Journal of Emerging Multidisciplinary Research
Volume 2, Issue No. 3, Jul – Sep, 2026 ISSN Number: 3139-3292
Frequency: Quarterly

Machine Learning-Based Prediction of Compressive Strength in Sustainable Concrete Incorporating Industrial By-Products: A Critical Review

Madhuraj Singh
Assistant Professor
School of Engineering and Technology, Vikrant University
Gwalior, Madhya Pradesh, India
soet_madhuraj@vikrantuniversity.ac.in

Abstract*— The partial replacement of Portland cement with industrial by-products such as fly ash, ground granulated blast furnace slag, silica fume and calcined clay is now a mainstream decarbonisation strategy in concrete technology. It also complicates strength prediction: by-product blends behave non-linearly, mature at different rates, and vary in composition from source to source, so the water–cement ratio rules embedded in conventional mix design lose much of their explanatory power. Machine learning (ML) has been widely proposed as the remedy. This paper reviews the development of ML-based compressive strength prediction for concretes containing industrial by-products, tracing the field from early artificial neural network models through ensemble and gradient boosting methods to hybrid and metaheuristic-optimised architectures, and to the recent turn toward interpretable models. The review argues that reported accuracy has plateaued while methodological quality has not kept pace. Four recurring weaknesses are identified: heavy dependence on a small number of reused public datasets, description of by-products by mass dosage rather than chemical or physical characteristics, near-universal absence of external validation on independently produced mixes, and reporting conventions that favour headline coefficients of determination over error distributions and uncertainty. Interpretability methods have improved transparency but are frequently used to confirm expectations rather than to test them. The paper concludes with a research agenda emphasising composition-aware feature sets, external and temporal validation, uncertainty quantification, and integration of prediction with multi-objective mix optimisation so that data-driven models contribute to sustainability outcomes and not only to prediction scores.

Keywords*— Compressive Strength; Machine Learning; Supplementary Cementitious Materials; Fly Ash; Ground Granulated Blast Furnace Slag; Sustainable Concrete; Model Interpretability

I. Introduction

Concrete is the most widely used manufactured material in the world, and Portland cement clinker production is among the largest single industrial sources of carbon dioxide. The most established route to reducing that burden is to replace part of the clinker with industrial by-products, generally described as supplementary cementitious materials (SCMs). Fly ash from coal combustion, ground granulated blast furnace slag (GGBS) from iron making, silica fume from ferrosilicon production and, more recently, calcined clays and recycled concrete powders are all used in this way. Alongside the environmental benefit, well-proportioned blended binders can improve durability, reduce heat of hydration and refine pore structure (Kanamarlapudi et al., 2020; Kosmatka et al., 2002).

The engineering difficulty is that these benefits are not free of consequence for strength development. Pozzolanic and latent hydraulic reactions proceed more slowly than clinker hydration, so blended concretes commonly show reduced early-age strength and enhanced later-age strength. The magnitude of both effects depends on replacement level, the reactivity and fineness of the by-product, curing conditions, and interactions between two or more SCMs used together. Fly ash content, for instance, has an optimum beyond which strength falls away, and that optimum shifts with the rest of the mix (Bouzoubaâ & Fournier, 2003). Classical mix design methods, built around a water–cement ratio law calibrated on plain Portland cement concrete, do not represent these interactions well. Empirical strength–maturity expressions extend the reach of the classical approach but remain restricted to the variables for which they were derived and cannot easily absorb a new input (Ben Chaabene et al., 2020).

Machine learning offers a different proposition. Rather than assuming a functional form, ML models infer the mapping from mixture proportions and curing age to strength directly from data, and can in principle be retrained as databases grow. The field traces back to Yeh (1998a), who trained a feed-forward neural network on high-performance concrete mixtures containing fly ash and slag and reported a coefficient of determination of roughly 0.91–0.92, far above the value obtainable from regression on water–cement ratio and age alone. The dataset compiled for that study was subsequently deposited in the UCI Machine Learning Repository (Yeh, 1998c) and has become the single most used benchmark in the field.

