ORIGINAL PAPER

Metaheuristic optimization of machine learning models for strength prediction of high-performance self-compacting alkali-activated slag concrete

Suraj Kumar Parhi[1] · Soumyaranjan Panda[1] · Saswat Dwibedy[1] · Saubhagya Kumar Panigrahi[1]

Abstract

The present study focuses on producing high-performance eco-efficient alternatives to conventional cement-based composites. The study is divided into two parts. The first part comprises of production of high-strength self-compacting alkali-activated slag concrete (SC-AASC) with GGBFS as a primary binder. The second part deals with the development of a prediction model to estimate the mechanical strength of developed concrete. In this study, to achieve high-performance SC-AASC, the alkali activator solution content varied from 220 to 190 kg/m^3, and the AAS/binder ratio varied between 0.47 and 0.36. The SP percentage fluctuated between 6 and 7%, while the additional water percentage was maintained between 21 and 24%. The approach used to obtain the high-performance SC-AASC was found to be competent as all the mix resulted in satisfactory performance for both fresh and hardened properties. For M45 graded SC-AASC, using 200 kg/m^3 of AAS with an AAS/binder ratio of 0.39 resulted in higher strength, while for M60 grade, 190 kg/m^3 of AAS with an AAS/binder ratio of 0.36 yielded stronger concrete. Additionally, a 6% SP and 24% extra water content enhanced workability for both M45 and M60 grade SC-AASC. A database of 135 observations was developed from the experimental study. The compressive strength and split tensile strength of SC-AASC were predicted using six machine-learning algorithms. The hyperparameters of all the models were optimized using the metaheuristic spotted hyena optimization technique. Optimized XGBoost outperformed other models scoring a higher R^2 of 0.97 and lower value of error parameters on both datasets. A comparison was drawn with previously published models to check the efficacy of the developed model. The Sobol and FAST global sensitivity analysis resulted in the AAS/binder ratio, AAS content, GGBFS content, and Curing days being most influential regarding the strength of SC-AASC.

Keywords High-strength · Self-compacting · Alkali-activated concrete · GGBFS · Machine learning · SHO · XGBoost · Sensitivity analysis

Abbreviations

GPC	Geopolymer concrete
SCGC	Self-compacting geopolymer concrete
SC-AASC	Self-compacting alkali-activated slag concrete
GGBFS	Ground granulated blast furnace slag
HS	High strength
HP	High performance
ML	Machine learning
AI	Artificial intelligence
AAS	Alkaline activator solution
SH	Sodium hydroxide
SS	Sodium silicate
SP	Superplasticizer
EW	Extra water
SF	Slump flow
VF	V-funnel
JR	J-Ring
T5	V-funnel at T5 minutes
CS	Compressive strength

✉ Saubhagya Kumar Panigrahi
skpanigrahi_ce@vssut.ac.in

Suraj Kumar Parhi
surajspeaks7@gmail.com

Soumyaranjan Panda
srpanda.ce@gmail.com

Saswat Dwibedy
saswatdwibedy@gmail.com

1 Department of Civil Engineering, VSSUT, Burla, India

STS   Split-tensile strength
EDA   Exploratory data analysis
SVR   Support vector regressor
RFR   Random forest regressor
BR   Bagging regressor
AB   AdaBoost
XGB   XGBoost
SHO   Spotted Hyena optimization

1 Introduction

Portland cement-based concrete is the second most-used material after water to accomplish rapid construction throughout the world (Feng et al. 2020). Portland cement manufacture produces large emissions in addition to requiring a lot of energy. In other words, the production of one tonne of cement produces roughly one tonne of carbon emissions (Alsalman et al. 2021). These vast carbon and other hazardous substances emission into the atmosphere creates various environmental concerns (Oliveira et al. 2019, 2020). To check on these environmental issues created by Portland cement production, various remedial measures have been explored by several researchers (Ahmed et al. 2022b, d, 2023b; Qaidi et al. 2022b, c; Smirnova et al. 2021; Unis Ahmed et al. 2022; Pradhan et al. 2022d, 2023a, b) and among those use of supplementary materials, use of alkali-activated materials, and geopolymer concrete are notable. Geopolymers are inorganic polymeric materials produced using some alumino-silicate materials as precursors and some reagents, which were discovered in the 1970s (Davidovits 1976). The world has seen enormous research since the 1990s on the development of geopolymer concrete (GPC) using various sources of precursors such as natural (clays, kaolin, metakaolin), industrial wastes (fly ash, ground granulated blast furnace slag or GGBFS, red mud), and agricultural wastes (sugarcane ash, rice husk ash, etc.) and various alkaline as well as acid solutions as chemical regents (Ahmed et al. 2023a, c; Das et al. 2023; Dwibedy and Panigrahi 2023; Pradhan et al. 2023a, b, 2024, Qaidi et al. 2022a, c; Unis Ahmed et al. 2023). To overcome the difficulty in the workmanship of GPC and alkali-activated concrete, a new form of GPC has also been explored which has a self-compacting nature and is termed Self-compacting geopolymer concrete and SC-AASC (Pradhan et al. 2022c). The SC-AASC can fill the dense reinforced formwork without segregation (EFNARC 2002). SC-AASC usage ensures a reduction in construction cost, labor, and energy consumption so it is highly beneficiary for modern-day infrastructural development. The GPC and alkali-activated concrete have

been established as a substitute for Portland cement concrete because of their mechanical and durability behaviors (Pradhan et al. 2022a). However, it is not in use in practical construction fields yet. The primary reason is the variations in the properties of starting materials as it creates difficulty in controlling the GPC properties. Prediction of strength characteristics from its constituting material has been practiced so that wastage of time, materials, energy, and repetition of tests can be avoided (Ahmed et al. 2021, 2022a, b, c, d, e, f; Faraj et al. 2022a). The factors governing the properties are vast (Pradhan et al. 2022b) and simple linear regressions can't provide an acceptable relation for the prediction of compressive strength. Machine Learning and Deep Learning techniques are extensively employed for tasks such as database classification and regression analysis. This is primarily due to their capacity to grasp intricate patterns within data (Ongsulee 2017). Various researchers have applied machine learning to predict concrete properties (Ahmed et al. 2022c, d, 2023b; Faraj et al. 2022a, b). Nhat-Duc (2023) used a deep learning approach to predict the compressive strength of GGBFS-based concrete. Zhang et al. (2023) used hybrid algorithms for compressive strength prediction of ultra-great workability concrete. Hu (2023) used a novel hybrid SVR model for compressive strength evaluation. Recently, various metaheuristic techniques have been used for hyperparameter optimization of machine learning models for concrete properties prediction (Dash et al. 2023; Parhi and Panigrahi 2023; Parhi and Patro 2023a, b; Singh et al. 2023). Compressive strength is the most important criterion in any type of concrete (Parhi et al. 2023). Compressive strength prediction of ecofriendly plastic sand paver blocks using gene expression programming (GEP) and multi-expression programming (MEP) has been done by various researchers (Iftikhar et al. 2023a, b). Chen et al. (2023a) employed GEP, decision tree (DT), multilayer perceptron neural network (MLPNN), and support vector machine (SVM) for predicting the reduction of compressive strength due to acid attack in egg-shell and glass waste incorporated cement-based composite. Zou et al. (2023) used GEP and MEP for slump and strength prediction of alkali-derived concrete. Utilization of GEP was further found in the strength prediction of waste-derived self-compacting concrete by Chen et al. (2023b). Alsharari et al. (2023) used a decision tree (DT), Adaboost (ADB), and bagging regressor (BR) approach for strength estimation of concrete at elevated temperatures. Random forest (RF) and GEP were used by Qureshi et al. (2023) for compressive strength prediction of preplaced aggregate concrete.

This research aims to meet stringent performance requirements in modern construction infrastructure by producing sustainable, green, and noise-free concrete. To facilitate these unique characteristics for a wide range of applications a method was proposed to develop HS-self-compacting

Table 1 Oxide composition of GGBFS (% wt)

CaO	SiO$_2$	Al$_2$O$_3$	MgO	TiO$_2$	MnO	Others
36.8	34.05	17.8	9.1	0.8	0.07	1–2

alkali activated slag concrete in this manuscript. Critical sustainable factors contributing to the workability and strength of concrete have been identified and varied to obtain high-performance concrete. Further different machine learning techniques were used to optimize the mix proportion for developed high-performance concrete by predicting the strength properties of its constituent materials.

1.1 Research significance and essence of the manuscript

The manuscript presents a pioneering study in the field of concrete technology and machine learning, focusing on the development of High-Performance Self-Compacting alkali-activated slag concrete. This research holds significant importance and contributes to both the concrete industry and the domain of predictive modeling. The development of High-Performance SC-AASC represents an innovative approach to sustainable construction materials. This has the potential to reduce carbon emissions and environmental impact, addressing critical concerns in the construction industry. From a base, M25 SC-AASC mix design, M45, and M60 high strength SC-AASC were prepared by exploring the variation of critical strength and workability factors. Understanding how these factors impact the concrete's performance is essential for optimizing its use in real-world construction applications. Metaheuristic Spotted Hyena-optimized Machine Learning models were used for predicting the strength properties of SC-AASC. The manuscript adopts advanced computational techniques to enhance the accuracy of strength predictions, potentially revolutionizing quality control and design processes in the construction industry. The research also employs global sensitivity analysis methods, such as Sobol and FAST, to identify the parameters that influence the strength of SC-AASC. This analysis provides invaluable insights into the underlying mechanisms governing the material's performance. The research aligns with global efforts to promote sustainability in construction. The development of a more environmentally friendly concrete mix and the use of advanced modeling techniques contribute to reducing the ecological footprint of construction projects. The manuscript bridges the gap between materials science, construction engineering, and data science, showcasing the potential of interdisciplinary research to address complex challenges in the construction sector.

2 Materials and methodology

2.1 Material collection

For the preparation of high-strength workable concrete, Ground Granulated Blast furnace slag (GGBFS) was used as the sole binder. GGBFS of specific gravity 2.8 was supplied by TATA Steel, Jajpur, Odisha. Table 1 represents the oxide composition of GGBFS. For the preparation of an alkaline activator solution (AAS), a combination of sodium hydroxide (SH) and sodium silicate (SS) of specific gravity 2.13 and 1.4 was used. Locally available natural coarse aggregate and fine aggregate were collected. SIKA VISCOFLOW 4005NS of specific gravity 1.11 was used as a superplasticizer (SP) for the production of higher-strength workable concrete. To maintain workability clean tap water was used as extra water (EW).

2.2 Mix proportion

In this study, the mix design procedure for self-compacting concrete as per IS 10262:2019 has been adopted. From the previous studies to prepare higher strength workable concrete, some parametric values have been identified like molarity of SH as 12 M, AAS ratio (SS/SH) as 2.5 (Nuruddin et al. 2011; Pradhan et al. 2022a, b, c) with a variation of percentage of superplasticizer as 6% and 7%, (Jithendra and Elavenil 2019; Memon et al. 2012) and extra water as 21% and 24% (Muraleedharan and Nadir 2021; Patel and Shah 2018). During the progress of high strength, mix designs were prepared from standard base SC-AASC i.e., M25 to M45 to M60 with the variation of two principal strength contributing parameters like AAS content and AAS/Binder ratio. Nine M25 mix designs were prepared with successive reductions of AAS content i.e., 220–210 kg/m^3, and AAS/Binder ratio i.e., 0.47–0.43. After getting workable M25 concrete, similarly, nine numbers of M45 and M60 mix designs were prepared with regular reduction of AAS content i.e., 210–200 kg/m^3 and 200–190 kg/m^3 along with a decrement of AAS/Binder ratio i.e., 0.41–0.39 and 0.37–0.36 respectively. Figure 1 represents the mix-design progress from M25 to M60. The detailed mix proportions of M25, M45, and M60 are shown in Tables 2, 3, and 4.
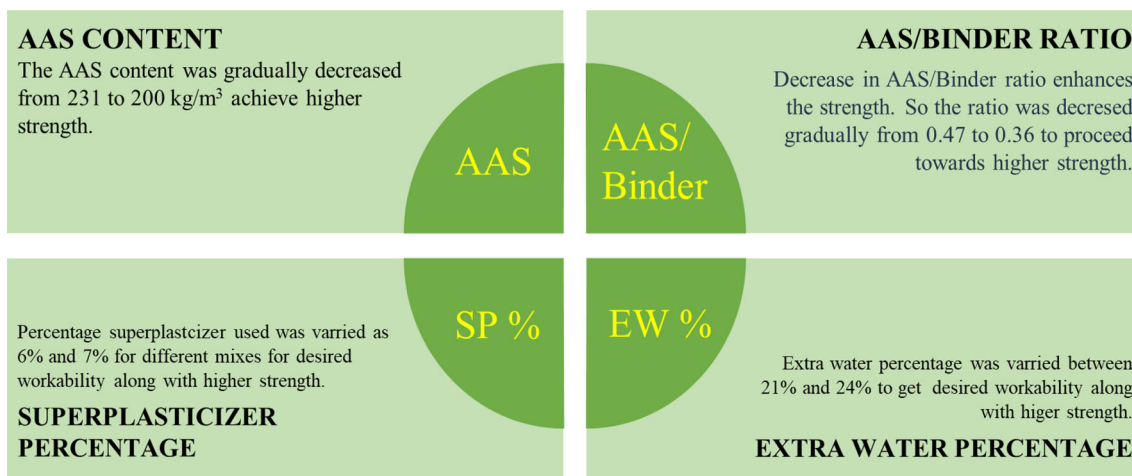
AAS CONTENT
The AAS content was gradually decreased from 231 to 200 kg/m³ achieve higher strength.

AAS

AAS/ Binder

AAS/BINDER RATIO
Decrease in AAS/Binder ratio enhances the strength. So the ratio was decresed gradually from 0.47 to 0.36 to proceed towards higher strength.

Percentage superplasticizer used was varried as 6% and 7% for different mixes for desired workability along with higher strength.
SUPERPLASTICIZER PERCENTAGE

SP %

EW %

Extra water percentage was varried between 21% and 24% to get desired workability along with higer strength.
EXTRA WATER PERCENTAGE

Fig. 1 Mix design progression from M25 to M60

2.3 Preparation, casting, and curing

For the preparation of fresh SC-AASC, materials like GGBFS, and aggregates were weighted properly. Aggregates were collected in a saturated surface dry condition. A dry mix of Ground Granulated Blast Furnace Slag (GGBFS) and these aggregates was prepared by mixing them for 2–3 min. A Sodium Hydroxide (SH) solution was prepared 24 h before casting. Two hours before casting, the Sodium Silicate (SS) solution was added to the SH solution to create an Alkali-Activated Slag (AAS) solution. The AAS solution was then added to the dry mix of GGBFS and aggregates. Extra water and a superplasticizer were added to the mixture concerning the mix's workability. The materials were thoroughly mixed for 3–5 min to ensure uniformity and homogeneity. Freshly prepared SC-AASC was tested for workability. SC-AASC was cast into cubes with sides measuring 150 mm and cylinders with a diameter of 150 mm and a height of 300 mm. The specimens were cured at 14-, 28-, and 56-day ambient temperature for mechanical property study. The detailed procedure for the production of SC-AASC is represented in Fig. 2.

2.4 Test methodology

The freshly prepared concrete was tested using four fresh property tests. The Slump Flow Test was used to evaluate the workability and flow characteristics of self-compacting AASC by measuring the spread of the concrete when it flows freely. The V-Funnel Test assesses the filling ability of SC-AASC by measuring the time it takes to flow through a V-shaped funnel, indicating its flow properties. The J-Ring Test measures the ability of SC-AASC to pass through obstructions, assessing its flowability in the presence of reinforcement. The V-Funnel at T5 Minutes Test accesses the

resistance to segregation during pouring. The compressive Strength Test was used to determine the maximum load SC-AASC can withstand under axial compression, indicating its overall strength, and split-tensile strength was used to evaluate its resistance to cracking. Figure 3 represents all the fresh and hardened tests conducted along with casted specimens.

