Ultra-Accurate CO_2 Emission Forecasting for the Cement Industry Using FbOA-Optimized Neural NODE Models

Omnia M. Osama[1,*], **Marwa M. Eid**[2,3] **, El-Sayed M. El-Rabaie**[4]

[1]Department of Communications & Electronics,Delta Higher Institute of Engineering & Technology, Mansoura, Egypt

[2]Faculty of Artificial Intelligence , Delta University for Science and Technology , Mansoura 35111, Egypt

[3] Jadara Research Center , Jadara University , Irbid 21110, Jordan

[4]Faculty of Electronic Engineering , Menoufia University , Department of Electronics and Communications , Menouf 32952, Egypt

Emails: Omnya.osama@dhiet.edu.eg; mmm@ieee.org; srabie1@yahoo.com

Abstract

The cement sector is a linchpin of global infrastructure and is also one of the world's most significant industrial sources of CO2 emissions, accounting for about 7-8% of anthropogenic emissions. The proper prediction of cement-generated emissions is thus essential for designing mitigation strategies, planning industrial transitions, and evaluating progress toward carbon-neutrality goals. This paper proposes a new time-series forecasting model that combines Neural Ordinary Differential Equations (NODE) with the Football Optimization Algorithm (FbOA) to enable automated, data-driven hyperparameter optimization. The performance of NODE is compared with Seq2Seq and ConvLSTM models for global CO2 emis-sions from cement production in baseline settings, and subsequently metaheuristically optimized using FbOA, PSO, MVO, WOA, and GA. The baseline experiments demonstrate that NODE, with an MSE of 0.00745, RMSE of 0.0863, MAE of 0.0515, and high levels of agreement (NSE = 0.91, WI = 0.905), outperforms both Seq2Seq and ConvLSTM. Upon hyperparameter optimization, the FbOA + NODE combination achieves significant performance improvement, with MSE of 3.95×10^{-7}, RMSE of 6.28×10^{-3}, and MAE of 3.42×10^{-4}, r = 0.977, R2 = 0.973, NSE = 0.975 and WI = 0.98. Competing optimizers (PSO, MVO, WOA, GA) also improve NODE's performance, and across all important metrics, they are consistently below FbOA. The findings indicate that integrating NODE and FbOA yields an accurate, sta-ble, and computationally inexpensive model for predicting cement-associated CO2 emissions, offering a potential avenue for data-driven climate and industrial planning.

Keywords: CO_2Emission Forecasting; Neural Ordinary Differential Equations (NODE); Metaheuristic Optimization; Football Optimization Algorithm (FbOA); Cement Industry Decarbonization

1 Introduction

The cement business world globally supports almost all aspects of contemporary socioeconomic growth and can be regarded as the main infrastructure of urbanization and industrialization. The cement-based building materials play a vital role in the development of transportation channels, domestic structures, power plants and water supply channels and mains [1]. With the ongoing urbanization of emerging economies, as well as the large-scale renovation and resilience efforts of industrialized countries, cement demand is expected to be strong and steady in the future as well [2]. Yet, this essential position has massive environmental implications. Combination of calcification of limestone and combustion of fuel in rotary kilns results in massive amounts of CO2 and cement production is classified as one of the most emissions-intensive industrial processes [3]. Recent estimates reveal that cement production is a major contributor of 7-8% of the anthropogenic global CO2 release into the atmosphere [4]. This scale makes the cement industry a key focus of the international process of decarbonization and technological development and change [5] [6]. Precise projection of emissions of CO2 by cement manufacturing is thus a critical endeavor in both science and policy making. Projections help to inform the national climate policy, influence investment in carbon-capture technologies, and help to plan the future industrial capacity, as well as to meet the international agreements like the Paris Accord [7]. Credible prediction systems can help governments and industries to consider possible routes to reduction of emissions, examine the potential of net-zero commitments, and develop active interventions which are based on empirical evidence. However, it is still difficult to have confident forecasting because emission dynamics are complex in nature. It is not only the volume of production which produces variability, but also the changing efficiencies of industries, changes in the type of fuel, regulatory influences, supply-chain upheavals and fluctuations in the macroeconomic environment, all contribute to variability as well [8]. The combination of these interacting factors produces very nonlinear and multiscale temporal behaviors which are beyond the modeling ability of most traditional methods [9]. Historically, classical statistical and econometric forecasting models, which include ARIMA, VAR and linear regression families, have been the common instruments in industrial and environmental prediction. Although they can be used effectively to capture the linear over a short-term analysis, they tend to fail at modeling the nonlinear, non-stationary and dynamically changing structure of long-horizon CO2 emission paths. In addition, they usually assume a lot when it comes to the understanding of temporal stability, error distributions and feature interaction [10]. These assumptions tend to be broken in the real-life industrial scenario especially in the global system where sudden policy interventions, technological disturbances, geopolitical restrictions and energy-price shocks are activated. Consequently, data-driven, representation-learning-based methods that have the capability to automatically detect complex dependencies are increasingly becoming popular without any restrictive parametric assumptions being made on them [11]. Recent deep learning (DL) developments have changed the landscape of time-series forecasting dramatically in terms of what it can do and how it can do it [12] [13]. Recurrent neural networks, convolutional temporal models, transformation sequence learners, and recently Neural Ordinary Differential Equations (NODEs) have shown good performance in the fields of environmental, economic, and industrial forecasting. Of them, Temporal NODEs (T-NODE) offer one of the most interesting structures as they model the dynamics of a time-dependent process using neural networks as parameters. The formulation enables T-NODE to deal with behavior in continuous time, long range dependencies and to represent irregular temporal behavior flexibly. Such characteristics are particularly beneficial to industrial emission systems that develop over a heterogeneous time scale and that have both slow and sudden variations [14]. The promise of NODE-based architectures notwithstanding, these designs are extremely delicate to design-time choices, including network depth, the configuration of a differential-equation solver, the dilation behavior, the dynamics of learning-rates, and regularization choices. The process of hyperparameter selection is critical in deciding whether a model can also be able to

generalize beyond historical trends, whether it becomes overfitted, and whether it will be stable to changes in time. Manual tuning is time-sensitive, relies on the intuition of experts and is likely to achieve less than optimal configurations. Also, the industrial time-series data is usually associated with noises, structural breaks, and regime shifts, which further complicate optimization. As a result, the combination of metaheuristic optimization algorithms and deep learning models has become a potential solution to automatically optimize hyperparameters and increase the robustness of the models. Even with these developments, there are a number of fundamental issues that hamper the reliability of forecasts. A single recurring problem is drift over time - systematic variation in the underlying data-generating process that is due to the passage of time, due to changing technology, regulatory regimes, fuel consumption, or macroeconomic changes. The temporal drift compromises the premise that historical data can be used to predict future behavior, which compromises generalization of the model. Moreover, the sensitivity of hyperparameters in NODE architectures is very high due to its high dimensionality and flexibility of structure. Small variations can bring significant changes in the stability of the training process, its convergence behavior, and predictions. Lastly, deep learning models have a latent problem of overfitting, particularly in areas with a small amount of data or a fast-evolving trend. The systematic optimization strategies that are capable of exploring the hyperparameter space systematically are needed to ensure good generalization of the model to future time periods. To overcome these difficulties, the current paper constructs a superior time-series predictive model, which incorporates the Football Optimization Algorithm (FbOA) as a newly developed population-based metaheuristic and the Temporal NODE model. This integration is intended to support automated, data-driven hyperparameter optimization that can be used to find effective architectural and training settings with no human effort. The framework is implemented on the popular dataset of global CO2 emission because of the cement manufacturing, providing the information on the interactions between the metaheuristic search mechanisms and the deep dynamical modeling. The suggested system aims to enhance the predictive accuracy, improve the time stability and offer a methodology that is reproducible and can be used in research, policy formulation and industry analytics. In addition to a better performance, the work will also be able to add to the overall methodological discussion by comparing the FbOA–T-NODE integration with the various tested and recognized metaheuristic-based-NODE combinations such as Particle Swarm Optimization (PSO), Multiverse Optimization (MVO), Whale Optimization Algorithm (WOA), and Genetic Algorithm (GA). These comparisons can be taken as an overall assessment of optimization behaviors and show the effects of different metaheuristic paradigms on the learning processes of NODE-based models. The study contributes to the creation of the next generation of forecasting systems of industrial emission systems by offering a scalable and transparent as well as computationally efficient pipeline. The significant contributions of this study are as follows.

- **Integrated Framework:** Introduction of a hybrid forecasting pipeline that unifies Temporal NODE with the Football Optimization Algorithm, enabling fully automated hyperparameter selection and enhancing the adaptability of NODE modeling to complex industrial time-series.
- **Systematic Benchmarking:** Comprehensive comparative evaluation involving nine established optimizers, including PSO, MVO, WOA, and GA, demonstrating how differing metaheuristic strategies influence NODE training behavior and predictive effectiveness.
- **Performance Advancement:** Demonstration of meaningful improvements in predictive consistency, robustness to temporal drift, and long-term forecast reliability, underscoring the value of the proposed approach for CO_2 emission analytics and broader environmental forecasting applications.

The rest of this paper is organized as follows. Section 2 offers a thorough literature review of the recent trends in industrial CO2 forecasting, temporal modeling based on deep learning and metaheuristic optimization. Section 3 identifies the materials and methodology, data formatting, modelling style, optimization set-up, and performance. The experimental evaluations and detailed analyses are given in Section 4. Section 5 gives final remarks on the major findings and sets forth the avenues of future research.

