University of Nottingham Ningbo China
Department of Civil Engineering

Master Dissertation

Prediction of Mechanical Strength through Machine Learning Models for Alkali-Activated Concretes

By

Hanfeng Su
20619014

Table of Contents

Table of Figures

Table of Tables

Abstract

Alkali-activated concrete has become an alternative construction material to overcome the environmental impacts of traditional Portland cement concrete, with its better performance, environmental friendliness, mechanical properties, and durability. In recent years, advanced machine learning techniques have been used to predict the mechanical properties of alkali-activated concrete; however, prediction accuracy still needs to be enhanced. In this innovative study, particle packing theory was used in mechanical properties prediction through machine learning techniques, where packing density was introduced as a new input variable in machine learning models. The dataset used in machine learning models training was collected from an experiment study, which involved 99 data points of compressive strength and 33 data points of flexural strength test. This study presents the possibility of predicting the compressive strength and flexural strength of alkali-activated concrete through mix proportions using four machine learning algorithms: Random Forest, Extreme Gradient Boosting, Support Vector Machine, and K-Nearest Neighbors. The results showed that all machine learning models performed with reasonable accuracy in the training set and testing set. Within four models, Extreme Gradient Boosting showed the best performance. According to the influence analysis for input variables, packing density showed an intermediate effect of both compressive strength and flexural strength, proving that packing density can be an influential factor in the mechanical properties prediction of alkali-activated concrete. This innovative research is able to reduce the materials, time, and costs in experiments, and it will also be beneficial in alkali-activated concrete mix design in the concrete industry.

1. Introduction

Concrete is a fundamental material in construction sites and plays a crucial role in developing infrastructure worldwide. Its versatile properties make it indispensable in various construction applications, from buildings and bridges to roads and dams. With its widespread usage around the world, every person in the world will produce approximately 1 ton of concrete on average. Modern concrete is a mixture of Portland cement, aggregates, and water (Aïtcin, 2000). Cement's versatility is evident in its wide range of applications across various sectors. In residential and commercial construction, it is used to create solid and durable structures, including foundations, walls, and floors. In infrastructure projects, cement is a key component in the construction of roads, bridges, and tunnels, generating the necessary strength and durability to resist heavy weight and environmental stress. Additionally, cement is used in the production of precast concrete products, such as pipes, blocks, and panels, which are essential for efficient and cost-effective construction (Mehta and Monteiro, 2006). Its ability to set and harden underwater makes it ideal for marine and underwater construction projects as well. However, the extensive use of cement also raises environmental concerns, including habitat destruction due to the quarrying of raw materials, energy-intensive production processes, and waste generation (Durastanti and Moretti, 2020) (Ivanov et al., 2018). With the high speed development in recent decades, the negative impacts of concrete usage have attracted more attention from researchers, especially in the manufacturing process of cement concrete.

The manufacturing process of cement involves several stages, each crucial in producing the fine, powdery material used in construction. Initially, source materials for instance, clay, limestone, and sand are extracted and crushed to

form a raw mix. This mixture is then heated in a rotary kiln at high temperatures, around 1450°C, in a process known as calcination. During calcination, the raw materials undergo chemical changes, resulting in the formation of clinker. Finally, cement was produced by ground cooled clinker with gypsum. The clinker is then cooled and ground with gypsum to produce the final product, cement. This intricate process ensures the quality and consistency required for cement to perform effectively in construction applications (Taylor, 1997). However, cement manufacturing is also associated with significant environmental impacts, mainly as a result of the great energy consumption and CO_2 emissions during calcination. Cement manufacturing is highly energy-intensive, and its industry consumes roughly 2–5% of the world's total energy demand (Omer, 2014). The main factors contributing to the high elevated CO_2 emissions associated with OPC production are identified as (1) the calcination process of limestone, a principal component, resulting in the generation and emission of CO_2; and (2) the intensive energy requirements for manufacturing, which involve heating raw materials in a rotating kiln at temperatures exceeding 1400 $°C$ (Gartner, 2004). Ordinary Portland cements are commonly used as the binding agent in concrete compositions. The production of Ordinary Portland cement accounts for an estimated 5–7% of worldwide anthropogenic CO_2 emissions (Turner and Collins, 2013). Nevertheless, the increasing demand for concrete, driven by rapid urbanization in recent years, brings significant environmental challenges and has emerged as a primary global concern. When cement and water are mixed together, a series of complex chemical reactions occur, collectively known as hydration. These reactions are essential for the setting and hardening of cement to form concrete. The primary compounds in cement, including tricalcium silicate (C_3S) and dicalcium silicate (C_2S), react with water to form calcium silicate hydrate ($C-S-H$) and calcium hydroxide ($Ca(OH)_2$). The $C-S-H$ gel is primarily responsible for the strength and

durability of the hardened concrete. At the same time, calcium hydroxide contributes to its alkalinity, providing a protective environment for embedded steel reinforcement (Neville, 1995). The hydration process continues over time, leading to increased strength and reduced permeability of the concrete, which enhances its performance in various construction applications. Although cement is an efficient product for creating paste in concrete, the high emission of manufacturing procedures causes many environmental issues, so developing an alternative cementitious material is highly urgent.

To reduce the environmental effects of Ordinary Portland cement concrete, alkali-activated concrete, also known as geopolymer concrete, is a sustainable replacement. Based on the superior properties and environmentally friendly impact of alkali-activated concrete, the researcher's attention was attracted when these alternative construction materials to the traditional Ordinary Portland cement concrete started to be used in construction. Glkhovsky (1959) was the first man to investigate using low elemental or accessible calcium alumino-silicate sources and alkaline activators to produce binders. Alkali-activated concrete is created by activating alumino-silicate precursor materials, such as fly ash, slag, or natural pozzolans, with an alkaline solution like sodium hydroxide or sodium silicate, leading to a binder that can rival or even surpass traditional cement in various applications. This reaction forms a gel that binds the aggregates, similar to the role of cement in traditional cement concrete. The chemical reaction underpinning alkali-activated concrete involves the dissolution of aluminosilicate materials in an alkaline solution, leading to the formation of a gel-like network primarily composed of sodium or potassium aluminosilicate hydrate. This process, known as geopolymerization, forms a three-dimensional network structure that provides AAC with its superior mechanical properties. During geopolymerization, the aluminosilicate materials

dissolve to release silicate and aluminate species, which then react with the alkaline activator to form a rigid, stable matrix (Provis and Van Deventer, 2013). Previous research already proved that the alkali-activated has the ability to reduce the environmental impact of concrete manufacturing.

Compared with traditional Portland cement concrete, the greatest advantage of alkali-activated concrete is the significant improvement in the environmental influence. The production of AAC generates significantly less CO_2 compared to OPC, mainly because it does not involve the calcination of limestone, a process responsible for substantial carbon dioxide emissions in traditional cement manufacturing (Habert and Ouellet-Plamondon, 2016). By utilizing industrial by-products such as fly ash and slag, AAC not only reduces greenhouse gas emissions but also helps in recycling waste materials, thereby mitigating the impact on landfills and conserving natural resources (Provis and Van Deventer, 2013). The application of alternative alkali-activated binders provides a sustainable solution for the environmental problems above and also offers a significant method to dispose of industrial waste, fly ash (FA), and ground granulated blast-furnace slag (GGBS), for example (Khan et al., 2021). According to a life cycle assessment, Weil et al. reported that FA-based alkali-activated concrete provides about a 70% decrease in the global warming potential compared with Ordinary Portland cement concrete (Weil et al., 2009). Another life cycle assessment investigation in Australia also shows that the alkali-activated method performed better in terms of carbon impact. McLellan et al. (2011) studied some cases of geopolymer concrete mixtures based on typical Australian feedstocks and concluded that there is a potential for a 44–64% decrease in CO_2 emissions compared with Ordinary Portland cement concrete. Mechanically, alkali-activated concrete demonstrates superior performance compared to OPC concrete. Published research has shown that alkali-activated

concrete performs better in mechanical strength, durability, and resistance to chemical attacks, for example. The high compressive strength of alkali-activated concrete is attributed to the dense microstructure formed during the geopolymerization process, which is more resistant to cracking and degradation over time (Caijun Shi, 2006). Additionally, the durability of alkali-activated concrete in hazardous environments, such as exposure to acids, sulfates, and chlorides, and the performance of alkali-activated concrete significantly improved when compared with OPC. This resistance is crucial for infrastructure exposed to harsh conditions, resulting in lower maintenance costs and longer service life (Bernal and Provis, 2014b). The superior mechanical properties make alkali-activated concrete an attractive choice for applications that require high strength concrete materials, not only including structural components, but also marine structures, chemical containment systems, etc. Because of the improvements in mechanical properties and reductions in environmental impact, nowadays alkali-activated concrete has become an alternative material to traditional Portland cement concrete in many construction applications. The higher compressive strength and superior durability of alkali-activated concrete make it suitable for comprehensive applications, while its lower carbon footprint and use of industrial waste materials address critical sustainability concerns in the construction industry.

Alkali-activated concrete (AAC) has emerged as a promising alternative to Ordinary Portland Cement (OPC) concrete, primarily due to its lower environmental impact and potential for utilizing industrial by-products such as fly ash and slag, significantly reducing CO_2 emissions by more than 68% when compared to OPC (Habert et al., 2011). Additionally, AAC has demonstrated superior mechanical properties, including rapid strength development and higher early compressive strength, making it suitable for applications requiring

accelerated construction schedules (Caijun Shi, 2006). Research indicates that AAC can achieve equivalent or even higher compressive and flexural strength compared to OPC, particularly when optimized mix proportions and curing conditions are applied (Bernal and Provis, 2014b). Furthermore, AAC is known for its exceptional durability, exhibiting enhanced resistance to aggressive environmental conditions such as sulfate attack, chloride penetration, and acid degradation, which are common challenges in conventional OPC structures (Caijun Shi, 2006). The improved chemical resistance of AAC is primarily due to its lower calcium content, which minimizes the risk of deleterious reactions, such as the formation of expansive ettringite under sulfate exposure (Bernal and Provis, 2014b). Despite these advantages, several challenges limit the widespread adoption of AAC in the construction industry. One of the most significant disadvantages is its cost, as the production of AAC often requires expensive alkali activators, leading to material costs estimated to be two to three times higher than OPC concrete. Moreover, the availability and quality control of raw materials, particularly fly ash and slag, vary significantly across different regions, affecting the consistency and performance of AAC mixes (Caijun Shi, 2006). Another critical disadvantage is the higher shrinkage and carbonation rate of AAC compared to OPC concrete, which can lead to increased cracking and long-term durability concerns if not adequately addressed through mix design optimization and curing techniques (Provis and Van Deventer, 2013). Additionally, the lack of standardized mix design procedures and long-term performance data hinders the acceptance of AAC as a mainstream construction material, as regulatory frameworks and building codes still predominantly cater to OPC-based systems. These limitations highlight the need for further research and development to optimize AAC formulations, improve cost efficiency, and establish standardized guidelines for its application in structural and non-structural elements. In conclusion, while AAC presents a viable and sustainable

alternative to OPC concrete with superior mechanical properties and durability, its widespread adoption is constrained by economic and practical challenges. Addressing these issues through technological advancements, regulatory support, and improved material standardization will be crucial for promoting AAC as a mainstream construction material. Future studies should focus on enhancing the cost-effectiveness of AAC production, refining mix design methodologies, and evaluating its long-term performance in real-world applications. By overcoming these challenges, AAC has the potential to play a pivotal role in the transition towards more sustainable and resilient construction practices.

Particle packing theory is a fundamental concept in concrete materials science, which plays a vital role in understanding and optimizing the properties of concrete. This theory focuses on the arrangement and distribution of particles within concrete to achieve the densest packing between different sizes of aggregates and cement particles. By minimizing voids to enhance material properties, particle packing theory can significantly improve concrete performance. This theory explores the optimal arrangement of particles of optimized different sizes and shapes particles in concrete mixture to achieve maximum density and minimal void space within concrete materials. In the context of concrete, this theory is applied to the mixture of cement, aggregates, and other additives. The primary objective is to enhance the concrete's workability, strength, durability, and overall performance by reducing the reduce the voids between aggregates. By correctly combining different sizes and shapes of particles in the mixture, voids between big particles have been filled with small particles, which will cause concrete to be thoroughly compacted with all sizes of particles and get the lowest void content inside.

The application of particle packing theory in concrete mix design can significantly enhance the mechanical and durability of concrete materials. Firstly, it improves workability by reducing friction between particles in concrete materials (Fennis and Walraven, 2012). Secondly, higher packing density means the minimum space of voids in concrete, which will increase compressive strength and make the concrete able to resist environmental factors, freeze-thaw cycles, chemical attacks, and abrasion, for example (Kumar and Santhanam, 2003). Previous research has significantly explained the knowledge of particle packing theory in concrete. Pioneering work published by Powers (1968) in 1969, which demonstrated the properties of fresh concrete, his research highlighted the importance of particle size distribution and its effect on concrete's rheology and strength. Later, in 1999, Larrard (1999) developed a comprehensive model for optimizing the particle packing density of granular materials, which has been widely adopted in concrete mix design. His work provided a theoretical foundation for understanding how different particle sizes interact and the connection between the packing density of each size of aggregate and concrete performance (De Larrard, 1999). Moreover, particle packing theory minimizes the voids between different sizes of aggregates to achieve efficient use of raw materials in concrete. This not only reduces the costs in the construction industry but also decreases the environmental impact associated with cement production, which is the mean origin of carbon dioxide emissions. Additionally, the decreasing of voids inside the concrete also improves the internal arrangement of the concrete, reduces the possibility of shrinkage and creep in concrete, and improves the combination of the aggregates, resulting in the performance of concrete in mechanical properties and durability (Zhu et al., 2020).

Machine learning has developed rapidly in recent years; this development has

been caused by upgrading in computational power, the feasibility of large datasets, and the expansion of computing algorithms; it transforms from a research area only in computer science into a powerful tool in the engineering field. Machine learning, which is a sub area of artificial intelligence, its purpose is to build a model with algorithms from existing datasets and recognize and learn the relationships between input variables to create a predictive model used to make decisions without human intervention. Adaptation and efficiency are the most significant advantages of machine learning techniques, especially in data science. Traditional data analysis methods often have difficulties with the complexity and volume of developing engineering datasets. However, machine learning algorithms can overcome these problems and discover the relationship between each pattern in the dataset. One of the primary reasons machine learning techniques are highly recommended when dealing with the amount of data, they are able to analyze the correlation between variables, and do classification and regression based on the data that has been given to the models. This ability is achieved through various methods, for instance, unsupervised learning, supervised learning, and reinforcement learning. A model is trained using a labeled dataset, to make predictions or classifications based on input variables in supervised learning. Unsupervised learning refers to identifying hidden patterns or groupings within data without pre-labeled outcomes, while reinforcement learning optimizes actions based on feedback from interactions with a given environment (Goodfellow et al., 2016). As an example, neural networks, especially the deep learning models, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), have an impressive ability to find out the non-linear in datasets and make prediction models, thus making them acceptable for a comprehensive data science applications (LeCun et al., 2015).

Machine learning involves a variety of techniques, each offering unique advantages for predictive modeling. Among the most widely used are decision trees, support vector machines, ensemble methods, etc. For instance, decision trees can efficiently handle both continuous and categorical variables and offer a clear and understandable depiction of decision-making processes. (Breiman, 2001). Support vector machine is another robust machine learning algorithm, especially effective in high-dimensional spaces. Support vector machines can provide robust predictions with limited data, making them suitable for applications where data availability is a constraint. The ability of the Support vector machine to maximize the margin between different classes in the data leads to high generalization performance, which is essential for accurate predictive modeling (Cortes and Vapnik, 1995a). Ensemble methods, such as random forests and gradient boosting machines, incorporate the predictions of multiple base models to improve precision and robustness. Ensemble methods are especially effective in dealing with complex and noisy datasets, a common challenge in many scientific and engineering domains. These machine learning algorithms have shown significant improvements in use, leveraging the strengths of different models for predictive performance and mitigating individual weaknesses (Friedman, 2001). Other algorithms such as the K-nearest neighbors algorithm, is a simple yet effective algorithm that makes predictions based on the proximity of data points. K-nearest neighbors are particularly useful for classification tasks and are valued for their simplicity and ease of implementation. However, its performance can degrade with high-dimensional data and large datasets (Cover and Hart, 1967). Layered structures of interconnected nodes (neurons) make up neural networks, this algorithm is modeled after the structure and operations of the human brain. These networks can catch complicated non-linear relationships in the data, which apply them the ability to become a powerful tool for a comprehensive scope of regression

tasks. Although neural networks can become computing intensive and consume a huge dataset in the training process, they are highly versatile and have been profitably applied in different domains (Bishop and Nasrabadi, 2006). The success of machine learning models, however, hinges on the quality and the amounts of data used for training. Datasets with high quality that accurately represent the variability and complexity of the natural world are essential for developing robust predictive models. Machine learning techniques offer powerful tools for predictive modeling, capable of analyzing complex datasets and identifying complex patterns. The continual research and development in machine learning promises to further expand its capabilities and applications, enhancing decision making processes and predictive accuracy across various fields. The integration of machine learning into traditional methodologies opens new possibilities for innovation and efficiency, paving the way for advancements in engineering.

Machine learning (ML) significantly enhances the prediction of alkali-activated concrete (AAC) properties by capturing complex nonlinear relationships between mix composition, curing conditions, and mechanical performance, offering greater accuracy than traditional empirical methods. ML algorithms such as artificial neural networks (ANN), support vector machines (SVM), random forests (RF), and extreme gradient boosting (XGBoost) can process large datasets to predict compressive strength, flexural strength, and durability with high precision, reducing the reliance on time-consuming and costly laboratory testing (Le et al., 2024). By analyzing vast experimental data, ML models identify key influencing factors, enabling the optimization of mix proportions to enhance performance while minimizing material costs. Additionally, ML applications extend to durability assessments, where predictive models estimate AAC's resistance to sulfate attack, chloride penetration, and

carbonation—critical parameters for ensuring long-term structural performance. Despite these advantages, challenges such as limited high-quality experimental datasets, variations in raw materials, and the interpretability of complex ML models hinder widespread adoption. Future research should focus on expanding databases, integrating hybrid AI techniques, and improving model transparency through explainable AI (XAI) methods to increase trust and applicability in the construction industry. As ML continues to evolve, it has the potential to revolutionize AAC research, leading to more sustainable and high-performance concrete materials.

This research tries to introduce packing density as a novel parameter in compressive strength and flexural strength prediction through machine learning models. Four machine learning models were chosen in this investigation, random forest, extreme gradient boosting, support vector machine, and k-nearest neighbor. The models were trained by a dataset collected from laboratory experiments. The goal of this investigation is to increase the precision of alkali-activated concrete's compressive and flexural strength predictions. It has the potential to significantly improve the understanding of the design and uses of alkali-activated concrete mixes in construction applications.

2. Literature review

2.1 Alkali-activated concrete

Alkali-activated concrete performs better than traditional Portland cement, not only in environmental aspects but also in mechanical properties and durability. Compressive strength is the most important mechanical property considered in

concrete materials. The development of compressive strength of alkali-activated concrete is much faster than Portland cement concrete. Angulo-Ramírez et al. (2017) reported that the alkali-activated hybrid cement, which is a mixture of a GGBS and Portland cement in proportion 80% and 20%, respectively, got 4.5 and 10.8 times higher compressive strength than the 100% Portland cement mixture. Alkali-activated concrete mixed with a binder content of 300, 400, and 500 $^{kg}/_{m^3}$ showed around 20 MPa compressive strength higher than the same binder content Portland cement concrete after 28 days and 90 days curing under conditions where the temperature of 25 ± 2 °C, and RH = 65 ± 5% (Bernal et al., 2011). Marvila et al. (2023) used blast furnace slag as cementitious material and activated by sodium hydroxide with different mass concentrations to make a $50 \times 100mm$ cylinder concrete sample. The compressive strength test at 7 days and 28 days demonstrates that the alkali-activated concrete presents a superior behavior compared with traditional Portland cement concrete. Based on the comprehensive review of Wang et al. (2020), the durability of alkali-activated concrete is better than Portland cement concrete in general.

Several factors affected the compressive strength of alkali-activated concrete, for example, the composition of the alkali-activator, the solution to binder ratio, the dosage of chemical components in the activator, the water to cementitious materials ratio, the modulus in the activator solution (mass ratio of SiO_2 to Na_2O), the type and dosage of the precursor materials, the curing conditions and curing time, etc. (Ding et al., 2016). The compressive strength of alkali-activated concrete increased with the increases in the molarity of $NaOH$ solution in the activator. Results from Singh et al. (2023) demonstrate the compressive strength of alkali-activated concrete increases about 81.4% with the molarity of $NaOH$ solution changed from 12 M to 14 M. According to an

experiment study published by Farhan et al., the compressive strength of 35.91 MPa for regular strength and 65.28 MPa for high performance geopolymer concrete, when the concentration of $NaOH$ solution is 12 M and 14 M (Farhan et al., 2019). Similar results were provided by previous research. (Aliabdo et al., 2019). In addition, other researchers investigated the influence of Na_2O content on the compressive strength of FA-based alkali-activated concrete. When the Na_2O content increases from 4% to 6%, the 28 day compressive strength increases to 348%. Nevertheless, when Na_2O content increases from 6% to 8% and 8% to 10%, the rate of compressive strength development drops to 181% and 115%, respectively (Cho et al., 2017).

Alkali-activated concrete is controlled by complex chemical mechanisms involving the dissolution, polymerization, and gelation of aluminosilicate materials under alkaline activation. The key reaction process, geopolymerization, begins with the breakdown of silica (SiO_2) and alumina (Al_2O_3) from precursor materials, such as fly ash or slag, facilitated by alkaline activators like sodium hydroxide ($NaOH$) and sodium silicate (Na_2SiO_3). This dissolution process releases silicate (SiO_4) and aluminate (AlO_4) species into the solution, where they undergo hydrolysis and polycondensation to form oligomeric species. These oligomers then polymerize into a three-dimensional sodium-aluminosilicate-hydrate ($N-A-S-H$) gel, which serves as the primary binding phase in AAC (Provis and Van Deventer, 2013). Sodium silicate plays a critical role by enhancing the dissolution of aluminosilicate precursors, increasing silicate availability, and promoting a denser geopolymer matrix, leading to improved mechanical performance and durability (Palacios et al., 2021). In addition to ($N-A-S-H$) gel, if calcium-rich materials such as ground granulated blast furnace slag (GGBS) are present, secondary reactions occur, leading to the formation of calcium-alumino-silicate-hydrate ($C-A-$

$S-H$) gel, which further enhances strength and reduces porosity (Bernal and Provis, 2014a). The SiO_2/Al_2O_3 ratio significantly influences the reaction kinetics, gel structure, and final material properties, where higher silica content promotes a well-crosslinked network, improving durability and chemical resistance (Davidovits, 2008). Furthermore, the Na_2O/SiO_2 ratio of the activator controls the viscosity of the geopolymer paste, affecting workability and setting time. Optimizing these parameters, alongside controlled curing conditions, ensures complete geopolymerization and enhances alkali-activated concrete's long-term mechanical stability, making it a promising alternative to conventional cementitious materials (Provis and Bernal, 2014).

In alkali-activated concrete, the activator is a main component that will influence the compressive strength, the effect can be demonstrated by a few ratios, M_s (modulus in activator also known as the ratio of SiO_2 to Na_2O), the ratio of sodium silicate solution to sodium hydroxide, etc. For the modulus in the activator, the compressive strength of alkali-activated concrete under constant curing conditions has been studied. With modulus value increases from 0.8 to 1.4, the compressive strength increased, however, they observed that the results when modulus is 2.0, the compressive strength decreased, and the value is much lower than those in 0.8 and 1.4 of modulus case (Cho et al., 2017). Soutsos (2016) found that the optimum range of values of the alkali modulus is determined by any value above and below, which will cause the strength to decrease. As the alkali dosage increases, the "sweet spot" expands to accommodate a higher alkali modulus. Generally, the best compressive strengths are obtained at alkali doses that have been studied in tests when the alkali modulus falls between 1 and 1.25. The reduction in available silica that can participate in the "reorganization–gelation–polymerization" steps of the alkali-activated concrete formation—that is, the development of a longer,

denser, and more integrity polymer chain—seems to be the cause of the compressive strength decrease with increasing alkali modulus. The decrease in compressive strengths at low modulus with a value of 0.5, can be attributed to the reaction system's pH value being lower when sodium silicate solution is the only alkali activator. For the ratio of sodium silicate to sodium hydroxide, the compressive strength of alkali-activated concrete increases with the increase of the ratio of sodium silicate to sodium hydroxide, up to 2.5, and then decreases (Joseph and Mathew, 2012).

In fly ash-slag based alkali-activated concrete, the compressive strength is influenced by the ratio of fly ash to ground granulated blast-furnace slag (GGBS). Lee and Lee (2013) demonstrated that the compressive strength of specimens with varying replacement ratios of slag to fly ash increased as the slag replacement ratio rose from 10% to 15%. Specifically, the compressive strength improved from 15.5 MPa to 23.0 MPa after 28 days of curing. Similar results were reported in the previous experiment; when the fly ash to GGBS ratio changed from 100:0 to 85:15, the compressive strength was 1.11, 1.16, 1.09, and 1.109 times higher when the specimens after 3 days, 7 days, 28 days, and 91 days of curing (Rashad, 2013). This phenomenon has been explained based on the chemical reaction in alkali-activated concrete, and the compressive strength was enhanced by incorporating slag, which contains a significant amount of CaO, into the mixture. This improvement is attributed to the formation of an amorphous $Ca-Al-Si$ gel, facilitated by the presence of CaO, which strengthened the geopolymer. This finding is consistent with results reported in the previous study (Yip, 2004). According to Kumar and his colleagues, the compressive strength increasing with the increase of GGBS could be on account of the formation of gel phases $C-S-H$ and $A-S-H$ and the compactness of the microstructure (Kumar et al., 2010). The chemical

composition of cementitious materials plays an essential part in compressive strength performance. Although the chemical composition of different slag is nearly the same, the chemical components content of each chemical composition in slag is very different (Fu et al., 2023). Shi et al. found that the SiO_2 and CaO contents in slags from different countries were similar, while the Al_2O_3, MgO, and TiO_2 contents varied significantly. These compositional differences affect the slag's reactivity and, consequently, its hydration process (Caijun Shi, 2006). The investigation reported that the quick early age compressive strength development can be attributed to the reaction of additional calcium bearing compounds present in the GGBS (Nath and Sarker, 2014). Another factor that affects the compressive strength of alkali-activated concrete is the binder content to activator ratio; based on Ibrahim et al., in the experiment study, the compressive strength increased by about 6.75% when the activator to binder ratio increased from 0.467 to 0.5. However, when this ratio still increased from 0.5 to 0.643, it showed a negative influence on compressive strength. They also suggest when the sodium silicate to sodium hydroxide ratio is 2.75, the best activator to binder ratio is 0.5 to achieve the biggest compressive strength (Ibrahim et al., 2017). Water to solids ratio also known as water to cementitious ratio in alkali-activated concrete, this ratio is similar to water to cement ratio in traditional Portland cement concrete. This ratio can be defined as the mass ratio of water in the activating solutions and additional water to the precursors (fly ash and GGBS) and alkali solids (sodium silicate and sodium hydroxide in the activator solution). The water to cementitious materials ratio has an adverse relationship with the compressive strength of alkali-activated concrete. The effect is more observable for pure fly ash alkali-activated concrete, a mixture with 100% fly ash, where the compressive strength and the ratio of water to cementitious materials are practically linear (Rafeet et al., 2017). Analogous tendencies have been

demonstrated in the experiment study (Joseph and Mathew, 2012). The addition of GGBS likely reduces the effect of water to cementitious materials ratio on the compressive strength, and this direction has also been proved by other published studies (Hung and Chang, 2013). Detrimental influences of water inclusion in 80:20 fly ash to GGBS ratio mixes were more obvious for higher water to solids ratios between 0.39 to 0.41 than they were for lower water to solids ratio in a range of 0.35 to 0.37, where there was no substantial decrease. For concrete mixtures with the fly ash to GGBS ratio of 60:40, greater values of water to solids will cause compressive values to show low values. When the water to solids ratio is less than 0.41, the compressive strength only shows a moderate reduction. When the water to solids ratio is higher than 0.41 and the paste content is equal to 30% of the entire mixture, compressive strength is reduction about 40%. However, when compressive strength was higher than 60 MPa, the increase in the water to solids ratio from 0.42 to 0.48 for mixes with fly ash to GGBS ratio of 30:70 did not seem to have a significant effect on the compressive strengths. And 24 hours compressive strengths as high as 30 MPa were accomplished for water to solids ratio of 0.42 (Rafeet et al., 2017).

Curing variations in the alkali-activated concrete, such as curing time, temperature of curing, and humidity of curing, affect the compressive strength. According to analyses of experimental results, Gomaa et al. found that curing temperatures have a positive trend with compressive strength. This is to be expected as fly ash dissolving kinetics and reaction product precipitation kinetics both improve with higher curing temperatures (Gomaa et al., 2021). Zhang et al. (2023) investigated the compressive strength under combined curing conditions (steam curing for 48h at 90℃, ambient curing until specific ages after steam curing) at the age of 3 days can achieve the compressive

strength value reached at 28 days under normal curing (≥95% relative humidity and the temperature is 20 ± 2 $°C$). The accelerated advancement of the compressive strength of ultra-high strength alkali-activated concrete specimens can be attributed to the steam curing for two days at 90 °C.

2.2 Particle packing theory in concrete

A concrete mixture has about 60% to 70% volumetric proportion of aggregate phase. Particle packing and size distribution in concrete highly affect the volume of voids between particles. Particle packing theory is a method to find out the optimum proportions of blended coarse aggregate and fine aggregate in the concrete mixture. Scientists advised that the concrete mix can be considered a combination of aggregate particles and cement paste, optimal aggregate arrangement can improve the volume of extra paste to increase the concrete performance (Powers, 1968). A packing model which able to maximize the strength of ultra-high-performance concrete was invented by researchers in 1994 (De Larrard and Sedran, 1994). A few famous researchers were continually advancing the aggregate distributions through particle packing theory, for example, the Modified Toufar Model (MTM), JD Dewar Model (JDD), and Compressible Packing Model (CPM); these models are combined models under a discrete approach (Toufar et al., 1976) (De Larrard, 1999) (Dewar, 1999). To minimize the number of cement paste and tightly packed aggregate and maximize the filling density, a multi-component aggregate mixture is used. A minimal number of slurries is only needed to permeate the voids within the massively packed aggregates, which lets the concrete mixture achieve the strength and durability required. Friction between the particles in the concrete mixture was generated by direct contact (Fuller and Thompson, 1907) (Kwan and Mora, 2001). Pallapothu et al. (2023) show the feasibility of using packing

density as a mean input variable in machine learning to forecast the compressive strength of traditional concrete. The analytical solution to maximize packing density by optimizing aggregate proportions was difficult to derive, however this work solved the problem. It offered a user-defined model that, by utilizing machine learning approaches, decreased the mathematical analysis for particle packing optimization.

The packing density of aggregates, which is known as the ratio of the volume of aggregates to the volume of voids between aggregates, is an essential factor in determining aggregate distribution degree in the concrete mixture; the packing density of aggregates in a concrete mixture can significantly influent the compressive strength of concrete. The compressive strength of recycled aggregate concrete increases with the packing density increasing (Li et al., 2017). Based on the effects of the study of aggregate packing density on the workability, rheological, and mechanical characteristics of self-consolidating concrete published by Khayat et al. The results illustrated that when the packing density of aggregate is 0.8, the superplasticizer usage is reduced by 40%, significantly (Khayat et al., 2000). Similar research was conducted; in this research, the same paste composition and paste volume were kept at a constant value; concrete specimens were produced with three different packing densities. For each 150 mm cube specimen group with the same packing density value, the compressive strength was measured after 28 days of curing. The observations showed that the packing density raised from 0.64 to 0.68, and the compressive strength of test samples increased from 40.5 to 45.6 $^{N}/_{mm^2}$. The reason for this is that the increased interlocking of the aggregate as the packing density increases and the porosity in concrete reduces, which causes higher strength. Additionally, with a development in packing density, the paste, more than the porosity, will help achieve advanced compatibility, which results

in higher strength. However, the packing density significantly affects the compressive strength of concrete (Nanthagopalan and Santhanam, 2012).

Niyazudddin and Umesh B (2023) investigated the compressive strength, split strength, flexural strength, and durability of geopolymer concrete incorporated with particle packing theory. The research shows that the geopolymer concrete mixtures with particle packing theory have a critical increase in compressive strength and durability, compared with the mixtures without particle packing theory. The compressive strength with the mix using particle packing theory is 40% higher than the mix without particle packing theory. The geopolymer concrete can achieve around 70Mpa after 28 days of curing. Karadumpa and Pancharathi (2021) developed a mixed design method using particle packing theory for composite cement concrete and optimizing the fly ash and GGBS contents. They found that the compressive strength and packing density are directly correlated, at water to binder ratios of 0.45, 0.50, and 0.55, the accuracy values $R^2$are 0.932, 0.957, and 0.979 respectively, show a significant positive relationship between the mechanical properties of sustainable concrete and the packing density of aggregates. At the same time, they found that the workability of concrete mixes was found to decrease drastically with an increase in the packing density of the aggregate system. The same result of workability affected by the packing density of concrete had been observed by Nanthagopalan et al. (Nanthagopalan et al., 2008). The previous research proved that the particle packing theory plays a vital role in alkali-activated concrete compressive strength performance, and packing density closely correlates with compressive strength.