2.5 Experimental test results

The freshly prepared concrete was tested for filling ability test (slump flow, V funnel test), Passing ability test (J ring), and resistance to segregation test (V-funnel at $T_{5minute}$) of workability as per EFNARC guidelines. After 14, 28, and 56 days of ambient curing, the specimens were tested for mechanical properties of compressive and split tensile tests. To avoid any misleading results, for each mix, three specimens were tested. The results of fresh and mechanical properties are shown in Table 5.

It is well observable that as the mix grade changes from M25 to M60, the slump flow generally decreases, indicating that higher-strength mixes are less flowable but still within an acceptable range for self-compacting concrete. The v-funnel test evaluates filling ability, Similar to slump flow results higher mix grades show slightly longer times (in seconds), indicating reduced filling ability with increasing strength. The J-ring test measures the ability of freshly prepared concrete to pass through the tight obstruction. It is a field test so can be performed onsite. The results show a slight increase in J-Ring values with higher mix grades, suggesting better resistance to segregation in stronger mixes. V-funnel test at T5 minutes provides insights into segregation resistance, stability, and viscosity. Similar to other workability test results this test also indicates reduced flowability over time for higher mix grades but within acceptable ranges. All the

Table 2 Details of mix proportions of M25 SC-AASC

MIX ID	AAS (kg/m^3)	AAS/binder	GGBFS (kg/m^3)	Fine aggregate (kg/m^3)	Coarse aggregate (kg/m^3)	NaOH (kg/m^3)	Na$_2$SiO$_3$ (kg/m^3)	SP %	Extra water %
M25SC1	220	0.47	468.09	1065.5	673.26	62.86	157.14	7	24
M25SC2	220	0.47	468.09	1065.5	673.26	62.86	157.14	7	21
M25SC3	220	0.47	468.09	1065.5	683.89	62.86	157.14	6	24
M25SC4	215	0.45	477.78	1043.00	693.90	61.43	153.57	7	24
M25SC5	215	0.45	477.78	1043.00	693.90	61.43	153.57	7	21
M25SC6	215	0.45	477.78	1043.00	704.75	61.43	153.57	6	24
M25SC7	210	0.43	488.37	1008.4	725.69	60.00	150.00	7	24
M25SC8	210	0.43	488.37	1008.4	725.69	60.00	150.00	7	21
M25SC9	210	0.43	488.37	1008.4	736.78	60.00	150.00	6	24

Table 3 Details of mix proportions of M45 SC-AASC

MIX ID	AAS (kg/m^3)	AAS/binder	GGBFS (kg/m^3)	Fine aggregate (kg/m^3)	Coarse aggregate (kg/m^3)	NaOH (kg/m^3)	Na$_2$SiO$_3$ (kg/m^3)	SP %	Extra water %
M45SC1	210	0.41	512.20	957.40	750.65	60.00	150.00	7	24
M45SC2	210	0.41	512.20	957.40	750.65	60.00	150.00	7	21
M45SC3	210	0.41	512.20	957.40	762.29	60.00	150.00	6	24
M45SC4	205	0.39	525.64	884.79	817.71	58.57	146.43	7	24
M45SC5	205	0.39	525.64	884.79	817.71	58.57	146.43	7	21
M45SC6	205	0.39	525.64	884.79	829.65	58.57	146.43	6	24
M45SC7	200	0.39	512.82	895.73	829.59	57.14	142.86	7	24
M45SC8	200	0.39	512.82	895.73	829.59	57.14	142.86	7	21
M45SC9	200	0.39	512.82	895.73	841.24	57.14	142.86	6	24

Table 4 Details of mix proportions of M60 SC-AASC

MIX ID	AAS (kg/m^3)	AAS/binder	GGBFS (kg/m^3)	Fine aggregate (kg/m^3)	Coarse aggregate (kg/m^3)	NaOH (kg/m^3)	Na$_2$SiO$_3$ (kg/m^3)	SP %	Extra water %
M60SC1	200	0.37	540.54	739.46	956.69	57.14	142.86	7	24
M60SC2	200	0.37	540.54	739.46	956.69	57.14	142.86	7	21
M60SC3	200	0.37	540.54	739.46	968.97	57.14	142.86	6	24
M60SC4	195	0.36	541.67	724.44	979.25	55.71	139.29	7	24
M60SC5	195	0.36	541.67	724.44	979.25	55.71	139.29	7	21
M60SC6	195	0.36	541.67	724.44	991.56	55.71	139.29	6	24
M60SC7	190	0.36	527.78	829.63	897.09	54.29	135.71	7	24
M60SC8	190	0.36	527.78	829.63	897.09	54.29	135.71	7	21
M60SC9	190	0.36	527.78	829.63	909.00	54.29	135.71	6	24

mix grades successfully pass the flowability and filling ability criteria for self-compacting concrete. The compressive strength and split-tensile strength results exhibited increasing mechanical strength with an increase in curing period and mix grades. The strong chemical bond formed due to the chemical activation of slag contributes to higher compressive strength compared to normal Portland-based concrete. Alkali-activated slag sometimes contains unreacted slag particles which makes the microstructure porous and prone to cracking (Sakulich et al. 2010). Due to this reason, a low value of split-tensile strength can be visualized in some mixed proportions. Also, self-compacting alkali-activated slag may require longer curing periods to achieve its optimal strength. In comparison, Portland cement concrete often gains strength more rapidly in the early stages. The results indicate satisfactory performance of SC-AASC for all the mix grades, hence the approach used to obtain the high-performance SC-AASC was found to be competent.

3 Database and exploratory data analysis

3.1 Database for model development

An adequate and dependable database holds equal significance alongside effective computational techniques. When simulating a model, the database must encompass a diverse array of input data. Variability in the experimental database arises from both epistemic and aleatoric errors, contributing to high variance. To serve the goals of research endeavors, a coherent database with 135 observations was generated through the Python interface of Google Colab. Refer to Table 6 for a summary of the descriptive statistics about the SC-AASC database, which was employed to construct models.

Machine learning (ML) approaches are used to extract data information, unknown patterns, and the relationship between the dataset. ML has two kinds of techniques for prediction and modeling; one is a single separate model and another is known as an ensemble (Chou and Pham 2013). For the prediction of strength characteristics of M25, M45, and M60 SC-AASC; 12 inputs were given i.e., along with all the constituent parameters that produce SC-AASC and the important fresh property results which defy the essential characteristic of SC-AASC i.e., flowability was also taken as input data while compressive strength and split tensile strength were taken as output data. Using both single and ensemble ML algorithms compressive and split tensile strength were predicted.
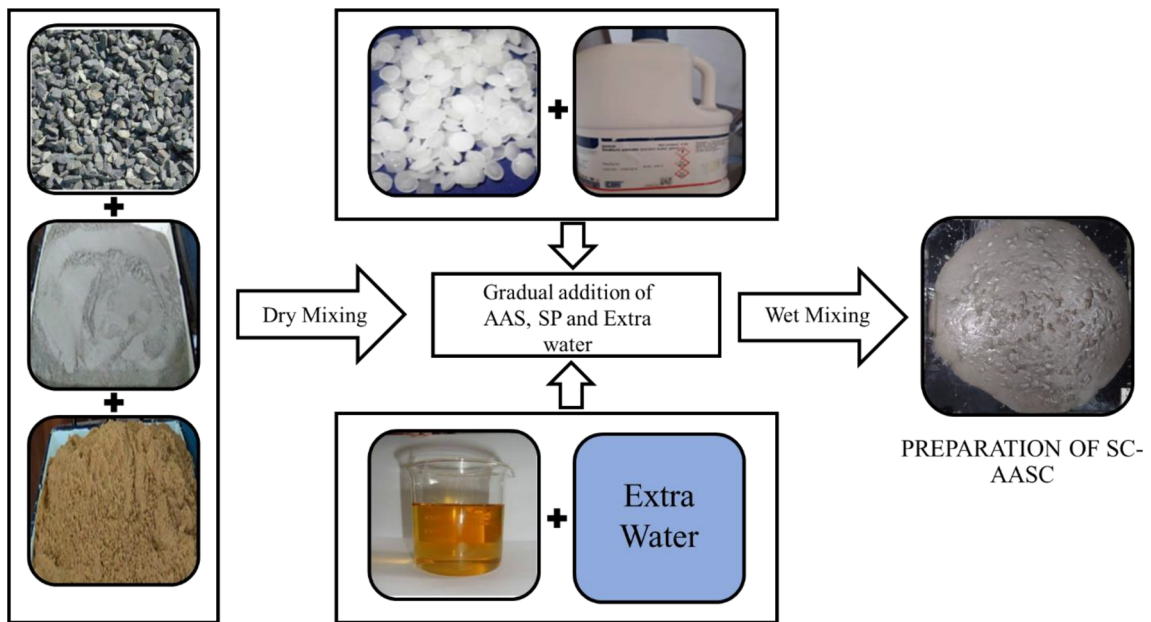
Fig. 2 Production procedure of self-compacting alkali-activated slag concrete

Fig. 3 Fresh and hardened tests conducted on specimens

Table 5 Fresh and mechanical results of all the mixes

MIX ID	Fresh property				Mechanical property					
	Slump flow (SF) in mm	V-funnel (VF) in sec	J-ring (JR) in mm	V-funnel at $T_{5minutes}$ (T5) in s	Compressive strength (CS) in MPa			Split tensile strength (STS) in MPa		
Days					14	28	56	14	28	56
M25SC1	752	8.16	1.75	10.21	26.65	32.70	39.1	1.52	1.68	1.82
M25SC2	740	8.34	1.78	10.25	27.23	33.82	40.24	1.6	1.75	1.9
M25SC3	746	8.32	1.83	10.28	25.44	31.8	38.13	1.5	1.67	1.82
M25SC4	742	8.46	1.85	10.51	29.54	36.23	43.53	1.72	1.89	2.07
M25SC5	733	8.61	1.96	10.58	30.98	38.16	45.79	1.78	1.97	2.19
M25SC6	735	8.67	2.01	10.64	28.76	35.87	42.96	1.75	1.92	2.06
M25SC7	729	8.83	2.06	10.89	32.28	40.28	48.18	1.9	2.12	2.31
M25SC8	718	8.96	2.2	10.92	33.29	41.00	49.23	2.06	2.28	2.48
M25SC9	723	9.02	2.16	10.97	32.97	40.09	48.03	1.91	2.08	2.26
M45SC1	730	8.87	2.09	10.98	38.84	48.59	57.4	2.65	2.92	3.17
M45SC2	721	9.04	2.18	11.07	40.93	50.78	60.75	2.8	3.07	3.33
M45SC3	724	9.01	2.22	11.12	42.71	52.16	63	2.92	3.21	3.49
M45SC4	710	9.17	2.26	11.25	43.94	54.08	64.74	3.01	3.32	3.6
M45SC5	702	9.31	2.35	11.3	45.96	56.72	67.93	2.93	3.24	3.52
M45SC6	700	9.29	2.33	11.36	45.72	56.02	67.09	2.98	3.28	3.56
M45SC7	691	9.45	2.42	11.47	46.23	56.94	68.16	3.12	3.43	3.72
M45SC8	685	9.52	2.47	11.45	46.55	57.12	68.54	3.05	3.35	3.64
M45SC9	682	9.51	2.5	11.58	47.63	58.32	69.87	3.07	3.38	3.67
M60SC1	695	9.56	2.52	11.54	51.91	63.71	72.69	3.19	3.51	3.81
M60SC2	688	9.64	2.57	11.55	52.43	64.58	73.59	3.32	3.66	3.95
M60SC3	683	9.57	2.64	11.60	53.47	65.56	73.4	3.16	3.59	3.91
M60SC4	673	9.68	2.66	11.79	54.07	66.29	74.08	3.23	3.65	3.96
M60SC5	664	9.76	2.59	11.87	54.56	66.95	74.16	3.46	3.79	4.12
M60SC6	667	9.81	2.74	12.02	55.18	68.47	74.95	3.55	3.91	4.25
M60SC7	661	9.94	2.85	12.18	55.38	67.78	73.21	3.59	3.95	4.28
M60SC8	654	10.02	2.98	11.97	56.04	68.89	75.68	3.56	3.92	4.26
M60SC9	656	9.97	2.92	11.89	56.35	69.14	75.93	3.62	3.98	4.32

3.2 Exploratory data analysis

Data analysis is an important step before proceeding with the creation of a model. Exploratory data analysis (EDA) helps in the initial investigation of data i.e., to identify patterns, get meaningful insights, check inconsistency and irregularity of data, and to check the hypothesis and parametric features with the assistance of statistics and graphs (Morgenthaler 2009). EDA was done before proceeding into ML model creation. Data analysis is an important step before proceeding with the creation of a model. Exploratory data analysis (EDA) helps in the initial investigation of data i.e., to identify patterns, get meaningful insights, check inconsistency and irregularity of data, and to check the hypothesis and parametric features with

the assistance of statistics and graphs (Morgenthaler 2009). EDA was done before proceeding into ML model creation. The dataset containing missing values and outliers significantly decreases the accuracy of the models. Outlier mining is a data analysis process associated with the detection and removal of outliers in a dataset (Angiulli and Pizzuti 2005). In a dataset outliers are data points situated at an uncommon distance or having significant deviation from well-structured data (Wang et al. 2019). Commutation of system, mechanical and software faults, and human and instrumental errors give emergence to outliers (Hodge and Austin 2004). Outliers can be viewed by plotting the attributes in a box plot which shows the distribution and range of data for each attribute (Petrovskiy 2003). If an outlier is present for any attribute that can

Table 6 Summary of experimental database used for modeling

Parameters	Mean	Std	Min	25%	50%	75%	Max
GGBFS (binder)	510.543333	25.475696	468.09	488.37	512.82	527.78	541.67
AAS	205.000000	9.185587	190.00	200.00	205.00	210.00	220.00
AAS/binder	0.403333	0.037650	0.36	0.37	0.39	0.43	0.47
CA	817.615926	106.844761	673.26	725.69	817.71	909.00	991.56
FA	905.367778	118.306723	724.44	829.63	895.73	1008.37	1065.53
SP	34.036296	2.959617	28.09	31.67	34.19	36.79	37.92
Water	117.598148	9.109216	98.30	112.34	117.21	126.15	130.00
SF	703.962963	29.999352	654.00	682.00	702.00	730.00	752.00
VF	9.201111	0.539266	8.16	8.83	9.29	9.64	10.02
T5	11.230617	0.561835	10.21	10.89	11.30	11.60	12.18
JR	2.328889	0.347387	1.75	2.06	2.33	2.59	2.98
Age	32.666667	17.569861	14.00	14.00	28.00	56.00	56.00
CS	52.183457	14.353802	25.44	40.28	52.43	65.56	75.93
STS	2.973580	0.829535	1.50	2.08	3.19	3.60	4.32
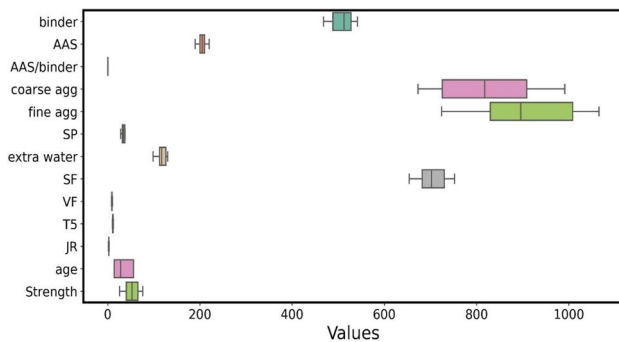
Fig. 4 Box plot of the compressive model dataset
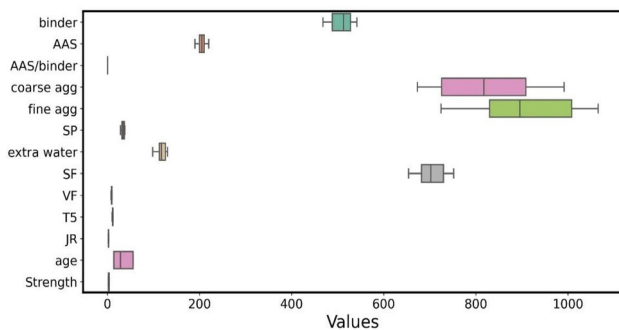
Fig. 5 Box plot of the split-tensile model dataset

be viewed with a dot mark in a box plot. Figures 4 and 5 show both datasets of compressive and split-tensile having no outliers.

