.

Table 1. Performance metrics

Performance metrics
$\text{MAPE} = \frac{100}{n}\sum_i^n \frac{\lvert y_i - f_i \rvert}{y_i}$
$\text{RMSE} = \sqrt{\frac{1}{n}\sum_i^n (y_i - f_i)^2}$
$\text{MAE} = \frac{1}{n}\sum_i^n \lvert y_i - f_i \rvert$
$\text{R} = \frac{n\sum_i^n y_i f_i - (\sum_i^n y_i)(\sum_i^n f_i)}{\sqrt{n(\sum_i^n y_i^2) - (\sum_i^n y_i)^2}\sqrt{n(\sum_i^n f_i^2) - (\sum_i^n fi)^2}}$
$\text{R}^2 = 1 - \frac{\sum_i^n (y_i - f_i)^2}{\sum_i^n (y_i - \bar{y})^2}$

Table 2. Results of correlation analysis

Code	Factor	Unit	Correlation
F1	Fly Ash	kg	◉
F2	Coarse Aggregate	kg	◉
F3	Fine Aggregate	kg	◉
F4	NaOH	kg	◉
F5	Na_2SiO_3	kg	◉
F6	Added Water	(%)	
F7	SiO_2	(%)	◉
F8	Na_2O	M	◉
F9	NaOH Molarity	Mpa	◉

◉ Statistically significant correlation at the 0.05 (two-tailed analysis)

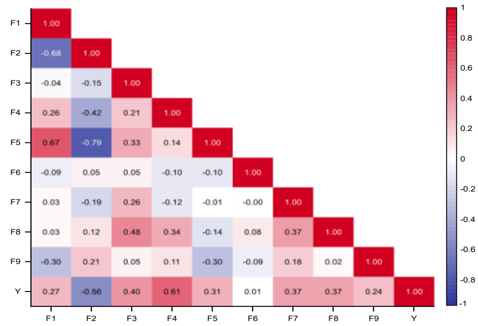


Figure 6. The result of Pearson's correlation method

Factor 6, which failed to meet the significance threshold, was excluded. The remaining factors, including fly Ash (F1), coarse aggregate (F2), fine aggregate (F3), NaOH (F4), Na2SiO3 (F5), SiO2 (F7), Na2O (F8), and NaOH Molarity (F9), were retained as significant predictors of GPC compressive strength. The dataset of 700 cases presented in Table 3 was used to develop the predictive model. Through feature selection, the study improved the predictive reliability and effectiveness of the developed model in estimating the compressive strength of GPC.

The box plot in Figure 7. illustrates the statistical distribution of the input factors (F1 to F8) and the output factor (Y), representing the strength of the GPC. Factors F1, F2, and F3 exhibited the highest variability with wide ranges and significant deviations, indicating their critical influence on the output. F2 showed the largest range, with a mean of 1033.62 and values ranging from 814.81 to 1685.00. In contrast, F6, F7, and F8 displayed narrow ranges and minimal variability.

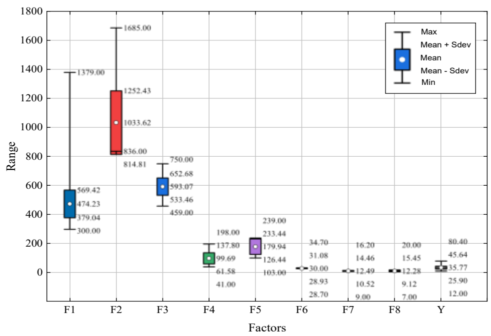


Figure 7. Dataset statistical measurement

The output factor reflects substantial variability, with a mean of 45.64 and values ranging from 12.00 to 80.40. Significant differences in the scale and variability of input factors indicate the necessity of normalization before training the machine learning model. Normalization ensures that all input features are scaled to a comparable range. This pre-processing step is crucial for improving the accuracy and stability of the predictive model and ensuring equitable treatment of all input features during training.

