Tanzania Journal of Engineering and Technology 2025, **44**(2):51-67
OPEN ACCESS articles distributed under Creative Commons
Attribution Licence [CC BY-ND]
Websites: https://ajol.org/tjet; https://tjet.udsm.ac.tz

ISSN 1821-536X (**print**); ISSN 2619-8789 (**electronic**)
https://doi.org/10.52339/tjet.v44i2.1266

Special Issue – 8th International Conference on Mechanical and Industrial Engineering, October 24 – 25, 2024 at The Nelson Mandela African Institute of Science and Technology, Arusha - Tanzania

Predicting the Influence of Aggregate Size and Distribution on Cementitious Concrete Properties: A Review

Habert Ayesiga, Mahamudu H. Mtebwa, and Innocent J. Macha†

Department of Mechanical and Industrial Engineering, University of Dar es Salaam, P.O Box 35131, Dar es Salaam, Tanzania.

†Corresponding Author: imacha@udsm.ac.tz; ORCID: 0000-0002-7517-4537

ABSTRACT

Cement concrete has been in use for centuries as one of the primary construction materials. Its demand in the construction industry is expected to continue for several centuries before the full development of alternative products. However, one of the main areas of research interest is understanding how its constituents can be tailored to make its properties predictable to reduce risks associated with structural failures, reconstruction and reduced durability. These hindrances associated with cementitious concrete result from several attributes, including constituent material characteristics, mixing ratios and workmanship. Understanding the predictability of cement concrete properties requires computer ***modelling*** *tools to provide reliable information for the mix design, construction, management and operation of cement concrete, and cement concrete structures. This paper reviews* ***progresses in*** *machine learning models for predicting cement concrete properties. Several algorithms have been reviewed, highlighting their applications, knowledge gaps and suggestions for future research. The paper provides a basis for selecting appropriate algorithms for predicting different concrete properties.*

ARTICLE INFO

1st Submitted: **Apr. 23, 2024**

Presented: **Oct. 25, 2024**

Revised: **Nov. 26, 2024**

Accepted: **Jan, 30, 2025**

Published: **June, 2025**

Keywords: *Cement concrete, Aggregate size and distribution, Machine learning, Supervised learning, Modeling*

INTRODUCTION

Concrete is one of the earliest materials ever used by human beings. The earliest use of concrete material dates back to 3000 BC, when the pyramids were constructed (Hasan, 2020; Brueckner and Lambert, 2017). Second only to water, concrete is the world's most consumed material and one of the most used construction materials (Vargas *et al.*, 2024). Its properties, such as high compressive strength, excellent durability, fire and water resistance, globally available materials, ease of processing, handling, placing, and low cost, make it one of the most attractive construction materials (Wangler *et al.*, 2019). Concrete mainly comprises cement, water, and aggregates mixed in proportions per desired mix design for a particular application (Bunyamin *et al.*, 2022). Of these, the aggregates (fine and coarse aggregates) make up the most significant proportion at 60% to 75% of the volume and 70% to 85% by mass, whereas cement binds the ingredients together and contributes to the overall strength of concrete (Adesina, 2018; Pawar *et al.*, 2016). These constituents affect its fresh properties, such as workability and pump-ability, and hardened state properties, such as compressive strength, water and fire resistance, and overall durability (Shah *et al.*, 2021).

On the other hand, the increased failure tendencies in cement concrete structures have triggered many scholars and stakeholders to find solutions to improve cementitious concrete material's efficiency, durability, reliability, and predictability (Kirthika and Singh, 2020). It is, therefore, essential to ensure the use of quality building materials if strong, durable, and cost-effective structures are to be established. This necessitates the use of the right type, quality, and size of aggregates in concrete mixes through careful selection of constituent materials to maintain set standards and the integrity of the structures (Savitha, 2012). According to Jiao *et al.* (2017), the quality of concrete depends on the characteristics and quantity of each constituent material (cement, water, aggregates, and any other additives) used in the concrete mixture. On the other hand, aggregates' quality depends on aggregates' size, shape, formation, and gradation. For practitioners to produce high-strength concrete, it is important to ensure the use of standard aggregates, quality of cement, and proper mix design. Mix design solely depends on aggregates' proper size, shape, and quality, whereas the aggregate's size, shape, and proper gradation dictates their quality.

However, in many areas around the globe, evidence shows that aggregates are sourced from different places with differences in the rocks from which the aggregates are formed, mined, and crushed (machine or hand-crushed), resulting in aggregates with diverse characteristics and unpredictable concrete properties. This makes it difficult to generalise concrete mix designs to achieve desired and uniform concrete for materials in a given locality, which contributes to continued structural failures. These results in erroneous decisions that are costly due to rework and life-threatening due to structural collapses (Okoye *et al.*, 2023).

To overcome such hindrances, most researchers and practitioners have carried out research based on the physical tests performed on specimens to test cement concrete properties at certain ages, such as compressive strength, workability, durability, etc., in an attempt to predict the behaviour of cement concrete to supplement the known mix designs and mitigate concrete failure (Kirthika and Singh, 2020). These methods are, however, labour-intensive, time-consuming, destructive to the specimen, erroneous and costly in the long run, making them unsustainable (John *et al.*, 2019). In recent years, with the advent of modern machine learning (ML) modeling techniques, several predictive models have been developed based on ML algorithms such as Artificial neural networks, decision trees, random forests, and support vector machine (SVM) to provide alternatives to traditional testing (Moein *et al.*, 2023).

Using computer modelling tools may enhance the understanding of concrete properties faster and enable the handling of nonlinear data to provide information for rational decisions. A good knowledge of cement concrete ingredients and properties is important for its improvement. This paper reviews the success of ML algorithms in predicting cement concrete properties based on different input parameters. It also seeks to identify the research gaps and highlights the future direction of ML in the design, construction and management of concrete and concrete structures.

Machine Learning

Over the years, several models have been proposed for different engineering operations like masonry work, ready-mix concreting, and infrastructure design to study and predict labour, materials, machine performance, etc. The development of analytical and numerical models in predicting the properties of concrete using ML techniques is, therefore, a resource,

time, and manpower-saving technique (Nasir *et al.*, 2020).
With the increasing number of researchers utilising different techniques to anticipate and evaluate various properties of cement concrete. Machine learning based methods are increasingly gaining popularity given the current trends in artificial intelligence (AI) and ML fields (Naderpour *et al.*, 2018). Machine learning aims to create a system with the capacity to learn from problem-specific training data and automate analytical model-building processes, giving it the capacity to solve similar tasks systematically and logically solve similar tasks (Janiesch *et al.*, 2021). A typical ML study (Figure 1) consists of six steps, i.e., problem definition, data collection, data pre-processing (or data cleaning), model development, model evaluation, and model deployment (Li *et al.*, 2022).