A density plot gives a fair assessment of the outlook of the data and if it is normally distributed or not so that the applied algorithm will not require any assumption about the organization of data which may cause consumption of time while running the model (Chen 2017). KDE is a probability density function that assists in the evaluation of the density plot (Kim and Park 2014). A steady curve was drawn at every data point and all the curves were summated to make a single smooth curve (Terrell and Scott 1992). The distribution of the data in the density plot should be like a Gaussian bell-shaped. If the data contains irregularity, then shape and distribution will vary. Figure 6 shows the perfect Gaussian distribution of all the input attributes of the dataset. The plot having the region with a higher peak resembles the density of maximum data points. Figure 7 shows the correlation matrix of all the features in the dataset. The Pearson correlation matrix indicates the relationship between all the attributes is linear or nonlinear. As it can be visualized from the figure the strength of SC-AASC shows a non-linear to moderately linear relationship with other parameters.

4 Machine learning models and model evaluation parameters

Most of the studies as documented in Sect. 1 employ ANN, GEP, Tree-based algorithms, boosting, and bagging regressors for strength prediction. In this study, for a comprehensive comparison between different types of AI-based models, six ML algorithms were employed. The hyperparameters of these models were optimized using a metaheuristic optimization algorithm. The ML algorithms include modified linear regression (LASSO), support vector regressor, bagging, and boosting regressors. These are very efficient ML algorithms that were used for strength prediction in this study.
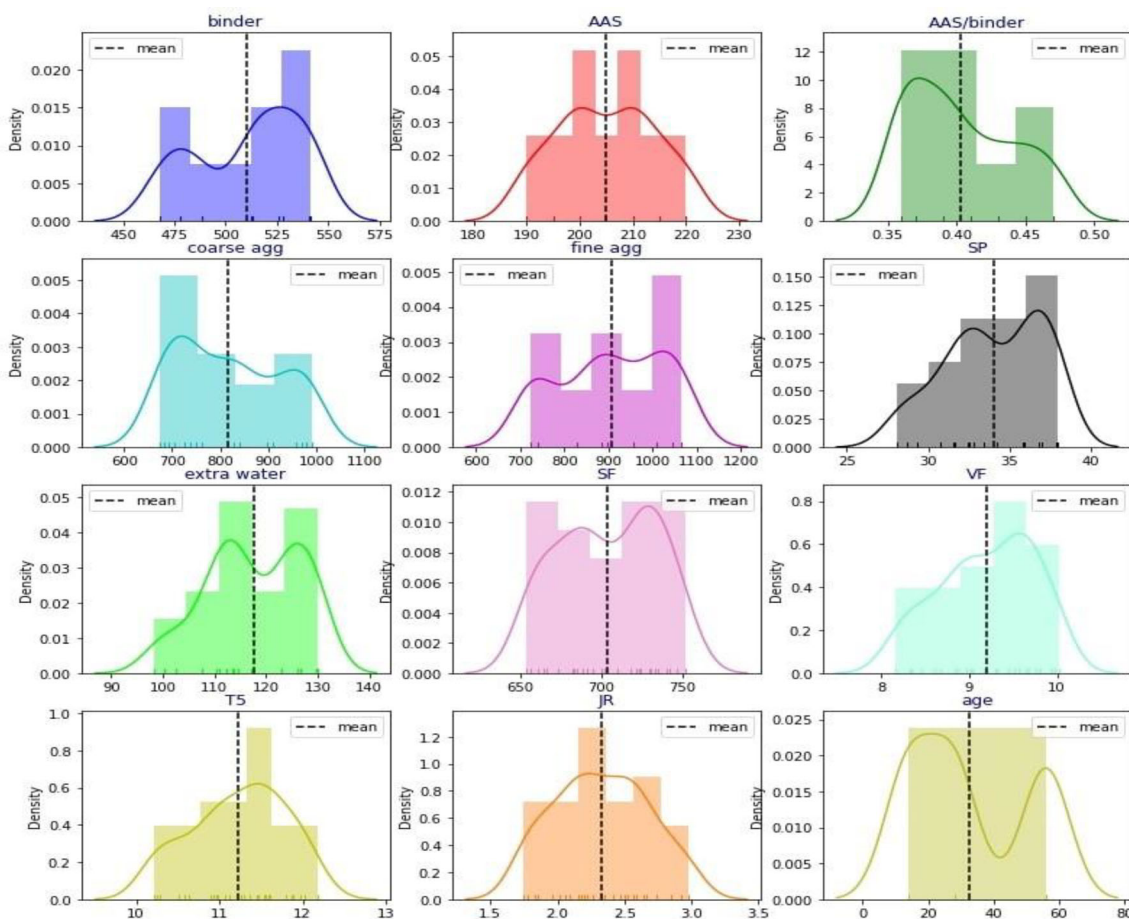
Fig. 6 Density-based distribution plot of the input dataset

4.1 LASSO regressor

The Least Absolute Shrinkage and Selection Operator (LASSO) is a modified linear regression method used as an ML algorithm for modeling purposes. It uses the shrinkage method to penalize the model. The values are shrunk towards a central point as the mean and the model is penalized for the sum of absolute values of the weights (Kang et al. 2021). LASSO uses regularisation and selection of variables or features of the dataset for enhancement of model accuracy i.e., introducing penalty factor to constraint the co-efficient and when the co-efficient is shrunk towards the mean of zero less important features are eliminated before the model proceeds to train (Mangalathu and Jeon 2018). The mathematical equation of the LASSO regressor is given below.

$$L(\beta) = \sum_{i=1}^{n} \left(y_i - \sum_j x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|. \tag{1}$$

Penalization factor i.e., the amount of shrinkage denoted by λ. An increase of λ has a direct proportion effect on

the bias but is indirectly proportioned toward variance. The LASSO regression method reduces the multi-collinearity of the dataset and simplifies the model. It also prevents overfitting.

4.2 Support vector regressor (SVR)

A support vector machine (SVM) is one of the popular ML algorithms used for classification and regression problems proposed by Cortes and Vapnik (1995). The fundamental thought behind SVM is that it takes a set of input data and using a non-linear mapping function transforms the data into high-dimensional feature space (Boser et al. 1992). SVM Regression referred to as SVR designated using the sparse solution, kernels, decision boundaries, and several support vectors; can predict the value for a non-linear database. SVR gives mobility to interpret the amount of acceptance of error in the model and find a suitable line i.e., a hyperplane in a higher dimension to fit the data. SVR trains utilize a symmetrical loss function which similar to SVM penalizes high and
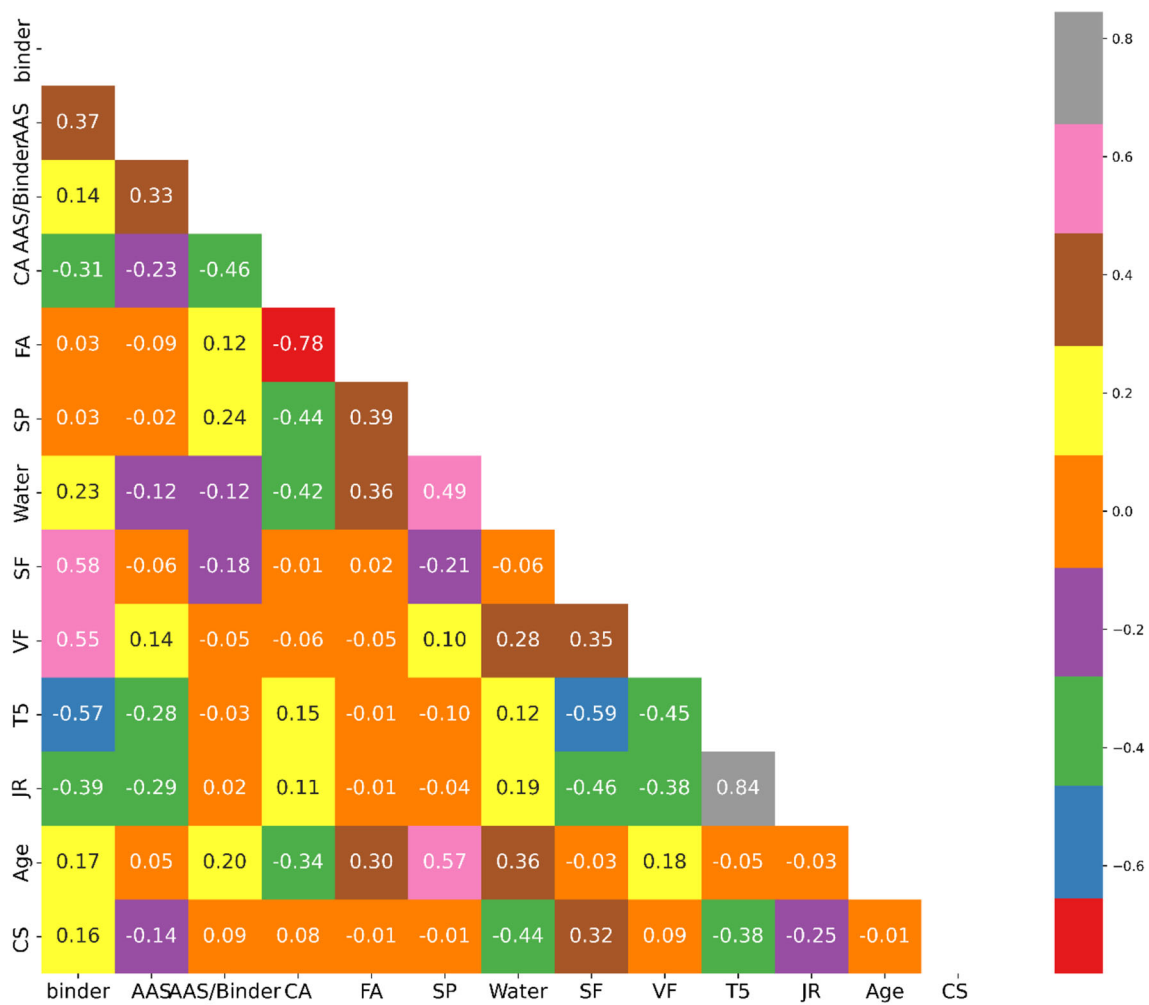
Fig. 7 Co-relation matrix

low misestimates (Awad and Khanna 2015). SVR's regression equation in the feature space can be approximated as

$$Z(a, w) = (w.\phi(a) + c). \tag{2}$$

The weight vector defined by w and $\phi(a)$ is the feature function. By the incorporation of Lagrangian multiplier β and β^*, using kernel function $K(x_i, x_j)$, and applying Karush–Kuhn–Tucker conditions (Parveen et al. 2016); the SVR function can be obtained as:

$$Z(a, \beta_i, \beta_i^*) = \sum_{i=1}^{N_{sv}} (\beta_i - \beta_i^*) K(a, a_i) + c. \tag{3}$$

4.3 Random Forest Regressor (RFR)

The Random Forest (RF) algorithm stands out as a versatile and user-friendly machine-learning technique. It excels in accuracy, often eliminating the need for intricate hyperparameter tuning. RF operates by constructing an ensemble of decision trees (DTs) referred to as a "Forest." These trees are generated from diverse dataset samples and trained using the bagging technique. The final output is obtained by amalgamating the outputs of these trees through averaging, yielding a dependable and precise prediction (Dou et al. 2019). Each tree within the RF is constructed independently, utilizing distinct subsets of attributes. This approach leads to a reduction in the dimensionality of the feature space. While developing these decision trees, RF doesn't prioritize the search for critical factors in node splitting; instead, it seeks the optimal feature among a subset of features. Consequently, it generates smaller DTs, thereby mitigating the risk of overfitting (Géron 2019).

$$\hat{y} = \frac{1}{B} \sum_{d=1}^{B} f_d(x). \tag{4}$$

4.4 Bagging regressor (BR)

A bagging regressor or bootstrap aggregating regressor is a bagging ensemble ML meta-estimator used for modeling that fits or resamples the original training dataset with replacement (Sutton 2005). Some data may be occurring multiple times while others left out in the resampled dataset. It is based on a powerful statistical method known as a bootstrap for estimating quantity from a data sample. As previously discussed in Sect. 5.2 this method decreases the variance of the black-box estimator by initiating randomization and then creating an ensemble. If for a given training set $B = \{(x_1, y_1) \dots (x_n, y_n)\}$ the bagging regressor converts it to quasi-replica training set T by replacement and the final regression equation can be given as

$$\overline{f}(x) = \sum_{i=1}^{T} f_i(x). \tag{5}$$

4.5 AdaBoost regressor (AB)

Adaptive Boosting, also referred to as AdaBoost, represents a meta-learning approach within the realm of machine learning (Wu et al. 2008). This pioneering boosting algorithm marked a significant advancement in enhancing the effectiveness of binary classification tasks, originally denoted as AdaBoost.M1. The fundamental principle behind AdaBoost involves its capacity to transform the errors made by weaker classifiers into strengths (Schapire 2013), thereby mitigating bias in supervised learning contexts. While contrasting with the Random Forest (RF) model, wherein Decision Trees (DTs) lack a predetermined depth and encompass an initial node connected to numerous leaf nodes, AdaBoost employs concise DTs featuring only one level or a solitary split. Such DTs with a solitary split are commonly referred to as 'Decision stumps' or simply 'stumps'. The algorithm's workflow entails constructing a model that initially attributes equal weights to all data points within a dataset. Subsequently, these weights are reassigned to favor weaker learners in each iteration, thus granting them higher importance in the subsequent rounds. This iterative training process persists until errors are minimized, leading to heightened accuracy.