Model testing

The effectiveness of the OLSMBM-FS model in predicting target outcomes was systematically evaluated by benchmarking it against five established ML models: BPNN, SVM, LSSVM, evolutionary least squares inference model (ELSIM), and Linear Regression (LR). This study undertook a comparative analysis to assess the accuracy and efficiency of the OLSMBM-FS model in predicting the compressive strength of GPC. All models were implemented in MATLAB, and their performances were assessed using performance evaluation metrics to provide a comprehensive evaluation of predictive capabilities.

The performance comparison among the six machine learning models revealed that OLSMBM-FS achieved the highest predictive accuracy. It recorded the lowest values for RMSE (4.279), MAE (2.291), and MAPE (6.59%), indicating minimal prediction errors. Additionally, it also achieved the highest R (0.901) and R^2 (0.813) values, reflecting a strong correlation between the target variable and the highest explained variance. Therefore, OLSMBM-FS secures the top rank with the highest RI score (1.000), demonstrating its robustness. ELSIM ranks second with competitive metrics, including an RI of 0.838, followed by LSSVM in third place, with an RI of 0.758.

BPNN and SVM demonstrated moderate performance, while LR showed the weakest predictive capacity. The results highlight the ability of OLSMBM-FS to outperform traditional and other advanced models in predicting the compressive strength of GPC. Its superior metrics underscore its effectiveness in handling the complexities of geopolymer concrete datasets, making it the most reliable model for this comparison.

Figure 8 presents a parallel coordinate plot that visualizes the comparative performance of the six machine learning models across six evaluation criteria. These include RMSE, MAE, MAPE, R, and R^2. Additionally, the RI, which represents the aggregated performance across all five metrics, is also included to provide a comprehensive assessment. Each line in the plot represents the performance of a specific model. In this evaluation, smaller RMSE, MAE, and MAPE values indicate better performance, while higher R, R^2, and RI values are desirable.

Table 3. Dataset of geopolymer concrete

No	F1	F2	F3	F4	F5	F6	F7	F8	Y
1	414	1091	588	69	138	32.9	15.3	10	33.80
2	414	1091	588	69	138	32.9	15.3	15	39.02
3	414	1091	588	69	138	32.9	15.3	20	46.69
4	523	1124	459	118	118	28.7	11.7	10	36.50
5	500	1166	475	113	113	28.7	11.7	10	33.00
...	...	...	...	...	...	...	...	...	...
698	550	838	600	95	239	30	12	14	40.20
699	550	838	600	95	239	30	12	14	39.60
700	550	838	600	95	239	30	12	14	39.00

The OLSMBM-FS demonstrates superior performance across all metrics, achieving the lowest RMSE, MAE, and MAPE while attaining the highest R, R^2, and RI values. These results confirm the robustness and predictive accuracy of OLSMBM-FS in the test dataset.

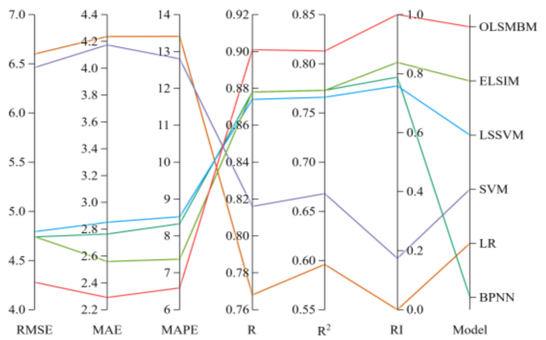
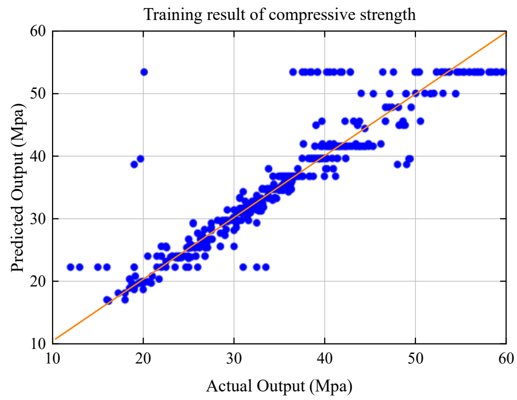