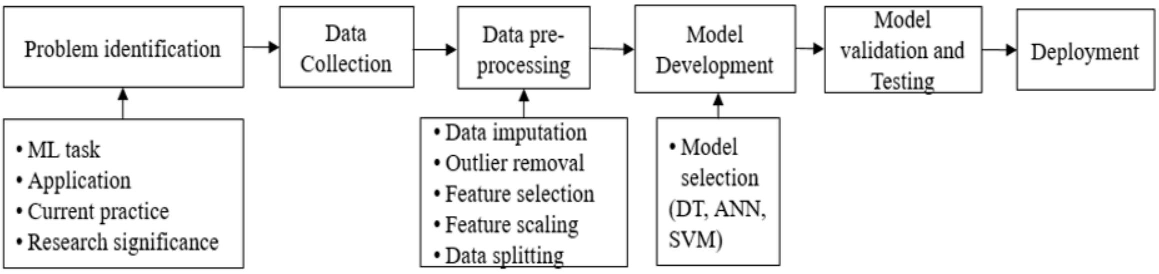

Figure 1: Typical machine learning workflow.

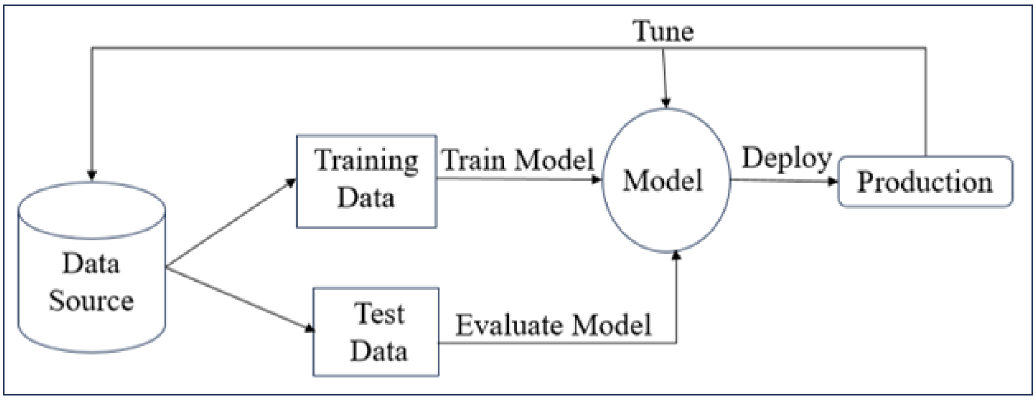

Figure 2: Supervised learning workflow (Mahesh, 2020).

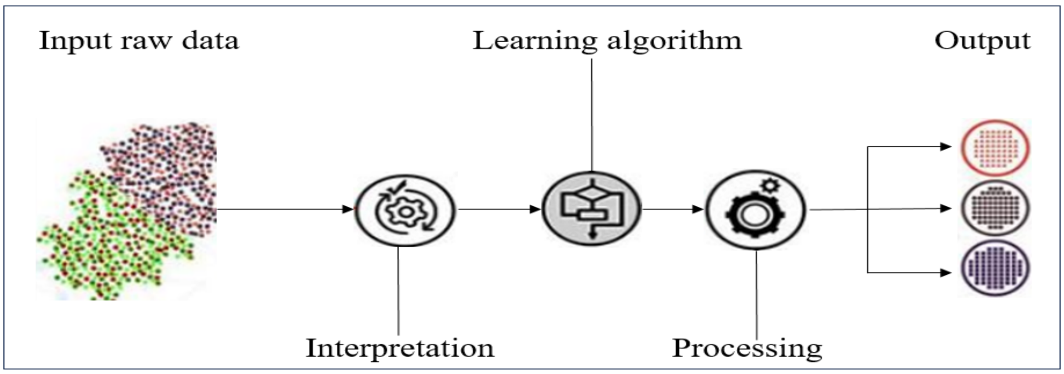

Figure 3: Unsupervised learning workflow (Mahesh, 2020).

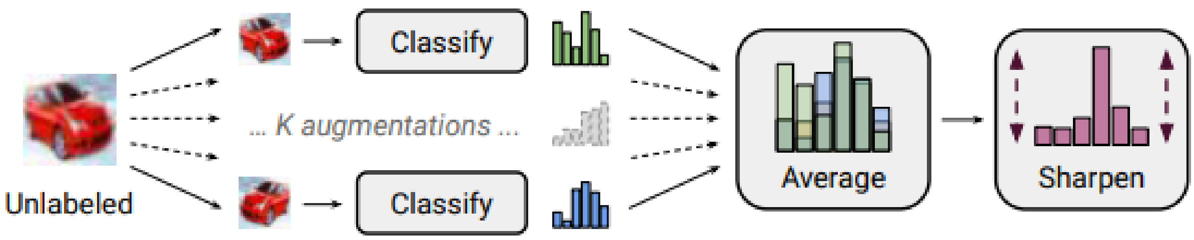

Figure 4: Mix match, an example of semi-supervised machine learning Berthelot *et al.*, 2019).

Model development uses several techniques depending on their learning capabilities, the nature of the data, and the target outcome to build effective models to solve a given task or problem. In concrete technology, ML has been utilised in predicting concrete properties like compressive strength, workability, and durability using several algorithms (Hu *et al.*, 2021). This paper outlines the ML techniques and algorithms utilised in concrete properties predictions.

Types of Machine Learning Techniques

According to Sarker (2021), four major ML categories exist; supervised, unsupervised, semi-supervised, and reinforcement learning. Each ML technique's mode of work and application is briefly discussed.

Supervised Machine Learning

Supervised learning is a ML technique that learns the relationship between input and output data. Data fed into the model are typically labelled as input and output data, enabling the model to learn from a sample of input-output pairs to map an input onto output data (Sarker, 2021; Mahesh, 2020). The input data is usually divided into training and test datasets to produce a prediction or classification. Supervised learning usually follows a systematic workflow (Figure 2) and is applicable in classification and regression tasks to fit the data (Janiesch *et al.*, 2021).

Supervised learning algorithms are mainly divided into two categories: regression and classification. The main regression algorithms include neural networks, decision trees, ensemble methods, and nonlinear regression (Mahesh, 2020). On the other hand, classification algorithms include the discriminate analysis, nearest neighbour, neural networks, decision trees, Naïve Bayes, and Support vector machine (Moein *et al.*, 2023). These have been used in models to predict cement concrete properties like compressive strength (Ahmad *et al.*, 2021; Varma *et al.*, 2023), compressive and tensile strength (Shang *et al.*, 2022) and high-temperature creep in concrete (Bouras and Li, 2023), to mention a few. It is the most commonly used method for most shallow machine learning applications.

Unsupervised Machine Learning

Instead of supervised learning, unsupervised learning is presented with unlabelled datasets, and devices are left to discover and present suitable data structures without human interference. They learn limited features in the data, which they apply to recognise new data (Mahesh, 2020) Unlike supervised learning, unsupervised ML follows specific trends (Figure 3). Unsupervised Machine Learning is mainly used for extracting generative features, identifying meaningful trends and structures, groupings in results, and for exploratory purposes with clustering, density estimation, feature learning, dimensionality reduction, finding association rules, anomaly detection, etc., being the most common unsupervised learning tasks (Sarker, 2021).

An example of unsupervised ML includes the K-Means Clustering models; Hierarchical Learning techniques, i.e., deep learning and artificial neural networks; latent variable models i.e., factor analysis, bind signal separation, and hidden Markov model (Usama *et al.*, 2019); data clustering models i.e., Bayesian clustering, partitional clustering, artificial neural networks, gaussian mixture, hidden Markov model, and hierarchical models (Moein *et al*., 2023) In concrete technology, unsupervised ML has been utilised to assess bridge damage indices (Akintunde *et al.,* 2021) and evaluate concrete's response to fire (Çiftçioğlu and Naser, 2022). This is, however, not commonly used because of unpredictable results during training, which sometimes may result in inaccurate and hard-to-interpret results.

