Original Study **Open Access**

Akhila Sheshadri, Shriram Marathe*, Anisha P Rodrigues, Martyna Nieświec

Predictive Modelling of Pavement Quality Fibre-Reinforced Alkali-Activated Nano-Concrete Mixes through Artificial Intelligence

https://doi.org/10.2478/sgem-2025-0007
received September 19, 2024; accepted January 22, 2025.

Abstract: Alkali-activated concrete (AAC) has emerged as a viable sustainable alternative for the building of roads and rigid pavements, particularly in India, where infrastructure development is a major priority. With the exponential growth of highway networks, there is a noteworthy emphasis on enhancing mechanical performance of concrete pavements to overcome their inherent brittleness and limited load-carrying capacity. This research examines the incorporation of nano-silica (NS) and nano-alumina (NA) to improve the mechanical properties of pavement quality alkali-activated concrete (PQAC). Additionally, polyvinyl alcohol fibres (PVAF) and Polypropylene Fibre (PPF) were integrated into the concrete mix to address the brittle nature of PQAC and improve the tensile strength of concrete. Given the challenges associated with optimising these material combinations, this research also leverages advanced machine learning models, including Multilinear Regression (MLR), Decision Tree (DT), Random Forest (RF), AdaBoost, Support Vector Regression (SVR), Gradient Boosting (GB), to predict the split tensile strength (STS) of PQAC. A thorough analysis of predicted performance was carried out utilising assessment measures. The findings demonstrate that the AdaBoost model outperforms other models in terms of test performance, achieving an R^2 value of 0.79. This surpasses the R^2 values of MLR (0.61), SVR (0.6), DT (0.79), GB (0.791) and RF (0.796). The remaining four error measures have the lowest values among all models, with MSE = 0.202, RMSE = 0.45, CV=0.714 and MAE = 0.38. The study highlights the superior performance of ensemble models in accurately predicting STS, underscoring their potential as reliable tools for optimising material compositions in pavement applications and thereby supporting or partly replacing laboratory split tension tests, thereby saving time and cost. This research contributes to the broader goal of developing more durable and sustainable concrete mixes for construction projects.

Keywords: Pavement quality alkali-activated concrete; split tensile strength; nano-admixtures; fibre; machine learning.

***Corresponding author: Shriram Marathe,** Department of Civil Engineering, Nitte (Deemed to be University), NMAM Institute ofTechnology (NMAMIT), Karkala Taluk, Udupi District, Karnataka, India-574110; Department of Materials Engineering and Construction Processes, Faculty of Civil Engineering,WroclawUniversity of Science and Technology: PolitechnikaWroclawska 27, 50-370, Wrocław, Poland, E-mail: shriram.marathe@pwr.edu.pl
Akhila Sheshadri, Department of Civil Engineering, Nitte (Deemed to be University), NMAM Institute ofTechnology (NMAMIT), Karkala Taluk, Udupi District, Karnataka, INDIA-574110.
Anisha P Rodrigues, Department of Computer Science and Engineering, Nitte (Deemed to be University), NMAM Institute of Technology (NMAMIT), Karkala Taluk, Udupi District, Karnataka, India-574110
Martyna Nieświec, Department of Materials Engineering and Construction Processes, Faculty of Civil Engineering, Wroclaw University of Science and Technology: PolitechnikaWroclawska 27, 50-370, Wrocław, Poland

List of Abbreviations

AAC	Alkali-Activated Concrete
AdaBoost	Adaptive Boosting
ANN	Artificial Neural Network
BPNN	Backpropagation Neural Networks
CV	Co-efficient of variation
DT	Decision Tree
GB	Gradient Boosting
GBM	Gradient Boosting Machine
GBR	Gradient Boosting Regression
GGBS	Ground Granulated Blast Furnace Slag
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLR	Multiple Linear Regression

Open Access. © 2025 Akhila Sheshadri et al., published by Sciendo. This work is licensed under the Creative Commons Attribution alone 4.0 License.

MSE	Mean Squared Error
NA	Nano-Alumina
NCA	Natural Coarse Aggregate
NS	Nano-Silica
OPC	Ordinary Portland Cement
PPF	Polypropylene Fibre
PQAC	Pavement Quality Alkali-Activated Concrete
PVAF	Polyvinyl Alcohol Fibre
R^2	Coefficient Of Determination
RF	Random Forest
RMSE	Root Mean Squared Error
RSFA	River Sand Fine Aggregate
SCM	Supplementary Cementitious Material
STS	Split Tensile Strength
SVR	Support Vector Regression

1 Introduction

Alkali-activated concrete (AAC) has emerged as a potentially promising alternative for sustainability in highway and rigid pavement construction. The development of highways and rigid pavement has been a significant focus in India and worldwide. The increase of the rigid pavement system in Thailand has led to the initiation of the Rigid Pavement Maintenance System by the Department of Rural Roads [1]. In India, the Golden Quadrilateral project has been a pivotal initiative, significantly impacting the manufacturing activities and contributing to the economic development of the country [2]. The National Highway Development Program in India has also been a major investment in upgrading and rehabilitating the country's major highways to international standards, reflecting the country's commitment to infrastructure development [3] [4][5]. There has been exponential growth in the national highways in India from the years 2014 to 2023 by an increase of about 47.85%, as shown in **Figure 1** [6][7][8]. However, one of the major disadvantages that is encountered in adapting the concrete pavements is its brittle nature and its ability to establish the greater load-carrying capacity[9], [10]. The incorporation of nano-additives in concrete has garnered significant interest due to their ability to enhance mechanical and durability properties [9], [10]. Each type of nano-additive offers specific advantages tailored to various applications. For example, compared to traditional additives like silica fume and fly ash, which primarily serve as supplementary cementitious materials (SCMs) for partial cement replacement [11], nano-silica (NS) and nano-alumina (NA) offer distinct advantages due to their nanoscale dimensions. Their larger surface area and higher reactivity enable effective penetration of micro-pores, resulting in superior matrix densification [12], while nano-TiO_2 is recognised for its self-cleaning and photocatalytic properties, although its mechanical benefits are limited. Nano-zinc oxide promotes early strength development but can lead to rapid setting and shrinkage issues, while nano-clay primarily improves rheological properties with minimal impact on strength. Nano-iron oxide enhances abrasion resistance and thermal stability but is less effective in reducing porosity. Carbon nanotubes and graphene oxides [12], [13] are expensive and require skilled handling [11], [12]. Unlike SCMs, which depend on calcium hydroxide (CH) for pozzolanic activity [11], NS and NA are effective even in AAC systems where CH is absent, making them vital for advancing sustainable, high-performance concrete solutions. Therefore, this research has incorporated NS and NA in order to improve the load-carrying capacity of PQAC.

There are numerous studies which analyses the effects of fibre in normal Ordinary Portland Cement (OPC) concrete. However, the incorporation of fibres in PQAC mixes is a seldom researched area. Therefore, in order to overcome the brittle nature of PQAC, fibres were incorporated in the concrete mix. Polyvinyl alcohol fibres (PVAF)are noted for their high tensile strength and Young's modulus, significantly enhancing the tensile load-carrying capacity of concrete. Their hydrophilic nature promotes strong bonding with the cementitious matrix, improving crack resistance and ductility. Polypropylene fibres (PPF), on the other hand, are effective in reducing early shrinkage cracks due to their high elongation capacity and flexibility, bridging micro-cracks and enhancing overall resistance to larger cracks. Both PVAF and PPF exhibit excellent chemical resistance, making them suitable for aggressive environments where traditional fibres, such as steel, may corrode. They are lightweight, are easy to incorporate and disperse uniformly throughout the concrete matrix, ensuring consistent performance and minimal disruption to workability. In contrast, steel fibres are known for significantly improving tensile strength and crack resistance but are susceptible to corrosion, particularly in harsh environments, which limits their long-term effectiveness. Glass fibres, while non-corrosive and exhibiting high tensile strength, are brittle and prone to alkali attacks unless treated, limiting their effectiveness in high-pH environments. Carbon fibres provide exceptional tensile strength but are often impractical for large-scale applications due to their high cost [14]. Thus, in order to overcome these drawbacks, PVAF and PPF were adapted in the current research work. The optimisation of these

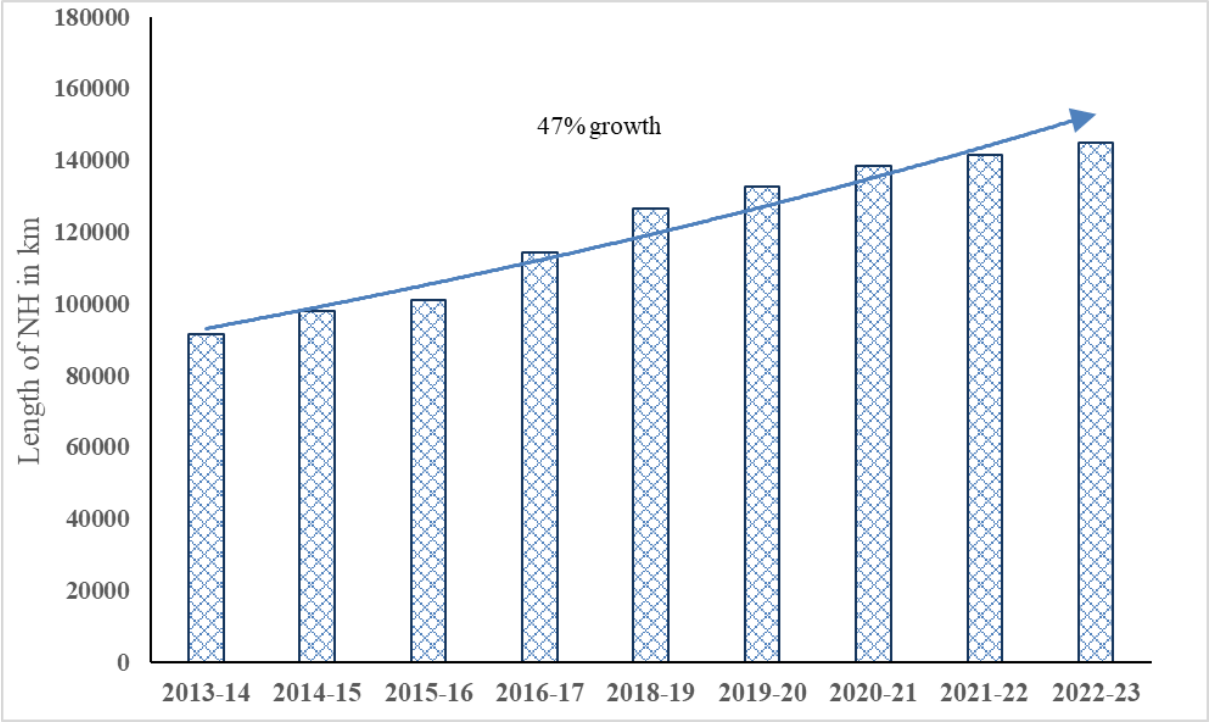


Figure 1: The yearly growth of National Highways in India [15].

various additives involves a lot of trial mixes, consequently resulting in utilisation of large quantity of materials. This poses a challenge in the field of building construction. Therefore, as artificial intelligence advances, a growing array of algorithms and models have emerged, offering fresh approaches that address these issues. Despite the comprehensive research conducted on nano-additives and fibres in OPC concrete, their integration into AAC has been less thoroughly investigated. The study addresses the gap by examining PQAC, which is specifically designed for sustainable pavement applications.

Nguyen and Dinh emphasised the importance of predicting the 28-day compressive strength of concrete from its initial ingredients, which is crucial for ensuring that concrete mixes meet strength requirements [16]. This predictive approach is echoed in the findings of Jamali et al., who applied AI methods to predict the compressive strength of fibre-reinforced polymer-confined concrete, further illustrating the adaptability of AI in diverse concrete scenarios [17]. Moreover, the use of machine learning techniques, including backpropagation neural networks (BPNNs), has been pivotal in predicting concrete strength. Li and Singh employed a BPNN to analyse the strength index of concrete with large recycled aggregates, achieving satisfactory prediction results [18]. This aligns with the findings of Haddad and Qarqaz, who conducted a comparative analysis of artificial neural network (ANN) for predicting bond strength in concrete, further validating the effectiveness of AI in structural applications [19]. Dao et al. explored novel hybrid AI techniques for predicting the compressive strength of geopolymer concrete, indicating that combining various AI methodologies can yield superior results compared to traditional methods [20]. This hybridisation approach is echoed in the work of Wang et al., who adapted ANNs for estimating concrete compressive strength, demonstrating the potential for integrating multiple AI techniques to improve predictive performance [21]. Wu et al. successfully predicted the tensile strength of high-performance concrete by combining ANN and Support Vector Regression (SVR) models using optimisation techniques [22]. The tensile strength of concrete was predicted by researchers using SVR and Gradient Boosting machine (GBM) models. The results showed that GBM provided superior prediction performance to SVR [23]. Four models were employed by Hammad et al., specifically M5P, model tree method, a gene expression programming, ANN and Random Forest (RF), to forecast the strengths of concrete comprising of metakaolin. Findings indicated that RF exhibited most accurate forecasting capability [24]. Nozar et al. investigated the compressive strength of concrete incorporating metakaolin by employing the Multi-Layer Perceptron (MLP) model. The results showed how the MLP network exhibited consistent accuracy in predicting

the compressive strength of metakaolin-infused concrete [25]. Additionally, an intuitive programme was created to simplify the utilisation of the suggested MLP network, which is grounded on machine learning techniques. Huang et al. introduced a model of hybrid machine learning that combines the firefly method with the RF algorithm to provide accurate forecasts of the compressive strength of cementitious materials when expanding clay is presents [26]. Abdulrahman et al.'s research assesses how well different ensemble and individual models predicted the compressive strength of binders made of expanding clays. The Decision Tree (DT) AdaBoost model and the enhanced bagging model showed the best prediction performance in the Strength of Metakaolin concrete, according to the findings of Bulbul et al [27]. While existing studies have primarily focused on OPC-based systems, this research extends ML applications to PQAC, addressing important gaps in the literature regarding its mix optimisation and mechanical performance. The study evaluates multiple ML algorithms to identify the most reliable model for predicting Split Tensile Strength (STS) in PQAC, ensuring robust and practical results. Nevertheless, there is a lack of research and study about the use of ML models for predicting the STS of PQAC that incorporates nano-additives and fibres. While most studies emphasise compressive strength prediction [28], this work addresses tensile properties critical for pavement applications, a relatively underexplored domain. This study employs advanced ensemble ML models, including RF, GBM and AdaBoost, to predict the STS of PQAC. Unlike simpler regression models or standalone ANNs, the ensemble models used here can handle complex, non-linear interactions between input features. By focusing on tensile properties, which are crucial for pavement applications, this research aims to optimise fibre dosages and nano-additive incorporation to enhance the tensile performance of PQAC. Experimental validation further confirms the relevance of ML predictions in real-world applications, enhancing the credibility of the proposed optimisation framework. Overall, this study fills a significant gap in current research by adapting machine learning to optimise PQAC properties, integrating the effects of fibres (PVAF and PPF) and nano-additives (NS and NA). The research uses a robust evaluation framework, including MAE, MSE, RMSE ,R^2, and CV metrics, to rigorously assess the predictive accuracy of the machine learning models. It also explores the effect of varying estimator counts in ensemble models, optimising them for practical applications. This comprehensive, data-driven approach aims to develop high-performance PQAC mixes tailored for sustainable construction applications. This approach minimises the need for extensive experimental trials, making it cost-effective and time-efficient. This combination of innovative material integration, focus on underexplored systems (PQAC), and the use of machine learning for predictive modelling sets the study apart in the field of fibre-reinforced alkali-activated nano-concretes.

2 Materials

The subsequent sections provide a description of all the materials used in the manufacturing of the PQAC mixes, which include ground granulated blast furnace slag (GGBS), river sand fine aggregate (RSFA), natural coarse aggregate (crushed granite) (NCA), liquid sodium silicate (LSS), sodium hydroxide (NaOH), water, NS, NA, PVAF and PPF.

