Data-Driven Analysis of Compressive Strength Development in Cementitious Systems Incorporating Slag-Based Non-Carbonated CaO

Bilguun Mend , Youngjun Lee , Jeong-Hwan Bang , Chan Woo Kim , Yong-Sik Chu *

Posted Date: 9 February 2026

doi: 10.20944/preprints202602.0671.v1

Keywords: supplementary cementitious materials; slag cement; compressive strength; multi-output prediction; deep neural networks

 doi:10.20944/preprints202602.0671.v1

Article

Data-Driven Analysis of Compressive Strength Development in Cementitious Systems Incorporating Slag-Based Non-Carbonated CaO

Bilguun Mend, Youngjun Lee, Jeong-Hwan Bang, Chan Woo Kim and Yong-Sik Chu *

Climate and Energy R&D Group, Korea Institute of Ceramic Engineering and Technology, South Korea
* Correspondence: yschu@kicet.re.kr

Abstract

The utilization of slag-based materials as non-carbonated CaO sources has attracted increasing attention as a strategy to reduce limestone consumption and CO_2 emissions in cement manufacturing. However, compressive strength development in cementitious systems incorporating slag and $CaCO_3$ replacement is governed by complex interactions among slag chemistry, mixture design, and physical properties, which makes systematic interpretation and prediction challenging. In this study, a structured dataset was compiled from previously published experimental investigations on clinker and cement systems incorporating various slag types and $CaCO_3$ replacement levels. The dataset integrates slag chemical composition, mixture design parameters, and physical properties with compressive strength measured at 3 and 28 days. Based on observed experimental trends, compressive strength prediction at multiple curing ages was formulated as a multi-output regression problem to explicitly account for the correlated nature of strength development over time. The results reveal clear nonlinear relationships between compressive strength, curing age, and multiple material parameters. Mixed-slag systems generally exhibit higher early-age strength compared to single-slag systems at comparable $CaCO_3$ replacement levels, while differences among systems become less pronounced at later ages. These findings indicate that compressive strength at different curing ages is interrelated and influenced by shared material characteristics rather than independent variables. Overall, this study provides a data-driven analysis of strength development trends in slag-blended cementitious materials and establishes a consistent literature-based dataset framework suitable for multi-age strength modeling. The proposed approach offers a transparent basis for future predictive modeling and optimization of cement systems incorporating slag-based non-carbonated CaO sources.

Keywords: supplementary cementitious materials; slag cement; compressive strength; multi-output prediction; deep neural networks

1. Introduction

The demand for sustainable cementitious materials has intensified in response to the need to reduce CO_2 emissions and natural resource consumption in the cement industry [1–4]. One widely adopted strategy is the partial replacement of clinker with supplementary cementitious materials, such as blast-furnace slag, together with mineral additives including calcium carbonate $CaCO_3$ [5–7]. Slag-based materials are particularly attractive due to their potential to reduce limestone usage while maintaining acceptable mechanical performance, especially at later curing ages [8–12]. However, the compressive strength development of slag–limestone blended cementitious systems is governed by complex interactions among slag chemistry, replacement ratio, and physical properties, making systematic interpretation challenging [13].

Compressive strength remains one of the most critical performance indicators for cement-based materials, as it directly influences structural design and serviceability [14]. Experimental studies have

 doi:10.20944/preprints202602.0671.v1

shown that both early-age and later-age compressive strength are sensitive to multiple interacting parameters, including chemical composition of slag, fineness, and $CaCO_3$ replacement level [15,16]. In particular, increasing $CaCO_3$ replacement generally leads to a reduction in early-age strength due to clinker dilution, while its influence on later-age strength is often less pronounced. Despite extensive experimental efforts, these effects are typically reported within isolated studies focusing on specific material combinations, which limits the extraction of general trends applicable across different cement systems.

In recent years, data-driven modeling techniques have emerged as useful tools for analyzing complex, nonlinear relationships in cementitious materials. Machine learning and deep learning approaches have been applied to predict compressive strength using chemical composition, mixture proportions, and curing age as input features [17,18]. However, most existing studies treat compressive strength at each curing age as an independent prediction target [19]. This single-output formulation neglects the inherent correlation between early-age and later-age strength, even though both responses originate from the same hydration and microstructural development processes [20].

From a materials perspective, compressive strength values measured at different curing ages are not independent properties but rather reflect different stages of a continuous strength evolution process [21]. Early-age hydration behavior influences the formation of the microstructural framework that governs later-age strength development. Consequently, compressive strength at early and later ages is expected to exhibit shared dependencies on material parameters, while also showing age-dependent sensitivity [22]. Modeling these strengths independently may therefore fail to capture the coupled nature of strength development in slag-blended cementitious systems.

At the same time, a large volume of high-quality experimental data on slag- and limestone-containing cement systems has already been reported in the literature. These datasets are often underutilized beyond their original experimental objectives, despite their potential to support broader trend analysis and modeling [23]. Systematically integrating published experimental results into a structured dataset provides an opportunity to extract additional insights, identify general behavior, and explore modeling strategies without the need for extensive new experimentation.

In this context, the present study proposes a data-driven analysis of compressive strength development in cementitious systems incorporating slag-based non-carbonated CaO and $CaCO_3$ replacement. A literature-based dataset is constructed by integrating slag chemical composition, mixture design parameters, and physical properties with compressive strength measured at 3 and 28 days. Based on observed experimental trends, compressive strength prediction at multiple curing ages is formulated as a multi-output regression problem to explicitly account for the correlated nature of strength development over time.

The objective of this study is not to establish a definitive predictive model with optimized accuracy, but rather to demonstrate the feasibility and usefulness of multi-age data-driven modeling for interpreting strength development trends in slag–limestone blended cementitious materials. By combining literature-based data integration with multi-output learning, this work aims to provide a transparent framework for analyzing early-age and later-age strength behavior and to offer insights that can support future experimental design and modeling efforts.

2. Materials and Methods

2.1. Dataset Compilation and Variables

The dataset used in this study was compiled from previously published experimental investigations on cementitious systems incorporating blast-furnace slag and calcium carbonate ($CaCO_3$) as partial cement replacements. Only studies reporting detailed information on material composition, mixture design parameters, and compressive strength were considered to ensure consistency and reliability of the collected data. By integrating experimental results from multiple sources, the constructed dataset covers a broad range of slag compositions and mixture conditions, enabling systematic trend analysis beyond individual experimental studies.

 doi:10.20944/preprints202602.0671.v1

The compiled dataset includes descriptors at both the raw material level and the cement level. At the raw material level, the chemical composition of slag is characterized by the mass fractions of SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, and SO_3, expressed in wt.%. These variables represent the intrinsic chemical characteristics of slag that influence hydration reactions and strength development. At the cement level, mixture design parameters such as the $CaCO_3$ replacement ratio and slag replacement ratio are included, together with physical properties represented by fineness and mortar flow.

Compressive strength measured at curing ages of 3 and 28 days is selected as the output variables of interest. These two curing ages are commonly reported in the literature and represent early-age and later-age mechanical performance, respectively. Rather than treating these strength values as independent responses, they are considered as correlated outputs reflecting different stages of a continuous strength development process.