Semi-supervised Machine Learning

Semi-supervised learning combines both supervised and unsupervised ML, enabling it to work on labelled and unlabelled data. Therefore, this can be learned both with and without supervision. It aims to improve the supervised and unsupervised ML models (Van Engelen and Hoos, 2020). Its main advantage comes with the use of unlabelled data (e.g. mix match, Figure 4), as opposed to labelled data ML methods. This is mainly because labelled data is expensive to collect and may involve expert knowledge like in medical fields, it is time-consuming and requires a lot of data which is hard to collect (Ouali *et al.*, 2020). Additionally, data may contain private information, making it difficult and expensive to obtain labelled data (Berthelot *et al.*, 2019).

The actual practical application of semi-supervised ML is still limited by empirical evidence. However, it has been applicable in such areas as machine translation, fraud detection, labelling data, and text classification (Sarker, 2021). Van-Engelen and Hoos (2020) also identified several semi-supervised ML methods, generally categorising them as inductive and transudative. The inductive methods include the wrapper methods i.e., self-training, co-training, and boosting methods; unsupervised pre-processing i.e., feature extraction, cluster-then-label, and pre-training; and intrinsically semi-supervised methods i.e., maximum margin, perturbation-based, manifolds and generative models. On the other hand, transudative methods are mainly graph-based methods that involve the model's construction, weighing and inferencing of the model (Ouali *et al.*, 2020). Application of semi supervised ML in concrete includes concrete crack detection (Liu and Yeoh, 2020), analysis of concrete defects (Karaaslan *et al.*, 2021), etc. This is however not commonly used by many researchers in preference to be supervised ML which is easy to use and less expensive.

Reinforcement Machine Learning

Reinforcement learning is a type of ML algorithm aimed at enabling software agents and machines to automate a given context's behavioural evaluation process to improve its efficiency (Sarker, 2021). Given its current trends in solving complex problems and its use in such intractable fields, reinforcement ML represents a step forward to automation based on artificial intelligence with a clear understanding of the visual world. However, reinforcement learning requires agents to deal with long-range time dependencies and transitions in the environment to come up with correct predictions (Arulkumaran *et al.*, 2017). Reinforcement ML has not seen much application in concrete properties predictions and materials science but its applications in other fields like medicine (Coronato *et al.*, 2020), self-driving cars (Elallid *et al.*, 2022; Li *et al.*, 2019; Aradi, 2020), robotics (Singh *et al.*, 2022; Liu *et al.,* 2021) and manufacturing industries in production scheduling (Waschneck *et al.*, 2018; Shiue *et al.*, 2018) is remarkable. This field still needs to be explored in terms of materials and concrete technology.

Types of Machine Learning Algorithms

There are several machine learning methods with, among others, support vector machine (SVM), Naïve Bayes classifier (NBC), decision tree (DT) and artificial neural network (ANN) being the most common in classification and regression-related tasks (Pineda_Jaramillo, 2019). They are characterised by low computational cost, short development cycles, powerful data processing and high prediction abilities, making them today's preferred choices over statistical and experimental models (Moein *et al.*, 2023). These are, however, limited to shallow machine learning associated with supervised ML, which is the focus for use in fields like materials science as opposed to deep learning.

On the other hand, deep machine learning is associated with more accuracy and prediction abilities than shallow machine learning, enabling it to outperform shallow machine learning models and traditional data analysis approaches. It works by combining several processing layers of the input, hidden and output layers, enabling it to learn from the data (Sarker, 2021). Deep learning resulted from the evolutions in ANN, which have led to the emergence of more advanced neural networks classified as Convolutional Neural Networks (CNNs), recurrent neural networks (RNN), deep belief networks (DBN) and deep coding network mainly used in deep machine learning (Wei *et al.*, 2019; Janiesch *et al.,* 2021). These have enabled their application in several fields, such as face recognition, automatic vehicles, self-service supermarkets, intelligent medical treatments and more accurate forecasts, which had been considered impossible in the past (Li *et al.,* 2021). Based on the no-free lunch machine learning theorem, comparing results from several ML algorithms is crucial as no one ML algorithm can outperform all other models for all data types (Zhang *et al.*, 2020; Moradi *et al.,* 2020). Therefore, machine-learning algorithms commonly used in shallow and deep learning are explained here.

Support Vector Machine

The support vector machine (SVM) is a shallow machine learning technique mainly used in data science as a generalised linear classifier for big data domains using linear models. Vapnik first proposed this to the computer science community in the 1990s to solve problems related to classification. However, it was later upgraded to include regression-related problems (Saha *et al.*, 2020). The SVMs (Figure 5) are equipped to classify sample data using linear models based on minimising structural risk and statistical learning theory through nonlinear mapping into high-dimensional feature space.

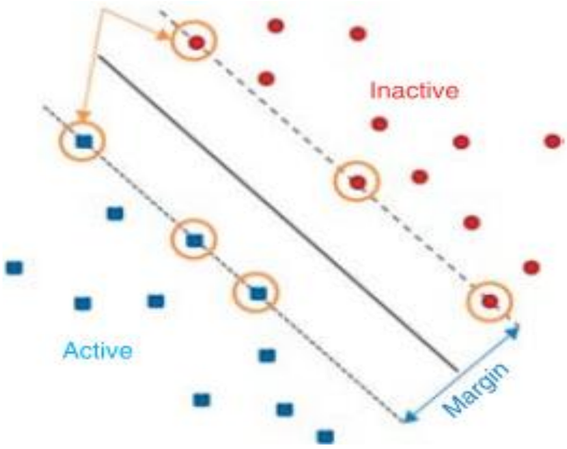

Figure 5: Support vector machine (Wei *et al.*, 2019).

Like any other ML technique, the SVMs also involve training and testing of data instances through which they try to minimise the upper bound of the generalisation error, enabling them to have better generalisability even when dealing with unseen data (Shih *et al.*, 2015).

The SVM's ability to learn from nonlinear decision surfaces, uniquely solve optimisation problems, be theoretically analysed using computational learning theory and excellently perform even with multiple predictors with a limited number of data instances makes it the best choice. This has made them an algorithm of the best option in many ML problems like speech recognition, text categorisation, image recognition, materials science and other critical fields like medicine classifications, enabling it to remain one of the most efficient ML algorithms to date (Montesinos *et al.,* 2022). In concrete properties prediction, SVM has been applicable in modeling concrete properties (Table 1). This, however, does not integrate aggregate sizes and distributions as inputs to the models.

Naïve Bayes Classifier

Naïve Bayesian Classifier (NBC) is one of the classification algorithms in ML and AI that works based on statistics (probability) or the Bayes rule. It assumes the independence of each attribute based on which the final decision is made, which limits the application of this algorithm in real-world situations. Therefore, it necessitates using other methods, such as the confusion matrix, to calculate and test the accuracy of the results (Prasetiyo and Muslim, 2019). According to Hassan *et al.* (2022), it applies to various problems like text classification, sentiment analysis, spam filtering, medical diagnosis, weather prediction, face recognition, and materials science, making it one of the most frequently used algorithms.

