Statistical and Machine-Learning Approaches for Predicting Concrete Compressive Strength: A Review of Methods, Predictors, and Practical Performance

Ali Hasan
University of Benghazi- Faculty of Engineering and Petroleum
Ali.hasan@uob.edu.ly
0009-0006-3202-5044

Wasfi Albadry
University of Benghazi- Faculty of Engineering
Wasfi.abdulhafith@uob.edu.ly
0009-0003-7368-8977

Wesam Al farsi
College of Science and Technology. Qaminis
wesam_alfarsi@qi.edu.ly
0009-0006-1035-1295

Abdulhamid Alhasi
University of Benghazi- Faculty of Engineering and Petroleum
Anasrazan2017@gmail.com
0009-0007-5394-8881

تاريخ الارسال 2026/9/3 تاريخ القبول 2023/9/6

Abstract

Concrete compressive strength is the primary criterion used to judge the quality, safety, and durability of concrete structures, yet its conventional 28-day determination delays quality decisions on site. Over the past two decades a substantial body of research has used statistical regression and, increasingly, machine-learning methods to predict compressive strength directly from mixture proportions and curing age. This review synthesizes that literature across three broad families of numerical approach: SPSS-based multiple linear regression on single-laboratory datasets, classical correlation and regression studies that established the field's core predictors, and more recent work using larger, pooled, or heterogeneous databases together with nonlinear and machine-learning models. Across more than thirty reviewed studies, the water-cement ratio and cement content emerge as the most consistent predictors of strength, while curing age contributes a positive but nonlinear, diminishing-returns effect. A clear and recurring pattern links reported model accuracy to dataset scale and homogeneity: small, single-source datasets tend to yield very high coefficients of determination ($R^2 > 0.90$), whereas large, pooled, multi-source databases yield more moderate values (R^2 between 0.50 to 0.65), which likely better represent the predictive accuracy achievable in general engineering practice. The review identifies persistent gaps around

dataset heterogeneity, limited treatment of supplementary cementitious materials and admixtures, sparse interaction-term modelling, and inconsistent independent validation practice, and proposes a research agenda addressing these gaps. The findings are intended to help practicing engineers and researchers to select an appropriate predictive approach and to calibrate their expectations of accuracy according to the scale and diversity of the underlying dataset.

These findings aim to guide practicing engineers and researchers in selecting appropriate predictive approaches, thereby calibrating accuracy expectations based on data diversity

Keywords*: concrete compressive strength; multiple linear regression; SPSS; machine learning; correlation analysis; water-cement ratio; model validation; literature review*

الأساليب الإحصائية وتقنيات التعلم الآلي للتنبؤ بمقاومة الضغط للخرسانة: مراجعة للطرق، والمتغيرات التنبؤية، والأداء العملي

المستخلص:

تُعد مقاومة الخرسانة للضغط المعيار الأساسي للحكم على جودة وسلامة ومتانة المنشآت الخرسانية، إلا أن الطريقة التقليدية لتحديد هذه المقاومة عند عمر 28 يوماً تؤدي إلى تأخير اتخاذ القرارات المتعلقة بالجودة في موقع العمل. وعلى مدى العقدين الماضيين، اعتمدت مجموعة كبيرة من الأبحاث على أساليب الانحدار الإحصائي، وبشكل متزايد على تقنيات تعلم الآلة، للتنبؤ بمقاومة الضغط مباشرةً بناءً على نسب مكونات الخلطة وعمر المعالجة. وتستعرض هذه الورقة تلك الأدبيات البحثية وتصنفها ضمن ثلاث فئات رئيسية من المناهج العددية: الانحدار الخطي المتعدد) باستخدام برنامج (SPSS المعتمد على مجموعات بيانات من مختبر واحد؛ ودراسات الارتباط والانحدار التقليدية التي أرست دعائم المتغيرات التنبؤية الأساسية في هذا المجال؛ والأعمال الأحدث التي تستخدم قواعد بيانات أكبر أو مجمعة أو غير متجانسة إلى جانب نماذج غير خطية ونماذج تعلم الآلة. ومن خلال مراجعة أكثر من ثلاثين دراسة، تبيّن أن نسبة الماء إلى الأسمنت ومحتوى الأسمنت هما أكثر المتغيرات ثباتاً في التنبؤ بالمقاومة، في حين أن عمر المعالجة يسهم بتأثير إيجابي ولكنه غير خطي ويتسم بتناقص العائد. كما يظهر نمط واضح ومتكرر يربط بين دقة النموذج المُبلغ عنها وحجم وتجانس مجموعة البيانات؛ إذ تميل مجموعات البيانات الصغيرة والمستمدة من مصدر واحد إلى إظهار معاملات تحديد (R^2) عالية جداً (أكبر من 0.90)، بينما تُظهر قواعد البيانات الكبيرة والمجمعة والمتعددة المصادر قيماً أكثر اعتدالاً (تتراوح بين 0.50 و0.65)، وهي قيم يُرجح أنها تعكس بشكل أدق مستوى الدقة التنبؤية الممكن تحقيقه في الممارسات الهندسية العامة. وتحدد المراجعة فجوات بحثية مستمرة تتعلق بعدم تجانس البيانات، ومحدودية تناول المواد الأسمنتية التكميلية والمضافات الكيميائية، وندرة نمذجة التفاعلات المتبادلة بين المتغيرات، وغياب الاتساق في ممارسات التحقق المستقل من النماذج، كما تقترح أجندة بحثية لمعالجة هذه الفجوات. وتهدف هذه النتائج إلى مساعدة المهندسين الممارسين