Table 1 summarizes the statistical characteristics of all input and output variables in the compiled dataset, including the minimum, maximum, mean, and standard deviation values. The $CaCO_3$ replacement ratio investigated in the collected studies ranges from 0 to 12 wt.%, while the slag replacement ratio spans a wide interval depending on the original experimental design. The resulting dataset provides a consistent basis for subsequent data-driven modeling and qualitative interpretation of compressive strength development in slag–limestone blended cementitious systems.

Table 1. Summary statistics of input and output variables used for data-driven modeling.

Chemical composition of slag (raw material level)					
Variable	**Unit**	**Min**	**Max**	**Mean**	**Std.Dev.**
SiO_2	wt. %	10.20	34.81	20.37	12.45
Al_2O_3	wt. %	3.37	23.10	12.33	8.42
Fe_2O_3	wt. %	0.94	34.94	14.73	15.03
CaO	wt. %	35.51	45.00	42.01	3.95
MgO	wt. %	3.34	6.30	4.66	1.23
SO_3	wt. %	0.06	3.87	1.32	1.52
Mixture design and physical properties (cement level)					
Variable	**Unit**	**Min**	**Max**	**Mean**	**Std.Dev.**
$CaCO_3$ replacement ratio	wt.%	0	12	6.0	4.74
Slag replacement ratio	wt.%	0	23.10	10.79	8.02
Fineness	cm^2/g	3610	4066	3860	125
Mortar flow	mm	185	207	196	7
Output variables (compressive strength)					
Compressive strength (3 d)	MPa	12.1	42.8	27.6	8.9
Compressive strength (28 d)	MPa	31.4	71.2	52.3	10.7

Note: The summarized variables represent the input and output features used for subsequent data-driven modeling and analysis.

2.2. Dataset Structure and Representative Examples

The compiled dataset encompasses a variety of cementitious systems, including ordinary Portland cement (OPC) as a reference, single-slag blended cements, and mixed-slag blended cements incorporating different slag types. These systems reflect the diversity of experimental conditions reported in the literature and allow the analysis of compressive strength development across a broad range of material combinations.

For single-slag systems, the dataset includes mixtures containing one dominant slag type with varying $CaCO_3$ replacement ratios and fineness levels. Mixed-slag systems consist of combinations of blast-furnace slag with other industrial slags, such as basic oxygen furnace slag, representing more complex raw material scenarios. This distinction enables the examination of strength development behavior under both relatively simple and more heterogeneous slag compositions.

 doi:10.20944/preprints202602.0671.v1

Rather than presenting the entire dataset in tabular form, representative examples are summarized to illustrate the structure and scope of the data used for modeling. These examples highlight key input variables, including slag type, $CaCO_3$ replacement ratio, fineness, and selected chemical composition indicators, together with the corresponding compressive strength values at 3 and 28 days. This approach provides transparency regarding the dataset organization while avoiding unnecessary redundancy.

Table 2 presents representative dataset entries employed for data-driven modeling. The selected entries demonstrate the variability in mixture design parameters and compressive strength outcomes across different cement systems. It should be noted that these examples are intended to illustrate the dataset structure rather than to serve as an exhaustive listing of all collected data.

Table 2. Representative dataset entries used for data-driven modeling of compressive strength.

System	Slag type	$CaCO_3$ repl.(%)	Fineness (cm^2/g)	Slag CaO (wt.%)	Slag SiO_2 (wwt.%)	3d (MPa)	28d (MPa)
OPC	-	0	3652	-	-	32.2	70.5
Single slag	LF	3	3610	45.00	10.20	29.5	67.9
Single slag	LF	6	3757	45.00	10.20	27.1	67.1
Mixed slag (H)	BF+KR+LF	3	3848	44.26	23.79	34.3	70.5
Mixed slag (H)	BF+KR+LF	6	3855	44.26	23.79	33.9	69.1
Mixed slag (H)	BF+KR+LF	9	3984	44.26	23.79	31.0	62.4
Mixed slag (H)	BF+KR+LF	12	3865	44.26	23.79	29.5	62.8
Mixed slag (P)	BF+BOF	3	3886	39.57	29.02	38.4	68.4
Mixed slag (P)	BF+BOF	6	3851	39.57	29.02	35.9	69.0
Mixed slag (P)	BF+BOF	9	4066	39.57	29.02	32.8	68.6
Mixed slag (P)	BF+BOF	12	3979	39.57	29.02	30.6	63.3

Note: The listed entries represent selected examples from the compiled literature-based dataset and are provided to illustrate the structure of the data used for modeling rather than the complete dataset.

By organizing the dataset in this manner, the present study ensures that the modeling framework is trained on diverse yet systematically structured information. This structure supports the subsequent analysis of strength development trends while maintaining consistency with the heterogeneous nature of literature-based experimental data.

2.3. Modeling Framework

To analyze compressive strength development in slag–limestone blended cementitious systems, a data-driven modeling framework was adopted. The primary objective of the modeling approach was not to achieve maximum predictive accuracy, but to explore general trends in strength development and to investigate the coupled behavior of early-age and later-age compressive strength based on literature-derived experimental data.

The compiled dataset was formulated as a regression problem, in which the input features consisted of slag chemical composition at the raw material level, mixture design parameters, and physical properties at the cement level. The output variables were defined as compressive strength measured at curing ages of 3 and 28 days. Given that these two strength values represent different stages of a continuous hydration and microstructural development process, they were treated as correlated outputs rather than independent prediction targets.

Based on this consideration, a multi-output learning strategy was employed. In this formulation, a single model simultaneously predicts compressive strength at both curing ages, allowing shared information between early-age and later-age responses to be exploited during training. This approach contrasts with conventional single-output models, which independently predict strength at each curing age and may overlook the intrinsic relationship between strength values measured at different times.

 doi:10.20944/preprints202602.0671.v1

Several representative modeling approaches were considered to evaluate the general behavior of data-driven prediction, including linear regression, tree-based ensemble methods, and a multi-task deep neural network. Linear regression serves as a baseline model to assess the limitations of linear assumptions, while tree-based methods are capable of capturing nonlinear interactions among input variables. The multi-task deep neural network is specifically designed to handle multi-output regression by learning shared and task-specific representations for strength prediction at different curing ages.

Figure 1 illustrates the overall workflow of the proposed modeling framework, including data acquisition from the literature, feature definition, multi-output learning, and subsequent evaluation and interpretation. This framework provides a consistent basis for examining strength development trends and assessing the feasibility of multi-age data-driven modeling for complex cementitious material systems.