Decision Tree

A decision tree (DT) is another prediction model common in many fields, from social science to astrophysics used to solve classification and regression tasks (Sarker, 2021). They are composed of nodes for testing independent attributes (Figure 6). These attributes are normally compared to a constant (i.e., values from control experiments) in a laboratory setup, making it a good algorithm fit for laboratory-based data. It aims at splitting the problems into small subsections where a new individual to be classified is routed down the tree.

This is done according to the values of the independent variables, which are tested in successive nodes by moving down the tree branch corresponding to the attribute values from the root node until the leaf is reached. The new individual is then classified according to the class assigned to the leaf (Montesinos *et al.,* 2022). According to Sarker (2021), the splitting in a decision tree is applicable using several methods, the "Gini" for Gini impurity and "entropy" for the information gain being the most common, as presented in subsequent equations (1 and 2).

Entropy: $$H(x) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \quad (1)$$

Gini $$Gini(E) = 1 - \sum_{i=1}^{c} p_i^2 \quad (2)$$

Artificial Neural Networks

An Artificial Neural Network (ANN) is a biologically inspired model constituting interconnected mathematical equations that mimic biological processes such as learning and memory in a controlled environment to acquire knowledge hidden in historical data for information-processing and computation purposes (Liu *et al.*, 2021). First developed in the 1950s, ANN modeling emerged as a branch of artificial intelligence (AI) with applications in several fields to solve numerous tasks.

Artificial Neural Networks are the most common in ML and AI applications like pattern, speech recognition, optimisation, control, forecasting and data analysis (Doroshenko, 2020) and material science (Duan *et al.*, 2013). This is mainly because they are easy to use, have high accuracy in predicting outcomes, are nonlinear and have high parallelism and generalisation abilities (Liu *et al.*, 2003).

Artificial Neural Networks (ANN) Architecture

Model development with ANN requires a thorough understanding of its structure, as different forms of ANN tend to differ in architecture due to differences in individual layers *(*Albawi *et al.*, 2017). ANNs mainly constitute several small neurons (Figure 7), similar to those of the human brain, which are attached and interconnected as training and recall algorithms to perform tasks (Kollár, 2021). These neurons are attached with connection weights in different layers divided into input, hidden and output layers, each performing a unique task.

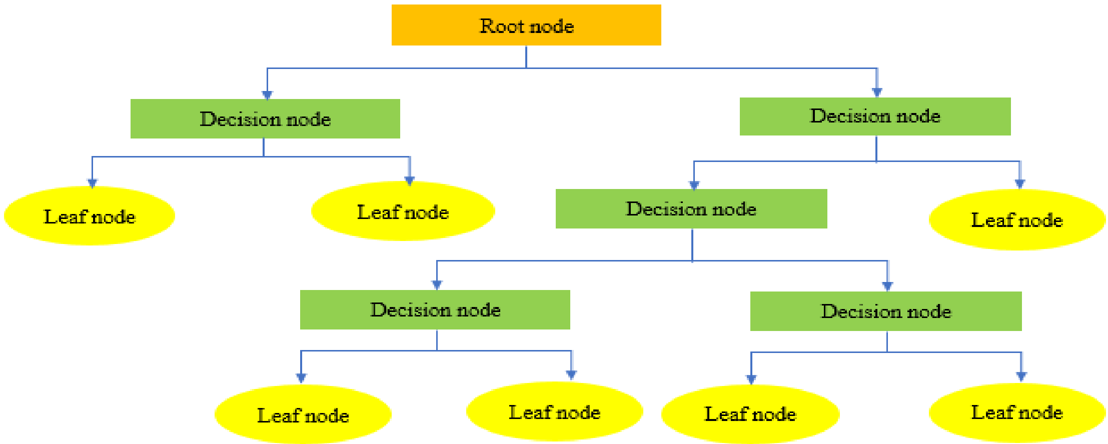

Figure 6: Typical decision tree structure.

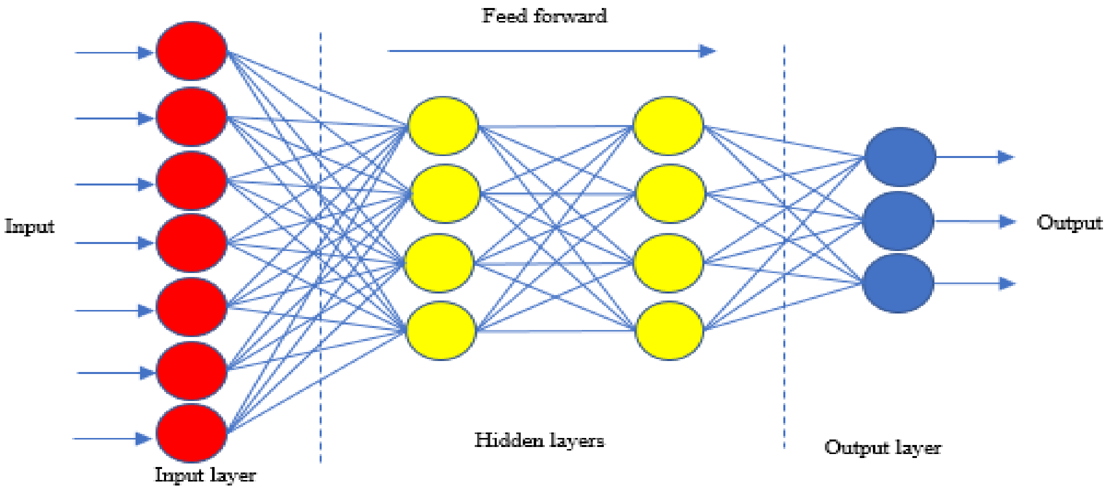

Figure 7: Typical artificial neural network architecture.

Input layer: The input layer receives input data, which are processed in the hidden layer and the result is obtained from the output layer as predictions. outcome or prediction (O'Shea and Nash, 2015).

Filtering Layer: The filtering layers mainly comprise the batch normalisation layer and the Rectified Linear Unit (ReLU) layer. The batch normalisation layer normalises a mini-batch of data across all observations for each channel independently, while the ReLU layer performs a threshold operation to each input element, where any value less than zero is set to zero. It aims to achieve an element activation function like a sigmoid between any two subsequent layers, which helps to improve performance (O'Shea and Nash, 2015).

Processing Layer: The pooling, dropout, flattening and fully connected layers make up the processing layer. The pooling layer does down-sampling along the spatial dimensionality of the given input, which reduces the number of parameters within a given activation (O'Shea and Nash, 2015). The dropout layer helps in model regularisation. The fully connected layer is essential for overfitting prevention (Bezdan and Džakula, 2019). The fully connected layers are the last. They are vital and they multiply the input by a weight matrix and then add a bias vector. The output of a fully connected layer is defined as per equation (3).

$$FC(x) = f(Wx + b) \in \rho^{m} \tag{3}$$

where: $FC(x)$ is the output of the Fully connected layer; f is the activation function; $W \in M_{m,n}(\rho)$ is the weight matrix, $b \in \rho_m$ is the bias vector and $x \in \rho_n$ is the input vector.