2.1 Ground Granulated Blast Furnace Slag (GGBS)

GGBS was supplied by JSW, Thoranagallu, India, and was acquired from a local supplier. The properties of GGBS, including its chemical and physical properties, are obtained from the research work of Sheshadri et al. [29] as this research work is a continuation of the previous research of Sheshadri et al. and are in accordance with the necessary BIS and BS standards (IS 12089-1987; BS 15167-1:2006).

2.2 Natural Coarse Aggregate (NCA) and River Sand Fine Aggregate (RSFA)

We procured NCA from local vendors. Crushed, well-graded, angular, clean granite aggregates up to the highest aggregate size of 20 mm in accordance with IRC:44-2017 [32] were employed. RSFA belonging to zone II as per IS:383, 2016 [33] constituted the fine aggregates used in the research work. The properties of the aggregate are obtained from research work of Sheshadri et al., [29] as it is an extension of the previous research performed.

2.3 Alkaline Activator Solution for PQAC Mixes

The alkaline activator solution is employed to activate the cementing property of pozzolanic binders in PQAC mixes. In the current study, LSS, that is, Na_2SiO_3, along

with NaOH, that is, NaOH chips (98% purity), served as the alkaline activator solution. These chemicals were procured from a local chemical supplier. The composition of LSS can be described as follows: It consists of 14.70% sodium oxide (Na_2O), 32.80% silicon dioxide (SiO_2) and 52.50% water (H_2O) as per IS:14212, 1995 [34]. The 'activator modulus' (Ms) defined as the 'ratio of SiO_2 to Na_2O' in LSS was 2.23. At the same time, LSS and NaOH had specific gravities of 2.10 and 1.57, respectively. The dosage of NaOH in the LSS was adjusted to generate a Ms value of 1.25 when LSS + NaOH were combined with water; this ratio was maintained throughout the PQAC investigations. To obtain the desired Ms value of 1.25, the alkaline activator solution was made at least 24 hours prior to the concrete production by combining the NaOH chips with LSS and water. At this point, the water-to-binder (w/b) ratio was 0.20; later, while preparing the PQAC concrete, more water was added to raise the w/b ratio to 0.4.

2.4 Nano-Silica (NS):

NS particles are utilised, which are spherical and nano-sized. They are in the form of a white, odourless powder with an average particle size of 17 nm and a specific gravity of 2.2. The NS particles, owing to their larger specific surface area and superior fineness, enhance the pore structure of the concrete matrix. NS is composed of SiO_2 content of 99.88%, carbon content of 0.06%, chloride of 0.009%, Al_2O_3 of 0.005%, TiO_2 of 0.004%, and Fe_2O_3 of 0.001%.

2.5 Nano-Alumina (NA):

NA is characterised by a spherical shape and is found in the form of a white, odourless and crystalline powder. Their mean particle size is 40 nm with a specific gravity of 3.4. The nanoparticles, because of their superior fineness and specific surface area, enhance the pore structure of the concrete matrix. Nano-alumina comprises pure Al_2O_3, which is in a crystalline form. The NA utilised in the experiments is made up of spherical particles with an Al_2O_3 concentration of 99.9%, Fe_2O_3 of 0.0012%, SiO_2 of 0.015% and NaO_2 of 0.45%.

2.6 Poly-vinyl Alcohol Fibre (PVAF)

PVAF was procured by Fibre Region supplier, Chennai, India, with 12 mm length, 40 μm diameter, 1600 MPa tensile strength, a specific gravity of 1.1, an elastic modulus of 40 GPa and a density of 1.29 g/cm^3.

2.7 Polypropylene Fibre

PPF was procured by Fibre Region supplier, Chennai, India, with a 6 mm length, 20 μm diameter, 500 MPa tensile strength, a specific gravity of 0.9, an elastic modulus of 4 GPa and a density of 0.91 g/cm^3.

3 Mix Design

3.1 Mix Design and PQAC Specimen Preparation

A total of 19 PQAC mixes were developed. The initial reference mix was developed based on the available literature studies [35], [36] to meet desired standards for high quality concrete pavements (28 days) and a slump value of 25–75 mm. For each mix ID, a total of 30 concrete cylinders were cast for testing. Therefore, the total number of samples that were cast was 540. Two nano-additives, that is, NS and NA, were adapted in the mix design to improve the strength of the PQAC. NS was used as a nano-additive at intervals of 0.5 starting from 0% to 2.0% by weight of the binder. NA was added at regular intervals of 0.25% starting from 0% to 1.25% by weight of the binder.

Furthermore, in order to overcome the brittleness of PQAC, fibres were incorporated into the concrete mix. The two fibres which were individually incorporated in the PQAC mixes are PVAF and PPF. Both the fibres were added at regular intervals of 0.4% starting from 0% to 2.0% by volume of the binder. The details of the mix design and mix IDs are presented in **Tables 1 and 2**. The two letters of the mix ID represent the nano-additive adapted in the mix design (**Table 1**). The three letters of the mix ID represent the fibre adapted in the mix design **(Table 2**).

4 Methodology

4.1 Experimental Tests

The quantities of GGBS, RSFA, NCA, NS, NA, PVAF and PPF were measured by weight. The base reference mix was designed to meet desired standards for high quality concrete pavements with a water-to-binder (w/b) ratio

Table 1: Mix Design of PQAC mixes with nano-additives.

Mix ID	A-0	AS-0.5	AS-1.0	AS-1.5	AS-2.0	AA-0.5	AA-0.75	AA-1.0	AA-1.25
% Addition of nano-additives (by weight of binder)	0%	0.5%	1.0%	1.5%	2.0%	0%	0.75%	1.0%	1.25%
Materials	in kg/m^3								
GGBS	493	493	493	493	493	493	493	493	493
NaOH flakes	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2
Liquid sodium silicate	75.12	75.12	75.12	75.12	75.12	75.12	75.12	75.12	75.12
Water	157.56	157.56	157.56	157.56	157.56	157.56	157.56	157.56	157.56
Natural coarse aggregate	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4
River sand fine aggregate	577.84	577.84	577.84	577.84	577.84	577.84	577.84	577.84	577.84
Nano-silica	0	2.47	4.93	7.39	9.86	-	-	-	-
Nano-alumina	-	-	-	-	-	2.47	3.69	4.93	6.16

Table 2: Mix Design of PQAC mixes with fibres.

Mix ID	APVA-0.4	APVA-0.8	APVA-1.2	APVA-1.6	APVA-2.0	APF-0.4	APF-0.8	APF-1.2	APF-1.6	APF-2.0
% Addition of fibres (by volume of binder)	0.4%	0.8%	1.2%	1.6%	2.0%	0.4%	0.8%	1.2%	1.6%	2.0%
Materials	in kg/m^3									
GGBS	493	493	493	493	493	493	493	493	493	493
NaOH flakes	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2
Liquid sodium silicate	75.12	75.12	75.12	75.12	75.12	75.12	75.12	75.12	75.12	75.12
Water	157.56	157.56	157.56	157.56	157.56	157.56	157.56	157.56	157.56	157.56
Natural coarse aggregate	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4	1071.4
River sand fine aggregate	577.84	577.84	577.84	577.84	577.84	577.84	577.84	577.84	577.84	577.84
PVA	0.748	1.49	2.24	2.99	3.74	-	-	-	-	-
PPF	-	-	-	-	-	0.612	1.224	1.836	2.448	3.06

of 0.4. NS and NA were added in increments of 0.5% and 0.25%, respectively, by weight of the binder. PVAF and PPF were added in increments of 0.4% by volume of the binder. GGBS, RSFA, NCA and the respective percentages of NS/NA were thoroughly mixed in a mechanical mixer for 2 minutes to ensure even distribution of dry components. The alkaline activator solution was slowly introduced to the dry mix while the mixer operated at low speed followed by additional water content. Mixing continued for 3 minutes to achieve homogeneity. PVAF or PPF fibres were gradually added over 1 minute to avoid clumping. The mixing was continued for additional 2 minutes to ensure uniform fibre dispersion. Cylindrical moulds (100 mm diameter × 200 mm height) were filled in three layers. Each layer was compacted using a vibrating table for 30 seconds to eliminate entrapped air. Specimens were demoulded after 24 hours and immediately transferred to curing conditions. The samples were water cured for 7 days and then air cured for the remaining period up to 28 days before the testing.

A flow chart representing the experimental framework is shown in **Figure 2**. Various tests were performed for the PQAC mixes developed as per the prescribed standards, the details of which are provided below.

4.2 Fresh Property

The compaction factor test was performed as per IS 1199 (Part II) [37] in order to assess the workability of each PQAC mix, the results of which are presented in **Figure 3**

4.3 Mechanical Property.

The STS test was performed on 30 cylindrical samples for each mix having a dimension of '200 mm height and 100 mm dia' as per the Indian Standard Codes IS 5816-1999 [38] for the PQAC mixes. The obtained results are presented in **Figure 4**

4.4 Experimental data

4.4.1 Compaction Factor

Figure 3 indicates the variation in compaction factor value (CFV), and the CFV exhibits a discernible trend of reduction with an increasing proportion of nano-additives and fibres. The CFV is documented as 0.923 for the reference mix (A-0-0), which is distinguished by the absence of nano-additives and fibres.

Nevertheless, as the concentration of NA is increased in 0.25% increments from 0% to 1.25%, the CFV progressively decreases from 0.923 to 0.807. The workability of PQAC reduces with NA incorporation as the NA particles tend to agglomerate due to their high surface energy. These agglomerates can act as barriers within the mix, inhibiting the movement of other particles causing increased viscosity, hindering flow and making the mixture less fluid for proper compaction and placement [39], [40]. In a similar manner, increasing the concentration of NS from 0% to 2% in 0.5% increment results in decrement of CFV values from 0.923 to 0.806. NS rapidly reacts with alkaline solution and water, producing a viscous liquid, which is the cause for decreased workability in PQAC mixes [41], [42]. In addition, the inclusion of fibres decreased the workability of PQAC mixtures. In case of incorporation of PVAF into the PQAC mix, the workability reduced from 0.923 to 0.797 as the percentage of PVAF increased from 0% to 2%, respectively. The decrease in workability is caused by the hindered movement of the paste due to the inclusion of fibres. The inclusion of PPF fibres resulted in a decrease in workability, with a decrease from 0.923 to 0.796 as the proportion of PPF introduced increased from 0% to 2% accordingly. The inclusion of fibres in the concrete matrix hinders the flow of cement paste, leading to a decrease in workability [43].

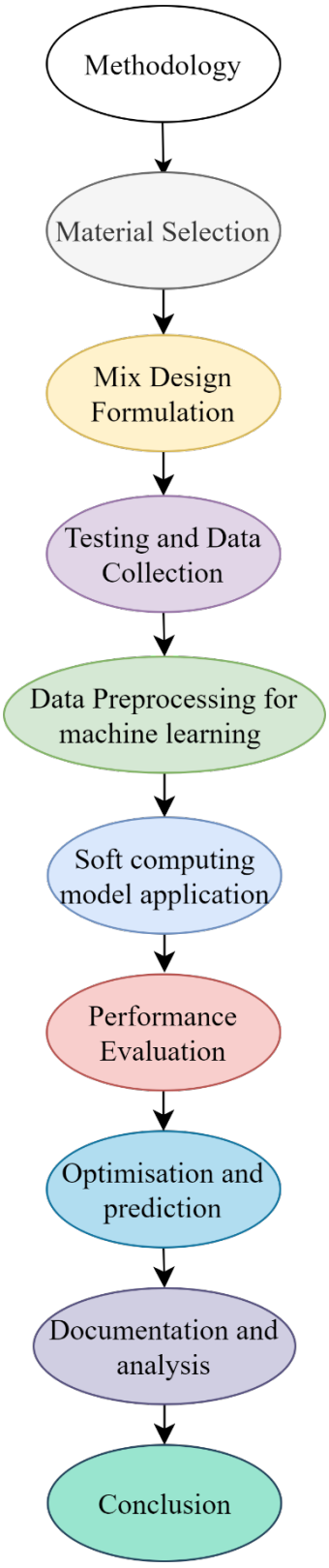


Figure 2: The detailed flowchart for the experimental programme and analysis.

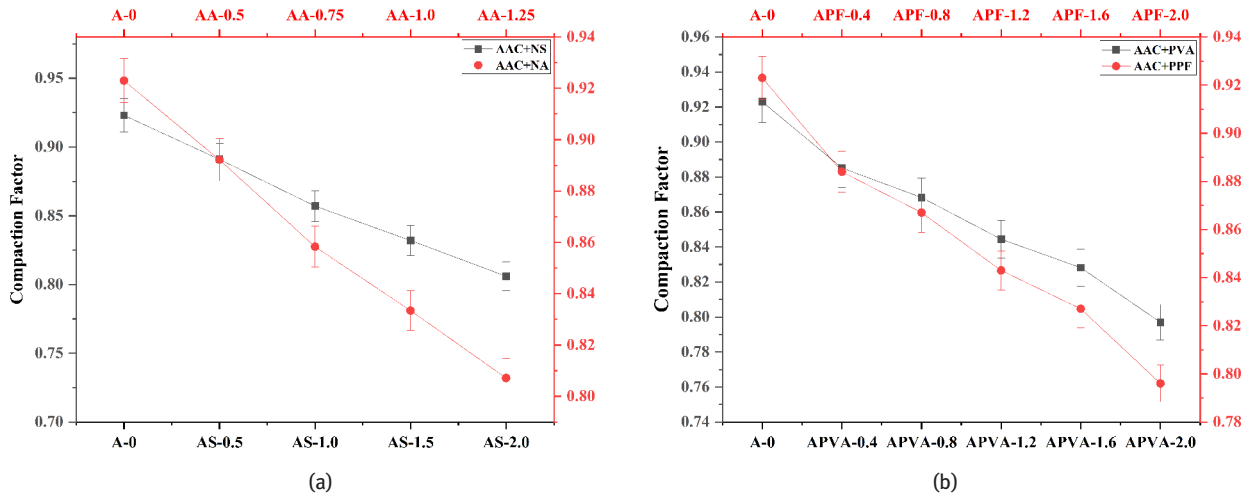


Figure 3: Compaction factor values of PQAC mixes: (a) PQAC+NA and PQAC+NS and (b) PQAC+PVA and PQAC+PPF.

4.4.2 Split Tensile Strength (STS)

The STS values for PQAC mixtures at 28-day curing are shown in **Figure 4.** The A-0-0 (Reference mix), had STS of 4.3 MPa. The STS in case of addition of NS was found to be 4.54, 4.78, 5.14 and 4.62 MPa, as the percentage addition increased from 0.5% to 2.0% at regular intervals of 0.5%, respectively. Similarly, the STS of PQAC mixes on addition of NA was found to be 4.58, 4.83, 5.26 and 4.81 as the percentage addition increased from 0.5% to 1.25% at regular intervals of 0.25%, respectively. The STS varied as the PVAF content varied from 0.4% to 2.0% at regular intervals of 0.4% as 4.98, 5.36, 5.82, 7.5 and 6.26 MPa.