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right). \tag{6}$$

4.6 XGBoost regressor (XGB)

XGBoost, short for Extreme Gradient Boosting, represents the pinnacle of ensemble machine learning algorithms, boasting both remarkable speed and formidable power. This algorithm provides a structured approach to harnessing the combined predictive prowess of multiple individual learners. Its unparalleled capacity for cache optimization renders it the swiftest performer in program execution. At its core, XGBoost embodies a distributed gradient-boosted decision tree (GBDT) framework for machine learning. Widely recognized as the vanguard of meta-learning algorithms, it commands the field of regression, classification, and ranking quandaries (Dong et al. 2020). Drawing upon the foundations of supervised machine learning, XGBoost leverages the principles of the classification and regression tree (CART), seamlessly integrating gradient boosting, stochastic gradient boosting, and regularized gradient boosting algorithms. These techniques orchestrate the execution of decision trees at each data point sequentially, while simultaneously training them in parallel. This strategy imparts more weight to weaker learners, which are subsequently fed into other decision trees. The cumulative output of these trees harmonizes to yield the ultimate result (Duan et al. 2021). XGBooster introduces a pivotal augmentation by integrating the predicted residuals and errors from preceding models via the gradient descent algorithm, thereby minimizing loss. The mathematical formulation encapsulating the essence of XGBoost is as follows:

$$\hat{y} = \sum_{k=1}^{k} f_k(x_i), \ f_k \varepsilon F. \tag{7}$$

4.7 Spotted Hyena optimization (SHO)

Dhiman and Kumar (2017) introduced a novel bio-inspired metaheuristic optimization technique known as the Spotted Hyena Optimizer (SHO). This approach draws its inspiration from the social behaviors observed in spotted hyenas, particularly in their hunting activities. The SHO algorithm is designed to replicate these natural interactions and utilize them for optimization tasks. To validate its effectiveness, the SHO algorithm was put to the test using practical engineering design problems that involve more than four variables (Dhiman and Kumar 2019). The results obtained from these experiments showcase the superior performance of SHO compared to other competing algorithms, particularly when dealing with real-world problems. The SHO algorithm revolves around four core components: searching, encircling, hunting, and attacking. These components correspond to the steps spotted hyenas take during their hunting process. The primary objective is to emulate the behavior

of the hyenas when they work together to catch their prey. In the algorithm, the targeted prey is akin to the ideal route for encircling the prey, and other search agents or hyenas adjust their positions based on this optimal solution. This optimal solution serves as a guide for the rest of the group, enabling efficient movement towards achieving the optimization goals. The initial focus of the SHO algorithm lies in executing the hunting strategy. This strategy involves generating an ideal response cluster for the most proficient search agent and adjusting the positions of the other search agents. This process is elucidated through a set of subsequent equations. Similar to a pack of spotted hyenas compelled to charge towards their prey, the algorithm orchestrates this movement. The efficacy of an algorithm's exploration capability is intricately linked to its search mechanism. This capability is harnessed by utilizing variables denoted as "E" and random numbers falling either above or below 1. Within the scope of this study, our efforts were directed at optimizing the hyperparameters inherent to all the proposed machine learning methodologies. These optimizations were geared towards enhancing the predictive accuracy for both compressive strength and split-tensile strength.

4.8 Model evaluation parameters

Cross-validation (CV) stands as a crucial technique to enhance the reliability of model outcomes (Xiong et al. 2020). The core concept behind CV involves partitioning the training set into two distinct subsets: one serves as the training dataset, facilitating model training, while the other functions as the validation or testing dataset, employed to assess the model's performance. The model's performance on the validation dataset acts as an estimate of its generalization error, becoming a pivotal metric in the process of model selection. In this specific study, a preference is given to k-fold cross-validation. This methodology involves splitting the dataset into k non-overlapping subsets of equal size. In each iteration, one of these subsets is utilized for validation, while the remaining $k − 1$ subsets are combined for training. This sequence is reiterated k times, each time utilizing a different subset for validation, allowing for extensive utilization of the dataset for training. The choice of k typically resides in the range of 3–10, with this study adopting a k value of 10. For the creation and assessment of prediction models utilizing machine learning algorithms, the k-fold cross-validation technique is applied in conjunction with the curated SC-AASC database. The evaluation of model accuracy is executed through efficiency metrics like the coefficient of determination (R^2) and adjusted R^2. Moreover, error metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are employed. R^2 gauges the variance proportion, adjusted R^2

offers insights into reliability and correlation, MAE computes the average deviation, MAPE addresses size-dependent errors not captured by MAE, while MSE and RMSE function as risk measures that quantify the deviation between data points and the fitted model. Additionally, two other methods namely SI (Scatter index) and StI (Structural index) were also used to access the machine learning models. An excellent model would typically have a StI value close to 0 or a very low StI value, indicating that the model's structural complexity is well-suited to the data and that it generalizes well. When considering the SI parameter, it's evident that a model's performance is classified as poor when SI exceeds 0.3. In cases where the SI falls within the range of 0.2–0.3, the performance is considered fair. Furthermore, a model can be deemed to exhibit good performance when its SI ranges from 0.1 to 0.2. Lastly, an exceptional level of performance is achieved when the SI is less than 0.1.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - y')^2} \tag{8}$$

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \tag{9}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y - y'| \tag{10}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y' - y}{y} \right| \times 100 \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y' - y)^2} \tag{12}$$

$$\text{SI} = \frac{\text{RMSE}}{t'} \tag{13}$$

$$\text{StI} = 1 - R^2. \tag{14}$$

5 Sensitivity analysis

Sensitivity assessments include changing a system's inputs to determine exactly how each input parameter affects the output. In this study, two global sensitivity methods were utilized. The method of allocating the uncertainty in outputs to the uncertainty in each input element over their whole range of interest is known as global sensitivity analysis. When all the input factors are changed at once and the sensitivity is assessed across the whole range of each input factor, the analysis is said to be global. Sobol and FAST global sensitivity analysis were used in this study.

5.1 Sobol sensitivity analysis

The goal of a Sobol sensitivity analysis is to identify the proportion of each input parameter's variability in the model output that is dependent on that parameter alone or on its interaction with other factors (Saltelli et al. 1999). The classical analysis of variance in a factorial design is used as the basis for the dissection of the output variance in a Sobol sensitivity analysis. It should be understood that the goal of a Sobol sensitivity study is not to determine what causes input variability. It merely indicates the effect and magnitude it will have on the model's output. This method is thought to decompose any given integrable function $f(x)$ into the sum of 2^n orthogonal terms.

$$f(x) = f_0 + \sum_{s=1}^{n} \sum_{i_1 < \cdots < i_s}^{n} f_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}). \quad (15)$$

5.2 Fourier amplitude sensitivity testing (FAST)

One of the most well-liked tools for uncertainty and sensitivity analysis is the Fourier Amplitude Sensitivity Test (FAST). It breaks down a model output's variance into partial variances supplied by various model parameters using a periodic sampling strategy and a Fourier transformation. Cukier et al. (1973) developed FAST in 1973. The FAST technique, which provides a measure of the impact of each of the input qualities on the model's output, was used in this work to numerically illustrate each feature's contribution. The effect on strength increases as the parameter's relevance value rises.

6 Result and discussion

6.1 Hyperparameter optimization

After a given number of iterations, the objective function's inputs are examined to locate the global optimum point using the metaheuristic SHO method. When there is a lot of domain knowledge that could sway or bias the optimization process, spotted hyena optimization enables the identification of novel solutions. The appropriate hyperparameter values for each machine-learning method that makes use of the SHO algorithm are listed in Table 7.

6.2 LASSO regressor

The prediction of strength properties through this model was done appropriately with training and testing of data. The optimum hyperparameter value of the lasso regressor is depicted in Table 7. The training regression plot for compressive data

Table 7 Optimal hyperparameters for machine learning algorithms determined by SHO

Machine learning algorithms	Hyperparameters	Optimal results
Lasso	No. of hyenas	10
	Maximum iteration	50
	alpha	0.01
SVR	No. of hyenas	10
	c	100
	epsilon	1
	Kernel function	rbf
RFR	No. of hyenas	10
	max_depth	5
	n_estimator	100
BR	No. of hyenas	10
	No. of trees	0.001
	Noise	0.1
AB	No. of hyenas	10
	n_estimator	110
	Learning_rate	0.8
XGB	No. of hyenas	10
	max_depth	7
	n_estimator	100
	Learning_rate	0.1

is shown in Fig. 8 and that for split-tensile strength is shown in Fig. 10. Likewise, Figs. 9 and 11 represent the regression plots in testing for both datasets. The testing accuracy of the compressive model is achieved at a somewhat 63% confidence level with a mean absolute error of 6.77 KN and that of the split-tensile model is at an 83% confidence level with and mean absolute error of 0.22 KN. As SC-AASC possesses a non-linear relationship between constituent material and strength it doesn't show a great level of accuracy in this modified linear regression model. Additional details of all the statistical parameters are listed in Tables 9 and 10.

6.3 Support vector regressor (SVR)

Support vector regression is a supervised learning algorithm that is used to predict discrete values. Support Vector Regression uses the same principle as the SVMs. The basic idea behind SVR is to find the best-fit line. Figures 8 and 10 demonstrate the training regression plot for compressive and split-tensile strength data, respectively. Similarly, Figs. 9 and 11 show the regression plots tested on both datasets. The testing accuracy of the compressive strength model is at 85%, and that for split-tensile is at 87% confidence level. In SVR the use of a non-linear kernel resulted in better performance
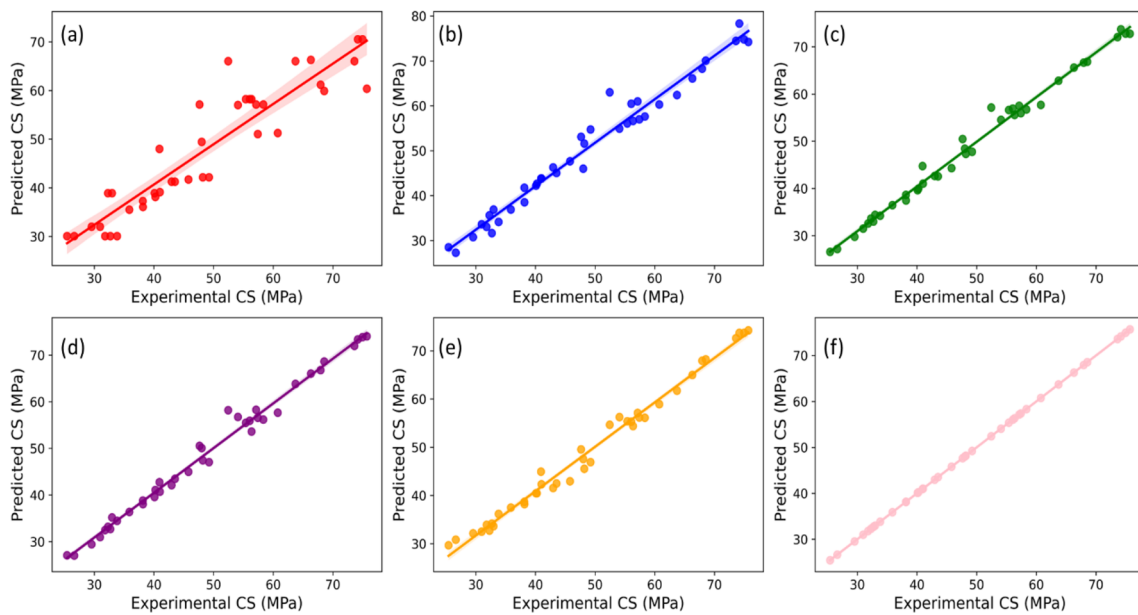
Fig. 8 Regression analysis in the training phase of CS models: **a** Lasso; **b** SVR; **c** RFR; **d** BR; **e** AB; and **f** XGB
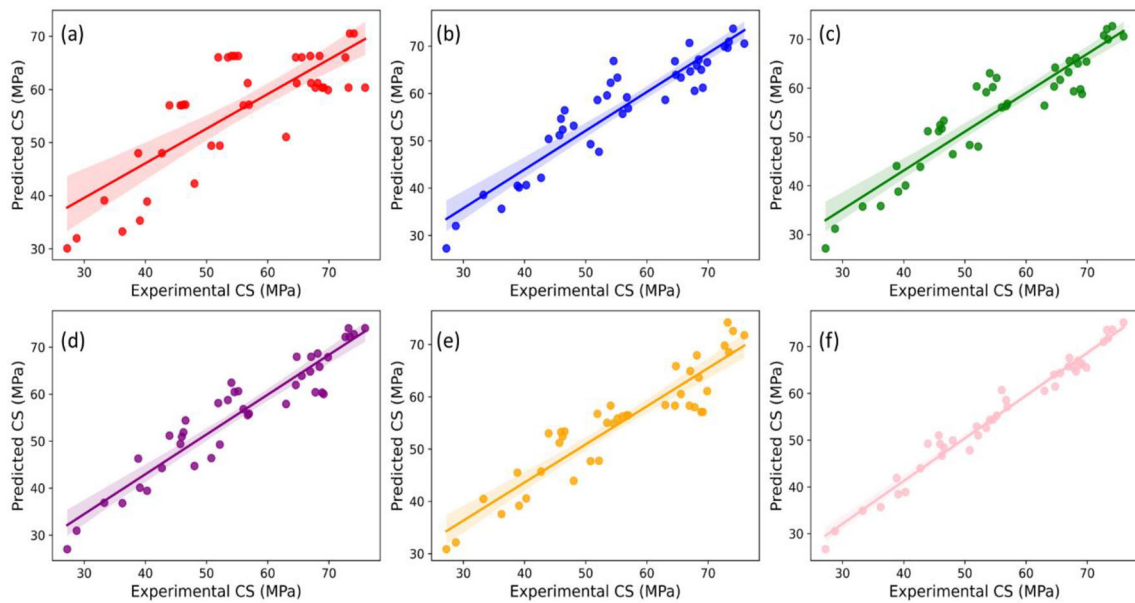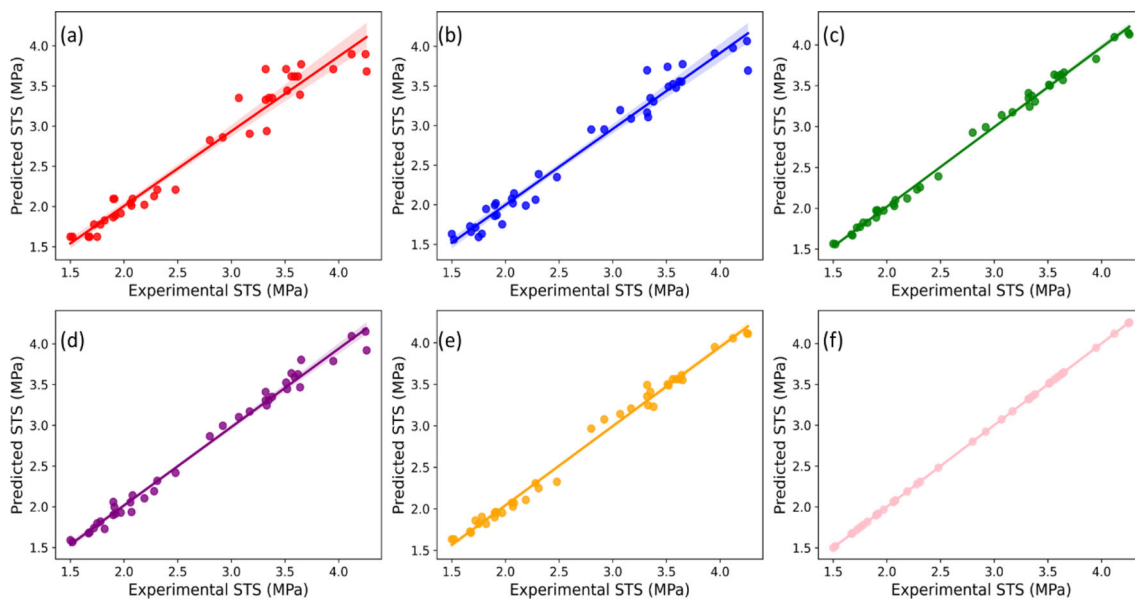

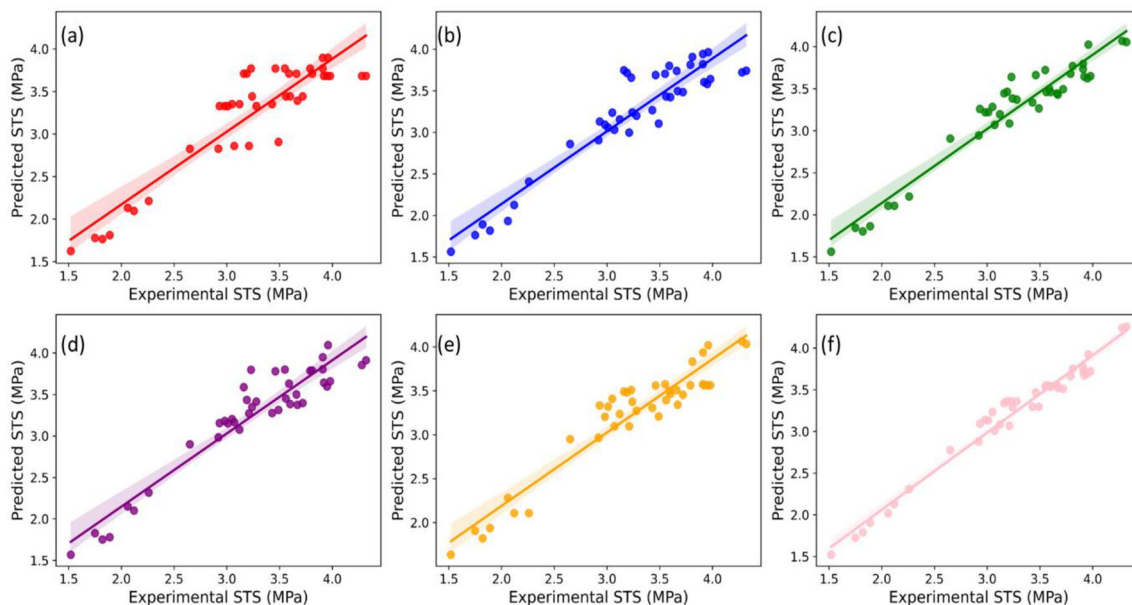**Fig. 9** Regression analysis in the testing phase of CS models: **a** Lasso; **b** SVR; **c** RFR; **d** BR; **e** AB; and **f** XGB

than the linear regression model. More information about each statistical parameter is listed in Tables 9 and 10.