Output layer: The output layer mainly comprises the Sigmoid and regression layers. The sigmoid layer applies a sigmoid function to the input such that the output is bounded in the interval (0,1). On the other hand, a regression layer computes the half-mean-squared-error loss for regression tasks. After studying the model structure, model preparation and development shall follow thereafter. The sigmoid layer applies a sigmoid function ($\sigma(x)$) as defined in equation (4).

$$\sigma(x) = \frac{1}{1+e^{-x}},\ x \in \rho \tag{4}$$

Model Performance

The model performance is normally evaluated based on the test and validation datasets necessary for assessing how the model generalises to new data. The validation of the model is executed by utilising the three statistical parameters i.e., Root Mean Square Error (RMSE), which shows a measure of how well the ANN is predicting the correct output, the mean absolute percentage error (MAPE) and the Pearson correlation coefficient, R^2 (Asteris and Mokos, 2019), as presented in equations 5,6,7 herein.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{5}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| \tag{6}$$

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{n}\left(x_i - y_i\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right) \tag{7}$$

where: "n" is the total number of datasets, "x_i" and "y_i" are the predicted and target values, respectively. At the end of the training, testing and validation, the predicted results of the model are compared to the actual results obtained from the field/ to actual data.

Shortcomings Associated with Machine Learning

The major shortcoming of most ML algorithms is overfitting. Over-fitting generally occurs when a model has learned patterns in a specific rather than a generic way. This causes it to perform poorly with the recall data fed into it, even when it performs well with training data. In a class of students, overfitting is related to memorising examples rather than understanding the concepts or general rules. Whenever it occurs, it is generally identified by looking at the performance graphs at the validation stage (Asteris *et al.*, 2016). The ANN model overfitting problem can be overcome by increasing the quality and quantity of data and implementing cross-validation. According to O'Shea and Nash (2015), reducing the complexity of ANNs can overcome overfitting as its likelihood reduces with a reduction in training parameters. In a decision tree, pruning tends to overcome or reduce overfitting tendencies.

The advent of machine learning algorithms has seen significant progress in several fields in solving several tasks, including robotics, self-driving vehicles, text and voice recognition, medicine, weather and materials science. In material science, their recent application in the study of cement concrete and its ingredients as a material is summarised.

Machine Learning Application in Concrete Prediction

The study's purpose was achieved by collecting and reviewing publications, including journal articles, conference proceedings and books. These were accessed through Google Scholar and Research Gate to understand cement concrete and its properties and machine learning algorithms, to identify research gaps. The publications were accessed by relying mainly on selected key terms, including "Cement concrete", "Machine learning", "Machine learning algorithms" and "Modeling". These publications were filtered by first reading their abstracts and then publications between 2000 and 2024 were selected. This is mainly because most of the publications were made in the last twenty-five years. At the end 118 publications were found to be relevant to the study. Thereafter, the articles were studied and machine learning algorithms used in predicting cement concrete properties were summarised with their subsequent input and out variables, highlighting their progresses in concrete property predictions and subsequent gaps (Table 1).

Table 1: Application of machine learning algorithms in concrete properties prediction

Algorithms	Input variables	Output	Ref.
Support vector machine	Water to binder ratio, water content, fine aggregate ratio, fly ash replacement ratio, silica fume replacement ratio, air entraining agent, superplasticiser	Compressive strength	Gupta (2007)
Support vector machine, artificial neural network	Cement/ fly ash dosages, water/ powder, superplasticiser, sand, coarse aggregate contents	Slump flow, T50, T60, V-funnel, Orimet, L-box ratio	Sonebi *et al.* (2016)
Support vector machine	Water/cement ratio, quantity of cement, volume of aggregate, density of aggregate	Compressive strength, young's modulus	Bonifácio *et al.* (2019)
Artificial neural network	Cement, Metakaolin, water, coarse aggregate, fine aggregate, MK specific area, SiO_2, Al_2O_3	Compression strength	Topçu and Sarıdemir (2008)
Decision tree algorithm, Adaboost algorithm	Water, cement, fine aggregate, natural coarse aggregate, recycled concrete aggregates, superplasticisers, water absorption of recycled concrete aggregates, maximum size of recycled concrete aggregates, density of recycled concrete aggregates	Compressive strength, splitting tensile strength	Shang *et al.* (2022)
Backpropagation neural network, the decision tree, random forest, linear regression, ridge regression	Water-cement ratios, thickness of concrete specimens, coarse aggregate fraction volume, environmental to maintenance standard temperature ratio, environmental humidity to relative humidity ratio	Chloride diffusivity	XuanRui (2022)
Adaptive neuro-fuzzy inference system, artificial neural networks, gene expression programming	Chemical composition of fine aggregates, mixing procedures, curing regime, activator content, fine	Compressive strength, slump	Nazar *et al.* (2023)

Algorithms	Input variables	Output	Ref.
	aggregate, coarse aggregate, extra water, activator-to-fine aggregate ratio, molarity of activators		
Genetic programming, artificial neural networks	Fly ash, ground granulated blast furnace slag, silica fume, slump flow, T50, L-box, V-funnel, J-ring, age	Compressive strength, split tensile strength, flexural strength	Awoyera *et al.* (2020)
Quadratic regression models, artificial neural networks	Water/cement ratio, in situ concrete temperature, curing method	Compressive and split tensile strength, pulse velocity, water penetration	Moein *et al.* (2023)
Decision tree, artificial neural network, bagging, gradient boosting	Water, cement, coarse aggregate, fine aggregate, fly ash, superplasticisers, silica fume, nano silica, temperature	Compressive strength of concrete	Ahmad *et al.* (2021)
Gene expression programming, artificial neural network, decision tree	Cement, fly ash, superplasticiser, coarse aggregate, fine aggregate, water, days	Compressive strength of concrete	Song *et al.* (2021)
Neural network, fuzzy-inference-system, random forest, gradient boosting	Cement, blast-furnace-slag, coarse aggregates, fine aggregates, fly ash, water, superplasticiser, curing days	Concrete compressive strength	Elshaarawy *et al.* (2024)
Linear Regression, support vector machine, decision tree, random forest, gradient boosting models	Quantity of: cement, fine, coarse aggregate, water	Comparing the properties of concrete	Jha *et al.* (2024)
Decision tree, multilayer perceptron neural network, support vector machines, random forest	Water, cement, fine aggregate, coarse aggregates, silica-furnace, plastic, superplasticiser, age	Compressive strength, tensile strength	Nafees *et al.* (2022)
Random forest, decision tree, support vector machine	Curing time, water to binder ratio, micro-silica dosage, nano-silica dosage	Compressive strength	Liu *et al.* (2023)
Decision tree, random forest, support vector machine, partial least squares, artificial neural networks, bootstrap aggregation, fuzzy logic models	Water/cement ratio, cement content, compressive strength of cement, fine aggregate, fineness module, chemical admixture, type of aggregate	Compressive strength, slump	Cihan (2019)
Gene expression programming, random forest regression, support vector machine	Sugarcane bagasse ash dosage, fine aggregate, coarse aggregate, water-cement ratio, cement content	Compressive strength	Shah *et al.* (2022)
Regression tree, k-nearest neighbors regressor, multi-layer perceptron neural	Water, cement, fine aggregate, moisture content of fine aggregate, source of	Compressive strength,	Vargas *et al.* (2024)