2 Literature Review

The integration of machine learning (ML) into the research of cement and concrete has transformed the scientific and industrial world with the possibility of complex material behaviour analysis, modelling of emissions, and optimization of mixtures. This trend is highly connected with the need to decrease the carbon footprint of the construction sector and keep its performance levels high. One of the most active research fields is ultra-high-performance concrete (UHPC) that, despite the high-mechanical and rheological properties, is severely reproached because of its high embodied carbon. To overcome these issues, a hybrid ML-based model has been developed as the first step involves training of Artificial Neural Network (ANN) to model compressive strength and slump flow. The model has less than 10% prediction error which is very strong. The outputs of these ANN predictions are introduced as inputs to a Genetic Algorithm (GA) which explores a high dimensional space of UHPC constituents subject to strength, workability, components proportions and volumetric consistency constraints [15]. The result is the optimization of UHPC mix with a significantly lower carbon footprint of 688 kg/m^3 which reveals the practical possibility of multi-objective optimization via ML to design more sustainable high-performance concretes. There is also a complementary research on the implementation of ML in the optimisation of high-strength concrete (HSC) and ultra-high-strength concrete (UHSC). The materials are very popular because of their superior strength characteristics, but they produce a lot of carbon emission owing to their high cement content. One of them collected a large experimental dataset of the literature, which created ML models that could predict compressive strength using constituent materials. Those predictions enabled the calculation of embodied carbon of each mix to be made [16]. Notably, the research found out that various mix designs can attain the same compressive strength but have great variance in embodied carbon. The authors suggested the use of benchmark intervals based on statistical analysis and ML-based trend mapping which characterizes the acceptable embodied-carbon of various strength grades. These insights play a vital role in guiding concrete professionals that aim to come up with low-carbon mixes without compromising the mechanical performance. The focus has also been on the calcification operation of production of clinker, which is one of the major contributors of CO2 in the process. Since the emissions throughout the calcinations are sensitive to chemical composition, particle size distribution, and thermal exposure, the traditional process models fail to account the nonlinear frameworks that determine the decarbonisation reactions. In order to eliminate these constraints, a number of advanced AI-based models, such as deep neural networks (DNN), ant colony optimization-enhanced ANN (ACO-ANN), and genetic algorithm-enhanced ANN (GA-ANN) have been tested in order to predict CO2 emissions of raw materials undergoing the calcination process [17]. The high level of modelling accuracy with R2 values greater than 0.99 and average relative errors lower than 1.04%, particularly of the DNN, confirms that ML is an effective method to model any complex thermochemical reactions in the clinker production. At the macro-level, national-level CO 2 emissions in the cement industry have been predicted using ML. An in-depth research used Improved Particle Swarm Optimization-Backpropagation (IPSO-BP) predictive model on 44 scenarios according to the second generation of dry cement production technologies [18]. The results show that Chinese emissions of cement-related CO2 can reach their maximum level long

before 2030 through measures aimed at reducing the capacity as well as introducing the latest technology of dry-process. Under different circumstances, the peak emissions may be postponed by almost twenty years as compared to the business-as-you-go scenario with a peak of 742 Mt. These long-term forecasting models cannot be ignored by the policy makers who want to adjust the industrial development towards the national carbon-neutrality targets. At the structural and mix-design stage, ML has facilitated the creation of innovations in cement-use optimization in those elements that gain in terms of deferred loading. Conventionally, mix proportioning is founded on a 28 day compressive strength; but most structural elements; foundations, pavements, etc. do not load until long after they have cured. In one of the studies, the ANN and regression techniques were implemented to estimate cement content needed to attain 28 day and 90 day strengths [19]. Elastic Net regression was found to be the most dependable predicting tool with prediction accuracy of as high as 94%. Putting these ML-based predictions on a real reinforced concrete project, the authors predicted a 10% cement consumption decrease and a 10% carbon emission decrease, which is a practical environmental benefit of ML-based design optimization. The other significant contribution to national scale modelling involved a national-level carbon-emission inventory of the cement industry in China to evaluate the predictive power of six ML models namely: ridge regression, polynomial regression, Random Forest, support vector regression, gradient-boosted regression trees and feed-forward neural networks [20]. The best predictive accuracy (R2 = 0.99) was achieved by using the models of the neural network and the polynomial regression. The analysis on feature-importance indicated that the activities of producing clinker and using coal were the factors contributing to almost 95% of the national cement-related emission. Spatial emission trends were also offered in the study, in which the country cement-sector emissions are mainly led by North, South and Southwest China. The findings indicate the importance of model interpretability in the development of regional decarbonization plans. Continuing on the use of ML to more detailed process modelling, another study used six ML methods on more than 6000 historical manufacturing data to investigate the degrees of oxidation and calcinations in cement calcinations [21]. The research determined process parameters that have the greatest impacts on CO2 release, by targeting the CO2 molecular composition and subjecting the sensitivity analysis to them. These lessons deepen the knowledge of process engineers regarding thermal processes and give specific channels of making kilns and calciners more efficient. A growing body of work also focuses on modelling CO_2 sequestration within cement-based materials. A study integrating extensive experimental datasets with literature data applied Decision Trees, Random Forests, and XGBoost to predict carbonation depth, a critical indicator of CO_2 uptake [22]. XGBoost significantly outperformed linear regression, demonstrating its ability to capture nonlinear and interaction-dominated relationships. With aid from SHAP interpretability tools, the study also revealed the previously underappreciated role of cement type—especially CEM II/B-LL and CEM II/B-M—in influencing carbonation behavior. These findings foster a deeper understanding of the chemical drivers of carbon sequestration. ML has also been used to model kiln-stack CO2 emissions. Based on 22 operational variables of a clinker manufacturing line in UAE, researchers tested models like KNN, linear regression, decision trees,Random Forest, Gradient Boosting and ANN models [23]. Random Forest and Gradient Boosting turned out to be the most dependable, with the R2 values of 0.984 and the small values of RMSE/MAE. It happens that kiln feed rate became the key variable influencing the amount of stack emissions, and consequently, focused operational changes might be directly linked to the enhancement of emission levels. The replacement of biochar (BC) with cement mortar has been gaining momentum. A comprehensive experimental study showed that BC increases compressive and tensile strengths, decreases water absorption and enhances the durability because of its pozzolanic action and porous microstructure [24]. The presence of the mechanical and durability properties under the influence of BC modifications in mortars was also tested through complementary ML analysis based on AdaBoost and linear regression, where AdaBoost was much more precise. This brings out the interplay of experimental and computational methods of developing sustainable cement mortars. The issue of uncertainty in combus-

tion systems also presents a problem to emission control. One study that solved this problem involved ANN surrogate models that were trained on 700 dynamic simulations of kiln and calciner operations with 10% uncertainty in the major input flows [25]. These obtained surrogates were applied to GA, PSO and hybrid GA-PSO optimizers. The hybrid model offered the best stable and correct answers, and sensitivity analysis later showed that the total coal and tertiary air flow were the most sensitive parameters on the CO2 and CO emissions. This paper has shown that intelligent optimization can be realised in real-time in industrial cement operation. ML is also useful in the use of calcined sludge as a partial replacement of cement. The six variables were experimented with six ML regressors, such as CNN, ensemble regression, MLP-ANN, SVR, and Random Forest, to predict compressive strength [26]. Ensemble regression and CNN were the most successful, and ML-based robustness assessment proved that curing age is the strongest variable affecting strengths. The results allow the preparation of sustainable waste-based binders that have a more predictable behavior. Another area where ML has won over deficiencies of semi-empirical kinetic models is in hydration kinetics of OPC combined with mineral additives. A single study showed the successful prediction of time-dependent observation of hydration by use of physiochemical attributes on the models with the help of the Random Forests methodology as an input variable [27]. Even the model could be used to produce mix designs which meet user-specified kinetic criteria to achieve targeted performance specifications without necessarily having mechanistic kinetic models. The sustainability usefulness of geopolymer mortars that include eggshell powder (ESP) and rice husk ash (RHA) was investigated with the help of both Response Surface Methodology (RSM) and the ML-based prediction one [28]. GPR, ANN, and Gradient Boosting models were trained on a dataset of 606 compressive strength results and GPR was the most accurate model. The experimental findings indicated that the strength and CO2 emission were better compared to the conventional binders and reflected that the mix optimization of alternative binders by use of ML was practical. Additional developments in the modeling of cement-based materials were also described in an article evaluating the five ML algorithms, which include SVM, Random Forest, decision tree, AdaBoost, and KNN, to predict compressive strength using a large experimental database [29]. The best performance was attained using the AdaBoost and the Random Forest models, and variable influence analysis gave good information on the influence of cement grade, curing age, water to binder ratio, and sand properties on strength development. Lastly, the use of ML has also been extensively used to maximize flexural strength (FS) of cementitious composites with waste glass powder (WGP). Experimental studies revealed that WGP improves FS as far as 15% replacement depending on the type of cement or fine aggregate substituted [30]. The six input parameters were trained on with two models of the ML- SVM and a bagging regressor, the bagging regressor provided the best predictive performance. This paper demonstrates that ML can make the assessment of waste-based cementitious composites faster and helps achieve the larger aim of sustainable materials.

3 Materials and Methods

2.1 Dataset Overview

The dataset used in this study is sourced from the publicly available Zenodo repository `https://zenodo.org/record/4738593` , which provides a long-term historical record of global CO_2 emissions from cement production. It integrates multiple authoritative sources—including the U.S. Geological Survey (USGS) and earlier editions of CDIAC's cement-emission reconstructions—to generate a consistent and continuous

time series across numerous countries and decades. This makes the dataset particularly suitable for temporal modeling, long-range forecasting, and comparative analysis across geographic regions.

The dataset's construction begins with production values back-calculated from CDIAC's 2019 methodology, which directly linked cement emissions to production volumes. Additional disaggregation is performed for former Soviet states to maintain country-level continuity prior to geopolitical transitions. Starting from 1990, USGS data overwrite earlier values to ensure modern accuracy. For countries lacking recent production records, simple extrapolation is used to extend the time series, and updated national statistics are incorporated for selected countries where superior data sources are available. Despite these refinements, the dataset retains several zero entries that likely represent missing rather than actual production values; such entries require careful preprocessing to avoid distorting model behavior.

The dataset is organized by year and country. A representative portion of the dataset's feature structure is shown in Table 1, presenting the exact feature labels as they appear in the raw dataset.

Table 1: Representative Feature Structure and Descriptions from the Cement Production Dataset

Feature	**Description**
Year	Calendar year corresponding to each recorded cement production entry.
Afghanistan	Annual cement production attributed to Afghanistan, expressed in metric tons as reported or reconstructed in the dataset.
Albania	Annual cement production for Albania, representing country-level output for each year.
Algeria	Yearly cement production quantities for Algeria, contributing to regional and global emission estimates.
Andorra	Recorded or estimated cement production values for Andorra, included for completeness in global coverage.
Angola	Annual cement production associated with Angola, derived from USGS and CDIAC-based reconstruction.
Anguilla	Cement production values for Anguilla, typically low or zero due to limited industrial activity; values may include reconstructed or missing-data proxies.
Antigua and Barbuda	Country-level annual cement production for Antigua and Barbuda, reflecting small-island reporting patterns.
Argentina	Annual cement production for Argentina, a major regional producer with relatively well-documented historical records.
Armenia	Cement production levels for Armenia, including reconstructed values for pre-1991 periods and direct observations thereafter.