In the quarter century since, the volume of published work has grown very rapidly, and reported accuracy has risen to the point where coefficients of determination above 0.95 on held-out test data are routine. This paper asks what that trajectory actually demonstrates. Its position is that the algorithmic side of the problem is largely settled — several families of model reach comparable accuracy on the same data — while the data and validation side has advanced much less. The review is organised accordingly: Section 2 sets out its scope; Section 3 summarises the material behaviour that models must capture; Section 4 traces the sequence of modelling approaches; Sections 5 and 6 examine the data foundations and the interpretability literature;

Journal of Emerging Multidisciplinary Research
Volume 2, Issue No. 3, Jul – Sep, 2026 **ISSN Number: 3139-3292**
Frequency: Quarterly

Section 7 offers a critical appraisal; Sections 8 and 9 consider the link to sustainable mix design and set out a research agenda.

II. Scope and Method of Review

This is a narrative critical review rather than a systematic one. Its purpose is to characterise how the field reasons, not to pool effect sizes, and it therefore prioritises depth of engagement with representative and highly cited studies over exhaustive coverage.

The literature considered comprises peer-reviewed journal articles that (i) predict the compressive strength of cementitious composites, (ii) include at least one industrial by-product among the input variables, and (iii) apply a learning algorithm trained on data rather than a fixed empirical expression. Studies of geopolymer and alkali-activated systems are cited where they illuminate a methodological point but are not treated as central, since their binder chemistry differs substantially. Work on other mechanical properties — tensile and flexural strength, elastic modulus — and on durability indicators is referenced where the modelling issues transfer.

Sources were identified from the reference networks of existing reviews in the area (Ben Chaabene et al., 2020) and from the citing literature of the foundational neural network studies, then extended by keyword searching across the major construction materials and computational engineering journals. Priority was given to studies that report their data provenance, validation protocol and error metrics in enough detail to be appraised. That criterion is itself informative: a substantial share of the published literature could not be evaluated on these terms, and this is discussed in Section 7.

III. Industrial By-Products and the Prediction Problem

Understanding why the prediction problem is hard requires a brief account of the materials involved.

Fly ash is the most widely studied SCM. Low-calcium (Class F) ashes are predominantly pozzolanic, consuming calcium hydroxide liberated by clinker hydration to form additional calcium silicate hydrate. The reaction is slow, and replacement levels above roughly 30 per cent typically depress 28-day strength while improving 90-day and later performance. High-calcium (Class C) ashes are partly hydraulic and behave differently. Crucially, both classes vary widely in glass content, fineness and oxide composition between and even within power stations, so nominally identical dosages of "fly ash" from two sources may perform very differently.

GGBS is latent hydraulic and can be used at higher replacement levels than fly ash, often 40–70 per cent. Its reactivity depends on glass content, fineness and the activation environment. Neural network models specific to GGBS concretes were reported comparatively early (Bilim et al., 2009), reflecting the practical importance of the material.

Silica fume is far finer and highly reactive, contributing both pozzolanically and by particle packing, and is normally used at low dosages. Its effects on strength are strongly non-linear and interact with superplasticiser demand.

Two further points matter for modelling. First, blends of two or more SCMs are common in practice and exhibit synergies not predictable from the individual components, which is precisely the regime in which linear or additive models fail. Second, the same nominal mixture cured differently — standard moist curing against accelerated or elevated-temperature regimes — will reach different strengths at the same age, and datasets that pool such records without labelling the conditioning regime encode a confound that no algorithm can resolve.

IV. Evolution of Predictive Modelling

A. Empirical and Statistical Baselines

The classical approach expresses strength as a function of water–cement ratio and age, with adjustment factors for admixtures. It remains valuable for its transparency and its grounding in physical reasoning. Its limitation is structural: adding a new variable requires recalibrating or replacing the expression, and interaction effects between multiple SCMs are not representable within it (Ben Chaabene et al., 2020). Multiple linear regression on mixture proportions improves flexibility but, as Yeh (1998a) demonstrated, explains only a modest fraction of variance in blended systems.