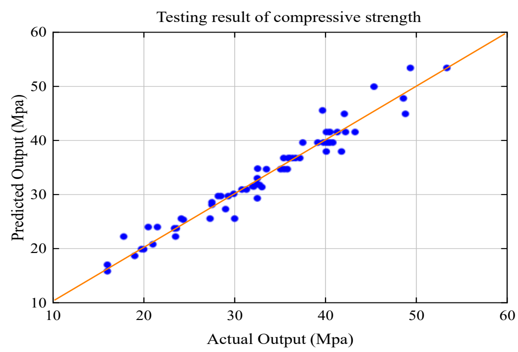
Figure 8. Performance evaluation

In contrast, ELSIM, LSSVM, SVM, LR, and BPNN show varying performance levels. Although ELSIM performs relatively well in some metrics, it falls short compared to OLSMBM-FS. BPNN and LR exhibited the weakest performance, with higher error values and lower correlation metrics, indicating their limitations in capturing complex relationships within the dataset. The parallel plot effectively visualizes these performance differences, highlighting the overall dominance of the OLSMBM-FS in terms of predictive accuracy and generalization. By excelling across all metrics, the OLSMBM-FS proved to be a highly reliable model for predicting the compressive strength of GPC.

Figure 9 illustrates the effectiveness of OLSMBM-FS in predicting the compressive strength of GPC for both the training and testing datasets. The graph on the left depicts the training results, whereas the graph on the right represents the testing results. In both graphs, the actual compressive strength is represented on the x-axis, while the predicted compressive strength is depicted on the y-axis. The orange diagonal line indicates the ideal scenario where the predicted values correspond precisely with the actual values. In the training results, most of the data points aligned closely with the diagonal line, indicating strong predictive accuracy during the training phase. This alignment demonstrates the ability of the OLSMBM-FS to capture complex patterns in the training dataset effectively. Similarly, the testing results similarly show high predictive accuracy, as evidenced by the clustering of data points near the diagonal line. The results confirm the robustness and generalizability of the model, maintaining strong predictive performance across the test dataset.



(a) Training dataset



(b) Testing dataset

Figure 9. Comparison of actual and predicted

Conclusion

This research successfully established the optimized least squares moment balanced machine with feature selection (OLSMBM-FS), a sophisticated artificial intelligence-based inference model designed to accurately predict the compressive strength of geopolymer concrete (GPC). The model incorporates a combination of sophisticated methods. BPNN are employed to assign importance-based weights to individual data points. LSSVM are utilized to determine the most suitable regression hyperplane. To further enhance predictive accuracy, the OMA is used to optimize the model's hyperparameters effectively.

The application of OLSMBM-FS leveraged a systematically dataset containing key input factors that influence the compressive strength of GPC. A data pre-processing, including normalization and feature selection, ensures the consistency and quality of the input data. The performance of the model was validated using a 10-fold cross-validation approach to ensure its generalizability. Comparative analysis with other machine learning models, such as BPNN, SVM, LSSVM, ELSIM, and LR, demonstrated the superior performance of OLSMBM-FS. The model demonstrated outstanding predictive performance by attaining the lowest error values, with RMSE of 4.279, MAE of 2.291, and MAPE of 6.59%. It also recorded the highest R of 0.901 and R^2 of 0.813. Moreover, the model achieved a perfect Reference Index (RI) score of 1.000, underscoring its high accuracy and superior performance relative to the other evaluated models.

The OLSMBM-FS model serves as an effective decision-support tool for predicting the compressive strength of GPC, helping to minimize reliance on expensive and time-intensive laboratory testing procedures. By enabling accurate predictions, the model supports material optimization and efficient resource allocation and facilitates the broader adoption of sustainable construction materials. While the results confirm the robustness and effectiveness of OLSMBM-FS, this study has certain limitations. Although the comparison was conducted with a limited set of machine learning models, future research should expand to include additional techniques to further validate its performance.