To study the tidiness of climate variables over time, one needs analytical instruments that can help to separate long-term trends and other short and medium-term variations. It can be very useful through the wavelet decomposition which is a strong multiresolution method to decompose a time-series into its own parts across multiple frequency bands. Such decomposition is of particular importance in global temperature variations, where one would wish to isolate persistent signals of warming in cyclical variations in the climate due to natural climate swings, volcanic eruption or temporary external perturbations. Figure 1 shows the

application of a discrete wavelet transform (Daubechies 4, level 4) onto a global temperature series, giving an idea of how the initial signal can be gradually broken down into its approximation and detail subsections. This representation gives a clear picture of long term climatic trends as well as multi-scale variability hence offering a better interpretation to the structural warming patterns and high frequency climate behaviour.

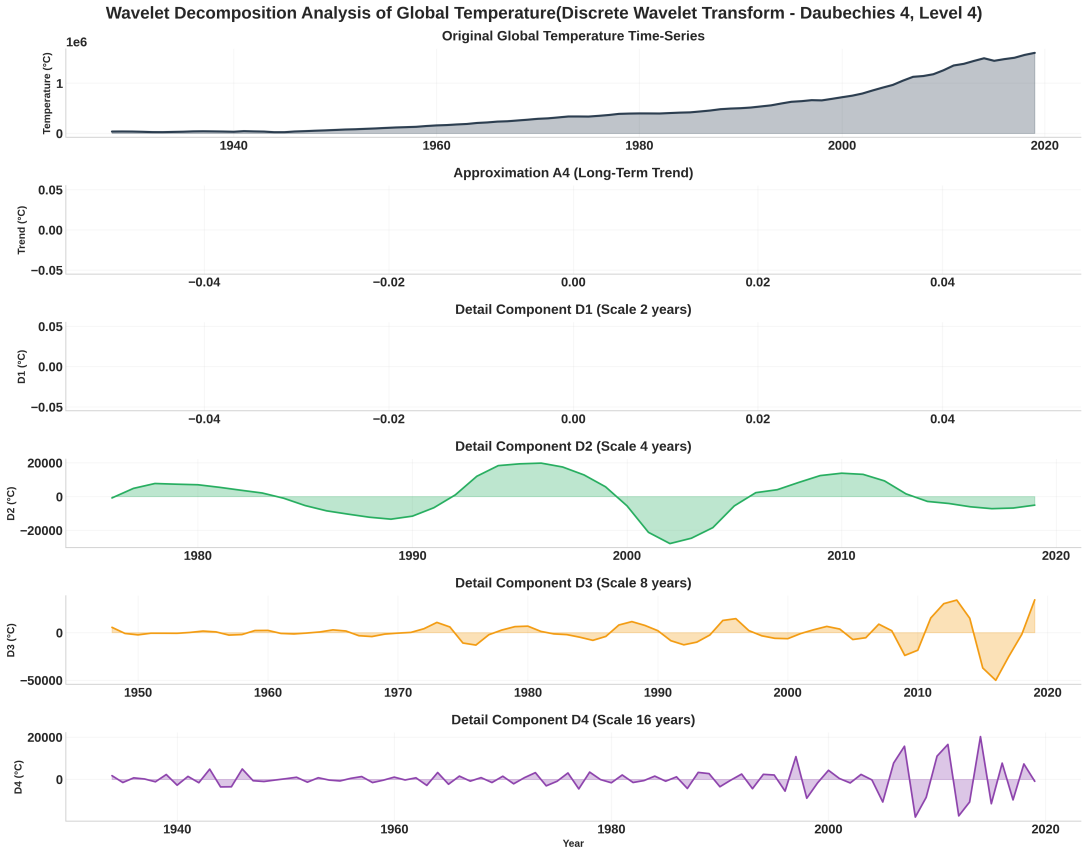

Figure 1: Wavelet decomposition analysis of global temperature using a Discrete Wavelet Transform (Daubechies 4, Level 4), illustrating the original series along with approximation (A4) and detail components (D1–D4).

The temporal dependence structure of climatic variables is important to study both short- and long-memory climatic time-series dynamics. Among these, the autocorrelation and partial autocorrelation diagnostics directly offers a method of the evaluation of the persistence, lag effects, and potential model forms that can be used in forecasting applications. Figure 2 shows the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the global temperature series, which provides insights as to how the signal is strongly positively correlated over a series of lags and to show how the signal is gradually decaying as a result of long-term climatic persistence. These relationships patterns provide essential understanding on the underlying temporal dynamics of global temperature fluctuations and forms a basis stage on the selection of correct predictive models.

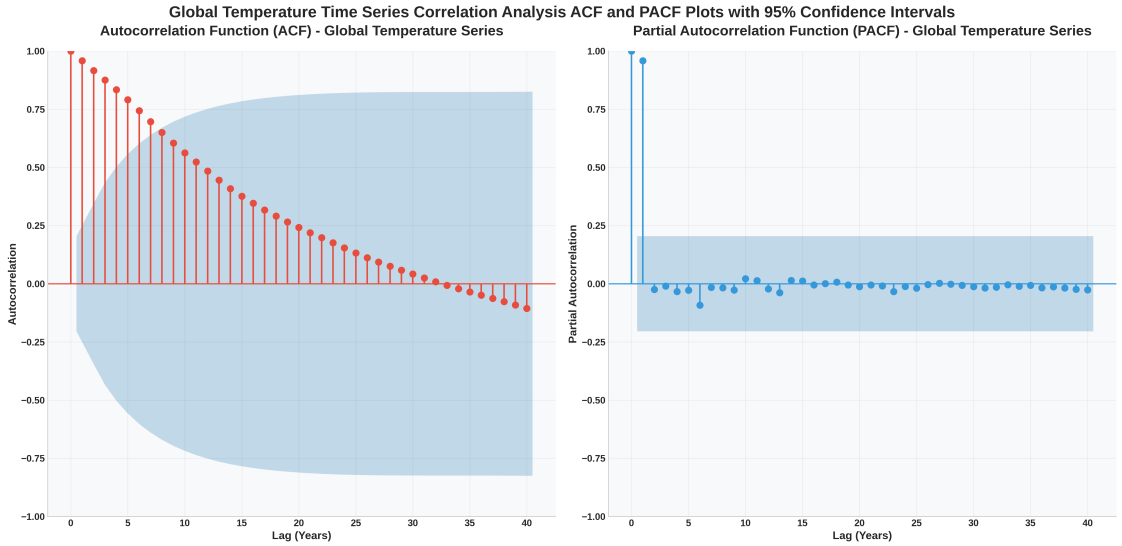

Figure 2: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for the global temperature time series, displayed with 95% confidence intervals.

2.2 Data Preprocessing

The preprocessing of the dataset to support temporal modeling involves a number of systematic preprocessing operations to resolve structural discrepancies, lost data and cross-country and cross-tempor heterogeneity in the manner data is being reported. Since the cement-production data is a combination of past reconstructions and current statistical reporting, the data in its raw form are irregular which need to be removed before training the model. These procedures are necessary to provide a stable, coherent and analytically significant input space upon which the forecasting framework is to be run. One main issue is how zero values brought forward by the previous CDIAC compilations should be treated. Most of these zero records, as recorded by the authors of the dataset, are not actual records of zero cement production but are instead indeterministic records of missed records or placeholders of missing data. Unattended, these values may cause artificial discontinuities, bias temporal dynamics, and have other negative impacts on the learning dynamics of deep models. Thus, all The zero values which do not correspond to confirmed production statistics are re-classified as missing values (NaN). Such method is consistent with best practices in time-series curation of environmental data, where the structural zero should be differentiated with the actual physical zero, otherwise the trend and seasonality elements may be skewed. Following the reclassification, missing values are imputed using a two-step imputation plan that aims at maintaining the statistical properties of the production path of each country. First, when short gaps occur in the time series of a country they are solved by means of linear interpolation because of the appropriateness of the method to gradual industrial trends and in consideration of the fact that monotonicity may be maintained when the underlying production varies smoothly with time. This method allows to give a consistent reconstruction of intermediate values without the introduction of unrealistic fluctuations. Second, in the case of longer gaps or intervals influenced by geopolitical transitions, e.g. dissolution events or incomplete national reporting, country-specific imputation is used.

This involves forward and backward filling using the nearest effective observations or where it is feasible, reconstruction using regional patterns and production intensities recorded by the metadata sources used in the dataset. Normalization is another processing step that is undertaken at the end of the preprocessing process to align production values among countries which have a significantly different industrial capacity. Minimum and maximum scaling strategy is used to maintain the relative form of the temporal variations as well as to ensure numerical stability in training the model. The scaling provides gradient convergence in the NODE architecture and avoids the dominance of high-output nations like China, India or the United States. Following scaling, the data is divided into training, validation and testing parts after an initial chronological partition to avoid temporal leakage, which is a frequent issue in time-series forecasting and can artificially improve predictive accuracy. By using this extensive preprocessing pipeline, the data is suitably formatted to be used in the forecasting structure of NODE. The transformation of the zeros into the missing values coupled with principled imputation and scaling processes will make the final data set analytically adequate and representative of the actual historical dynamics of production. This design indicates a fine spatial resolution of the dataset and could be used to conduct country-level, regional, and global forecasting studies. The characteristics that are depicted here will represent the format that is utilized in all of the featured nations and facilitate multi-variate temporal modeling frameworks.

2.3 Deep Learning Models

Three deep learning networks are used to assess the performance of forecasting over time in various temporal modeling paradigms, namely: Neural Ordinary Differential Equations (NODE), Sequence-to-Sequence (Seq2Seq) recurrent models, and Convolutional Long Short-Term Memory networks (ConvLSTM). Both models are unique methods of modelling temporal relationships, non-linear dynamics and structural tendencies found in long-term cement production data. Their joint analysis gives a strong comparative foundation on analyzing the sensitivity of models with the nature of data, optimization processes, and time variations.