B. Artificial Neural Networks

The neural network era began with Yeh (1998a, 1998b) and expanded quickly. Ni and Wang (2000) and Lee (2003) confirmed that multilayer perceptrons could learn the strength surface from mixture and age inputs. Fuzzy and neuro-fuzzy variants were introduced to handle imprecision in input characterisation (Akkurt et al., 2004). Applications specific to by-product concretes followed, including GGBS concretes (Bilim et al., 2009) and self-compacting concretes with mineral additives under elevated temperature (Uysal & Tanyildizi, 2012).

Neural networks established the central claim of the field — that data-driven models substantially outperform regression baselines on blended concretes — but exposed two weaknesses that persist. They require careful architecture selection, and they offer no native account of why a prediction was made.

C. Evolutionary and Kernel-Based Methods

Two responses developed in parallel. One used evolutionary computation to generate explicit symbolic expressions: genetic programming and gene expression programming produce closed-form equations that can be inspected and, in principle, written into a specification (Castelli et al., 2013). The other applied kernel methods, particularly support vector regression, which behaves well on the small datasets typical of laboratory campaigns. Cheng et al. (2012) combined evolutionary optimisation with fuzzy support vector machines for high-performance concrete, and comparative studies began to appear systematically (Chopra et al., 2018). Evolutionary training of network weights was also explored as an alternative to gradient descent (Nikoo et al., 2015; Chandwani et al., 2015).

D. Ensembles and Gradient Boosting

The most consequential shift followed the maturation of tree ensembles. Random forests, and subsequently gradient boosting implementations including XGBoost (Chen & Guestrin, 2016), proved well suited to tabular mixture data: they capture non-linearity and interactions without feature engineering, tolerate heterogeneous scales, and are comparatively robust to modest dataset size.

Han et al. (2019) reported an improved random forest formulation for high-performance concrete, and Kaloop et al. (2020) demonstrated gradient tree boosting on the same problem. Ahmad et al. (2021) applied bagging, AdaBoost, gene expression programming and decision trees to concretes containing fly ash and blast furnace slag, reporting the bagging ensemble as strongest with a coefficient of determination of 0.92 under k-fold cross-validation, together with a sensitivity analysis of input contributions. Al-Shamiri et al. (2020) pursued a complementary objective, showing that competitive accuracy is attainable without extensive hyperparameter tuning — a practically important result given that tuning protocols are frequently under-reported.

By this point the accuracy differences between well-implemented model families on a common dataset had narrowed considerably. This is the plateau referred to in the introduction: the marginal return from a new algorithm applied to an existing benchmark is small.

E. Hybrid and Metaheuristic-Optimised Models

The subsequent direction has been hybridisation, in which a metaheuristic optimiser tunes the hyperparameters or structure of a base learner. Biswas et al. (2022) applied this approach to carbonation depth in fly ash concrete, and analogous constructions have been reported across strength, durability and elevated-temperature response (Zhou et al., 2024). Hybrid models often report improved test metrics. Whether the improvement reflects a better model or a more thoroughly optimised fit to a particular data partition is frequently impossible to determine from the information published, an issue returned to in Section 7.

V. DATA FOUNDATIONS AND THE BENCHMARK MONOCULTURE

The single most consequential feature of this literature is its data. Yeh's dataset of 1,030 records with eight input variables — cement, blast furnace slag, fly ash, water, superplasticiser, coarse aggregate, fine aggregate and age — is used in a very large share of published studies, either alone or as an external validation set (Yeh, 1998c). Its accessibility has been valuable: it permits direct comparison across algorithms, and it lowered the barrier to entry for computational researchers without laboratory access.

The costs are less often acknowledged. Three deserve emphasis.

First, repeated evaluation on a single fixed benchmark by a large community produces optimistic bias at the level of the field. Each study selects the configuration that performs best on that data; published results are therefore drawn from the upper tail of a very large implicit search. High reported accuracy on this dataset is now weak evidence about performance on new mixes.

Second, the feature set encodes by-products purely as mass dosages. As Section 3 argued, fly ash from different sources at the same dosage is not the same material. Models trained on dosage alone cannot represent this and will transfer poorly to by-products outside the compositional range of the training data. Studies that add oxide composition or physical characterisation to the input vector consistently find these variables informative, and the direction of travel in recent work is toward composition-aware feature sets.