6.4 Random forest (RF)

This is an advanced version of the bagging ensemble regressor and creates DTs at every data point which decreases variance and when ensembled gives better accuracy. For this reason, the model predicts compressive strength at an 88% confidence level and split-tensile strength at a 91%

confidence level in the testing phase. The ideal RFR hyperparameter value is shown in Table 6. The fitting curve shown in Figs. 8, 9, 10, and 11. Tables 9 and 10 represent details of all the additional model performance parameters.

6.5 Bagging regressor (BR)

This bagging ensemble model predicts compressive strength at an 80% confidence level and predicts split-tensile strength at a 91.5% confidence level in the testing phase. The hyperparameters optimized by SHO are listed in Table 6. The fitting

Fig. 10 Regression analysis in the training phase of STS models: **a** Lasso; **b** SVR; **c** RFR; **d** BR; **e** AB; and **f** XGB

Fig. 11 Regression analysis in the testing phase of CS models: **a** Lasso; **b** SVR; **c** RFR; **d** BR; **e** AB; and **f** XGB

curve is shown in Figs. 8, 9, 10, and 11 which includes the training and testing phase for both datasets. It shows greater performance than prior models as it decreases the variance of the data. Details of all the additional model performance parameters are shown in Tables 9 and 10.

6.6 AdaBooster (AB)

This boosting ensemble model uses fix-shaped decision stumps at every data point which reduces bias and simplifies the model ultimately increasing accuracy. This model

forecasts compressive strength at an 84.5% confidence level and split-tensile strength at a 90.4% confidence level in the testing phase achieving a generalized MAE value. The fitting curve, which includes the training and testing phases for both datasets, is depicted in Figs. 8, 9, 10, and 11.

6.7 XGBooster (XGB)

Spotted hyena-optimized hyperparameters of XGB are shown in Table 6. XGB is the most advanced boosting ensemble model and using CART and gradient boosting methods

Table 8 Training and testing scores of all the CS models

Model	Phase	R^2	Adj. R^2	SI	StI	MAE	MAPE	RMSE
Lasso	Train	0.85	0.842	0.18	0.15	4.22	8.88	15.41
	Test	0.63	0.618	0.32	0.37	6.77	12.7	19.21
SVR	Train	0.93	0.91	0.085	0.07	3.14	5.12	8.81
	Test	0.85	0.83	0.162	0.15	4.05	7.4	15.62
RFR	Train	0.98	0.97	0.064	0.02	1.12	2.21	1.81
	Test	0.88	0.87	0.148	0.12	3.6	6.5	13.41
BR	Train	0.975	0.96	0.068	0.025	1.34	2.91	1.78
	Test	0.80	0.78	0.21	0.20	4.63	8.5	9.88
AB	Train	0.979	0.962	0.067	0.021	1.71	3.95	2.05
	Test	0.845	0.836	0.186	0.155	4.17	7.9	8.62
XGB	Train	0.99	0.98	0.051	0.01	0.0015	0.35	1.14
	Test	0.97	0.964	0.068	0.03	1.78	2.5	2.26

Table 9 Training and testing scores of all the STS models

Model	Phase	R^2	Adj. R^2	SI	StI	MAE	MAPE	RMSE
Lasso	Train	0.95	0.93	0.09	0.05	0.13	4.78	0.035
	Test	0.83	0.81	0.19	0.17	0.22	6.8	0.082
SVR	Train	0.96	0.95	0.08	0.04	0.12	4.49	0.025
	Test	0.87	0.86	0.15	0.13	0.18	5.4	0.060
RFR	Train	0.99	0.98	0.05	0.01	0.052	1.94	0.004
	Test	0.91	0.90	0.11	0.09	0.17	5	0.0414
BR	Train	0.99	0.982	0.052	0.01	0.06	2.31	0.084
	Test	0.915	0.904	0.108	0.085	0.16	5.11	0.19
AB	Train	0.99	0.98	0.053	0.01	0.067	2.7	0.085
	Test	0.904	0.892	0.12	0.096	0.18	5.8	0.21
XGB	Train	0.99	0.985	0.049	0.01	0.0006	0.2	0.0009
	Test	0.968	0.96	0.071	0.032	0.1	3	0.12

increases the predictive power of the model. The model estimates compressive strength at a 97% confidence level and split-tensile strength at a 96% confidence level; performing with less MAE, MAPE, and RMSE. For both datasets, the training and testing phases of the fitting curve are depicted in Figs. 8, 9, 10, and 11. It uses self-compatible regulatory functions i.e., under-sampling and column shrinkage that makes it more reliable and outperforms other ML models. Tables 8 and 9 include all the statistical indices of the XGB model.

6.8 Comparison between the machine learning models

This section presents a thorough evaluation of the model's robustness and generalizability by contrasting each model with the others. A comparison of the models was made by plotting the errors of each data point, analyzing the maximum errors, and performance metrics scores for each model. A 3D plot is used to visualize the errors of each data point in correspondence to the experiment and predicted values. Figure 12 represents the 3D error plot for the compressive dataset while Fig. 13 represents the split-tensile dataset. From the plot, it can be seen that XGB shows the least error per data point indicating superior prediction performance. Ensemble ML models also perform better in comparison to single ML models. The maximum error for each model is listed in Table 10. Lasso performs least accurately with the highest maximum error while XGB outperforms all. Likewise, for both datasets among all the model's Lasso regressor model shows the least R^2 and adjusted R^2 value while the XGBoost model shows the highest R^2 and adjusted R^2 score. Ensemble ML models show good generalizability in comparison to single ML models and compared to bagging ensemble models boosting ensemble models showed good reliability.

Accuracy and error statistics of all the models in training and testing are shown in Tables 8 and 9. Figures 14 and 15 represent the line plot for actual and predicted values in
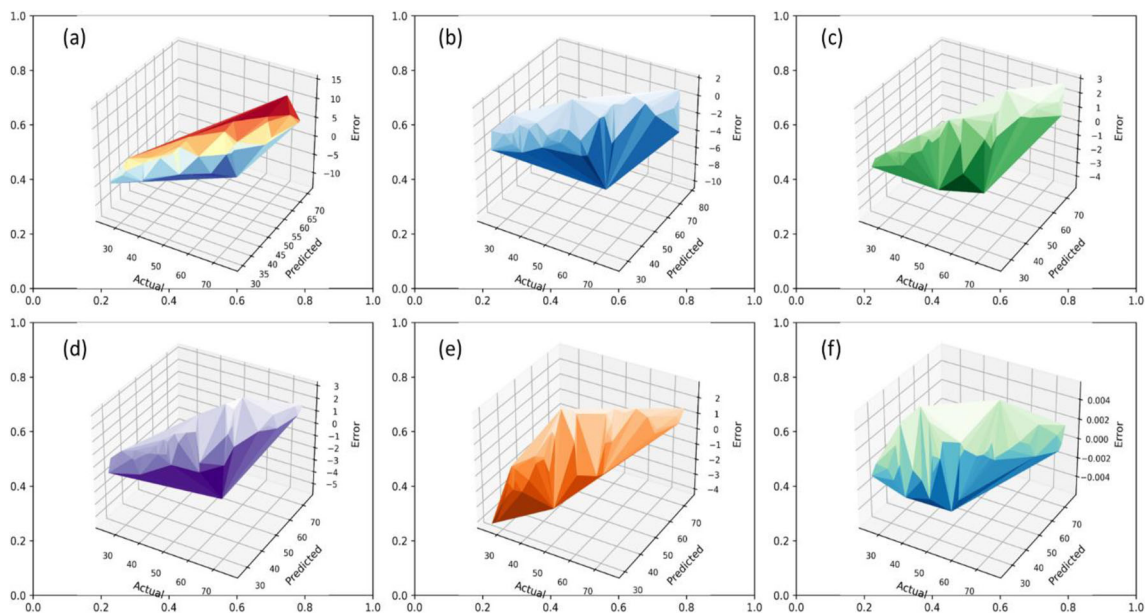
Fig. 12 3D plot for error in prediction for CS models: **a** Lasso; **b** SVR; **c** RFR; **d** BR; **e** AB; and **f** XGB
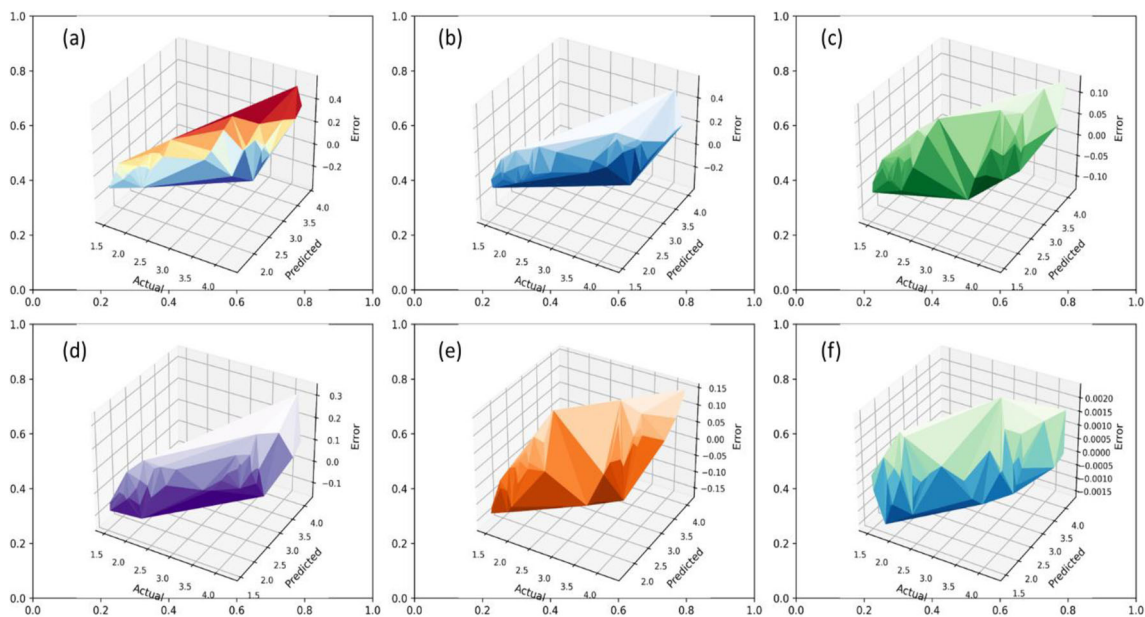

Fig. 13 3D plot for error in prediction for STS models: **a** Lasso; **b** SVR; **c** RFR; **d** BR; **e** AB; and **f** XGB

the testing phase. For both CS and STS models, the Lasso model with a higher value of error metrics indicates lower generalization. SVR shows the least error among the single ML models as it shows effectiveness in modeling. Overall boosting ensemble models show fewer errors in comparison to the bagging model as ensemble ML models perform at higher accuracy than single ML models as it prioritizes weak learners but as previously discussed RF model shows the best error statistically than AdaBooster for the split-tensile dataset. XGB model which can be labeled as the best ML

model shows excellent generalizability and robust predictive capability.

Figure 16 represents a Taylor diagram of all the models tested on both the CS and STS datasets. The radial distance in the Taylor diagram represents the standard deviation and the angular axis represents the co-relation. The performance of all the models was visualized with a reference. It can be seen that the SHO-XGB performs well compared to all the models for both datasets.

Table 10 Maximum error of each model

Algorithms	Max. error in CS prediction (MPa)	Max. error in STS prediction (MPa)
Lasso	9.8	0.42
SVR	8.7	0.40
RFR	2.8	0.14
BR	3.6	0.21
AB	4.8	0.16
XGB	0.004	0.015

6.9 Comparison among previously published results

The spotted hyena-optimized XGBoost was found to be performing exceptionally well in both the CS and STS datasets. In this section, a comparison is drawn between previously published research articles that have worked on self-compacting geopolymer concrete and alkali-activated slag concrete. Awoyera et al. (2020) used ANN and GEP for the prediction of compressive strength of SCGC. ANN model was found to be more accurate. Shahmansouri et al. (2020) used GEP for CS prediction of GGBS-based alkali-activated slag concrete. Faridmehr et al. (2021) used Bat-optimized ANN for SCGC CS estimation, and Basilio and Goliatt (2022) utilized gradient-boosting hybridization for CS prediction. Figure 17 represents a comparison of SHO-XGB with all the models. It can be ascertained that the developed model gives satisfactory results.

6.10 Sensitivity analysis

Two global sensitivity approaches were used to evaluate the relative influence of each input parameter on output. Figure 18 represents the total order of Sobol indices of all the input parameters. The total sampling size of the Sobol sensitivity method was determined to be 77,784. From Fig. 18, it can be seen that AAS/binder ratio, AAS content, binder content, and curing days greatly affect the compressive strength of alkali-activated slag concrete. Similar observations have been made earlier using slag as binder material (Kumar Dash et al. 2023; Saini and Vattipalli 2020; Shahmansouri et al. 2020). So, the alignment of the current study with earlier published results establishes the importance of these factors in controlling the strength development of alkali-activated concrete. The AAS content and AAS-binder ratio which resembles the water content and water-cement ratio are the two most sensitive factors governing strength development (Fig. 18). So, we can adopt the approach of controlling these two factors to have the desired strength properties of alkali-activated slag concrete. The FAST sensitivity analysis (Fig. 19) was used to overcome the cost-effectiveness and time factor associated with the Sobol method. The FAST method takes less time while giving almost equal results. The sampling size in the FAST method was 77,848. AAS/binder ratio, binder content, and AAS content are also presented with maximum index. The fresh property results which were taken as input scored minimum sensitivity value as they possess' a linear relationship with strength.