Neural Ordinary Differential Equations (NODE) Colloquially, neural Ordinary Differential Equations (NODE) by Chen et al. (2018) are the reinterpretation of deep networks not as discrete layered transformations but as continuous-time dynamical systems. NODE does not stack many layers, as it learns the dynamics of a hidden state in an ordinary differential equation that is of the form:

$$\frac{dh(t)}{dt} = f(h(t), t; \theta),$$

where $f(\cdot)$ is a neural network parameterized by θ. The framework allows the model to model smooth temporal trajectories, and continuous-time behaviour which are especially well adapted to long-range industrial and environmental time-series. NODEs offer adaptive computation, as well, in ODE solvers where step sizes are varied in response to the complexity of the system to enable effective modeling of periods with sudden variations to those with steady trends. NODE has benefits in cement production forecasting in terms of dealing with non-stationary effects that are caused by technological change, change in regulations and macroeconomic shocks. It is especially useful in irregular temporal intervals, extrapolation, and complicated nonlinear processes models owing to its capability to model temporal dynamics as continuous flows.

Sequence-to-Sequence (Seq2Seq) The Sequence-to-Sequence architecture which was initially designed to translate neural networks into machine translation now forms a core architecture of multi-step time-series prediction. A typical Seq2Seq model is made out of two recurrent modules:

- an **encoder** that processes an input sequence and compresses its temporal information into a context vector, and
- a **decoder** that generates the forecasted sequence conditioned on the learned context.

The use of Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) cells is due to the fact that these models can alleviate vanishing and exploding gradients. Seq2Seq models are also capable of capturing long-range dependencies and input sequences with variable length to output sequences with variable length, and therefore are useful in multi-horizon forecasting tasks. To handle data on cement production, Seq2Seq allows the model to pick up temporal regularities like growth trends, cyclicality and structural discontinuities in past emissions. The autoregressive decoding process enables the model to make refinements stepwise with the help of contextual memory of years past. This can be applied especially in forecasting in areas where structural dynamics are slow changing or inertial.

Convolutional Long Short-Term Memory (ConvLSTM) Convolutional Long Short-Term Memory (ConvLSTM) networks are based on the same pattern as in the LSTM networks, except that they add convolutional operations to the gating mechanisms. As opposed to regular LSTMs, which use fully connected transformations, ConvLSTM uses spatially structured convolutions, and thus it is especially good at capturing time-wise and spatial correlations. ConvLSTM has been effective in any field that involves the temporal evolution that is governed by the structured spatial relationships although it was initially developed to predict precipitation nowcasting and spatiotemporal prediction. In the cement production dataset, the spatial dimension is the group of countries, which is a spatial unit with the possible correlation of production trends in the region because of the economic integration of regional economies, similarity of industrial activity, or geopolitical cooperation. ConvLSTM can learn to capture:

- regionally correlated production behaviors,
- cross-country interactions,
- multi-scale temporal patterns,
- and shared economic or regulatory dynamics.

This space-time dualization allows the architecture to acquire more distinct representations than recurrent-only or convolution-only models, and ConvLSTM is a good candidate to operate at the scale of global predictions. Combined, NODE, Seq2Seq and ConvLSTM are complementary modeling paradigms of continuous-time dynamics, recurrent sequence learning and spatiotemporal convolutional modeling. The two-part assessment is a holistic evaluation of the various deep learning processes to the multiplicity of long-term cement production forecasting.

3.3 Optimization Strategy

Effective hyperparameter optimization is essential for achieving robust performance in deep learning–based forecasting models, particularly when dealing with complex, nonlinear, and non-stationary temporal data such as global cement-production records. Manual tuning is often inefficient, subjective, and insufficient for navigating high-dimensional search spaces. This study therefore employs metaheuristic optimization techniques that enable systematic, global exploration of hyperparameter configurations for the NODE, Seq2Seq, and ConvLSTM models.

Role of Metaheuristics Metaheuristic algorithms offer a general, gradient-free way to optimize objective functions which have a nonlinear appearance, many local minima and nonlinear variable interactions. By using their population-based searching methods they are able to balance both exploration and exploitation, adaptively refine their candidate solutions and efficiently search several large parts of the hyperparameter space. The above properties make metaheuristics particularly useful to deep learning, where the performance of a model tends to vary significantly with the learning rate, depth of architecture, solver choices (in the case of NODE), sequence length (in the case of Seq2Seq), and convolutional parameters (in the case of ConvLSTM). Metaheuristics are useful in forecasting time series of industry and the environment to enable models to be generalizable to regimes of variability, structural breaks and gradual long-term shifts. They also help avoid trial-and-error methods of hyperparameter selection and increase results reproducibility across experiments by automating the process of hyperparameter selection. This is of particular significance to NODE because the optimization path can be highly dependent on solver behavior and starting point; hence, their capacity to leave local minima.

3.3.1 Comparative Algorithms

In order to measure the performance of the proposed optimization framework, a few well established metaheuristic algorithms are taken as some baselines. These optimizers are based on different biological, physical and evolutionary paradigms and a wide range of global search mechanisms are availed.

Particle Swarm Optimization (PSO). PSO is informed by the flock movement of birds and the school of fish. It moves its particles through the search space using both the personal and global best experience to update the velocities. PSO has been noted to be fast converging and has been commonly used in tuning hyperparameters of neural networks because it balances between exploration and exploitation.

Multiverse Optimization (MVO). MVO is an algorithm based on physics with multiverse theory. It imagines solutions as being composed of universes, which are linked together by white holes, black holes and wormholes which direct information exchange and perturbation. This design enables MVO to search very multimodal landscapes with efficiency and prevents early convergence.

Whale Optimization Algorithm (WOA). WOA is a behavior that resembles the hunting of humpback whales with the use of a bubble-net. It switches between encircling prey, spiral updates and random search, which allows the intensification around promising regions and diversification where necessary. The adaptive exploitation mechanisms of WOA make it suitable in fine-tuning the neural network parameters.

Genetic Algorithm (GA). GA is an evolving algorithm which is founded on natural selection and genetics. It mutates, selects and crossovers a population of candidate solutions. GA is diverse in solutions and able to efficiently search through complex spaces in recombination such that it represents a strong and long-standing basis on which optimization is performed. All these algorithms together provide a powerful reference on the performance of the suggested metaheuristic strategy. Their presence will guarantee a wide benchmark between evolutionary, swarm-based and physics-based optimization strategies.

3.3.3 FbOA Optimization Proposed

FbOA is an optimization that is motivated by co-ordinated directional motion and force-based motion that is dynamic in both exploration and exploitation. The algorithm spreads agents in the space of hyperparameter search and refines their locations step by step on the basis of the quality of the solutions they propose. This framework is suitable to the nonlinear and multimodal topologies of NODE hyperparameter tuning. The process starts with the creation of a population of candidate vectors. Each of the vectors is a codification of a complete set of hyperparameters of the NODE and is measured by the losses of the forecasts. Agents then revise their positions based on the application of force-based movement rules that exhibit exploratory or exploitative behavior based on the stage of iteration. The model of exploration is described by velocity update that helps the agents traverse the search space in large areas. One simplified version of the update to the exploration in the form of the algorithm described is as below.

$$V_n = F_{\max}\left(b_x\, a_i\left(F_{\text{ext}} - F_{\min}\right) + r\, b_y\, a_j\left(F_{\text{best}} - F_{\min}\right)\right)\cos\left(\frac{\pi}{t}\right),$$

where V_n is the exploratory velocity at iteration n, $F_{\max}$ and $F_{\min}$ define upper and lower force bounds, F_{ext} reflects external influence on movement, F_{best} is the force associated with the best solution, a_i, a_j are acceleration terms, b_x, b_y are directional coefficients, r introduces randomness, and t is the current iteration.

This formulation allows agents to move outward along one of the directional vectors, promoting wide coverage of the hyperparameter domain during early iterations.

The best force term is updated dynamically to guide the search toward increasingly promising regions:

$$F_{\text{best}} = \frac{1}{K}\sum_{n=0}^{K}\left(\frac{F_{\max}^{n^2}}{(2n+1)^2}\right),$$

where K grows over successive iterations to gradually shift the balance from exploration toward more focused refinement.

Exploitation is modeled through a position update that emphasizes convergence around strong solutions. A general form of the exploitation update is

$$S(t+1) = F_i + z_3\, S(t) + K\,\sin\left(\frac{\pi}{t}\right),$$

where $S(t)$ is the current position, F_i represents the influence of a local force acting on the agent, z_3 controls the contribution of the current solution, and the sinusoidal term modulates the refinement intensity over time.

To further increase diversity and prevent stagnation, a mutation mechanism perturbs the solution during later iterations:

$$S(t) = K\, a_q \left(\frac{2n+1}{x}\right) + K\, \cos\left(\frac{\pi}{t}\right),$$

where a_q scales the mutation effect and x normalizes the perturbation. All these mechanisms (directional movement, force-based exploration, exploitative refinement, and mutation) allow FbOA to thoroughly explore the hyperparameter space and find parameter configurations that provide higher forecasting performance. The directional structure and dynamic force terms enable the algorithm to automatically switch to local convergence as optimization occurs and the global search.

4. Results

4.1 Baseline Model Evaluation

To get a baseline performance, each deep learning model was initially trained with its default or those hyperparameter settings common to all models before metaheuristic optimization was applied. This gives an object of comparison through which it is possible to gauge the improvements through optimization. The baseline analysis contains a complete range of statistical and forecasting measures: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Pearson correlation coefficient (r), coefficient of determination (R2), Relative RMSE (RRMSE), NashSutcliffe Efficiency (NSE) and the Willmott Index of Agreement (WI). A combination of these metrics captures different dimensions of predictive performance such as accuracy, dispersion, bias and general agreement between the predicted and observed value. Tables 2 present the baseline of NODE, Seq2Seq, and ConvLSTM. Of the three architectures, the unoptimized NODE model has the best performance in almost all the evaluation measures. It has the lowest MSE, RMSE, and MAE indicating a smaller average and squared errors to observed values. NODE also presents the largest correlation coefficient and R 2 which means that it explains a large percentage of the variation in trends in cement-production over time. Moreover, both NSE and WI values prove that NODE gives more credible agreement between the simulated and observed output than Seq2Seq and ConvLSTM. Seq2Seq is also mediocre with a higher performance compared to ConvLSTM and lower compared to NODE. It has a greater MAE and RMSE, which means that it does not predict the consequences of a multi-step forecasting as well. Although ConvLSTM can capture the spatiotemporal dependence, it exhibits the most undesirable baseline properties in this context. Its large error values, and smaller values of agreement indicators, imply that it is hard to model the long-term temporal irregularities in the dataset without fitted hyperparameters. Comprehensively, these initial findings demonstrate the need of hyperparameter optimization and support the use of metaheuristic strategies to improve the forecasting ability of all the three model classes.