2.3 Machine learning in Alkali-activated concrete

The high speed advancement of artificial intelligence (AI) technology has led to its integration across diverse sectors for enhancement purposes. Numerous researchers have applied AI to the field of civil engineering construction materials. Machine learning models can make good accurate predictions based on a big dataset without understanding the complex physical mechanism. Many publications have shown the feasibility of predicting the mechanical properties of traditional Ordinary Portland cement concrete by machine learning algorithms in recent decades, for example, random forest (RF), support vectors machine (SVM), gradient boosting (GB), etc. (Cheng et al., 2012, Chou and Pham, 2013). To forecast the compressive strength of high strength concrete, Al-Shamiri et al. used an extreme learning machine (ELM) and backpropagation ANN (BP) with 324 data records from laboratory experiments. Five parameters were chosen as input values in the machine learning model: fine aggregate, coarse aggregate, water, cement, and water reducer. The simulation result shows the strong potential for using ELM and BP to forecast the compressive strength of high strength concrete (Al-Shamiri et al., 2019). A laboratory dataset includes 99 data points for self-compacting concrete. Python machine learning models were used to train the Lasso regression, Linear regression, Ridge regression, decision tree regression, multi-layer perceptron regression, and random forest regression; the results show the feasibility of the machine learning model for self-compacting concrete compressive strength prediction (Rajakarunakaran et al., 2022). Tavares et al. present an innovative and efficient methodology, this methodology not only optimizes the mixture design of ultra-high-performance-concrete but also reduces experimental runs, enabling concrete manufacturers to quickly create efficient models using their own experiment data and raw materials. The dataset for machine learning

models training and testing was collected from orthogonal arrays. Machine learning techniques in K-Nearest Neighbors and random forest are able to produce Performance Density Diagrams in this methodology, results demonstrate that the Performance Density Diagrams from machine learning models predicted the trends, magnitude, and ranking of most mixtures stored with high efficiency (Tavares et al., 2022). Explainable Boosting Machine performed with a high accuracy with a $R^2 = 0.93$ which was trained and tested by a comprehensive dataset has 1030 compressive strength data points. The significant time consumption reduction in the machine learning model creating procedure can be attributed to the use of Bayesian optimization algorithms to iteratively build the algorithmic and hyperparametric model iteratively, also identifying the optimal point in the search space (Liu and Sun, 2023).

Machine learning models are also an efficient tool for predicting the mechanical properties of alkali-activated concrete (Huo et al., 2022). Thomas and Peethamparan used a linear regression algorithm to analyze the optimum mix design for slag-based geopolymer and class C fly ash concrete material (Thomas and Peethamparan, 2017). In the research published by Sun et al., 193 strength data of alkali-activated concrete were collected to build a random forest model for compressive strength prediction; the training set and testing set have 0.96 and 0.92 accuracy of prediction, respectively (Sun et al., 2023b). According to Gomaa et al., their research also created a random forest model with around 200 data to predict the compressive strength of alkali-activated concrete. Results showed the model, after being meticulously trained and optimized was able to perform with high accuracy in prediction, with the Pearson correlation coefficient, $R \gtrsim 0.95$ (Gomaa et al., 2021). The Bayesian linear regression algorithm can be used as an innovative technique to investigate the actual influence of the chemical interaction between raw materials on the mechanical

properties of fly ash based geopolymer concrete. For the training set and testing set, the model's R factor values were 0.89 and 0.826, respectively (Toufigh and Jafari, 2021). Other machine learning models can be applied to alkali-activated concrete compressive strength prediction as well, gradient boosting machine, decision tree, and support vector machine were trained to predict the compressive strength of metakaolin based geopolymer concrete. Test results demonstrated that the gradient boosting machine has exceptional performance, with a mean absolute error of 1.615 Mpa and a coefficient of determination of 0.983 (Afzali et al., 2024). Peng and Unluer illustrated three different machine learning algorithms, which are Support Vector Machine (SVM), Extreme Learning Machine (ELM), and Backpropagation Neural Network (BPNN), have the ability to predict 28 days compressive strength of geopolymer concrete through mix proportions and pre-curing conditions. The difference between each model has been analyzed, highlighting variations in prediction accuracy (Peng and Unluer, 2022). In addition, the machine learning technique shows good performance in workability, slump flow, flowability, flow consistency, etc., and prediction of alkali-activated materials (Kong and Kurumisawa, 2023).

Recently, many researchers have frequently applied machine learning models in compressive strength prediction of alkali-activated concrete. The relations of mix parameters of alkali-activated concrete such as SiO_2: Na_2O ratio, $NaOH$ solution to Na_2SiO_3 solution ratio, alkaline solution to cementitious materials ratio, etc. The machine learning models show the significance and influence of the chemical compositions of cementitious materials and alkali activator on the compressive strength of fly ash based alkali activated concrete (Toufigh and Jafari, 2021). Research chose the mass of the Na_2O dosage to the mass of the GGBS and fly ash ratio; the mass of the SiO_2 dosage to the mass of the Na_2O dosage ratio; the mass of additional water; the mass of GGBS and fly ash; the

mass of GGBS to the mass of cementitious materials ratio and curing time as input value in machine learning models training (Sun et al., 2023a). Precursor content, GGBS ratio, Na_2O content, modulus in activator, water content, fine aggregate, and coarse aggregate were chosen as parameters in alkali-activated mixture proportions for building random forest machine learning dataset to forecast the compressive strength (Sun et al., 2023b). Gomaa et al. (2021) trained machine models to predict the compressive strength of alkali-activated concrete, which uses a comprehensive input value focus on the chemical composition in cementitious materials and curing conditions. Three dissimilar machine learning models, Random Forest (RF), Gradient Boosting (GB), and Back Propagation Neural Network (BPNN), show that they have outstanding prediction performance in compressive strength of alkali-activated concrete with R^2 over 0.85 and 0.70 for training and testing set. Eight input variables in the dataset are used in machine learning model training and testing (Li et al., 2023). Investigation proposes a machine learning model according to their mix formula and the chemical elements of their cementitious materials and activators to estimate the compressive strength of alkali-activated concrete. The dataset in this study includes 676 mixture design samples collected from peer-reviewed investigations. Eleven input parameters which mainly focus on chemical components of raw materials in mixture design were selected for machine learning models training and testing. Four machine learning algorithms, support vector machine, extra trees, random forest, and gradient boosting all got significant performance predictions of compressive strength in alkali-activated concrete (Zhang et al., 2022).

Different machine learning models also show strengths and weaknesses in concrete properties prediction. Random Forest excels in modeling non-linear relationships inherent in concrete datasets, interactions between mix

proportions, curing conditions, and alkali activators, for example. While resisting overfitting, making it robust for predicting compressive and flexural strength (Chou et al., 2014). However, its tendency to average extreme strength values can limit accuracy in predicting high-performance or low-strength alkali-activated concrete. XGBoost often achieves superior accuracy in concrete strength prediction tasks due to gradient boosting and regularization, effectively capturing hierarchical relationships such as the impact of SiO_2/Na_2O ratio, but its reliance on meticulous hyperparameter tuning poses challenges for noisy or imbalanced concrete datasets (Feng et al., 2020). Support Vector Machines are advantageous for small, high-dimensional concrete datasets using non-linear kernels to model complex strength trends, though their computational inefficiency with large datasets limits scalability in industrial applications (Chou and Pham, 2013). K-Nearest Neighbors is rarely favored in concrete strength prediction due to its sensitivity to irrelevant features and poor extrapolation beyond training data clusters, despite its simplicity for small-scale studies. Overall, RF and XGBoost dominate in alkali-activated concrete research for their balance of accuracy and interpretability while SVM and KNN are solutions depending on dataset characteristics (Ansari et al., 2024).

2.4 Summary

Alkali-activated concrete shows not only advances in mechanical properties but also reduces the environmental impact in the global civil engineering industry. Researchers completed extensive investigations to show that alkali-activated has many benefits as an alternative material for traditional Portland cement concrete in construction and building materials. Based on the literature study by previously published research journal articles, machine learning techniques such as support vector machines, extra trees, random forests, natural networks,

extreme learning machines, etc., performed with high accuracy in the prediction of compressive strength of alkali-activated concrete. However, researchers mainly focus on the cementitious materials, alkaline activators, water, and chemical composition of each raw material. Additionally, comprehensive studies on the particle packing theory of concrete materials have established a strong correlation between aggregate packing density and mechanical properties. Despite this, no research has explored the application of machine learning techniques to predict the mechanical properties of alkali-activated concrete based on particle packing theory. Given the critical role of aggregate packing density in influencing compressive and flexural strength, incorporating this parameter into machine learning based prediction models presents a novel approach. This study aims to address this gap by integrating aggregate packing density as an input variable in machine learning models to enhance the predictive accuracy of alkali-activated concrete's mechanical properties, thereby advancing the understanding and optimization of sustainable concrete mix designs. While this study seeks to bridge the identified research gap, it also faces several methodological challenges. The dataset for model training is derived from laboratory experiments conducted specifically for this research, making the dataset relatively small. The limited sample size may affect the generalizability of the machine learning models, requiring careful validation techniques for model creation. Machine learning techniques have demonstrated strong predictive capabilities, but model optimization remains a challenge. Further improvements, such as hyperparameter tuning, feature selection, and advanced ensemble learning methods, may enhance model performance. Addressing these challenges is crucial for developing a more robust and reliable machine learning based prediction framework for alkali-activated concrete mechanical properties. Machine learning techniques have demonstrated strong predictive capabilities, but model optimization remains a

challenge.

3. Research Significance

In this research, the packing density of aggregates has been chosen as a new parameter in machine learning model development; this is an innovative work in this research field. This investigation selects random forest, extreme gradient boosting, support vector machine, and K-Nearest Neighbors as four different machine learning algorithms to figure out the most adequate and accurate model for alkali-activated concrete compressive strength prediction and consolidate packing density as an important input variable. A dataset involving 99 compressive strength data and 33 flexural data was used to train and test the machine learning models; all data in this dataset were collected from an experiment study. By training machine learning models on this experimental data, analyzing the feature importance of each model, subsequently evaluating their performance, and comparing the prediction accuracy with previous models, this research endeavors to stretch the limits of compressive strength prediction and flexural strength prediction in alkali-activated concrete with high precision.

The overall aim of this research is to create a novel approach by integrating machine learning models with particle packing theories. This approach seeks to improve the accuracy of compressive and flexural strength predictions in alkali-activated concrete. It holds great promise for enhancing our understanding of alkali-activated concrete mix design and applications in civil engineering construction programs. By reducing the environmental impact of traditional Portland cement, it also increases the feasibility of alkali-activated concrete in real-world construction practices.

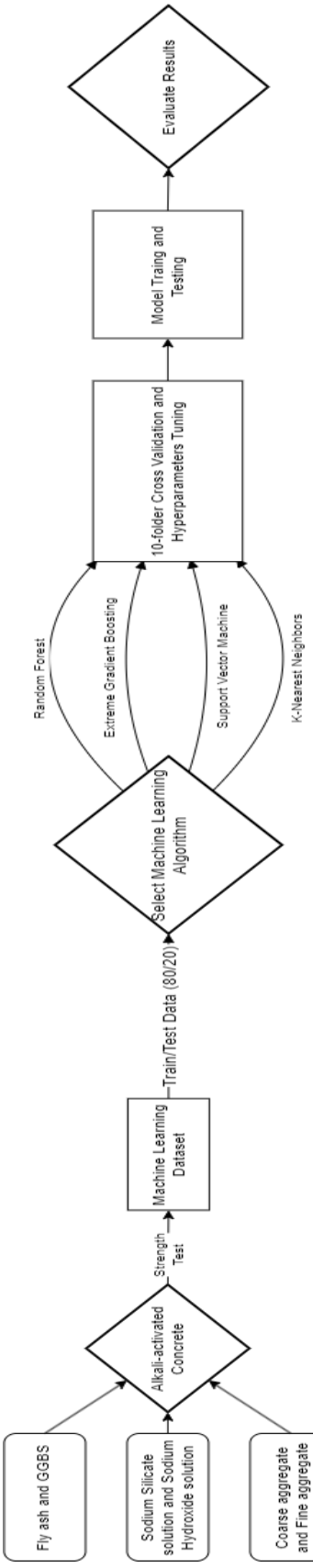

Figure 1. Flow chart of the use of machine learning model to predict mechanical properties of alkali-activated concrete

4. Experimental program

In this study, all experiments including raw materials preparation, raw materials properties measurement, concrete mix, compressive strength test, flexural strength test, and data collection were completed by the author of this dissertation.

4.1 Raw materials

Alkali-activated concrete was formulated using two cementitious materials: ground granulated blast-furnace slag (GGBS) and class F fly ash, water, $NaOH$ and Na_2SiO_3 solution, fine aggregates, and coarse aggregates. The following subsections describe detailed information about the precursors, alkali activators, and basic properties of aggregates.

4.1.1 Collection of coarse aggregates and fine aggregates

Table 1. Physical properties of the aggregates

Properties	19-12.5mm	12.5-9.5mm	9.5-4.75mm	River sand
Specific gravity	2.65	2.62	2.60	2.55
Water absorption (%)	0.91	1.01	0.71	0.28
Bulk density (kg/m^3)	1522.3	1450.1	1481.3	1668.8
Void content (%)	42.4	44.5	42.9	34.4
Packing density	57.6	55.5	57.1	65.6

Table 2. Particle size distribution of fine aggregates

Sieve size (mm)	River sand passing (%)
9.5	100
4.75	99.21

2.36	86.95
1.18	76.06
0.6	51.22
0.3	6.73
0.15	2.07
0.08	0.79

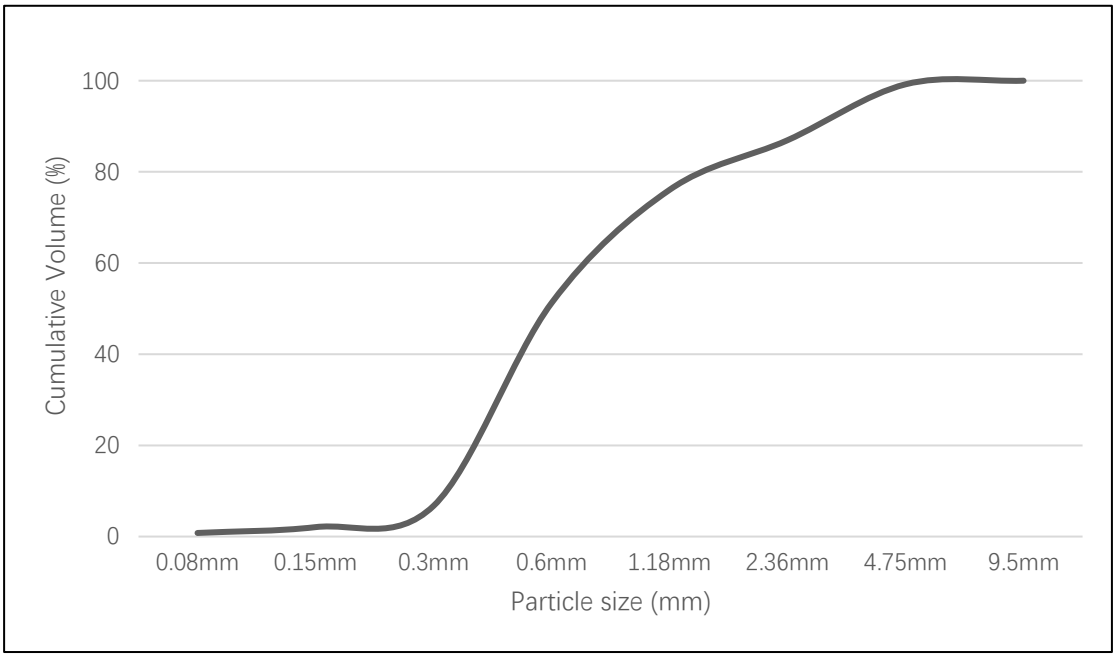

Figure 2. Particle size distribution of fine aggregates

Figure 3. Sieving machine

The present experiment used river sand as fine aggregates in the concrete mix. Three distinct sizes of coarse aggregates with maximum diameters of 19mm, 12.5mm, and 9.5mm were used in the experiment. The collection of coarse aggregates and fine aggregates is based on the standard ASTM C33. The selection of four aggregate sizes (19–12.5 mm, 12.5–9.5 mm, 9.5–4.75 mm, and less than 4.75 mm) was based on the size separation specified in ASTM C33. This choice enables precise control and measurement of packing density in alkali-activated concrete mixtures. By varying the proportions of each aggregate size, the packing density can be optimized to reduce voids and enhance the mechanical properties of the concrete. Additionally, this approach provides a standardized framework for measuring packing density in each mixture, ensuring the results are aligned with international guidelines.

4.1.2 Specific gravity test of coarse aggregate

The specific gravity test in this experiment study of coarse aggregate was completed before the experiment by using ASTM C127. The test will follow the procedure below: (1) Sampling the coarse aggregate, 2 kilograms for each nominal maximum size aggregate less or equal to 12.5mm, and 3 kilograms for 19mm aggregate. (2) Place the test coarse aggregate specimen in the oven at 105℃, then cool the aggregate sample at room temperature, which is 23℃, for 2 hours. (3) Saturate the coarse aggregate sample in water at 23℃, for 24 hours. (4) Take out the test sample from the water, and clean every water film that is apparent on the aggregate surface. Measure the weight of the test coarse aggregate sample under its saturated surface-dry condition. Use the masses measured to the bigger value of closest 0.5g or 0.05% of the sample mass. (5) Put the coarse aggregate sample with the saturated surface-dry condition in a container; next, determine the apparent mass in water at room temperature.

(6) After allowing the sample to cool for one to three hours at room temperature after drying it at 105°C to a consistent mass, weigh the aggregate sample. The specific gravity and absorption of coarse aggregate can be calculated by this equation:

$$Specific\ gravity = \frac{A}{B - C} \qquad \text{Equation 1}$$

$$Absoption, \% = \left(\frac{B - A}{A}\right) \times 100 \qquad \text{Equation 2}$$

Where A is the weight determined from the oven-dry test specimen in air g, B is the weight determined from the saturated surface-dry test specimen in air g, and C is the apparent weight determined from the saturated test specimen in water g.

4.1.3 Specific gravity test of fine aggregate

In this experiment, ASTM C128 will be used to measure the specific gravity of fine aggregate. The test will follow the following processes: (1) Sample the aggregate of approximately 1 kilogram. (2) Make the fine aggregate arid in the oven until the mass is still in a constant value at a temperature of 105℃, and measure the mass. (3) Cover the sample with water for 24 hours, then remove the fine aggregate specimen from the water, locate the sample on a level nonabsorbent surface such as a plastic table, expose the sample to slowly moving warm air, and mix it frequently to make sure homogeneous drying to get saturated surface-dry condition, then determine the mass. (4) Use water to fill the pycnometer, put 500g of saturated surface-dry fine aggregate sample from the previous step in the pycnometer, and continually add water into the pycnometer to reach 90% of its total capacity. Roll it with a hand and invert the pycnometer to remove visible air bubbles between fine aggregate particles. (5) Take out the fine aggregate samples from the pycnometer, place them in the

oven at a temperature of 105℃ until they achieve a constant mass, and measure the mass of the sample. (6) Use water to fill the pycnometer to its calibrated capacity at room temperature, and measure the weight of it. The specific gravity and absorption of fine aggregate can be calculated by this equation:

$$Specific\ gravity = \frac{A}{B + S - C} \qquad \text{Equation 3}$$

$$Absoption, \% = \left(\frac{S - A}{A}\right) \times 100 \qquad \text{Equation 4}$$

Where A is the weight determined from oven dry fine aggregate sample in the air g, B is the weight determined from the pycnometer filled with water to calibrated capacity in g, C is the mass of the pycnometer filled with fine aggregate specimen and water to the calibrated capacity, S represents the mass of saturated surface-dry test fine aggregate sample in the air g.

4.1.4 Test bulk density and voids of fine and coarse aggregates

The bulk density and voids of both coarse and fine aggregates in this investigation were determined according to ASTM C29. The materials used included fine and coarse aggregate samples, a calibrated cylindrical container of known volume, a balance accurate to 0.05% of the sample mass, a steel tamping rod with a hemispherical tip, a scoop, and a funnel. First of all, determine the volume of the container, determine the weight of the container weight as T_1, and determine the weight of the container with water as T_2. Calculate the volume of the container by using this equation: $V = \frac{T_2 - T_1}{997.54}$. Where the V is the volume of the container, the 997.54 is the density of water at room

temperature. The aggregates were first oven-dried at 110 ± 5°C until constant mass was achieved and then cooled to room temperature. A representative sample of the aggregate was obtained through quartering or using a sample splitter. For fine aggregates, the container was filled approximately one-third full of the aggregate using a scoop, and the layer was rodded 25 times uniformly over the surface with the tamping rod. This process was repeated for subsequent layers until the container was filled to overflowing. The excess aggregate was struck off level with the top of the container using a straightedge, and the filled container was weighed to the nearest 0.05% of the total mass. For coarse aggregates, the procedure was similar. The container was again filled in one-third increments, with each layer rodded or vibrated as needed, struck off, and weighed. So, the bulk density and void content of aggregates can be calculated as following equations:

$$M = (G - T)/V \qquad \text{Equation 5}$$

$$Void\ content = 100 \times [(S \times W) - M]/(S \times W) \qquad \text{Equation 6}$$

Where, M represent the aggregate's bulk density in $^{kg}/_{m^3}$, G is the mass of aggregate plus the measure in kg, T is the mass of measure in kg, V is the volume of the measure in m^3, S represent the aggregate's specific gravity, W is the water's density in $^{kg}/_{m^3}$.

4.1.5 Cementitious materials

Figure 4. Fly ash

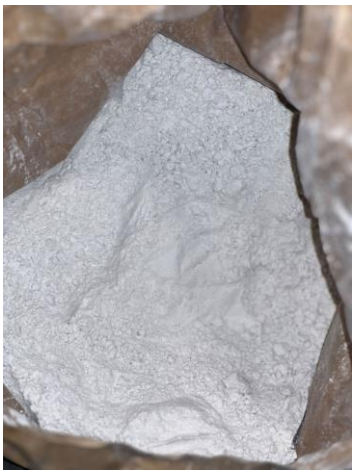

Figure 5. Ground granulated blast furnace slag

Table 3. Particle size distribution of Fly ash

Sieve size (mm)	Fly ash particles Passing (%)
0.6	100
0.3	99.847
0.15	76.471
0.08	23.682

Table 4. Particle size distribution of Ground granulated blast furnace slag

Sieve size (mm)	GGBS particles Passing (%)
0.6	100

0.3	21.455
0.15	71.063
0.08	7.482

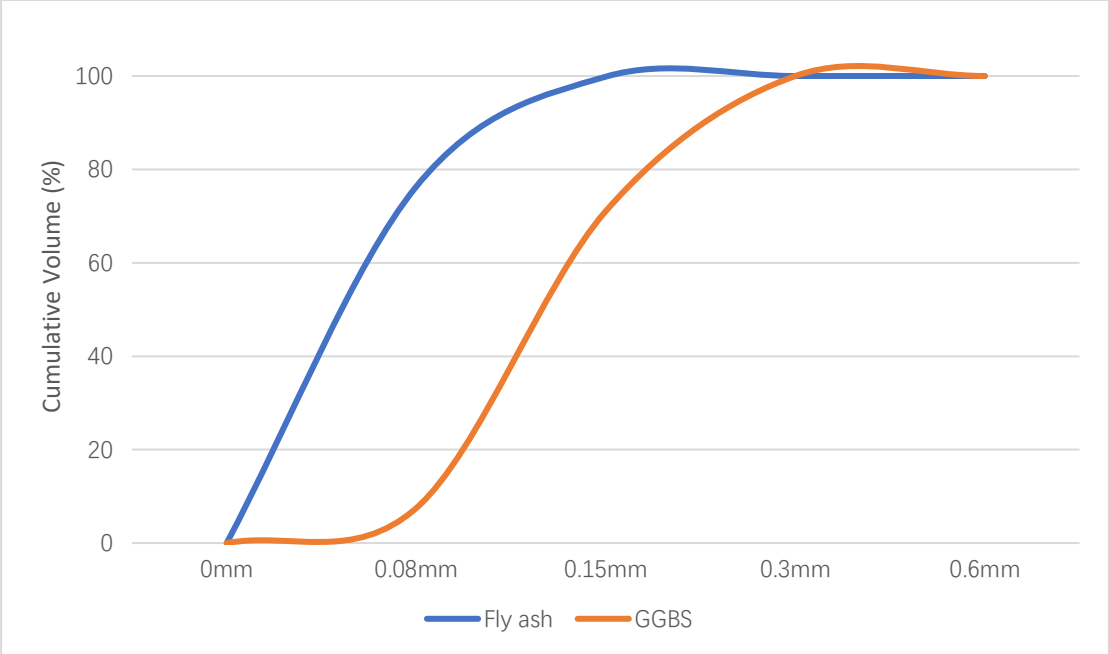

Figure 6. Particle size distribution of Fly ash and GGBS

In the experiment, ground granulated blast furnace slag and coal fly ash (ASTM type F) were used as cementitious materials in the alkali-activated concrete mixture.

4.1.4 Alkali activator solution

The alkali activators in this experiment were combined with a sodium hydroxide solution and water glass solution. In this experiment study $NaOH$ (solid) with >99% purity. The 14 M molarity of the $NaOH$ solution was made up using solid $NaOH$ pellets. In this process, 560 g of the solid $NaOH$ pellets were combined with 1000 g of top water at a room temperature of 23°C until all of the pellets had completely dissolved. Since sodium hydroxide dissolves in water and releases much heat, this process needs to be completed 24 hours before starting to mix the concrete to allow the sodium hydroxide solution to cool to

room temperature. Na_2SiO_3(liquidous solution), with ${}^{SiO_2}/_{Na_2O}$ of 2.33, had Na_2O content of 13.32%, SiO_2 content of 30.12%, and H_2O content of 56.56% by mass.

4.2 Mix proportion

The following table describes the detailed mix proportions used in this experiment study. For each mix, three prism samples and ten cylinder samples were made. In this experiment, 33 mix proportions were designed to make alkali-activated concrete samples. The mix proportion in order to build a machine learning dataset for compressive strength and flexural strength prediction based on various input value. The total mix proportion can be divided into three groups. In each group, the packing density in each mix is adjusted by changing the percentage of each aggregate. In addition, the Fly ash to GGBS ratio, ${}^{Na_2SiO_3}/_{NaOH}$ ratio, M_s and water to cementitious ratio of each mix is different. These variables are usually the key factors affecting alkali-activated concrete. The purpose of such an experimental design is to explore the influence of various variables on the strength of alkali-activated concrete through the powerful data analysis capabilities of machine learning. The mix design also cites the mix design process from pass investigation (Pavithra et al., 2016).

Table 5. Alkali-activated concrete mix proportion

Mix	Fly ash $({}^{Kg}/_{m^3})$	GGBS $({}^{Kg}/_{m^3})$	Sodium silicate $({}^{Kg}/_{m^3})$	Sodium hydroxide $({}^{Kg}/_{m^3})$	Additional water $({}^{Kg}/_{m^3})$	Aggregate 19-12.5mm $({}^{Kg}/_{m^3})$	Aggregate 12.5-9.5mm $({}^{Kg}/_{m^3})$	Aggregate 9.5-4.75mm $({}^{Kg}/_{m^3})$	Fine aggregate $({}^{Kg}/_{m^3})$
1	259	111	105.68	42.27	56.43	337.345	122.878	418.08	854.25

Mix	Fly ash $(^{Kg}/_{m^3})$	GGBS $(^{Kg}/_{m^3})$	Sodium silicate $(^{Kg}/_{m^3})$	Sodium hydroxide $(^{Kg}/_{m^3})$	Additional water $(^{Kg}/_{m^3})$	Aggregate 19-12.5mm $(^{Kg}/_{m^3})$	Aggregate 12.5-9.5mm $(^{Kg}/_{m^3})$	Aggregate 9.5-4.75mm $(^{Kg}/_{m^3})$	Fine aggregate $(^{Kg}/_{m^3})$
2	185	185	105.68	42.27	56.43	301.835	140.432	435.5	854.25
3	111	259	105.68	42.27	56.43	319.59	175.54	383.24	854.25
4	259	111	105.87	65.25	55.75	177.55	351.08	522.6	683.4
5	259	111	105.87	65.25	55.75	532.65	175.54	348.4	683.4
6	259	111	105.87	65.25	55.75	355.1	526.62	174.2	683.4
7	259	111	105.87	65.25	55.75	532.65	175.54	174.2	854.25
8	259	111	105.87	65.25	55.75	177.55	175.54	522.6	854.25
9	259	111	105.87	65.25	55.75	177.55	526.62	174.2	854.25
10	259	111	105.87	65.25	55.75	355.1	175.54	174.2	1025.1
11	259	111	105.87	65.25	55.75	177.55	175.54	174.2	1195.95
12	259	111	105.87	65.25	55.75	355.1	175.54	696.8	512.55
13	170	170	153.24	51.08	69.69	177.55	351.08	522.6	683.4
14	170	170	153.24	51.08	69.69	532.65	175.54	348.4	683.4
15	170	170	153.24	51.08	69.69	355.1	526.62	174.2	683.4
16	170	170	153.24	51.08	69.69	532.65	175.54	174.2	854.25
17	170	170	153.24	51.08	69.69	177.55	175.54	522.6	854.25
18	170	170	153.24	51.08	69.69	177.55	526.62	174.2	854.25
19	170	170	153.24	51.08	69.69	355.1	175.54	174.2	1025.1
20	170	170	153.24	51.08	69.69	177.55	175.54	174.2	1195.95
21	170	170	153.24	51.08	69.69	355.1	175.54	696.8	512.55
22	106.5	248.5	118.50	59.27	55.75	355.1	175.54	696.8	512.55
23	106.5	248.5	118.50	59.27	55.75	532.65	175.54	348.4	683.4
24	106.5	248.5	118.50	59.27	55.75	355.1	526.62	174.2	683.4
25	106.5	248.5	118.50	59.27	55.75	532.65	175.54	174.2	854.25
26	106.5	248.5	118.50	59.27	55.75	177.55	175.54	522.6	854.25
27	106.5	248.5	118.50	59.27	55.75	177.55	526.62	174.2	854.25
28	106.5	248.5	118.50	59.27	55.75	355.1	175.54	174.2	1025.1
29	106.5	248.5	118.50	59.27	55.75	177.55	175.54	174.2	1195.95
30	106.5	248.5	118.50	59.27	55.75	177.55	351.08	522.6	683.4
31	259	111	105.87	65.25	55.75	303.01	172.78	360.58	908.22
32	170	170	153.24	51.08	69.69	303.01	172.78	360.58	908.22
33	106.5	248.5	118.50	59.27	55.75	303.01	172.78	360.58	908.22

4.3 Concrete specimens

Figure 7. Cylinder concrete mold

Figure 8. Prism concrete mold

In this experiment study, two kinds of mold were used to produce concrete specimens. As shown in Figure 7 and Figure 8, the cylinder mold has a size of 100*200mm, and the prism mold has a size of 100*100*400mm. For each mix formula shown in the previous section, ten cylinder concrete specimens were used to complete the compressive strength test, and three prism concrete specimens were used to complete the flexural strength test.

4.3.1 Concrete mixing process

Figure 9. Concrete mix machine

To produce the alkali-activated concrete samples in this experiment study, the procedure of mixing was demonstrated as following procedures: (1) Combined all aggregates, and mixed for 1 min in the mixing machine; (2) Fly ash, and GGBS were gradually added into the mixing machine, continually mix with coarse and fine aggregates for another 1 min in the machine; (3) Combined both Na_2SiO_3 and $NaOH$ solution, added into the mixing machine, and mixed with other materials for 1 min; (4) The additional water was putted into the mixing machine for 1 min; (5) Once all components of alkali-activated concrete were added into the mixer, continually mixed for 3 min. After completion of the mixing process, transfer the fresh alkali-activated concrete to the cylinder mold and prism mold, smooth the surface of concrete samples, cover it with plastic film, and cure it in a stable environment at around 20°C for 24 hours until demold.

4.3.2 Concrete casting

Figure 10. Vibrating table

First, to mold concrete specimens using a vibration machine, ensure the cylinder molds $(100\ mm\ x\ 200\ mm)$ and prism molds $(100\ mm\ x\ 100\ mm\ x\ 400\ mm)$ are clean, and a thin release agent layer is applied to their inner surfaces. Fill each mold with fresh concrete in layers; for cylinder molds, fill in three layers; and for prism molds, fill in two layers. After adding each layer, place the molds on the vibration machine to compact the concrete and eliminate air voids, ensuring even distribution and proper bonding between layers. After filling and compacting the final layer, level the surface with a trowel to achieve a smooth surface, even finish. For initial curing, keep the molds in a stable environment at around 20°C for 24 hours.

4.3.3 Curing

Figure 11. Concrete curing container

The alkali-activated concrete prism specimens covered with plastic bags were cured temperature of $23 \pm 2\,°C$ for a day. The cylinder concrete specimens were demolded and covered with plastic bags; then the settled cylinder specimens in a moisture curing container maintained at $23 \pm 2\,°C$ and the relative humidity of $95 \pm 5\%$ until testing date.

4.4 Particle packing theory

Particle packing theory in concrete refers to a foundational concept extensively employed to intricately optimize the spatial arrangement of particles within the concrete mix, thus significantly enhancing its performance characteristics. This theory delves deeply into the intricate interplay of various particle sizes, ranging from coarse aggregates to fine cement particles. It describes how their strategic organization can lead to the achievement of maximum density and minimal void content. The particle packing arrangement plays a crucial role in concrete microstructure and, consequently, affects concrete material's mechanical properties and durability. The main of particle packing theory is to optimize the

packing in aggregates and other particles, to minimize the voids between different particles in concrete material, reduce the demand for cementitious material and water usage in the final. They not only improve the mechanical properties of concrete, but also enhance the sustainability of concrete. Otherwise, the best packing arrangement in concrete improves durability by reducing voids, for example, heightened resistance to freeze-thaw cycles, chemical attack, and permeability to deleterious agents, thereby elongating the lifespan of concrete structures and maintenance and repair costs. Packing density refers to the packing fraction or compacity, which is the ratio of the volume of particles to the total volume of the material they are packed in; it is an important parameter in particle packing theory. The higher packing density will cause a denser particle arrangement in concrete. Achieving high packing density is important in the concrete mix design. Particle packing theory provides the fundamental strategy and method for arranging particles to achieve optimal packing density. This approach seeks to reduce the void content in concrete by carefully selecting and positioning particles of various sizes and shapes to effectively fill the gaps. In this study, alkali-activated concrete is the target material, which is made of cementitious materials, coarse aggregates, fine aggregates, and alkali activators; during the experiment, only concern about the particle packing of fine aggregates and coarse aggregates. Niyazuddin and Umesh (2023) showed that as the packing density of concrete aggregates increases, the compressive strength increases. According to Nanthagopalan and Santhanam (2012), packing density had a positive influence on the compressive strength of self-compacting concrete made from cement and fly ash.