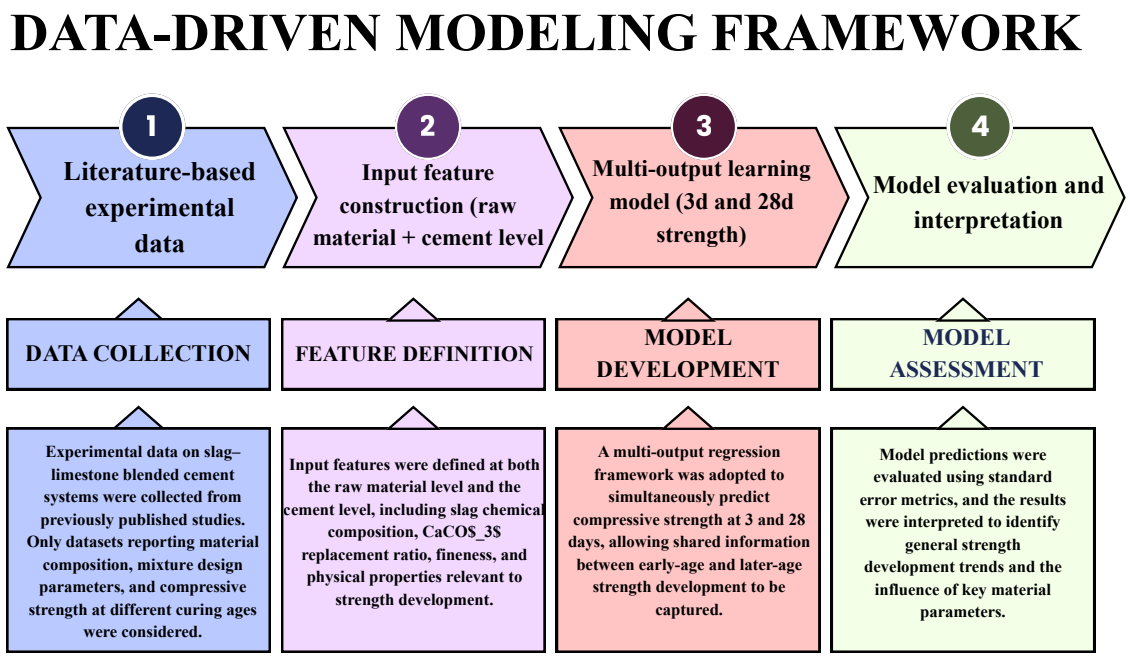

Figure 1. Data-driven modeling framework illustrating the workflow from literature-based data collection and input feature definition to multi-output learning for compressive strength prediction and subsequent model evaluation and interpretation.

2.4. Model Training and Evaluation Strategy

Prior to model training, all input variables were preprocessed to ensure numerical consistency across different feature scales. Continuous input features, including chemical composition parameters, mixture design variables, and physical properties, were normalized to eliminate scale-related bias during model learning. This preprocessing step is particularly important for literature-based datasets, where input variables originate from heterogeneous experimental sources.

The compiled dataset was randomly divided into training and testing subsets. The training set was used to fit the data-driven models, while the testing set was reserved for independent evaluation of model behavior. This strategy allows the generalization capability of the models to be examined without relying on the same data used for training.

Model performance was assessed using commonly adopted error metrics, including mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2). These metrics were selected to provide complementary information on prediction accuracy and overall agreement between predicted and measured compressive strength values. Performance evaluation was

 doi:10.20944/preprints202602.0671.v1

conducted separately for compressive strength at 3 and 28 days to capture age-dependent prediction behavior.

It should be emphasized that the primary purpose of model evaluation in this study is not rigorous performance benchmarking, but rather the identification of general prediction trends and the assessment of the feasibility of multi-output learning for compressive strength analysis. Consequently, model comparison is interpreted in a qualitative and relative manner, focusing on consistency and robustness of prediction behavior across different curing ages rather than on absolute optimization of accuracy.

2.5. Scope and Limitations

The scope of the present study is limited to the analysis of compressive strength development in slag–limestone blended cementitious systems based on previously published experimental data. The constructed dataset integrates results from multiple literature sources, which enables broad trend analysis but also introduces inherent variability associated with differences in experimental conditions, testing standards, and material sources.

It should be noted that the compiled dataset is heterogeneous in nature, reflecting the diversity of slag types, mixture designs, and curing conditions reported in the literature. Although this heterogeneity may influence quantitative prediction accuracy, it is considered appropriate for the primary objective of this study, which is to identify general strength development trends rather than to establish a strictly optimized predictive model.

Furthermore, the analysis focuses on compressive strength measured at curing ages of 3 and 28 days, as these ages are most commonly reported in published studies. Strength development at other curing ages, as well as additional durability-related properties, are not considered in the present framework. Consequently, the applicability of the proposed modeling approach is restricted to early-age and later-age strength behavior within the investigated age range.

Finally, the data-driven models employed in this study are intended to support qualitative interpretation and comparative trend analysis. The modeling results should therefore be interpreted as complementary to experimental investigations rather than as a replacement for detailed laboratory testing. Future studies may extend the proposed framework by incorporating larger datasets, additional curing ages, and more comprehensive material descriptors to further enhance the robustness and interpretability of data-driven analysis in cementitious material research.

3. Results and Discussion

3.1. Strength Development Trends in the Compiled Dataset

The strength development behavior of slag–limestone blended cementitious systems was first examined at the dataset level to establish baseline trends prior to data-driven modeling. Figure 2 presents multiple complementary perspectives on the influence of $CaCO_3$ replacement on compressive strength development at early and later curing ages.

Figure 2(a) shows the absolute compressive strength values measured at curing ages of 3 and 28 days. At early age, compressive strength generally decreases with increasing $CaCO_3$ replacement, indicating a pronounced sensitivity of early-age strength to clinker dilution and limited early hydration of slag. In contrast, 28-day compressive strength exhibits a comparatively milder reduction over the same replacement range, suggesting partial compensation through continued slag hydration and filler effects at later ages.

 doi:10.20944/preprints202602.0671.v1

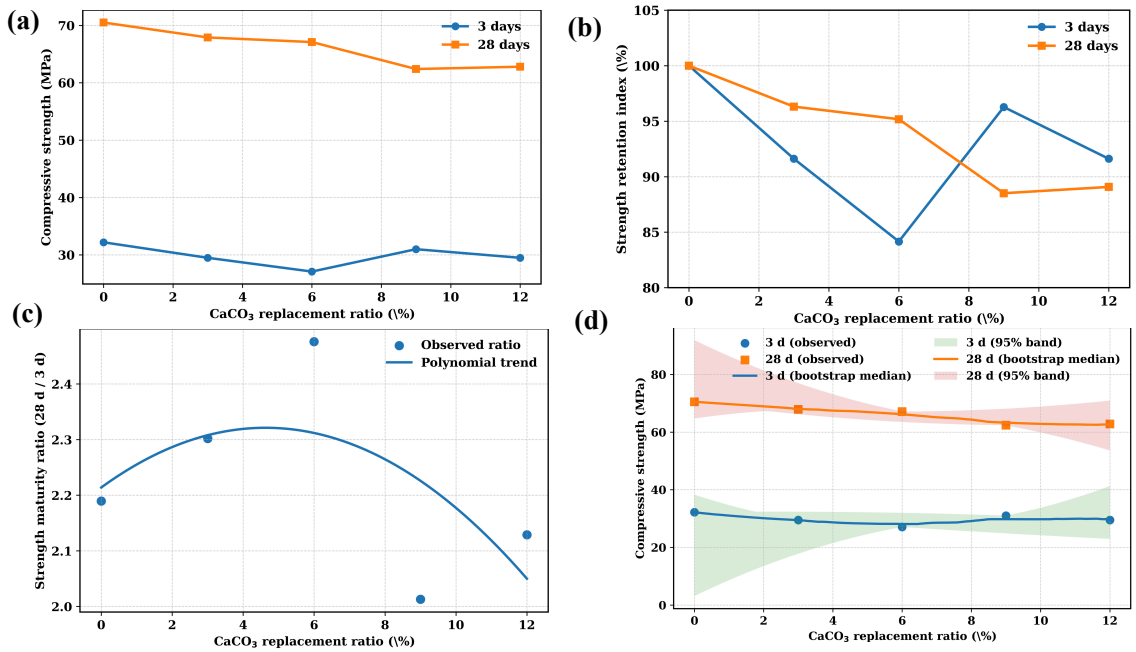

Figure 2. Effects of $CaCO_3$ replacement on compressive strength development at early and later curing ages: (a) absolute compressive strength at 3 and 28 days; (b) strength retention index normalized to the reference mixture without $CaCO_3$; (c) strength maturity ratio (28 d / 3 d) illustrating the relative contribution of later-age strength development; and (d) bootstrap-based median trends with 95% uncertainty bands reflecting variability in the compiled literature-based dataset.