Algorithms	Input variables	Output	Ref.
network regressor, support vector machine regressor, random forest regressor, gradient boosting regressor, extreme gradient boosting regressor	fine aggregates, coarse aggregate, moisture content of coarse aggregates, source of coarse aggregates, supplementary cementitious materials	flexural strength, slump	
Minimax probability machine regression, emotional neural network, hybrid artificial neural network-particle swarm optimisation	Cement, fly ash, silica fume, lime stone powder, water, superplasticiser, coarse aggregate, fine aggregate	Slump, compressive strength	Kaloop *et al.* (2020)
XGBoost regressor, gradient boosting regressor, Adaboost regressor, random forest regressor, decision tree regressor, k nearest neighbours, linear regressor, multi-layer perceptron, ridge regressor, support vector regressor, lasso regressor	Water-to-cement (W/C) ratio, sand-to-aggregate (S/ a) ratio, coarse aggregate size, superplasticiser, silica fume, hooked steel fibre volume fraction, fiber aspect ratio	Compressive, flexural strengths	Kang *et al.* (2021)
Gradient boosting, extreme gradient boosting, support vector regression, three particle swarm optimisation (PSO): GB_PSO, XGB_PSO, PSO	Cement, water, natural aggregate, recycled concrete aggregate, sand, water absorption rates of natural aggregates	Compressive strength of recycled concrete	Tran *et al.* (2022)
Random forest, k-nearest neighbor, extremely randomised trees, extreme gradient boosting, gradient boosting, light gradient boosting machine, category boosting, generalised additive models: inverse gaussian and poisson	Cement, water, mineral admixture, fine aggregates, coarse aggregates, superplasticisers	Compression strength	de-Prado-Gil *et al.* (2022)

Knowledge Gaps

This study reveals that all these models have not considered aggregate sizes and gradation. A practical application of machine learning in predicting the influence of aggregate size and distribution requires models with "aggregate size and distribution patterns" as inputs to the model.

The standards (e.g., ACI, BS, AASTHO, ASTM, etc) followed in most mixed designs are based on material characteristics in certain locations, leaving out the fact that materials possess unique characteristics per area depending on their chemical composition, methods of extraction and crushing, storage and quality assurance standards onsite.

The existing literature does not emphasise the need for multi-model evaluations in predicting concrete behaviour, as most studies rely on one algorithm to predict concrete properties.

CONCLUSIONS AND RECOMMENDATIONS

The reviewed documents show the application of machine learning in predicting cement concrete properties.

Results further reveal that machine learning can gain momentum for use in predicting the properties of cement concrete.

This study concludes that it is possible to use machine learning to predict cement concrete properties in both fresh and hardened states. Further concludes that machine learning can be used to supplement the known traditional methods, reducing the time and costs spent with the use associated with the traditional laboratory-based tests.

There is a need for continued re-adjustment of cement concrete mixed designs. These should be based on the concrete ingredients used in a given locality. Hence, practitioners need to test the individual characteristics of each concrete constituent to be used in a given locality based on the available samples to close the continued failure tendencies in cement concrete structures.

The present and traditional methods of predicting cement concrete need to be integrated with the current trends in machine learning with cheaper, faster and higher accuracies in predicting concrete properties behaviour as opposed to conventional statistical and empirical models, which are generally inaccurate, costly and time-consuming.

More studies should be focused on understanding the effects of individual concrete ingredient characteristics like aggregate size, aggregate type, distribution, mineralogy, the shape of aggregates, kind of cement, water quality, etc., as these greatly impact cement concrete.

Acknowledgements

This review was conducted as part of the research project titled "Quantification of the influence of aggregate size and distribution on fresh and hardened properties of cementitious concrete". The authors are grateful to God and the support from cohorts, friends and family who have made this possible. Gratitude is significantly extended to the funder of this study, Hon. John Baptist Baguma Bitera. May your effort be rewarded.

Declaration of Conflict of Interest

The authors declare no conflict of interest regarding this paper's publication.

REFERENCES

Adesina, A. (2018). "Concrete Sustainability Issues". *In: Cement and Concrete Science Conference* (pp. 24-26). Coventry University, London, United Kingdom: ISBN 978-1-84600-088-1.

Ahmad, A., Ostrowski, K. A., Maślak, M., Farooq, F., Mehmood, I. and and Nafees, A. (2021). "Comparative Study of Supervised Machine Learning Algorithms for Predicting the Compressive Strength of Concrete at High Temperature". *Materials, 14*(15), 19.

Ahmad, M., Hu, J. L., Ahmad, F., Tang, X. W., Amjad, M., Iqbal, M. J., . . . Farooq, A. (2021). "Supervised Learning Methods for Modeling Concrete Compressive Strength Prediction at High Temperature". *Materials, 14*(8), 19.

Akintunde, E., Azam, S. E., Rageh, A. and and Linzell, D. G. (2021). "Unsupervised Machine Learning for Robust Bridge Damage Detection: Full-Scale Experimental Validation". *Engineering Structures, 249*, 51.

Albawi, S., Mohammed, T. A. and Al-Zawi, S. (2017). "Understanding of a Convolutional Neural Network". *In 2017 International Conference on Engineering and Technology (ICET)* (pp. pp. 1-6). Antalya, Turkey: EEE.

Aradi, S. (2020). "Survey of Deep Reinforcement Learning for Motion Planning of Autonomous Vehicles". *IEEE Transactions on Intelligent Transportation Systems, 23*(2), 740-759.

Arulkumaran, K., Deisenroth, M. P., Brundage, M. and Bharath, A. A. (2017). "Deep Reinforcement Learning: A Brief Survey". *IEEE Signal Processing Magazine, 34*(6), 26-38.

Asteris, P. G., & Mokos, V. G. (2019). "Concrete Compressive Strength using Artificial Neural Networks". *Neural Computing and Applications, 32*(15), 11807–11826.

Asteris, P. G., Kolovos, K. G., Douvika, M. G. and Roinos, K. (2016). "Prediction of Self-Compacting Concrete Strength using Artificial Neural Networks". *European Journal of Environmental and Civil Engineering, 1*, 102-122.

Awoyera, P. O., Kirgiz, M. S., Viloria, A. and Ovallos-Gazabon, D. (2020). "Estimating Strength Properties of Geopolymer Self-Compacting Concrete using Machine Learning Techniques". *Journal of Materials Research and Technology, 9*(4), 9016-9028.

Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A. and Raffel, C. A. (2019). "MixMatch: A Holistic Approach to Semi-Supervised Learning". *Advances in Neural Information Processing Systems, 32*, 11.

Bezdan, T. and Džakula, N. B. (2019). "Convolutional Neural Networks and Architectures". *International Science Conference on Information Technology and Data Related Research* (pp. 445-451). Belgrade, Serbia: Data Science and Digital Broadcasting Systems.

Bonifácio, A. L., Mendes, J. C., Farage, M. C., Barbosa, F. S., Barbosa, C. B. and Beaucour, A.-L. (2019). "Application of Support Vector Machine and Finite Element Method to Predict the Mechanical Properties of Concrete". *Latin American Journal of Solids and Structures, 16*(07), 11.

Bouras, Y. and Li, L. (2023). "Prediction of High-Temperature Creep in Concrete using Supervised Machine Learning Algorithms". *Construction and Building Materials, 400*(2), 1-14.

Brueckner, R. and Lambert, P. (2017). "Unexpected Effects of Historic Concrete Innovations". *International Journal of Heritage Architecture, 1*(4), 549-563.