4.4.1 Measurement of packing density of the aggregates by experiment method

Table 6. Packing density measurement of each mixture

Mix	Void content	Packing density
1	0.24289	0.75711
2	0.27783	0.72217
3	0.29307	0.70693
4	0.23867	0.76133
5	0.32176	0.67824
6	0.35475	0.64525
7	0.29017	0.70983
8	0.30025	0.69975
9	0.31998	0.68002
10	0.27898	0.72102
11	0.30137	0.69863
12	0.36199	0.63801
13	0.24555	0.75445
14	0.31093	0.68907
15	0.33393	0.66607
16	0.28307	0.71693
17	0.29375	0.70625
18	0.31656	0.68344
19	0.28563	0.71437
20	0.30465	0.69535
21	0.36595	0.63405
22	0.35797	0.64203
23	0.30877	0.69123
24	0.32776	0.67224
25	0.29035	0.70965
26	0.29578	0.70422
27	0.31037	0.68963
28	0.28269	0.71731
29	0.30465	0.69535
30	0.25537	0.74463
31	0.20676	0.79324
32	0.19988	0.80012
33	0.20059	0.79941

The packing density of the aggregates was measured by a modified edition of

the test process described in ASTM C29. This procedure involved several steps designed to minimize subjective errors and ensure consistent and precise measurements. The experiment was conducted using a setup that included a steel bucket and a cylindrical container, ensuring accurate and repeatable results.

Initially, three different sizes of coarse aggregates (19mm, 12.5mm, and 9.5mm maximum sizes) and fine aggregate (river sand) were prepared. The aggregates were measured in separate trays, each with a mass equivalent to 12 liters according to their volume proportions. These aggregates were then manually mixed to achieve a homogenous blend, ensuring a uniform distribution of particle sizes. A bucket with a bottom radius of 70 mm, a top radius of 170 mm, and a height of 310 mm were used to carry the combined aggregates. The bucket was placed on a stand, and a cylindrical container with a diameter of 270 mm and a height sufficient to hold 10 liters was positioned below it. This setup ensured that the aggregates could fall freely into the container without any compaction or additional handling. After filling, any extra aggregates container were removed above the top. The container, now filled with aggregates, was weighed using a digital scale. The empty weight of the container was subtracted from this measurement to determine the exact mass of the combined aggregates.

The packing density was calculated using the measured mass and the specific gravity of each aggregate type. The void content was first determined using the equation:

$$Void\ content = \left(\frac{V_c - \left(\frac{M_1}{S_1} + \frac{M_2}{S_2} + \frac{M_3}{S_3} + \frac{M_{Fine}}{S_{Fine}}\right)}{V_c}\right) \qquad \text{Equation 7}$$

Where V_c is the volume of the container, M_1, M_2, M_3, M_{Fine} are the masses of

each coarse aggregate, 19-12.5mm, 12.5mm-9.5mm, 9.5-4.75mm, and fine aggregates, respectively. And S_1, S_2, S_3, S_{Fine} are the specific gravities of the different aggregate types with maximum diameters of 19mm, 12.5mm, 9.5mm, and fine aggregate. The packing density was then calculated as follows:

$$Packing\ density = 1 - Void\ content \qquad \text{Equation 8}$$

4.4.2 Use Continues models to achieve the biggest packing density in the concrete mixture

To achieve the biggest packing density in this experiment, mathematical approaches and packing models were used to establish optimal aggregate proportions analytically. The pioneering work of Fuller and Thomsen demonstrated that the arrangement of all types of aggregates in concrete influences the properties of the resulting concrete. They found that aggregates, in combination with a geometric continuous grading, could improve the properties of concrete (Fuller and Thompson, 1907). Following the research by Fuller and Thomsen, it was found that the minimum value of voids could theoretically be reached through an optimal particle size distribution (PSD) of all sizes of aggregates used in the concrete mix, as indicated in the equation shown as below. (Andreasen, 1930)

$$P(D) = (\frac{D}{D_{Max}})^q \qquad \text{Equation 9}$$

where $P(D)$ represents a fraction of the entire aggregates smaller than the size D, D represents the particle dimension, D_{Max} represents the biggest particle dimension and q is the distribution modulus.

However, the equation above does not account for the minimum particle size, although engineering practice must have a finite lower size limit. Therefore, an altered model according to the Andreasen and Andersen Equation was

suggested by Funk and Dinger. In this study, three concrete mixtures are designed according to this modified Andreasen and Andersen model, which is presented below (Funk and Dinger, 2013):

$$P(D) = \frac{D^q - D_{min}^q}{D_{max}^q - D_{min}^q} \qquad \text{Equation 10}$$

Table 7. Fraction of each size of aggregates from Modified Andreasen and Andersen Equation

Aggregate size in mm	Fraction of aggregate
0.08	0
0.15	3.61053
0.3	8.79935
0.6	15.646
1.18	24.4303
2.36	36.2712
4.75	52.0592
9	72.7277
12.5	82.6314
19	100

where D_{min}^q is the minimum particle size. In the methods of Fuller and Thompson, 0.37 was the packing factor; in the methods of Funk and Dinger, it was 0.37; and in the method of Andreasen and Andersen, it could have been any suitable value between 0.33 and 0.5. In this experiment study, the value q will be set as 0.4 in this experiment. Distribution modulus q is used to determine the fraction between each size of aggregates in the concrete mixture. Different types of concrete can be obtained by setting different values for the q. A coarse mixture normally has a distribution modulus with a value higher than 0.5, while loose concrete mixes with more fine aggregates result in lower values of distribution modulus lower than 0.25.

In this experiment, modified Andreasen and Andersen

4.5 Compressive strength test procedures

Figure 12. Concrete cylinder for compressive strength test

Figure 13. Compressive strength test machine

Figure 14. Compressive strength test machine

Before the compressive strength test, use gypsum water mixture to make the surface of the concrete cylinder smooth enough to continue the test. The gypsum and water were mixed with a ratio of 100g gypsum to 35g water for each concrete cylinder sample. For the compressive strength test of 100 mm x 200 mm cylindrical concrete specimens as per ASTM C39/C39M, first measure the diameter at mid-height to the nearest 0.2 mm and calculate the cross-sectional area. Place the specimen in the testing machine, ensuring it is centered under the loading head and aligned properly. Apply a continuous, shock-free load at a rate of 0.25 ± 0.05 MPa/s until failure occurs, so the load rate set for the machine was 2KN/s. When the test machine stops loading, record the maximum load the specimen carries. The compressive strength of a specimen is calculated by dividing its highest load by its cross-sectional area. In order to ensure the accuracy and reliability of the compressive strength and flexural strength, three concrete samples have been tested. After the test data are obtained, the three data are averaged to obtain the final strength. The outlier analysis was conducted according to the following procedure to ensure the reliability of experiment results (Bernal et al., 2011, Sun et al., 2023a, Zhang et al., 2023, Fennis and Walraven, 2012, Soutsos et al., 2016, Pavithra et al.,

2016, Rafeet et al., 2017) Once detected, each outlier is evaluated by comparing it with similar studies in literature to determine if similar trends in strength ranges, material composition, and failure modes have been reported. If the experimental data is abnormal, repeat the experiment in the same way.

4.6 Flexural strength test procedures

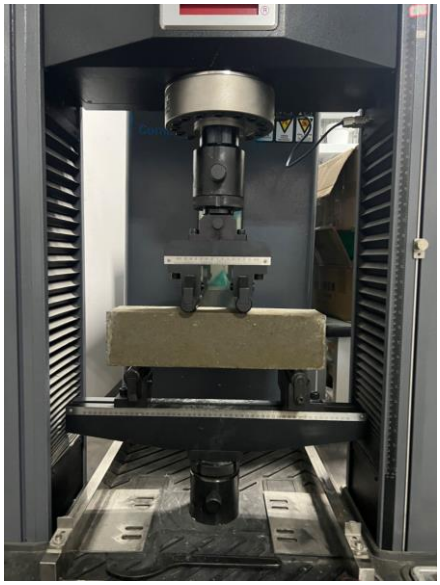

Figure 15. Flexural strength test set up

Figure 16. Flexural strength test machine

Figure 17. Concrete prism after test

The ASTM C78/C78M details the method for using a simple beam with third-point loading to determine the modulus of rupture. This test method evaluates the flexural strength of concrete specimens, which reflects the concrete's ability to resist bending. The testing procedure involves placing the 100 mm x 100 mm x 400 mm concrete prism in the testing machine with its tension face in contact with the support blocks. Up until the beam fails, the weight is applied steadily, continuously, and without shock at a steady rate. The rate of loading is

calculated to ensure it increases the maximum stress on the tension face at a steady rate. In this experiment, the loading rate is 40N/s as a constant. After failure, the dimensions of the fractured sections are measured to calculate the modulus of rupture. Post-test measurements include the specimen's width and depth at the fracture point, taken to the nearest 1 mm. The span length and the distances from the fracture line to the nearest support are also measured. The modulus of rupture is calculated based on these measurements and the maximum load applied during the test. The calculation considers whether the fracture happens outside or inside the center third of the span length, and using specific equations provided in the standard. The fracture occurs within the central third of the span length:

$$R = \frac{PL}{bd^2} \qquad \text{Equation 11}$$

For the fracture is not more than 5% of the span length outside the center third:

$$R = \frac{3Pa}{bd^2} \qquad \text{Equation 12}$$

Where, R represents flexural strength, in MPa. P is the maximum load applied to the specimen, in N. L is span length, in mm. b is the specimen's average width at the fracture, in mm. d represents the specimen's average depth at the fracture, in mm. a is average length from the line of fracture to the nearest support, in mm. To ensure the accuracy and reliability of the compressive strength measurements, three concrete prisms will be tested. The results from these tests will be averaged to determine the final flexural strength. Outlier analysis will be performed based on methodologies established in previously published experimental studies. If any experimental data are deemed abnormal, the tests will be repeated following the same procedures.

5. Machine learning models

5.1 Machine Learning Models Dataset

The entire dataset includes 99 data points; machine learning models were trained by the dataset, which was collected from previous experiments; the compressive strength and flexural strength were set as target values in the machine learning models. The following table shows the dataset's entire data used to predict alkali-activated concrete compressive strength and flexural strength in machine learning models. The compressive strength was set as the target output value, which is a function of the eight input values, packing density of aggregates, fly ash to GGBS ratio, weight of cementitious materials, activator to cementitious materials ratio, sodium silicate to sodium hydroxide ratio, modulus of activator, water to cementitious materials ratio and curing time. In the flexural strength dataset, the seven input values are packing density of aggregates, fly ash to GGBS ratio, weight of cementitious materials, activator to cementitious materials ratio, water glass solution to sodium hydroxide solution ratio, modulus of activator and water to cementitious materials ratio.

Table 8. Machine learning training dataset used for training and testing (compressive strength)

Compressive strength (MPa)	Packing density	Fly ash to GGBS ratio	Weight of Cementitious materials	Activator to Cementitious ratio	Sodium silicate to Sodium hydroxide ratio	Ms	Water to Cementitious material ratio	Curing time
18.93	0.75711	2.333	370	0.4	2.5	1.08	0.39	3
35.18	0.75711	2.333	370	0.4	2.5	1.08	0.39	7
50.34	0.75711	2.333	370	0.4	2.5	1.08	0.39	28
26.70	0.72217	1.000	370	0.4	2.5	1.08	0.39	3
40.66	0.72217	1.000	370	0.4	2.5	1.08	0.39	7
61.25	0.72217	1.000	370	0.4	2.5	1.08	0.39	28
31.27	0.70693	0.429	370	0.4	2.5	1.08	0.39	3
49.36	0.70693	0.429	370	0.4	2.5	1.08	0.39	7

Compressive strength (MPa)	Packing density	Fly ash to GGBS ratio	Weight of Cementitious materials	Activator to Cementitious ratio	Sodium silicate to Sodium hydroxide ratio	Ms	Water to Cementitious material ratio	Curing time
69.35	0.70693	0.429	370	0.4	2.5	1.08	0.39	28
27.87	0.76133	2.333	370	0.4	2.5	1	0.32	3
39.92	0.76133	2.333	370	0.4	2.5	1	0.32	7
58.47	0.76133	2.333	370	0.4	2.5	1	0.32	28
22.59	0.67824	2.333	370	0.4	2.5	1	0.32	3
36.82	0.67824	2.333	370	0.4	2.5	1	0.32	7
50.11	0.67824	2.333	370	0.4	2.5	1	0.32	28
26.49	0.64525	2.333	370	0.4	2.5	1	0.32	3
37.34	0.64525	2.333	370	0.4	2.5	1	0.32	7
51.58	0.64525	2.333	370	0.4	2.5	1	0.32	28
22.71	0.70983	2.333	370	0.4	2.5	1	0.32	3
37.93	0.70983	2.333	370	0.4	2.5	1	0.32	7
55.45	0.70983	2.333	370	0.4	2.5	1	0.32	28
23.64	0.69975	2.333	370	0.4	2.5	1	0.32	3
39.64	0.69975	2.333	370	0.4	2.5	1	0.32	7
54.39	0.69975	2.333	370	0.4	2.5	1	0.32	28
20.55	0.68002	2.333	370	0.4	2.5	1	0.32	3
35.87	0.68002	2.333	370	0.4	2.5	1	0.32	7
50.12	0.68002	2.333	370	0.4	2.5	1	0.32	28
24.75	0.72102	2.333	370	0.4	2.5	1	0.32	3
37.38	0.72102	2.333	370	0.4	2.5	1	0.32	7
53.79	0.72102	2.333	370	0.4	2.5	1	0.32	28
21.09	0.69863	2.333	370	0.4	2.5	1	0.32	3
34.33	0.69863	2.333	370	0.4	2.5	1	0.32	7
50.91	0.69863	2.333	370	0.4	2.5	1	0.32	28
19.73	0.63801	2.333	370	0.4	2.5	1	0.32	3
20.52	0.63801	2.333	370	0.4	2.5	1	0.32	7
48.84	0.63801	2.333	370	0.4	2.5	1	0.32	28
30.53	0.75445	1.000	340	0.6	3	1.12	0.42	3
39.71	0.75445	1.000	340	0.6	3	1.12	0.42	7
62.14	0.75445	1.000	340	0.6	3	1.12	0.42	28
27.32	0.68907	1.000	340	0.6	3	1.12	0.42	3
41.56	0.68907	1.000	340	0.6	3	1.12	0.42	7
57.70	0.68907	1.000	340	0.6	3	1.12	0.42	28
29.72	0.66607	1.000	340	0.6	3	1.12	0.42	3
42.55	0.66607	1.000	340	0.6	3	1.12	0.42	7
53.92	0.66607	1.000	340	0.6	3	1.12	0.42	28
26.97	0.71693	1.000	340	0.6	3	1.12	0.42	3
39.85	0.71693	1.000	340	0.6	3	1.12	0.42	7
58.67	0.71693	1.000	340	0.6	3	1.12	0.42	28
29.46	0.70625	1.000	340	0.6	3	1.12	0.42	3

Compressive strength (MPa)	Packing density	Fly ash to GGBS ratio	Weight of Cementitious materials	Activator to Cementitious ratio	Sodium silicate to Sodium hydroxide ratio	Ms	Water to Cementitious material ratio	Curing time
41.39	0.70625	1.000	340	0.6	3	1.12	0.42	7
61.07	0.70625	1.000	340	0.6	3	1.12	0.42	28
27.02	0.68344	1.000	340	0.6	3	1.12	0.42	3
40.56	0.68344	1.000	340	0.6	3	1.12	0.42	7
55.70	0.68344	1.000	340	0.6	3	1.12	0.42	28
31.61	0.71437	1.000	340	0.6	3	1.12	0.42	3
39.80	0.71437	1.000	340	0.6	3	1.12	0.42	7
59.97	0.71437	1.000	340	0.6	3	1.12	0.42	28
24.25	0.69535	1.000	340	0.6	3	1.12	0.42	3
33.31	0.69535	1.000	340	0.6	3	1.12	0.42	7
56.15	0.69535	1.000	340	0.6	3	1.12	0.42	28
23.14	0.63405	1.000	340	0.6	3	1.12	0.42	3
34.25	0.63405	1.000	340	0.6	3	1.12	0.42	7
55.97	0.63405	1.000	340	0.6	3	1.12	0.42	28
40.06	0.64203	0.429	355	0.5	2	0.72	0.37	3
52.83	0.64203	0.429	355	0.5	2	0.72	0.37	7
63.53	0.64203	0.429	355	0.5	2	0.72	0.37	28
41.55	0.69123	0.429	355	0.5	2	0.72	0.37	3
53.17	0.69123	0.429	355	0.5	2	0.72	0.37	7
65.71	0.69123	0.429	355	0.5	2	0.72	0.37	28
42.34	0.67224	0.429	355	0.5	2	0.72	0.37	3
50.76	0.67224	0.429	355	0.5	2	0.72	0.37	7
64.87	0.67224	0.429	355	0.5	2	0.72	0.37	28
42.79	0.70965	0.429	355	0.5	2	0.72	0.37	3
57.16	0.70965	0.429	355	0.5	2	0.72	0.37	7
70.84	0.70965	0.429	355	0.5	2	0.72	0.37	28
47.07	0.70422	0.429	355	0.5	2	0.72	0.37	3
53.75	0.70422	0.429	355	0.5	2	0.72	0.37	7
69.16	0.70422	0.429	355	0.5	2	0.72	0.37	28
39.98	0.68963	0.429	355	0.5	2	0.72	0.37	3
53.74	0.68963	0.429	355	0.5	2	0.72	0.37	7
66.13	0.68963	0.429	355	0.5	2	0.72	0.37	28
48.62	0.71731	0.429	355	0.5	2	0.72	0.37	3
54.81	0.71731	0.429	355	0.5	2	0.72	0.37	7
68.53	0.71731	0.429	355	0.5	2	0.72	0.37	28
40.87	0.69535	0.429	355	0.5	2	0.72	0.37	3
49.74	0.69535	0.429	355	0.5	2	0.72	0.37	7
68.31	0.69535	0.429	355	0.5	2	0.72	0.37	28
44.81	0.74463	0.429	355	0.5	2	0.72	0.37	3
55.55	0.74463	0.429	355	0.5	2	0.72	0.37	7
72.20	0.74463	0.429	355	0.5	2	0.72	0.37	28

Compressive strength (MPa)	Packing density	Fly ash to GGBS ratio	Weight of Cementitious materials	Activator to Cementitious ratio	Sodium silicate to Sodium hydroxide ratio	Ms	Water to Cementitious material ratio	Curing time
29.74	0.79324	2.333	370	0.4	2.5	1	0.32	3
39.87	0.79324	2.333	370	0.4	2.5	1	0.32	7
60.42	0.79324	2.333	370	0.4	2.5	1	0.32	28
33.92	0.80012	1.000	340	0.6	3	1.12	0.42	3
47.18	0.80012	1.000	340	0.6	3	1.12	0.42	7
67.76	0.80012	1.000	340	0.6	3	1.12	0.42	28
45.32	0.79941	0.429	355	0.5	2	0.72	0.37	3
59.31	0.79941	0.429	355	0.5	2	0.72	0.37	7
75.67	0.79941	0.429	355	0.5	2	0.72	0.37	28

Table 9. Machine learning training dataset used for training and testing (flexural strength)

Flexural Strength (MPa)	Packing density	Fly ash to GGBS ratio	Weight of Cementitious materials	Activator to Cementitious ratio	Sodium silicate to Sodium hydroxide ratio	Ms	Water to Cementitious material ratio
5.3641	0.75711	2.333	370	0.4	2.5	1.08	0.39
5.9834	0.72217	1.000	370	0.4	2.5	1.08	0.39
6.4639	0.70693	0.429	370	0.4	2.5	1.08	0.39
6.8177	0.76133	2.333	370	0.4	2.5	1	0.32
6.613	0.67824	2.333	370	0.4	2.5	1	0.32
6.5768	0.64525	2.333	370	0.4	2.5	1	0.32
6.7787	0.70983	2.333	370	0.4	2.5	1	0.32
6.5064	0.69975	2.333	370	0.4	2.5	1	0.32
6.4775	0.68002	2.333	370	0.4	2.5	1	0.32
6.5595	0.72102	2.333	370	0.4	2.5	1	0.32
6.3778	0.69863	2.333	370	0.4	2.5	1	0.32
6.2121	0.63801	2.333	370	0.4	2.5	1	0.32
6.6986	0.75445	1.000	340	0.6	3	1.12	0.42
6.1725	0.68907	1.000	340	0.6	3	1.12	0.42
6.0902	0.66607	1.000	340	0.6	3	1.12	0.42
6.6675	0.71693	1.000	340	0.6	3	1.12	0.42
6.536	0.70625	1.000	340	0.6	3	1.12	0.42
6.6934	0.68344	1.000	340	0.6	3	1.12	0.42
6.1281	0.71437	1.000	340	0.6	3	1.12	0.42
5.8624	0.69535	1.000	340	0.6	3	1.12	0.42

Flexural Strength (MPa)	Packing density	Fly ash to GGBS ratio	Weight of Cementitious materials	Activator to Cementitious ratio	Sodium silicate to Sodium hydroxide ratio	Ms	Water to Cementitious material ratio
6.0014	0.63405	1.000	340	0.6	3	1.12	0.42
7.1034	0.64203	0.429	355	0.5	2	0.72	0.37
7.4049	0.69123	0.429	355	0.5	2	0.72	0.37
7.2084	0.67224	0.429	355	0.5	2	0.72	0.37
7.79	0.70965	0.429	355	0.5	2	0.72	0.37
7.4049	0.70422	0.429	355	0.5	2	0.72	0.37
7.1245	0.68963	0.429	355	0.5	2	0.72	0.37
7.6656	0.71731	0.429	355	0.5	2	0.72	0.37
7.2157	0.69535	0.429	355	0.5	2	0.72	0.37
7.5391	0.74463	0.429	355	0.5	2	0.72	0.37
6.8231	0.79324	2.333	370	0.4	2.5	1	0.32
6.7842	0.80012	1.000	340	0.6	3	1.12	0.42
7.7852	0.79941	0.429	355	0.5	2	0.72	0.37

5.2 Random Forest Model

Known as an ensemble learning method, the random forest technique is applied to regression and classification tasks. During model training, it creates a large number of decision trees. The choice made by the majority of trees is the random forest model's output when it comes to classification problems. The mean value, or average prediction value, that is derived from each individual tree in a regression task is returned and used as the output. During training, the core of a Random Forest model builds several decision trees and aggregates their outputs to produce a final forecast (Breiman, 2001). Unlike the decision trees model, the random forest model has advantages in avoiding overfitting the given training set. This model uses the power of multiple algorithms to reduce overfitting and improve the performance in target value prediction. A random forest model is an ensemble of decision trees built from a different

subset of training data made through the bootstrapping process. In order to generate multiple bootstrap samples, data points from the initial dataset are substituted during this process. Subsequently, each decision tree in the forest is trained using one of these bootstrap samples, which gives the trees more variance. The fact that Random Forests split nodes based on a random subset of features rather than all features during tree generation further encourages variance among the trees and prevents any one factor from dictating the model. The prediction process makes use of a Random Forest. The final outcome is the sum of the forecasts made by each individual tree. The class that receives the most votes overall across all trees is chosen as the final prediction.

Random Forest is a suitable machine learning model for this study due to its ability to handle small datasets while reducing the risk of overfitting. As an ensemble learning method, random forest builds multiple decision trees during training, using random subsets of the data and features, and averages their predictions, resulting in a robust model (Breiman, 2001). This robustness is particularly beneficial for small datasets, as it prevents overfitting by combining weak learners and ensuring that individual data points do not dominate the prediction process (Biau, 2010). Furthermore, random forest is capable of modeling non-linear relationships, which are common in material property predictions, while providing interpretable feature importance rankings.

5.3 Extreme Gradient Boosting Model (XGBoost)

Extreme Gradient Boosting or XGBoost is a stable and very effective gradient boosting framework that has been carefully designed to advance the speed and efficiency of machine learning models. The XGBoost model is highly recommended in data science because of its ability to build an ensemble of

decision trees sequentially, and the XGBoost model has advantages based on its ability to correct the errors created by the previous tree (Zhang et al., 2021). Unlike random forests, where all trees are built independently, XGBoost builds trees sequentially with the goal that each tree reduces the residue left by the previous tree. This approach is known for its superior scalability, accuracy, and ability to handle a wide range of data science problems. This complex process is based on the principle of gradient boosting, where the model iteratively merges trees to minimize a specified loss function. During the prediction phase, XGBoost combines the results of each tree into the ensemble. In the context of a classification problem, the scores generated by the individual trees are combined, and then these scores are converted into probability estimates using a logistic function; the maximum probability obtained is the prediction of the final classification. Conversely, for regression tasks, the predictions of all trees are added together to produce the final result. This iterative mechanism allows XGBoost to improve predictions by incorporating information about errors found in previous iterations (Chen et al., 2015).

XGBoost is a highly effective algorithm for small datasets because it employs boosting techniques to optimize model performance by iteratively refining weak learners. Its regularization features, including L1 and L2 penalties, reduce overfitting, which is crucial for small datasets with limited variability (Chen and Guestrin, 2016). Additionally, XGBoost handles missing data seamlessly and is computationally efficient, allowing for flexible experimentation with hyperparameters to adapt to the unique characteristics of small datasets, such as the 99 compressive strength and 33 flexural strength data points in this study.

5.4 Support vectors machine model

Support Vector Machine (SVM) represents a highly efficient supervised learning algorithm widely utilized for the purpose of both classification and regression tasks within the realm of machine learning. The fundamental principle underlying SVM revolves around the identification of an optimal hyperplane that effectively segregates the dataset into distinct classes. This characteristic renders SVM particularly proficient in managing datasets characterized by high dimensionality, exhibiting a remarkable degree of resilience and accuracy in its operations (Jakkula, 2006). Specifically, SVM is geared towards establishing a hyperplane or a series of hyperplanes within a multi-dimensional space strategically positioned to maximize the margin between the various classes present in the dataset. The margin essentially denotes the spatial separation between the hyperplane and the nearest data points belonging to each class, commonly referred to as support vectors. These support vectors play a pivotal role in delineating the precise location and orientation of the hyperplane within the dataset. When applied to classification tasks, SVM leverages the decision boundary defined by the optimal hyperplane to effectively categorize novel data points. By discerning the side of the hyperplane on which a new data point falls, SVM accurately assigns it to a specific class. Conversely, in scenarios involving regression tasks, denoted as Support Vector Regression (SVR), SVM attempts to formulate a function that closely approximates the target variable within a designated tolerance margin. The primary objective here is to devise the most suitable function within the permissible margin while concurrently minimizing errors in the process (Cortes and Vapnik, 1995b).

Support Vector Machine (SVM) is particularly effective for small datasets due to its reliance on maximizing the margin between data points and the decision

boundary, which improves generalization even with limited data (Cortes and Vapnik, 1995a). Additionally, the use of kernel functions, such as the radial basis function (RBF), enables SVM to capture complex, non-linear relationships between input features and target variables (Smola and Schölkopf, 2004). This capability ensures accurate modeling of material properties like compressive and flexural strength, even when the dataset size is restricted.

5.5 K-Nearest Neighbors model

A popular and essential machine learning technique, K-Nearest Neighbors (KNN) is based on the idea of instance-based learning, predicting outcomes based on the 'k' most similar instances in the feature space. Since it is non-parametric, no assumptions on the distribution of the data are made, contributing to its flexibility and ease of implementation for classification and regression tasks (Mucherino et al., 2009). The algorithm involves storing the entire training dataset and calculating the distance between new data points and existing points using metrics such as Euclidean, Manhattan, or Minkowski distances. Predictions are made by majority voting for classification tasks or averaging values for regression tasks. The benefits of k-NN include its simplicity, adaptability to various distance metrics, and robustness to noisy data when an appropriate 'k' value is chosen (Kramer, 2013). In civil engineering, KNN has been effectively applied in predicting monthly pan evaporation to assist in water resource management, and evaluated compressive strength of high-performance concrete (Hsieh, 2021) (Abed et al., 2023). Despite its computational cost and memory requirements, especially with large datasets, k-NN remains a valuable tool due to its intuitive approach and effectiveness across diverse applications.

The K-Nearest Neighbors (KNN) algorithm is particularly suitable for small datasets because it is an instance-based learning method that relies directly on the data rather than requiring a training phase, making it inherently advantageous for limited data points (Altman, 1992). Its non-parametric nature allows KNN to effectively capture non-linear relationships without assuming a specific data distribution (Peterson, 2009). This flexibility makes it an excellent choice for the prediction of material properties where small datasets, such as the one used in this study, are common.

5.6 Performance indicators for machine learning validation

Evaluating a predictive model's performance is essential to ensure its reliability and effectiveness. In regression tasks, several metrics can be used to assess how well a model predicts target values. This section delves into four commonly used evaluation metrics: MAE, RMSE, MAPE, and R^2. Each metric provides different insights into the model's performance, helping to identify strengths and weaknesses.

5.6.1 Mean Absolute Error (MAE)

The mean absolute error (MAE), which disregards the direction of the errors, is a metric used to determine the average magnitude of the errors in a set of forecasts. The average absolute discrepancies between the expected and actual values are computed. MAE is easy to understand and interpret. It offers a linear score, which implies that the average is equally weighted for each individual difference. A model that fits the data better is indicated by a lower MAE. It is beneficial when the costs of errors are directly proportional to their magnitude. Lower MAE corresponds to higher accuracy, while higher MAE corresponds to lower accuracy. The following function shows the calculation of MAE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad \text{Equation 13}$$

Where y_i is the observed value, $\hat{y}_i$ is the forecast value from models, and n is the number of observations.

5.6.2 Root Mean Squared Error (RMSE)

The average error magnitude is expressed as Root Mean Squared Error. The average squared difference between the expected and actual numbers is its square root. RMSE is more sensitive to large errors than MAE due to the squaring of the differences before averaging, which gives greater weight to significant errors, making it susceptive to outliers. A lower RMSE indicates a better fit of the model. It is advantageous when large errors are highly undesirable and must be penalized more. Lower RMSE corresponds to higher accuracy, while higher RMSE corresponds to lower accuracy. The calculation of RMSE is shown below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad \text{Equation 14}$$

Where y_i represents the observed value, $\hat{y}_i$ represents the forecast value from models, and n is the amounts of observations.

5.6.3 R-squared (R^2)

R-squared, is defined as the coefficient to measure the percentage of the volatility of the dependent parameters that can be predicted according to the independent variables. It offers an indicator of how closely the model matches the actual results. R^2 has a value in the range of 0 to 1. An R^2 value of 1 demonstrates that the model predicts the target variable with extreme accuracy,

while an R^2 of 0 demonstrates that the model has extremely bad performance. Negative values of R^2 can also occur, indicating that the model performs worse than a horizontal line which is the mean of the target. R^2 provides an overall measure of the performance of the model, making it a useful summary statistic for regression models. R^2 calculated by the following formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad \text{Equation 15}$$

Where $\bar{y}_i$ represents the mean of the observed values.

5.6.4 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error measures the accuracy of predictions as a percentage, representing the average absolute percentage error among predicted and observed values. It provides an institutionalized measure of prediction performance. Mean Absolute Percentage Error clearly indicates prediction accuracy in percentage terms. However, it can be problematic when actual values are minimal, leading to extremely high percentage errors. Lower RMSE corresponds to higher accuracy, while higher RMSE corresponds to lower accuracy. MAPE is beneficial for understanding the error in the context of the scale of the actual values, making it easier to interpret in a real-world context.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \qquad \text{Equation 16}$$

Where y_i represents the actual value, $\hat{y}_i$ presents the predicted value from models, and n is the amount of observations.

5.6.5 Mean Squared Error (MSE)

Mean Squared Error (MSE) is an essential indicator for assessing regression models' effectiveness in machine learning. It quantifies the average squared

difference between the actual values and the predicted values generated by a model. By squaring these differences, MSE ensures that larger errors have a disproportionately higher impact, which helps highlight significant deviations between the forecast and observed values. There is an inverse relationship between MSE and model accuracy: lower MSE corresponds to higher accuracy, while higher MSE corresponds to lower accuracy. The calculation formula of MSE is shown below,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \text{Equation 18}$$

Where y_i is the observed value, $\hat{y}_i$ is the predicted value from models, and n is the amount of observations.

5.7 Model training and testing of machine learning models

5.7.1 Dataset splitting

The dataset is divided using an 80/20 train-test split, where 80% of the data is used for training and 20% is reserved for testing. Within the training set, 10-fold cross-validation is employed to further split the data into subsets. This process ensures that the model is trained and validated on different folds of the data, providing a robust evaluation of its generalization capability. Additionally, hyperparameter tuning is conducted using GridSearchCV, which systematically explores predefined hyperparameter combinations during the 10-fold cross-validation process. The best-performing parameters are selected based on the cross-validation results, and the final model is evaluated on the untouched test set. The dataset is divided using an 80/20 train-test split, where 80% of the data is used for training and 20% is reserved for testing. Within the training set, 10-fold cross-validation is employed to further split the data into subsets. This process ensures that the model is trained and validated on different folds of the

data, providing a robust evaluation of its generalization capability. Additionally, hyperparameter tuning is conducted using GridSearchCV, which systematically explores predefined hyperparameter combinations during the 10-fold cross-validation process. The best-performing parameters are selected based on the cross-validation results, and the final model is evaluated on the untouched test set. This approach ensures efficient use of the dataset while maintaining the integrity of the test set for unbiased evaluation. The 10-fold cross-validation provides a reliable measure of performance by reducing variability associated with a single train-test split. GridSearchCV further enhances the model's predictive capability by optimizing hyperparameters, ensuring the selected configuration generalizes well across folds. Together, these techniques improve model accuracy and robustness while minimizing overfitting, making models particularly suitable for scenarios with complex data and small or middle dataset sizes, the dataset used in this investigation, for example.

This approach ensures efficient use of the dataset while maintaining the integrity of the test set for unbiased evaluation. The 10-fold cross-validation provides a reliable measure of performance by reducing variability associated with a single train-test split. GridSearchCV further enhances the model's predictive capability by optimizing hyperparameters, ensuring the selected configuration generalizes well across folds. Together, these techniques improve model accuracy and robustness while minimizing overfitting, making models particularly suitable for scenarios with complex data and small or middle dataset sizes, the dataset used in this investigation, for example.