Third, many laboratory-derived datasets are small — often one to three hundred records — which places them in a regime where variance across data partitions is large and single train–test splits are unreliable (Shaikhina & Khovanova, 2017). Repeated cross-validation is the appropriate response and is increasingly reported, though not universally.

VI. INTERPRETABILITY

The opacity of ensemble and neural models has motivated a substantial interpretability literature, with Shapley additive explanations (SHAP) as the dominant method. Ekanayake et al. (2022) provided an early demonstration for concrete strength prediction, and the approach has since been applied across CO2-cured concretes (Chu et al., 2025) and elevated-temperature systems, among others.

The findings are broadly consistent and broadly reassuring: cement content and age typically dominate, water content acts negatively, and SCM contributions display the threshold behaviour that materials science would predict. This consistency is genuinely useful, because a model whose attributions contradict established hydration chemistry should not be trusted regardless of its test metrics.

Two cautions apply. Attribution methods explain the model, not the material; a model that has learned a spurious association will produce a confident and entirely misleading explanation of it. And correlated inputs — cement content and water–binder ratio, for example — distribute attribution between themselves in ways that depend on the estimator, so feature importance rankings from these systems should be read as indicative rather than as measured physical influence. In much of the published work, SHAP analysis functions as confirmation of expectations rather than as a test capable of failing.

VII. CRITICAL APPRAISAL

Drawing the preceding sections together, four weaknesses recur across the literature.

External validation is rare. The great majority of studies validate on a random partition of the same dataset used for training. Because records within these datasets frequently originate from the same laboratory campaigns, random partitioning leaves training and test sets more similar than a genuinely new mix would be. Validation on independently produced mixtures, or partitioning by source campaign rather than at random, is uncommon; where it has been attempted, accuracy degrades relative to internal test performance, which is the expected and honest result.

Reporting conventions obscure error structure. The coefficient of determination is reported almost universally and often alone. It is a poor summary for engineering purposes: it is sensitive to the spread of the test sample and says nothing about where errors fall. A model with excellent aggregate accuracy that systematically under-predicts high-replacement, low-early-strength mixes is unsafe in exactly the region where the design question is difficult. Error distributions conditioned on replacement level and age are far more informative and are seldom presented.

Uncertainty is not quantified. Predictions are almost always reported as point estimates. Structural design operates on characteristic values with defined exceedance probabilities, so a prediction without an interval cannot enter a design workflow. Prediction intervals from quantile regression, conformal methods or ensemble variance are readily available and largely unused.

Comparisons are frequently not like-for-like. Where a study reports that model A outperforms models B and C, the tuning effort devoted to each is rarely stated. Given that the proposed model is normally the one the authors have optimised most thoroughly, such comparisons carry limited weight. Publication practice compounds this: negative results and failed transfers are underrepresented, so the aggregate literature overstates reliability.

None of these is unique to concrete technology; all are recognised difficulties in applied machine learning. They are noted here because the consequences in this domain are physical, and because the field's reporting norms have not yet adjusted to them.

VIII. From Prediction to Sustainable Mix Design

Prediction is an intermediate goal. The applied objective is mixture design that satisfies strength and durability requirements at minimum environmental and economic cost, and this is a constrained multi-objective optimisation problem in which a strength model serves as one constraint among several (DeRousseau et al., 2018). Optimisation of concrete structures against combined cost and carbon objectives has an independent history (Camp & Huq, 2013), and the two strands are now converging: a learned strength surrogate is coupled to an emissions and cost model, a Pareto front is generated, and a preference-ranking method selects among non-dominated solutions.

This integration is where data-driven methods can deliver value that conventional design cannot, because the search space of multi-SCM blends is too large to explore experimentally. It also raises the evidential bar. A surrogate used for interpolation within a well-sampled region is a modest extrapolation of existing practice; a surrogate used to nominate an optimal blend is being asked to extrapolate toward the edges of its training distribution, which is where the weaknesses identified in Section 7 bite hardest. Optimised mixes proposed on this basis require experimental confirmation before use, and reporting that confirmation — including where it fails — would materially strengthen the literature.

IX. Research Agenda

Five priorities follow from this review.

Composition-aware features. By-products should be represented by oxide composition, fineness and reactivity indicators alongside dosage. This is the most direct route to models that transfer across material sources.