Figures 20 and 21 represent the effect of various parameters on the compressive strength and split-tensile strength of
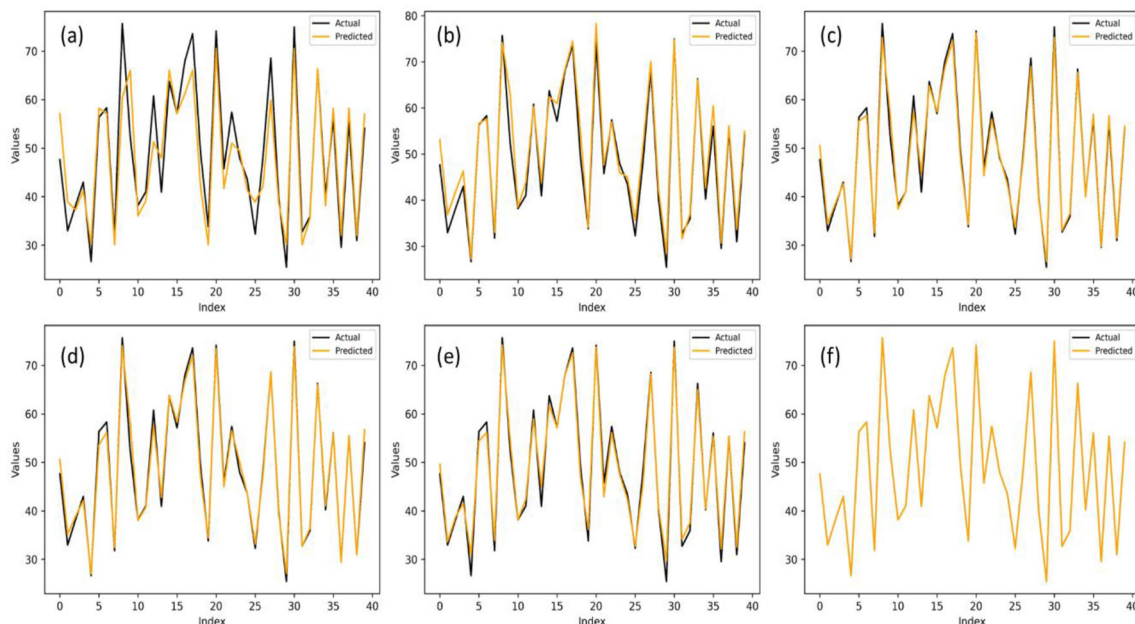

Fig. 14 Line plot between actual and predicted CS models in testing: **a** Lasso; **b** SVR; **c** RFR; **d** BR; **e** AB; and **f** XGB
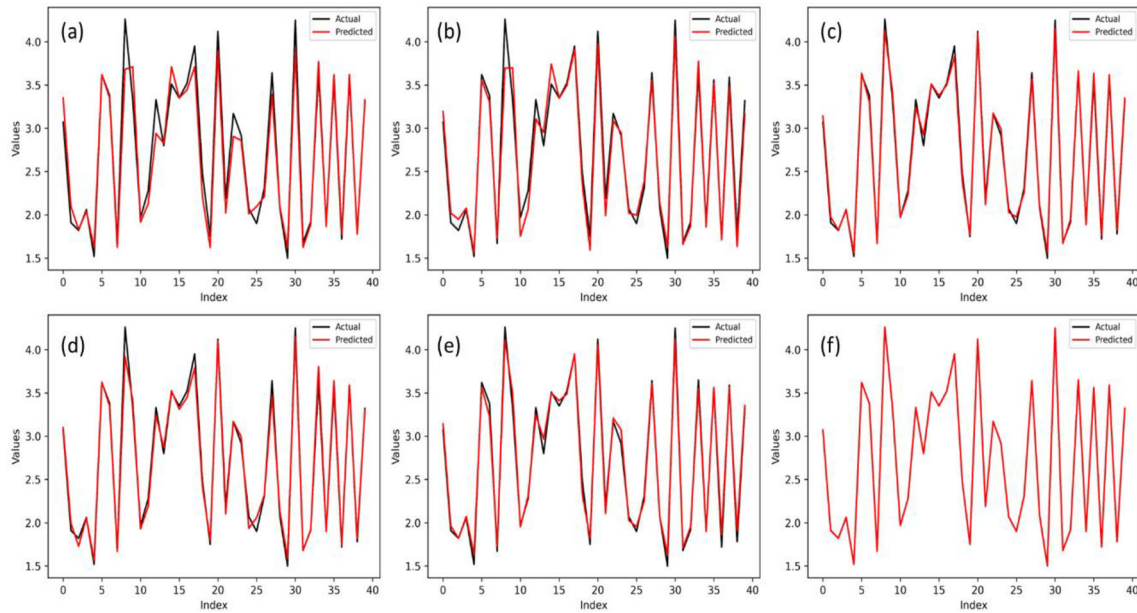
Fig. 15 Line plot between actual and predicted STS models in testing: **a** Lasso; **b** SVR; **c** RFR; **d** BR; **e** AB; and **f** XGB
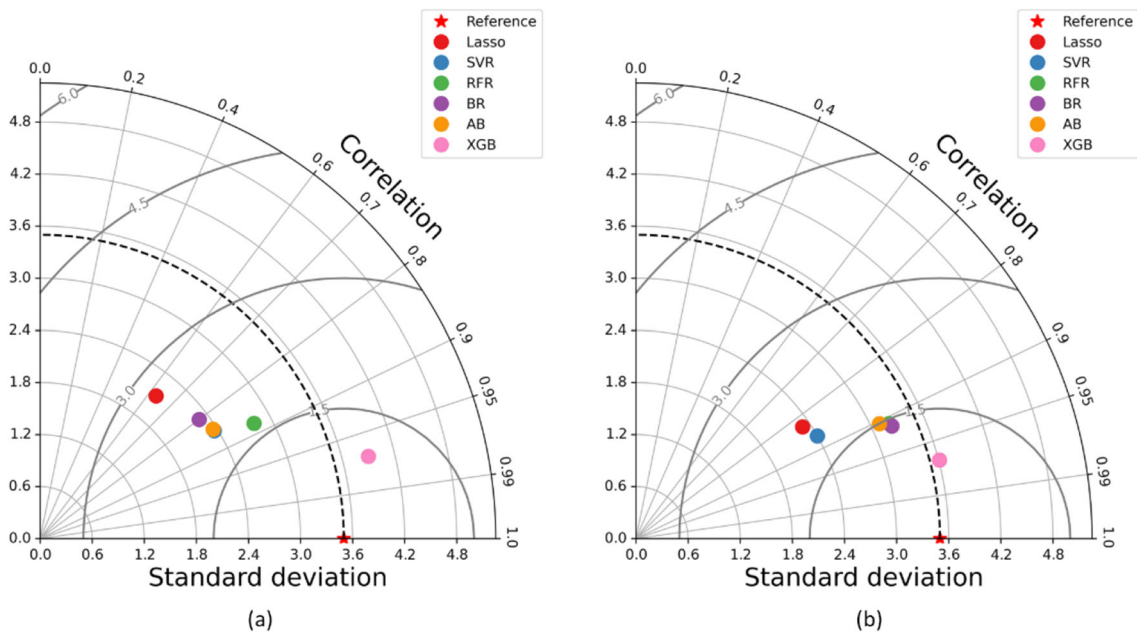

Fig. 16 Taylor diagram of all the models: **a** CS dataset, **b** STS dataset

SC-AASC. It's noted that as the content of GGBFS increases, the compressive and split-tensile strengths of SC-AASC also increase. This suggests that GGBFS is contributing positively to the strength properties of the concrete mix. GGBFS is known for its pozzolanic properties, which can enhance the strength and durability of concrete. Conversely, reducing the content of AAS also results in higher strength. AAS is used as an activator in alkali-activated concretes, and a lower content implies that the mixture might be more optimized in terms of activator dosage. It can also be observed that an increase

in GGBFS content and a decrease in AAS content led to SC-AASC with higher strength. GGBFS content in the range of 500–540 kg/m^3 and AAS in the range of 190–200 kg/m^3 resulted in higher strength. A lower AAS/binder ratio resulted in higher strength while a higher content of coarse aggregate also increased the strength. This implies that reducing the amount of activator relative to the total binder content is beneficial for higher strength development, however, the increase in coarse aggregate has a negative effect on the flow property of SC-AASC. Lower fine aggregate content seems to

Fig. 17 Comparison of SHO-XGB with published models
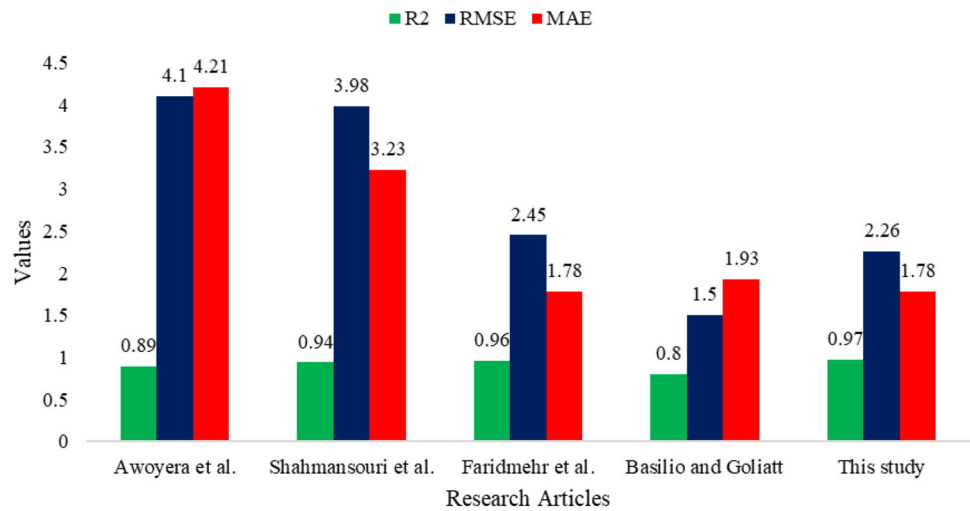
Fig. 18 Sobol total order index of input parameters
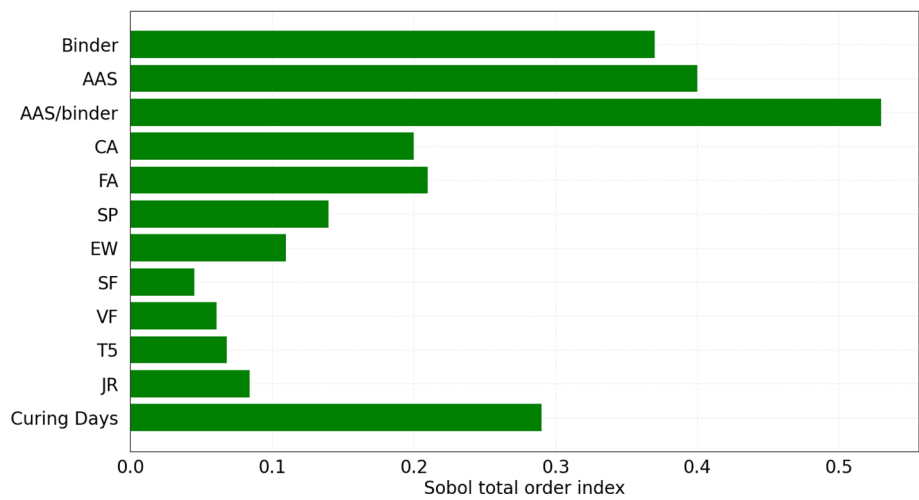
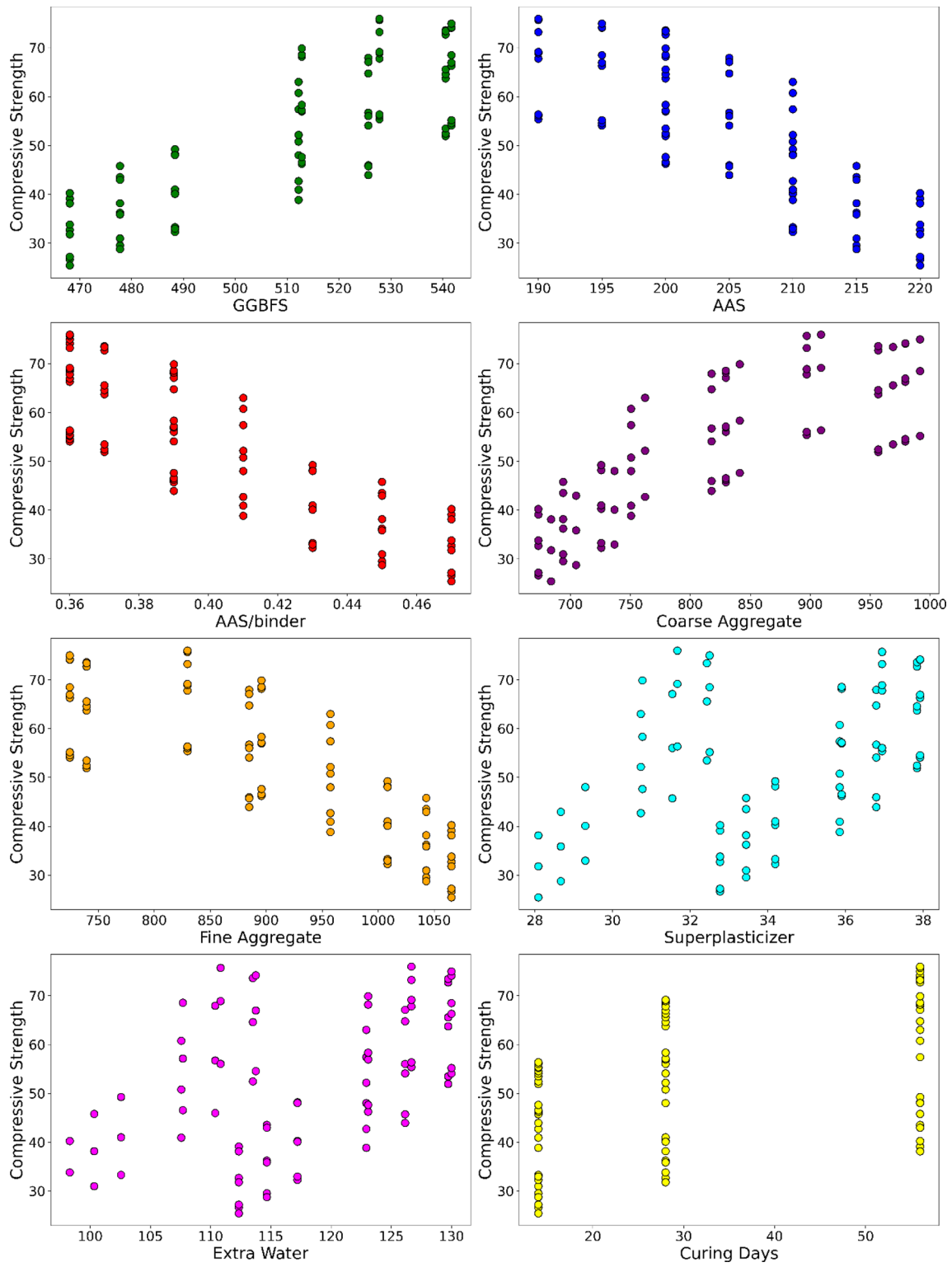**Fig. 19** FAST First order index of input parameters
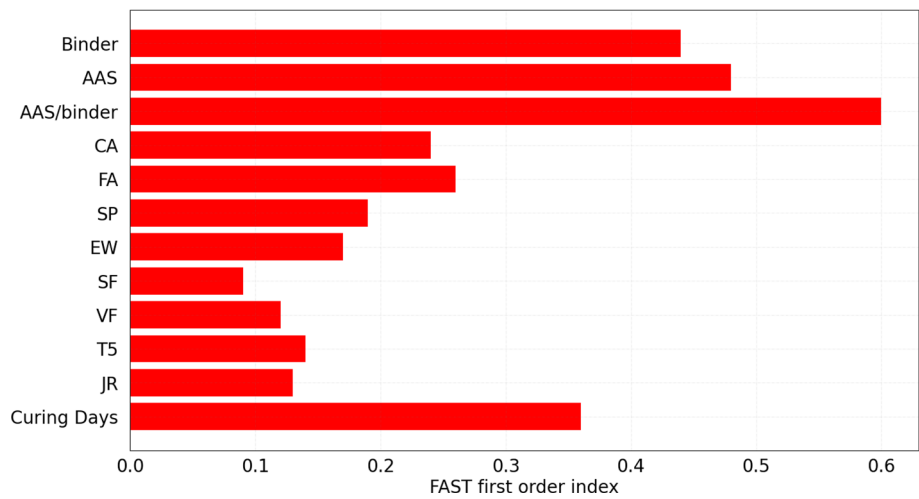
Fig. 20 Effect of various parameters on compressive strength
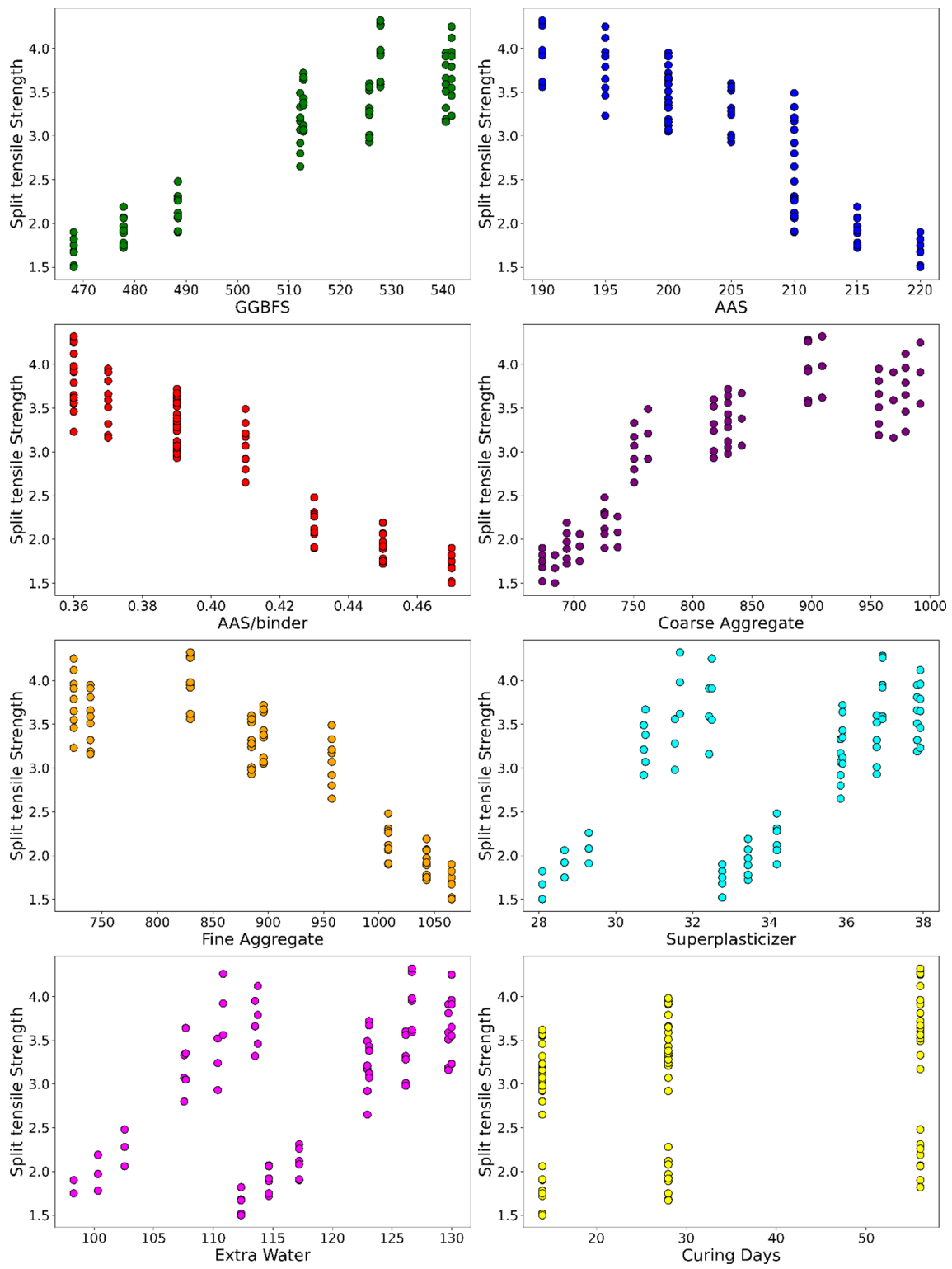
Fig. 21 Effect of various parameters on split-tensile strength