5.7.2 Random Forest

Random Forest algorithms have established themselves as powerful tools in machine learning for both classification and regression tasks due to their robustness and high performance. The procedures of creating an RF model begin with data preparation. This involves handling missing values, removing

duplicates, and ensuring that all data types are correct. The prepared data is then separated into training and testing sets, typically in an 80:20 ratio, to aid model training and evaluation. The random forest model is an improved edition of bagging that combines both bagging and a random subset of input variables to decrease the relations of grown trees and, at the same time, keep reasonable prediction accuracy. (Breiman, 2001) A training set of $\{x_i, y_i\}_1^N$ given for random forest model, the characteristic N is the quantity of training samples, random forest model summaries the predictions from each individual tree which can present as $T_j(x_i; \theta_j)$, and by calculating the average value of them using the following equation: $f(X) = \frac{1}{M}\sum_{J=1}^{M} T_J(X: \theta_j)$ where M means the entire amount of base learners, each θ_j includes basic information knowledge about re-arrangement of samples, splitting locations and separating features. By using a Lasso regularization process, the base learners who are deemed irrelevant and uninformative will be eliminated from the ensemble. This also includes assessing the weights of the base learners according to their performance, dependency, and involvement in the total predictive power. When a specific base learner has a value zero of its weight α_j, it may be eliminated. Finally, we can get a compact-random forest model, which contains much fewer base learners, and has an improved generalization capability compared with the original random forest model. The optimal weights αj are used to minimize the penalty of ensemble loss, as below: $minimize\ \sum_{n=1}^{N} \varphi\left(y_i, \sum_{J=1}^{M} \alpha_j T\left(X: \theta_j\right)\right) + \lambda \sum_1^M \left|\alpha_j\right|$, in the ensemble model the standard loss function (mean squared error was normally used) is represented by $\varphi()$, Lasso parameter is given as λ, $T(X; \theta_j)$ is a simple base learner categorized by θ_j. λ value was identified by using the k-fold cross-validation technique in this investigation. (Afzali et al., 2024)

Once the data is prepared, the Random Forest model is constructed by importing necessary libraries and loading the dataset. The dataset is then separated into features and target variables, and the model is initialized using default parameters. This random forest model is trained on the training data and evaluated on the testing data using the metrics shown in the previous section. Hyperparameter tuning is conducted to optimize the model. Key hyperparameters in this random forest model in this study are (N_estimators), (Max_depth), (Min_samples_split), (Min_samples_leaf), (Max_features) and (bootstrap). The hyperparameters of a Random Forest model significantly influence its performance by controlling the complexity and generalization capability of the model. Specifically, the number of estimators (N_estimators) affects the model's robustness, with an increased number of trees generally reducing variance and improving accuracy, though at the cost of computational efficiency. The maximum depth (Max_depth) of the trees determines the complexity of the decision boundaries, with deeper trees capturing more intricate patterns but increasing the risk of overfitting. Meanwhile, the minimum samples required to split an internal node (Min_samples_split), and the minimum samples required at a leaf node (Min_samples_leaf) regulate the tree's growth, with higher values leading to simpler models that reduce overfitting but may underfit if too restrictive. The maximum number of features considered for a split (Max_features) plays a crucial role in controlling tree correlation, with smaller values promoting diversity among trees, thus enhancing generalization. Lastly, the bootstrap parameter, which controls whether sampling with replacement is applied, can impact the variance of the model by influencing the diversity of trees and improving robustness when set to true. Therefore, fine-tuning these hyperparameters is essential to balance the trade-off between bias and variance, optimizing model performance while preventing overfitting or underfitting. The ranges of hyperparameters tuning

are also important, it needs to design to explore the hyperparameter space while considering the computation time and resources.

The value of each hyperparameter was explored using Grid Search. This exhaustive search across parameter combinations helps identify the optimal configuration that maximizes model performance. To ensure the model's robustness and ability to generalize to new data, 10-fold cross-validation is applied. Ten subsets are created from the data in 10-fold cross-validation. Nine subsets are used to train the model, and the last subset is used to validate it. Every subset is used as the validation set once during the ten repetitions of this process. To get a trustworthy estimation of the model's performance, the results are averaged. After the random forest model finished hyperparameter tuning, it is evaluated using the testing data. Performance metrics such as Mean squared error and R-squared (R^2) score are computed to assess the model's effectiveness and predictive power. Creating and refining a Random Forest model is an iterative and comprehensive process. It involves meticulous data preparation, detailed feature importance analysis, strategic feature selection, rigorous hyperparameter tuning, and robust cross-validation. These steps collectively enhance the model's predictive power, robustness, and applicability to various machine learning tasks. Continuous evaluation and optimization are paramount in achieving a high-performance machine learning model that can deliver accurate and reliable predictions of the mechanical properties of alkali-activated concrete.

5.7.3 Extreme Gradient Boosting

The effective machine learning technique Extreme Gradient Boosting (XGBoost) is well-known for its exceptional performance in regression and classification applications. Chen and Guestrin published this model, which is an optimized

and improved gradient boosting algorithm. (Li and Song, 2022) The aim function for the XGBoost model can be explained below: $F_{obj}^{(t)} = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + K$, where F_{obj} is the aim function, L represents the loss function, Ω represents the regularization terms (L_1 and L_2), K represents the constant term. The initial phase includes careful data preparation, addressing missing values, removing duplicates, and ensuring accurate data types. Feature engineering creates and modifies features that enrich the dataset. To aid in model development and assessment, the data is then divided into training and test sets, often in an 80:20 ratio (Shen et al., 2022)

After data preparation, the XGBoost model is built. The required libraries are imported, and the dataset is loaded. The features and target variables are separated, and the XGBoost model is initialized with default parameters. The model is then trained on the training data and evaluated on the test data using metrics. Hyperparameter tuning is fundamental to optimizing the performance of the XGBoost model. Important hyperparameters include N_estimators, learning_rate, max_depth, min_child_weight, subsample, colasample_bytree. Grid Search Cross Validation (GridSearchCV) explores different hyperparameter combinations and identifies the optimal set. The hyperparameters of the XGBoost model significantly influence its performance by controlling the trade-off between model complexity, generalization ability, and computational efficiency. The number of estimators (N_estimators) dictates the number of trees, with a higher value, such as 300 for flexural strength, capturing more complex patterns but at the risk of overfitting and higher computation costs. The learning rate determines the step size during optimization, where smaller values like 0.01 ensure gradual learning, reducing overfitting, while larger values like 0.2 accelerate convergence but risk overshooting. The maximum depth (Max_depth) limits tree complexity, with smaller values for example 3

preventing overfitting while maintaining generalization. The minimum weight of child (Min_child_weight) ensures that nodes have a sufficient number of samples, with higher values improving robustness by reducing overfitting. Subsample and Colasample_bytree introduce randomness in sampling data and features, respectively, with values like 0.8 and 0.9 enhancing model diversity and robustness. Proper tuning of these hyperparameters ensures that the model captures relevant patterns in the data while avoiding overfitting or underfitting, optimizing its predictive performance.

To ensure the robustness and generalization of the model, 10-fold cross validation is used. In 10-fold cross validation, the data is divided into 10 subsets. On nine subgroups, the model is trained, and on the final subset, it is validated. Each subset is utilized as a validation set once, and this process is repeated ten times. To obtain a trustworthy approximation of the model's performance, the results are averaged. The test data is used to assess the improved model following hyperparameter adjustment. Developing and refining XGBoost models is an iterative and comprehensive process. It involves meticulous data preparation, rigorous hyperparameter tuning, and robust cross validation. Together, these steps enhance the model’s predictive ability, robustness, and applicability to a variety of machine learning assignments. Continuous evaluation and optimization are critical to achieving high-performance machine learning models that can provide accurate and reliable predictions of alkali-activated concrete's mechanical properties.

5.7.4 Support Vector Machine

Support vector machine is a resilient supervised machine learning method that has shown significant efficacy in improving the ability to generalize for both regression and classification assignments (Cortes and Vapnik, 1995b). In a

training dataset $\{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i)\}$, where i means the amount of training data, x_i represents the input vector, and y_i is the relevant actual target value. The core of the support vector machine when solving a regression problem is to create a linear function; this function has the ability to minimize the imparity within predictions and observation values in a set threshold (ε), maximizing the flatness at the same time. The following equation shows the linear regression function required in support vector machine: $f(x) = \langle w, x\rangle + b$, in this formula w is the vector of weight, scalar threshold vector represented by b, between vector w and x, the $\langle, \rangle$ calculate the dot product operation of it. The goal of maximizing flatness can be explained by the following formula, which minimizes the weight vector's norm: $minimize\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{i}(\xi + \xi_i^*)\ subjected\ to\ \begin{cases} y_i - \langle w, x\rangle - b \le \varepsilon + \xi \\ \langle w, x\rangle + b - y_i \le \varepsilon + \xi^* \\ \xi^*\xi_i \ge 0 \end{cases}$

In this equation, where C represents the regularization parameter that governs the trade-off between the flatness of the function $f(x)$ and the tolerance for deviations exceeding ε. Slack variables ξ and $\xi *$ are introduced to solve the convex optimization problem above, which allows for the ignore of deviations bigger than ε utilizing an ε-insensitive loss function, the equation can explain this shown as follows: $|\xi|_\varepsilon = \begin{cases} 0\ if\ f|y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon\ otherwise \end{cases}$. To provide the solution for the non-linearity problems, support vector machine use a linear function in a feature space with high dimensions using Lagrange multipliers αi and αi^* : $f(x, \alpha i, \alpha i^*) = \sum_{i=1}^{nsv}(\alpha i - \alpha i^*)\langle\varphi_{(x_i)}\varphi_{(x)}\rangle + b$, in this linear function, φ represent a non-linear map in the range between input space to the high-dimensional feature space, $\langle\varphi_{(x_i)}\varphi_{(x)}\rangle$ is the dot product between training patterns in high-dimensional feature space, nsv is the amount of support vectors. Nevertheless, expansive computational costs are an issue in mapping training patterns to a higher-dimensional space.

A kernel function can reduce this high computational cost $K(x_i, x_j)$, polynomial functions or radial basis, for example, without explicitly computing the form of $\varphi(x_i)$, as shown below: $f(x, \alpha i \cdot, \alpha i^*) = \sum_{i=1}^{nsv}(\alpha i - \alpha i^*)K(x_i, x) + b.$ The magnitude of the input variable affects the support vector machine, which is classified as one kind of geometric algorithm. When a specific feature is more important than others, the support vector machine may give this feature more weight, thus causing a mistake or an error in explaining its importance (Afzali et al., 2024) (Smola and Schölkopf, 2004).

In this investigation, hyperparameter tuning and cross validation are used to improve the reliability of the SVM algorithm. Important hyperparameters include the type of kernel, regularization parameter, epsilon, degree, and gamma. The hyperparameters of the Support Vector Machine (SVM) model significantly influence its ability to generalize, handle data complexity, and optimize prediction performance. The C parameter controls the trade-off between achieving a low error on the training data and ensuring good generalization. Higher values reduce bias but increase the risk of overfitting, while lower values promote simpler models with better generalization. The Kernel defines the function to transform input data into a higher-dimensional space for better separability. The linear kernel performs well for data, which is linearly separable, while more complex kernels like RBF are suited for non-linear relationships. The Gamma parameter, used in non-linear kernels, defines the influence of individual data points, a smaller value scale, generalizes better, while higher values focus more on specific points, increasing the risk of overfitting. The Degree parameter, relevant to polynomial kernels, controls the complexity of the decision boundary, with lower degrees, favoring simpler models. Lastly, the Epsilon parameter in regression tasks determines the margin of tolerance around predictions, higher values reduce model sensitivity to

outliers but may underfit the data. Together, these hyperparameters must be carefully tuned to balance model complexity, accuracy, and generalization.

To ensure the robustness and generalization of the model, 10-fold cross-validation is used. In 10-fold cross-validation, the data is divided into 10 subsets. On nine subgroups, the model is trained, and on the final subset, it is validated. This approach is performed ten times, using each subset as a validation set once. To obtain a trustworthy approximation of the model's performance, the results are averaged. The testing dataset is used to evaluate the improved model following hyperparameter adjustment. After hyperparameter tuning, the optimized model is assessed by the test dataset. Performance metrics such as MSE and R^2 scores are calculated for regression tasks, and accuracy is calculated for classification tasks to evaluate the effectiveness of the model.

Multicollinearity occurs when two or more independent variables in dataset are highly correlated, in this research activator to cementitious ration with water to cementitious material ratio, sodium silicate to sodium hydroxide ratio with M_s, for example. It will be affecting the performance of non-parametric models like SVM and KNN. In SVM, multicollinearity can cause instability in feature importance, affecting the construction of the optimal hyperplane and increasing overfitting risks, particularly in non-linear kernels. To mitigate these issues, incorporating L1 (LASSO) and L2 (Ridge) regularization techniques into Support Vector Machines (SVMs) effectively mitigates the adverse effects of multicollinearity by controlling feature weights and enhancing model stability. L1 regularization addresses multicollinearity by shrinking some coefficients to zero, effectively performing feature selection and removing redundant variables. L2 regularization, on the other hand, deals with multicollinearity by constricting the coefficients of correlated variables, thereby distributing their effects more

evenly across the features. By applying these regularization methods, SVMs can achieve improved predictive accuracy and robustness, reducing the bias and overfitting associated with multicollinearity. (Chan et al., 2022) (Dedieu and Mazumder, 2019)

5.7.5 k-nearest neighbors

After the completion of data preparation, the construction of the k-nearest neighbors (KNN) model commences. This process involves importing the necessary libraries and loading the dataset into the computational environment. Subsequently, the features and target variables are segregated to prepare for the KNN model's initialization with default parameters. It is imperative to engage in hyperparameter tuning to enhance the KNN model's performance. Key hyperparameters that necessitate optimization encompass the number of neighbors, the weight function, the algorithm utilized for computing the nearest neighbors, the leaf size and P parameter. The hyperparameters of the K-Nearest Neighbors (KNN) model play a crucial role in determining its performance by balancing simplicity, accuracy, and computational efficiency. The N_neighbors parameter specifies the number of nearest neighbors considered for predictions, with smaller values focusing on local patterns and potentially overfitting, while larger values provide smoother predictions but may underfit. The Weight parameter determines how neighbors contribute to the prediction, with "uniform" assigning equal importance to all neighbors and "distance" giving more weight to closer neighbors, which can enhance performance on data with spatial significance. The Algorithm parameter specifies the algorithm for finding nearest neighbors, with automatically choosing the most suitable method, ball_tree and kd_tree being efficient for low to moderate dimensions, and brute being computationally expensive but effective for high-dimensional datasets. The Leaf_size controls the size of the leaf in the tree-based search

algorithms, affecting the speed and memory efficiency of the model; smaller values ensure faster search but may increase computational overhead for large datasets. Finally, the P parameter defines the power parameter for the Minkowski distance, with 1 and 2 affecting how distances are calculated and influencing how neighbors are determined. Proper tuning of these hyperparameters is essential to optimize KNN's performance by reducing overfitting, ensuring accurate predictions, and maintaining computational efficiency.

Grid Search Cross Validation (GridSearchCV) is a technique utilized in machine learning to systematically explore various combinations of hyperparameters and determine the most optimal set for a given model. In order to ensure the reliability and generalizability of the model, a common practice is to employ 10-fold cross-validation. This technique involves splitting the data into 10 equal subsets or folds. Next, nine of these subsets are used to train the model, while the remaining fold is used to validate it. Ten times, this process is repeated, with each subset serving as the validation set alternately. A more accurate and stable estimation of the model's performance can be achieved by averaging the results obtained from each iteration. The model is evaluated using a separate test dataset following the hyperparameter tuning process. Key performance metrics such as Mean Squared Error (MSE) and R^2 scores are typically computed for regression tasks, while accuracy is calculated for classification tasks to assess the model's effectiveness. Developing and refining KNN models is a multifaceted and iterative process that demands careful attention to detail. This process encompasses thorough data preprocessing, meticulous hyperparameter optimization, and robust cross-validation techniques. By diligently following these steps, the predictive capabilities, robustness, and versatility of the model can be significantly enhanced, thereby improving its

performance across a wide array of machine learning tasks. It is essential to emphasize the continuous evaluation and refinement of models, as this is paramount in achieving high-performance machine learning models capable of generating precise and dependable predictions, particularly in the realm of predicting mechanical properties of alkali-activated concrete.

As mentioned above, multicollinearity affects the performance of KNN and SVM. In KNN, multicollinearity affects model by distorting distance-based calculations, leading to biased similarity measures. Since KNN depends on distance metrics such as Euclidean or Manhattan distance, correlated variables create redundant information, causing the model to overemphasize certain features. This results in misleading neighbor assignments, increased computational complexity, and reduced interpretability (Jolliffe and Cadima, 2016). To mitigate these effects, Principal Component Analysis (PCA) is a powerful technique that transforms correlated variables into a new set of uncorrelated features called principal components. PCA reduces dimensionality by capturing the maximum variance in fewer components, allowing KNN to operate efficiently in a lower-dimensional space while retaining essential information. This approach eliminates redundant information, improves model interpretability. By using PCA, KNN can make more accurate and robust predictions, particularly in high-dimensional datasets (Abdi and Williams, 2010).

6. Results and discussions

6.1 Experiment results

The entire mix proportion for the experiment in this investigation is shown in section 4.2, and the mix will be divided into three groups and a trial group.

Group 1 has a fly ash to GGBS ratio of 70:30, a total weight of cementitious materials of 370kg/m^3, an alkali activator to cementitious materials ratio of 0.4, the water glass solution to $NaOH$ solution ratio is 2.5, the Modulus ratio in alkali activator is 1, and the water to cementitious materials ratio is 0.32. Other mixes can be categorized with the same mix parameters in the other two groups, group 2 and group 3. Each group of alkali-activated has 10 mixtures that complete both the compressive strength and flexural strength test. In each group, the packing density of the fine aggregate and coarse aggregate combination is the only parameter. Because in different groups, the fly ash to GGBS ratio, total weight of cementitious materials, alkali activator to cementitious materials ratio, water glass solution to $NaOH$ solution ratio, modulus ratio in alkali activator, and water to cementitious materials ratio are not maintained, the same value, therefore the influence of these values on alkali-activated concrete will be described and discussed in the machine learning result part.

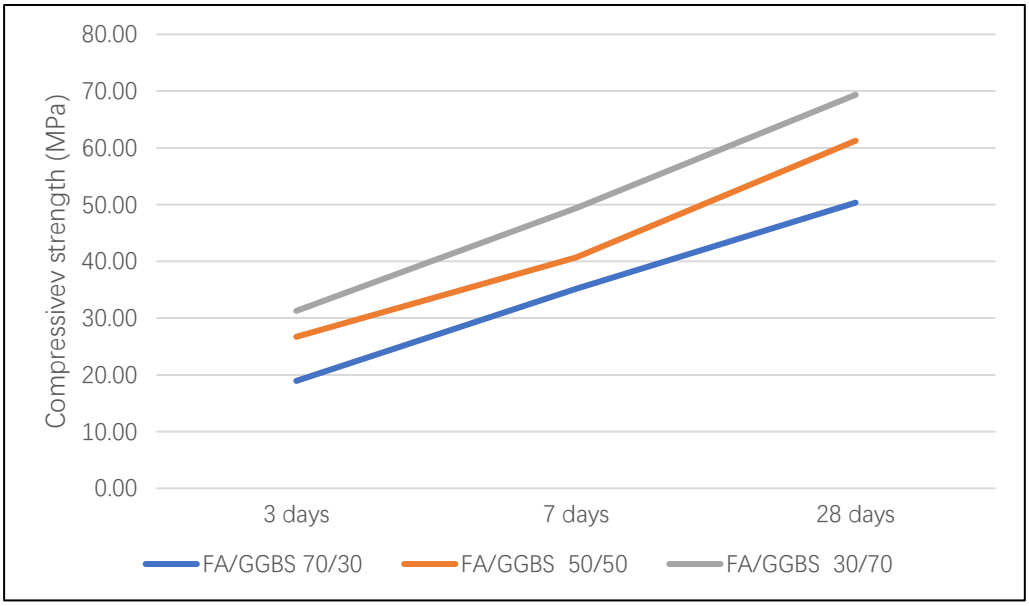

Figure 18. Compressive strength of alkali-activated concrete (Trial Group)

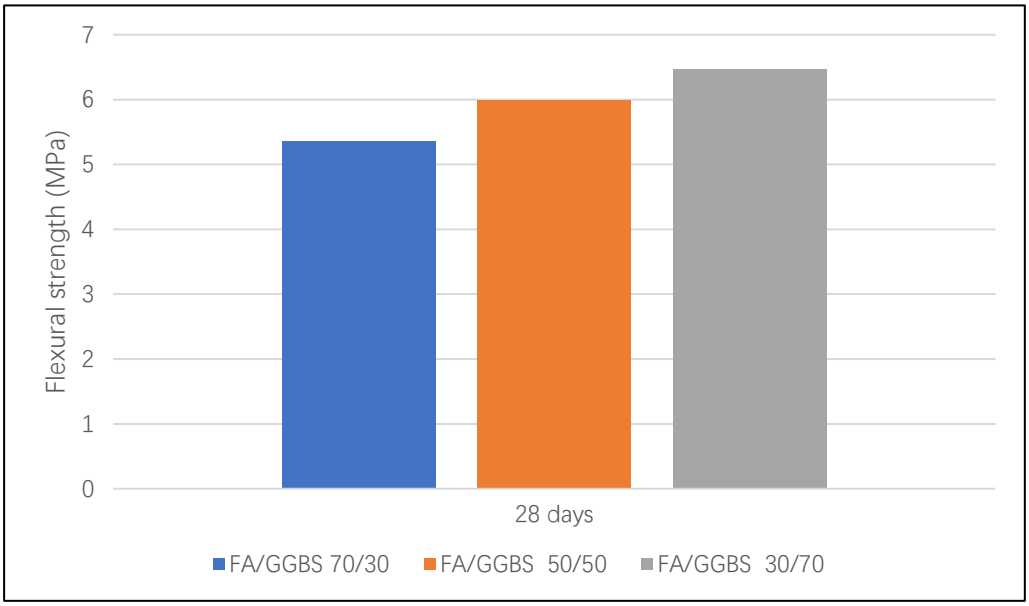

Figure 19. 28days Flexural strength of alkali-activated concrete (Trial Group)

In trial group, contain all parameters except fly ash to GGBS ratio. Figure 18 shows the compressive strength development of alkali-activated concrete at 3, 7, and 28 days for three different fly ash to GGBS ratios: 70/30, 50/50, and 30/70. The results show a clear trend of increasing compressive strength with curing time across all mixtures. Among the three ratios, the 30/70 mix demonstrates the highest compressive strength at all curing stages, achieving over 70 MPa at 28 days. The 50/50 mix follows with intermediate strength values, while the 70/30 mix consistently exhibits the lowest compressive strength. These results indicate that the GGBS content plays a critical role in strength development, with higher GGBS proportions contributing significantly to improved mechanical performance, likely due to its higher calcium content and reactivity in the alkali-activation process. (Rashad, 2013) (Lee and Lee, 2013) Figure 19 displays the flexural strength at 28 days for alkali-activated concrete prepared with varying fly ash to GGBS ratios: 70/30, 50/50, and 30/70. A clear trend is observed, where an increase in the proportion of GGBS correlates with higher flexural strength. The 30/70 mix achieves the greatest flexural strength, approximately 6.5 MPa, followed by the 50/50 mix with a strength of around

6.0 MPa. The 70/30 mix records the lowest flexural strength at approximately 5.5 MPa.

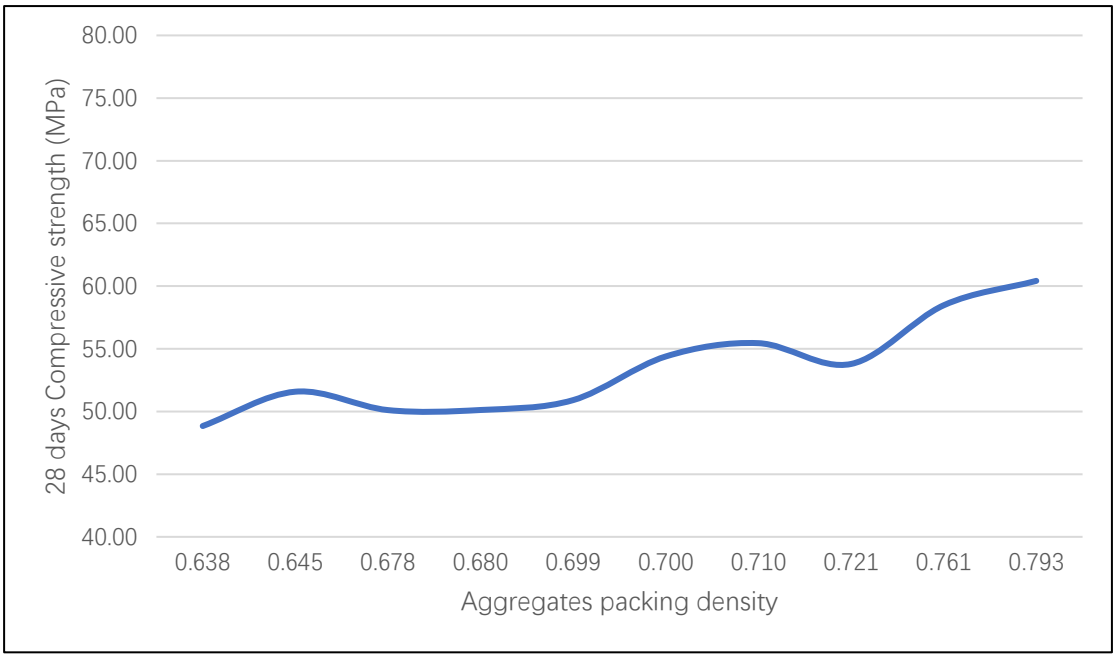

Figure 20. 28 days compressive strength of alkali-activated concrete (Group 1)

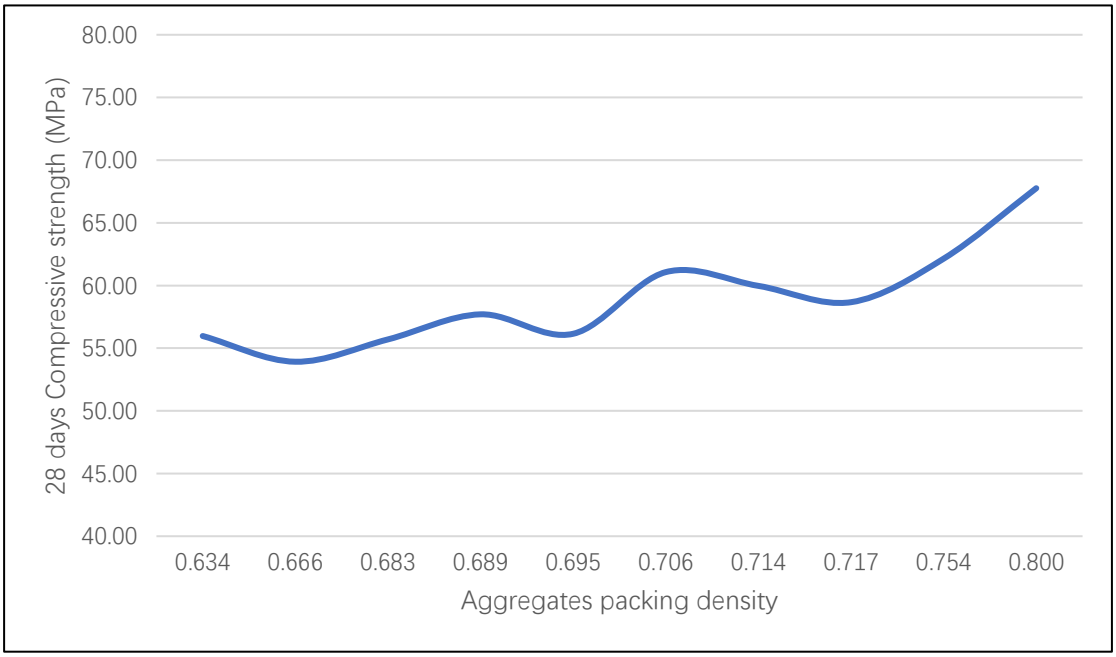

Figure 21. 28 days compressive strength of alkali-activated concrete (Group 2)

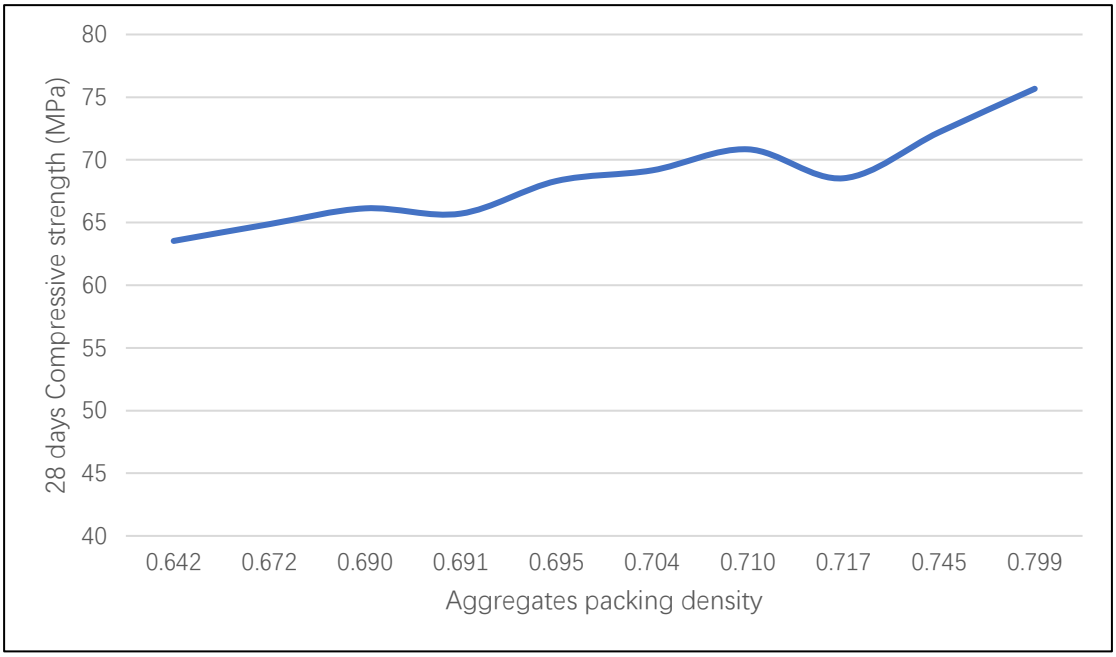

Figure 22. 28 days compressive strength of alkali-activated concrete (Group 3)

Figure 20, Figure 21, and Figure 22 show the compressive strength of alkali-activated concrete after 28 days of curing. In each group, increasing the packing density of aggregates results in an increase in compressive strength; in group 1, the compressive strength increases by 23.7% with an increase in packing density of 24.3%. For group 2 and group 3, the compressive strength increases by about 25.7% and 16.7%, with increases in packing density at 20.1% and 18.9%, respectively. From the diagrams above, it is notable that in all groups of alkali-activated mixtures, the compressive strength demonstrated an upward trend as the packing density value of the aggregate system increased. The relationship between packing density and compressive strength of alkali-activated concrete shows the same trend in the experimental investigation completed by Niyazuddin and Umesh (Niyazuddin and B, 2023). Also, this is in agreement with Karadumpa and Pancharathi (Karadumpa and Pancharathi, 2021). This can be explained by particle packing theory: when packing density increases, the aggregate in concrete will be denser, the interlocking of aggregates will increase, and void content will be reduced, so the concrete

achieves higher compressive strength. Furthermore, as the packing density rises, the left paste beyond the void volume contributes to improved compatibility, resulting in enhanced strength. The previous studies also proved a similar result, which was that the compressive strength would increase as the packing density increased (Nanthagopalan and Santhanam, 2012).