والباحثين في اختيار المنهج التنبؤي المناسب وضبط توقعاتهم بشأن الدقة بما يتناسب مع حجم وتنوع مجموعة البيانات المستخدمة.

تهدف هذه النتائج إلى توجيه المهندسين الممارسين والباحثين لاختيار مناهج تنبؤية مناسبة، ومن ثم ضبط توقعات الدقة بناءً على تنوع البيانات.

الكلمات المفتاحية: مقاومة الخرسانة للضغط؛ الانحدار الخطي المتعدد؛ SPSS؛ تعلم الآلة؛ تحليل الارتباط؛ نسبة الماء إلى الأسمنت؛ التحقق من صحة النموذج؛ مراجعة الأدبيات.

1. Introduction

Concrete is the most widely used construction material in the world, and its compressive strength is the single most important mechanical property used to judge whether a batch or structural member meets its design requirements. Compressive strength is conventionally obtained by testing standard cubes or cylinders after a 28-day curing period, in accordance with international standards such as ACI 318, ASTM C39, and EN 206. The resulting delay between placement and strength confirmation is a long-standing practical problem, since any deficiency discovered after 28 days can trigger costly demolition, repair, or strengthening works, while significantly disrupting the construction programme.

To mitigate this limitation, researchers have developed numerical prediction methods that estimate compressive strength directly from mixture proportions, curing age, and, in some cases, early-age test results, without waiting for the full curing period. These methods range from simple bivariate correlations, through multiple linear regression performed in general-purpose statistical software such as SPSS, to nonlinear regression and modern machine-learning algorithms. Each approach involves a trade-off between transparency and accessibility on one hand, and predictive accuracy on the other, and this trade-off is strongly mediated by the size and homogeneity of the dataset used to build the model.

This review synthesizes the numerical (statistical and machine-learning) branch of the compressive-strength prediction literature. It does not attempt to cover the theoretical and purely experimental branches in depth, except where they provide the physical basis for interpreting statistical results. The specific

objectives are to: (1) classify the reviewed studies by dataset scale, predictor set, and modelling technique; (2) identify which mixture variables are most consistently reported as dominant predictors and why; (3) compare reported accuracy metrics across studies and relate them systematically to dataset characteristics; (4) identify methodological gaps, particularly around independent validation and dataset heterogeneity; and (5) propose directions for future research and practical guidance for engineers selecting a prediction approach.

2. Review Methodology

This is a narrative, structured literature review rather than a formal systematic review with pre-registered search strings, reflecting the breadth of publication venues (journals, conference proceedings, and institutional reports) in this applied engineering field. Sources were identified through targeted searches of the concrete-technology and civil-engineering literature on regression- and machine-learning-based compressive strength prediction, cross-referenced against the reference lists of recent primary studies to capture earlier foundational work. Studies were included if they (a) developed a statistical or machine-learning model relating mixture proportions and/or curing age to measured compressive strength, and (b) reported at least one quantitative goodness-of-fit or error metric, such as (R, R^2, MAE, RMSE, or MAPE). Purely experimental studies without a predictive modelling component, and purely theoretical derivations without empirical calibration, were excluded except where cited for physical context.