have a positive impact on both compressive and split-tensile strength while 7% of superplasticizer dosage increased the strength of SC-AASC. 21% of extra water dosage and longer curing period observed to be giving maximum strength.

7 Practical application of the research

The research on predicting mechanical strength properties in high-performance SC-AASC using metaheuristic optimized machine learning models has multiple practical applications that can offer advantages to various stakeholders in the construction industry. This research has explored the development of High-Performance SC-AASC formulated with GGBFS can be utilized as an eco-friendly alternative to traditional concrete in various construction projects. Its reduced carbon footprint and efficient use of industrial byproducts make it an ideal choice for environmentally conscious builders. Its self-compacting properties can simplify the construction of complex structures, such as bridges, tunnels, and high-rise buildings, where ensuring proper compaction is traditionally challenging. This can lead to faster infrastructure development. The optimization of High-Performance SC-AASC mixtures enhances their durability and longevity, reducing the need for frequent maintenance and repair of concrete structures. This can result in significant long-term cost savings for infrastructure owners and operators. The major drawback is the cost associated with the production of this type of concrete as there are no specific codes and regulations available to optimize its performance. A prediction model can help decrease the cost of the concrete by optimizing various aspects of the manufacturing process like mix design optimization, process efficiency, quality control, and materials selection. The application of metaheuristic-optimized machine learning models for predicting its strength properties offers a powerful tool for engineers and construction professionals. They can use these models to assess material performance quickly and accurately during the design phase, leading to better-informed decisions. The use of predictive modeling reduces the reliance on extensive and time-consuming laboratory testing, speeding up material characterization. This not only accelerates the research and development process but also lowers testing costs, benefiting both researchers and construction projects for the utilization of SCGC and SC-AASC. The utilization of GGBFS as a primary binder encourages collaboration between the construction and steel industries, promoting the repurposing of industrial byproducts and fostering sustainable practices.

8 Conclusions

The high-performance SC-AASC represents an eco-conscious alternative to conventional cement-based composites. By utilizing industrial waste materials and emitting lower quantities of carbon dioxide, AASC emerges as a sustainable construction material, effectively minimizing its environmental influence. Predicting the mechanical strength in SC-AASC presents a notable hurdle, necessitating the use of advanced techniques for precise estimation. This study tackles this challenge through the creation of innovative machine-learning models optimized with metaheuristic methods. The subsequent section presents the concluding insights derived from this research endeavor.

- The study proposes a method of production of high-strength SC-AASC from a base M25 mix grade to an M45 and M60 mix grade.
- The AAS content varied from 220 to 190 kg/m^3 and the AAS/binder varied from 0.47 to 0.36 to produce high-performance SC-AASC. The percentage of SP varied between 6–7% while the percentage of extra water kept at 21% and 24%.
- AAS content of 200 kg/m^3 and AAS/binder of 0.39 yielded greater strength for M45 graded SC-AASC while M60 grade SC-AASC showed higher strength with AAS content of 190 kg/m^3 and AAS/binder of 0.36.
- SP of 6% and extra water content of 24% facilitated good workability of M45 and M60 grade SC-AASC.
- A database containing 135 datapoints was developed for the creation of machine learning models by taking 12 input parameters and compressive strength and split-tensile strength as output.
- Data analysis resulted in the database having no outlier but having non-linear co-relation.
- Six machine learning models were developed consisting of two single supervised algorithms and four ensembled supervised algorithms.
- The hyperparameters of all the machine learning models were optimized using the metaheuristic bio-based spotted hyena optimization method.
- The ensembled ML algorithm (Bagging and Boosting) showed better accuracy and tolerable error in comparison to the single ML algorithm (Lasso and SVR). Error metrics showed boosting ensemble models (AdaBooster and XGB) showed better accuracy than bagging ensemble models (BR and RF).

- Non-linear single ML models (SVR) showed better performance than modified linear single ML models (Lasso; which can be termed as the worst-performing model). The optimized XGB was found to be best best-performing model.
- Sobol and FAST sensitivity methods were used for the assessment of relative influence. The AAS/binder, AAS content, binder content, and curing days were found to be the primary influencing factors concerning the strength properties of high-strength SC-AASC.

9 Limitations and future scope of the study

9.1 Limitations

The following are the limitations of this study.

- The research primarily focuses on specific mix grades (M25 to M60) and may lack broader applicability to other concrete grades or different environmental conditions.
- While the study emphasizes reduced environmental impact, the broader environmental implications can be derived using life cycle assessments or environmental footprint analysis.
- Although the study explores variations in critical parameters like AAS content, AAS/binder ratio, and curing days, the scope might not cover all potential combinations or extreme ranges of these parameters. This limitation might restrict the model's predictive capacity in scenarios not fully represented in the dataset.
- The size of the dataset needs to be enhanced with more parametric variation and experimental testing for better ML performance.
- The sensitivity analysis, while valuable, might not encompass the entire spectrum of potential influencing factors. Factors not included in the analysis might have an impact on the material properties, affecting the comprehensiveness of the conclusions drawn.

9.2 Future scope

The future scope of the study includes

- Researching novel materials or hybrid systems that combine various sustainable elements, such as recycled aggregates, nano-materials, or bio-based additives, could lead to the development of even more advanced and sustainable construction materials.
- Conducting a comprehensive life cycle assessment (LCA) or environmental footprint analysis to evaluate the holistic environmental impact of SC-AASC, including its production, use, and disposal stages, would provide valuable insights into its sustainability.
- Expanding the study to encompass a wider range of parameters could provide a more comprehensive understanding of their impact on the properties of high-strength SC-AASC.
- Conducting extensive external validation by applying the developed models to real construction projects or experimental testing would validate the models' performance in practical scenarios, ensuring their reliability and applicability.
- Investigating optimization strategies beyond metaheuristic methods, such as evolutionary algorithms or reinforcement learning, could enhance the precision of parameter optimization for high-strength SC-AASC production.
- Collaborating with industry stakeholders and standardization bodies to develop guidelines or standards for the production and application of high-strength SC-AASC would facilitate its widespread adoption in the construction sector.

Data availability Data can be made available on request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahmed HU, Mohammed AS, Mohammed AA, Faraj RH (2021) Systematic multiscale models to predict the compressive strength of fly ash-based geopolymer concrete at various mixture proportions and curing regimes. PLoS One 16(6):e0253006. https://doi.org/10.1371/journal.pone.0253006

Ahmed HU, Abdalla AA, Mohammed AS, Mohammed AA, Mosavi A (2022a) Statistical methods for modeling the compressive strength of geopolymer mortar. Materials 15(5):5. https://doi.org/10.3390/ma15051868

Ahmed HU, Mohammed AA, Mohammed A (2022b) Soft computing models to predict the compressive strength of GGBS/FA- geopolymer concrete. PLoS One 17(5):e0265846. https://doi.org/10.1371/journal.pone.0265846

Ahmed HU, Mohammed AA, Mohammed AS (2022c) The role of nanomaterials in geopolymer concrete composites: a state-of-the-art review. J Build Eng 49:104062. https://doi.org/10.1016/j.jobe.2022.104062

Ahmed HU, Mohammed AS, Faraj RH, Qaidi SMA, Mohammed AA (2022d) Compressive strength of geopolymer concrete modified with nano-silica: experimental and modeling investigations. Case Stud Constr Mater 16:e01036. https://doi.org/10.1016/j.cscm.2022.e01036

Ahmed HU, Mohammed AS, Mohammed AA (2022e) Multivariable models including artificial neural network and M5P-tree to forecast the stress at the failure of alkali-activated concrete at ambient curing condition and various mixture proportions. Neural Comput Appl 34(20):17853–17876. https://doi.org/10.1007/s00521-022-07427-7

Ahmed HU, Mohammed AS, Mohammed AA (2022f) Proposing several model techniques including ANN and M5P-tree to predict the compressive strength of geopolymer concretes incorporated with nano-silica. Environ Sci Pollut Res 29(47):71232–71256. https://doi.org/10.1007/s11356-022-20863-1

Ahmed HU, Mohammed AS, Mohammed AA (2023a) Effectiveness of silicon dioxide nanoparticles (Nano SiO$_2$) on the internal structures, electrical conductivity, and elevated temperature behaviors of geopolymer concrete composites. J Inorg Organomet Polym Mater. https://doi.org/10.1007/s10904-023-02672-2

Ahmed HU, Mohammed AS, Faraj RH, Abdalla AA, Qaidi SMA, Sor NH, Mohammed AA (2023b) Innovative modeling techniques including MEP, ANN and FQ to forecast the compressive strength of geopolymer concrete modified with nanoparticles. Neural Comput Appl 35(17):12453–12479. https://doi.org/10.1007/s00521-023-08378-3

Ahmed HU, Mohammed AS, Mohammed AA (2023c) Engineering properties of geopolymer concrete composites incorporated recycled plastic aggregates modified with nano-silica. J Build Eng 75:106942. https://doi.org/10.1016/j.jobe.2023.106942

Alsalman A, Assi LN, Kareem RS, Carter K, Ziehl P (2021) Energy and CO$_2$ emission assessments of alkali-activated concrete and ordinary Portland cement concrete: a comparative analysis of different grades of concrete. Clean Environ Syst 3:100047. https://doi.org/10.1016/j.cesys.2021.100047

Alsharari F, Iftikhar B, Uddin MA, Deifalla AF (2023) Data-driven strategy for evaluating the response of eco-friendly concrete at elevated temperatures for fire resistance construction. Results Eng 20:101595. https://doi.org/10.1016/j.rineng.2023.101595

Angiulli F, Pizzuti C (2005) Outlier mining in large high-dimensional data sets. IEEE Trans Knowl Data Eng 17(2):203–215. https://doi.org/10.1109/TKDE.2005.31

Awad M, Khanna R (2015) Support vector regression. In: Awad M, Khanna R (eds) Efficient learning machines: theories, concepts, and applications for engineers and system designers. Apress, Berkeley, pp 67–80. https://doi.org/10.1007/978-1-4302-5990-9_4

Awoyera PO, Kirgiz MS, Viloria A, Ovallos-Gazabon D (2020) Estimating strength properties of geopolymer self-compacting concrete using machine learning techniques. J Mark Res 9(4):9016–9028. https://doi.org/10.1016/j.jmrt.2020.06.008

Basilio SA, Goliatt L (2022) Gradient boosting hybridized with exponential natural evolution strategies for estimating the strength of geopolymer self-compacting concrete. Knowl Based Eng Sci 3(1):1. https://doi.org/10.51526/kbes.2022.3.1.1-16

Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. COLT'92: Proceedings of the fifth annual workshop on computational learning theory, USA. pp 144–152. https://doi.org/10.1145/130385.130401

Chen Y-C (2017) A tutorial on kernel density estimation and recent advances. Biostat Epidemiol 1(1):161–187. https://doi.org/10.1080/24709360.2017.1396742

Chen Z, Amin MN, Iftikhar B, Ahmad W, Althoey F, Alsharari F (2023a) Predictive modelling for the acid resistance of cement-based composites modified with eggshell and glass waste for sustainable and resilient building materials. J Build Eng 76:107325. https://doi.org/10.1016/j.jobe.2023.107325

Chen Z, Iftikhar B, Ahmad A, Dodo Y, Abuhussain MA, Althoey F, Sufian M (2023b) Strength evaluation of eco-friendly waste-derived self-compacting concrete via interpretable genetic-based machine learning models. Mater Today Commun 37:107356. https://doi.org/10.1016/j.mtcomm.2023.107356

Chou J-S, Pham A-D (2013) Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength. Constr Build Mater 49:554–563. https://doi.org/10.1016/j.conbuildmat.2013.08.078

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297. https://doi.org/10.1007/BF00994018