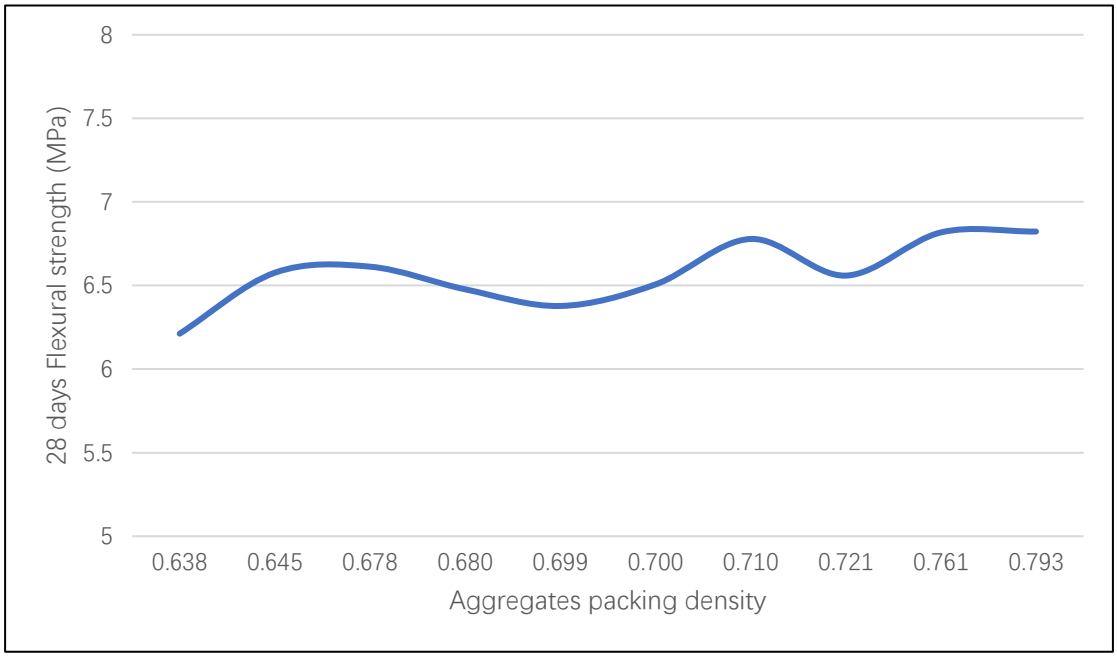

Figure 23. 28 days flexural strength of alkali-activated concrete (Group 1)

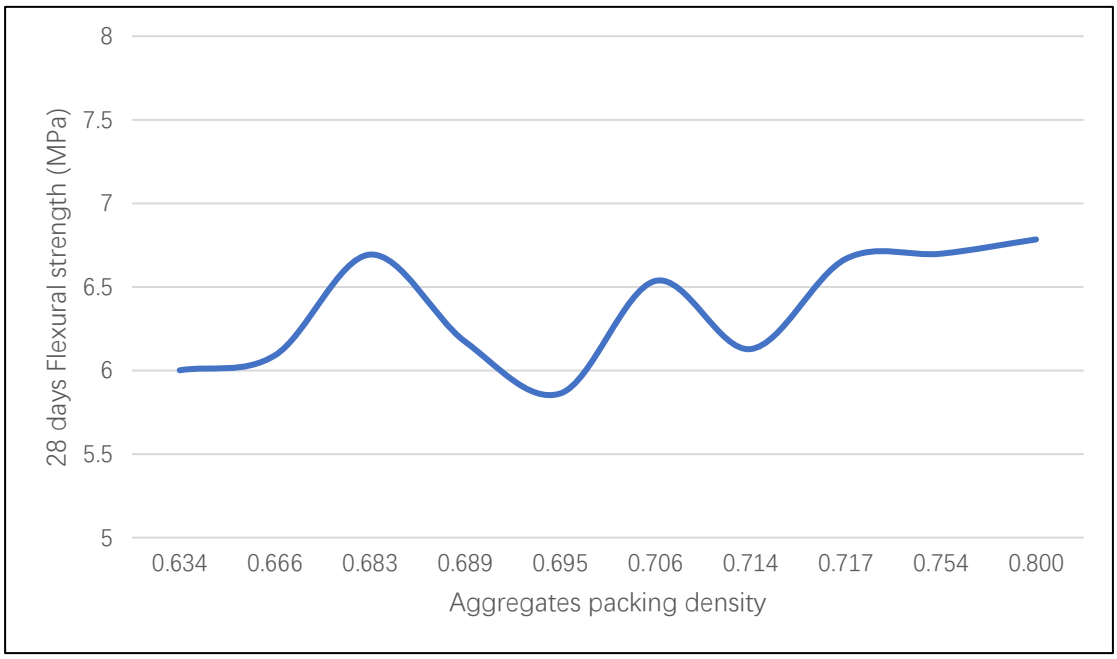

Figure 24. 28 days flexural strength of alkali-activated concrete (Group 2)

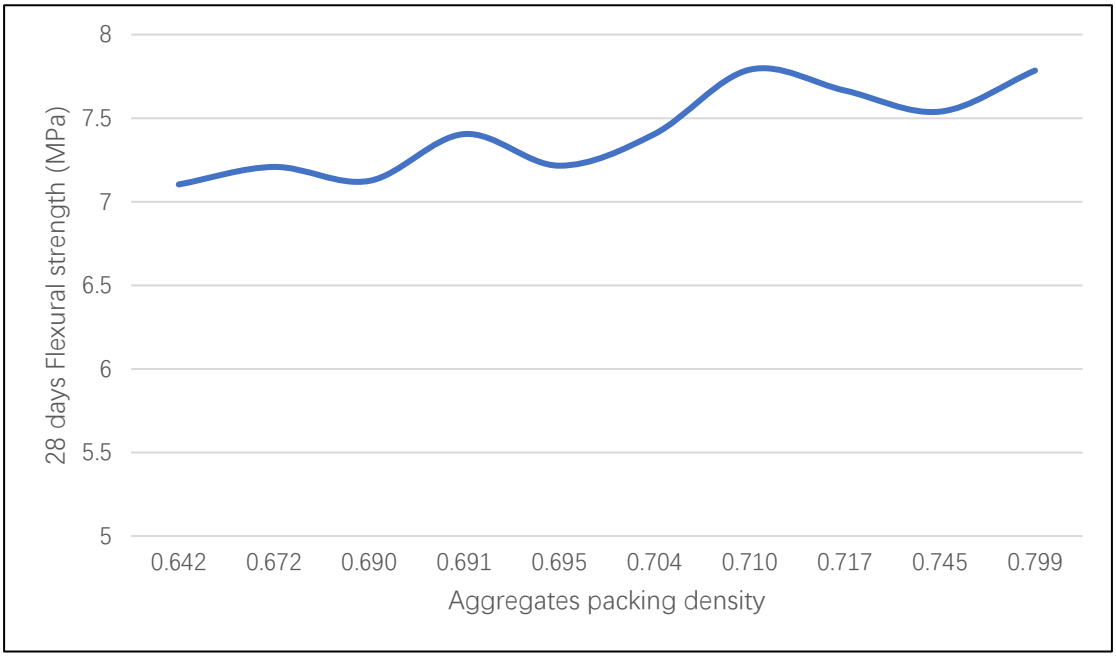

Figure 25. 28 days flexural strength of alkali-activated concrete (Group 3)

Figure 23, Figure 24, and Figure 25 show the relationship between 28 days of flexural strength and aggregate packing density for the flexural strength of alkali-activated concrete. For each group of alkali-activated concrete prism, the flexural strength increases by 3.7%, 11.4%, and 8%, while packing density increases by about 25.7%, 20.1%, and 18.9%, respectively. Fennis and Walraven found the same trend between flexural strength and packing density (Fennis and Walraven, 2012).

6.2 Machine learning dataset

Table 10. Statistical descriptive analysis of compressive strength dataset

Variables	Weight of CEM	SS to SH ratio	FA to GGBS ratio	Activator to CEM ratio	Water to CEM ratio	Ms	Packing density	Curing time	Compressive strength
Unit	kg/m^3	None	None	None	None	None	None	days	MPa
Mean	356.364	2.5	1.254	0.491	0.372	0.958	0.707	12.667	44.932
St.D.	12.512	0.040	0.802	0.083	0.400	0.165	0.042	11.020	14.529
Min	340	2.0	0.429	0.4	0.32	0.72	0.634	3	18.93
Max	370	3.0	2.333	0.6	0.42	1.12	0.800	28	75.67

Table 11. Statistical descriptive analysis of flexural strength dataset

Variables	Weight of CEM	SS to SH ratio	FA to GGBS ratio	Activator to CEM ratio	Water to CEM ratio	Ms	Packing density	Flexural strength
Unit	kg/m^3	None	None	None	None	None	None	MPa
Mean	356.364	2.5	1.254	0.491	0.372	0.959	0.707	6.710
St.D.	12.641	0.395	0.810	0.084	0.400	0.167	0.042	0.584
Min	340	2.0	0.429	0.4	0.320	0.72	0.634	5.364
Max	370	3.0	2.333	0.6	0.420	1.12	0.800	7.790

In the dataset used for compressive strength prediction, as shown in Table 10 it includes various variables that affect the compressive strength of alkali-activated concrete. Eight variables have been chosen as input variables for machine learning models training and testing, the weight of cementitious materials (CEM) $in\ kg/m^3$, the sodium silicate to sodium hydroxide ratio (SS to SH ratio), the fly ash to ground granulated blast-furnace slag ratio (FA to GGBS ratio), the activator to cementitious materials ratio (Activator to CEM ratio), the

water to cementitious materials ratio (Water to CEM ratio), the modulus in activator (Ms), packing density, and curing time in days. The dataset's statistical properties are as follows: the mean weight of cementitious materials is 356.364 kg/m^3 with a standard deviation of 12.512 kg/m^3, the SS to SH ratio averages at 2.5 with a standard deviation of 0.040, and the FA to GGBS ratio has a mean of 1.254 and a standard deviation of 0.802. The activator to cementitious materials ratio averages 0.491 with a standard deviation of 0.083, while the water to cementitious materials ratio has a mean of 0.372 and a standard deviation of 0.400. The activator solution's modulus has an average of 0.958 with a standard deviation of 0.165, the packing density averages 0.707 with a standard deviation of 0.042, and the curing time averages 12.667 days with a standard deviation of 11.020 days. The compressive strength, the target variable, has a mean of 44.932 MPa with a standard deviation of 14.529 MPa. The entire dataset involves 99 data points for machine learning model training and testing. Table 11 shows the statistical descriptive analysis of the flexural strength dataset, with the same input value as the compressive strength dataset except for curing time; for flexural strength, the investigation only considers the flexural strength measure from a specimen after 28 days of curing. The statistical properties of the dataset demonstrate that the mean weight of cementitious materials is 356.364 kg/m^3 and the standard deviation is 12.641 kg/m^3, the mean of the ratio of SS to SH is 2.5, and the standard deviation is 0.395. The mean of the ratio of FA to GGBS is 1.254, and the standard deviation is 0.810. The mean of the ratio of activator to cementitious materials is 0.491, and the standard deviation is 0.084; the mean of the water-cementitious materials ratio is 0.372, and the standard deviation is 0.400. The modulus of the activator solution has a mean of 0.959 and a standard deviation of 0.167; the mean of the bulk density is 0.707, and the standard deviation is 0.042. The mean of the target variable flexural strength is 6.710 , and the standard

deviation is 0.584 MPa.

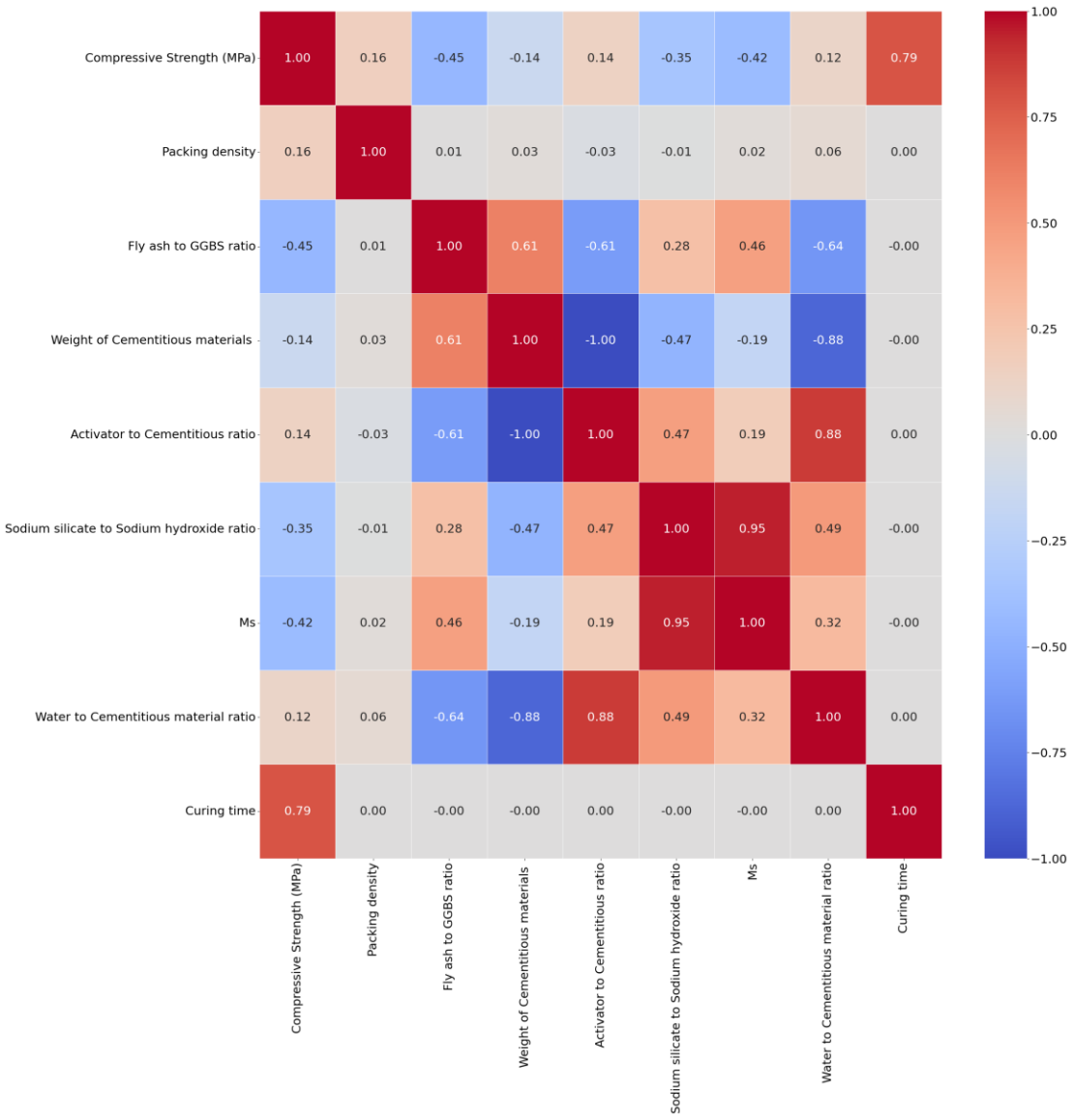

Figure 26. Feature correlation heat-map of compressive strength dataset

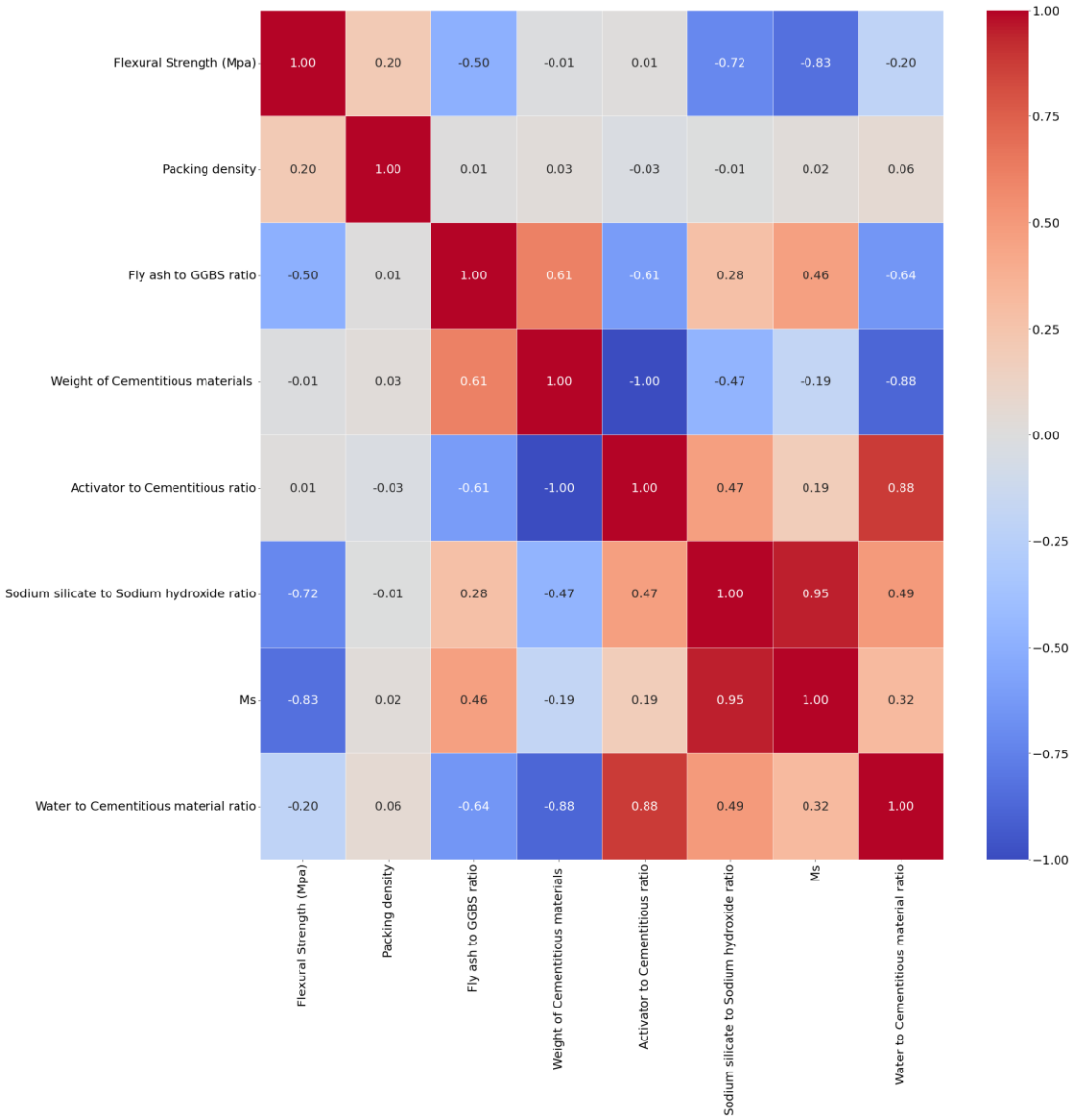

Figure 27. Feature correlation heat-map of flexural strength dataset

Figure 26 provides a visual presentation of the relationship in various variables; in the heat map, the color intensity presents the strength and direction of the correlation in variables; the darker the color, the stronger the correlation. Also, the blue color means a negative relationship, and the red color represents a positive relationship. A significant positive correlation exists between compressive strength and curing time; the correlation coefficient is 0.79, which is the highest of all variables; this means the longer curing times result in higher compressive strength of alkali-activated concrete. This relationship has been

proved by previous study (Sun et al., 2023b). Alkali-activated concrete's compressive strength has a mild negative connection with the ratio of sodium silicate to sodium hydroxide and a moderate negative correlation with the fly ash to GGBS ratio; the correlation coefficient for these factors are -0.45 and -0.35, respectively. The increases in fly ash to GGBS ratio in the mixture will decrease the compressive strength of alkali-activated concrete, and the increases in sodium silicate to sodium hydroxide ratio will reduce the compressive strength. The modulus in the alkali activator has a negative correlation with compressive strength either, with a correlation coefficient of -0.42. The other four input values, packing density, weight of cementitious materials, activator to cementitious ratio, and water to cementitious material ratio, have the correlation coefficient with compressive strength of alkali-activated concrete are 0.16, -0.14, 0.14, and 0.12, respectively. According to Sun et al., the fly ash to GGBS ratio in alkali-activated concrete plays a crucial role in compressive strength development, and the water to cementitious materials ratio has less importance in compressive strength, which is the adverse of traditional cement based materials (Sun et al., 2023a). The packing density of aggregates in alkali-activated concrete shows a similar correlation as water to cementitious ratio and weight of cementitious ratio. In traditional cement concrete, these three variables similarly influence the compressive strength of concrete (Pallapothu et al., 2023). The correlation between the compressive strength of alkali-activated concrete and the packing density of aggregates shows that packing density is one of the influential factors in alkali-activated concrete, and it will affect the prediction of compressive strength by using machine learning models.

Figure 27 shows the heat map for the flexural dataset. The graph indicates that the flexural strength of alkali-activated concrete has a strong negative

correlation with modulus in alkali activator and sodium silicate to sodium hydroxide ratio, where the correlation coefficient is -0.83 and -0.72. The fly ash to GGBS ratio has a moderate negative relationship with flexural strength with a -0.5 correlation coefficient. The packing density has a positive correlation coefficient of 0.2, which means that when packing density increases, the flexural strength will increase. Moreover, the water to cementitious material ratio negatively correlates with the target value. The influence of activator to cementitious ratio and weight of cementitious materials ratio only has a negligible impact on flexural strength in alkali-activated concrete, which can be ignored in mix design.

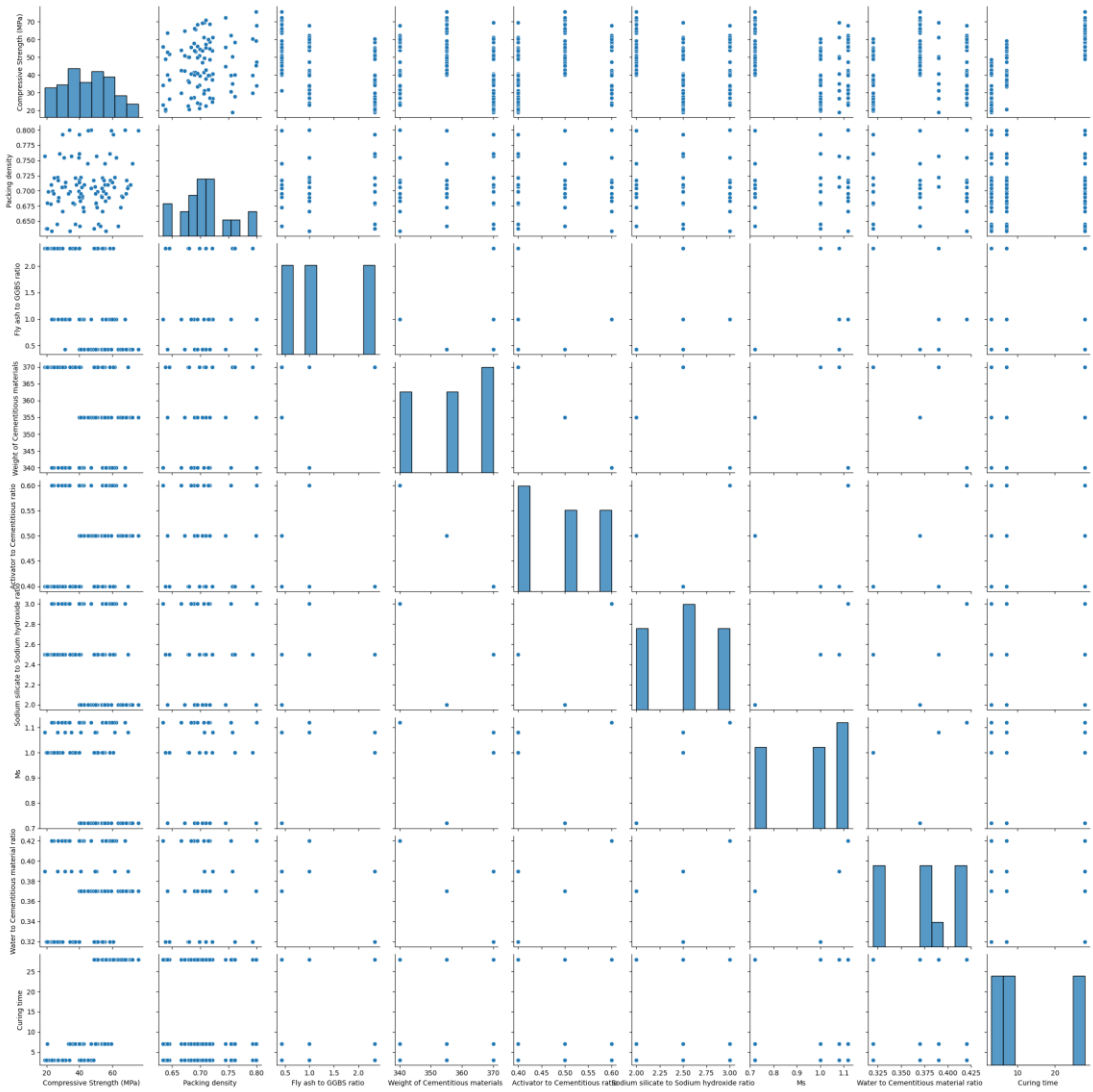

Figure 28. Pair plot of compressive strength dataset

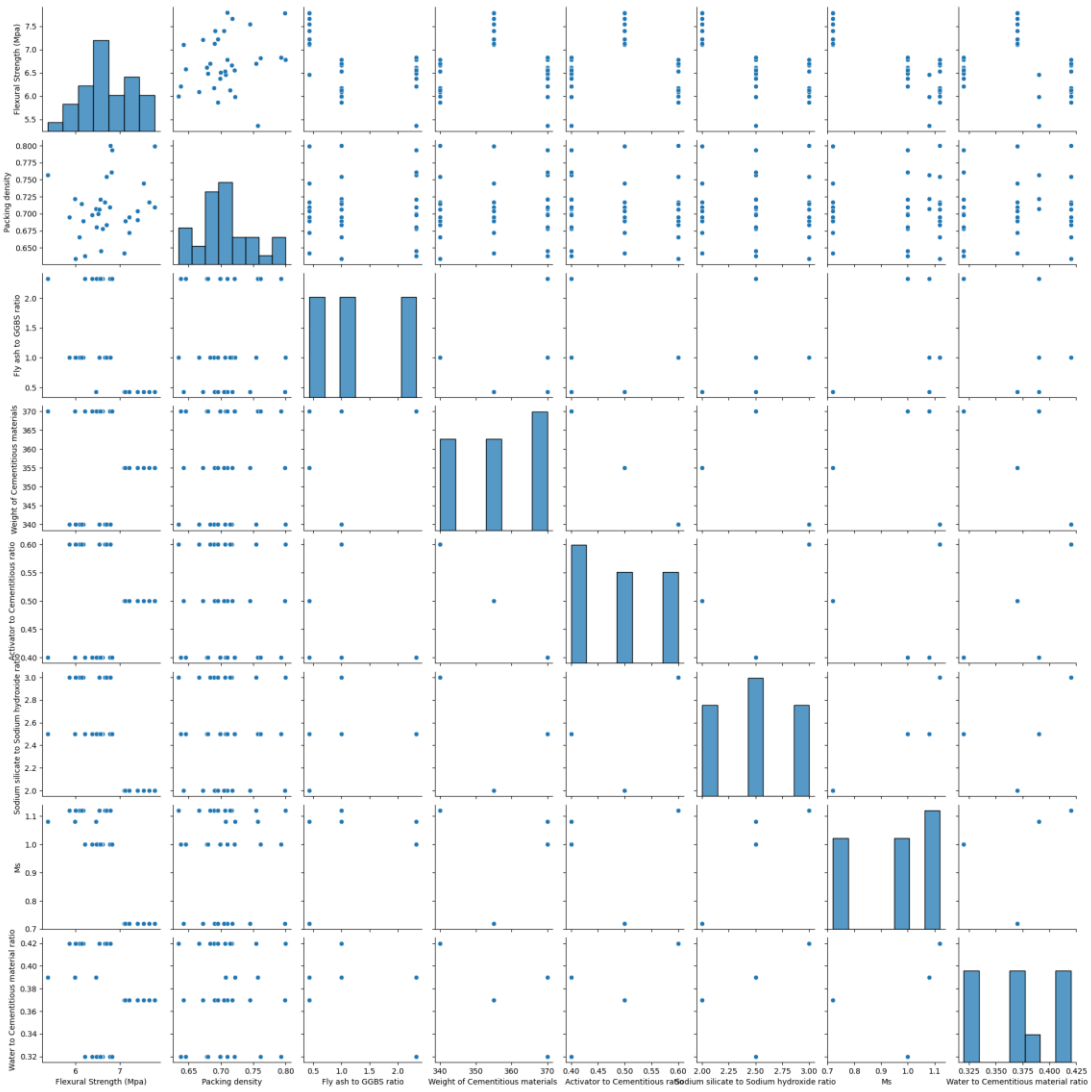

Figure 29. Pair plot of flexural strength dataset

Figure 28 and Figure 29 are the pair plots of the datasets, a visualization technique that can discover the pairwise correlations within these two datasets. This method creates an axes grid with the y-axis representing the rows and the x-axis representing the columns. While the diagonal plots show the distribution of a single variable, each plot inside the grid shows the correlation between two variables. The illustration demonstrates the connections among the different features present in the dataset. The information is condensed to facilitate the identification of potential patterns and relationships among the variables.

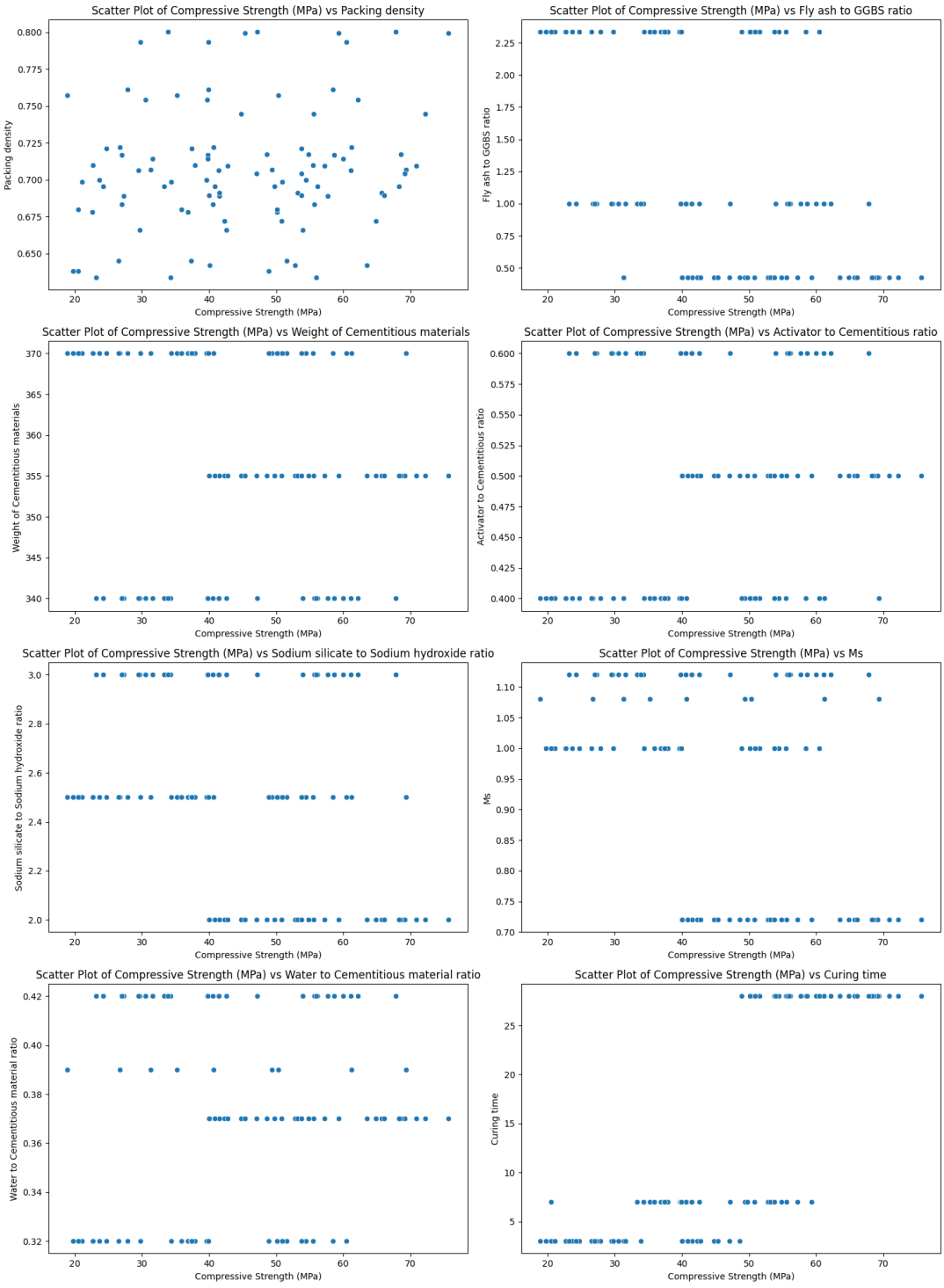

Figure 30. Scatter plot of compressive strength dataset

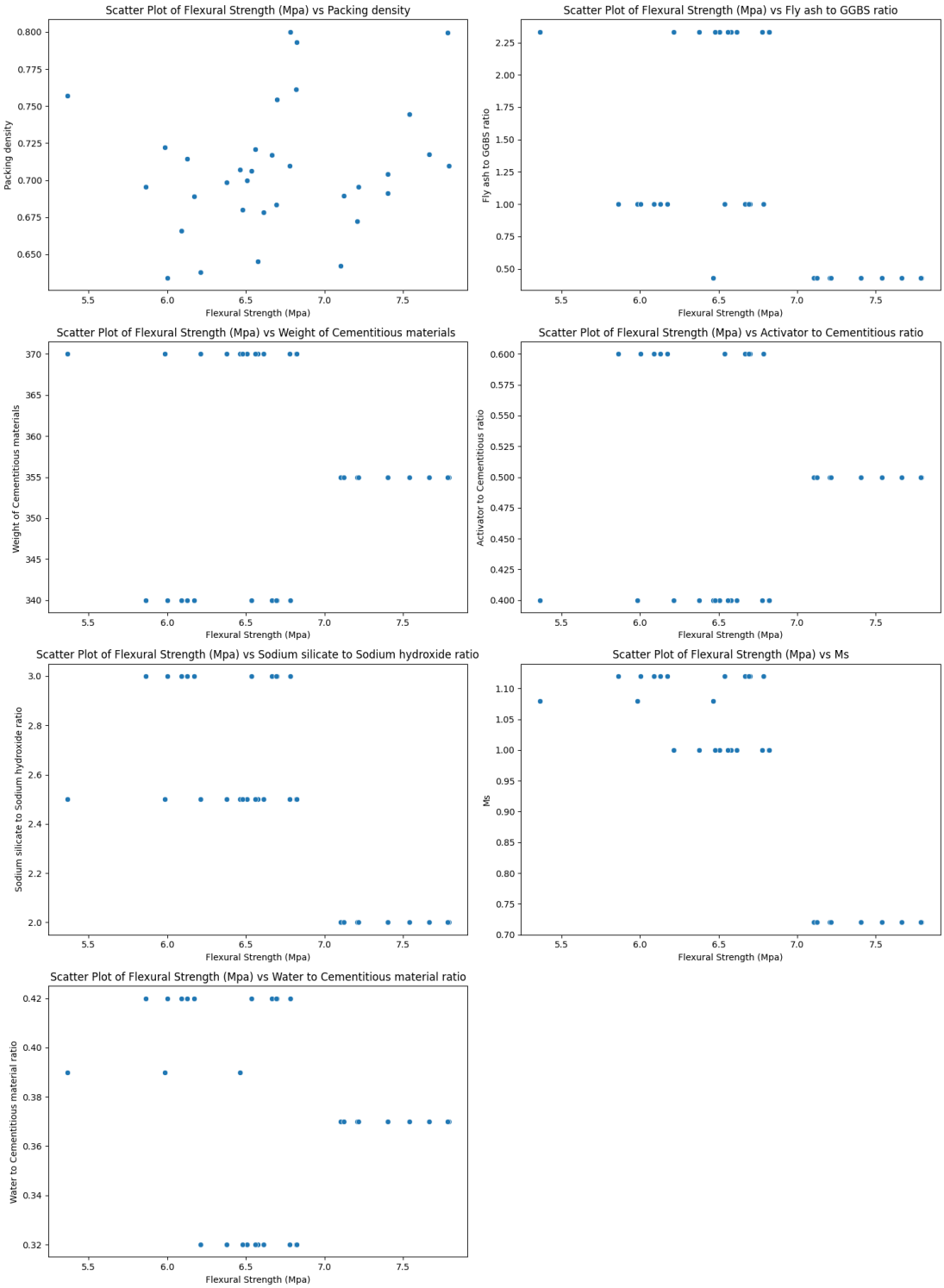

Figure 31. Scatter plot of flexural strength dataset

Figure 30 is the scatter plot of the compressive strength dataset. A scatter plot is a powerful tool in data analysis to demonstrate the distribution of each input variable with target value; it also provides a clear visualization of the dataset,

which highlights the concentration and sparsity area of data points. In the dataset, the input variables fly ash to GGBS ratio, the weight of cementitious materials, activator to cementitious ratio, sodium silicate to sodium hydroxide ratio, modulus of alkali activator, and curing time can be grouped into three parts; these input variables have three different values in the dataset, for packing density, which has a comprehensive value between the range of 0.6 to 0.8. The dataset has three different values in the fly ash to GGBS ratio, the weight of cementitious ratio, the activator to cementitious ratio, and the sodium silicate to sodium hydroxide ratio. In addition, the modulus in the activator and water to cementitious ratio had four different values in the dataset. Curing times were chosen as 3 days, 7 days, and 28 days after alkali-activated concrete demold. The compressive strength of alkali-activated concrete is within the range of 19 MPa to 75 MPa. Figure 31 shows the scatter plot of the flexural strength dataset; according to the flexural strength only measured after 28 days of curing, lesser data were included in the flexural strength dataset. The flexural strength dataset has the same input variables as the compressive strength dataset except for curing time.