For each included study the review extracted, where reported: dataset size and source (single-laboratory versus pooled/multi-source), predictor variables used, modelling technique, calibration goodness-of-fit, and independent validation performance. Together, these parameters underpin the thematic synthesis in Sections 3-5 and the comparative table in Section 6.

3. Theoretical and Historical Basis

The theoretical starting point for almost all statistical prediction work is Abram's water-cement ratio law, which links strength inversely to the ratio of water to cement in the fresh mix through the porosity of the resulting hardened cement paste (Neville, 2011). Neville's classical treatment of concrete properties (2011) provides the mechanistic explanation that later regression

studies consistently reproduced empirically: for a fixed cement content, additional water dilutes the paste and increases capillary porosity, thereby reducing strength, while for a fixed water content, additional cement improves the paste-to-void ratio and raises strength.

Early statistical work built directly on this theoretical foundation. Pande and Gupta (2003) confirmed experimentally that the water-cement ratio is the single most influential parameter governing strength development. Namyong et al. (2004), analysing a large in-situ dataset of 1,442 test results from 59 mixtures, used correlation analysis to identify water-cement ratio and cement content as the dominant predictors and derived regression equations from 7- and 28-day data, showing that multiple linear regression using water-cement ratio, cement content, and the cement-to-aggregate ratio provides a practical quality-control tool. This early study established both the core predictor set and the basic multiple-regression methodology that later SPSS-based studies would adopt.

4. SPSS-Based Multiple Linear Regression Studies

A substantial cluster of studies has specifically used IBM SPSS Statistics to build multiple linear regression equations relating mix proportions to compressive strength, reflecting the software's accessibility to civil engineers without programming backgrounds. Hussain et al. (2020) developed a multiple linear regression equation for high-strength concrete based on cement and water content, reporting a very strong fit ($R^2 = 0.949$) on a dataset of 56 cube specimens. Idris and Bello (2021) modelled strength as a function of water-cement ratio and cement content over the range of W/C = 0.40–0.70 and likewise reported a strong correlation. Unamba et al. (2021) and Saif et al. (2024) applied SPSS regression to smaller, purpose-built datasets of 36 and 14 mixtures respectively, obtaining models with low prediction errors on their own data. Baihaqi and Rolia (2025) used a ten-year laboratory record of 28-day strength results to build an SPSS-based decision-support model for ongoing quality monitoring, illustrating a longitudinal, single-source application of the same technique. Bahoria and Vairagade (2021) extended the SPSS-regression approach to concrete containing quarry dust and waste plastic as sand replacements, confirming both the software choice and the general predictor set in a non-conventional mix context.

A common feature of this group of studies is that the datasets were small (14 to several hundred specimens) and drawn from a single laboratory or a narrow mix-design range. As discussed further in Section 6, this combination tends to produce very high R^2 values that describe the sampled range well but may not generalise to more heterogeneous conditions encountered across different projects, materials, and testing programmes.

5. Larger Databases, Nonlinear Models, and Machine Learning

5.1 Pooled and large-scale databases

A second cluster of studies has moved toward considerably larger and more heterogeneous databases, trading some calibration fit for greater representativeness. DeRousseau et al. (2019) combined 1,681 field-produced mixtures with over 1,000 laboratory mixtures and found that even the best-performing regression and polynomial models achieved only a moderate coefficient of determination ($R^2 \approx 0.51$) when predicting field strength. Furthermore, models trained purely on laboratory data tended to over-estimate field strength due to uncontrolled site conditions. Young et al. (2018) used a database exceeding 10,000 observations and found that machine-learning models outperformed classical regression for large, heterogeneous datasets, while still confirming cement content, water content, and age as dominant predictors. Similarly, Assegie et al. (2022) and Bansal et al. (2023), working with the widely used UCI concrete dataset of 1,030 records, reported higher R^2 values around 0.90-0.94 for optimized regression and machine-learning models, again identifying cement content, curing age, and water content as the most influential variables — a reminder that dataset size alone does not determine achievable accuracy; homogeneity of sourcing and mix-design conventions matters at least as much.