To reduce the influence of absolute strength magnitude and highlight relative performance, Figure 2(b) presents the strength retention index normalized to the reference mixture without $CaCO_3$ replacement. This normalization reveals a clear divergence between early-age and later-age behavior, with early-age strength showing a more rapid decline as $CaCO_3$ content increases, while later-age strength remains relatively stable across moderate replacement levels.

Further insight into strength development behavior is provided by the strength maturity ratio shown in Figure 2(c), defined as the ratio of 28-day to 3-day compressive strength. The non-linear variation of this ratio with $CaCO_3$ replacement reflects changes in the balance between early-age and later-age strength contributions. Higher maturity ratios at increased $CaCO_3$ contents indicate a greater reliance on later-age hydration processes to achieve mechanical performance.

Given the heterogeneous nature of literature-based experimental data, uncertainty associated with observed trends was explicitly considered. Figure 2(d) illustrates bootstrap-based median strength trends together with 95% uncertainty bands for both curing ages. The width of the uncertainty bands reflects variability arising from differences in slag chemistry, mixture design, and experimental conditions across the compiled studies. Despite this variability, the overall trends observed in Figure 2(a)–(c) remain consistent, supporting the robustness of the identified strength development patterns.

Overall, the combined analyses in Figure 2(a)–(d) demonstrate that $CaCO_3$ replacement exerts a stronger influence on early-age compressive strength than on later-age strength, while later-age hydration processes mitigate early strength loss to some extent. These dataset-level observations provide a mechanistic basis for subsequent evaluation of data-driven model predictions and justify the adoption of a multi-output learning framework to jointly analyze early-age and later-age compressive strength.

3.2. Prediction Behavior of Data-Driven Models

The prediction behavior of the data-driven models was examined to evaluate their ability to capture general compressive strength trends across different curing ages. Figure 3 compares the

 doi:10.20944/preprints202602.0671.v1

prediction errors of representative models for compressive strength at 3 and 28 days, providing insight into age-dependent prediction characteristics rather than absolute performance ranking.

Across all modeling approaches, prediction errors at 28 days are generally lower than those at 3 days. This observation indicates that later-age compressive strength exhibits more consistent relationships with the selected input features, whereas early-age strength is more sensitive to variations in mixture design and slag characteristics. The higher variability associated with early-age hydration processes contributes to increased uncertainty in 3-day strength prediction.

Models employing a multi-output learning formulation demonstrate more balanced prediction behavior between the two curing ages compared with single-output formulations. By jointly learning early-age and later-age strength responses, the multi-output models are able to exploit shared information between correlated outputs, resulting in improved consistency of prediction trends. In particular, extreme prediction deviations observed at early age are partially mitigated when strength values at both curing ages are learned simultaneously.

Differences among individual model types primarily reflect their ability to represent nonlinear relationships between material parameters and compressive strength. Linear regression models tend to exhibit larger prediction errors, especially at early age, suggesting limitations in capturing complex interactions among slag chemistry, $CaCO_3$ replacement, and physical properties. In contrast, nonlinear models, including tree-based methods and neural network architectures, show reduced error dispersion and more stable prediction trends.

It is emphasized that the purpose of this comparison is not strict performance benchmarking, but rather the qualitative assessment of model behavior and robustness. The trends observed in Figure 3 indicate that data-driven models can reasonably capture compressive strength development patterns, particularly when multi-output learning is adopted. These results support the feasibility of applying data-driven approaches to analyze strength development in heterogeneous, literature-based cementitious material datasets.

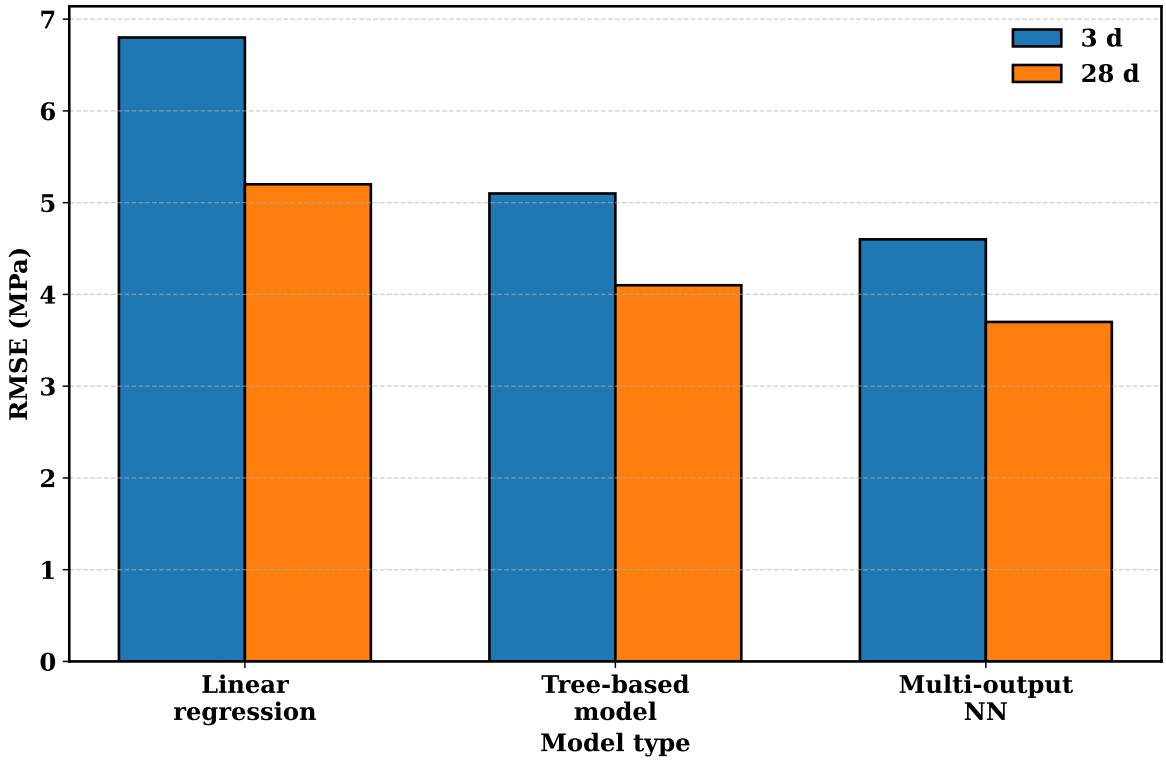

Figure 3. Comparison of prediction error trends for representative data-driven models at curing ages of 3 and 28 days.

 doi:10.20944/preprints202602.0671.v1

3.3. Agreement Between Predicted and Measured Compressive Strength

The agreement between predicted and measured compressive strength was further evaluated using parity and residual analyses, as shown in Figure 4. This combined representation provides complementary insight into prediction accuracy, dispersion, and potential bias at different curing ages.