Table 2: Baseline performance metrics for unoptimized deep learning models

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
NODE	0.00745	0.0863	0.0515	0.0328	0.89	0.891	1.48	0.91	0.905
Seq2Seq	0.0103	0.1015	0.0738	0.0580	0.887	0.876	1.88	0.895	0.898
ConvLSTM	0.0112	0.1058	0.0830	0.0642	0.871	0.868	2.85	0.873	0.865

Comparison of performance using various metrics of forecasting are critical to the study of model behavior, source of error, and how predictive systems are robust. Overviews are an intuitive way of evaluating such differences in a variety of criteria at a time. Figure 3 is a bar chart of the key performance indicators namely MSE, RMSE, MAE, MBE, correlation values, and agreement indexes with error bars to show variability or uncertainty in the estimate. The visualization is used to emphasize the relative strengths and weaknesses of the model in various statistical dimensions providing a holistic view of the accuracy, bias, consistency and predictive reliability.

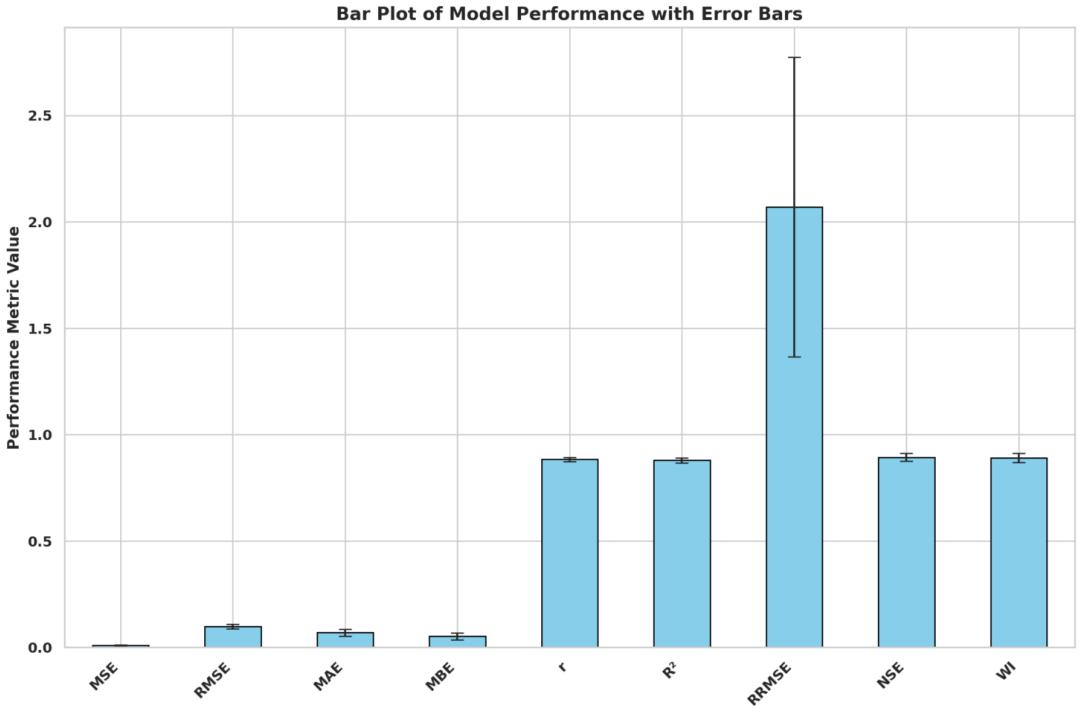

Figure 3: Bar plot of model performance metrics with associated error bars.

Control of the distributional characteristic of forecasting measures gives a better understanding of the consistency, variability, and reliability of model performance. Although single-value measures like MSE, RMSE or R^2 summarize accuracy, they can hide resampling variations, re-training effects or sensitivity to starting conditions. Box plots provide an all-encompassing view of these distributional properties by displaying and

highlighting both medians, quartiles, and possible outliers yet also comparing the dispersion of different measures of evaluation. Figure 4 shows box plots of nine important performance measures, each with marks of the mean and standard deviation, as a result of which the immediate evaluation of the central tendency and dispersion is possible. Such representation allows developing a more detailed vision of the consistency of the model in metrics, as well as indicating those indicators that have a greater uncertainty or variability in training. Distributional analysis of forecasting measures gives further understanding of how stable, variable and general reliable the model is performing. Single-value measures of accuracy like MSE, RMSE, or $R2$ can be used to summarize the accuracy, however, they can hide resampling or repeated train-run variability or sensitivity to initialization. Box plots provide a detailed graphical representation of these distributional aspects showing medians, quartiles, and possible outliers and at the same time comparing the dispersion of the different measures of evaluation. Figure 4 shows box plots of nine main performance metrics with mean and standard deviation markers, which allows one to check central tendency and dispersion directly. It is through this representation that a sensitive interpretation of the consistency of the model between metrics can be achieved and an insight into indicators that are more uncertain or variable during training is gained.

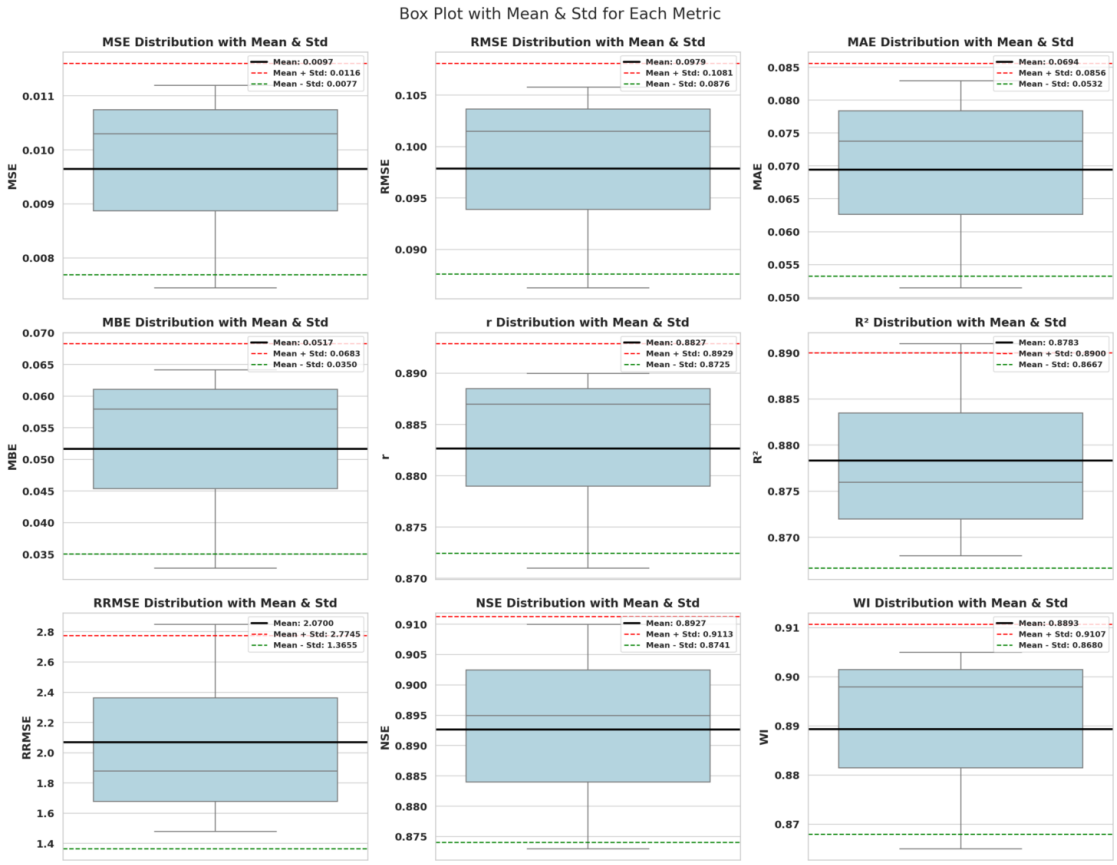

Figure 4: Box plots displaying the distribution, mean, and standard deviation for each performance metric.

The investigation of the joint relations between forecasting measures gives essential information on the interactions of the various measures of model performance, their correlations or deviations under different circumstances. Pairwise visualizations, using kernel density estimation (KDE) not just to show linear relationships, but also to identify underlying structural patterns including clustering patterns, distributional overlap and nonlinear relationships. Figure 5 shows a complete pairplot of all model performance metrics with bivariate contour densities and univariate KDE curves on the diagonal. This combined form has allowed systematic analysis of the correlations between error-based and agreement-based measures to provide a subtle insight into the consistency, trade-offs and internal consistency of model performance.