5.2 Nonlinear age formulations

Because strength gain with curing age follows a diminishing-returns pattern rather than a straight line, several studies have explored nonlinear age formulations. Qadi et al. (2011) fitted fully quadratic response-surface models to cement and water content and found that quadratic terms improved fit over simple linear terms at early ages. Kazberuk and Lelusz (2007) and Abd et al. (2008) developed regression models relating strength to replacement

proportions and curing age for both cement paste and concrete, reporting strong correlation coefficients that reinforce the role of age as a predictor once its nonlinearity is accounted for. Nipatsat and Tangtermsirikul (2004) and Daregholi and Hosseinzadeh (2021) specifically developed mathematical formulations for estimating strength at various curing ages. In particular, this included a mix-specific, project-calibrated approach for mass-concrete dam construction that achieved notably better accuracy than generic multi-source equations — underscoring the value of project-specific calibration highlighted further in Section 7.

5.3 Machine-learning and hybrid approaches

Jaskulski et al. (2017) built five first-order models with two to six input parameters and found that prediction accuracy improved as aggregate quality and additional variables were incorporated. Hematibahar and Kharun (2024) compared linear, ridge, and polynomial regression on more than 45 mix designs and found ridge and polynomial regression outperformed ordinary linear regression (R^2 = 0.93 and 0.91 respectively). Wang et al. (2022) and Ahmed et al. (2025) used probabilistic and neural-network-based approaches on smaller, curated datasets and achieved very high R^2 values (0.966 and 0.977), reinforcing the general finding that model accuracy tends to improve as dataset homogeneity increases, even when the underlying mathematical form is comparatively simple. More recent applied studies, including Wang et al. (2024), Ismail et al. (2024), Ezenkwa et al. (2025), and Ramadhan and Raharjo (2025), continue to confirm cement content, water-cement ratio, and curing age as first-order predictors while extending the methodology toward interpretable machine learning and toward accelerated, sub-28-day quality assessment.

6. Synthesis: Predictors, Dataset Scale, and Reported Accuracy

Across the reviewed literature, three patterns recur with notable consistency. First, the water-cement ratio and cement content are almost universally identified as the strongest predictors of compressive strength, whether considered individually through bivariate correlation or jointly within a multivariable regression (Idris & Bello, 2021; Namyong et al., 2004; Pande & Gupta, 2003; Neville, 2011; Assegie et al., 2022; Bansal et al., 2023; Wang et al., 2024). Second, curing age contributes a positive but nonlinear effect, best

captured through a logarithmic or quadratic transformation rather than a linear term (Daregholi & Hosseinzadeh, 2021; Nipatsat & Tangtermsirikul, 2004; Qadi et al., 2011). Third, and most importantly for practitioners selecting a modelling approach, reported model accuracy is strongly mediated by dataset size and source homogeneity: small, single-source datasets consistently yield very high R^2 values, while large, pooled, multi-source datasets yield more moderate values that likely better represent real-world predictive performance.

Table 1. Indicative comparison of reported goodness-of-fit across studies of differing dataset scale

Study	Dataset scale / source	Technique	Reported R^2
Hussain et al. [7]	56 specimens, single lab	SPSS multiple linear regression	0.949
Unamba et al. [9]	36 mixtures, single lab	SPSS multiple linear regression	High (single-source)
Wang et al. [22]	Curated dataset	PCA-based neural network	0.966
Ahmed et al. [23]	Curated dataset	Probabilistic / regression hybrid	0.977
Assegie et al. [18]; Bansal et al. [19]	UCI dataset, 1,030 records	Optimised regression / ML	0.90 - 0.94
Hematibahar & Kharun [21]	>45 mix designs	Ridge / polynomial regression	0.91 - 0.93
DeRousseau et al. [16]	>2,600 records, field + lab, pooled	Regression / polynomial models	≈ 0.51
Young et al. [17]	>10,000 observations, heterogeneous	Machine learning vs. classical regression	ML > regression

This inverse relationship between dataset scale/heterogeneity and reported R^2 is not a contradiction between studies but a structural feature of pooled, observational data: as the sampled range of materials, aggregate sources, and testing programs widens, sources of variance that are not captured by the six or seven routinely recorded mixture variables (cement type, aggregate mineralogy, mixing procedure, curing environment) increasingly dominate the residual, unexplained variance. Practitioners should therefore calibrate their accuracy expectations to the scale and diversity of the dataset underlying

any given model, rather than assuming that a high R^2 reported for a small, single-source study will transfer to a different set of materials or a different project.