The STS varied as the PPF content varied from 0.4% to 2.0% at regular intervals of 0.4% as 4.71, 5.23, 5.70, 6.85 and 6.12 MPa. Notable observations are identified based on the analysis of the results. The incorporation of NS improved the STS of PQAC mixes up to 1.5% beyond which there was decrement in strength. A similar pattern is observed in case of NA where the STS increased up to an addition of 1.0% beyond that optimum point and further increment in the NA caused a decrease in the strength of PQAC mixes. Nano-additives have high surface area-to-volume ratios, allowing them to form strong bonds with the surrounding cementitious matrix [39]. This enhanced interfacial bonding improves the load transfer mechanism within the material, leading to increased stiffness and modulus of elasticity; they also act as fillers within the ITZ between the aggregate particles and the cementitious gel, which improves the bonding between these zones, creating a more continuous and stronger network, and good interfacial adhesion is responsible for overcoming local failures and providing higher resistance to bending and fracture forces. The existence of evenly distributed nano-particles impedes the spreading of fractures when subjected to tensile stress. By bridging micro-cracks or providing additional nucleation sites for further gel formation around the crack tip, NA can help prevent the growth of cracks and improve overall tensile strength [51 52]. Nevertheless, at higher nano-additive contents, if the particles are not well dispersed in the mix, they may cause agglomeration and weak bonding, resulting in weaker zones and a decrease in strength. Hakamy proposed that the agglomeration of the surplus unreacted fine materials functions as both stress concentrators and defect sites, weakening the material [45], and therefore, an optimised content of the nano-additive needs to be considered for incorporation in the mix. The PVAF and PPF also followed the same trend that the STS increased substantially up to a certain percentage addition of fibres, that is, 1.6% both in case of PVAF and PPF beyond which the STS reduced. Fibres in the PQAC matrix act as discrete reinforcements, effectively bridging cracks and improving the tensile load-bearing capacity of the material. As the fibre content increases, the number of fibres available to intersect and arrest the propagation of cracks also increases, leading to enhanced tensile strength. At lower fibre contents, they are typically well dispersed and distributed within the concrete matrix, maximising their reinforcement effectiveness. However, as the fibre content increases beyond the optimal point, agglomeration and poor dispersion can occur, leading to the formation of weak zones and stress concentration points, ultimately reducing the tensile strength [46]. Furthermore, the workability and compaction of the PQAC mix reduce, making it challenging to achieve a homogeneous and

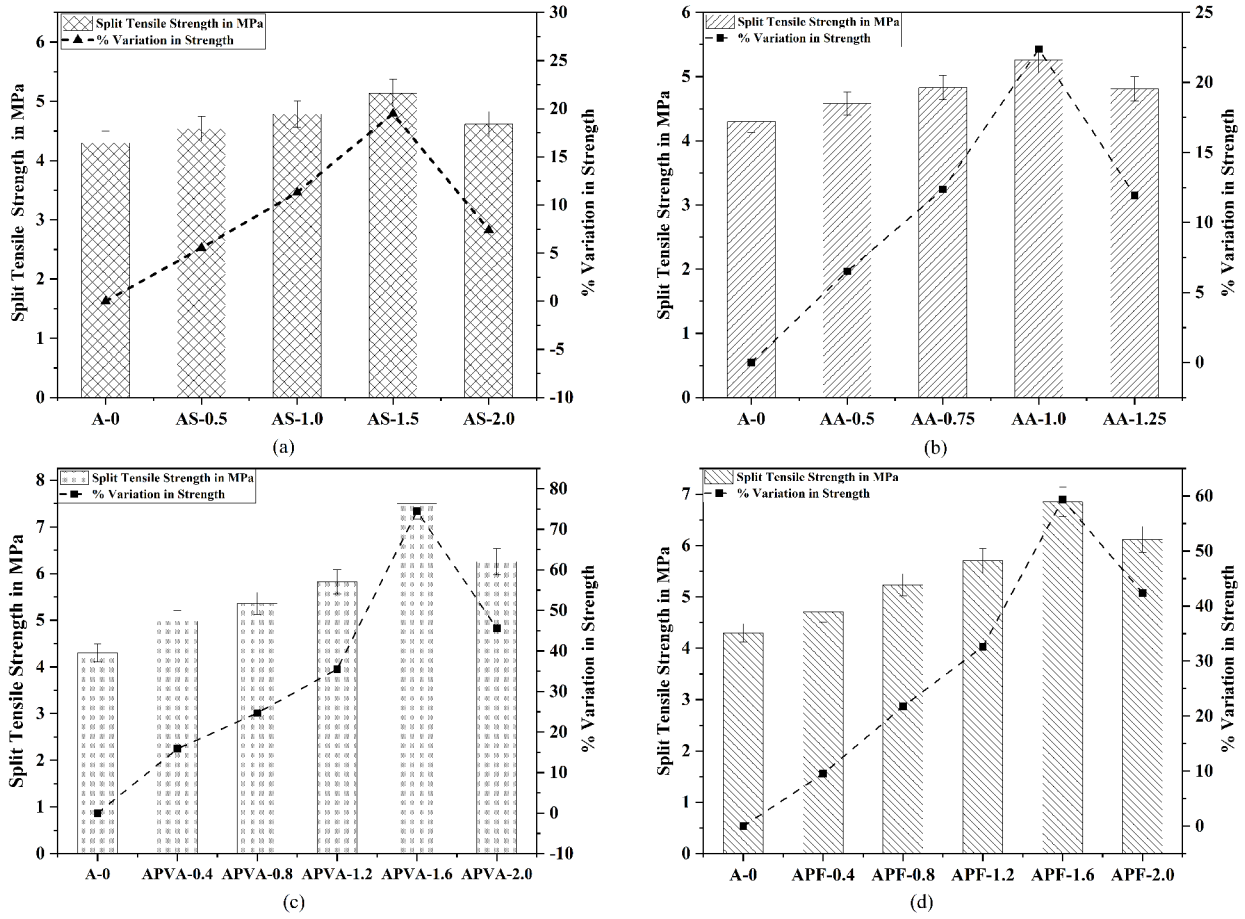


Figure 4: Split tensile strength and percentage variation in split tensile strength: (a) PQAC+NS, (b) PQAC+NA, (c) PQAC+PVAF and (d) PQAC+PPF.

void-free matrix. Improper compaction can result in the formation of voids and defects, which can act as stress concentrators and reduce the overall tensile strength [43].

4.5 Data Collection and Description

The area of ML is widely applied across several domains and is now undergoing extensive study to enhance its practicality. The present study use 399 samples, with 70% randomly designated as the training set and the remaining 171 samples, constituting 30%, utilised as the test set for the trained model. The STS has been forecasted utilising ML techniques, such as MLR, DT, RF, AdaBoost, SVR and GB. This study employs many tools, including DT, MLR, RF, ADA, SVR and GB, within the Sci-kit-Learn software package using Python.

A representative dataset is essential and crucial for ML models. Hence, this study gathered a grand total of 570 samples through its experimental works. **Table 3** shows the statistical features of the variables found in these samples. The input variables include GGBS, NaOH, LSS, RSFA, NCA, Water, NS, NA, PVAF and PPF. It can be observed that the content of NS, NA, PVAF and PPF varies in the entire data set. The aim of this computational modelling work is to utilise ML methods to create metamodels that represent the STS as a mathematical function of 10 variables.

4.6 Machine Learning Methods

4.6.1 Multiple Linear Regression (MLR)

MLR is an expanded form of the basic linear regression model. MLR models are used to estimate the degree of correlation between the input variable and the response variable [47], [48]. It is a ML approach that selects a predictor variable based on a variety of input factors.

Many studies employed linear regression or MLR to predict the output variables efficiently [49]. The study incorporates many independent factors and one or more dependent variables, which can be either categorical or continuous in nature. It may be used to simulate the effect of independent factors on the dependent variable, to test hypotheses regarding the coefficients of the independent variables and to generate predictions based on the fitted model [50], [51].

The equation for multiple linear regression is represented by Eq. 1:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \beta 3X3 + \cdots + \beta pX\,p + \in \qquad (1)$$

where Y is the dependent variable, X1, X2, X3..., Xp are the independent variables, β0 is the intercept, β1, β2, β3..., βp are the regression coefficients and ∈ is the error term.

The aforementioned prognosis of STS of PQAC, here, STS, is considered as the dependent variable. The quantities of GGBS, NaOH, LSS, RSFA, NCA, Water, NS, NA, PVAF and PPF are considered as the independent variables. The fundamental objective of MLR is to allocate the coefficient and intercept values in a sequence of iterations in a manner that minimises the error or cost function, ensuring that the predicted values are as close as possible to the actual values. The dataset was divided into training and testing subsets in a 70:30 ratio, assigning 70% of the data to the training group. Thereafter, the co-efficient and constants were calculated. Additionally, the formulation was developed by altering the values of constants, and the intercept was subsequently assessed using the remaining 30% of the information. Nevertheless, if the method fails to implement appropriate sample and selection of independent criteria, it will not yield dependable findings for the end users.

4.6.2 Decision Tree Regression (DT)

DT is a supervised ML model that is capable of modelling both classification and regression problems [52]. In order to generate a DT, the DT model divides the samples according to their features. It is a rule-based model that predicts the mapped target value by generating logical divisions in data [53]. DT regression uses the characteristics of an item to train a model inside a hierarchical framework, facilitating the estimation of future data and the generation of meaningful continuous outputs. The procedure entails analysing the data via a series of enquiries, each aimed at refining the prospective values till a dependable model is developed. DT regression is a technique that partitions a dataset into more compact subgroups. The structure comprises a decision leaf that divides into two or more branches, each of which denotes the value of the attribute under investigation. The DT model comprises of three nodes: the root node, the decision node (or internal node) and the terminal node (or leaf node). The root node is the most suitable predictor and is located at the topmost level. The algorithm initiates at the root node, encompassing all training data, subsequently leading to the division of the root node into decision nodes. The process persists through the following levels until the tree reaches a specified maximum depth or when the nodes contain only a single sample from the training data. The approach aims to reduce the loss function, particularly metrics such as mean squared error (MSE). The model is calibrated to independent factors in conjunction with the target variable and is partitioned at designated places referred to as leaf nodes within a dataset for every variable that is independent. The loss function is observed, and the partition with the least error is chosen, perpetuating this procedure iteratively. However, the DT can become excessively large and overfit the data, leading to reduced generalisation on the test set and potentially impairing prediction accuracy. A threshold is established for multiple splits, or a minimum number of data points is required for each leaf node in order to construct the DT and mitigate the risk of increased misclassification [54], [55].

4.6.3 Random Forest Regression (RF)

RF algorithm is a method of ensemble estimation. It is utilisable for both regression and classification tasks. An ensemble method is superior to a single DT methodology since it reduces over-fitting by averaging the results. The RF algorithm employs basic DTs as the fundamental learners [56]. RF is an integrated learning model that comprises multiple DTs. The fundamental concept is to enhance prediction accuracy and stability through the construction of multiple DTs. Each DT is constructed using random samples and features, which enables RF to mitigate overfitting and exhibit strong robustness. The following benefits include: RF can employ multiple DTs for prediction, resulting in greater accuracy compared to a single DT. RFs are capable of managing a substantial quantity of input features, making them suitable for classification and regression tasks involving high-dimensional data. They are developed through the use of random samples and random features, which mitigates the issue of overfitting. RF represents an enhanced iteration of the DT model, employing the ensemble method to integrate

several DTs, thereby diminishing the overall model's variance [57]. Every DT is constructed from the sample by employing a bootstrap/bagging technique. Utilising the voting and averaging approach, the output of all the DT is consolidated into a single value for classification and regression tasks. Bagging methodology iteratively chooses B random samples (with replacement) using a specified training dataset (X) to train DTs. During the construction process, these DTs are formed independently and do not interact with each other. Once trained, these trees are capable of predicting the unknown data. In general, DTs have a quite large variation. Nevertheless, when they are merged simultaneously to develop the RF, the total variance decreases. Addition of randomisation serves to reduce the correlation among individual trees in the ensemble. Selecting the variable for that particular node minimises the error for the target variable [58]. This work implements the RF algorithm using the scikit-learn Python tools [59], [60]. The prediction of this algorithm is dependent on the RF's multiple trees, which are more generalised [61]. Conversely, RF is more challenging to interpret and visualise than Decision Trees [58], [61].

4.6.4 Support Vector Regression (SVR)

Many real-world datasets may exhibit a non-linear connection between input and output properties, which the linear regression technique may not be capable of detecting. The SVR approach is commonly employed for constructing input–output model mappings due to its efficient resolution of non-linear regression issues. SVR has shown several effective applications in the domain of civil engineering [62]. SVR is a linear approach, it is a supervised algorithm from the machine learning family that was developed by Vapnik in 1995, and it is considered one of the most advanced methods in the field of machine learning in order to resolve classification and regression issues [63]. The objective of this supervised learning approach is to optimise the distance between the separator hyperplane and the nearest train point (support vector) for each class in order to get the required performance for training data. Once linear data separation becomes unattainable, a mapping to a higher space is employed to facilitate this separation. An algorithm called Kernel is used to do this mapping, enabling the determination of a non-linear decision boundary in a high-dimensional space. Subsequently, the margin boundary lines are drawn from the hyperplane at a given error threshold distance, represented by the symbol epsilon. We call this kind of loss 'ε-insensitive loss'. Outside of the margin mapped data points are not included in the training process. The goal of SVR for regression is to determine a hyperplane that best fits the data within a specified error margin, ensuring minimal deviation from actual values while maintaining generalization [60], [64]. There are numerous kernels available, including the linear function, Radial Basis Function and polynomial function [65]. A linear function was employed in the investigation. The SVR algorithm is executed utilising Scikit-learn tools [60], [65].

4.6.5 ADA Boost Regression

Freund and Schapire were the first to suggest AdaBoost. It is an ensemble technique. The learning rate and the number of estimators are the primary parameters of the AdaBoost algorithm. AdaBoost operates by integrating multiple weak learners into a singular robust learner. In general, AdaBoost employs single divided decisions, termed 'decision stumps', to assist feeble learners [66]. The fundamental principle of the AdaBoost learning method is to construct a robust classifier with excellent detection performance by combining weak classifiers. It is a learning method that performs two functions involving repeated computations: function selection and classifier conditioning. By iteratively refining computations, the overall classification performance improves, as weak classifiers, which contribute to performance evaluation, are integrated into the strengthened iterative classifier [67]. The AdaBoost algorithm assigns weights to classification examples based on their difficulty level, meaning that instances that are difficult to classify are assigned a higher weight than those that are simple to classify. It adjusts to the mistakes produced by classifiers in prior rounds [68], [69] adaptively. With each iteration, adjusting the weight of training examples compels the learning algorithms to prioritise instances that were previously misclassified and downplay instances that were previously successfully classified. Put simply, weights are increased for misclassified cases and minimised for correctly classified ones. Consequently, the subsequent iterations will assign a higher weight to misclassification mistakes for those misclassified occurrences. The use of adaptive boosting extends to both regression and classification problems. Transitioning from Linear Regression to AdaBoost enables the mapping of a wider range of non-linear connections, leading to improved estimate and, consequently, increased accuracy.

Table 3: Statistical summary of the dataset.

	Count	Mean	Std	Min	25%	50%	75%	Max
GGBS	570	493	0	493	493	493	493	493
NaOH	570	11.2	0	11.2	11.2	11.2	11.2	11.2
LSS	570	75.12	0	75.12	75.12	75.12	75.12	75.12
Water	570	157.56	0	157.56	157.56	157.56	157.56	157.56
NCA	570	1071.4	0	1071.4	1071.4	1071.4	1071.4	1071.4
RSFA	570	577.84	0	577.84	577.84	577.84	577.84	577.84
NS	570	1.3	2.81	0	0	0	0	9.86
NA	570	0.91	1.87	0	0	0	0	6.16
PVAF	570	0.59	1.13	0	0	0	0.75	3.74
PPF	570	0.48	0.92	0	0	0	0.61	3.06
STS	570	5.34	0.94	3.69	4.64	5.17	5.81	8.2

4.6.6 Gradient Boosting Regression (GBR)

Friedman introduced Gradient Boosting (GB) in 1999 as an ensemble method for doing regression and classification tasks. Among the several boosting techniques, GB is unique in that it is exclusively applicable to regression tasks. The GBR algorithm employs many base learners in order to minimise the residual error [70]. GBR employs a gradient descent technique to address the problem, where each weak learner iteratively decreases the prediction error according to the established learning rate [71]. A sequential base learner is employed, where each subsequent learner minimises the error made by the preceding learner, therefore resulting in the development of a generalised model at the conclusion. Within this technique, every iteration of the randomly chosen training set is verified against the base model. By randomly subsampling the training data, the speed and precision of GBR for execution may be enhanced, therefore mitigating the risk of overfitting. The approach operates based on the boosting concept, in which several weak learners are together combined to create a robust learner. In general, DTs are employed as the weakest learners. First, single-node basis trees are built. The construction of subsequent trees is contingent upon the mistakes made by the preceding trees. The scaling of the trees is achieved by selecting the learning rate, therefore guaranteeing the equitable contribution of each tree to the total prediction. The next trees are combined with the previous trees in order to forecast the reaction. This procedure is iterated until the maximum number of trees is achieved or the ensuing answers do not show improvement [72]. GBR is applicable for predicting numerical outputs, so the response variable must be numerical.