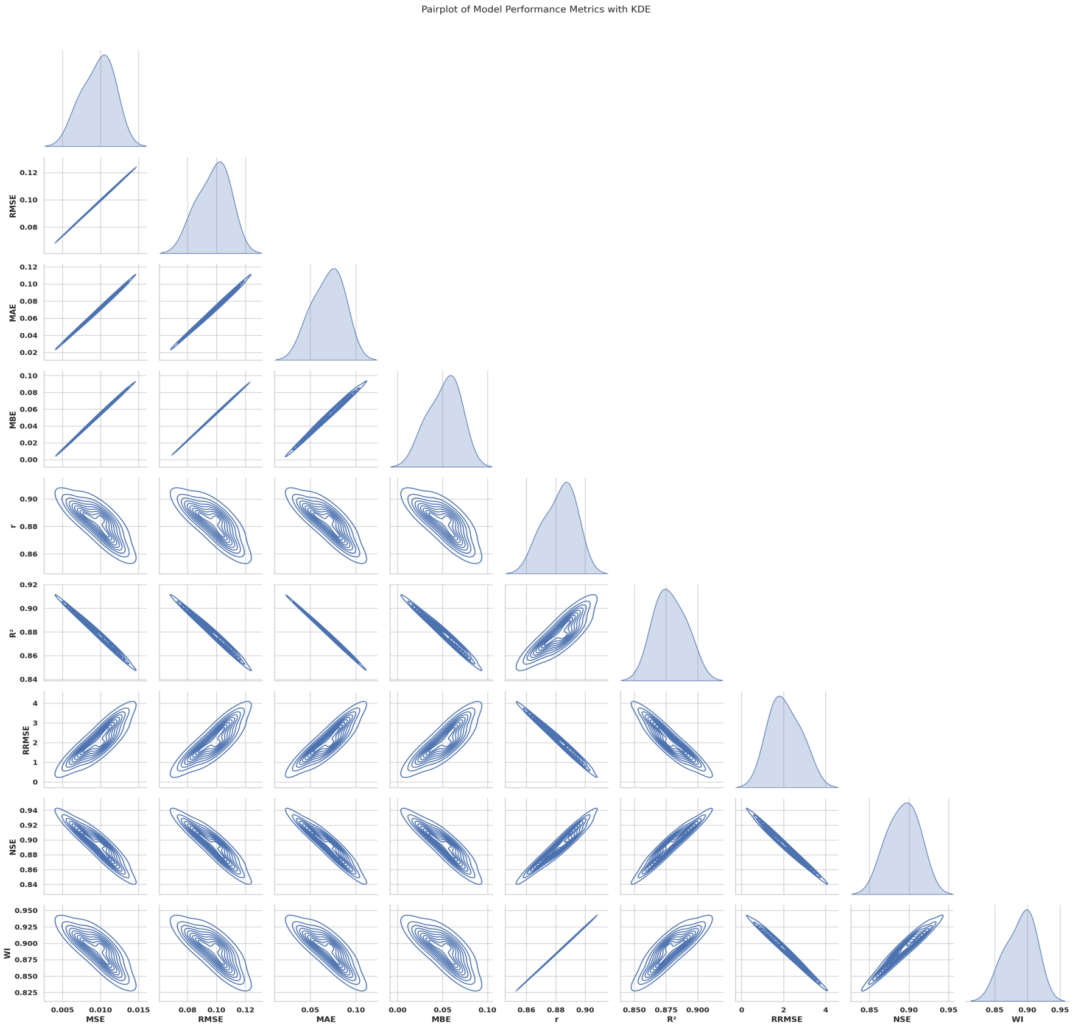

Figure 5: Pairplot with KDE for all performance metrics, illustrating joint distributions, correlations, and univariate density structures.

4.2 Hyperparameter Optimization Results

Once the baseline performance of the unoptimized models had been determined, the metaheuristic algorithms of Section 3.3 were used to tune the NODE architecture on hyperparametric. This step was intended to measure the degree to which automated search through the hyperparameter space can be used to improve the forecasting performance and to compare the relative performance of optimizers. The analysis is once again based on MSE, RMSE, MAE, MBE, r, R2, RRMSE, NSE, and WI which makes them directly comparable to the results of the baseline. Table 3 provides a summary of the performance of NODE as optimized by FbOA, PSO, MVO, WAO and GA. FbOA-guided optimization is the most significant in all metrics. The FbOA + NODE combination has the minimum values of MSE, RMSE, MAE and MBE values when compared to any other optimizer, implying that the random and systematic error elements are reduced significantly. The correlation coefficient is 0.977 and the value of R2 is 0.973 which shows that the opti-mized model can be able to account a very high percentage in the movements of the observed series. The large NSE (0.975) and WI (0.98) also indicate that FbOA + NODE provides highly close predictions and observed emissions. PSO + NODE is placed at the second position in the majority of metrics, and then comes MVO + NODE, WAO + NODE, and GA + NODE. Despite the fact that all of the metaheuristic strategies are an improvement over the NODE baseline, their performance varies in terms of the error level and their performance score. As an example, GA + NODE continues to provide significant improvements to unoptimized NODE model but has higher error values and smaller agreement indices than FbOA, PSO and MVO. Such ranking means that the search dynamics and directional processes of FbOA are especially useful to explore the NODE hyperparameter space in the given forecasting problem. Altogether, these findings indicate the fact that metaheuristic optimization can significantly optimize NODE-based forecasting, and FbOA can make the most significant gains in all metrics of the entire portfolio.

Table 3: Performance of NODE after hyperparameter optimization by different metaheuristic algorithms

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
FbOA + NODE	3.95E-07	6.28E-03	3.42E-04	8.10E-05	0.977	0.973	0.085	0.975	0.98
PSO + NODE	5.50E-07	7.42E-03	3.90E-04	1.05E-04	0.967	0.964	0.11	0.967	0.975
MVO + NODE	7.10E-07	8.43E-03	4.20E-04	1.20E-04	0.959	0.956	0.16	0.959	0.97
WAO + NODE	8.50E-07	9.21E-03	4.55E-04	1.32E-04	0.952	0.949	0.20	0.952	0.965
GA + NODE	1.03E-06	1.02E-02	4.75E-04	1.50E-04	0.946	0.943	0.25	0.944	0.96

Evaluating the distributional characteristics of performance measures is critical in testing model postulations as well as in providing the statistical credibility of the forecast outcomes. QuantileQuantile (QQ) plots can be used to give a good diagnostic of whether the distribution of each measure is actually a theoretical normal distribution or not, to indicate deviations (skewness, heavy-tailed, systematic bias). Figure 6 demonstrates the QQ plot of all the metrics that are being evaluated, and it is possible to visually compare the order of the observed values with their theoretical quantile. Such representation provides a traditionally brief but rigorous way of assessing normality assumptions and facilitates the further inferential or comparative analysis of the modeling workflow.

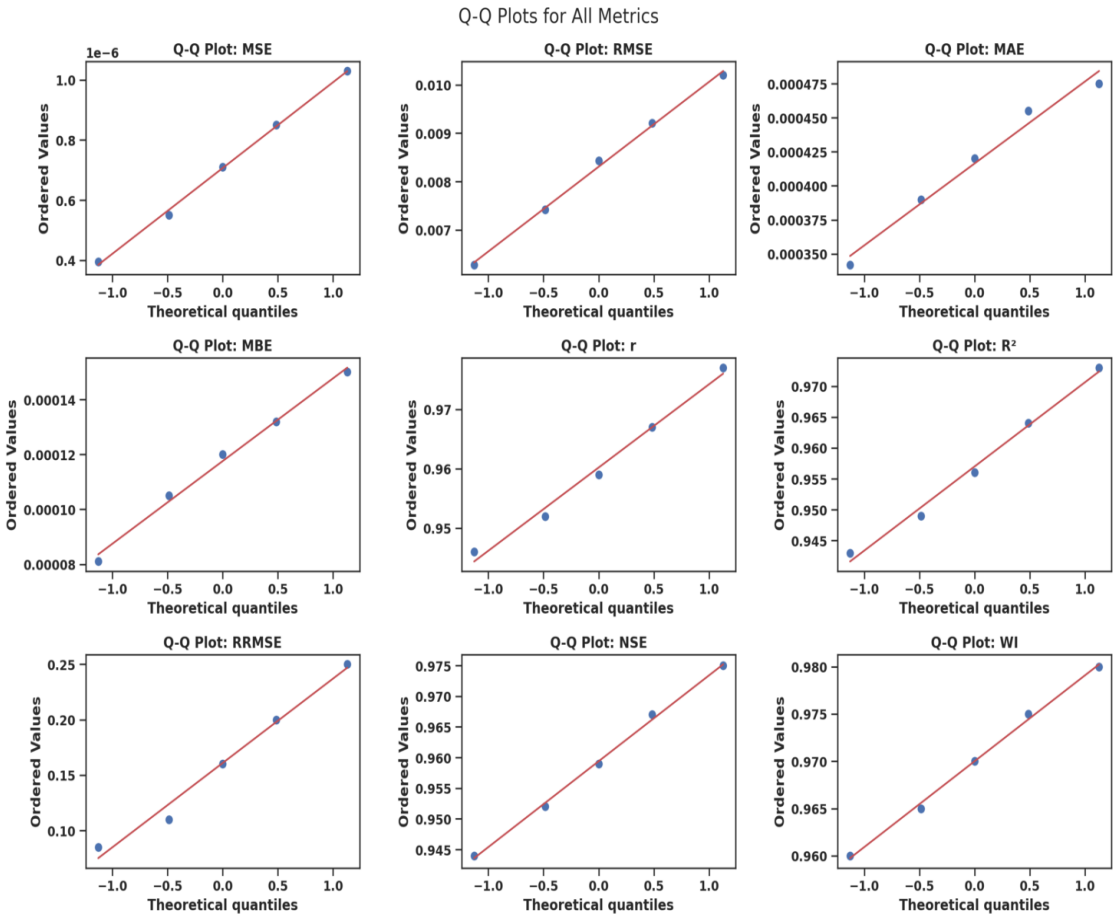

Figure 6: Q–Q plots for all performance metrics comparing empirical ordered values with theoretical normal quantiles.

Here, the complete distributional character of forecasting measures can only be grasped using visuals that outline not just central tendency and spread, but also form, density, and point-wise variance of the information. Mixed plots with swarm, violin, and boxplot demonstrations are useful as they provide a comprehensive diagnostic picture and at the same time show individual observations, the smoothness of distribution, and strong statistical summaries. Such an integrated visualization of all the performance metrics is provided in Figure 7 where dispersion, symmetry, clustering patterns, and possible outliers can be easily compared. Such coherence to representation makes it more interpretable by showing finer details to the metric distributions that might not be evident with one technique of visualization.