7. Research Gaps and Future Directions

- Independent validation practices are inconsistent: many of the SPSS-based single-source studies reviewed report only calibration-set R^2 and do not test the fitted equation against an independent hold-out sample, which risks over-stating practical predictive accuracy.
- Supplementary cementitious materials (fly ash, slag, silica fume) and chemical admixtures, particularly superplasticisers, are largely absent from the reviewed regression datasets, despite their increasing use in practice; extending predictor sets to include these materials is a clear opportunity for future work.
- Interaction effects between predictors (for example, between cement content and curing age, or between water-cement ratio and aggregate proportioning) are rarely modelled explicitly in the linear-regression literature, even though several studies note nonlinear or conditional relationships qualitatively.
- Reporting of dataset provenance (single-source versus pooled, aggregate type, testing standard used) is uneven across studies, which complicates like-for-like comparison of reported accuracy metrics and motivates more standardised reporting practice.
- Comparative studies that apply both classical regression and machine-learning techniques to the same dataset, with consistent independent validation, remain relatively rare outside of a few large-database studies (Young et al., 2018; Bansal et al., 2023), limiting the evidence base for choosing between transparency and raw predictive accuracy in a given practical context.
- Mix-specific or project-specific calibration, as demonstrated for mass-concrete dam construction (Daregholi & Hosseinzadeh, 2021), appears to substantially outperform generic multi-source equations but has not yet been systematically compared across a range of project types and scales.

8. Practical Implications for Engineering Practice

For practicing engineers, the reviewed literature supports a differentiated use of statistical prediction tools depending on the scale and homogeneity of the available calibration data. Where a firm or ready-mix producer can calibrate a regression equation on its own historical batch records, using materials and mix designs specific to its operations, the reviewed single-source SPSS studies suggest that very high accuracy (R^2 above 0.90) is achievable, making such equations suitable for closer, near-decision-grade quality screening. Where only generic, literature-derived or multi-source equations are available, as in large pooled databases, the achievable accuracy is materially lower (R^2 between 0.50 and 0.65, with errors sometimes exceeding 30%). Therefore, such equations are best used as early, approximate screening tools that flag batches for closer follow-up rather than as a substitute for standard 28-day compliance testing under ACI, ASTM, or EN provisions.

In either case, the water-cement ratio and cement content should be treated as the primary variables to record and control, given their consistent dominance across the reviewed literature, while curing age should be incorporated using a nonlinear (logarithmic or quadratic) transformation rather than a linear term to better represent its diminishing-returns relationship with strength gain.

9. Conclusions

- This review has synthesized more than thirty studies applying statistical regression and machine-learning techniques to the prediction of concrete compressive strength based on mixture proportions and curing age.
- The water-cement ratio and cement content emerge as the most consistently dominant predictors across the literature, with curing age contributing a positive but nonlinear effect.
- A clear and recurring relationship links reported model accuracy to dataset scale and homogeneity: small, single-source SPSS-based studies report very high goodness-of-fit, whereas large, pooled, multi-source databases report more moderate, but arguably more realistic, accuracy.

- Persistent gaps remain around independent validation practice, the treatment of supplementary cementitious materials and admixtures, interaction-effect modelling, and the standardized reporting of dataset provenance.
- Addressing these gaps, together with wider adoption of mix-specific calibration and more systematic head-to-head comparison between regression and machine-learning techniques, represents a productive agenda for future research and offers practical guidance for engineers selecting an appropriate early-age strength prediction approach for their own materials and projects.

10. Recommendations

The following recommendations translate the review's findings into concrete actions for practising engineers and for researchers extending this line of work.