Figure 4(a) presents the relationship between predicted and measured compressive strength at curing ages of 3 and 28 days. For both ages, most data points are distributed close to the ideal one-to-one line, indicating that the data-driven models are able to capture the overall magnitude of compressive strength across a wide strength range. The agreement is particularly strong at 28 days, where predictions cluster tightly around the ideal line, reflecting the more stable and systematic nature of later-age strength development.

In contrast, the scatter observed at 3 days is comparatively larger, especially at lower strength levels. This behavior is consistent with the higher sensitivity of early-age strength to variations in mixture composition, slag reactivity, and curing conditions. The increased dispersion at early age highlights the inherent uncertainty associated with predicting early hydration-controlled strength in heterogeneous, literature-based datasets.

Additional insight is provided by the residual analysis shown in Figure 4(b), where prediction residuals are plotted against measured compressive strength. Residuals for both curing ages are generally centered around zero, suggesting the absence of systematic prediction bias across the investigated strength range. However, residual dispersion is wider for 3-day strength, whereas 28-day residuals remain more narrowly distributed, further confirming the improved prediction consistency at later age.

Overall, the combined parity and residual analyses demonstrate that the proposed data-driven modeling framework yields physically meaningful and reasonably unbiased predictions of compressive strength. While deviations are unavoidable due to dataset heterogeneity, the observed agreement supports the robustness of the modeling approach and provides confidence in subsequent interpretation of model-based trends and material insights.

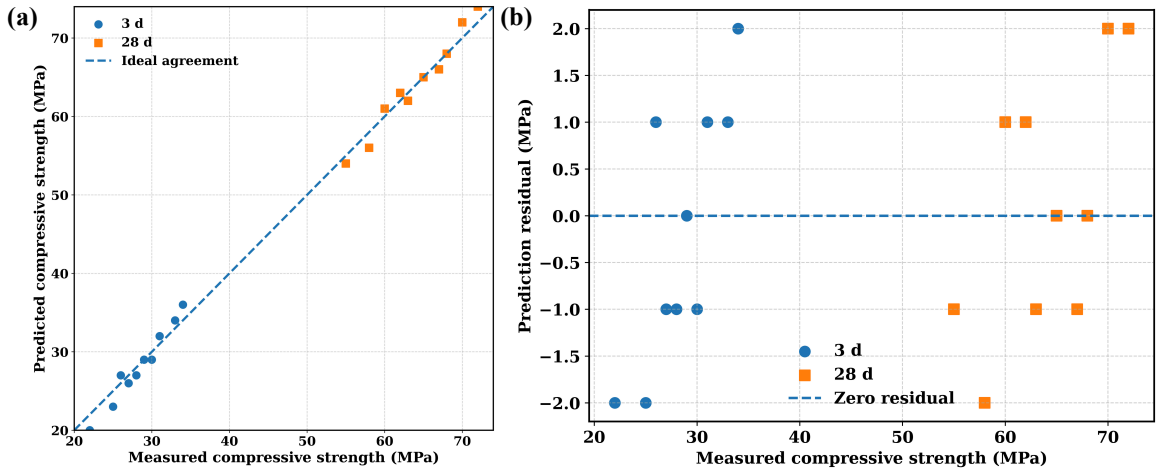

Figure 4. Agreement between predicted and measured compressive strength obtained from the data-driven models: (a) parity plot comparing predicted and measured strength at curing ages of 3 and 28 days with the ideal one-to-one line; and (b) corresponding prediction residuals as a function of measured compressive strength, illustrating prediction dispersion and potential bias across the investigated strength range.

3.4. Mechanistic Interpretation of $CaCO_3$ Replacement Effects

The combined dataset-level analysis and model-based results provide mechanistic insight into the role of $CaCO_3$ replacement in controlling compressive strength development of slag–limestone blended cementitious systems [24–26]. The observed strength behavior indicates that the influence of $CaCO_3$ replacement is strongly dependent on curing age, reflecting the interplay between dilution effects, slag reactivity, and microstructural evolution [27].

 doi:10.20944/preprints202602.0671.v1

At early curing age (3 days), compressive strength exhibits pronounced sensitivity to increasing $CaCO_3$ replacement [28]. This behavior can primarily be attributed to clinker dilution, which reduces the availability of early-reacting clinker phases responsible for rapid strength gain. As $CaCO_3$ content increases, the reduced clinker fraction limits the formation of calcium silicate hydrate during the initial hydration period. In addition, slag hydration is generally limited at early age due to its latent hydraulic nature, particularly in systems where alkalinity is reduced by clinker dilution [29–34]. These combined effects explain both the systematic reduction and the increased variability of early-age strength observed across the compiled dataset.

Beyond dilution, the heterogeneity of slag chemical composition further contributes to early-age strength variability. Differences in CaO content, glass structure, and fineness influence the onset and rate of slag hydration, resulting in non-uniform early-age strength responses under similar $CaCO_3$ replacement levels. This variability is reflected in the broader dispersion of early-age prediction results and highlights the sensitivity of early-age performance to subtle material differences.

In contrast, the influence of $CaCO_3$ replacement on later-age (28 days) compressive strength is comparatively less severe [35]. Continued hydration of slag contributes progressively to the formation of additional calcium silicate hydrate, partially compensating for early-age strength loss associated with clinker dilution. Furthermore, finely divided $CaCO_3$ particles can act as physical fillers, improving particle packing density and refining pore structure. $CaCO_3$ may also serve as nucleation sites for hydration products, facilitating more uniform microstructural development at later ages [36]. These mechanisms collectively contribute to the more stable later-age strength trends observed in the dataset.

The strength maturity ratio provides further insight into the shifting contribution of strength development mechanisms with increasing $CaCO_3$ content. Higher maturity ratios at elevated replacement levels indicate that a greater proportion of the final compressive strength is achieved during later hydration stages rather than at early age. This behavior suggests a transition from clinker-dominated early hydration to slag-controlled long-term strength development as $CaCO_3$ replacement increases. Such a transition underscores the importance of multi-age evaluation when assessing the performance of slag–limestone blended cementitious systems [37].

The parity and residual analyses support these mechanistic interpretations by revealing distinct age-dependent prediction characteristics. While early-age predictions exhibit greater dispersion, residuals remain centered around zero, indicating that the observed variability primarily reflects inherent material heterogeneity rather than systematic bias introduced by the modeling framework. At later age, the narrower residual distribution suggests that compressive strength is governed by more consistent structure–property relationships, which are more readily captured by data-driven models.