5 Results and Discussion

5.1 Pair Plots

The compiled database is examined to reveal the statistical attributes, as shown in **Table 3,** of all input and output parameters. The STS of the evaluated concrete mixes ranges from 3.69 to 8.2 MPa, reflecting the variability across the test samples in the dataset. **Figure 5** presents pair plots illustrating the distribution of data points across different levels. The pair plot of the dataset provides a comprehensive visualisation of the relationships between the various input variables and the output variable, STS. The diagonal elements of the pair plot illustrate the distribution of individual variables, revealing that certain variables, such as GGBS, water content (W), RSFA, LSS, NaOH and NCA, exhibit low variability because of their categorical nature or restricted range in the experimental design. The off-diagonal scatter plots showcase the relationships between variable pairs, emphasising the sparse distribution of many variables, which suggests limited unique values or discrete levels within the dataset. This sparsity is especially evident in the integer-based variables. While some scatter plots display distinct patterns suggesting potential correlation such as clustering many variable pairs do not show strong correlations, implying that certain factors may have a more pronounced influence on STS than others. Overall, the pair plot serves as a preliminary exploratory tool, providing insights into the underlying data structure and guiding subsequent detailed analyses. Research has demonstrated that PVAF and PPF contribute positively to the STS of PQAC. The findings suggest that optimising PVAF and PPF content could lead to enhanced STS, as they not only

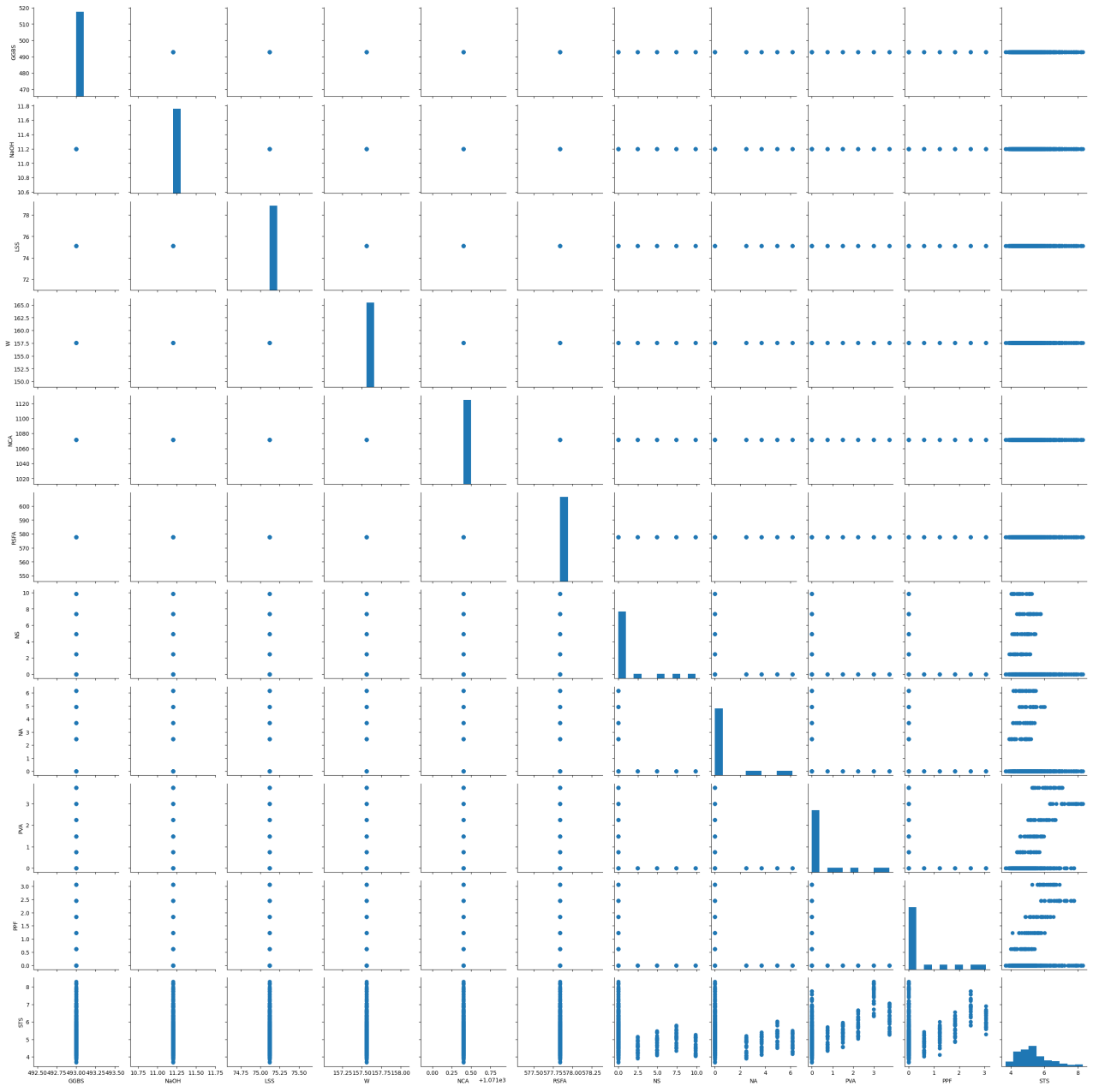


Figure 5: Pair plots for the input parameters and observed responses.

serve as a reinforcement but also interact synergistically with other components, improving the overall mechanical properties of the concrete. In contrast, parameters such as GGBS, NaOH, LSS, Water, RSFA and NCA remain constant across the dataset, indicating limited variability and thus a negligible impact on STS. The lack of variation in these parameters suggests that they are unlikely to exhibit significant relationships with STS, as their fixed values do not allow for the exploration of potential correlations. The scatter plots involving NA/NS and STS suggest a non-linear relationship, where higher dosages of NA/NS may lead to specific changes in STS. However, the overall weak correlation indicates that while NA/NS can influence STS, its effect may not be as pronounced as that of PVAF/PPF. This aligns with literature that discusses the complex interactions between various nano-additives and fibres and their contributions to the mechanical properties of concrete. Furthermore, the variability of nano-additives and fibres in the dataset may indicate their potential significance as input parameters. These parameters show distinct clustering or trends in relation to STS, and they are also considered as important factors influencing STS.

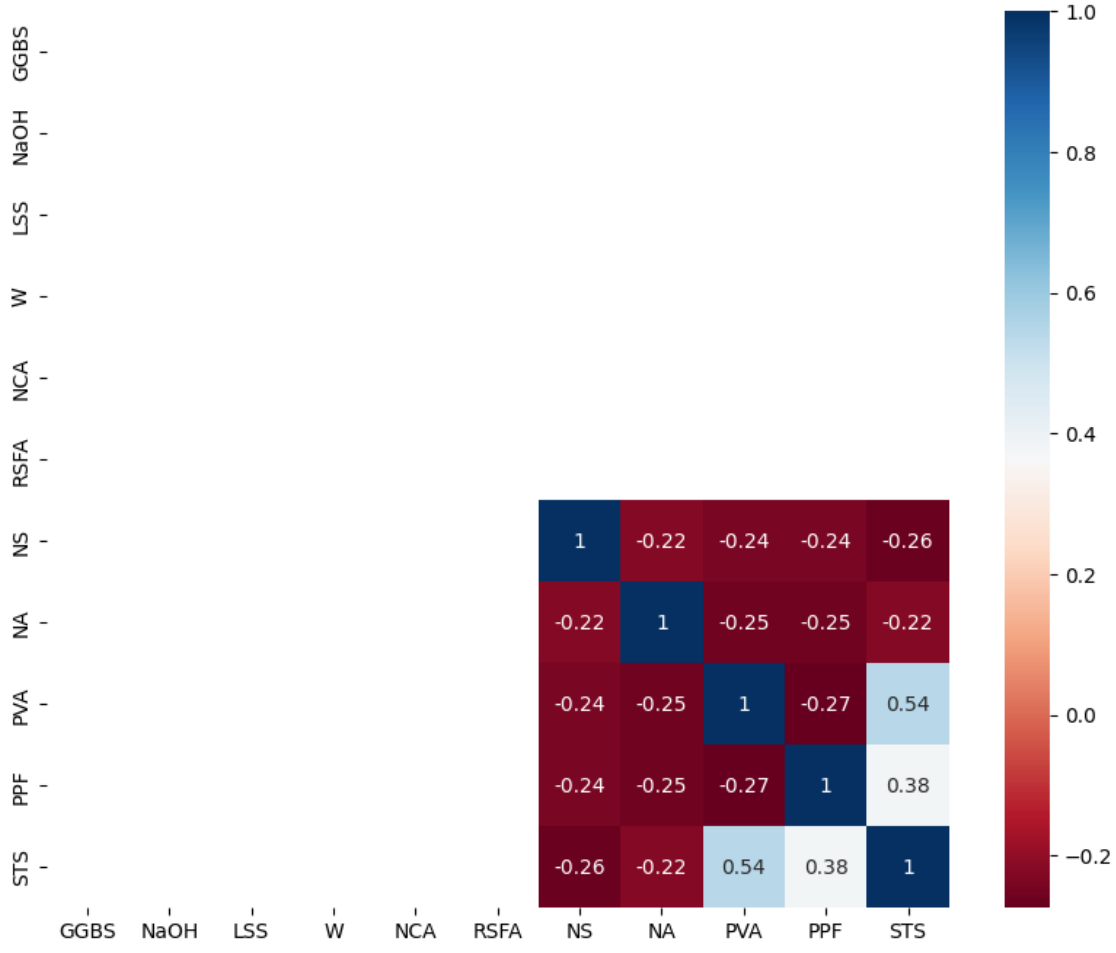


Figure 6: Correlation heatmap of input parameters and observed responses.

In conclusion, the analysis underscores PVAF as the most significant parameter affecting STS in PQAC, with PPF, NA and NS also showing potential influence depending on their variability and interactions with other materials.

5.2 Heatmap Co-relation

The co-relation matrix is a square matrix that displays each dataset's pairwise co-relation coefficients between variables [73], [74], [75]. Every value in the matrix corresponds to the co-relation between two variables, whereas the diagonal components indicate the co-relation between each variable and itself, which is 1 [76]. Data co-relation analysis is a valuable technique for examining the connections between variables in a dataset and detecting patterns or trends. It can also be employed to detect multicollinearity, a problem that arises in statistical models when two or more predictor variables exhibit a high co-relation with each other.

Figure 6 displays the graphical representation of the correlation matrix between the input and output parameters, known as a Heatmap. Evidently, out of the 11 input characteristics mentioned, the correlation coefficients of 0.54 and 0.38 for 'PVAF' and 'PPF', respectively, constitute the most robust correlation. The parameter PVAF shows a strong positive correlation (0.54) with STS. This suggests that increases in PVAF are associated with increases in STS, highlighting a potentially significant relationship worth further exploration. Several parameters exhibit moderate negative correlations, such as STS and NS (-0.26) or NA and STS (-0.22). These relationships indicate that as one parameter increases, the other tends to decrease, which could be relevant for understanding input–output dynamics in the system being studied. Nevertheless, the association remains feeble, suggesting that making predictions using MLR for STS is not dependable because of the presence of intricate non-linear connections between these factors and the result.

5.3 Relative Frequency Distribution and Statistical and Scalar Analysis

Figure 7 illustrates the relative frequency distribution of the predictions-to-test ratio across six predictive modelling techniques; each panel (a–f) includes histograms that depict the frequency of observed prediction ratios, accompanied by vertical lines indicating mean values (in orange), median values (in green) and standard deviation (SD) (in blue) for each model. Desirable circumstances are characterised by a lower standard deviation and a mean value approaching 1. A mean number above 1.0 defines over-prediction, whereas a mean value below 1.0 shows under-prediction. The AdaBoost model consistently delivers the most precise predictions, as seen by its mean value of 0.99 and matching standard deviation of 0.084. The DT model follows closely with a mean value of 1.0026 and an SD of 0.0855, followed by RF with a mean value of 1.0027 and an SD of 0.0858. The MLR model has a mean value of 1.003 and a standard deviation of 0.108. The SVR and GBR models exhibit comparable performance, with mean values of 1.0048 and 1.0045, respectively, and SD values of 0.109 and 0.087, respectively. The greatest mean values and SD values in the SVR model indicate the weakest performance. The histograms show that the predictions-to-test ratios vary significantly across different models. Each model's predictions cluster around different ranges, reflecting their distinct predictive capabilities. The vertical lines representing the mean highlight the central tendency of each model. Notably, RF and GB demonstrate a higher mean prediction ratio, indicating better performance in comparison to other models. The width of the distributions suggests variability in the models' predictions. For instance, SVR exhibits a narrower distribution range, implying more consistent predictions, while models like DT and AdaBoost show broader distributions, indicating variability in their performance. Certain models, such as AdaBoost, reveal potential outliers in the predictions, suggesting instances where the model performs significantly better or worse than the average. The mean prediction ratio varies across models, with GB exhibiting the highest mean (1.004576), followed closely by AdaBoost (0.999269). This suggests that GB consistently produces higher predictions relative to test scores. The median values also reflect this trend, aligning closely with the mean, indicating a symmetrical distribution for most models. The standard deviation values indicate the variability of the predictions. MLR (0.108407) and DT (0.085500) show lower variability, while AdaBoost (0.0845896) and GB (0.08731) have higher dispersion, suggesting greater variability in prediction accuracy. The skewness values highlight the asymmetry of the distributions. SVR shows a moderate positive skew (0.076785), while GB has a more moderate negative skew (-0.416545), indicating that a greater number of predictions are clustered around lower values with a tail extending towards higher values. Based on the relative frequency distribution observed in Figure 7, GB emerges as the best method among the compared models. Its higher mean prediction ratio, coupled with a reasonably spread distribution, reflects a strong balance between accuracy and consistency in predictions. This finding underscores the effectiveness of GB in capturing complex patterns in the data while minimising prediction error, making it a robust choice for achieving high predictive performance in this context.

Table 4: Relative frequency distribution of the prediction-to-test STS ratio.

Model	Mean	Median	Standard deviation	Skew
Linear Regression	1.003337	0.99848	0.108407	0.154706
Decision Tree	1.002638	0.98704	0.0855	0.399816
Random Forest	1.002764	0.98737	0.085843	0.40379
Support Vector	1.004801	0.99847	0.109484	0.076785
ADA Boost	0.999269	0.98854	0.084598	0.367988
Gradient Boosting	1.004576	0.99069	0.08731	0.416545

5.4 Violin Diagram for Relative Error

Figure 8 displays a violin graph of the different models' relative error percentages for various predictive models, including MLR, DT, RF, SVR, AdaBoost, and GBR. It is evident that in the test dataset, the GBR model shows a nearly zero relative prediction error and a higher concentration than other models. The mistakes' statistical analysis emphasises the positive prediction performance even more. The plot reveals that the MLR model exhibits the widest distribution of errors, with a large spread from approximately -35% to 30%, indicating significant variability in its predictions. Simpler models, such as MLR, struggle with the complex, non-linear relationships inherent in the dataset. This limitation is evident from the wider distribution of errors, indicating that MLR fails to capture significant variances in the data effectively.