10.1 Recommendations for Practising Engineers

- Calibrate SPSS-based multiple linear regression equations on in-house historical batch records rather than relying on generic literature-derived equations; very high accuracy (R^2 above 0.90) is realistically achievable only when the calibration data are drawn from the firm's own materials and mix designs.
- Treat water-cement ratio and cement content as the primary variables to record and control on site, and represent curing age using a nonlinear (logarithmic or quadratic) transformation rather than a linear term, consistent with its diminishing-returns effect on strength gain.
- Use generic, multi-source, or pooled-database equations only as early, approximate screening tools that flag batches for closer follow-up, not as a substitute for standard 28-day compliance testing under ACI, ASTM, or EN provisions.
- Where supplementary cementitious materials or admixtures (fly ash, slag, silica fume, superplasticisers) are used on site, extend or recalibrate in-house prediction models accordingly, since these materials are largely absent from the regression datasets reviewed.

- Record and retain dataset provenance (single-source versus pooled, aggregate type, testing standard used) whenever developing or adopting a predictive equation, to support later validation and to avoid misapplying a model beyond the conditions it was calibrated on.

10.2 Recommendations for Researchers

- Report independent hold-out validation performance alongside calibration-set R^2, since many of the SPSS-based single-source studies reviewed report only calibration-set fit and may over-state practical predictive accuracy.
- Extend predictor sets to include supplementary cementitious materials and chemical admixtures, particularly superplasticisers, which remain largely absent from the regression datasets reviewed despite their increasing use in practice.
- Model interaction effects explicitly (for example, between cement content and curing age, or between water-cement ratio and aggregate proportioning) rather than relying solely on additive linear terms.
- Adopt standardised reporting of dataset provenance and scale (single-source versus pooled, aggregate type, testing standard) to enable like-for-like comparison of reported accuracy metrics across studies.
- Conduct more comparative studies applying both classical regression and machine-learning techniques to the same dataset, with consistent independent validation, to build a stronger evidence base for choosing between transparency and raw predictive accuracy.
- Systematically compare mix-specific or project-specific calibration, as demonstrated for mass-concrete dam construction (Daregholi & Hosseinzadeh, 2021), against generic multi-source equations across a wider range of project types and scales.

References

Abd, K., et al. (2008). Mathematical regression models to predict the compressive strength of cement and concrete at multiple curing ages.

Aggarwal, R., Kumar, M., Sharma, R. K., & Sharma, M. K. (2015). Predicting compressive strength of concrete. International Journal of Applied Science and Engineering, 13(2), 171-185.

Ahmed, M. S., Tasnim, A., Hasan, M. F., & Kabir, G. (2025). Dynamic probabilistic modeling of concrete strength: Markov chains and regression for sustainable mix design. Infrastructures, 10(8), Article 219.

Ahn, N., & Fowler, D. W. (2001). Utilisation of manufactured or complete aggregates as an effective concrete filler material.

Assegie, T. A., et al. (2022). Evaluation of regression models for concrete compressive strength prediction using the UCI database.

Bahoria, B. V., & Vairagade, V. S. (2021). Regression analysis using SPSS for compressive strength of concrete containing quarry dust and waste plastic as sand replacement. International Journal of Advanced Research in Engineering and Technology, 12(4), 141-150.

Baihaqi, A., & Rolia, N. (2025). Development of a predictive model for concrete compressive strength as a decision-support tool. Muhammadiyah Metro University.

Bansal, A., et al. (2023). Comparative analysis of concrete compressive strength prediction using machine learning techniques with hyperparameter optimisation.

Chopra, P., Sharma, R. K., & Kumar, M. (2014). Prediction of compressive strength of concrete using artificial neural network and genetic programming. Advances in Materials Science and Engineering, 2014.

Daregholi, B., & Hosseinzadeh, N. (2021). Prediction of long-term compressive strength of mass concrete using early-age test results: a case study of a dam construction site.

DeRousseau, M. A., Laftchiev, E., Kasprzyk, J. R., Balaji, R., & Srubar III, W. V. (2019). Machine learning methods for predicting the field compressive

strength of concrete (Technical Report No. TR2019-096). Mitsubishi Electric Research Laboratories.