Curated multi-source datasets with provenance. The field would benefit from larger open datasets that record source campaign, curing regime and material characterisation, enabling partitioning by source and honest assessment of transfer.

External and temporal validation as standard. Reporting performance on independently produced mixes, and on mixes postdating the training data, should become a normal expectation rather than an exception.

Uncertainty quantification. Prediction intervals should accompany point estimates, with calibration assessed and reported.

Physically constrained and hybrid formulations. Embedding known constraints — monotonicity in age, the form of the water–binder relationship — restricts the hypothesis space to physically admissible functions and tends to improve extrapolation, offering a middle path between empirical expressions and unconstrained learners.

X. Conclusion

Machine learning has demonstrably improved the prediction of compressive strength in concretes containing industrial by-products relative to classical empirical methods, and the argument for data-driven modelling of these systems is settled. The specific choice of algorithm, however, has become a secondary consideration: several model families achieve comparable accuracy on common benchmarks, and the marginal value of a further algorithmic comparison on existing data is low.

The binding constraints now lie elsewhere. Progress depends on richer material characterisation in the input space, datasets whose provenance permits meaningful partitioning, validation protocols that test transfer rather than interpolation, and reporting that conveys uncertainty and error structure rather than a single aggregate coefficient. Interpretability tools are a valuable addition but should be deployed as tests that a model can fail, not as post hoc reassurance. Addressing these issues would move the field from demonstrating that prediction is possible — which is no longer in doubt — toward models trustworthy enough to inform the mix design decisions that determine whether the sustainability benefits of industrial by-products are realised in practice.

References

Ahmad, W., Ahmad, A., Ostrowski, K. A., Aslam, F., Joyklad, P., & Zajdel, P. (2021). Application of advanced machine learning approaches to predict the compressive strength of concrete containing supplementary cementitious materials. *Materials*, 14(19), 5762. https://doi.org/10.3390/ma14195762

Akkurt, S., Tayfur, G., & Can, S. (2004). Fuzzy logic model for the prediction of cement compressive strength. *Cement and Concrete Research*, 34(8), 1429–1433. https://doi.org/10.1016/j.cemconres.2004.01.020

Al-Shamiri, A. K., Yuan, T. F., & Kim, J. H. (2020). Non-tuned machine learning approach for predicting the compressive strength of high-performance concrete. *Materials*, 13(5), 1–15.

Ben Chaabene, W., Flah, M., & Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. *Construction and Building Materials*, 260, 119889. https://doi.org/10.1016/j.conbuildmat.2020.119889

Bilim, C., Atiş, C. D., Tanyildizi, H., & Karahan, O. (2009). Predicting the compressive strength of ground granulated blast furnace slag concrete using artificial neural network. *Advances in Engineering Software*, 40(5), 334–340. https://doi.org/10.1016/j.advengsoft.2008.05.005

Biswas, R., Li, E., Zhang, N., Kumar, S., Rai, B., & Zhou, J. (2022). Development of hybrid models using metaheuristic optimization techniques to predict the carbonation depth of fly ash concrete. *Construction and Building Materials*, 346, 128483. https://doi.org/10.1016/j.conbuildmat.2022.128483

Bouzoubaâ, N., & Fournier, B. (2003). Optimization of fly ash content in concrete. Part I: Non-air-entrained concrete made without superplasticizer. *Cement and Concrete Research*, 33(7), 1029–1037. https://doi.org/10.1016/S0008-8846(03)00004-8

Camp, C. V., & Huq, F. (2013). CO2 and cost optimization of reinforced concrete frames using a big bang-big crunch algorithm. *Engineering Structures*, 48, 363–372. https://doi.org/10.1016/j.engstruct.2012.09.004

Castelli, M., Vanneschi, L., & Silva, S. (2013). Prediction of high performance concrete strength using genetic programming with geometric semantic genetic operators. *Expert Systems with Applications*, 40(17), 6856–6862. https://doi.org/10.1016/j.eswa.2013.06.037

Chandwani, V., Agrawal, V., & Nagar, R. (2015). Modeling slump of ready mix concrete using genetic algorithms assisted training of artificial neural networks. *Expert Systems with Applications*, 42(2), 885–893. https://doi.org/10.1016/j.eswa.2014.08.048