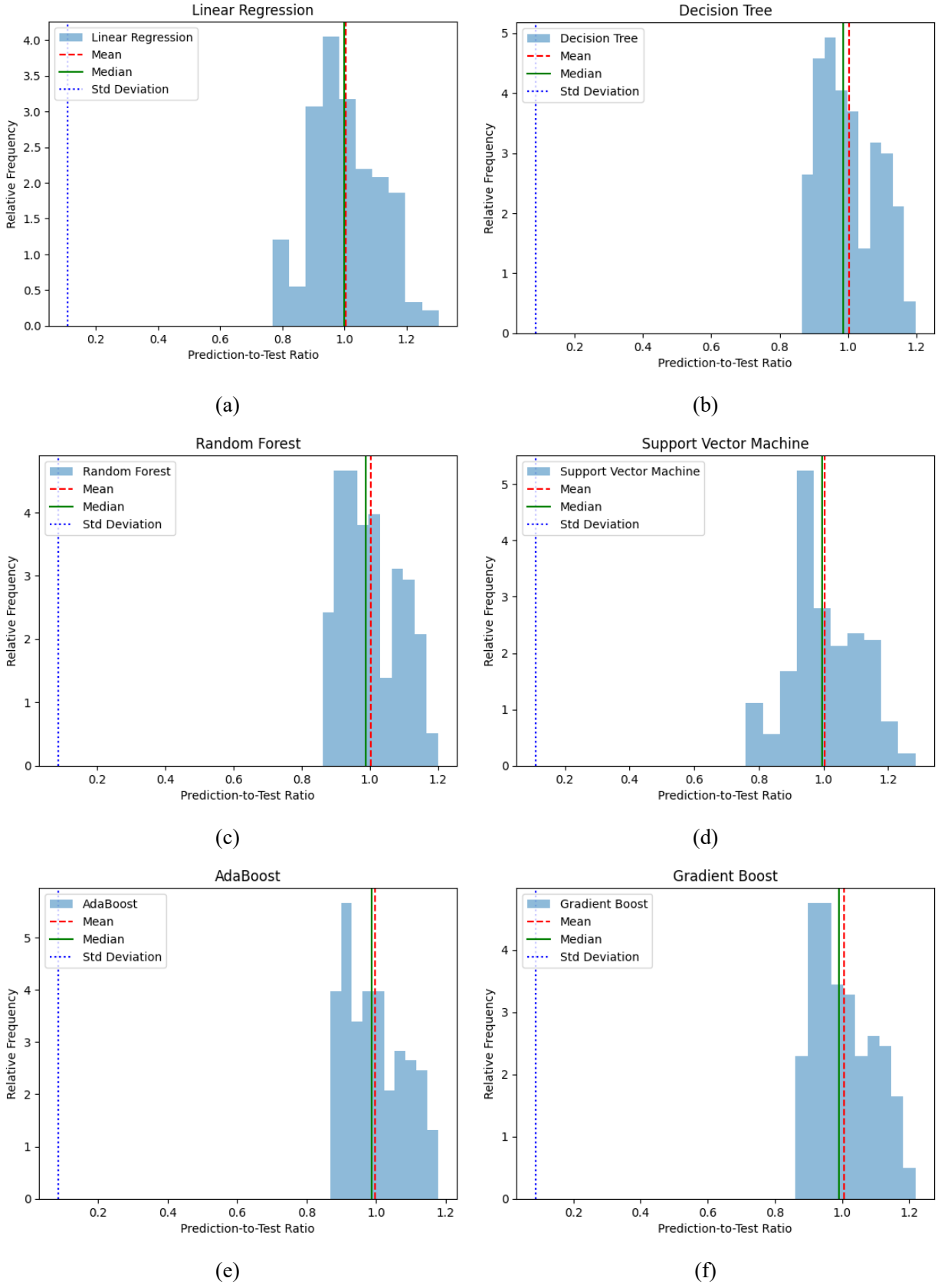


Figure 7: Relative frequency distribution of the prediction-to-test STS ratio.

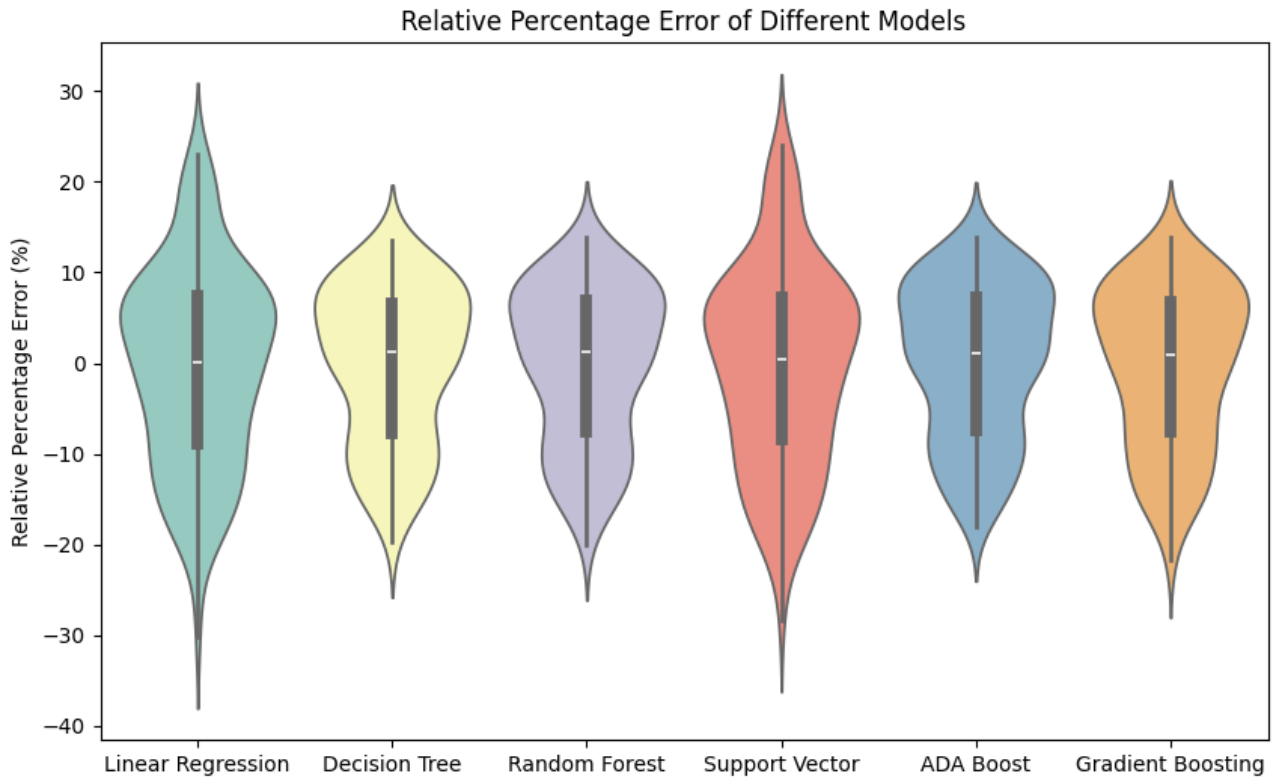


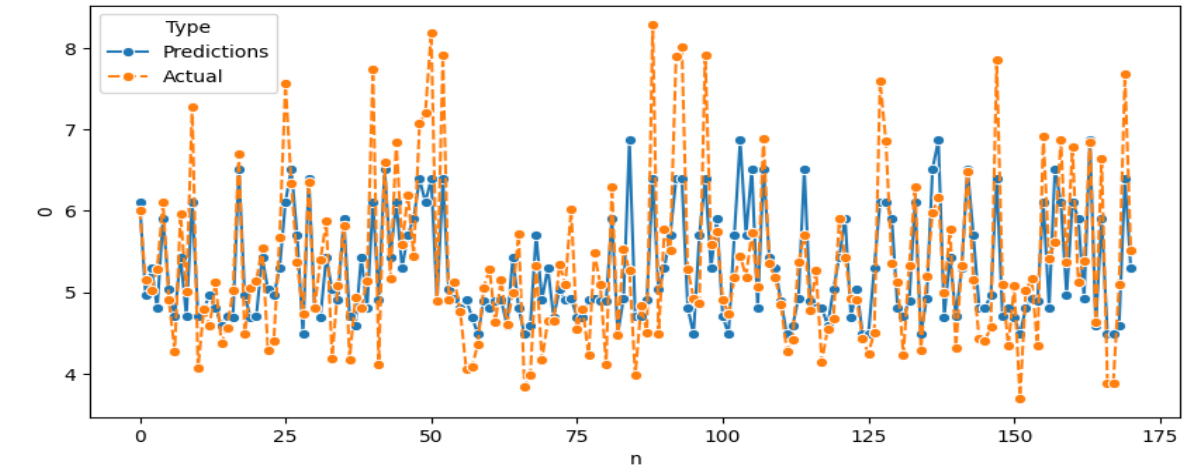
Figure 8: The violin plot illustrating the relative error percentages of different models.

Such deficiencies make MLR unsuitable for accurately predicting mixes with non-linear behaviours. The SVR presents a slightly wider distribution than the ensemble models but remains more stable than MLR. In contrast, models like RF and DT show more compact distributions, suggesting that they produce more consistent and accurate predictions with errors generally confined between -25% and 25%. These are better equipped to handle non-linearity. However, they can still face challenges. For instance, RF, with its narrower error distribution and a median close to zero, demonstrates consistent performance across various mixes, effectively managing both bias and variance. However, it may exhibit larger errors for mixes with unique characteristics not adequately represented in the training data. This issue is attributed to overfitting to the training set. Certain mixes, acting as outliers in the dataset, contribute significantly to larger errors in models like DT. Mixes with unusual or extreme properties skew the model's ability to generalise effectively. This sensitivity to outliers is reflected in the variability of error distributions across mixes. The GBR and AdaBoost models also demonstrate a relatively narrow error distribution, though slightly broader than that of DT. These exhibits increased errors for underrepresented or imbalanced mixes within the dataset. These models, while robust in general, perform less accurately for mixes lying outside common data patterns. Overall, the ensemble methods, particularly DT, appear to offer more reliable performance, as evidenced by their tighter error distributions, while the MLR shows greater susceptibility to prediction errors.

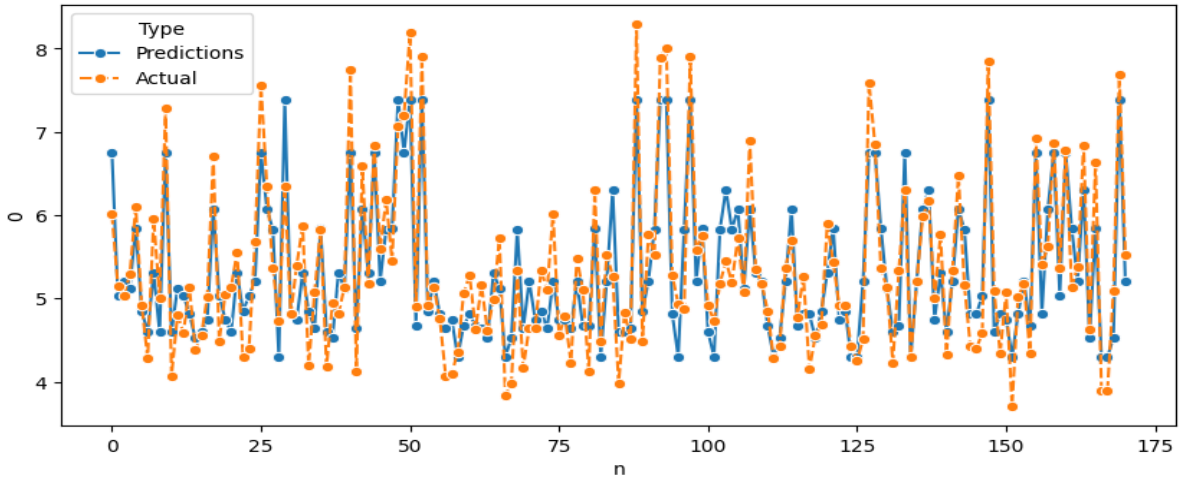
5.5 Performance Results

Mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), R squared (R^2) and cross-validation (CV) mean were calculated for all the models and are presented in **Table 5**. A greater R^2 value denotes a better match between the projected values of the analytical and gradient-boosting regressions, while lower values of MAE, MSE, RMSE and CV mean suggest a more accurate prediction result [28].

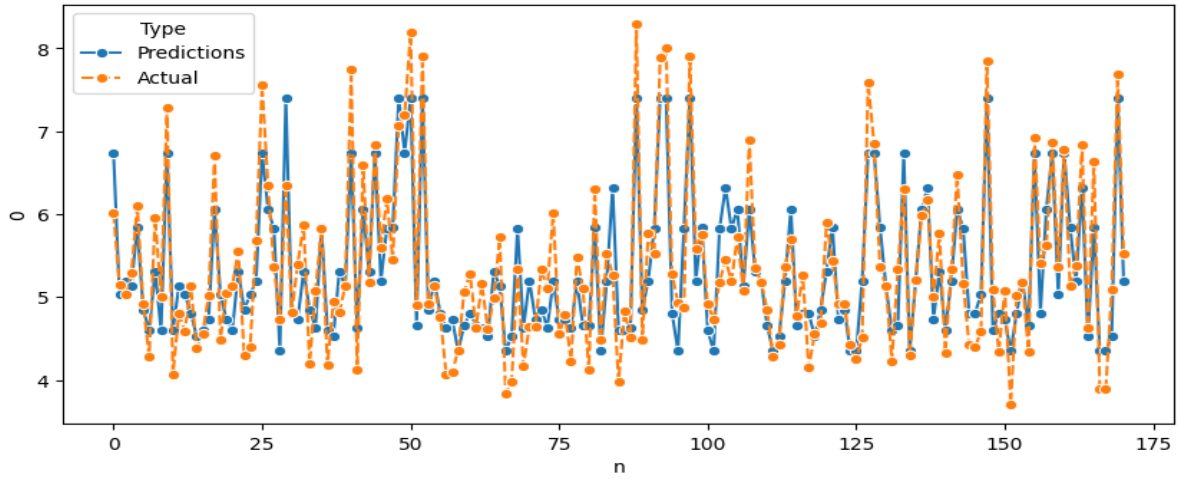
The variations between the projected outcomes and the observed outcomes of each sample for various models are depicted in **Figure 9**. Based on the coefficient of determination (R^2), all six models have strong predictive accuracy. Within the models, the DT model attained the



(a)

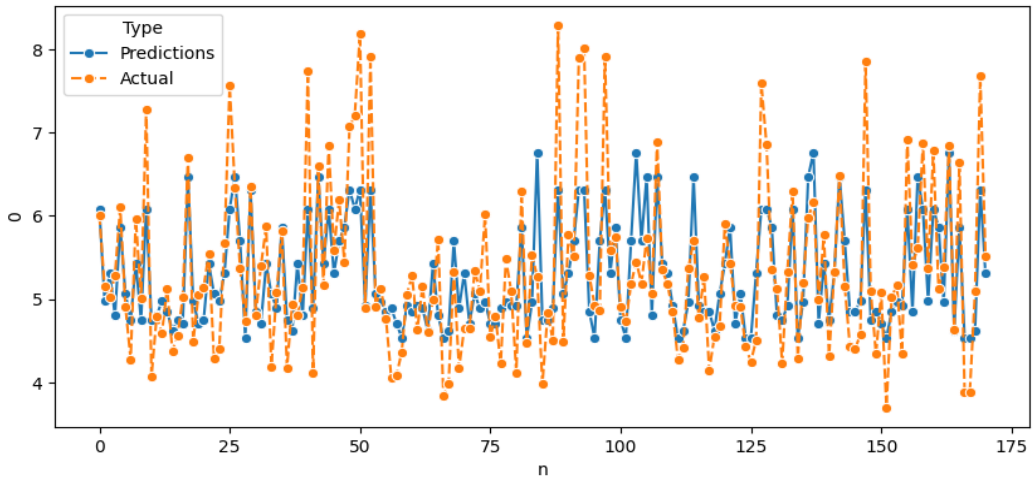


(b)

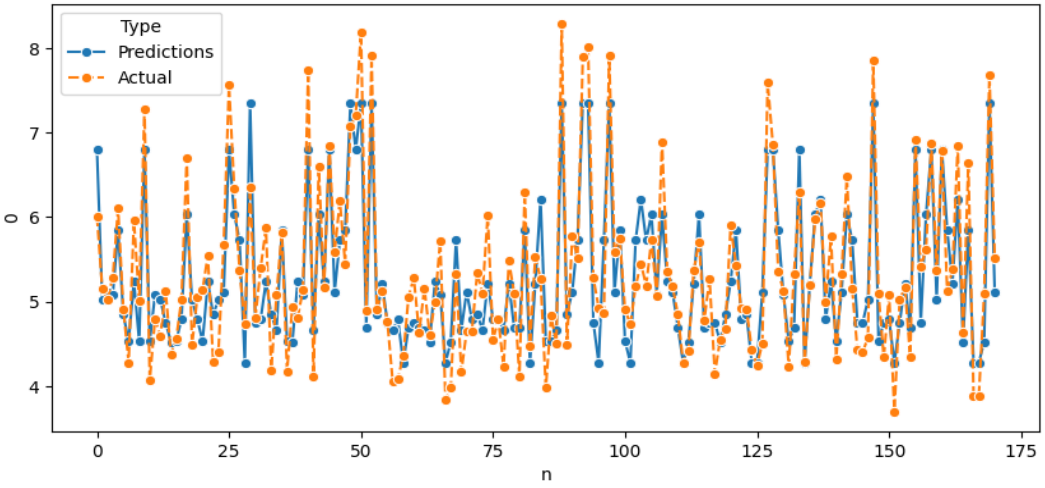


(c)