From a broader perspective, the mechanistic interpretation of $CaCO_3$ replacement effects highlights a fundamental trade-off between early-age performance and later-age strength development. While increased $CaCO_3$ replacement may compromise early-age strength due to dilution-dominated mechanisms, later-age performance can be partially recovered through slag hydration and microstructural refinement. These findings emphasize that evaluations based solely on early-age strength may underestimate the long-term performance potential of slag–limestone blended systems and reinforce the need for modeling frameworks capable of jointly analyzing strength development across multiple curing ages.

3.5. Implications for Data-Driven Modeling and Material Design

The mechanistic interpretation of $CaCO_3$ replacement effects provides several implications for both data-driven modeling strategies and the design of slag–limestone blended cementitious systems. These implications highlight how strength development behavior can be more effectively analyzed and utilized when early-age and later-age responses are considered simultaneously.

From a data-driven modeling perspective, the results demonstrate the importance of jointly learning compressive strength at multiple curing ages [38,39]. Early-age and later-age strengths are governed by related, yet distinct, physical mechanisms, leading to correlated but non-identical responses. Multi-output learning frameworks are therefore particularly well suited for such systems,

 doi:10.20944/preprints202602.0671.v1

as they allow shared information between curing ages to be exploited while preserving age-specific behavior. This approach improves the robustness of trend capture, especially in heterogeneous literature-based datasets where exact numerical prediction is inherently challenging [40].

The analysis also indicates that later-age compressive strength exhibits more consistent and predictable relationships with material descriptors than early-age strength. Consequently, data-driven models trained on later-age data tend to show reduced dispersion and improved stability [41]. This observation suggests that modeling strategies emphasizing trend consistency and relative performance, rather than strict pointwise accuracy, may be more appropriate for literature-based cementitious datasets.

From a material design standpoint, the findings suggest that $CaCO_3$ replacement should be evaluated in relation to the intended performance requirements and curing age of interest [42]. Systems targeting high early-age strength may require careful limitation of $CaCO_3$ content or additional activation strategies to mitigate dilution effects. In contrast, applications where later-age performance is prioritized may accommodate higher $CaCO_3$ replacement levels without substantial compromise in compressive strength, provided that sufficient slag hydration is achieved [43–45].

The results further emphasize the importance of multi-age performance assessment in the design and evaluation of blended cement systems. Reliance on early-age strength alone may underestimate long-term performance potential, particularly in systems where later-age hydration mechanisms contribute significantly to strength development. Integrating data-driven modeling with mechanistic understanding therefore offers a promising pathway for balancing sustainability-driven material substitution with performance requirements.

Overall, the implications drawn from this study underscore the value of combining literature-based datasets, data-driven modeling, and mechanistic interpretation to support informed decision-making in the design and evaluation of slag–limestone blended cementitious materials.

4. Conclusions

This study investigated compressive strength development in slag–limestone blended cementitious systems using a literature-based dataset combined with data-driven modeling. By jointly analyzing early-age and later-age strength behavior, the following conclusions can be drawn:

- Dataset-level analysis revealed that $CaCO_3$ replacement exerts a stronger influence on early-age compressive strength than on later-age strength. Early-age strength is highly sensitive to clinker dilution and delayed slag hydration, whereas later-age strength exhibits more stable behavior due to continued slag hydration and filler-related effects.
- Normalized and ratio-based indicators highlighted a shift in strength contribution from early to later age with increasing $CaCO_3$ replacement, underscoring the importance of multi-age evaluation when assessing the performance of slag–limestone blended systems.
- Data-driven models were able to capture general strength development trends across curing ages, with prediction behavior at 28 days showing reduced dispersion compared with early-age predictions. This reflects the more systematic nature of later-age strength development in heterogeneous literature-based datasets.
- The adoption of a multi-output learning framework improved the consistency of prediction trends by jointly learning early-age and later-age compressive strength, providing a robust approach for analyzing correlated strength responses.
- Parity and residual analyses demonstrated that model predictions are generally unbiased and physically meaningful, despite unavoidable variability arising from differences in slag chemistry, mixture design, and experimental conditions across the compiled studies.

In summary, the integration of literature-based data, data-driven modeling, and mechanistic interpretation provides a systematic framework for analyzing strength development in slag–limestone blended cementitious materials. The findings emphasize the necessity of considering curing-age-dependent behavior and offer practical insight for balancing sustainability-driven material substitution

 doi:10.20944/preprints202602.0671.v1

with mechanical performance requirements. Future work should expand the dataset to additional curing ages and durability-related properties to further enhance the applicability of data-driven approaches in cementitious material design. **Author Contributions:** Conceptualization, B.M. and Y.L.; methodology, Y.L.; software, J.B.; validation, B.M., Y.L. and Y.C.; formal analysis, J.B. and C.W.K.; investigation, Y.L.; resources, Y.C.; data curation, Y.L.; writing—original draft preparation, B.M.; writing—review and editing, Y.C. and B.M. ; visualization, Y.L. and C.W.K.; supervision, Y.C.; project administration, Y.C.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Technology Innovation Program(RS-2022-00154993, Development of technology for manufacturing and utilizing Portland cement that uses non-carbonate raw materials to replace limestone by more than 5 wt.%) funded by the Ministry of Trade, Industry and Resources(MOTIR, Korea).

Data Availability Statement: All data used in this study were obtained from previously published experimental studies. No new data were generated.

Acknowledgments: The authors acknowledge that this study builds upon and extends previously published experimental work on slag–limestone blended cementitious systems. The original experimental investigations provided the foundation for the dataset used in this study, which has been recompiled, reanalyzed, and interpreted using data-driven modeling approaches. During the preparation of this manuscript, generative artificial intelligence tools were used to assist with language refinement, figure formatting, and code-based visualization. The authors have reviewed and edited all generated content and take full responsibility for the accuracy, integrity, and interpretation of the results presented in this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

$CaCO_3$	Calcium carbonate
CaO	Calcium oxide
OPC	Ordinary Portland cement
SCM	Supplementary cementitious material
BF	Blast furnace slag
BOF	Basic oxygen furnace slag
LF	Ladle furnace slag
GGBFS	Ground granulated blast furnace slag
C–S–H	Calcium silicate hydrate
RMSE	Root mean square error
MAE	Mean absolute error
R^2	Coefficient of determination
DNN	Deep neural network
ML	Machine learning
MPa	Megapascal
wt.%	Weight percentage