Cukier RI, Fortuin CM, Shuler KE, Petschek AG, Schaibly JH (1973) Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. J Chem Phys 59(8):3873–3878. https://doi.org/10.1063/1.1680571

Das R, Panda S, Sahoo AS, Panigrahi SK (2023) Effect of superplasticizer types and dosage on the flow characteristics of GGBFS based self-compacting geopolymer concrete. Mater Today Proc 62:1–11

Dash PK, Parhi SK, Patro SK, Panigrahi R (2023) Influence of chemical constituents of binder and activator in predicting compressive strength of fly ash-based geopolymer concrete using firefly-optimized hybrid ensemble machine learning model. Mater Today Commun 37:107485. https://doi.org/10.1016/j.mtcomm.2023.107485

Davidovits J (1976) Solid-phase synthesis of a mineral blockpolymer by low temperature polycondensation of alumino-silicate polymers: Na-poly (sialate) or Na-PS and characteristics IUPAC Symposium on Long-Term Properties of Polymers and Polymeric Materials, Stockholm. Topic III: New Polymers of High Stability

Dhiman G, Kumar V (2017) Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications. Adv Eng Softw 114:48–70. https://doi.org/10.1016/j.advengsoft.2017.05.014

Dhiman G, Kumar V (2019) Spotted hyena optimizer for solving complex and non-linear constrained engineering problems. In: Yadav N, Yadav A, Bansal JC, Deep K, Kim JH (eds) Harmony search and nature inspired optimization algorithms. Springer, Berlin, pp 857–867. https://doi.org/10.1007/978-981-13-0761-4_81

Dong W, Huang Y, Lehane B, Ma G (2020) XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. Autom Constr 114:103155. https://doi.org/10.1016/j.autcon.2020.103155

Dou J, Yunus AP, Tien Bui D, Merghadi A, Sahana M, Zhu Z, Chen C-W, Khosravi K, Yang Y, Pham BT (2019) Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci Total Environ 662:332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221

Duan J, Asteris PG, Nguyen H, Bui X-N, Moayedi H (2021) A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model. Eng Comput 37(4):3329–3346. https://doi.org/10.1007/s00366-020-01003-0

Dwibedy S, Panigrahi SK (2023) Factors affecting the structural performance of geopolymer concrete beam composites. Constr Build Mater 409:134129. https://doi.org/10.1016/j.conbuildmat.2023.134129

EFNARC (2002) Specification and guidelines for self-compacting concrete, European federation of specialist construction chemicals and concrete systems, Syderstone, Norfolk

Faraj RH, Mohammed AA, Mohammed A, Omer KM, Ahmed HU (2022a) Systematic multiscale models to predict the compressive strength of self-compacting concretes modified with nanosilica at different curing ages. Eng Comput 38(3):2365–2388. https://doi.org/10.1007/s00366-021-01385-9

Faraj RH, Mohammed AA, Omer KM, Ahmed HU (2022b) Soft computing techniques to predict the compressive strength of green self-compacting concrete incorporating recycled plastic aggregates and industrial waste ashes. Clean Technol Environ Policy 24(7):2253–2281. https://doi.org/10.1007/s10098-022-02318-w

Faridmehr I, Nehdi ML, Huseien GF, Baghban MH, Sam ARM, Algaifi HA (2021) Experimental and informational modeling study of sustainable self-compacting geopolymer concrete. Sustainability 13(13):13. https://doi.org/10.3390/su13137444

Feng D-C, Liu Z-T, Wang X-D, Chen Y, Chang J-Q, Wei D-F, Jiang Z-M (2020) Machine learning-based compressive strength prediction for concrete: an adaptive boosting approach. Constr Build Mater 230:117000. https://doi.org/10.1016/j.conbuildmat.2019.117000

Géron A (2019) Hands-on machine learning with scikit-learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media Inc, Sebastopol

Hodge V, Austin J (2004) A survey of outlier detection methodologies. Artif Intell Rev 22(2):85–126. https://doi.org/10.1023/B:AIRE.0000045502.10941.a9

Hu X (2023) Evaluation of compressive strength of the HPC produced with admixtures by a novel hybrid SVR model. Multiscale Multidiscip Model Exp Des 6(3):357–370. https://doi.org/10.1007/s41939-023-00150-3

Iftikhar B, Alih SC, Vafaei M, Javed MF, Rehman MF, Abdullaev SS, Tamam N, Khan MI, Hassan AM (2023a) Predicting compressive strength of eco-friendly plastic sand paver blocks using gene expression and artificial intelligence programming. Sci Rep 13(1):1. https://doi.org/10.1038/s41598-023-39349-2

Iftikhar BC, Alih S, Vafaei M, Javed MF, Ali M, Gamil Y, Rehman MF (2023b) A machine learning-based genetic programming approach for the sustainable production of plastic sand paver blocks. J Mater Res Technol 25:5705–5719. https://doi.org/10.1016/j.jmrt.2023.07.034

Jithendra C, Elavenil S (2019) Role of superplasticizer on GGBS based geopolymer concrete under ambient curing. Mater Today Proc 18:148–154. https://doi.org/10.1016/j.matpr.2019.06.288

Kang M-C, Yoo D-Y, Gupta R (2021) Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete. Constr Build Mater 266:121117. https://doi.org/10.1016/j.conbuildmat.2020.121117

Kim JS, Park J (2014) An experimental evaluation of development length of reinforcements embedded in geopolymer concrete. Appl Mech Mater 578–579:441–444. https://doi.org/10.4028/www.scientific.net/AMM.578-579.441

Kumar Dash P, Kumar Parhi S, Kumar Patro S, Panigrahi R (2023) Efficient machine learning algorithm with enhanced cat swarm optimization for prediction of compressive strength of GGBS-based geopolymer concrete at elevated temperature. Constr Build Mater 400:132814. https://doi.org/10.1016/j.conbuildmat.2023.132814

Mangalathu S, Jeon J-S (2018) Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques. Eng Struct 160:85–94. https://doi.org/10.1016/j.engstruct.2018.01.008

Memon FA, Nuruddin MF, Demie S, Shafiq N (2012) Effect of superplasticizer and extra water on workability and compressive strength of self-compacting geopolymer concrete. Res J Appl Sci Eng Technol 8:407–414

Morgenthaler S (2009) Exploratory data analysis. Wires Comput Stat 1(1):33–44. https://doi.org/10.1002/wics.2

Muraleedharan M, Nadir Y (2021) Factors affecting the mechanical properties and microstructure of geopolymers from red mud and granite waste powder: a review. Ceram Int 47(10):13257–13279. https://doi.org/10.1016/j.ceramint.2021.02.009

Nhat-Duc H (2023) Estimation of the compressive strength of concretes containing ground granulated blast furnace slag using a novel regularized deep learning approach. Multiscale Multidiscip Model Exp Des 6(3):415–430. https://doi.org/10.1007/s41939-023-00154-z

Nuruddin F, Demie S, Memon FA, Shafiq N (2011) Effect of superplasticizer and NaOH molarity on workability, compressive strength and microstructure properties of self-compacting geopolymer concrete. World Acad Sci Eng Technol 75:187–194

Oliveira MLS, Izquierdo M, Querol X, Lieberman RN, Saikia BK, Silva LFO (2019) Nanoparticles from construction wastes: a problem to health and the environment. J Clean Prod 219:236–243

Oliveira MLS, Tutikian BF, Milanes C, Silva LFO (2020) Atmospheric contaminations and bad conservation effects in Roman mosaics and mortars of Italica. J Clean Prod 248:119250. https://doi.org/10.1016/j.jclepro.2019.119250

Ongsulee P (2017) Artificial intelligence, machine learning and deep learning. 2017 15th international conference on ICT and knowledge engineering (ICT&KE). pp 1–6. https://doi.org/10.1109/ICTKE.2017.8259629

Parhi SK, Panigrahi SK (2023) Alkali–silica reaction expansion prediction in concrete using hybrid metaheuristic optimized machine learning algorithms. Asian J Civ Eng. https://doi.org/10.1007/s42107-023-00799-8

Parhi SK, Patro SK (2023a) Compressive strength prediction of PET fiber-reinforced concrete using Dolphin echolocation optimized decision tree-based machine learning algorithms. Asian J Civ Eng. https://doi.org/10.1007/s42107-023-00826-8

Parhi SK, Patro SK (2023b) Prediction of compressive strength of geopolymer concrete using a hybrid ensemble of grey wolf optimized machine learning estimators. J Build Eng 71:106521. https://doi.org/10.1016/j.jobe.2023.106521

Parhi SK, Dwibedy S, Panda S, Panigrahi SK (2023) A comprehensive study on controlled low strength material. J Build Eng. https://doi.org/10.1016/j.jobe.2023.107086

Parveen N, Zaidi S, Danish M (2016) Support vector regression model for predicting the sorption capacity of lead (II). Perspect Sci 8:629–631. https://doi.org/10.1016/j.pisc.2016.06.040

Patel YJ, Shah N (2018) Enhancement of the properties of ground granulated blast furnace slag based self compacting geopolymer concrete by incorporating rice husk ash. Constr Build Mater 171:654–662

Petrovskiy MI (2003) Outlier detection algorithms in data mining systems. Program Comput Softw 29(4):228–237. https://doi.org/10.1023/A:1024974810270

Pradhan P, Dwibedy S, Pradhan M, Panda S, Panigrahi SK (2022a) Durability characteristics of geopolymer concrete—progress and perspectives. J Build Eng 59:105100. https://doi.org/10.1016/j.jobe.2022.105100

Pradhan P, Panda S, Kumar Parhi S, Kumar Panigrahi S (2022b) Effect of critical parameters on the fresh properties of Self Compacting geopolymer concrete. Mater Today Proc. https://doi.org/10.1016/j.matpr.2022.02.506

Pradhan P, Panda S, Kumar Parhi S, Kumar Panigrahi S (2022c) Factors affecting production and properties of self-compacting geopolymer concrete—a review. Constr Build Mater 344:128174. https://doi.org/10.1016/j.conbuildmat.2022.128174

Pradhan P, Panda S, Kumar Parhi S, Kumar Panigrahi S (2022d) Variation in fresh and mechanical properties of GGBFS based self-compacting geopolymer concrete in the presence of NCA and

RCA. Mater Today Proc. https://doi.org/10.1016/j.matpr.2022.03.337

Pradhan J, Panda S, Dwibedy S, Pradhan P, Panigrahi SK (2023a) Production of durable high-strength self-compacting geopolymer concrete with GGBFS as a precursor. J Mater Cycles Waste Manag. https://doi.org/10.1007/s10163-023-01851-0

Pradhan J, Panda S, Kumar Mandal R, Kumar Panigrahi S (2023b) Influence of GGBFS-based blended precursor on fresh properties of self-compacting geopolymer concrete under ambient temperature. Mater Today Proc. https://doi.org/10.1016/j.matpr.2023.06.338

Pradhan J, Panda S, Parhi SK, Pradhan P, Panigrahi SK (2024) GGBFS-based self-compacting geopolymer concrete with optimized mix parameters established on fresh, mechanical, and durability characteristics. J Mater Civ Eng 36(2):04023578. https://doi.org/10.1061/JMCEE7.MTENG-16669

Qaidi SMA, Tayeh BA, Isleem HF, de Azevedo ARG, Ahmed HU, Emad W (2022a) Sustainable utilization of red mud waste (bauxite residue) and slag for the production of geopolymer composites: a review. Case Stud Constr Mater 16:e00994. https://doi.org/10.1016/j.cscm.2022.e00994

Qaidi S, Najm HM, Abed SM, Ahmed HU, Al Dughaishi H, Al Lawati J, Sabri MM, Alkhatib F, Milad A (2022b) Fly ash-based geopolymer composites: a review of the compressive strength and microstructure analysis. Materials 15(20):20. https://doi.org/10.3390/ma15207098

Qaidi S, Yahia A, Tayeh BA, Unis H, Faraj R, Mohammed A (2022c) 3D printed geopolymer composites: a review. Mater Today Sustain 20:100240. https://doi.org/10.1016/j.mtsust.2022.100240

Qureshi HJ, Alyami M, Nawaz R, Hakeem IY, Aslam F, Iftikhar B, Gamil Y (2023) Prediction of compressive strength of two-stage (preplaced aggregate) concrete using gene expression programming and random forest. Case Stud Constr Mater 19:e02581. https://doi.org/10.1016/j.cscm.2023.e02581

Saini G, Vattipalli U (2020) Assessing properties of alkali activated GGBS based self-compacting geopolymer concrete using nano-silica. Case Stud Constr Mater 12:e00352. https://doi.org/10.1016/j.cscm.2020.e00352

Sakulich AR, Miller S, Barsoum MW (2010) Chemical and microstructural characterization of 20-month-old alkali-activated slag cements. J Am Ceram Soc 93(6):1741–1748. https://doi.org/10.1111/j.1551-2916.2010.03611.x

Saltelli A, Tarantola S, Chan KP-S (1999) A Quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41(1):39–56. https://doi.org/10.1080/00401706.1999.10485594

Schapire RE (2013) Explaining AdaBoost. In: Schölkopf B, Luo Z, Vovk V (eds) Empirical inference: festschrift in honor of vladimir N. Vapnik. Springer, Berlin, pp 37–52. https://doi.org/10.1007/978-3-642-41136-6_5

Shahmansouri AA, Bengar HA, Ghanbari S (2020) Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method. J Build Eng 31:101326

Singh S, Patro SK, Parhi SK (2023) Evolutionary optimization of machine learning algorithm hyperparameters for strength prediction of high-performance concrete. Asian J Civ Eng. https://doi.org/10.1007/s42107-023-00698-y

Smirnova O, Kazanskaya L, Koplík J, Tan H, Gu X (2021) Concrete Based on clinker-free cement: selecting the functional unit for environmental assessment. Sustainability 13(1):1. https://doi.org/10.3390/su13010135

Sutton CD (2005) 11—Classification and regression trees, bagging, and boosting. In: Rao CR, Wegman EJ, Solka JL (eds) Handbook of statistics, vol 24. Elsevier, New York, pp 303–329. https://doi.org/10.1016/S0169-7161(04)24011-1

Terrell GR, Scott DW (1992) Variable kernel density estimation. Ann Stat 20(3):1236–1265

Unis Ahmed H, Mahmood LJ, Muhammad MA, Faraj RH, Qaidi SMA, Hamah Sor N, Mohammed AS, Mohammed AA (2022) Geopolymer concrete as a cleaner construction material: an overview on materials and structural performances. Clean Mater 5:100111. https://doi.org/10.1016/j.clema.2022.100111

Unis Ahmed H, Mohammed AS, Mohammed AA (2023) Fresh and mechanical performances of recycled plastic aggregate geopolymer concrete modified with Nano-silica: Experimental and computational investigation. Constr Build Mater 394:132266. https://doi.org/10.1016/j.conbuildmat.2023.132266

Wang H, Bah MJ, Hammad M (2019) Progress in outlier detection techniques: a survey. IEEE Access 7:107964–108000. https://doi.org/10.1109/ACCESS.2019.2932769

Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14(1):1–37. https://doi.org/10.1007/s10115-007-0114-2

Xiong Z, Cui Y, Liu Z, Zhao Y, Hu M, Hu J (2020) Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. Comput Mater Sci 171:109203. https://doi.org/10.1016/j.commatsci.2019.109203

Zhang Y, Bai Z, Zhang H (2023) Compressive strength estimation of ultra-great workability concrete using hybrid algorithms. Multiscale Multidiscip Model Exp Des 6(3):389–400. https://doi.org/10.1007/s41939-023-00145-0

Zou B, Wang Y, Nasir Amin M, Iftikhar B, Khan K, Ali M, Althoey F (2023) Artificial intelligence-based optimized models for predicting the slump and compressive strength of sustainable alkali-derived concrete. Constr Build Mater 409:134092. https://doi.org/10.1016/j.conbuildmat.2023.134092