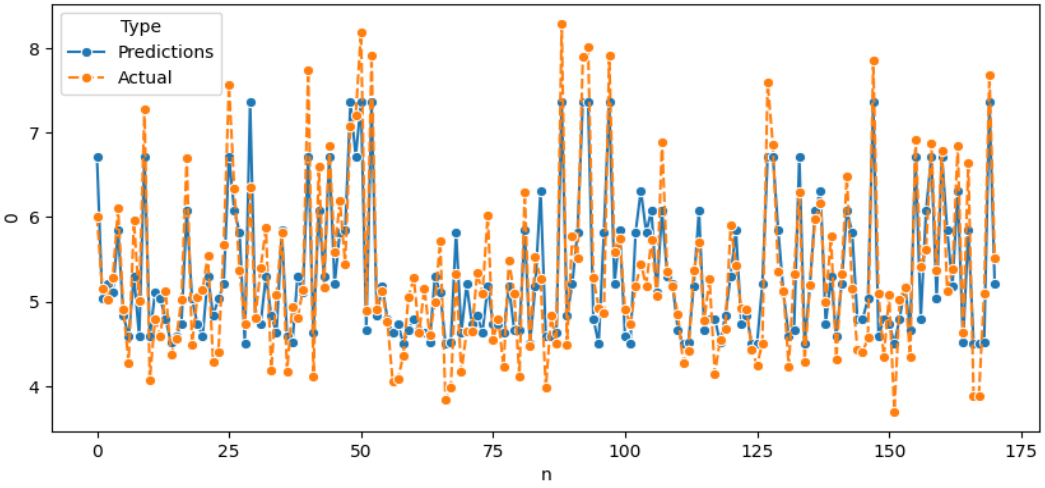
Figure 9: Comparative analysis between actual and predicted values: a) MLR, b)DT, c) RF, d) SVR, e) AdaBoost and f) GBR.



(d)



(e)



(f)

Continued **Figure 9:** Comparative analysis between actual and predicted values: a) MLR, b)DT, c) RF, d) SVR, e) AdaBoost and f) GBR.

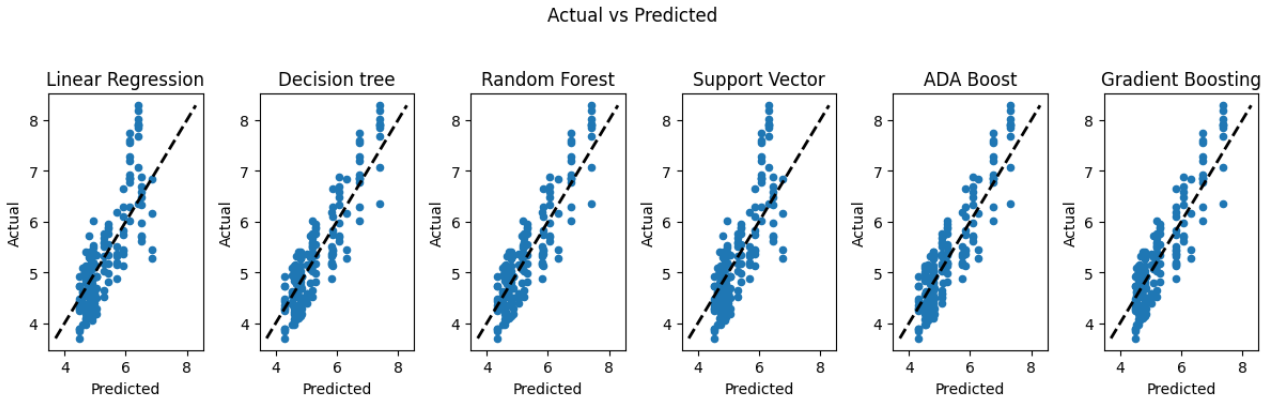


Figure 10: Correlation between expected and experimental values of STS for PQAC.

Table 5: Performance results of the predictive models.

Model	RMSE	MAE	MSE	R^2 score	CV mean
Linear Regression	0.629266	0.488546	0.395976	0.608906	0.445094
Decision Tree	0.452297	0.382320	0.204573	0.797950	0.713217
Random Forest	0.453657	0.383848	0.205805	0.796733	0.707697
Support Vector	0.636715	0.492057	0.405406	0.599593	0.454470
ADA Boost	0.454821	0.384150	0.206862	0.795688	0.714149
Gradient Boosting	0.459522	0.390104	0.211160	0.791443	0.71143811

greatest correlation value of 0.7979, with RF following at 0.7967, AdaBoost at 0.7956, GBR at 0.7914, MLR at 0.608 and SVR at 0.599. Generally, depending on a solitary measure for assessment may result in unreliability. Hence, five error metrics were computed for the predictions of each model, as presented in **Table 4**. By comparison to the other five models, it is evident that the DT model attains lower error metrics. The DT model has an MSE of 0.204, RMSE of 0.452, MAE of 0.382, and CV mean of 0.713. In this study, the best-performing models (DT, RF, AdaBoost) achieved R^2 values around 0.79, indicating that these models explain approximately 79% of the variability in STS. A high R^2 ensures that predictions of STS are closely aligned with actual experimental results. For pavement applications, this accuracy means engineers can confidently optimise mix designs without extensive physical testing and saving time and resources while ensuring material performance meets design requirements. The MSE values for the best-performing models ranged from 0.202 to 0.211. This low value reflects the squared difference between predicted and observed STS values. For a typical range of STS values (3.69–8.2 MPa), an MSE of 0.202 implies that the average squared prediction error is minor compared to the overall STS range. In practical terms, such precision ensures that pavement mixes designed using these models will perform as expected, avoiding under- or over-designed materials that could lead to failures or inefficiencies. MAE values for the top models (e.g., DT, RF, AdaBoost) were in the range of 0.382–0.384, indicating the average absolute deviation between predicted and actual STS. In real-world scenarios, an MAE of 0.38 MPa means that on average, the predicted STS differs from the true STS by less than 0.4 MPa. For pavement engineers, this level of accuracy is sufficient to ensure that materials meet safety standards without excessive conservatism, which could inflate costs. RMSE values for these models were 0.45–0.46, slightly higher than MAE but still low relative to the STS range. RMSE penalises larger errors more than smaller ones. A low RMSE confirms that significant deviations between predicted and observed values are minimal, which is crucial in pavement design to prevent localised weaknesses or overestimated capacities in the final structure. Models showed CV mean values around **0.71**, indicating consistency across different data subsets. High CV consistency suggests the model's generalisability to new mix designs beyond those tested in the study.

This robustness ensures that engineers can use these models for diverse pavement projects without retraining or overfitting concerns. The high predictive accuracy and low error rates of these models translate to reduced experimental testing and improved confidence in mix designs, especially in large-scale pavement projects. Additionally, the adoption of these predictive tools aligns with sustainable engineering practices by reducing material wastage and the carbon footprint associated with extensive experimental trials.

Figure 10 displays the values derived from the six models together with the experimental data. Using six ML models, it displays scatter graphs showing the expected STS value against the experimental value. In general, the plots display a positive correlation, indicating that the models are capable of capturing the underlying trends in the data. MLR and DT tend to exhibit a closer clustering of points around the line of equivalence, suggesting a more accurate fit to the data. Conversely, the RF and GBR, while exhibiting some non-linear patterns, show greater scatter, indicating that they may struggle with certain instances, potentially leading to higher prediction errors in those cases. The SVR and AdaBoost models also demonstrate moderate scatter but showcase the ability to capture complex relationships in subsets of the data. All of the established ML models show a strong connection between the observed and anticipated STS with R^2 >79%. It may be deduced from this figure that employing the DT model may be enough and provide satisfactory accuracy with 11 input variables for estimating the STS. Ultimately, it is evident that the DT model offers superior performance accuracy compared to other models. This work may help engineers choose suitable supervised learning models and parameters for the manufacturing of PQAC. DT excels in capturing non-linear relationships and interactions between input variables. It segments the data into smaller, more homogeneous groups through a hierarchical structure, which is particularly effective for datasets with complex patterns like those in PQAC mixes. The simplicity of DT allows it to make precise splits in regions of the dataset where relationships between variables are most apparent. The dataset in this study had discrete and continuous variables, with some non-linear dependencies between inputs (e.g., fibre proportions, nano-material dosages) and the output (STS). DT effectively identified these dependencies. RF reduces overfitting, a common limitation of standalone DTs, by averaging the outputs of multiple trees trained on random subsets of the data. This process increases robustness and generalisation. The ensemble nature of RF enables it to smooth out inconsistencies caused by noisy or sparse data points, yielding better predictions. The dataset featured variability in STS based on diverse input combinations. RF leveraged bootstrapping to create diverse training subsets, improving its ability to generalise and handle potential outliers or noise. It consistently predicted STS due to its capability to aggregate diverse 'opinions' from individual trees. AdaBoost combines multiple weak learners (often simple decision stumps) to create a strong learner, adaptively focusing on errors in previous iterations. It is particularly effective for imbalanced data or data with some challenging-to-fit regions. In the PQAC dataset, certain input combinations (e.g., specific fibre proportions and nano-materials) have been harder to predict accurately. AdaBoost adapted by assigning higher weights to these harder-to-predict instances in subsequent iterations improves overall accuracy. Its iterative nature allowed it to fine-tune predictions, particularly for the tail-end cases where errors were higher for other models. While GB also showed competitive performance, it is more prone to overfitting than RF due to its sequential approach to reducing residuals. SVR, with its reliance on kernel functions to model non-linear relationships, struggled due to the high-dimensional feature space and sparse relationships in some input combinations. MLR assumes linear relationships, which limited its effectiveness given the complex and non-linear nature of the input–output relationships in the PQAC dataset. The superior performance of DT, RF and AdaBoost can be attributed to their ability to capture non-linear relationships and interactions effectively and handle variability and noise in the data through ensemble averaging (RF, AdaBoost). Adapt to challenging regions of the dataset where traditional models like MLR and SVR failed to generalise. These characteristics made ensemble methods (especially RF and AdaBoost) well-suited to the complexity and variability of predicting STS in PQAC mixes.

5.6 Effect of the number of estimators

The ML metamodels are initially calibrated to optimise the primary hyperparameters. This is accomplished by adjusting the quantity of estimators and training the ML metamodels.

Figure 11 presents that the MAE, MSE, R^2, CV and RMSE decrease as the number of estimators increases, indicating that the model's accuracy improves with a higher number of estimators in RF. As the number of estimators increases, a notable decrease in MAE, MSE and RMSE is observed, indicating improved predictive accuracy and reduced error in the model's predictions. The MAE reaches a

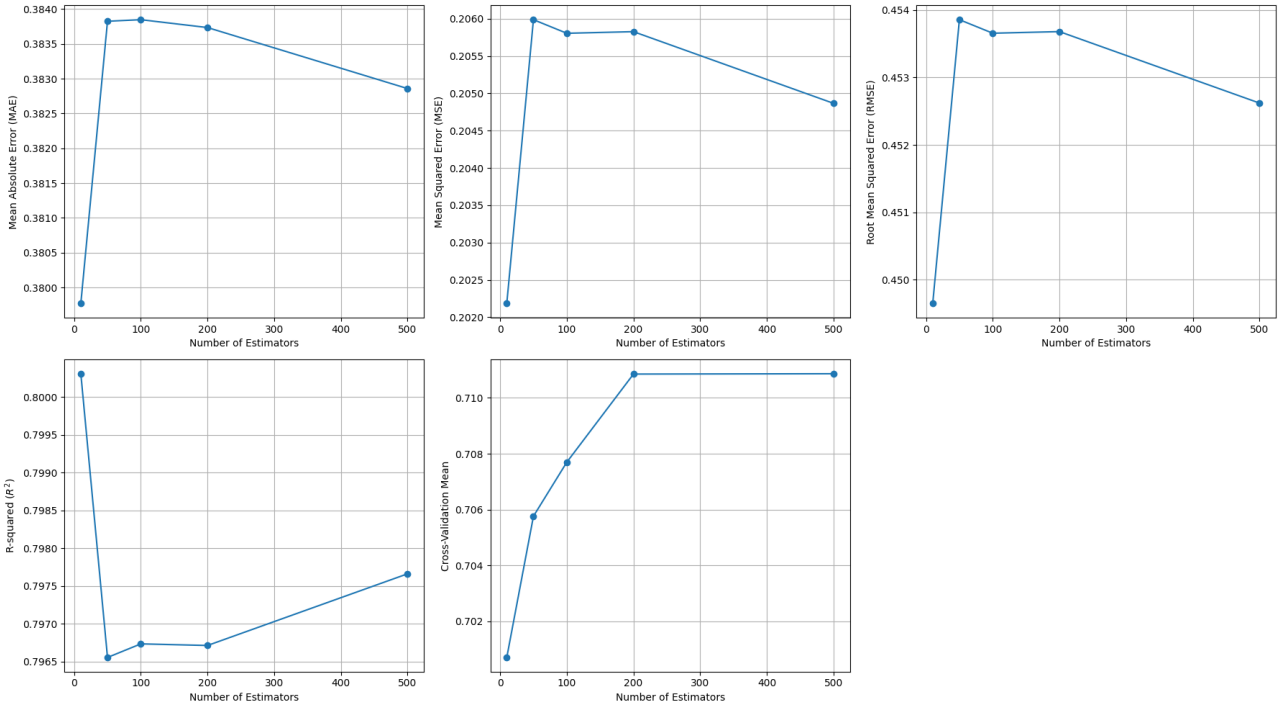


Figure 11: Effect of the number of estimators on RF's performance in terms of (a) MAE, (b) MSE, (c) RMSE, (d) R^2 and (e) cross-validation mean.

plateau of around 300–400 estimators, suggesting that further increasing the number of estimators beyond this point may not significantly improve the model's performance. Specifically, the MAE and MSE metrics show a marked decline up to a certain threshold, after which the improvements plateau, suggesting diminishing returns with additional estimators. Conversely, R^2 values exhibit an initial increase, reflecting enhanced model fit, before stabilising, which further supports the notion of optimal estimator count. The R^2 value, which represents the proportion of the variance in the target variable that is explained by the model, increases as the number of estimators increases. This indicates that the model's ability to capture the underlying patterns in the data improves with a higher number of estimators. The cross-validation mean, which represents the average performance of the model across multiple validation folds, also shows a similar trend, decreasing as the number of estimators increases and reaching a plateau of around 300–400 estimators. These findings underscore the importance of selecting an appropriate number of estimators to balance model complexity and performance, ultimately leading to more robust and generalisable predictions in RF.

Figure 12 presents the performance of the AdaBoost algorithm using various metrics, including MAE, Mean MSE, RMSE, R^2 and CV, as the number of estimators varied. The results, as depicted in the plots, demonstrate that all the error metrics significantly decrease as the number of estimators increases from 0 to approximately 50, after which they stabilise and show little to no further improvement. This indicates that the model's accuracy increases with more estimators up to a certain point, beyond which adding more estimators does not significantly enhance the model's performance. The R^2 metric shows a corresponding increase as the number of estimators rises, reaching a plateau around the same point, suggesting a limit to the explanatory power of additional estimators. Similarly, the CV initially shows an improvement with an increase in the number of estimators but stabilises after approximately 50 estimators, indicating consistent model performance beyond this point. These findings suggest that while adding estimators improves model performance initially, the benefit diminishes, highlighting the importance of selecting an optimal number of estimators to balance performance and computational efficiency.