References

Ahmad, A., W. Ahmad, F. Aslam, and P. Joyklad. (2022). Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. *Case Studies in Construction Materials*, 16: e00840. https://doi.org/10.1016/j.cscm.2021.e00840.

Ahmad, A., W. Ahmad, K. Chaiyasarn, K. A. Ostrowski, F. Aslam, P. Zajdel, and P. Joyklad. (2021). Prediction of Geopolymer Concrete Compressive Strength Using Novel Machine Learning Algorithms. *Polymers*, 13 (19): 3389. https://doi.org/10.3390/polym13193389.

Alaneme, G. U., K. A. Olonade, and E. Esenogho. (2023). Critical review on the application of artificial intelligence techniques in the production of geopolymer-concrete. *SN Applied Sciences*, 5 (8): 217. https://doi.org/10.1007/s42452-023-05447-z.

Amin, M. N., K. Khan, M. F. Javed, F. Aslam, M. G. Qadir, and M. I. Faraz. (2022). Prediction of Mechanical Properties of Fly-Ash/Slag-Based Geopolymer Concrete Using Ensemble and Non-Ensemble Machine-Learning Techniques. *Materials*, 15 (10): 3478. https://doi.org/10.3390/ma15103478.

Cheng, M.-Y., and R. R. Khasani. (2024a). Moment balanced machine: a new supervised inference engine for on-site construction productivity prediction. *Applied Intelligence*, (0123456789). Springer US. https://doi.org/10.1007/s10489-024-05419-9.

Cheng, M. Y., and R. R. Khasani. (2024b). *Least Square Moment Balanced Machine: a New Approach To Estimating Cost To Completion for Construction Projects*. *Journal of Information Technology in Construction*.

Cheng, M. Y., and R. R. Khasani. (2024c). A Novel Approach for Advancing Asphalt Pavement Temperature and Flow Number Predictions Using Optical Microscope Algorithm–Least Square Moment Balanced Machine. *Journal of Computing in Civil Engineering*, 38 (6): 1–15. https://doi.org/10.1061/JCCEE5.CPENG-5923.

Cheng, M. Y., and M. N. Sholeh. (2023). Optical microscope algorithm: A new metaheuristic inspired by microscope magnification for solving engineering optimization problems. *Knowledge-Based Systems*, 279: 110939. Elsevier B.V. https://doi.org/10.1016/j.knosys.2023.110939.

Chindaprasirt, P., and W. Chalee. (2014). Effect of sodium hydroxide concentration on chloride penetration and steel corrosion of fly ash-based geopolymer concrete under marine site. *Construction and Building Materials*, 63: 303–310. https://doi.org/10.1016/j.conbuildmat.2014.04.010.

Diaz-Loya, E. I., E. N. Allouche, and S. Vaidya. (2011). Mechanical Properties of Fly-Ash-Based

Geopolymer Concrete. *ACI Materials Journal*, 108 (3). https://doi.org/10.14359/51682495.

Embong, R., A. Kusbiantoro, N. Shafiq, and M. F. Nuruddin. (2016). Strength and microstructural properties of fly ash based geopolymer concrete containing high-calcium and water-absorptive aggregate. *Journal of Cleaner Production*, 112: 816–822. https://doi.org/10.1016/j.jclepro.2015.06.058.

Gad, M. A., E. Nikbakht, and M. G. Ragab. (2024). Predicting the compressive strength of engineered geopolymer composites using automated machine learning. *Construction and Building Materials*, 442: 137509. https://doi.org/10.1016/j.conbuildmat.2024.137509.

Kupwade-Patil, K., and E. N. Allouche. (2013). Impact of Alkali Silica Reaction on Fly Ash-Based Geopolymer Concrete. *Journal of Materials in Civil Engineering*, 25 (1): 131–139. https://doi.org/10.1061/(ASCE)MT.1943-5533.0000579.

Kusbiantoro, A., M. F. Nuruddin, N. Shafiq, and S. A. Qazi. (2012). The effect of microwave incinerated rice husk ash on the compressive and bond strength of fly ash based geopolymer concrete. *Construction and Building Materials*, 36: 695–703. https://doi.org/10.1016/j.conbuildmat.2012.06.064.