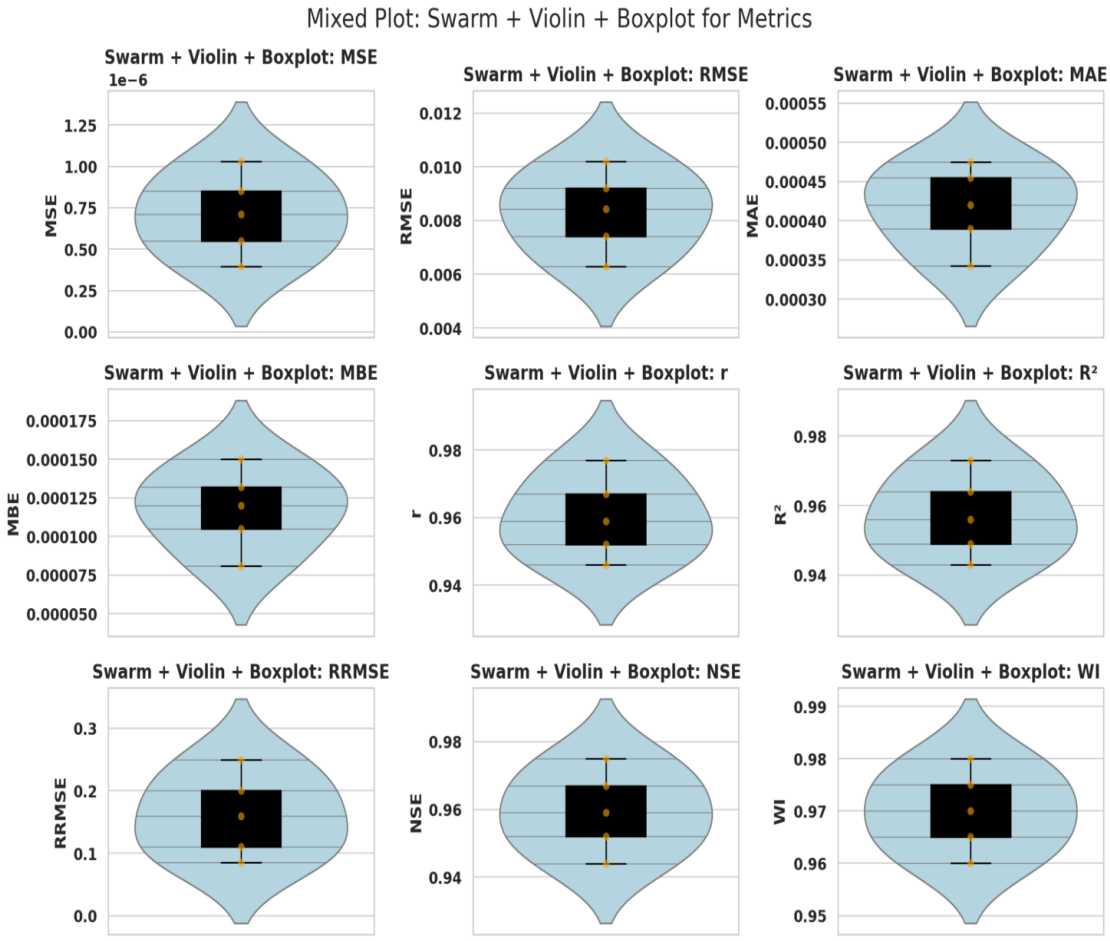

Figure 7: Mixed swarm–violin–boxplot visualizations for all performance metrics, showing distribution shape, point-level variability, and statistical summaries.

The need to compare several forecasting models with a wide range of evaluation measures would be helped by a visualization method that has the ability to summarize the performance of multidimensional measurements in a single understandable format. Radar charts offer this structure, visualizing every metric with a radial axis, which allows comparing accuracy, bias, the strength of correlation and the model agreement overall at once. Figure 8 depicts a radar chart of the performance profiles of NODE as optimized using five metaheuristic algorithms. This display presents the comparative performance of each optimizer in the entire metric suite, enabling one to easily determine the performance dominance, patterns of consistency, and trade-offs between error minimization and correlation based metrics.

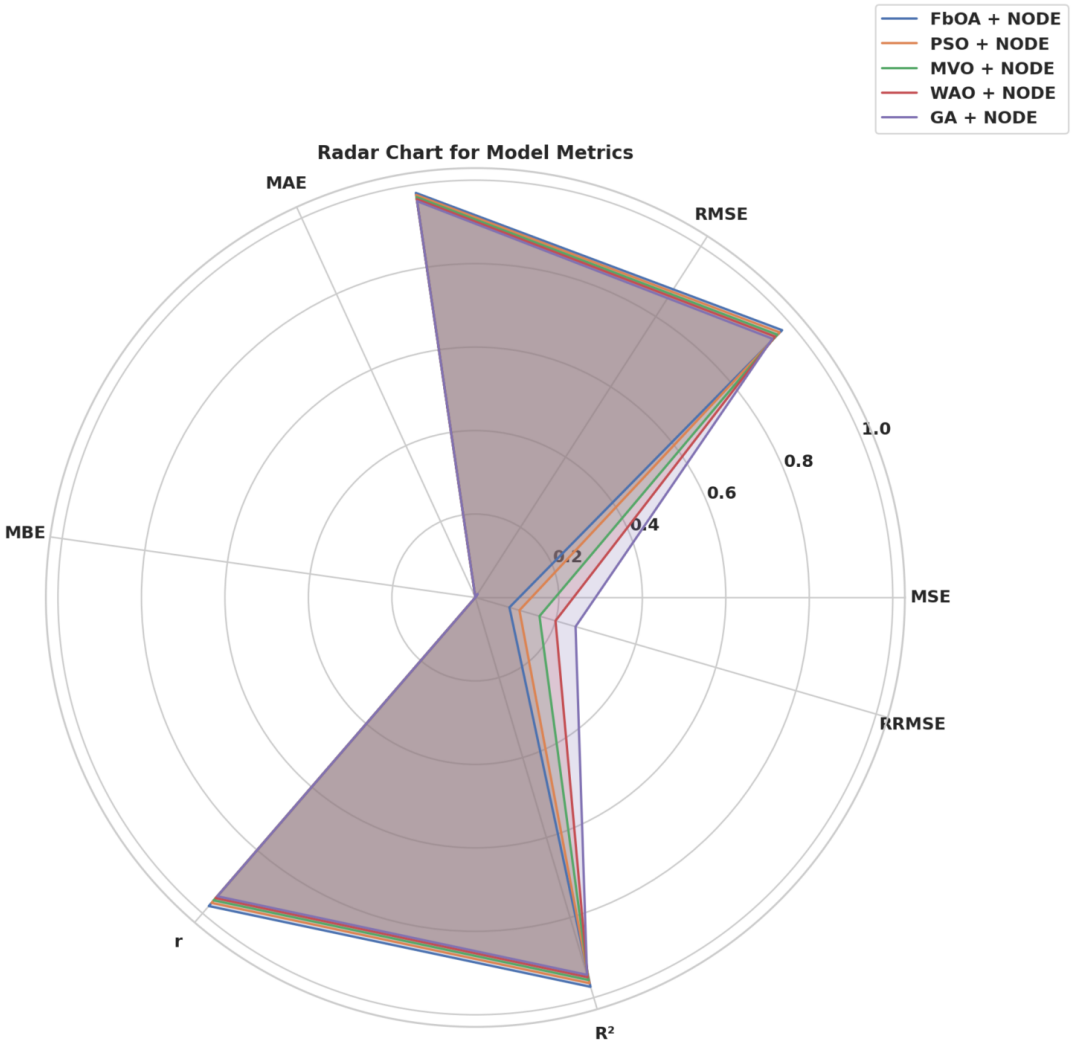

Figure 8: Radar chart illustrating the performance metrics of NODE optimized using five metaheuristic algorithms.

4 Discussion

This research paper shows the significant benefits of combining metaheuristic optimization and deep learning architectures to predict complex industrial time-series like the world emissions of CO2 as a result of cement production. In all the base tests, the performance of NODE was better compared to Seq2Seq and ConvLSTM, which is consistent with the theoretical advantages of continuous-time modeling in the process of representing non-stationary and irregular temporal dynamics. The high baseline performance of NODE

points to its ability to capture the long-term and smooth curves of industrial emission tendencies and retain sensitiveness to variations caused by external sources of variability of regulatory changes or macroeconomic cycles. However, the apparent change in specific metrics when repeated runs are used shows that most architectures of high performance are sensitive to hyperparameter settings, and thus require sophisticated optimization techniques. When the Football Optimization Algorithm was introduced as the main optimizer, significant gains were attained in all the assessment measures, which is better than the tested metaheuristics PSO, MVO, WOA, and GA. Such level of performance improvement by FbOA indicates that the directional search mechanism and force-based update principles FbOA have are especially effective to explore the hyperparameter space that NODE represents is highly nonlinear and multi-modal. The multi-metric decrease in the error values, and the increase in the agreement indices depict that FbOA does not only increase the precision but also increases the model robustness and generalization. The visualization analyses (box plot, Q-Q diagnostics, violin-swarm plot and radar chart) also confirm that FbOA-optimized NODE has a narrower distributional variability and more consistent statistical performance. These features are essential in prediction systems to be used in long run climate and industrial planning where minimizing uncertainty is important as much as accuracy. In addition to model-specific information, a multi-visual analytical framework, used in the current research, gives a deeper picture of performance trends and inter-metric relationships. KDE pairplot analyses indicate that the error-based metrics (MSE, RMSE, MAE) and agreement-based metrics (R2, NSE, WI) are intimately coupled and that progress in one area spreads uniformly in all others. This internal consistency gives confidence in the underlying learning dynamics and reliability of the optimised model. Also, the ACF/PACF and the wavelet decomposition findings put the forecasting task into perspective through the observation of long-term persistence, multi-scale variability, and smooth trend in the global temperature data. Combined with the two diagnostic views, it can be demonstrated that deep learning models when optimized appropriately could address both the global and fine-temporal variations successfully. The more general implication is that metaheuristic-directed NODE models have a high potential of being integrated into pipeline-based real-world decision support that aims at emissions control, scenario evaluation, and strategic decarbonization.