Figure 13 presents the performance of the GBR algorithm by varying the number of estimators, with the results captured across multiple performance metrics: MAE, MSE, RMSE, R^2 and CV mean. As shown in the plots, the MAE, MSE and RMSE metrics all demonstrate a sharp decline as the number of estimators increases

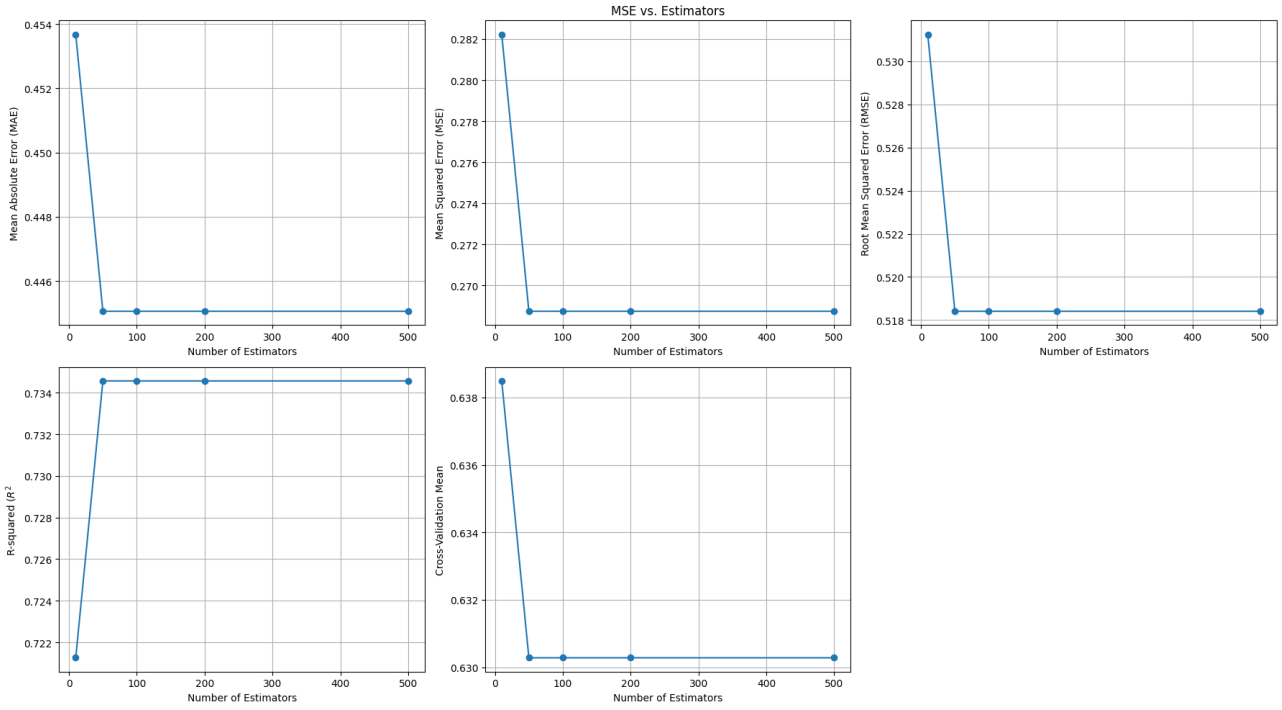


Figure 12: Effect of the number of estimators on AdaBoost's performance in terms of (a) MAE, (b) MSE, (c) RMSE, (d) R^2 and (e) cross-validation mean.

from 0 to approximately 100, indicating significant improvements in model accuracy. Beyond this point, the error metrics stabilise, suggesting diminishing returns from adding more estimators. The R^2 metric, which measures the proportion of variance explained by the model, shows a marked increase as the number of estimators increases, plateauing around 100 estimators, indicating that most of the model's predictive power is captured early on. Similarly, the cross-validation mean improves with the number of estimators, reaching its peak around 100 estimators and stabilising afterwards, reflecting consistent model performance across different folds of the data. These results suggest that while increasing the number of estimators initially leads to better model performance, there is a threshold beyond which additional estimators provide negligible benefits, emphasising the importance of choosing an optimal number of estimators for Gradient Boosting to balance accuracy and computational cost.

6 Practical Applications

Scalability is a critical consideration for the practical implementation of any new material in construction, particularly in the context of widespread use in infrastructure projects. The proposed PQAC mix leverages the advantages of nano-materials and fibres, which are increasingly available, although they may come at a higher cost compared to traditional additives. Notably, while PVAF is effective, PPF is more cost-effective and widely accessible, making them a promising choice for large-scale projects. The established manufacturing and supply chains for PPF facilitate their integration into large batching operations. One of the primary challenges in scaling this mix is achieving a uniform dispersion of nano-materials and fibres during the mixing process. To address this, advanced mixing techniques and chemical dispersants are required to prevent agglomeration, ensuring consistent quality throughout the batch. The use of industrial mixing equipment with high shear capabilities can effectively tackle these challenges, promoting uniformity and reliability in the final product. The mix design closely aligns with traditional concrete practices, which aids in its integration into existing construction workflows. Minor adjustments to current processes can facilitate the adoption of this innovative mix, making it more feasible for large-scale applications. However, it is important to note that the incorporation of nano-materials and fibres can reduce workability, as evidenced by lower CFV, such as 0.807 for nano-additive at 1.25% and 0.796 for PPF at 2.0%. This reduction in

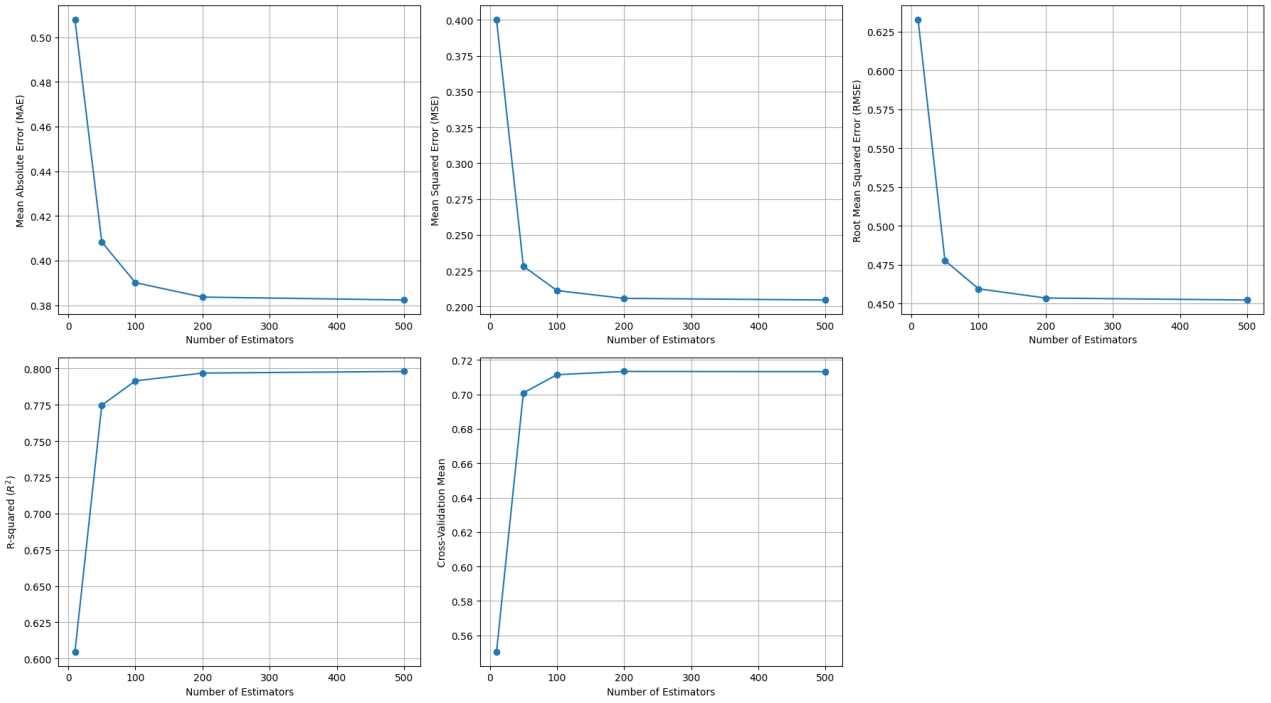


Figure 13: Effect of the number of estimators on GBR's performance in terms of (a) MAE, (b) MSE, (c) RMSE, (d) R^2 and (e) cross-validation mean.

workability may pose challenges for placement and compaction, particularly in thick pavement layers. Therefore, careful consideration of placement techniques and potential adjustments to the mix design may be necessary to ensure optimal performance in large-scale applications. From a performance perspective, the significant improvements in mechanical properties, such as a split tensile strength increase of up to 7.5 MPa with PVA fibres, indicate that this mix is well-suited for high-stress pavement applications. In addition to performance benefits, the environmental advantages of this mix are noteworthy. The replacement of OPC with alkali-activated materials, can lead to a reduction in carbon emissions by 40–80%. This aligns with sustainable construction practices and makes the mix particularly appealing for environmentally conscious infrastructure projects, which are increasingly prioritised by government and industry initiatives. Moreover, the use of predictive modelling techniques, such as DT, RF, GB, MLR, SVR and AdaBoost, allows for efficient optimisation of the mix. These models can reduce the need for extensive trial-and-error testing at scale by simulating and predicting performance across various site conditions and material availabilities. This capability not only enhances scalability but also expedites the deployment of optimised mixes tailored to diverse project requirements. In conclusion, the proposed PQAC mix demonstrates significant potential for scalability in large-scale pavement applications, particularly for high-performance and sustainable infrastructure. However, practical challenges related to material costs, workability and the need for advanced mixing and quality control must be addressed to fully realise this potential. Continued innovations in predictive modelling will further enhance the scalability and cost-effectiveness of this approach, paving the way for its widespread adoption in the construction industry.

7 Conclusion

This study presents a method for predicting STS utilising concrete with nano-additions and fibres through individual and ensemble learning models. These ML models exhibit strong efficacy in representing the intricate non-linear correlations between input and output parameters in predicting STS for PQAC. The present study enhances engineers' comprehension in research about the optimal selection of input variables and regressors for the accurate execution of ML models to forecast outputs, with high accuracy, minimising reliance on extensive experimental testing.

- By replacing OPC with alkali-activated materials, this research supports the global push for greener construction practices, reducing the carbon footprint of rigid pavements.

- The adoption of predictive modelling reduces material wastage and optimises resource utilisation, aligning with sustainable engineering goals.
- The enhanced tensile strength and crack resistance of the proposed fiber reinforced, nano-concrete mix contribute to longer-lasting and more resilient pavement structures.
- This innovation addresses the inherent brittleness of conventional AAC, making it more suitable for high-stress applications such as highways and heavy-duty pavements.
- The use of machine learning in mix optimisation introduces a paradigm shift, allowing engineers to predict performance under diverse conditions. This capability could lead to the development of smart construction materials and automated quality control systems.
- In accordance with additional error metrics and the co-relation coefficient between expected and actual values, the DT ensemble model demonstrates superior predictive ability and is endorsed as an effective approach for STS prediction.
- The co-relation plot shows that PVAF and PPF are highly co-related with the STS.
- It is clear from comparing the six models that DT, RF and AdaBoost have a strong command of strength forecasting, indicating their suitability for the redesign of PQAC.
- The DT model exhibited superior performance compared to the other machine learning models, while the RF model was shown to yield the second-best performance. Furthermore, the AdaBoost, GB and MLR exhibited commendable performance. The SVR model had the lowest level of prediction accuracy.
- The statistical indicators show that DT regression models outperformed AdaBoost, RF, MLR, SVR and GB approaches in terms of minimising the error difference between the targeted and projected values, with reduced errors.
- The low values of AdaBoost, DT and RF models for MAE (0.38, 0.382 and 0.383), MSE (0.202, 0.204 and 0.205) and RMSE (0.45, 0.452 and 0.453) verify that these machine learning models are more accurate than the others.
- The R^2 for the forecasted strength of MLR, DT, RF, SVR, AdaBoost and GB is observed to be 0.61, 0.797, 0.797, 0.6, 0.799 and 0.791, respectively.
- The effect of a number of estimators was studied on three different ML models, and the optimum number of estimators was found to be around 300–400 for RF, around 50 for AdaBoost and in the range of 50–100 for GB.

Overall, Artificial intelligence (AI) has significantly transformed predictive modelling in civil engineering, improving precision, simplifying construction procedures, and potentially replacing or augmenting conventional experimental approaches. The mix design can be extended to other infrastructure elements such as bridge decks, airport runways and industrial flooring, where tensile strength and durability are critical. Its adaptability to local materials and site-specific conditions makes it viable for diverse geographies and construction scenarios. ML algorithms like Adaboost, DT and RF are effective for analysing extensive information, making them useful for forecasting real strength characteristics under varying complexity settings. Ensemble learning methods, like the one used in this study, can consistently mature and adjust to fresh data, ensuring improved accuracy over time. Integrating explainable AI techniques like Shapley Additive Explanations (SHAP) can provide a deeper understanding of how input factors impact the projected result.

The results show that ensemble models of ML simulated more realistically than other algorithms, which can be effectively used at the building site to reduce laboratory usage and save time and costs. With further optimisation, the efficacy of these algorithms could be enhanced to yield more accurate predictions. The enhanced accuracy of the models signifies their significance in civil engineering applications, especially in forecasting concrete strength, as the experimental method is time consuming and labour intensive.

Furthermore, reliable and high-quality experimental data are crucial for the model's real performance. Future research should focus on augmenting the dataset with additional examples and using SHAP analysis to further examine the influence of input parameters on output.

References

[1] S. Suanmali and V. Ammarapala, "Maintenance Budget Planning: A Case Study for Rigid Pavement Maintenance System in Thailand," 2010, doi: 10.1109/icsssm.2010.5530086.

[2] E. Ghani, A. G. Goswami, and W. R. Kerr, "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *The Economic Journal*, 2015, doi: 10.1111/ecoj.12207.

[3] A. Das, E. Ghani, A. G. Goswami, W. Kerr, and R. Nanda, "Infrastructure and Finance: Evidence from India's GQ Highway

Network," *SSRN Electronic Journal*, 2019, doi: 10.2139/ssrn.3399447.

[4] R. Kumar and S. Sharma, "Perpetual Flexible Pavement vs. Rigid Pavement: An Economic and Environmental Cost Comparison," *IOP Conf Ser Earth Environ Sci*, 2022, doi: 10.1088/1755-1315/1084/1/012053.

[5] K. Sharma and M. K. Trivedi, "Statistical Analysis of Delay-Causing Factors in Indian Highway Construction Projects Under Hybrid Annuity Model," *Transportation Research Record Journal of the Transportation Research Board*, 2023, doi: 10.1177/03611981231161594.

[6] "Year End Review-2022 : Ministry of Road Transport and Highways."

[7] P. Solanki and S. Sharma, "Performance Rating of National Highways Based on Road User Perspective – A Case Study of Kalka–Shimla Highway," *IOP Conf Ser Earth Environ Sci*, 2022, doi: 10.1088/1755-1315/1084/1/012055.

[8] C. Aditya, D. Irawan, and S. Silviana, "Implementation of Marble Waste as Aggregate Material Rigid Pavement," *Eureka Physics and Engineering*, 2021, doi: 10.21303/2461-4262.2021.001932.

[9] T. Phoo-ngernkham, P. Chindaprasirt, V. Sata, S. Hanjitsuwan, and S. Hatanaka, "The effect of adding nano-SiO2 and nano-Al2O3 on properties of high calcium fly ash geopolymer cured," 2014, doi: 10.1016/J.MATDES.2013.09.049.

[10] H. Assaedi, F. U. A. Shaikh, and I. M. Low, "Effect of nano-clay on mechanical and thermal properties of geopolymer," *Journal of Asian Ceramic Societies*, 2016, doi: 10.1016/J.JASCER.2015.10.004.