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Cheng, M.-Y., Chou, J.-S., Roy, A. F. V., & Wu, Y.-W. (2012). High-performance concrete compressive strength prediction using time-weighted evolutionary fuzzy support vector machines inference model. *Automation in Construction*, 28, 106–115. https://doi.org/10.1016/j.autcon.2012.07.004

Chopra, P., Sharma, R. K., Kumar, M., & Chopra, T. (2018). Comparison of machine learning techniques for the prediction of compressive strength of concrete. *Advances in Civil Engineering*, 2018, 5481705. https://doi.org/10.1155/2018/5481705

Chu, J., Guo, B., Zhong, T., Guan, Q., & Wang, Y. (2025). Explainable machine learning model for predicting compressive strength of CO2-cured concrete. *Case Studies in Construction Materials*, 22, e04589. https://doi.org/10.1016/j.cscm.2025.e04589

DeRousseau, M. A., Kasprzyk, J. R., & Srubar, W. V., III. (2018). Computational design optimization of concrete mixtures: A review. *Cement and Concrete Research*, 109, 42–53.

Ekanayake, I. U., Meddage, D. P. P., & Rathnayake, U. (2022). A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Studies in Construction Materials*, 16, e01059. https://doi.org/10.1016/j.cscm.2022.e01059

Han, Q., Gui, C., Xu, J., & Lacidogna, G. (2019). A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. *Construction and Building Materials*, 226, 734–742. https://doi.org/10.1016/j.conbuildmat.2019.07.315

Kaloop, M. R., Kumar, D., Samui, P., Hu, J. W., & Kim, D. (2020). Compressive strength prediction of high-performance concrete using gradient tree boosting machine. *Construction and Building Materials*, 264, 120198.

Kanamarlapudi, L., Jonalagadda, K. B., Jagarapu, D. C. K., & Eluru, A. (2020). Different mineral admixtures in concrete: A review. *SN Applied Sciences*, 2(4), 1–10. https://doi.org/10.1007/s42452-020-2533-6

Kosmatka, S. H., Kerkhoff, B., & Panarese, W. C. (2002). *Design and control of concrete mixtures*. Portland Cement Association.

Lee, S.-C. (2003). Prediction of concrete strength using artificial neural networks. *Engineering Structures*, 25(7), 849–857. https://doi.org/10.1016/S0141-0296(03)00004-X

Ni, H.-G., & Wang, J.-Z. (2000). Prediction of compressive strength of concrete by neural networks. *Cement and Concrete Research*, 30(8), 1245–1250. https://doi.org/10.1016/S0008-8846(00)00345-8

Nikoo, M., Torabian Moghadam, F., & Sadowski, Ł. (2015). Prediction of concrete compressive strength by evolutionary artificial neural networks. *Advances in Materials Science and Engineering*, 2015, 849126. https://doi.org/10.1155/2015/849126

Shaikhina, T., & Khovanova, N. A. (2017). Handling limited datasets with neural networks in medical applications: A small-data approach. *Artificial Intelligence in Medicine*, 75, 51–63.

Uysal, M., & Tanyildizi, H. (2012). Estimation of compressive strength of self compacting concrete containing polypropylene fiber and mineral additives exposed to high temperature using artificial neural network. *Construction and Building Materials*, 27(1), 404–414. https://doi.org/10.1016/j.conbuildmat.2011.07.028

Yeh, I.-C. (1998a). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, 28(12), 1797–1808. https://doi.org/10.1016/S0008-8846(98)00165-3

Yeh, I.-C. (1998b). Modeling concrete strength with augment-neuron networks. *Journal of Materials in Civil Engineering*, 10(4), 263–268.

Yeh, I.-C. (1998c). *Concrete compressive strength* [Data set]. UCI Machine Learning Repository. https://doi.org/10.24432/C5PK67

Zhou, J., Tian, Q., Ahmad, A., & Huang, J. (2024). Exploring the resilience of supplementary cementitious materials-based concrete to elevated temperatures via modern computing techniques. *Materials Today Communications*, 38, 107894. https://doi.org/10.1016/j.mtcomm.2023.107894