Ezenkwa, C. S., Onwuka, D. O., Okere, C. E., Arimanwa, J. I., & Onwuka, S. U. (2025). Compressive strength and statistical investigation of concrete made with bituminous sand. International Journal of Science and Research Archive, 14(3), 417-429.

Galetakis, M., & Raka, S. (2004). Effect of natural sand replacement proportions on fresh and hardened concrete properties.

Hematibahar, M., & Kharun, M. (2024). Prediction of concrete mixture design and compressive strength through data analysis and machine learning. Journal of Mechanics of Continua and Mathematical Sciences, 19(3), 1-21.

Hussain, I., Ehsan, A., Rehman, U. A., & Tahir, S. M. T. (2020). Prediction of compressive strength of high strength concrete using multiple linear regression equation (MLRE) and numerical model. International Journal of Latest Engineering Research and Applications, 5(9), 26-36.

Idris, M. H., & Bello, N. (2021). Statistical model for predicting of concrete compressive strength from water-cement ratio and cement content. USEP: Journal of Science and Engineering Production, 3(2), 281-290.

Ismail, A., et al. (2024). Prediction of concrete compressive strength using regression analysis and ANOVA-based model validation.

Jaskulski, R., Kubissa, W., Koteš, P., & Brodňan, M. (2017). Predicting of the compressive strength of RCA concrete. MATEC Web of Conferences, 17, Article 00066.

Kazberuk, M., & Lelusz, H. (2007). Regression models for predicting the compressive strength of concrete containing recycled constituents.

Lamb, D. (2005). Compressive strength performance of structural concrete containing 12% sand content relative to C35 grade specifications.

Namyong, J., Yoon, S., & Cho, H. (2004). Prediction of compressive strength of in-situ concrete based on mixture proportions. Journal of Asian Architecture and Building Engineering, 3(1), 9-16.

Neville, A. M. (2011). Properties of Concrete (5th ed.). Pearson.

Nipatsat, N., & Tangtermsirikul, S. (2004). Development of mathematical equations for estimating compressive strength of concrete at various curing ages.

Palika, S., et al. (2014). Prediction of concrete compressive strength using multiple regression analysis based on water/cementitious and aggregate/cementitious ratios.

Pande, C. B., & Gupta, R. (2003). Effect of concrete mix proportions on compressive strength development: an experimental study.

Qadi, A. N. S., et al. (2011). Effect of key admixtures on compressive strength using fully quadratic statistical models.

Ramadhan, F., & Raharjo, D. (2025). Machine learning models for predicting 28-day cement compressive strength from chemical and physical properties.

Ramugade, R. (2012). Estimation of compressive strength of concrete cubes at 3, 7, and 28 days using multiple regression analysis.

Saif, M., Al Hariri, M. O. R., Hamad, H., Serag, A., El-Hafez, G. A., & Fergany, M. (2024). Predictive modeling of concrete water penetration depth based on material properties. Engineering Research Journal, 53(4), 290-299.

Sayed-Ahmed, M. (2012). Statistical model for predicting compressive strength based on mix-design parameters and curing age.

Sevinç, A. H. (2024). Determination of optimal concrete mix components and curing time for target compressive strength using an advanced statistical model.

Unamba, U. K., Nwajagu, E. S., Abutu, J., & Agbo-Anike, O. J. (2021). Predictive model of the compressive strength of concrete containing coconut shell ash as partial replacement of cement using multiple regression analysis. International Journal of Innovative Science and Research Technology, 6(8), 600-605.

Wang, W., Zhong, Y., Liao, G., Ding, Q., Zhang, T., & Li, X. (2024). Prediction of compressive strength of concrete specimens based on interpretable machine learning. Materials, 17(15), Article 3661.

Wang, H., Zhao, K., Zhang, Y., & Zhang, X. (2022). Prediction of compressive strength of concrete using the Spearman and PCA-based BP

neural network. Wireless Communications and Mobile Computing, 2022, Article 7248561.

Young, B. A., Hall, A., Pilon, L., Gupta, P., & Sant, G. (2018). Can the compressive strength of concrete be estimated from knowledge of the mixture proportions? New insights from statistical analysis and machine learning methods. Cement and Concrete Research, 115, 379-388.