Bunyamin, B., Kurniasari, F. D., Munirwan, R. P. and Putra Jaya, R. (2022). "Effect of Coral Aggregates of Blended Cement Concrete Subjected to Different Water Immersion Condition". *Advances in Civil Engineering, 2022, 2919167*(1), 10.

Çiftçioğlu, A. Ö. and Naser, M. Z. (2022). "Hiding in Plain Sight: What Can Interpretable Unsupervised Machine Learning and Clustering Analysis Tell Us About the Fire Behavior of Reinforced Concrete Columns?". *In Structures, 40*, 920-935.

Cihan, M. T. (2019). "Prediction of Concrete Compressive Strength and Slump by Machine Learning Methods". *Advances in Civil Engineering, Volume 2019*(1), 11.

Coronato, A., Naeem, M., De Pietro, G. and Paragliola, G. (2020). "Reinforcement Learning for Intelligent Healthcare Applications: A Survey". *Artificial Intelligence in Medicine, 109, 101964*, 71.

de-Prado-Gil, J., Palencia, C., Silva-Monteiro, N. and Martínez-García, R. (2022). "To Predict the Compressive Strength of Self Compacting Concrete With Recycled Aggregates Utilising Ensemble Machine Learning Models". *Case Studies in Construction Materials, 16, e01046*, 17.

Doroshenko, A. (2020). "Applying Artificial Neural Networks in Construction". *In E3S Web of Conferences* (pp. Vol. 143, p. 01029). Moscow: EDP Sciences.

Duan, Z. H., Kou, S. C. and Poon, C. S. (2013). "Prediction of Compressive Strength of Recycled Aggregate Concrete using Artificial Neural Networks". *Construction and Building Materials, 40*, 1200-1206.

Elallid, B. B., Benamar, N., Hafid, A. S., Rachidi, T. and Mrani, N. (2022). "A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving". *Journal of King Saud University-Computer and Information Sciences, 34*(9), 7366-7390.

Elshaarawy, M. K., Alsaadawi, M. M. and Hamed, A. K. (2024). "Machine Learning and Interactive GUI for Concrete Compressive Strength Prediction". *Scientific Reports, 14*(1), 26.

Gupta, S. M. (2007). "Support Vector Machines Based Modelling of Concrete Strength". *Int. J. Intel. Technol, 3*, 12-18.

Hasan, N. (2020). *"Durability and Sustainability of Concrete"*. Springer International Publishing: Cham, Switzerland.

Hu, X., Li, B., Mo, Y. and Alselwi, O. (2021). "Progress in Artificial Antelligence-Based Prediction of Concrete Performance". *Journal of Advanced Concrete Technology, 19*(8), 924-936.

Janiesch, C., Patrick, Z. and Heinrich, K. (2021). "Machine Learning and Deep Learning". *Electronic Markets., 31*, 685–695.

Jiao, P., Borchani, W., Hasni, H. and Lajnef, N. (2017). "A New Solution of Measuring Thermal Response of Prestressed Concrete Bridge Girders for Structural Health Monitoring". *Measurement Science and Technology, 28*(8), 12, 085005.

John, V. M., Quattrone, M., Abrao, P. C. and Cardoso, F. A. (2019). "Rethinking Cement Standards: Opportunities For a Better Future". *Cement and Concrete Research, 124, 105832*(2019), 18.

Kaloop, M. R., Samui, P., Shafeek, M. and Hu, J. W. (2020). "Estimating Slump Flow and Compressive Strength of Self-Compacting Concrete using Emotional Neural Networks". *Applied Sciences, 10, 8543*(23), 17.

Kang, M. C., Yoo, D. Y. and Gupta, R. (2021). "Machine Learning-Based Prediction for Compressive and Flexural Strengths of Steel Fiber-Reinforced Concrete". *Construction and Building Materials, 266, 121117*, 13.

Karaaslan, E., Bagci, U. and Catbas, F. N. (2021). "Attention-Guided Analysis of Infrastructure Damage With Semi-Supervised Deep Learning". *Automation in Construction, 125, 103634*, 1-10.

Kirthika, S. K. and Singh, S. K. (2020). "Durability Studies on Recycled Fine Aggregate Concrete". *Construction and Building Materials, 250*, 14.

Kollár, A. (2021). *"Betting models using AI: A review on ANN, SVM, and Markov Chain"*. Munich: Munich Personal RePEc Archive-MPRA.

Li, D., Zhao, D., Zhang, Q. and Chen, Y. (2019). "Reinforcement Learning and Deep Learning Based Lateral Control for Autonomous Driving [Application Notes]". *IEEE Computational Intelligence Magazine, 14*(2), 83-98.

Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J. (2021). "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects". *IEEE Transactions on Neural Networks and Learning Systems, 33*(12), 21.

Li, Z., Yoon, J., Zhang, R., Rajabipour, F., Srubar III, W. V., Dabo, I. and Radlińska, A. (2022). "Machine Learning in Concrete Science: Applications, Challenges, and Best Practices". *npj Computational Materials, 8*(1), 17.

Liu, G., Zhao, H., Amin, M. N., Zaman, A., Hassan, A. M., Ali, M. and Rehman, M. F. (2023). "Strength Predictive Models of Cementitious Matrix by Hybrid Intrusion of Nano and Micro Silica: Hyper-Tuning with Ensemble Approaches". *Journal of Materials Research and Technology, 26*, 1808-1832.

Liu, J., Savenije, H. H. and Xu, J. (2003). "Forecast of Water Demand in Weinan City in China using WDF-ANN Model". *Physics and Chemistry of the Earth 28, 28*(Issues 4–5), 219–224.

Liu, R., Nageotte, F., Zanne, P., de Mathelin, M. and Dresp-Langley, B. (2021). "Deep Reinforcement Learning for the Control of Robotic Manipulation: A Focussed Mini-Review". *Robotics, 10*(1), 22.

Liu, S., Chang, R., Zuo, J., Webber, J. R., Xiong, F. and Dong, N. (2021). "Application of Artificial Neural Networks in Construction Management: Current Status and Future Directions". *Applied Sciences Appl. Sci. 2021, 11, 9616.*, 19.

Liu, Y. and Yeoh, J. K. (2020, March 9). "Vision-Based Semi-Supervised Learning Method for Concrete Crack Detection". *In Construction Research Congress 2020. Reston, VA: American*

Society of Civil Engineers, pp. pp. 527-536.

Mahesh, B. (2020). "Machine Learning Algorithms- A Review". *International Journal of Science and Research (IJSR), 9*(1), 381-386.

Moein, M. M., Saradar, A., Rahmati, K., Mousavinejad, S. H., Bristow, J., Aramali, V. and Karakouzian, M. (2023). "Predictive Models for Concrete Properties Using Machine Learning and Deep Learning Approaches: A Review". *Journal of Building Engineering, 63, 105444*, 1-41.

Montesinos, L. O., Montesinos, L. A. and Crossa, J. (2022). *"Multivariate Statistical Machine Learning Methods for Genomic Prediction (p. 691)".* Cham, Switzerland: Springer International Publishing.

Moradi, R., Berangi, R. and Minaei, B. (2020). "A Survey of Regularization Strategies for Deep Models". *Artificial Intelligence Review, 53*(6), 3947-3986.