6.3 Random Forest

Table 12. Model evaluation results of compressive strength prediction

	MAE	RMSE	MSE	R^2	MAPE
Training set	1.720	2.355	2.546	0.973	4.761
Testing set	2.187	3.325	11.056	0.948	6.245

Table 13. Model evaluation results of flexural strength prediction

	MAE	RMSE	MSE	R^2	MAPE
Training set	0.164	0.226	0.051	0.864	2.598
Testing set	0.199	0.228	0.052	0.775	2.941

Table 14. Hyperparameter configuration for the Random Forest model

Hyperparameter	Search space	Best parameter (compressive strength)	Best parameter (flexural strength)
N_estimators	[100, 200, 300]	300	300
Max_depth	[10, 20, 30]	10	10
Min_samples_split	[2, 5, 10]	2	2
Min_samples_leaf	[1, 2, 4]	2	2
Max_feature	[sqrt, log2, None]	None	sqrt
Bootstrap	[True, False]	Ture	False

Regarding the random forest model, Table 14 shows the hyperparameter tuning search space and the best parameters for compressive and flexural strength predictions. For compressive strength prediction and flexural strength prediction, in tuning hyperparameters in random forest, the search space for each hyperparameter is [100, 200, 300], [10, 20, 30], [2, 5, 10], [1, 2, 4], [the square root of the number of features, the logarithm base 2 of the number of features, all feature], [bootstrap sampling, without bootstrap sampling], respectively. Table 14 also provides the best parameters from Grid search; in compressive strength prediction, the hyperparameters were tuned and selected to be [300, 10, 2, 2, None, Ture]; for flexural strength, they were [300, 10, 2, 2, sqrt, false], respectively. Table 13 presents the performance metrics of compressive strength forecast of the random forest model. The values of MAE, RMSE, MSE, R^2, and MAPE of the training set are 1.720, 2.355, 2.546, 0.973, and 4.761. For the testing set, the performance metrics are 2.187, 3.325, 11.056, 0.948, and 6.245, respectively. In flexural strength prediction from the random

forest model, the best parameters of Max_feature and Bootstrap were tuned as Sqrt and False; other parameters were the same as compressive strength prediction. Table 13 describes the random forest model evaluation results of flexural strength prediction; the values of MAE, RMSE, MSE, cap R squared, and MAPE are 0.164, 0.226, 0.051, 0.864, 2.598, and for the testing set, the values are 0.199, 0.226, 0.051, 0.775, 2.941, respectively.

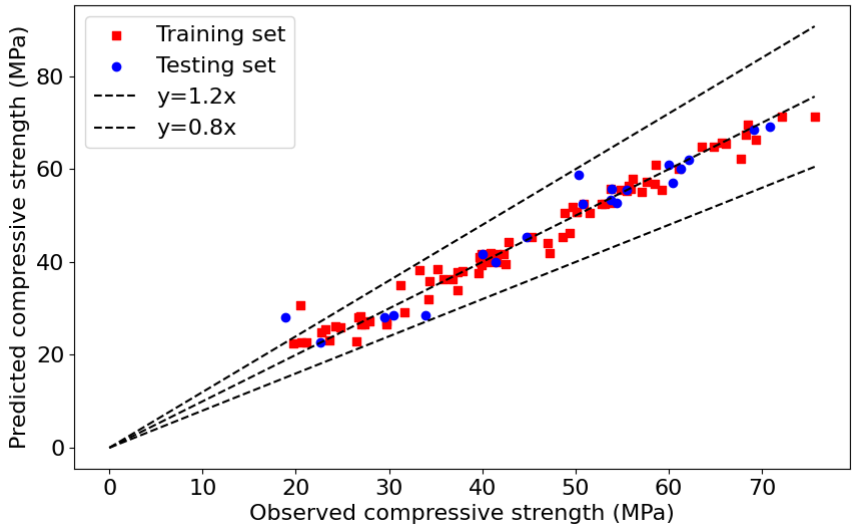

Figure 32. Regression plot of predicted versus observed values in compressive strength prediction

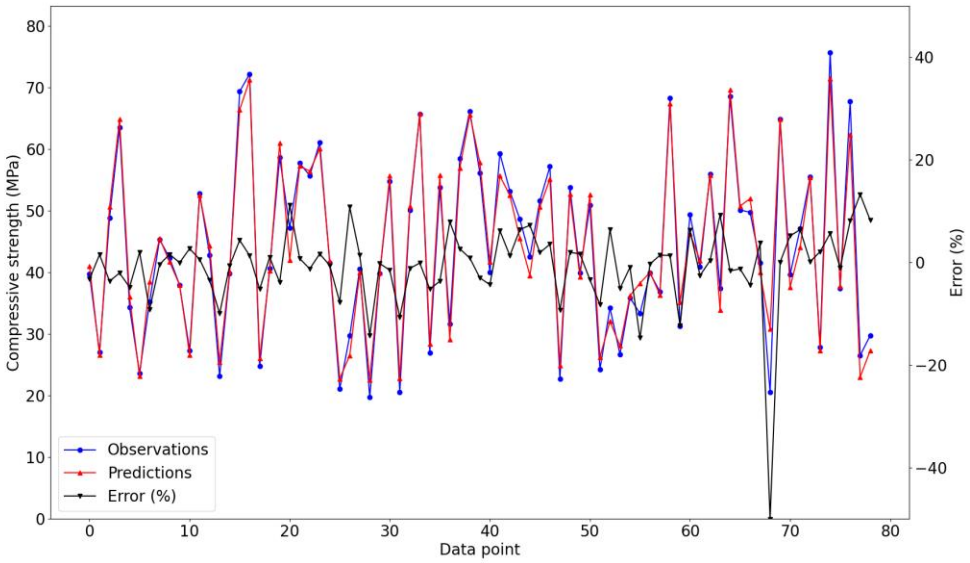

Figure 33. Predictive performance of the Random Forest model in compressive strength prediction

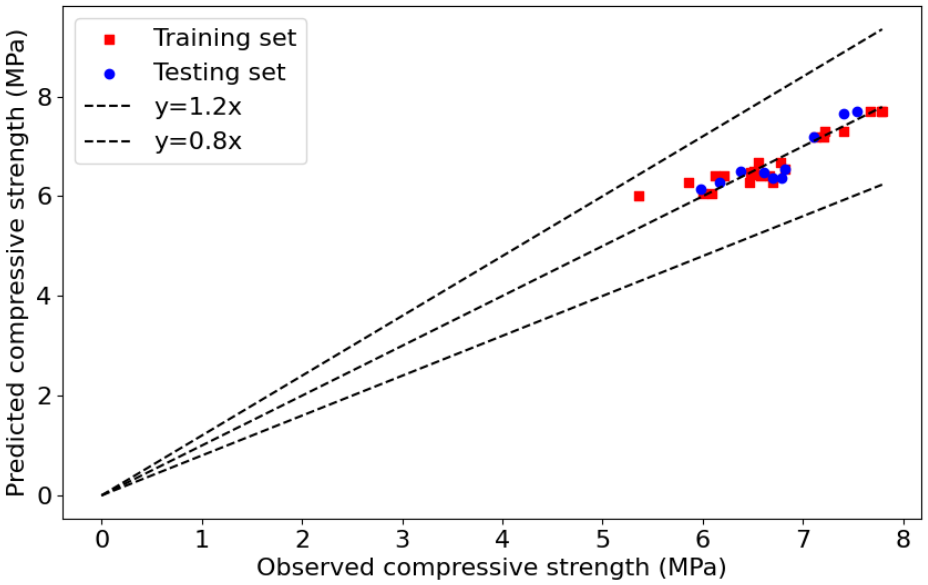

Figure 34. Regression plot of predicted versus observed values in flexural strength prediction

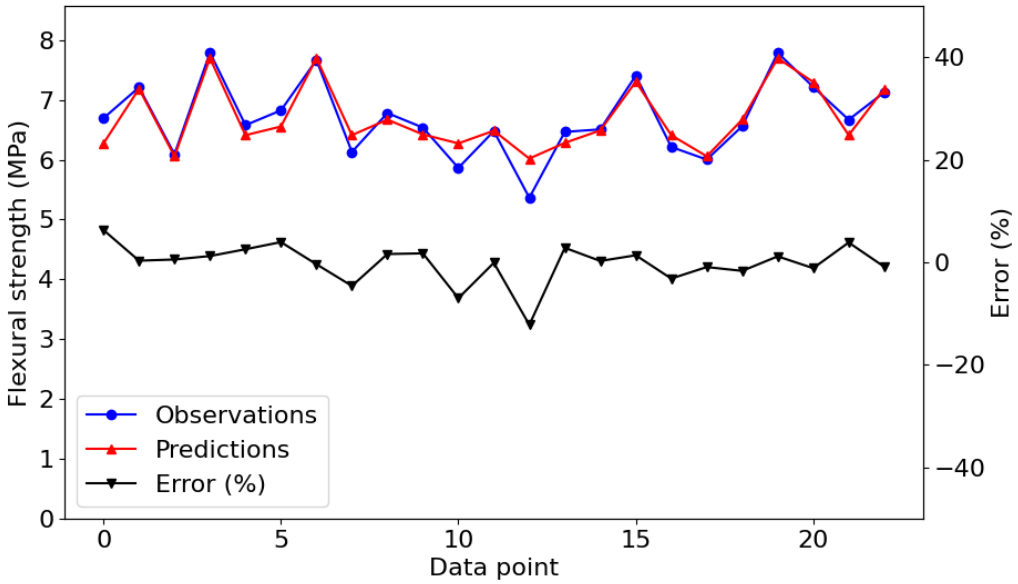

Figure 35. Predictive performance of the Random Forest model in flexural strength prediction

In terms of compressive strength, the random forest findings show an excellent fit between the observation and predictions. This trend can also be found in Figure 32. Most of all, the prediction value is close to the line of ideality, and only two predictions are outside the range of ±20% bound. Figure 33 presents the prediction results versus observed values and errors; the blue line represents the observed values, the red line represents the prediction values, and the black line represents the error in the percentage of observations. It demonstrates that, for the random forest machine learning method, the difference between the experimental value and the predictive value of the majority of samples is really small. The most common error in predictions is located in the range of 0% to 10% of the observation value. It verifies the feasibility of using packing density as an input variable to develop a random forest model for compressive strength prediction of alkali-activated concrete. The random forest model created in this investigation, which set packing density as a new input value, predicts an accurate value in compressive strength of alkali-activated concrete, compared with published models developed for

alkali-activated concrete compressive strength prediction using random forest model; it is able to provide equivalent or superior performance. The random forest from Gomaa et al., Sun et al., and Li et al. got the accuracy of compressive strength prediction are 0.944, 0.85, and 0.94, respectively (Gomaa et al., 2021) (Sun et al., 2023a) (Li et al., 2023).

In flexural strength prediction, Figure 34 shows the regression plot of predicted versus observed values, all training set predictions, and testing values within the range of plus or minus20% of the ideality line, which means a good accuracy of random forest model in flexural strength prediction of alkali-activated concrete when curing time is 28 days. Figure 35 presents the predictive performance of the random forest model in flexural strength prediction, as the black line in the figure is the error between predictions and observations in percentage; nearly all of the predictions have an error of 10%. This result proves that the random forest model developed in this investigation has good accuracy in flexural strength prediction with the packing density of aggregates as an input parameter in the machine learning model. Previous studies used equations to predict the flexural strength of GGBS/fly ash-based alkali-activated concrete, the standard error is 0.03, and the R^2 value is 0.35 (Sun et al., 2023a). Compared with flexural strength prediction, the random forest machine learning model in this investigation shows advanced accuracy. However, according to the experiment limitations in this research, the dataset in flexural strength prediction is not comprehensive enough, the more data points need to be added in further model training.

6.4 Extreme Gradient Boosting

Table 15. Model evaluation results of compressive strength prediction

	MAE	RMSE	MSE	R^2	MAPE
Training set	0.775	1.025	1.050	0.995	2.124
Testing set	2.185	3.065	9.392	0.956	5.274

Table 16. Model evaluation results of flexural strength prediction

	MAE	RMSE	MSE	R^2	MAPE
Training set	0.025	0.052	0.003	0.992	0.383
Testing set	0.153	0.234	0.055	0.748	2.367

Table 17. Hyperparameter configuration in Extreme Gradient Boosting model

Hyperparameter	Search space	Best parameter (compressive strength)	Best parameter (flexural strength)
N_estimators	[100, 200, 300]	100	300
Learning_rate	[0.01, 0.1, 0.2]	0.2	0.01
Max_depth	[3, 4, 5]	3	3
Min_child_weight	[1, 2, 3]	2	3
Subsample	[0.8, 0.9, 1.0]	0.8	0.8
Colasample_bytree	[0.8, 0.9, 1.0]	0.9	0.8

For the extreme gradient boosting model, Table 17 shows the hyperparameter turning search space and optimal parameters for mechanical properties prediction using extreme gradient boosting. In compressive and flexural strength prediction, few hyperparameters were selected in the extreme gradient boosting model. The search Spaces for each hyperparameter are [100, 200, 300], [0.01, 0.1, 0.2], [3, 4, 5], [1, 2, 3], [0.8, 0.9, 1.0], [0.8, 0.9, 1.0], respectively. After Grid search, the best hyperparameters of extreme gradient boosting are shown below: N_estimator are 100 and 300, Learning_rate 0.2 and 0.01, Max_depth are both 3, Min_child_weight are 2 and 3, Subsample are both

0.8, Colasample_bytree are 0.9 and 0.8 respectively. Table 15 describes the evaluation results of the extreme gradient boosting model for the prediction of compressive strength. The values of MAE, RMSE, MSE, R^2, and MAPE are 0.775, 1.025, 1.050, 0.995, and 2.124, respectively. For the test set, these values are 2.185, 3.065, 9.392, 0.956, and 5.274. In the training and testing of the extreme gradient boosting model for flexural strength, which is shown in Table 16, the model performance indexes MAE, RMSE, MSE, R^2, and MAPE were 0.025, 0.052, 0.003, 0.992, and 0.383 in the training set, and 0.153, 0.234, 0.055, 0.748 and 2.367 in the test set, respectively. These performance metrics indicate that the extreme gradient boosting model in this investigation is able to provide accurate predictions of the compressive strength of alkali-activated concrete. Compared with previous studies, a gradient boosting regression tree model has been developed by Sun et al. to predict the compressive strength of alkali-activated concrete; the R^2 of the prediction model is 0.94 (Sun et al., 2023a). Li et al. created a machine learning model with gradient boosting and trained with 177 data points in the dataset, the accuracy of the model shows R^2 very close to 1 (Li et al., 2023). And Afzali et al. used a gradient boosting model that performed superior results in compressive strength prediction of metakaolin based geopolymer concrete, with 0.983 of R^2 (Afzali et al., 2024). It denotes that this extreme gradient boosting model can achieve the same or even better accuracy in compressive strength predictions.

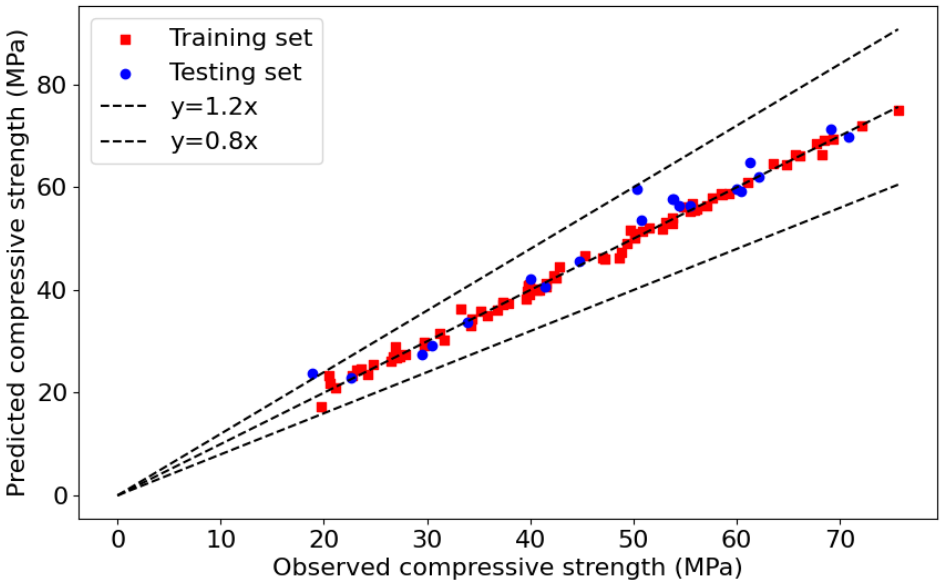

Figure 36. Regression plot of predicted versus observed values in compressive strength prediction by extreme gradient boosting model

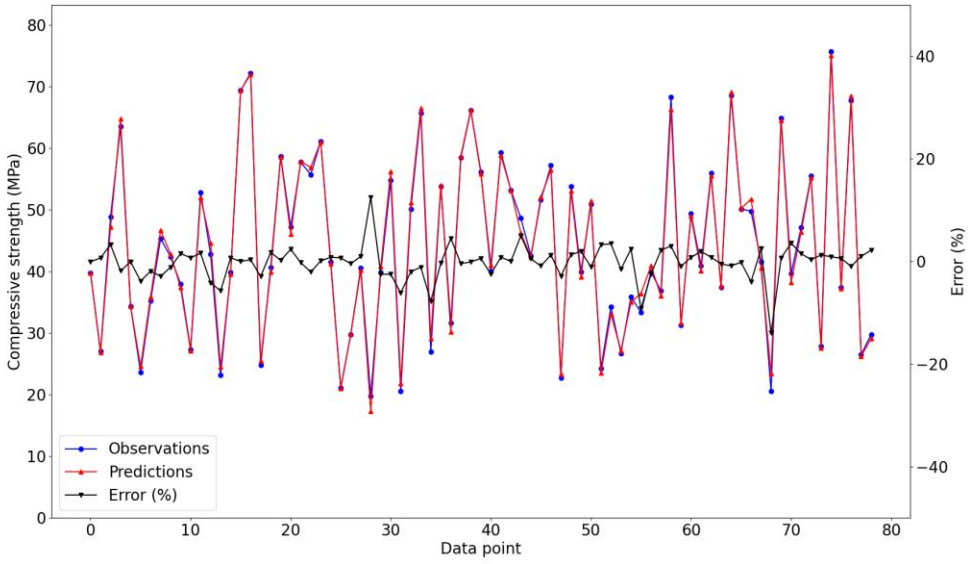

Figure 37. Predictive performance of the extreme gradient boosting model in compressive strength prediction

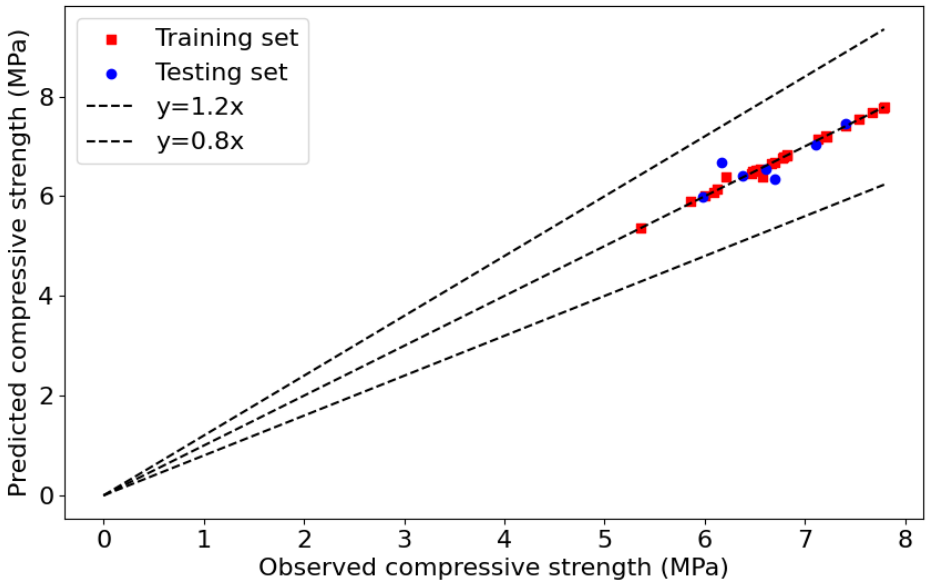

Figure 38. Regression plot of predicted versus observed values in flexural strength prediction

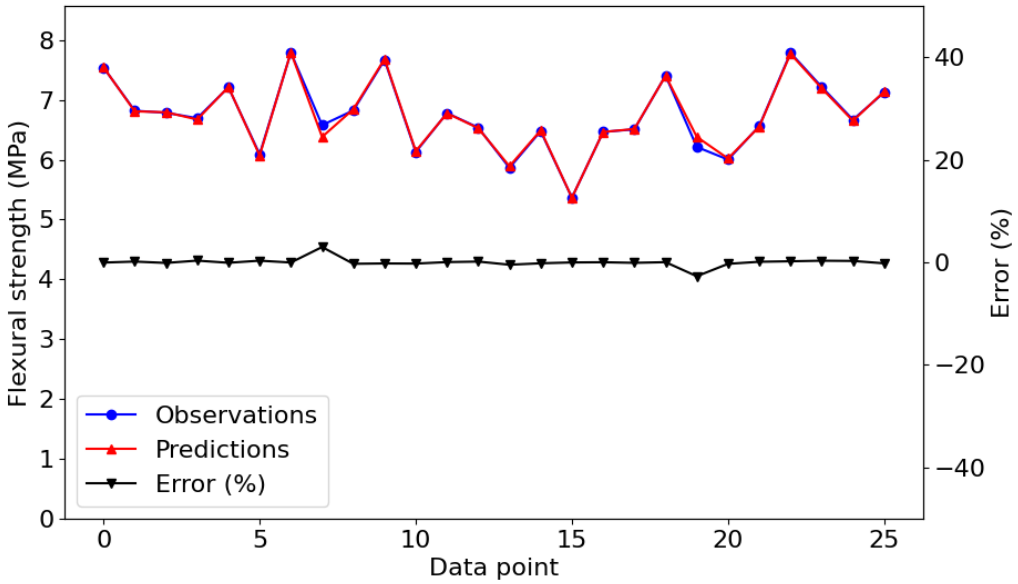

Figure 39. Predictive performance of the Random Forest model in flexural strength prediction

This conclusion is also proved by Figure 36 and Figure 37. The first figure shows that in the training set, nearly all results were close to the ideality line and all

within the range of a plus or $\pm$20% bound to the ideality line; in the testing set, only one point was outside the range. In the second figure, except for one error bigger than 10% of observation, others all keep an error in 10% of observation, which indicates a favorable performance of extreme gradient boosting in compressive strength prediction when packing density is an important input variable in the dataset. For flexural strength prediction, Figure 38 presents the regression plot of flexural strength predicted versus observed values in both the training set and testing set; in this figure, the vast majority of both the training set and testing set data fell into a $\pm$20% error band deviated from experimental results. Figure 39 also shows the same performance; almost all predictions contained an error of less than $\pm$10%. However, in flexural strength prediction, the performance of the testing data is significantly worse than that of the training data. The error values are much higher, and the R^2 value drops to 0.748, indicating that the model does not generalize as well. A significant drop in performance between the training and testing sets shown in flexural strength prediction (R^2 from 0.992 to 0.748). For this case is likely due to a combination of factors. The small dataset which only includes 33 samples limits the model's ability to generalize as normally, even cross-validation and regularization were applied. The dataset may also be affected from inadequate feature relevance, high variance in the target variable, or distribution imbalances between the training and testing sets. Additionally, the presence of noise or outliers in the experimental data could cause the model to overfit the training set while failing to capture meaningful patterns for prediction. The dataset may also suffer from inadequate feature relevance, high variance in the target variable, or distribution imbalances between the training and testing sets. Additionally, the presence of noise or outliers in the experimental data could cause the model to overfit the training set while failing to capture meaningful patterns for prediction. The issues above all influence the performance of XGBoost model

on unknown data.

6.5 Support Vector Machine

Table 18. Model evaluation results of compressive strength prediction

	MAE	RMSE	MSE	R^2	MAPE
Training set	4.371	5.260	27.670	0.865	12.394
Testing set	3.576	4.374	19.128	0.911	9.761

Table 19. Model evaluation results of flexural strength prediction

	MAE	RMSE	MSE	R^2	MAPE
Training set	0.319	0.369	0.136	0.621	4.868
Testing set	0.272	0.354	0.124	0.424	4.268

Table 20. Hyperparameter configuration in support vector machine model

Hyperparameter	Search space	Best parameter (compressive strength)	Best parameter (flexural strength)
C	[0.001, 0.01, 0.1, 0.5, 1, 2]	2	2
Kernel	[linear, rbf]	linear	Linear
Gamma	[scale, auto]	scale	scale
Degree	[2, 3, 4]	2	2
Epsilon	[0.001, 0.01, 0.1, 0.5, 1, 2]	2	0.5

From Table 18 and Table 19, the model performance evaluation metrics results for the support vector machine model predicting compressive and flexural

strength show differing performance levels. For compressive strength prediction, the training set has an MAE of 4.371, RMSE of 5.260, MSE of 27.670, R^2 of 0.865, and MAPE of 12.394%. The testing set shows improved performance with an MAE of 3.576, RMSE of 4.374, MSE of 19.128, R^2 of 0.911, and MAPE of 9.761. Otherwise, in flexural strength prediction, the training set has an MAE of 0.319, RMSE of 0.369, MSE of 0.136, R^2 of 0.621, and MAPE of 4.868%. The testing set has an MAE of 0.272, RMSE of 0.354, MSE of 0.124, R^2 of 0.424, and MAPE of 4.268. The performance metrics presented the performance with low prediction accuracy. In previous studies, Peng and Unluer (2022) archived 0.915 of R^2 using a support vector machine in alkali-activated concrete compressive strength prediction. Afzali got a similar accuracy in support vector machine prediction with a 235 data points size dataset (Afzali et al., 2024). Overall, while the SVM model performs well for compressive strength prediction with high accuracy and generalization, the prediction of flexural strength could benefit from further investigation in the dataset to enhance its predictive power.

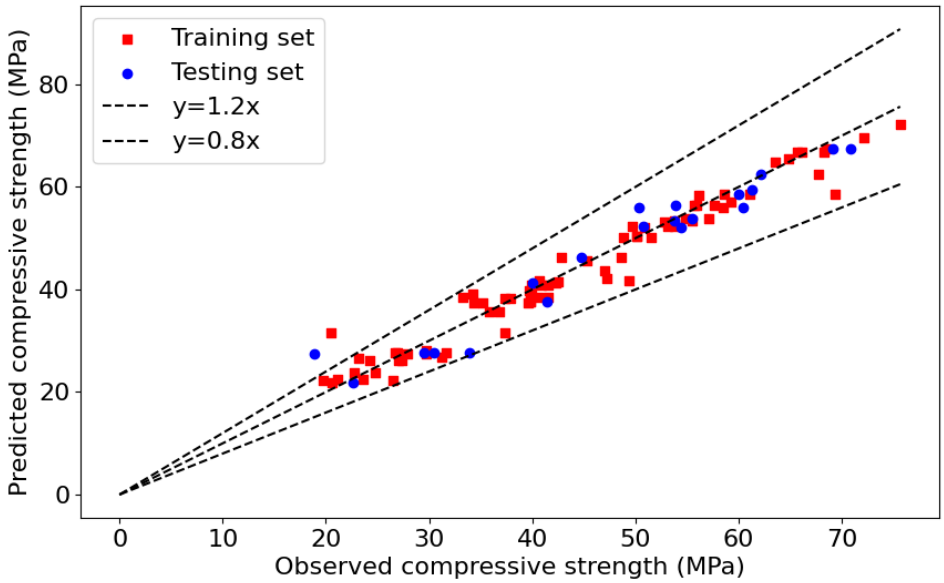

Figure 40. Regression plot of predicted versus observed values in compressive strength prediction by support vector machine model

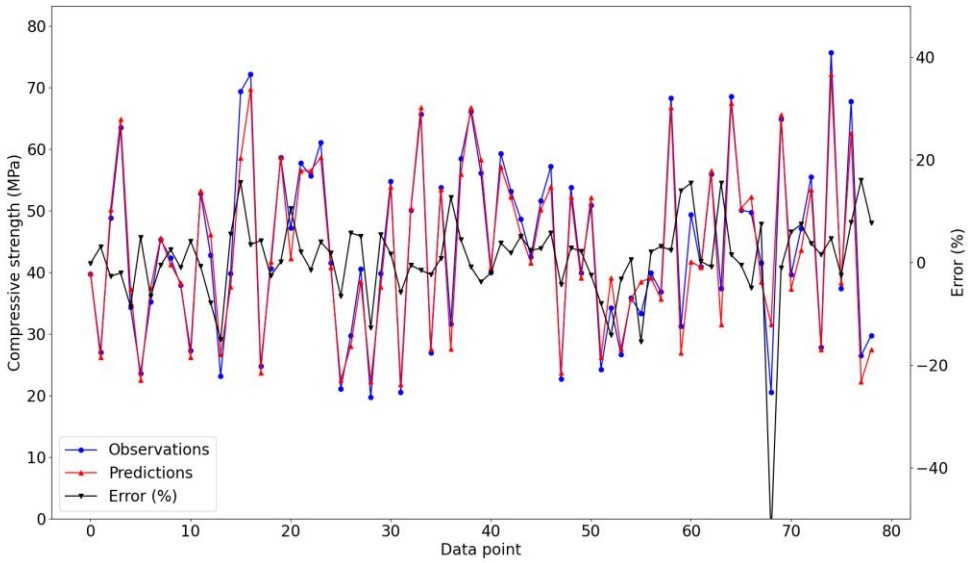

Figure 41. Predictive performance of the Random Forest model in compressive strength prediction by support vector machine model

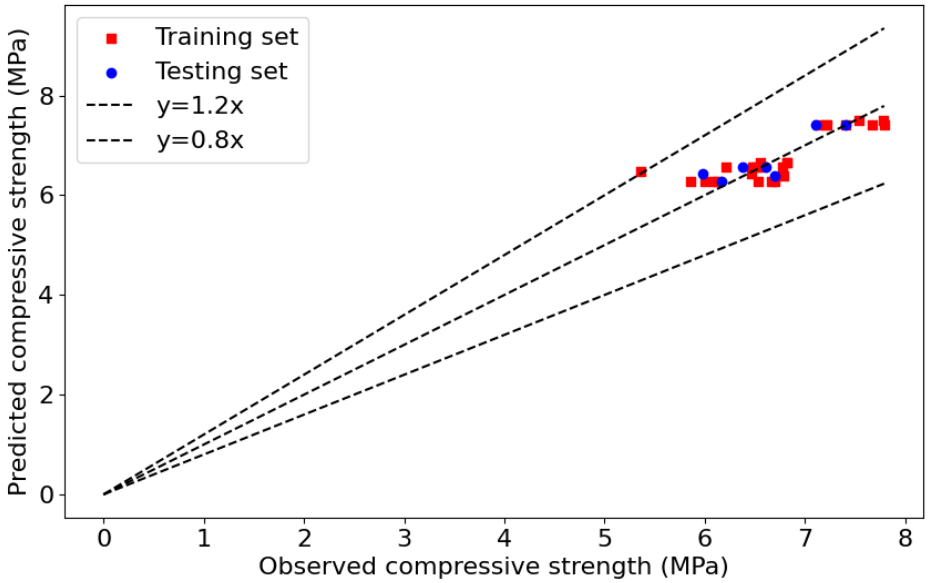

Figure 42. Regression plot of predicted versus observed values in flexural strength prediction by support vector machine model

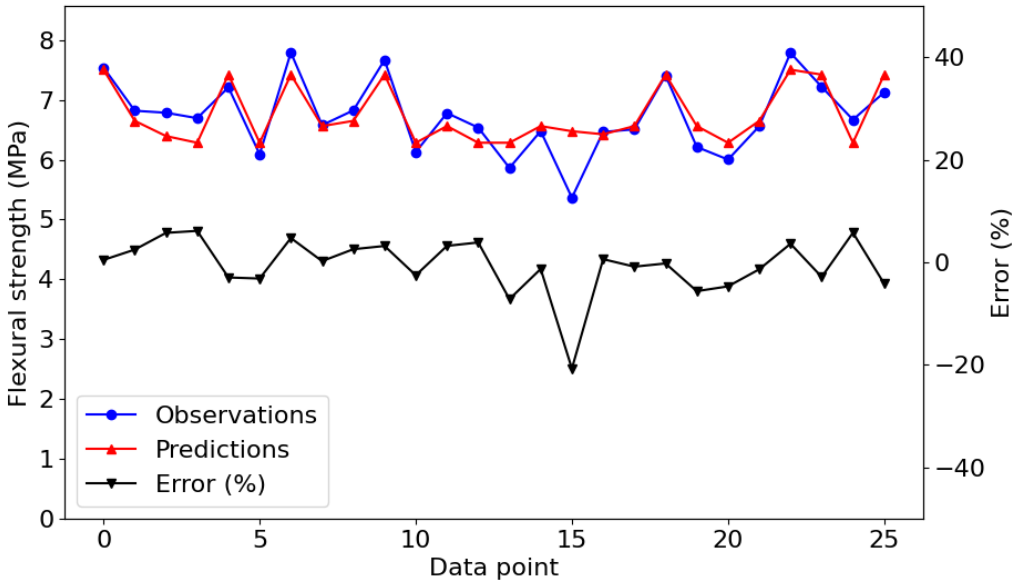

Figure 43. Predictive performance of the Random Forest model in flexural strength prediction by support vector machine model

Table 20 lists all the hyperparameter configurations, search space, and results in the support vector machine model used to predict the compressive strength and flexural strength of alkali-activated concrete. The hyperparameters tuned include the regularization parameter C, the kernel type, gamma, degree, and epsilon. The search space for C was [0.001, 0.01, 0.1, 0.5, 1, 2]; for the kernel it was [linear, rbf]; for gamma, it was [scale, auto]; for the degree it was [2, 3, 4], and for epsilon it was [0.001, 0.01, 0.1, 0.5, 1, 2].For compressive strength prediction, the best parameters were C=2, kernel=linear, gamma=scale, degree=2, and epsilon=2. For flexural strength prediction, the best parameters were C=2, kernel=linear, gamma=scale, degree=2, and epsilon=0.5. These optimal values were found through hyperparameter tuning to enhance the model's performance in predicting both types of strength. Figure 40 represents the observations versus predictions with errors in prediction values. When the

observation value and prediction value are compared, it can be seen that the forecasts nearly match the observations, demonstrating the support vector machine model's strong performance and general agreement with the data. The error fluctuates around 0%, indicating that the model predictions are sometimes above and sometimes below the exact values. However, there are instances where the error exceeds 20%, and a few cases in which it goes below -20%, suggesting occasional significant deviations between predictions and observations. Figure 41 shows the predicted versus observed plot. In this figure, most points clustered around the ideal line and within the acceptable error range between $\pm$20% line; however, when the compressive strength of the alkali-activated sample has an observed value between 20 to 30 MPa, the predictions located beyond the acceptable error range, which demonstrate in low compressive strength range, the performance of support vector machine in compressive strength prediction are much lower than in high compressive strength range. This may be because of the noise in the data and limitations of the support vector machine model to generalize from the dataset. For flexural strength prediction, Figure 42 and Figure 43 show the observations and prediction results. In the first plot, the predicted values (red points) closely follow the observed values (blue points), with the percentage error (black line) remaining relatively low and stable, mainly within a ±20% range. The second plot, displaying the predicted versus observed flexural strength, shows that training (red squares) and testing (blue circles) data points are tightly clustered around the ideal line and within the bounding lines.