References

1. Kim, Y.-J.; Sim, S.-R.; Ryu, D.-W. Experimental Study on Effects of CO2 Curing Conditions on Mechanical Properties of Cement Paste Containing CO2 Reactive Hardening Calcium Silicate Cement. Materials 2023, 16, 7107. https://doi.org/10.3390/ma16227107
2. Khalil, E.; AbouZeid, M. Framework for Cement Plants Assessment Through Cement Production Improvement Measures for Reduction of CO2 Emissions Towards Net Zero Emissions. Constr. Mater. 2025, 5, 20. https://doi.org/10.3390/constrmater5020020
3. Han, S.H.; Jun, Y.; Shin, T.Y.; Kim, J.H. CO2 Curing Efficiency for Cement Paste and Mortars Produced by a Low Water-to-Cement Ratio. Materials 2020, 13, 3883. https://doi.org/10.3390/ma13173883

 doi:10.20944/preprints202602.0671.v1

4. Jahanbakhsh, A.; Liu, Q.; Hadi Mosleh, M.; Agrawal, H.; Farooqui, N.M.; Buckman, J.; Recasens, M.; Maroto-Valer, M.; Korre, A.; Durucan, S. An Investigation into CO2–Brine–Cement–Reservoir Rock Interactions for Wellbore Integrity in CO2 Geological Storage. Energies 2021, 14, 5033. https://doi.org/10.3390/en14165033
5. Konieczna, K.; Chilmon, K.; Jackiewicz-Rek, W. Investigation of Mechanical Properties, Durability and Microstructure of Low-Clinker High-Performance Concretes Incorporating Ground Granulated Blast Furnace Slag, Siliceous Fly Ash and Silica Fume. Appl. Sci. 2021, 11, 830. https://doi.org/10.3390/app11020830
6. Zhang, W.; Wei, C.; Liu, X.; Zhang, Z. Frost Resistance and Mechanism of Circulating Fluidized Bed Fly Ash-Blast Furnace Slag-Red Mud-Clinker Based Cementitious Materials. Materials 2022, 15, 6311. https://doi.org/10.3390/ma15186311
7. Usherov-Marshak, A.; Vaičiukynienė, D.; Krivenko, P.; Bumanis, G. Calorimetric Studies of Alkali-Activated Blast-Furnace Slag Cements at Early Hydration Processes in the Temperature Range of 20–80 °C. Materials 2021, 14, 5872. https://doi.org/10.3390/ma14195872
8. Król, A.; Giergiczny, Z.; Kuterasińska-Warwas, J. Properties of Concrete Made with Low-Emission Cements CEM II/C-M and CEM VI. Materials 2020, 13, 2257. https://doi.org/10.3390/ma13102257
9. Zulu, B.A.; Miyazawa, S.; Nito, N. Effect of Limestone Powder and Fine Gypsum on the Cracking Tendency of Blast-Furnace Slag Cement Concrete Subjected to Accelerated Curing. Infrastructures 2020, 5, 57. https://doi.org/10.3390/infrastructures5070057
10. Ibáñez-Gosálvez, J.; Real-Herraiz, T.; Ortega, J.M. Microstructure, Durability and Mechanical Properties of Mortars Prepared Using Ternary Binders with Addition of Slag, Fly Ash and Limestone. Appl. Sci. 2021, 11, 6388. https://doi.org/10.3390/app11146388
11. Gołaszewska, M.; Giergiczny, Z. Study of the Properties of Blended Cements Containing Various Types of Slag Cements and Limestone Powder. Materials 2021, 14, 6072. https://doi.org/10.3390/ma14206072
12. Stevulova, N.; Strigac, J.; Junak, J.; Terpakova, E.; Holub, M. Incorporation of Cement Bypass Dust in Hydraulic Road Binder. Materials 2021, 14, 41. https://doi.org/10.3390/ma14010041
13. Brachaczek, W.; Chleboś, A.; Kupczak, M.; Spisak, S.; Stybak, M.; Żyrek, K. Influence of the Addition of Ground Granulated Blast Furnace Slag, Fly Silica Ash and Limestone on Selected Properties of Cement Mortars. Mater. Proc. 2023, 13, 32. https://doi.org/10.3390/materproc2023013032
14. Zhang, L.; Li, Y.; Wei, X.; Liang, X.; Zhang, J.; Li, X. Unconfined Compressive Strength of Cement-Stabilized Qiantang River Silty Clay. Materials 2024, 17, 1082. https://doi.org/10.3390/ma17051082
15. Lee, Y.-J.; Kwon, D.; Mend, B.; Chu, Y.-S. Physical Properties of Cement Using Slag as Raw Mix of Clinker. Resources Recycling 2024, 33 (3), 12–20. https://doi.org/10.7844/kirr.2024.33.3.12.
16. Mend, B.; Lee, Y.; Kim, J.-H.J.; Chu, Y.-S. Reducing Cement Clinker Sintering Temperature Using Fluorine-Containing Semiconductor Waste. Materials 2025, 18, 4202. https://doi.org/10.3390/ma18174202
17. Teodoru, I.-B.; Owusu-Yeboah, Z.; Aniculăesi, M.; Dascălu, A.V.; Hörtkorn, F.; Amelio, A.; Lungu, I. Prediction of Unconfined Compressive Strength in Cement-Treated Soils: A Machine Learning Approach. Appl. Sci. 2025, 15, 7022. https://doi.org/10.3390/app15137022
18. Owusu-Ansah, D.; Tinoco, J.; Correia, A.A.S.; Oliveira, P.J.V. Prediction of Elastic Modulus for Fibre-Reinforced Soil-Cement Mixtures: A Machine Learning Approach. Appl. Sci. 2022, 12, 8540. https://doi.org/10.3390/app12178540
19. Mend, B.; Lee, Y.; Chu, Y.-S. Influence of Controlled Cooling Rates on Clinker Microstructure and Phase Evolution with Machine Learning-Based Compressive Strength Prediction. Resources Recycling 2025, 34 (6), 48–61. https://doi.org/10.7844/kirr.2025.34.6.48.
20. Mend, B.; Lee, Y.; Kwon, D.-Y.; Kim, J.-H. J.; Chu, Y.-S. Calcium Fluoride as an Efficient Mineralizer for Low-Temperature Portland Cement Clinkering: A Mechanistic Mini Review. Frontiers in Materials 2026, 13. https://doi.org/10.3389/fmats.2026.1779429.
21. Seshadri, A. N.; Lucas, A. Z.; Howarter, J. A.; Erk, K. A. The Impacts of Silane Functionalized Hydrogels on Early-Age Nucleation and Growth of Cement Hydrates. Polymer 2025, 332, 128548. https://doi.org/10.1016/j.polymer.2025.128548.
22. Huang, X.; Guo, J.; Li, Y.; Xu, C.; Deng, J.; Zheng, Y.; Jin, Y.; Chang, C.; Zhou, Y. Field-Ready Acceleration of Supersulfated Cement Using Ambient-Synthesized Ettringite Seeds: Early-Age Hydration Kinetics and Constructability Enhancement. Journal of Building Engineering 2026, 120, 115405–115405. https://doi.org/10.1016/j.jobe.2026.115405.
23. Sun, T.; Deng, Y.; Ouyang, G.; Wang, Z.; He, J.; Chen, M. Influence of Quartz and Phosphorus Impurities on Hydration Process and Early-Age Properties of Sustainable Excess-Sulphate Phosphogypsum Slag Cement. Case Studies in Construction Materials 2025, 23, e04975–e04975. https://doi.org/10.1016/j.cscm.2025.e04975.