[11] M. A. Uddin, M. T. Bashir, A. M. Khan, F. Alsharari, F. Farid, and R. Alrowais, "Effect of Silica Fume on Compressive Strength and Water Absorption of the Portland Cement–Silica Fume Blended Mortar," *Arab J Sci Eng*, vol. 49, no. 4, pp. 4803–4811, Apr. 2024, doi: 10.1007/s13369-023-08204-x.

[12] F. Matalkah, A. Ababneh, and R. Aqel, "Effects of nanomaterials on mechanical properties, durability characteristics and microstructural features of alkali-activated binders: A comprehensive review," *Constr Build Mater*, vol. 336, p. 127545, Jun. 2022, doi: 10.1016/j.conbuildmat.2022.127545.

[13] M. Rezaei Shahmirzadi, A. Gholampour, A. Kashani, and T. D. Ngo, "Geopolymer mortars for use in construction 3D printing: Effect of LSS, graphene oxide and nanoclay at different environmental conditions," *Constr Build Mater*, vol. 409, p. 133967, Dec. 2023, doi: 10.1016/j.conbuildmat.2023.133967.

[14] M. T. Bashir, S. Muhammad, M. J. Butt, M. Alzara, and M. S. El-kady, "Aspect ratio effect of multi-walled carbon nanotubes and carbon fibers on high-performance cement mortar matrices," *Innovative Infrastructure Solutions*, vol. 5, no. 2, Aug. 2020, doi: 10.1007/s41062-020-00290-2.

[15] "MoRTH Annual Report for the Year 2022-23."

[16] T. T. Nguyen and K. Dinh, "An artificial intelligence approach for concrete hardened property estimation," *Journal of Science and Technology in Civil Engineering (STCE) - NUCE*, vol. 14, no. 2, pp. 40–52, Apr. 2020, doi: 10.31814/stce.nuce2020-14(2)-04.

[17] F. Jamali, S. R. Mousavi, A. B. Peyma, and Y. Moodi, "Prediction of compressive strength of fiber-reinforced polymers-confined cylindrical concrete using artificial intelligence methods," *Journal of Reinforced Plastics and Composites*, vol. 41, no. 17–18, pp. 679–704, Sep. 2022, doi: 10.1177/07316844211068116.

[18] T. Li, J. Xiao, and A. Singh, "Strength index analysis of concrete with large size recycled aggregate based on back propagation neural network," *Advances in Structural Engineering*, vol. 25, no. 1, pp. 133–145, Jan. 2022, doi: 10.1177/13694332211046348.

[19] R. Haddad and N. Qarqaz, "Predicting NSMR–concrete bond strength using artificial neural networks: A comparative-analysis study," *Structural Concrete*, vol. 24, no. 5, pp. 6421–6435, Oct. 2023, doi: 10.1002/suco.202200630.

[20] D. Van Dao, S. H. Trinh, H.-B. Ly, and B. T. Pham, "Prediction of Compressive Strength of Geopolymer Concrete Using Entirely Steel Slag Aggregates: Novel Hybrid Artificial Intelligence Approaches," *Applied Sciences*, vol. 9, no. 6, p. 1113, Mar. 2019, doi: 10.3390/app9061113.

[21] Y. R. Wang, L. T. Q. Ngo, Y. F. Shih, Y. L. Lu, and Y. M. Chen, "Adapting ANNs in SONREB Test to Estimate Concrete Compressive Strength," *Key Eng Mater*, vol. 792, pp. 166–169, Dec. 2018, doi: 10.4028/www.scientific.net/KEM.792.166.

[22] Y. Wu and Y. Zhou, "Splitting tensile strength prediction of sustainable high-performance concrete using machine learning techniques," *Environmental Science and Pollution Research*, vol. 29, no. 59, pp. 89198–89209, Dec. 2022, doi: 10.1007/s11356-022-22048-2.

[23] S. Ray, M. M. Rahman, M. Haque, M. W. Hasan, and M. M. Alam, "Performance evaluation of SVR and GBM in predicting compressive and splitting tensile strength of concrete prepared with ceramic waste and nylon fiber," *Journal of King Saud University - Engineering Sciences*, vol. 35, no. 2, pp. 92–100, Feb. 2023, doi: 10.1016/j.jksues.2021.02.009.

[24] H. A. Shah *et al.*, "Application of Machine Learning Techniques for Predicting Compressive, Splitting Tensile, and Flexural Strengths of Concrete with Metakaolin," *Materials*, vol. 15, no. 15, p. 5435, Aug. 2022, doi: 10.3390/ma15155435.

[25] N. Moradi, M. H. Tavana, M. R. Habibi, M. Amiri, M. J. Moradi, and V. Farhangi, "Predicting the Compressive Strength of Concrete Containing Binary Supplementary Cementitious Material Using Machine Learning Approach," *Materials*, vol. 15, no. 15, p. 5336, Aug. 2022, doi: 10.3390/ma15155336.

[26] J. Huang, M. Zhou, H. Yuan, M. M. S. Sabri, and X. Li, "Prediction of the Compressive Strength for Cement-Based Materials with Metakaolin Based on the Hybrid Machine Learning Method," *Materials*, vol. 15, no. 10, p. 3500, May 2022, doi: 10.3390/ma15103500.

[27] A. M. R. Bulbul *et al.*, "In-Depth Analysis of Cement-Based Material Incorporating Metakaolin Using Individual and Ensemble Machine Learning Approaches," *Materials*, vol. 15, no. 21, p. 7764, Nov. 2022, doi: 10.3390/ma15217764.

[28] S. Marathe and A. P. Rodrigues, "Intelligent Models for Prediction of Compressive Strength of Geopolymer Pervious Concrete Hybridized with Agro-Industrial and Construction-Demolition Wastes," *Studia Geotechnica et Mechanica*, 2024, doi: 10.2478/sgem-2024-0020.

[29] A. Sheshadri, S. Marathe, M. Bettadapura Manjunath, A. Jayasimhan, and Ł. Sadowski, "Effective Utilization of Foundry Waste as Aggregates in Developing Eco-Friendly Alkali-Activated and Conventional Concretes for Sustainable Pavement Infrastructure," *Practice Periodical on Structural*

Design and Construction, vol. 29, no. 3, Aug. 2024, doi: 10.1061/PPSCFX.SCENG-1501.

[30] B. S. E. N. 15167-1:2006, *Ground granulated blast furnace slagfor use in concrete, mortar and grout*, vol. 3. The European Standard EN, 2006.

[31] I. 12089-1987, *Specification for granulated slag for the manufacture of Portland slag cement*. Bureau of Indian Standards, New Delhi, 1987.

[32] "IRC44".

[33] N. Delhi. Bureau of Indian Standards, "IS:383 (2016) Coarse and Fine Aggregate for Concrete — Specification. 1–21.," 1970.

[34] IS:14212, "Indian Standard Sodium and Potassium Silicates - Methods of Test," 1995, *Bureau of Indian Standards, New Delhi*.

[35] S. Marathe, I. R. Mithanthaya, B. M. Mithun, S. Shetty, and A. P. K, "Performance of slag-fly ash based alkali activated concrete for paver applications utilizing powdered waste glass as a binding ingredient," *International Journal of Pavement Research and Technology*, 2020, doi: 10.1007/s42947-020-0173-2.

[36] B. M. Mithun, "Performance of alkali activated slag concrete mixes incorporating copper slag as fine aggregate," 2017.

[37] IS-1199(Part2), "Fresh Concrete— Methods of Sampling, Testing and Analysis; Part 2 Determination of Consistency of Fresh Concrete," 2018, *Bureau of Indian Standards, New Delhi*.

[38] B. of Indian Standards, "IS 5816 (1999): Method of Test Splitting Tensile Strength of Concrete".

[39] N. Anil Kumar Reddy and K. Ramujee, "Comparative study on mechanical properties of fly ash & GGBFS based geopolymer concrete and OPC concrete using nano-alumina," *Mater Today Proc*, vol. 60, pp. 399–404, 2022, doi: 10.1016/j.matpr.2022.01.260.

[40] E. Mohseni, M. J. Kazemi, M. Koushkbaghi, B. Zehtab, and B. Behforouz, "Evaluation of mechanical and durability properties of fiber-reinforced lightweight geopolymer composites based on rice husk ash and nano-alumina," *Constr Build Mater*, vol. 209, pp. 532–540, Jun. 2019, doi: 10.1016/j.conbuildmat.2019.03.067.

[41] S. Shahbazpanahi, M. K. Tajara, R. H. Faraj, and A. Mosavi, "Studying the C–H Crystals and Mechanical Properties of Sustainable Concrete Containing Recycled Coarse Aggregate with Used Nano-Silica," *Crystals (Basel)*, vol. 11, no. 2, p. 122, Jan. 2021, doi: 10.3390/cryst11020122.

[42] K. P. Bautista-Gutierrez, A. L. Herrera-May, J. M. Santamaría-López, A. Honorato-Moreno, and S. A. Zamora-Castro, "Recent Progress in Nanomaterials for Modern Concrete Infrastructure: Advantages and Challenges," *Materials*, vol. 12, no. 21, p. 3548, Oct. 2019, doi: 10.3390/ma12213548.

[43] M. A. Faris *et al.*, "Properties of Hooked Steel Fibers Reinforced Alkali Activated Material Concrete," *MATEC Web of Conferences*, vol. 78, p. 01068, Oct. 2016, doi: 10.1051/matecconf/20167801068.

[44] H. A. Abdel-Gawwad, M. S. Mohammed, and T. Alomayri, "Single and dual effects of magnesia and alumina nano-particles on strength and drying shrinkage of alkali activated slag," *Constr Build Mater*, vol. 228, p. 116827, Dec. 2019, doi: 10.1016/j.conbuildmat.2019.116827.

[45] A. Hakamy, F. U. A. Shaikh, and I. M. Low, "Characteristics of nanoclay and calcined nanoclay-cement nanocomposites," *Composites Part B-engineering*, 2015, doi: 10.1016/J.COMPOSITESB.2015.03.074.

[46] J. J. Ekaputri, C. Fujiyama, N. Chijiwa, T. D. Ho, and H. T. Nguyen, "Improving Geopolymer Characteristics with Addition of Poly-Vinyl Alcohol (PVA) Fibers," *Civil Engineering Dimension*, vol. 23, no. 1, pp. 28–34, Apr. 2021, doi: 10.9744/ced.23.1.28-34.

[47] A. Ashrafian, M. J. Taheri Amiri, M. Rezaie-Balf, T. Ozbakkaloglu, and O. Lotfi-Omran, "Prediction of compressive strength and ultrasonic pulse velocity of fiber reinforced concrete incorporating nano silica using heuristic regression methods," *Constr Build Mater*, vol. 190, pp. 479–494, Nov. 2018, doi: 10.1016/j.conbuildmat.2018.09.047.

[48] M. F.M. Zain and S. M. Abd, "Multiple Regression Model for Compressive Strength Prediction of High Performance Concrete," *Journal of Applied Sciences*, vol. 9, no. 1, pp. 155–160, Dec. 2008, doi: 10.3923/jas.2009.155.160.

[49] G.-H. Kim, J.-M. Shin, S. Kim, and Y. Shin, "Comparison of School Building Construction Costs Estimation Methods Using Regression Analysis, Neural Network, and Support Vector Machine," *Journal of Building Construction and Planning Research*, vol. 01, no. 01, pp. 1–7, 2013, doi: 10.4236/jbcpr.2013.11001.

[50] A. F. Hayes, "Statistical Methods for Communication Science."

[51] S. V. Patil, K. Balakrishna Rao, and G. Nayak, "Prediction of recycled coarse aggregate concrete mechanical properties using multiple linear regression and artificial neural network," *Journal of Engineering, Design and Technology*, vol. 21, no. 6, pp. 1690–1709, Nov. 2023, doi: 10.1108/JEDT-07-2021-0373.

[52] H. Almuallim, S. Kaneda, and Y. Akiba, "3 DEVELOPMENT AND APPLICATIONS OF DECISION TREES," 2002.

[53] B. A. Young, A. Hall, L. Pilon, P. Gupta, and G. Sant, "Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods," *Cem Concr Res*, vol. 115, pp. 379–388, Jan. 2019, doi: 10.1016/j.cemconres.2018.09.006.

[54] T. Hastie, R. Tibshirani, and J. Friedman, "Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction."

[55] M. Timur Cihan, "Prediction of Concrete Compressive Strength and Slump by Machine Learning Methods," *Advances in Civil Engineering*, vol. 2019, no. 1, Jan. 2019, doi: 10.1155/2019/3069046.

[56] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.

[57] L. Breiman, "Random Forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[58] M. A. DeRousseau, E. Laftchiev, J. R. Kasprzyk, B. Rajagopalan, and W. V. Srubar, "A comparison of machine learning methods for predicting the compressive strength of field-placed concrete," *Constr Build Mater*, vol. 228, p. 116661, Dec. 2019, doi: 10.1016/j.conbuildmat.2019.08.042.

[59] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html

[60] H. Nguyen, T. Vu, T. P. Vo, and H.-T. Thai, "Efficient machine learning models for prediction of concrete strengths," *Constr*

Build Mater, vol. 266, p. 120950, Jan. 2021, doi: 10.1016/j.conbuildmat.2020.120950.
[61] R. Cook, J. Lapeyre, H. Ma, and A. Kumar, "Prediction of Compressive Strength of Concrete: Critical Comparison of Performance of a Hybrid Machine Learning Model with Standalone Models," *Journal of Materials in Civil Engineering*, vol. 31, no. 11, Nov. 2019, doi: 10.1061/(ASCE)MT.1943-5533.0002902.
[62] K. Yan and C. Shi, "Prediction of elastic modulus of normal and high strength concrete by support vector machine," *Constr Build Mater*, vol. 24, no. 8, pp. 1479–1485, Aug. 2010, doi: 10.1016/j.conbuildmat.2010.01.006.
[63] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer New York, 2000. doi: 10.1007/978-1-4757-3264-1.
[64] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans Intell Syst Technol*, vol. 2, Aug. 2007.
[65] F. Nelli, "Machine Learning with scikit-learn," in *Python Data Analytics*, Berkeley, CA: Apress, 2018, pp. 313–347. doi: 10.1007/978-1-4842-3913-1_8.
[66] D. P. Solomatine and D. L. Shrestha, "AdaBoost.RT: a boosting algorithm for regression problems," in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, IEEE, pp. 1163–1168. doi: 10.1109/IJCNN.2004.1380102.
[67] Y. Freund and R. E. Schapire, "A Short Introduction to Boosting," 1999. [Online]. Available: https://api.semanticscholar.org/CorpusID:9621074
[68] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm," in *International Conference on Machine Learning*, 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:1836349
[69] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 37–52. doi: 10.1007/978-3-642-41136-6_5.
[70] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *The Annals of Statistics*, vol. 29, no. 5, Oct. 2001, doi: 10.1214/aos/1013203451.
[71] R. Zemel and T. Pitassi, "A Gradient-Based Boosting Algorithm for Regression Problems," Aug. 2001.
[72] J. H. Friedman, "Stochastic gradient boosting," *Comput Stat Data Anal*, vol. 38, no. 4, pp. 367–378, Feb. 2002, doi: 10.1016/S0167-9473(01)00065-2.
[73] D. N. Lawley, "Tests of Significance for the Latent Roots of Covariance and Correlation Matrices," *Biometrika*, vol. 43, no. 1/2, p. 128, Jun. 1956, doi: 10.2307/2333586.
[74] J. H. Steiger, "Tests for comparing elements of a correlation matrix.," *Psychol Bull*, vol. 87, no. 2, pp. 245–251, Mar. 1980, doi: 10.1037/0033-2909.87.2.245.
[75] Q. Ma, J. Li, M. Aamer, and G. Huang, "Effect of Chinese Milk Vetch (Astragalus sinicus L.) and Rice Straw Incorporated in Paddy Soil on Greenhouse Gas Emission and Soil Properties," *Agronomy*, vol. 10, no. 5, p. 717, May 2020, doi: 10.3390/agronomy10050717.
[76] R. Alyousef *et al.*, "Machine learning-driven predictive models for compressive strength of steel fiber reinforced concrete subjected to high temperatures," *Case Studies in Construction Materials*, vol. 19, Dec. 2023, doi: 10.1016/j.cscm.2023.e02418.