5 Conclusion and Future Work

The study proposed an extensive model for predicting CO2 emissions of the global cement industry, concluding that the combination of Temporal Neural Ordinary Differential Equations (NODE) and the Football Optimization Algorithm (FbOA) yields the best results. Since cement production accounts for about 7 to 8% of anthropogenic CO2 emissions, the need to develop precise and consistent forecasting mechanisms is critical for climate planning, industrial policy, and long-term decarbonization directions. Analysis has started with the evaluation of three models of deep learning, i.e., NODE, Seq2Seq and ConvLSTM in the conditions of baseline. NODE was the best performer with the highest MSE of 0.00745, RMSE of 0.0863, MAE of 0.0515, and good values of NSE = 0.91 and WI = 0.905. These findings demonstrated the benefits of continuous-time models for addressing the heterogeneous time variation in data on world cement production. The research then implemented metaheuristic optimization to improve NODE's performance. PSO, MVO, WOA, and GA were compared with the proposed FbOA-based framework. FbOA was the most optimal algorithm with an MSE of 3.95×10^{-7}, RMSE of 6.28×10^{-3}, MAE of 3.42×10^{-4}, correlation coefficient of 0.977, and an R2 = 0.973. These values are much better than baseline NODE and NODE optimized by other metaheuristics. The directional movement approach to fbOA and the rules for updating its forces demonstrated high efficiency in navigating the hyperparameter space of NODE, increasing convergence and

minimizing the algorithm's error across all evaluation measures. Altogether, the suggested FbOA-NODE model is a robust, evidence-based method for emissions prediction that offers both methodological and practical utility. It provides high predictive precision, good generalization, and clear optimization behaviour, hence it can be applied to long-term industrial and environmental decisions. The framework can be extended in several ways in the future. An avenue worth pursuing is the development of a model for multi-sector or multi-gas forecasting, enabling a springboard for an integrated analysis of industrial emissions not only of the cement sector but also of other sectors. Further prospects lie in using spatial structure in forecasting models to capture both cross-country interactions and geographically clustered patterns of emissions, as well as regional patterns. The other option, the development of uncertainty quantification methods, including probabilistic NODE models or ensemble-based methods, would also be an essential avenue, enabling confidence intervals and risk-aware planning. The optimization process itself can be extended to multi-objective specifications that balance accuracy, computational cost, interpretability, and model complexity. Forecasts may be adjusted to align with real-world conditions by integrating domain-specific constraints or policy scenarios into the optimization loop. Lastly, real-time or online learning extensions are another significant direction of the future, especially with the development of industrial monitoring systems and the growing availability of continuous data streams. Overall, this paper has already shown that NODE, in combination with FbOA, is an effective predictive instrument that can be used to implement more informed and proactive climate and industrial policy. The deep dynamical modeling, combined with the metaheuristic optimization, provides a solid basis for future development in data-based emissions analysis and decision support.

References

[1] Susan K Onsongo, John Olukuru, and Onesmus Mwabonje. Circular economy in the cement industry: a systematic review of sustainability assessment and justice considerations in local community development. *Circular Economy and Sustainability*, pages 1–21, 2025.

[2] Qiang Su, Ruslan Latypov, Shuyi Chen, Lei Zhu, Lixin Liu, Xiaolu Guo, and Chunxiang Qian. Life cycle assessment and environmental load management in the cement industry. *Systems*, 13(7), 2025.

[3] Alina Bărbulescu and Kamal Hosen. Cement industry pollution and its impact on the environment and population health: A review. *Toxics*, 13(7), 2025.

[4] Oluwafemi Ezekiel Ige and Musasa Kabeya. Decarbonizing the cement industry: Technological, economic, and policy barriers to co2 mitigation adoption. *Clean Technologies*, 7(4), 2025.

[5] Amruta A. Yadav, Sneha G. Hirekhan, Pranita S. Bhandari, Rajesh M. Bhagat, Amit B. Ranit, Sagar Shelare, Manzoore Elahi M. Soudagar, Shubham Sharma, V.K. Bupesh Raja, Abinash Mahapatro, Sarabjit Singh, Abhinav Kumar, and Ehab El Sayed Massoud. Ai-driven sustainable concrete mix design: Hybrid deep q-learning and genetic algorithms-based multi-objective machine learning optimizations for high structural strength, low cost, and low carbon footprints. *Structures*, 82:110443, 2025.

[6] Negin Marandi and Sharareh Shirzad. Sustainable cement and concrete technologies: A review of materials and processes for carbon reduction. *Innovative Infrastructure Solutions*, 10(9):408, 2025.

[7] Mingyue Jing, Haonan Jia, Quansheng Liu, Kai Zhang, Shuzhan Xu, Xiquan Zheng, and Chunlei Wang. Multi-objective optimization design of cement-based materials for low-carbon goals. *Materials Today Communications*, 44:112135, 2025.

[8] Esraa Khalil and Mohamed AbouZeid. A global assessment tool for cement plants improvement measures for the reduction of co2 emissions. *Results in Engineering*, 26:104767, 2025.

[9] Xiao Liu, Li Yang, Jinhong Du, Hao Zhang, Jingnan Hu, Aizhong Chen, and Wei Lv. Carbon and air pollutant emissions forecast of china's cement industry from 2021 to 2035. *Resources, Conservation and Recycling*, 204:107498, 2024.

[10] Spyros Giannelos, Federica Bellizio, Goran Strbac, and Tai Zhang. Machine learning approaches for predictions of co2 emissions in the building sector. *Electric Power Systems Research*, 235:110735, 2024.

[11] Lapyote Prasittisopin. Machine learning (ml) and deep learning (dl) in sustainable concrete construction: review, trend and gap analyses. *Journal of Asian Architecture and Building Engineering*, pages 1–29, 2025.

[12] Amirali Hosseinnia, Mohammadreza Noori Sichani, Babak Enami Alamdari, Pariya Aghelizadeh, and Amirehsan Teimortashlu. Machine learning formulation for predicting concrete carbonation depth: A sustainability analysis and optimal mixture design. *Structures*, 76:109036, 2025.

[13] Lusy Widowati, Mohammad Gousuddin, Mohammad Nizamuddin Inamdar, and Emil R. Kaburuan. Toward indonesia cement decarbonization: A survey on machine learning model. In *2024 International Conference on Orange Technology (ICOT)*, pages 1–7, 2024.

[14] Ayaz Ahmad, Waqas Ahmad, Fahid Aslam, and Panuwat Joyklad. Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. *Case Studies in Construction Materials*, 16:e00840, 2022.

[15] Min Wang, Mingfeng Du, Yue Jia, Cheng Chang, and Shuai Zhou. Carbon emission optimization of ultra-high-performance concrete using machine learning methods. *Materials*, 17(7), 2024.

[16] P.S.M. Thilakarathna, S. Seo, K.S. Kristombu Baduge, H. Lee, P. Mendis, and G. Foliente. Embodied carbon analysis and benchmarking emissions of high and ultra-high strength concrete using machine learning algorithms. *Journal of Cleaner Production*, 262:121281, 2020.

[17] Yakoub Boukhari. Applying artificial intelligence techniques for predicting amount of co2 emissions from calcined cement raw materials. In Brahim Lejdel, Eliseo Clementini, and Louai Alarabi, editors, *Artificial Intelligence and Its Applications*, pages 231–243, Cham, 2022. Springer International Publishing.

[18] Wei Li and Shubin Gao. Prospective on energy related carbon emissions peak integrating optimized intelligent algorithm with dry process technique application for china's cement industry. *Energy*, 165:33–54, 2018.

[19] Mohammadsadegh Shahrokhishahraki, Mohammadhossein Malekpour, Sajjad Mirvalad, and Gloria Faraone. Machine learning predictions for optimal cement content in sustainable concrete constructions. *Journal of Building Engineering*, 82:108160, 2024.

[20] Chao-qiang Wang, An-ping Zuo, and Yan-yan Liu. Machine learning-based carbon emission prediction and influence factor analysis discussion in china cement industry. *Carbon Letters*, pages 1–24, 2025.

[21] Kwaku Boakye, Kevin Fenton, and Steve Simske. Machine learning algorithm to predict co2 using a cement manufacturing historic production variables dataset: A case study at union bridge plant, heidelberg materials, maryland. *Journal of Manufacturing and Materials Processing*, 7(6), 2023.

[22] Yanjie Sun, Chen Zhang, Yuan-Hao Wei, Haoliang Jin, Peiliang Shen, Chi Sun Poon, He Yan, and Xiao-Yong Wei. Machine learning for efficient co2 sequestration in cementitious materials: a data-driven method. *npj Materials Sustainability*, 3(1):9, 2025.

[23] Mohammed Wajahat Khan, Mohammed Haroon, and Nishant Pandya. Co2 emission prediction in the cement industry using machine learning techniques. In *2024 International Conference on Modeling, Simulation Intelligent Computing (MoSICom)*, pages 484–489, 2024.

[24] Md. Habibur Rahman Sobuz, Mahmudur Hossain Khan, Md. Kawsarul Islam Kabbo, Ali Hussain Alhamami, Fahim Shahriyar Aditto, Md. Saziduzzaman Sajib, U. Johnson Alengaram, Walid Mansour, Noor Md. Sadiqul Hasan, Shuvo Dip Datta, and Arafat Alam. Assessment of mechanical properties with machine learning modeling and durability, and microstructural characteristics of a biochar-cement mortar composite. *Construction and Building Materials*, 411:134281, 2024.

[25] Muhammad Usman, Iftikhar Ahmad, Muhammad Ahsan, and Hakan Caliskan. Prediction and optimization of emissions in cement manufacturing plant under uncertainty by using artificial intelligence-based surrogate modeling. *Environment, Development and Sustainability*, pages 1–24, 2024.

[26] Jinrui Zhang, Wenjun Niu, Youzhi Yang, Dongshuai Hou, and Biqin Dong. Machine learning prediction models for compressive strength of calcined sludge-cement composites. *Construction and Building Materials*, 346:128442, 2022.

[27] Rachel Cook, Taihao Han, Alaina Childers, Cambria Ryckman, Kamal Khayat, Hongyan Ma, Jie Huang, and Aditya Kumar. Machine learning for high-fidelity prediction of cement hydration kinetics in blended systems. *Materials Design*, 208:109920, 2021.

[28] Mahmoud Abdellatief, Mohamed Abdellatief, Ezzat Elfadaly, and Hassan Hamouda. Carbon footprint reduction and performance optimization of sustainable free cement concrete with eggshell powder and rice husk ash using machine learning. *International Journal of Sustainable Development and Science*, 7(1):195–213, 2024.

[29] Panagiotis G Asteris, Mohammadreza Koopialipoor, Danial J Armaghani, Evgenios A Kotsonis, and Paulo B Lourenço. Prediction of cement-based mortars compressive strength using machine learning techniques. *Neural Computing and Applications*, 33(19):13089–13121, 2021.

[30] Muhammad Nasir Amin, Hassan Ali Alkadhim, Waqas Ahmad, Kaffayatullah Khan, Hisham Alabduljabbar, and Abdullah Mohamed. Experimental and machine learning approaches to investigate the effect of waste glass powder on the flexural strength of cement mortar. *PloS one*, 18(1):e0280761, 2023.