6.6 k-nearest neighbors

Table 21. Model evaluation results of compressive strength prediction

	MAE	RMSE	MSE	R^2	MAPE
Training set	2.105	2.951	8.707	0.957	5.547
Testing set	2.677	3.384	11.454	0.946	7.248

Table 22. Model evaluation results of flexural strength prediction

	MAE	RMSE	MSE	R^2	MAPE
Training set	0.248	0.327	0.107	0.702	3.854
Testing set	0.203	0.250	0.062	0.713	3.149

Table 23. Hyperparameter configuration in k-nearest neighbors model

Hyperparameter	Search space	Best parameter (compressive strength)	Best parameter (flexural strength)
N_neighbors	[3, 5, 7, 9]	3	7
Weight	[uniform, distance]	uniform	uniform
Algorithm	[auto, ball_tree, kd_tree, brute]	auto	auto
Leaf_size	[20, 30, 40, 50]	20	20
P	[1, 2]	1	1

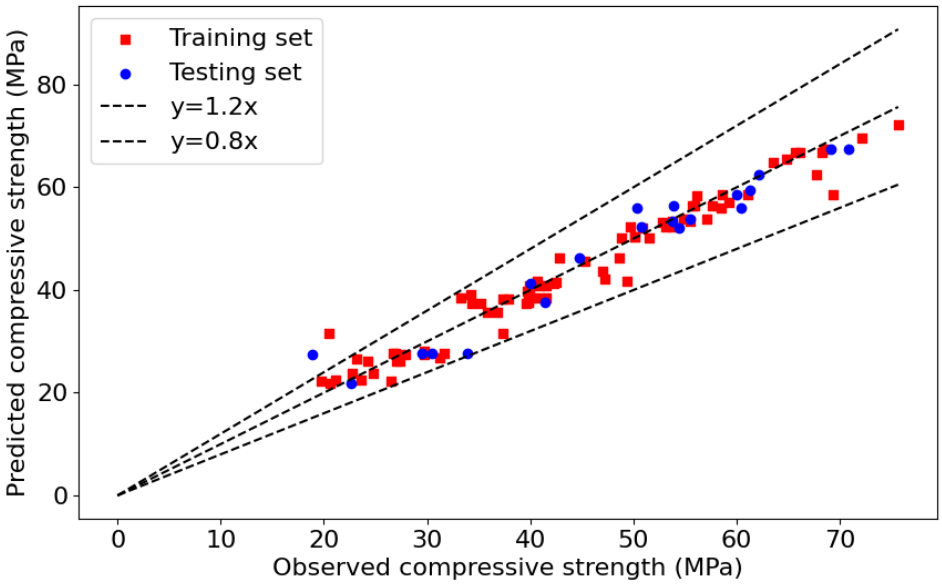

Figure 44. Regression plot of predicted versus observed values in compressive strength prediction by k-nearest neighbors model

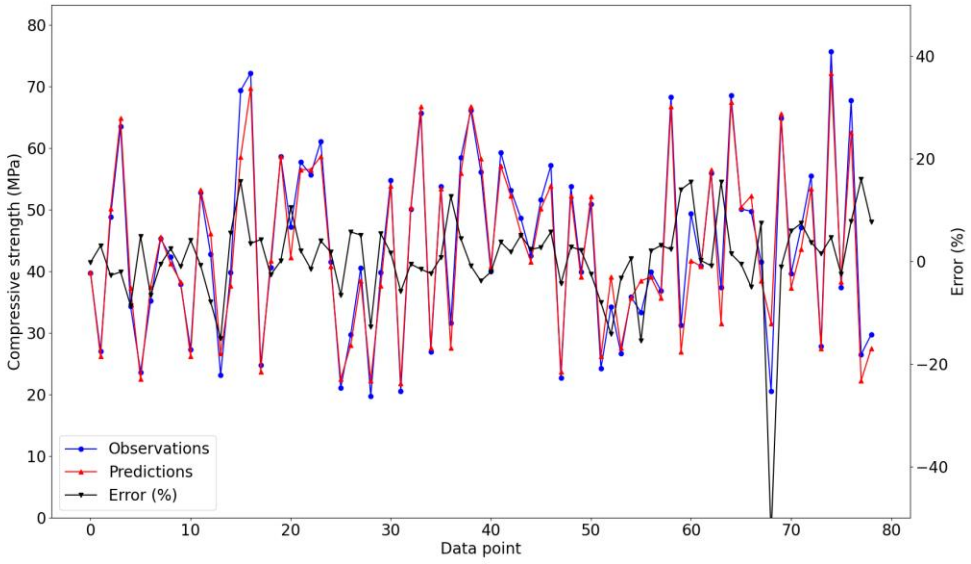

Figure 45. Predictive performance of the Random Forest model in flexural strength prediction by k-nearest neighbors model

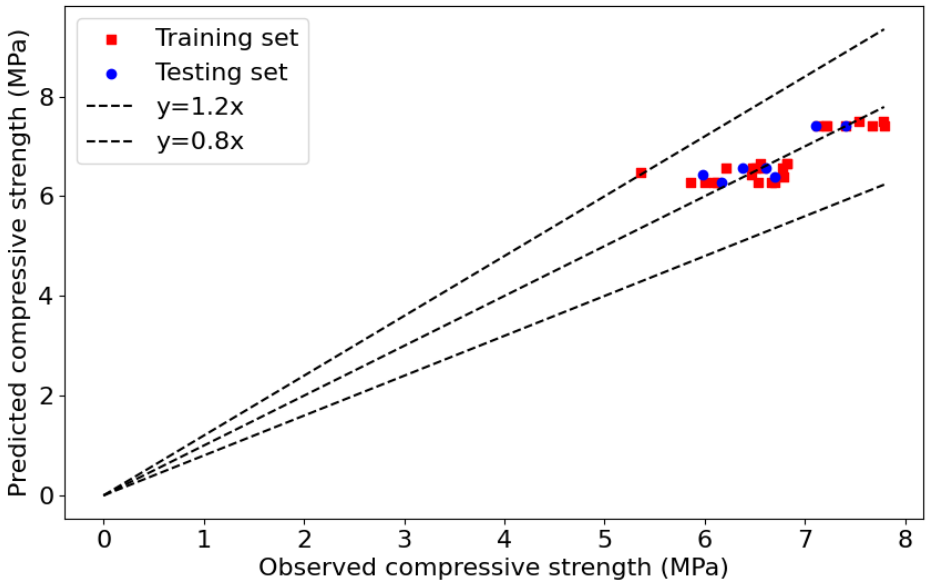

Figure 46. Regression plot of predicted versus observed values in compressive strength prediction by k-nearest neighbors model

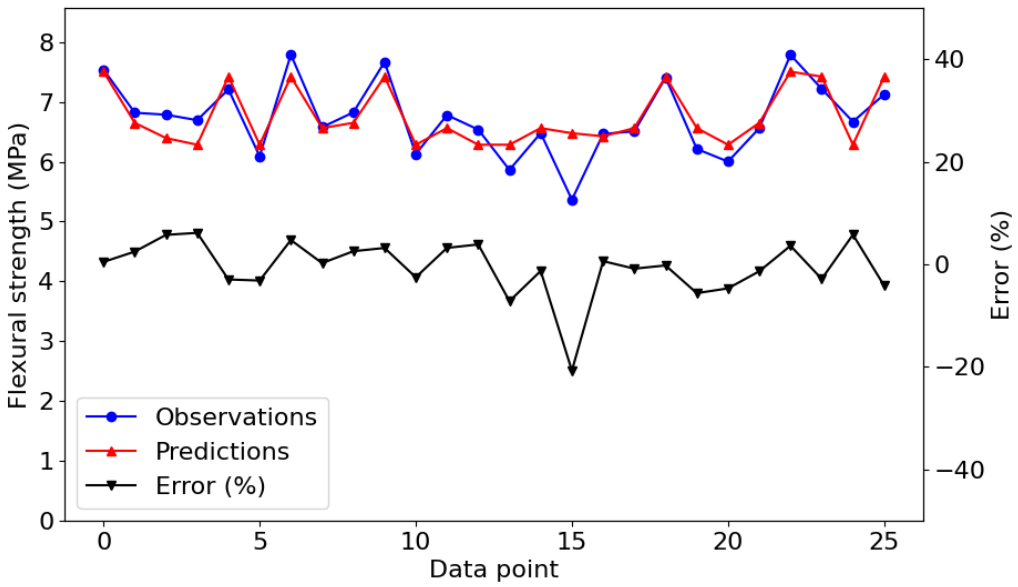

Figure 47. Predictive performance of the Random Forest model in flexural strength prediction by k-nearest neighbors model

The evaluation of the K-Nearest Neighbors model for predicting compressive and flexural strengths is demonstrated by the comprehensive metrics provided in Table 21 and Table 22. For compressive strength prediction, the R^2 of training and testing values of 0.957 and 0.946, respectively, indicating high accuracy of the K-Nearest Neighbors model in compressive prediction of alkali-activated concrete with packing density as a parameter in the dataset. The training set metrics MAE of 2.105, RMSE of 2.951, MSE of 8.707, and MAPE of 5.547, alongside the testing set metrics MAE of 2.677, RMSE of 3.384, MSE of 11.454, and MAPE of 7.248, these model performance metrics also provide the same results with R^2. Conversely, the flexural strength prediction model achieves moderate values of R^2, the values are 0.702 for training and 0.713 for testing, not only R^2, but also other metrics of performance evaluation, MAE (0.248 for training, 0.203 for testing), RMSE (0.327 for training, 0.250 for testing), MSE (0.107 for training, 0.062 for testing), and MAPE (3.854 for training, 3.149 for testing). The optimal hyperparameters for compressive strength prediction model include N_neighbors: 3, Weight: uniform, Algorithm: auto, Leaf size: 20 and P:1, which the Grid Search space are [3, 5, 7, 9], [uniform, distance], [auto, ball_tree, kd_tree, brute], [20, 30, 40, 50] and [1, 2], respectively. In the hyperparameters tuning in flexural strength prediction takes the same hyperparameters search space with compressive strength prediction, however, the best results are N_neighbors: 7, Weight: uniform, Algorithm: auto, Leaf size: 20 and P:1. However, in flexural strength prediction after compared with model performance metrics, metrics have greater value in training set than testing set, which is not normal in typical scenarios. This may be explained by the size of the dataset in flexural strength prediction; random variability can cause the test set to appear easier to predict than the training set, especially in a small dataset; this situation is less common with large datasets. Similar observations occur in predicting dynamic yield stress and static yield stress, as given by Sun et al. (Sun

et al., 2023b). Also, in compressive strength prediction by using the support vector machine model, Peng and Unluer (2022) found that the performance evaluation metrics showed higher accuracy in the testing set than in the training set. After considering this case, more experiment studies need to be done to advance the dataset of flexural strength prediction. Figure 45 shows the observed verse predicted compressive strength values with a parallel plot of the prediction error percentages. The strong correspondence between the observed (blue) and observed (red) values indicates that the KNN model accurately captures the underlying pattern in the data. The error rates (black line) are mostly within a range between 20% higher or less than observed values, which suggests that the model has accurate predictions of the compressive strength of alkali-activated concrete. Figure 44 is a scatter plot comparing predicted compressive strength to observed values for training and testing sets. The data points cluster around the ideal y = x line, the dashed lines at y = 1.2x and y = 0.8x serve as acceptable deviation boundaries, and most of the data points are located within this boundary, further validating the model's predictions. Figure 46 illustrates the observed versus predicted flexural strength values alongside the prediction error percentages in flexural strength predictions. Similar to the compressive strength plot, the observed and predicted values are closely aligned, indicating accurate predictions by the k-nearest neighbors model. The error percentages are generally low, demonstrating consistent performance. Figure 47 shows the relationship between predicted and observed flexural strength values. The data points are closely grouped around the y = x line, indicating predictive solid performance. Figure 45, a scatter plot for flexural strength predictions akin to the compressive strength plot, includes dashed boundary lines (y = 1.2x and y = 0.8x), within which most data points reside. Still, compared with compressive strength predictions, the accuracy does not show well. These visualizations and

quantitative metrics demonstrate that the k-nearest neighbors model performs robustly in predicting the compressive strength of alkali-activated concrete with packing density as a critical parameter in the dataset. After comparison with the k-nearest neighbors model from Karademupa and Pancharathi, Khan et al., the k-nearest neighbors model in this study shows similar or even better performance than previous studies (Karadumpa and Pancharathi, 2021) (Khan et al., 2024). Nonetheless, the flexural strength prediction does not achieve the same accuracy, and further experiments can enhance the dataset and ensure the model's reliability in flexural strength prediction.

6.7 Model comparison

6.7.1 Compressive strength prediction

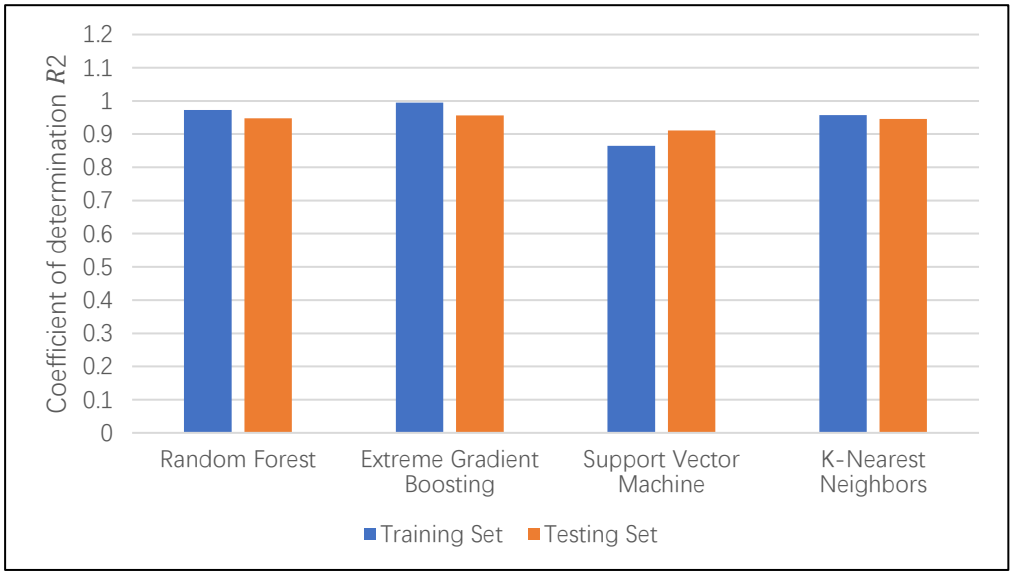

Figure 48. Coefficient of determination R^2 of four machine learning models in compressive strength prediction

As shown in Figure 48, the performance of four machine learning models was evaluated in predicting the compressive strength of alkali-activated concrete

using the coefficient of determination R^2 for both training and testing sets. Overall, the Extreme Gradient Boosting model showed the highest predictive accuracy, strong performance, and minimal difference between the training and testing sets, indicating excellent generalization capability. The Random Forest and K-Nearest Neighbors models also demonstrated high accuracy, with slightly lower performance than Extreme Gradient Boosting but still maintaining strong predictive power. The Support Vector Machine model showed comparatively lower accuracy, with a noticeable difference between the training and testing sets. It may not capture the complex relationships in the data as effectively as the ensemble methods. The results indicate that ensemble methods, particularly Extreme Gradient Boosting, provide superior predictive accuracy for the compressive strength of alkali-activated concrete. These models' ability to generalize well to new data. With its lower accuracy and greater discrepancy between training and testing performance, the Support Vector Machine model appears less adept at capturing nonlinear relationships within the data. These findings highlight the efficacy of ensemble learning methods in this application, suggesting that further exploration of hyperparameter optimization and additional features could enhance predictive accuracy. And the high performance of machine learning models proved that the packing density can be an input variable in alkali-activated concrete compressive strength prediction. Previous studies have already found the possibility of packing density used in the dataset in traditional Portland cement concrete compressive strength prediction (Pallapothu et al., 2023). The four machine learning models achieve equivalent or even superior performance in compressive strength in alkali-activated concrete compared with existing machine learning models (Gomaa et al., 2021) (Sun et al., 2023b) (Sun et al., 2023a) (Peng and Unluer, 2022) (Khan et al., 2024) (Zhang et al., 2022) (Rajakarunakaran et al., 2022). Future research might explore the dataset to collect more data points to

enhance the packing density contribution in predicting alkali-activated concrete compressive strength, to further improve model performance, and to contribute valuable insights to civil engineering practices in alkali-activated concrete properties.

6.7.2 Flexural Strength Prediction

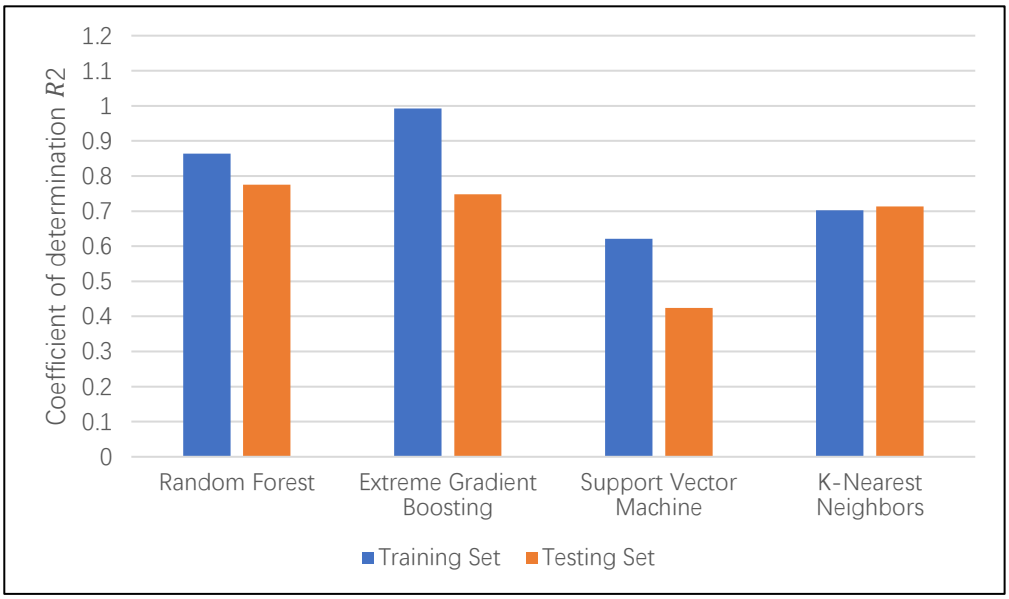

Figure 49. Coefficient of determination R^2 of four machine learning models in flexural strength prediction

Similar to the performance in compressive strength prediction, Figure 49 shows the performance of the same four machine learning models assessed for predicting the flexural strength of alkali-activated concrete. Extreme Gradient Boosting achieved the highest accuracy again, with consistent performance across training and testing sets. Random Forest and K-Nearest Neighbors also performed well, although there was a slightly larger gap between training and testing set performance compared to Extreme Gradient Boosting. The Support Vector Machine model showed the least accuracy, with significant discrepancies between the training and testing sets, indicating potential overfitting and less robustness in capturing the flexural strength characteristics.

The results indicate that while ensemble methods such as Extreme Gradient Boosting and Random Forest provide strong predictive accuracy for compressive strength, the prediction of flexural strength proves to be more challenging. The models exhibit a noticeable decrease in accuracy for flexural strength, particularly with the Support Vector Machine showing significant discrepancies between training and testing sets. This suggests potential overfitting and a lack of robustness in capturing the complex behavior of flexural strength in alkali-activated concrete.

To enhance the performance of flexural strength prediction models, expanding the dataset is a critical step. Collecting additional experimental data would provide the models with more diverse patterns to learn from, thereby improving their generalizability. Alternatively, synthetic data generation methods, such as the Synthetic Minority Oversampling Technique (SMOTE), or data augmentation strategies could be employed to artificially increase the dataset size. Incorporating data from relevant external studies or publicly available datasets, provided they align with the experimental context, may also supplement the dataset effectively and reduce overfitting risks. Implementing ensemble or advanced machine learning models can further enhance predictive accuracy. Ensemble approaches, such as stacking or bagging, which combine algorithms like Random Forest, XGBoost, and KNN, leverage the complementary strengths of individual models while mitigating their limitations. Additionally, advanced gradient boosting methods such as LightGBM or CatBoost are particularly suitable for small datasets due to their efficiency and robustness. Exploring neural network architectures, including Multi-Layer Perceptrons or 1D-Convolutional Neural Networks (1D-CNNs), can enable the capture of complex, non-linear interactions within the data. Furthermore, hybrid models

that integrate machine learning with domain-specific knowledge may offer a more comprehensive and contextually accurate prediction framework.

6.8 Effects of Different Input Parameters

Analyzing the potential influence and importance of the prediction of each input variable is helpful in alkali-activated concrete mix design. This research can identify those input variables that have a big impact on the final results and those with little impact. To make a measurement of different input variables in machine learning model results, permutation feature importance is a normal method (Almustafa and Nehdi, 2020). In the permutation feature importance method, a certain input variable X_i will shuffle individually, and other input variables remain in the previous arrangement; if this procedure results from the prediction accuracy changing significantly, it means this input variable X_i has a high importance in all machine learning input variables. Based on the core idea of the permutation feature importance method, in this study, an equation was provided below to calculate the permutation feature importance value of a specific input variable by using Mean Absolute Error as the error coefficient.

$$Permutation\ Feature\ Importance = MAE_{permute} - MAE_{orignial} \quad \text{Equation 18}$$

In this equation, $MAE_{permute}$ represents the Mean Absolute Error of the machine learning model before permuting an input variable, $MAE_{orignial}$ represents the Mean Absolute Error of the machine learning model after permuting an input variable. According to this method, the value of permutation feature importance of an input variable will describe the influence of the machine learning model. When this value is close to zero, a low impact of this variable on the output value has been proved; while this was big, the

impact of this variable on the output value was serious.

Table 24. Permutation feature importance values of different input variables in the compressive strength prediction by the Extreme Gradient Boosting model

Notation	Input variable	Permutation feature importance
X1	Modulus in activator solution	0.003
X2	Activator to Cementitious materials ratio	0.080
X3	Water to Cementitious materials ratio	0.3326
X4	Weight of Cementitious materials	0.4144
X5	Packing density	1.2811
X6	Sodium silicate to Sodium hydroxide ratio	1.2922
X7	Fly ash to GGBS ratio	2.9767
X8	Curing time	12.2166

Table 25. Permutation feature importance values of different input variables in the flexural strength prediction by the Extreme Gradient Boosting model

Notation	Input variable	Permutation feature importance
X1	Modulus in activator solution	0.2680
X2	Activator to Cementitious materials ratio	0.0094
X3	Water to Cementitious materials ratio	0.0507
X4	Weight of Cementitious materials	0.0312
X5	Packing density	0.0573
X6	Sodium silicate to Sodium hydroxide ratio	0.1243
X7	Fly ash to GGBS ratio	0.0912

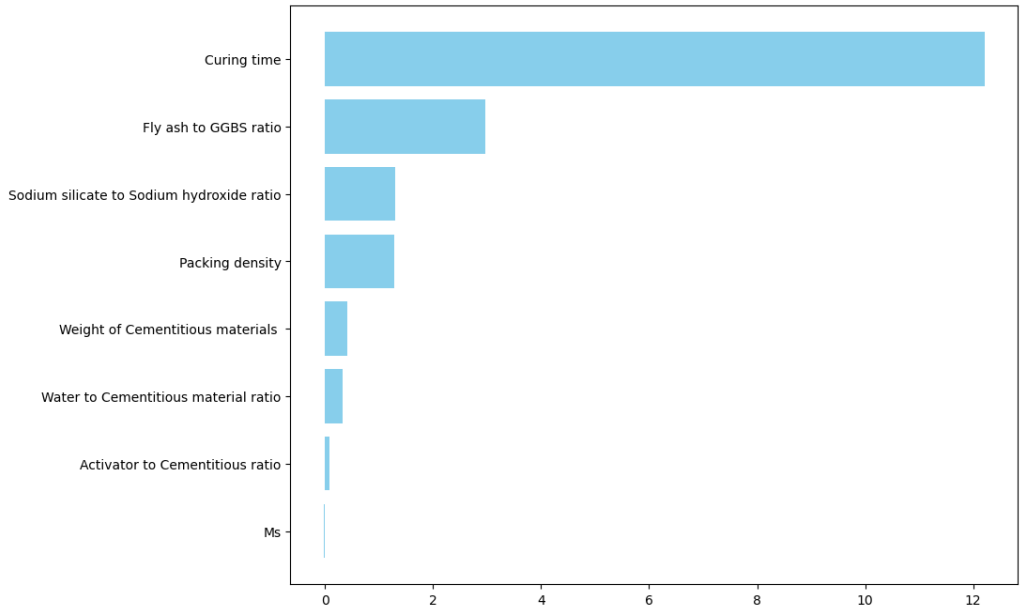

Figure 50. Importance of features in compressive strength prediction by Extreme Gradient Boosting model

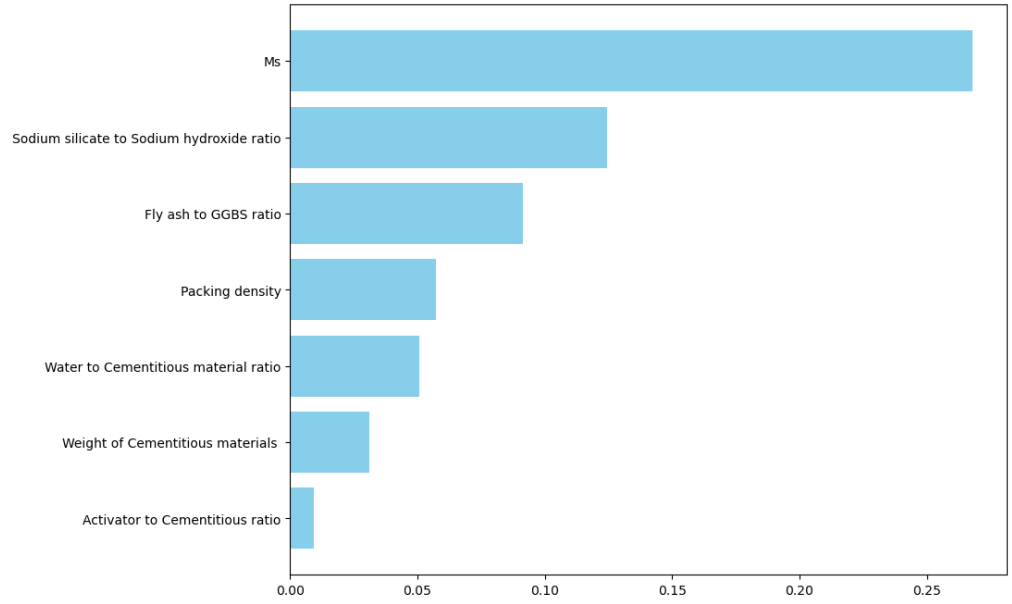

Figure 51. Importance of features in flexural strength prediction by Extreme Gradient Boosting model

Table 26. Feature importance values of different input variables in the compressive strength prediction by the Random Forest model

Notation	Input variable	Permutation feature importance

X1	Modulus in activator solution	0.055949
X2	Activator to Cementitious materials ratio	0.002982
X3	Water to Cementitious materials ratio	0.008839
X4	Weight of Cementitious materials	0.003058
X5	Packing density	0.047344
X6	Sodium silicate to Sodium hydroxide ratio	0.053000
X7	Fly ash to GGBS ratio	0.132631
X8	Curing time	0.696197

Table 27. Feature importance values of different input variables in the flexural strength prediction by the Random Forest model

Notation	Input variable	Permutation feature importance
X1	Modulus in activator solution	0.474312
X2	Activator to Cementitious materials ratio	0.000389
X3	Water to Cementitious materials ratio	0.024902
X4	Weight of Cementitious materials	0.001716
X5	Packing density	0.040041
X6	Sodium silicate to Sodium hydroxide ratio	0.381120
X7	Fly ash to GGBS ratio	0.077519

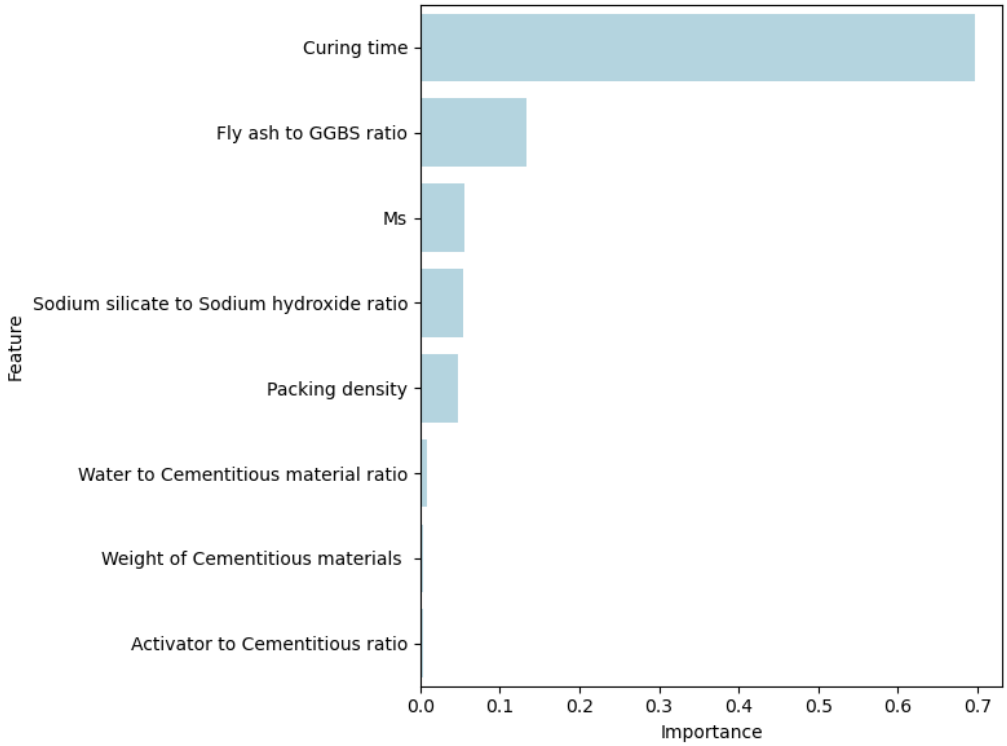

Figure 52. Importance of features in compressive strength prediction by Random Forest model

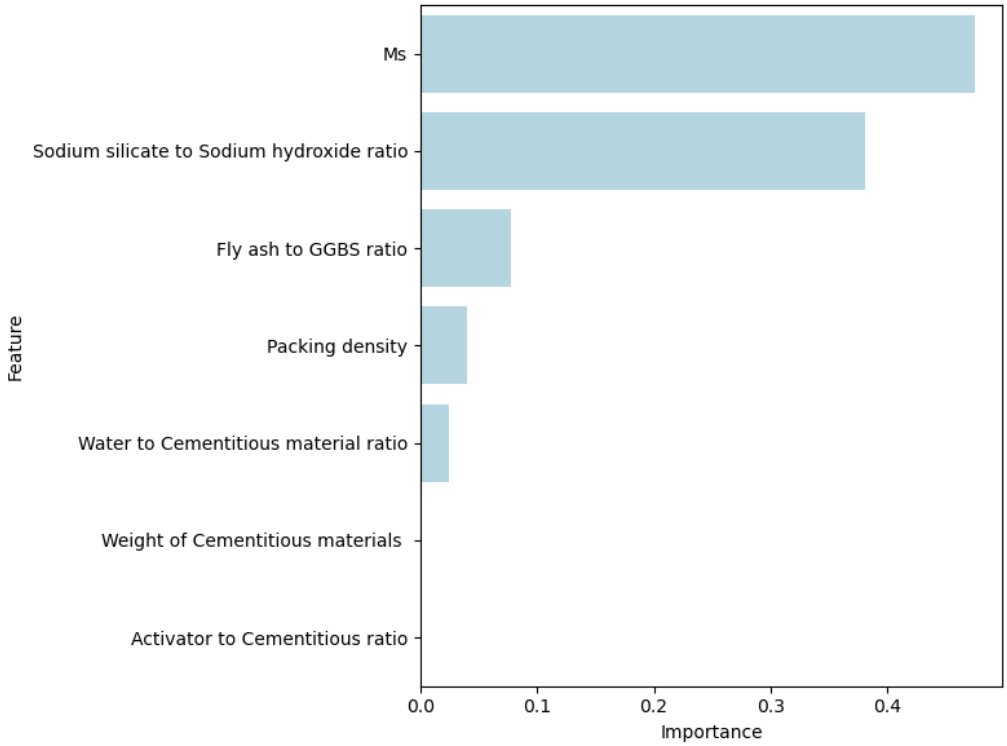

Figure 53. Importance of features in flexural strength prediction by Random Forest model

As mentioned in the previous section, in this study, the Extreme Gradient Boosting model and Random Forest model have the best performance in both compressive strength prediction and flexural strength prediction. To analyze the feature importance of each input parameter, Table 24 and Figure 50 show the results after calculating the permutation feature importance value of the Extreme Gradient Boosting model. In compressive strength prediction, curing time is the most influential parameter which shows a significant impact on compressive strength, this is the common phenomenon in concrete materials. The following factors are fly ash to GGBS ratio, sodium silicate to sodium hydroxide ratio, and packing density; each has an important permutation feature of 2.9767, 1.2922, and 1.2811, respectively. This result proved that the packing density is an important factor in alkali-activated compressive strength. Modulus in activator solution and activator to cementitious materials ratio have the smallest impact on prediction. The water to cementitious materials ratio and weight of cementitious materials has a permutation feature importance value of 0.3326 and 0.4144, which show intermediate influence in prediction. According to the Random Forest built-in feature importances analysis program, feature importances in Random Forest model were provided in Table 26 and Figure 52, for compressive strength, the Random Forest model indicates that curing time is the most significant contributor, followed by the fly ash to GGBS ratio, the sodium silicate to sodium hydroxide ratio, the modulus in the activator solution, and packing density. Meanwhile, the water-to-cementitious ratio, weight of cementitious materials, and activator-to-cementitious ratio show relatively smaller impacts. However, the permutation features important value is limited to analyzing the connection between input variables and output values from a single correspondence; it cannot show how a factor affects the final prediction outcome; this is due to the compressive strength of alkali-activated concrete affected by many different influencing factors. The literature

already proved that when the different input variables change in the alkali-activated concrete mixture, the ultimate compressive strength will be influenced, but the effects are very different. According to the permutation feature importance results, obviously, curing time, sodium silicate to sodium hydroxide ratio, and packing density all have significant effects on weight (Niyazuddin and B, 2023) (Sun et al., 2023a) (Sun et al., 2023b) (Zhang et al., 2022) (Peng and Unluer, 2022). Due to machine learning models not being able to provide a formula directly between input variables and output variables, a sensitivity analysis should be done in further investigation.