Naderpour, H., Rafiean, A. H. and Fakharian, P. (2018). "Compressive Strength Prediction of Environmentally Friendly Concrete Using Artificial Neural Networks". *Journal of Building Engineering*, *16*, 213-219.

Nafees, A., Khan, S., Javed, M. F., Alrowais, R., Mohamed, A. M., Mohamed, A. and Vatin, N. I. (2022). "Forecasting the Mechanical Properties of Plastic Concrete Employing Experimental Data using Machine Learning Algorithms: DT, MLPNN, SVM, and RF". *Polymers*, *14*(8), 40.

Nasir, M., Gazder, U., Maslehuddin, M., Baghabra Al-Amoudi, O. S. and Syed, I. A. (2020). "Prediction of Properties of Concrete Cured Under Hot Weather using Multivariate Regression and ANN Models". *Arabian Journal for Science and Engineering*, *45*, 4111-4123.

Nazar, S., Yang, J., Amin, M. N., Khan, K., Ashraf, M., Aslam, F., . . . Eldin, S. M. (2023). "Machine Learning Interpretable-Prediction Models to Evaluate the Slump and Strength of Fly Ash-Based Geopolymer". *Journal of Materials Research and Technology*, *24*, 100-124.

Okoye, J. U., Apeh, S. T., Olaye, E. and Osuji, S. O. (2023). "Towards Automation of Building Integrity Tracking: Review of Building Collapse in Nigeria". *NIPES-Journal of Science and Technology Research*, *5*(1), pp. 179-194.

O'Shea, K. and Nash, R. (2015). "An introduction to convolutional neural networks". *ArXiv Preprint ArXiv, 1511*, 1-11.

Ouali, Y., Hudelot, C. and Tami, M. (2020). "An Overview of Deep Semi-Supervised Learning". *ArXiv Preprint ArXiv, 2006.05278.*, 1-43.

Pawar, C., Sharma, P. and Titiksh, A. (2016). "Gradation of Aggregates and its Effects on Properties of Concrete". *International Journal of Trend in Research and Development, Volume 3*(2), PP. 580-584. ISSN: 2394-9333.

Pineda_Jaramillo, J. D. (2019). "A review of Machine Learning (ML) Algorithms Used for Modeling Travel Mode Choice". *Dyna, 86*(211), 32-41.

Prasetiyo, B. and Muslim, M. A. (2019). "Analysis of Building Energy Efficiency Dataset using Naive Bayes Classification Classifier". *Journal of Physics: Conference Series*, (Vol. 1321, No. 3, p. 032016). IOP Publishing.

Saha, P., Debnath, P. and Thomas, P. (2020). "Prediction of Fresh and Hardened Properties of Self-Compacting Concrete using Support Vector Regression Approach". *Neural Computing and Applications, 32*(12), 7995-8010.

Sarker, I. H. (2021). "Machine Learning: Algorithms, Real-World Applications and Research Directions". *SN Computer Science, 2, 160*(3), 1-21.

Savitha, R. (2012). *"Importance of Quality Assurance of Materials for Construction Work".* Building Materials Research and Testing Division, National Building Research Organization, 1-5.

Shah, H. A., Yuan, Q. and Zuo, S. (2021). "Air Entrainment in Fresh Concrete and Its Effects on Hardened Concrete-A Review". *Construction and Building Materials, 274, 121835*, 17.

Shah, M. I., Javed, M. F., Aslam, F. and Alabduljabbar, H. (2022). "Machine

Learning Modeling Integrating Experimental Analysis for Predicting the Properties of Sugarcane Bagasse Ash Concrete". *Construction and Building Materials, 314*(3), 125634.

Shang, M., Li, H., Ahmad, A., Ahmad, W., Ostrowski, K. A., Aslam, F., . . . Majka, T. M. (2022). "Predicting the Mechanical Properties of RCA-Based Concrete Using Supervised Machine Learning Algorithms". *Materials, 15*(2), 27.

Shih, Y. F., Wang, Y. R., Lin, K. L. and Chen, C. W. (2015). "Improving Non-Destructive Concrete Strength Tests Using Support Vector Machines". *Materials, 8*(10), 7169–7178.

Shiue, Y. R., Lee, K. C. and Su, C. T. (2018). "Real-Time Scheduling for a Smart Factory Using a Reinforcement Learning Approach". *Computers and Industrial Engineering, 125*, 604-614.

Singh, B., Kumar, R. and Singh, V. P. (2022). "Reinforcement Learning in Robotic Applications: A Comprehensive Survey". *Artificial Intelligence Review, 55*(2), 945-990.

Sonebi, M., Cevik, A., Grünewald, S. and Walraven, J. (2016). "Modelling the Fresh Properties of Self-Compacting Concrete using Support Vector Machine Approach". *Construction and Building materials, 106*, 55-64.

Song, H., Ahmad, A., Farooq, F., Ostrowski, K. A., Maślak, M., Czarnecki, S. and Aslam, F. (2021). "Predicting the Compressive Strength of Concrete With Fly Ash Admixture Using Machine Learning Algorithms". *Construction and Building Materials, 308, 125634*, 15.

Topçu, İ. B. and Sarıdemir, M. (2008). "Prediction of Rubberized Concrete Properties Using Artificial Neural Network and Fuzzy Logic". *Construction and Building Materials, 22*(4), 532-540.

Tran, V. Q., Dang, V. Q. and Ho, L. S. (2022). "Evaluating Compressive Strength of Concrete Made With Recycled Concrete Aggregates Using Machine Learning Approach". *Construction and Building Materials, 323*(2), 126578.

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.-l. A., Elkhatib, Y., . . . Al-Fuqaha, A. (2019). "Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges". *IEEE Access, 7*, 65579-65615.

Van Engelen, J. E. and Hoos, H. H. (2020). "A survey on Semi-Supervised Learning". *Machine Learning, 109*(2), 373-440.

Vargas, J. F., Oviedo, A. I., Ortega, N. A., Orozco, E., Gómez, A. and Londoño, J. M. (2024). "Machine-Learning-Based Predictive Models for Compressive Strength, Flexural Strength and Slump of Concrete". *Applied Sciences, 14*(11), 24.

Varma, B. V., Prasad, E. V. and Singha, S. (2023). "Study on Predicting Compressive Strength of Concrete using Supervised Machine Learning Techniques". *Asian Journal of Civil Engineering, 24*(7), 2549-2560.

Wangler, T., Roussel, N., Bos, F. P., Salet, T. A. and Flatt, R. J. (2019). "Digital Concrete: A Review". *Cement and Concrete Research, 123, 105780*, 37.

Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A. and Kyek, A. (2018). "Optimisation of Global Production Scheduling With Deep Reinforcement Learning". *Procedia Cirp, 72*, 1264-1269.

Wei, J., Chu, X., Sun, X. Y., Xu, K., Deng, H. X., Chen, J., . . . Lei, M. (2019). "Machine Learning in Materials Science". *InfoMat, 1*(3), 338-358.

XuanRui, Y. (2022). "Developing an Artificial Neural Network Model to Predict the Durability of the RC Beam by Machine Learning Approaches". *Case Studies in Construction Materials, 17*(2022), 14, e01382.

Zhang, J., Huang, Y., Wang, Y. and Ma, G. (2020). "Multi-Objective Optimization of Concrete Mixture Proportions using Machine Learning and Metaheuristic Algorithms". *Construction and Building Materials, 253, 119208*, 15.