Lavanya, G., and J. Jegan. (2015). Durability Study on High Calcium Fly Ash Based Geopolymer Concrete. *Advances in Materials Science and Engineering*, 2015: 1–7. https://doi.org/10.1155/2015/731056.

Le, Q.-H., D.-H. Nguyen, T. Sang-To, S. Khatir, H. Le-Minh, A. H. Gandomi, and T. Cuong-Le. (2024). Machine learning based models for predicting compressive strength of geopolymer concrete. *Frontiers of Structural and Civil Engineering*, 18 (7): 1028–1049. https://doi.org/10.1007/s11709-024-1039-5.

Ma, G., A. Cui, Y. Huang, and W. Dong. (2022). A Data-Driven Influential Factor Analysis Method for Fly Ash–Based Geopolymer Using Optimized Machine-Learning Algorithms. *Journal of Materials in Civil Engineering*, 34 (7). https://doi.org/10.1061/(ASCE)MT.1943-5533.0004266.

Mehta, A., and R. Siddique. (2017). Sulfuric acid resistance of fly ash based geopolymer concrete. *Construction and Building Materials*,146:136–143. https://doi.org/10.1016/j.conbuildmat.2017.04.077.

Muthadhi, A., and V. Dhivya. (2017). Investigating Strength Properties of Geopolymer Concrete with Quarry Dust. *ACI Materials Journal*, 114 (3). https://doi.org/10.14359/51689674.

Nuaklong, P., V. Sata, A. Wongsa, K. Srinavin, and P. Chindaprasirt. (2018). Recycled aggregate high calcium fly ash geopolymer concrete with inclusion of OPC and nano-SiO2. *Construction and Building Materials*, 174: 244–252. https://doi.org/10.1016/j.conbuildmat.2018.04.123.

Pane, I., I. Imran, and B. Budiono. (2018). Compressive strength of fly ash-based geopolymer concrete with a variable of sodium hydroxide (NaOH) solution molarity. *MATEC web of conferences*, 1004.

Phoo-ngernkham, T., C. Phiangphimai, N. Damrongwiriyanupap, S. Hanjitsuwan, J. Thumrongvut, and P. Chindaprasirt. (2018). A Mix Design Procedure for Alkali-Activated High-Calcium Fly Ash Concrete Cured at Ambient Temperature. *Advances in Materials Science and Engineering*, (M. Lavorgna, ed.), 2018 (1). https://doi.org/10.1155/2018/2460403.

Rathnayaka, M., D. Karunasinghe, C. Gunasekara, K. Wijesundara, W. Lokuge, and D. W. Law. (2024). Machine learning approaches to predict compressive strength of fly ash-based geopolymer concrete: A comprehensive review. *Construction and Building Materials*, 419: 135519. https://doi.org/10.1016/j.conbuildmat.2024.135519

Tchadjie, L. N., and S. O. Ekolu. (2018). Enhancing the reactivity of aluminosilicate materials toward geopolymer synthesis. *Journal of Materials Science*, 53 (7): 4709–4733. https://doi.org/10.1007/s10853-017-1907-7.

Topark-Ngarm, P., P. Chindaprasirt, and V. Sata. (2015). Setting Time, Strength, and Bond of High-Calcium Fly Ash Geopolymer Concrete. *Journal of Materials in Civil Engineering*, 27 (7). https://doi.org/10.1061/(ASCE)MT.1943-5533.0001157.

Wang, R., J. Zhang, Y. Lu, S. Ren, and J. Huang. (2024). Towards a Reliable Design of Geopolymer Concrete for Green Landscapes: A Comparative Study of Tree-Based and Regression-Based Models. *Buildings*, 14 (3): 615. https://doi.org/10.3390/buildings14030615.

Wong, L. S. (2022). Durability Performance of Geopolymer Concrete: A Review. *Polymers*, 14 (5): 868. https://doi.org/10.3390/polym14050868