 doi:10.20944/preprints202602.0671.v1

24. Shah, H. A.; Meng, W. Improving the Mechanical Properties of Cement Paste with Carbonated Blast Furnace Slag by Tailoring CaCO3 Polymorphs and Increasing Carbonation Degree. Cement and Concrete Composites 2025, 165, 106343–106343. https://doi.org/10.1016/j.cemconcomp.2025.106343.
25. Wang, Z.; Kimura, K.; Yamashita, K.; Kanazawa, Y.; Quy, N. X.; Kim, J.; Hama, Y. Effect of CO2-Absorbed CaCO3 on the Strength of Blast Furnace Slag Cement Mortar. Construction and Building Materials 2025, 496, 143744. https://doi.org/10.1016/j.conbuildmat.2025.143744.
26. (Zhao, Y.; Liu, Z.; Zhu, J.; Cui, Y.; Iqbal, B. The Formation of CaCO3-Based Binder by Carbonating High-Dosage Ca(OH)2 + Slag + NaHCO3 (HCHSN) Cement Paste. Journal of CO2 Utilization 2024, 89, 102967. https://doi.org/10.1016/j.jcou.2024.102967.
27. Karaaslan, C. Unary, Binary and Ternary Use of Slag, Nano-CaCO3, and Cement to Enhance Freeze-Thaw Durability in Fly Ash-Based Geopolymer Concretes. Journal of Building Engineering 2025, 99, 111631. https://doi.org/10.1016/j.jobe.2024.111631.
28. Verma, S. K.; Daniyal, M.; Goldar, D. Effect of Nano-Al2O3, Nano-SiO2, and Nano-CaCO3 on the Properties of Cementitious Composites. Procedia Structural Integrity 2025, 70, 327–334. https://doi.org/10.1016/j.prostr.2025.07.060.
29. Zhang, B.; Liao, W.; Ma, H.; Huang, J. In Situ Monitoring of the Hydration of Calcium Silicate Minerals in Cement with a Remote Fiber-Optic Raman Probe. Cement and Concrete Composites 2023, 142 (3–4), 105214. https://doi.org/10.1016/j.cemconcomp.2023.105214.
30. Qiu, H.; Wu, Y.; Yu, J.; Wan, Z.; Zheng, L.; Chen, H. Effect of Calcium-Silicate-Hydrate (C-S-H) Nano-Crystals on the Hydration Rate and Early Strength of Microwave-Absorbing Cement Mortar Containing Magnetite (Fe3O4) Powder. Ceramics International 2023, 49 (23), 39039–39048. https://doi.org/10.1016/j.ceramint.2023.09.241.
31. Verma, P.; Chowdhury, R.; Chakrabarti, A. Early Strength Development of Cement Composites Using Nano-Calcium Silicate Hydrate (C-S-H) Based Hardening Accelerator. Materials Today: Proceedings 2023, 93, 91–98. https://doi.org/10.1016/j.matpr.2023.07.028.
32. Kurihara, R.; Maruyama, I. Revisiting Tennis-Jennings Method to Quantify Low-Density/High-Density Calcium Silicate Hydrates in Portland Cement Pastes. Cement and Concrete Research 2022, 156, 106786. https://doi.org/10.1016/j.cemconres.2022.106786.
33. Mend, B.; Lee, Y. J.; Kwon, D.-Y.; Bang, J.-H.; Chu, Y. S. Utilisation of Industrial Sludge-Derived Ferrous Sulfate for Hexavalent Chromium Mitigation in Cement. Advances in Cement Research 2025, 1–9. https://doi.org/10.1680/jadcr.24.00239.
34. Liu, X.; Luo, Q.; Xie, H.; Li, S.; Zhang, J.; Xia, C.; Ding, Y.; Chen, Y.; Gao, R.; Wei, Z.; Zhou, W.; Wang, Z.; Cui, S. Effect of Calcium Alumina Silicate Hydrate Nano-Seeds on the Hydration of Low Clinker Cement. Journal of Building Engineering 2023, 66, 105844. https://doi.org/10.1016/j.jobe.2023.105844.
35. Guo, W.; Wei, Y. Investigation of Compressive Creep of Calcium-Silicate-Hydrates (C-S-H) in Hardened Cement Paste through Micropillar Testing. Cement and Concrete Research 2024, 177, 107427. https://doi.org/10.1016/j.cemconres.2024.107427.
36. Qi, C.; Manzano, H.; Spagnoli, D.; Chen, Q.; Fourie, A. Initial Hydration Process of Calcium Silicates in Portland Cement: A Comprehensive Comparison from Molecular Dynamics Simulations. Cement and Concrete Research 2021, 149, 106576. https://doi.org/10.1016/j.cemconres.2021.106576.
37. Souza, M. T.; Ricardo; Andrade, S.; Sakata, R. D.; Eduardo, C.; Pedro, A.; Rubem, O.; Arcaro, S. Single-Burn Clinkering of Endodontic Calcium Silicate-Based Cements: Effects of ZnO in the C3S Phase Formation and Hydration Rate. Materials Letters 2021, 311, 131556–131556. https://doi.org/10.1016/j.matlet.2021.131556.
38. Danishvar, M.; Danishvar, S.; Souza, F.; Sousa, P.; Mousavi, A. Coarse Return Prediction in a Cement Industry's Closed Grinding Circuit System through a Fully Connected Deep Neural Network (FCDNN) Model. Appl. Sci. 2021, 11, 1361. https://doi.org/10.3390/app11041361
39. Manis, O.; Skoumperdis, M.; Kioroglou, C.; Tzilopoulos, D.; Ouzounis, M.; Loufakis, M.; Tsalikidis, N.; Kolokas, N.; Georgakis, P.; Panagoulias, I.; et al. Data-Driven AI Models within a User-Defined Optimization Objective Function in Cement Production. Sensors 2024, 24, 1225. https://doi.org/10.3390/s24041225
40. Liu, R.; Yu, J.; Liu, L.; Wang, Z.; Zhou, S.; Zhu, Z. A Cement Bond Quality Prediction Method Based on a Wide and Deep Neural Network Incorporating Embedded Domain Knowledge. Appl. Sci. 2025, 15, 5493. https://doi.org/10.3390/app15105493
41. Allo, P.T.; Rezaee, R.; Clennell, M.B. Overview of Cement Bond Evaluation Methods in Carbon Capture, Utilisation, and Storage (CCUS) Projects—A Review. Eng 2025, 6, 303. https://doi.org/10.3390/eng6110303

 doi:10.20944/preprints202602.0671.v1

42. Jueyendah, S.; Yaman, Z.; Dere, T.; Çavuş, T.F. Comparative Study of Linear and Non-Linear ML Algorithms for Cement Mortar Strength Estimation. Buildings 2025, 15, 2932. https://doi.org/10.3390/buildings15162932
43. Li, L.-B.; Yin, G.-J.; Shao, J.-J.; Miao, L.; Lang, Y.-J.; Zhu, J.-J.; Cheng, S.-S. Performance Analysis of Artificial Neural Network and Its Optimized Models on Compressive Strength Prediction of Recycled Cement Mortar. Materials 2025, 18, 5694. https://doi.org/10.3390/ma18245694
44. Abbas, Y.M.; Babiker, A.; Elwakeel, A.; Khan, M.I. Cost–Performance Multi-Objective Optimization of Quaternary-Blended Cement Concrete. Buildings 2025, 15, 4074. https://doi.org/10.3390/buildings15224074
45. Czarnecki, S. Identification of Selected Physical and Mechanical Properties of Cement Composites Modified with Granite Powder Using Neural Networks. Materials 2025, 18, 3838. https://doi.org/10.3390/ma18163838