In flexural strength prediction, Table 25 and Figure 51 present the permutation feature importance of input variables in the Extreme Gradient Boosting model. The importance of input variables from high to low is modulus in activator, sodium silicate to sodium hydroxide ratio, fly ash to GGBS ratio, packing density, water to cementitious materials ratio, weight of cementitious materials and activator to cementitious materials ratio. Table 27 and Figure 53 provide the feature importance of flexural strength prediction in Random Forest model. The modulus in the activator solution dominates, followed by the sodium silicate to sodium hydroxide ratio. The fly ash to GGBS ratio, packing density, and water-to-cementitious ratio play moderate roles, whereas the weight of cementitious materials and the activator-to-cementitious ratio have minimal influence. In both models, packing density shows moderate impact to flexural strength prediction

7. Limitations and further study

This investigation is based on a scientific analysis of predicting mechanical properties of alkali-activated concrete through machine learning models. Some

limitations need to be addressed. The key restrictions are specified as follows:

Using certain input variables provided in the concrete mix design, the trained machine learning model can predict the compressive strength and flexural strength of alkali-activated concrete using four distinct machine learning algorithms. Although all machine learning models show high accuracy in compressive strength prediction, the machine learning models still need to be verified by a more comprehensive dataset. In machine learning models, aggregates' packing density influences prediction performance, so validating this observation is necessary in future studies. In future studies, more comprehensive experimental designs and more experimental data collection are needed to verify the accuracy of machine learning models in predicting compressive strength in actual alkali-activated concrete design and research. This can reduce the material and time costs required in the alkali-activated concrete design process.

Regarding using machine learning models to predict the flexural strength of alkali-activated concrete, its prediction accuracy is not high after training and testing with the existing data set. This is due to the small size of the data set. In the future, more experimental data will be needed to establish a comprehensive data set for training and testing to demonstrate the performance of machine learning methods in predicting the flexural strength of alkali-activated concrete.

8. Conclusion

In this study, machine learning regression techniques were used to make the models predict the compressive strength and flexural strength of alkali-

activated concrete, which is made of ground granulated blast-furnace slag and fly ash. The dataset for model training and testing was collected from laboratory experiments. 99 compressive strength data points and 33 flexural strength data points were included in the dataset. Packing density of aggregates, fly ash to ground granulated blast-furnace slag weight ratio, weight of cementitious materials, activator to cementitious materials ratio, sodium silicate to sodium hydroxide ratio, modulus of activator, water to cementitious materials ratio. From the experiment study and the performance of machine learning models, the following conclusions can be drawn:

1. In experiments, both the compressive strength and the flexural strength of alkali-activated concrete show high relevance with the packing density of aggregates. With the increasing aggregate packing density, the compressive strength or flexural strength increases.

2. Random Forest, Extreme Gradient Boosting, Support Vector Machine, and K-Nearest Neighbors model all have the ability to predict the compressive strength of alkali-activated concrete. In these four models, Extreme Gradient Boosting demonstrated the best performance both in the training and testing sets. The effective rank of four machine learning models in prediction from high to low is Extreme Gradient Boosting, Random Forest, K-Nearest Neighbors, and Support Vector Machine.

3. Machine learning models developed in this research achieve standard accuracy in predicting compressive strength with published research papers.

4. In flexural strength prediction, the Extreme Gradient Boosting Random Forest shows better performance than the K-Nearest Neighbors and Support Vector Machine. Overall, all predicted accuracy does not show well in flexural strength compared to compressive strength prediction. The lack of data points in the dataset is the crucial influence of flexural strength prediction; further experiment study needs to be carried out to build a comprehensive dataset for

machine learning model development.

5. Feature importance from XGBoost and RF models demonstrate that curing time is the strongest influence on compressive strength, while the modulus of the activator solution dominates flexural strength prediction. Among the other factors, the fly ash to GGBS ratio, sodium silicate to sodium hydroxide ratio, and notably, the packing density play a significant role in enhancing both compressive and flexural strength. In contrast, the water to cementitious ratio, weight of cementitious materials, and activator to cementitious materials ratio have relatively moderate or minimal impacts. However, because feature importance does not illuminate the complex interactions within these variables, further sensitivity analysis is recommended to capture the synergistic effects that ultimately govern alkali-activated concrete performance.

6. This study successfully demonstrated the application of machine learning techniques, including Random Forest, XGBoost, KNN, and SVM, to predict the compressive and flexural strength of alkali-activated concrete. The findings revealed that the packing density play a curial role in compressive strength and flexural strength prediction of alkali-activated concrete, the machine learning models perform high accuracy in strength prediction. This investigation provided valuable insights into sustainable construction material properties prediction through machine learning techniques. These outcomes are instrumental for advancing predictive models in the field of sustainable construction materials. However, the study has limitations, such as the relatively small dataset for strength predictions, which might affect the generalizability of the models. Furthermore, the dataset used in the study was limited to specific aggregate sizes and material combinations, limiting the broad applicability of the model. Despite these challenges, the strengths of this research, including its innovative approach and practical relevance, outweigh its limitations. Future studies could focus on expanding the dataset and

exploring additional variables to enhance model accuracy and applicability in industrial practice.

7. This research developed machine learning models to predict the compressive strength and flexural strength of alkali-activated concrete, these models reduce the laboratory experiment spend during research in the further study, also improve the accuracy of strength prediction in alkali-activated concrete. However, according to the limitation of laboratory experiment, the dataset used in this investigation, this may cause the machine learning models lack of generalizability. To improve this investigation, future research should focus on expanding the dataset, particularly for strength, to enhance model robustness and generalizability. Incorporating a wider range of material properties, such as diverse aggregate types and alkali activator compositions, would increase applicability across various scenarios. Advanced machine learning techniques, like neural networks and hybrid models, should be explored to improve predictive accuracy, especially for smaller datasets. Sensitivity analysis and external validation using other datasets would provide deeper insights into influential factors and model reliability. Additionally, integrating long-term durability studies, analyzing environmental and economic impacts, and applying models to broader structural and property contexts would strengthen the practical relevance and sustainability of the findings.

9. Acknowledgements

I thank my lead supervisor, Assistant Professor Weizhuo Shi, for his support in experiments, computer programming, and dissertation writing. I am grateful to him for his supervision, expert engineering research experience, and encouragement in completing this dissertation. I thank my co-supervisor, Professor Bo Li, for his professional advice on my dissertation, his valuable time, and for allowing me to use his construction materials laboratory for experiment

study. I thank all administrative staff in the Faculty of Science and Engineering, University of Nottingham Ningbo, China. I acknowledge all PhD students in Dr. Weizhuo's and Prof Bo's groups for their support in my experiment.

10. Reference

ABED, M., IMTEAZ, M., AHMED, A. N. & HUANG, Y. F. Application of k-nearest neighbors (KNN) technique for predicting monthly pan evaporation. AIP Conference Proceedings, 2023. AIP Publishing.

AFZALI, S. A. E., SHAYANFAR, M. A., GHANOONI-BAGHA, M., GOLAFSHANI, E. & NGO, T. 2024. The use of machine learning techniques to investigate the properties of metakaolin-based geopolymer concrete. *Journal of Cleaner Production*, 141305.

AiTCIN, P.-C. 2000. Cements of yesterday and today: Concrete of tomorrow. *Cement and Concrete Research,* 30, 1349-1359.

AL-SHAMIRI, A. K., KIM, J. H., YUAN, T.-F. & YOON, Y. S. 2019. Modeling the compressive strength of high-strength concrete: An extreme learning approach. *Construction and Building Materials,* 208, 204-219.

ALIABDO, A. A., ABD ELMOATY, A. E. M. & EMAM, M. A. 2019. Factors affecting the mechanical properties of alkali activated ground granulated blast furnace slag concrete. *Construction and Building Materials,* 197, 339-355.

ALMUSTAFA, M. K. & NEHDI, M. L. 2020. Machine learning model for predicting structural response of RC slabs exposed to blast loading. *Engineering Structures,* 221, 111109.

ANDREASEN, A. 1930. Ueber die Beziehung zwischen Kornabstufung und Zwischenraum in Produkten aus losen Körnern (mit einigen Experimenten). *Kolloid-Zeitschrift,* 50, 217-228.

BERNAL, S. A., DE GUTIéRREZ, R. M., PEDRAZA, A. L., PROVIS, J. L., RODRIGUEZ, E. D. & DELVASTO, S. 2011. Effect of binder content on the performance of alkali-activated slag concretes. *Cement and concrete research,* 41, 1-8.

BERNAL, S. A. & PROVIS, J. L. 2014. Durability of alkali-activated materials: progress and perspectives. *Journal of the American Ceramic Society,* 97, 997-1008.

BISHOP, C. M. & NASRABADI, N. M. 2006. *Pattern recognition and machine learning*, Springer.

BREIMAN, L. 2001. Random forests. *Machine learning,* 45, 5-32.

CAIJUN SHI, D. R., PAVEL KRIVENKO 2006. *Alkali-Activated Cements and Concretes*, CRC Press.

CHEN, T., HE, T., BENESTY, M., KHOTILOVICH, V., TANG, Y., CHO, H., CHEN, K., MITCHELL, R., CANO, I. & ZHOU, T. 2015. Xgboost: extreme gradient boosting. *R package version 0.4-2,* 1, 1-4.

CHENG, M.-Y., CHOU, J.-S., ROY, A. F. & WU, Y.-W. 2012. High-performance concrete

compressive strength prediction using time-weighted evolutionary fuzzy support vector machines inference model. *Automation in Construction,* 28**,** 106-115.

CHO, Y.-K., YOO, S.-W., JUNG, S.-H., LEE, K.-M. & KWON, S.-J. 2017. Effect of Na2O content, SiO2/Na2O molar ratio, and curing conditions on the compressive strength of FA-based geopolymer. *Construction and Building Materials,* 145**,** 253-260.

CHOU, J.-S. & PHAM, A.-D. 2013. Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength. *Construction and Building Materials,* 49**,** 554-563.

CORTES, C. & VAPNIK, V. 1995a. Support-vector network-. machine learning 20: 273–297. *Portfolio Selection, Journal of Global Optimization,* 43.

CORTES, C. & VAPNIK, V. 1995b. Support-Vector Networks. *Machine Learning,* 20**,** 273-297.

COVER, T. & HART, P. 1967. Nearest neighbor pattern classification. *IEEE transactions on information theory,* 13**,** 21-27.

DE LARRARD, F. 1999. *Concrete mixture proportioning: a scientific approach*, CRC Press.

DE LARRARD, F. & SEDRAN, T. 1994. Optimization of ultra-high-performance concrete by the use of a packing model. *Cement and concrete research,* 24**,** 997-1009.

DEWAR, J. 1999. *Computer modelling of concrete mixtures*, CRC press.

DING, Y., DAI, J.-G. & SHI, C.-J. 2016. Mechanical properties of alkali-activated concrete: A state-of-the-art review. *Construction and Building Materials,* 127**,** 68-79.

DURASTANTI, C. & MORETTI, L. 2020. Environmental Impacts of Cement Production: A Statistical Analysis. *Applied Sciences* [Online], 10.

FARHAN, N. A., SHEIKH, M. N. & HADI, M. N. S. 2019. Investigation of engineering properties of normal and high strength fly ash based geopolymer and alkali-activated slag concrete compared to ordinary Portland cement concrete. *Construction and Building Materials,* 196**,** 26-42.

FENNIS, S. & WALRAVEN, J. 2012. Using particle packing technology for sustainable concrete mixture design. *Heron,* 57**,** 73-101.

FRIEDMAN, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics***,** 1189-1232.

FU, Q., BU, M., ZHANG, Z., XU, W., YUAN, Q. & NIU, D. 2023. Hydration Characteristics and Microstructure of Alkali-Activated Slag Concrete: A Review. *Engineering,* 20**,** 162-179.

FULLER, W. B. & THOMPSON, S. E. 1907. The laws of proportioning concrete. *Transactions of the American Society of Civil Engineers,* 59**,** 67-143.

FUNK, J. E. & DINGER, D. R. 2013. *Predictive process control of crowded particulate suspensions: applied to ceramic manufacturing*, Springer Science & Business Media.

GARTNER, E. 2004. Industrially interesting approaches to "low-CO2" cements. *Cement and Concrete research,* 34**,** 1489-1498.

GOMAA, E., HAN, T., ELGAWADY, M., HUANG, J. & KUMAR, A. 2021. Machine learning to predict properties of fresh and hardened alkali-activated concrete. *Cement and*

Concrete Composites, 115**,** 103863.

GOODFELLOW, I., BENGIO, Y. & COURVILLE, A. 2016. *Deep learning*, MIT press.

HABERT, G. & OUELLET-PLAMONDON, C. 2016. Recent update on the environmental impact of geopolymers. *RILEM technical Letters,* 1**,** 17-23.

HSIEH, S.-C. 2021. Prediction of Compressive Strength of Concrete and Rock Using an Elementary Instance-Based Learning Algorithm. *Advances in Civil Engineering,* 2021**,** 6658932.

HUNG, C.-C. & CHANG, J.-J. 2013. The influence of mixture variables for the alkali-activated slag concrete on the properties of concrete. *Journal of Marine Science and Technology,* 21**,** 1.

HUO, W., ZHU, Z., SUN, H., MA, B. & YANG, L. 2022. Development of machine learning models for the prediction of the compressive strength of calcium-based geopolymers. *Journal of Cleaner Production,* 380**,** 135159.

IBRAHIM, M., MEGAT JOHARI, M. A., RAHMAN, M. K. & MASLEHUDDIN, M. 2017. Effect of alkaline activators and binder content on the properties of natural pozzolan-based alkali activated concrete. *Construction and Building Materials,* 147**,** 648-660.

IVANOV, Y. V., KARTASHOV, A. V., IVANOVA, A. I., IVANOV, V. P., MARCHENKO, S. I., NARTOV, D. I. & KUZNETSOV, V. V. 2018. Long-term impact of cement plant emissions on the elemental composition of both soils and pine stands and on the formation of Scots pine seeds. *Environmental Pollution,* 243**,** 1383-1393.

JAKKULA, V. 2006. Tutorial on support vector machine (svm). *School of EECS, Washington State University,* 37**,** 3.

JOSEPH, B. & MATHEW, G. 2012. Influence of aggregate content on the behavior of fly ash based geopolymer concrete. *Scientia Iranica,* 19**,** 1188-1194.

KARADUMPA, C. S. & PANCHARATHI, R. K. 2021. Developing a novel mix design methodology for slow hardening composite cement concretes through packing density approach. *Construction and Building Materials,* 303**,** 124391.

KHAN, A. Q., NAVEED, M. H., RASHEED, M. D. & MIAO, P. 2024. Prediction of Compressive Strength of Fly Ash-Based Geopolymer Concrete Using Supervised Machine Learning Methods. *Arabian Journal for Science and Engineering,* 49**,** 4889-4904.

KHAN, M. A., ZAFAR, A., FAROOQ, F., JAVED, M. F., ALYOUSEF, R., ALABDULJABBAR, H. & KHAN, M. I. 2021. Geopolymer concrete compressive strength via artificial neural network, adaptive neuro fuzzy interface system, and gene expression programming with K-fold cross validation. *Frontiers in Materials,* 8**,** 621163.

KHAYAT, K., HU, C. & LAYE, J. Influence of aggregate grain-size distribution on workability of self-consolidating concrete. International Symposium on High Performance Concrete-Workability, Strength and Durability, 2000. 641-658.

KONG, Y. & KURUMISAWA, K. 2023. Application of machine learning in predicting workability for alkali-activated materials. *Case Studies in Construction Materials,* 18**,** e02173.

KRAMER, O. 2013. K-Nearest Neighbors. *In:* KRAMER, O. (ed.) *Dimensionality Reduction with Unsupervised Nearest Neighbors*. Berlin, Heidelberg: Springer Berlin Heidelberg.

KUMAR, S., KUMAR, R. & MEHROTRA, S. 2010. Influence of granulated blast furnace slag on the reaction, structure and properties of fly ash based geopolymer. *Journal of materials science,* 45**,** 607-615.

KWAN, A. & MORA, C. 2001. Effects of various shape parameters on packing of aggregate particles. *Magazine of concrete Research,* 53**,** 91-100.

LECUN, Y., BENGIO, Y. & HINTON, G. 2015. Deep learning. *Nature,* 521**,** 436-444.

LI, L., LIN, C., CHEN, G., KWAN, A. & JIANG, T. 2017. Effects of packing on compressive behaviour of recycled aggregate concrete. *Construction and Building Materials,* 157**,** 757-777.

LI, Q.-F. & SONG, Z.-M. 2022. High-performance concrete strength prediction based on ensemble learning. *Construction and Building Materials,* 324**,** 126694.

LI, Y., SHEN, J., LIN, H. & LI, Y. 2023. Optimization design for alkali-activated slag-fly ash geopolymer concrete based on artificial intelligence considering compressive strength, cost, and carbon emission. *Journal of Building Engineering,* 75**,** 106929.

LIU, G. & SUN, B. 2023. Concrete compressive strength prediction using an explainable boosting machine model. *Case Studies in Construction Materials,* 18**,** e01845.

MEHTA, P. K. & MONTEIRO, P. 2006. Concrete: microstructure, properties, and materials. *(No Title)*.

MUCHERINO, A., PAPAJORGJI, P. J. & PARDALOS, P. M. 2009. k-Nearest Neighbor Classification. *In:* MUCHERINO, A., PAPAJORGJI, P. J. & PARDALOS, P. M. (eds.) *Data Mining in Agriculture.* New York, NY: Springer New York.

NANTHAGOPALAN, P., HAIST, M., SANTHANAM, M. & MüLLER, H. S. 2008. Investigation on the influence of granular packing on the flow properties of cementitious suspensions. *Cement and Concrete Composites,* 30**,** 763-768.

NANTHAGOPALAN, P. & SANTHANAM, M. 2012. An empirical approach for the optimisation of aggregate combinations for self-compacting concrete. *Materials and Structures,* 45**,** 1167-1179.

NATH, P. & SARKER, P. K. 2014. Effect of GGBFS on setting, workability and early strength properties of fly ash geopolymer concrete cured in ambient condition. *Construction and Building Materials,* 66**,** 163-171.

NEVILLE, A. M. 1995. *Properties of concrete*, Longman London.

NIYAZUDDIN & B, U. 2023. Mechanical and durability properties of standard and high strength geopolymer concrete using particle packing theory. *Construction and Building Materials,* 400**,** 132722.

OMER, A. 2014. Energy Efficiency Improvement Utilising High Technology: An Assessment of Energy Use in Industry, Buildings Development and Environment. *Buildings Development and Environment (March 1, 2014)*.

PALLAPOTHU, S. N. R. G., PANCHARATHI, R. K. & JANIB, R. 2023. Predicting concrete strength through packing density using machine learning models. *Engineering Applications of Artificial Intelligence,* 126**,** 107177.

PENG, Y. & UNLUER, C. 2022. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. *Construction and Building Materials,* 316**,** 125785.

POWERS, T. 1968. Properties of Fresh Concrete, John Wiley and Sons. *Inc., New York*, 301.

PROVIS, J. L. & VAN DEVENTER, J. S. 2013. *Alkali activated materials: state-of-the-art report, RILEM TC 224-AAM*, Springer Science & Business Media.

RAFEET, A., VINAI, R., SOUTSOS, M. & SHA, W. 2017. Guidelines for mix proportioning of fly ash/GGBS based alkali activated concretes. *Construction and Building Materials,* 147, 130-142.

RAJAKARUNAKARAN, S. A., LOURDU, A. R., MUTHUSAMY, S., PANCHAL, H., JAWAD ALRUBAIE, A., MUSA JABER, M., ALI, M. H., TLILI, I., MASELENO, A., MAJDI, A. & ALI, S. H. M. 2022. Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques. *Advances in Engineering Software,* 173, 103267.

RASHAD, A. M. 2013. Properties of alkali-activated fly ash concrete blended with slag. *Iran. J. Mater. Sci. Eng,* 10, 57-64.

SHEN, J., LI, Y., LIN, H., LI, H., LV, J., FENG, S. & CI, J. 2022. Prediction of compressive strength of alkali-activated construction demolition waste geopolymers using ensemble machine learning. *Construction and Building Materials,* 360, 129600.

SMOLA, A. J. & SCHöLKOPF, B. 2004. A tutorial on support vector regression. *Statistics and computing,* 14, 199-222.

SUN, B., DING, L., YE, G. & DE SCHUTTER, G. 2023a. Mechanical properties prediction of blast furnace slag and fly ash-based alkali-activated concrete by machine learning methods. *Construction and Building Materials,* 409, 133933.

SUN, Y., CHENG, H., ZHANG, S., MOHAN, M. K., YE, G. & DE SCHUTTER, G. 2023b. Prediction & optimization of alkali-activated concrete based on the random forest machine learning algorithm. *Construction and Building Materials,* 385, 131519.

TAVARES, C., WANG, X., SAHA, S. & GRASLEY, Z. 2022. Machine learning-based mix design tools to minimize carbon footprint and cost of UHPC. Part 1: Efficient data collection and modeling. *Cleaner Materials,* 4, 100082.

TAYLOR, H. F. 1997. *Cement chemistry*, Thomas Telford London.

THOMAS, R. J. & PEETHAMPARAN, S. 2017. Stepwise regression modeling for compressive strength of alkali-activated concrete. *Construction and Building Materials,* 141, 315-324.

TOUFAR, W., BORN, M. & KLOSE, E. 1976. Contribution of optimisation of components of different density in polydispersed particles systems. *Freiberger booklet A,* 558, 29-44.

TOUFIGH, V. & JAFARI, A. 2021. Developing a comprehensive prediction model for compressive strength of fly ash-based geopolymer concrete (FAGC). *Construction and Building Materials,* 277, 122241.

TURNER, L. K. & COLLINS, F. G. 2013. Carbon dioxide equivalent (CO2-e) emissions: A comparison between geopolymer and OPC cement concrete. *Construction and Building Materials,* 43, 125-130.

WEIL, M., DOMBROWSKI, K. & BUCHWALD, A. 2009. Life-cycle analysis of geopolymers. *Geopolymers*. Elsevier.

YIP, C. K.-B. 2004. *The role of calcium in geopolymerisation.* University of Melbourne,

Department of Chemical and Biomolecular Engineering.

ZHANG, L. V., MARANI, A. & NEHDI, M. L. 2022. Chemistry-informed machine learning prediction of compressive strength for alkali-activated materials. *Construction and Building Materials,* 316**,** 126103.

ZHANG, W., WU, C., ZHONG, H., LI, Y. & WANG, L. 2021. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers,* 12**,** 469-477.

ZHANG, W., WU, C., ZHONG, H., LI, Y. & WANG, L. 2021. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers,* 12**,** 469-477.

WANG, A., ZHENG, Y., ZHANG, Z., LIU, K., LI, Y., SHI, L. & SUN, D. 2020. The durability of alkali-activated materials in comparison with ordinary Portland cements and concretes: a review. *Engineering,* 6**,** 695-706.

ABDI, H. & WILLIAMS, L. J. 2010. Principal component analysis. *WIREs Computational Statistics,* 2**,** 433-459.

ABED, M., IMTEAZ, M., AHMED, A. N. & HUANG, Y. F. Application of k-nearest neighbors (KNN) technique for predicting monthly pan evaporation. AIP Conference Proceedings, 2023. AIP Publishing.

AFZALI, S. A. E., SHAYANFAR, M. A., GHANOONI-BAGHA, M., GOLAFSHANI, E. & NGO, T. 2024. The use of machine learning techniques to investigate the properties of metakaolin-based geopolymer concrete. *Journal of Cleaner Production***,** 141305.

AiTCIN, P.-C. 2000. Cements of yesterday and today: Concrete of tomorrow. *Cement and Concrete Research,* 30**,** 1349-1359.

AL-SHAMIRI, A. K., KIM, J. H., YUAN, T.-F. & YOON, Y. S. 2019. Modeling the compressive strength of high-strength concrete: An extreme learning approach. *Construction and Building Materials,* 208**,** 204-219.

ALIABDO, A. A., ABD ELMOATY, A. E. M. & EMAM, M. A. 2019. Factors affecting the mechanical properties of alkali activated ground granulated blast furnace slag concrete. *Construction and Building Materials,* 197**,** 339-355.

ALMUSTAFA, M. K. & NEHDI, M. L. 2020. Machine learning model for predicting structural response of RC slabs exposed to blast loading. *Engineering Structures,* 221**,** 111109.

ALTMAN, N. S. 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician,* 46**,** 175-185.

ANDREASEN, A. 1930. Ueber die Beziehung zwischen Kornabstufung und Zwischenraum in Produkten aus losen Körnern (mit einigen Experimenten). *Kolloid-Zeitschrift,* 50**,** 217-228.

ANSARI, S., ANSARI, H., KHATEEB, A. & IBRAHIM, S. M. 2024. Comparative study of machine learning models for predicting the compressive strength of concrete using Non-Destructive Testing methods. *Materials Today: Proceedings*.

BERNAL, S. A., DE GUTIéRREZ, R. M., PEDRAZA, A. L., PROVIS, J. L., RODRIGUEZ, E. D. &

DELVASTO, S. 2011. Effect of binder content on the performance of alkali-activated slag concretes. *Cement and concrete research,* 41**,** 1-8.

BERNAL, S. A. & PROVIS, J. L. 2014a. Durability of Alkali-Activated Materials: Progress and Perspectives. *Journal of the American Ceramic Society,* 97**,** 997-1008.

BERNAL, S. A. & PROVIS, J. L. 2014b. Durability of alkali-activated materials: progress and perspectives. *Journal of the American Ceramic Society,* 97**,** 997-1008.

BIAU, G. 2010. Analysis of a Random Forests Model. *Journal of Machine Learning Research,* 13.

BISHOP, C. M. & NASRABADI, N. M. 2006. *Pattern recognition and machine learning*, Springer.

BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.

CAIJUN SHI, D. R., PAVEL KRIVENKO 2006. *Alkali-Activated Cements and Concretes*, CRC Press.

CHAN, J. Y.-L., LEOW, S. M. H., BEA, K. T., CHENG, W. K., PHOONG, S. W., HONG, Z.-W. & CHEN, Y.-L. 2022. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics,* 10**,** 1283.

CHEN, T. & GUESTRIN, C. 2016. *XGBoost: A Scalable Tree Boosting System*.

CHEN, T., HE, T., BENESTY, M., KHOTILOVICH, V., TANG, Y., CHO, H., CHEN, K., MITCHELL, R., CANO, I. & ZHOU, T. 2015. Xgboost: extreme gradient boosting. *R package version 0.4-2,* 1**,** 1-4.

CHENG, M.-Y., CHOU, J.-S., ROY, A. F. & WU, Y.-W. 2012. High-performance concrete compressive strength prediction using time-weighted evolutionary fuzzy support vector machines inference model. *Automation in Construction,* 28**,** 106-115.

CHO, Y.-K., YOO, S.-W., JUNG, S.-H., LEE, K.-M. & KWON, S.-J. 2017. Effect of Na2O content, SiO2/Na2O molar ratio, and curing conditions on the compressive strength of FA-based geopolymer. *Construction and Building Materials,* 145**,** 253-260.

CHOU, J.-S. & PHAM, A.-D. 2013. Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength. *Construction and Building Materials,* 49**,** 554-563.

CHOU, J.-S., TSAI, C.-F., PHAM, A.-D. & LU, Y.-H. 2014. Machine learning in concrete strength simulations: Multi-nation data analytics. *Construction and Building Materials,* 73**,** 771-780.

CORTES, C. & VAPNIK, V. 1995a. Support-vector network-. machine learning 20: 273–297. *Portfolio Selection, Journal of Global Optimization,* 43.

CORTES, C. & VAPNIK, V. 1995b. Support-Vector Networks. *Machine Learning,* 20**,** 273-297.

COVER, T. & HART, P. 1967. Nearest neighbor pattern classification. *IEEE transactions on information theory,* 13**,** 21-27.

DAVIDOVITS, J. 2008. *Geopolymer Chemistry and Applications*.

DE LARRARD, F. 1999. *Concrete mixture proportioning: a scientific approach*, CRC Press.

DE LARRARD, F. & SEDRAN, T. 1994. Optimization of ultra-high-performance concrete by the use of a packing model. *Cement and concrete research,* 24**,** 997-1009.

DEDIEU, A. & MAZUMDER, R. 2019. *Solving large-scale L1-regularized SVMs and cousins: the surprising effectiveness of column and constraint generation*.

DEWAR, J. 1999. *Computer modelling of concrete mixtures*, CRC press.

DING, Y., DAI, J.-G. & SHI, C.-J. 2016. Mechanical properties of alkali-activated concrete: A state-of-the-art review. *Construction and Building Materials,* 127**,** 68-79.

DURASTANTI, C. & MORETTI, L. 2020. Environmental Impacts of Cement Production: A Statistical Analysis. *Applied Sciences* [Online], 10.

FARHAN, N. A., SHEIKH, M. N. & HADI, M. N. S. 2019. Investigation of engineering properties of normal and high strength fly ash based geopolymer and alkali-activated slag concrete compared to ordinary Portland cement concrete. *Construction and Building Materials,* 196**,** 26-42.

FENG, D.-C., LIU, Z.-T., WANG, X.-D., CHEN, Y., CHANG, J.-Q., WEI, D.-F. & JIANG, Z. 2020. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials,* 230**,** 117000.

FENNIS, S. & WALRAVEN, J. 2012. Using particle packing technology for sustainable concrete mixture design. *Heron,* 57**,** 73-101.

FRIEDMAN, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics***,** 1189-1232.

FU, Q., BU, M., ZHANG, Z., XU, W., YUAN, Q. & NIU, D. 2023. Hydration Characteristics and Microstructure of Alkali-Activated Slag Concrete: A Review. *Engineering,* 20**,** 162-179.

FULLER, W. B. & THOMPSON, S. E. 1907. The laws of proportioning concrete. *Transactions of the American Society of Civil Engineers,* 59**,** 67-143.

FUNK, J. E. & DINGER, D. R. 2013. *Predictive process control of crowded particulate suspensions: applied to ceramic manufacturing*, Springer Science & Business Media.

GARTNER, E. 2004. Industrially interesting approaches to "low-CO2" cements. *Cement and Concrete research,* 34**,** 1489-1498.

GOMAA, E., HAN, T., ELGAWADY, M., HUANG, J. & KUMAR, A. 2021. Machine learning to predict properties of fresh and hardened alkali-activated concrete. *Cement and Concrete Composites,* 115**,** 103863.

GOODFELLOW, I., BENGIO, Y. & COURVILLE, A. 2016. *Deep learning*, MIT press.

HABERT, G., D'ESPINOSE DE LACAILLERIE, J. B. & ROUSSEL, N. 2011. An environmental evaluation of geopolymer based concrete production: reviewing current research trends. *Journal of Cleaner Production,* 19**,** 1229-1238.

HABERT, G. & OUELLET-PLAMONDON, C. 2016. Recent update on the environmental impact of geopolymers. *RILEM technical Letters,* 1**,** 17-23.

HSIEH, S.-C. 2021. Prediction of Compressive Strength of Concrete and Rock Using an Elementary Instance-Based Learning Algorithm. *Advances in Civil Engineering,* 2021**,** 6658932.

HUNG, C.-C. & CHANG, J.-J. 2013. The influence of mixture variables for the alkali-activated slag concrete on the properties of concrete. *Journal of Marine Science and Technology,* 21**,** 1.

HUO, W., ZHU, Z., SUN, H., MA, B. & YANG, L. 2022. Development of machine learning models for the prediction of the compressive strength of calcium-based geopolymers. *Journal of Cleaner Production,* 380**,** 135159.

IBRAHIM, M., MEGAT JOHARI, M. A., RAHMAN, M. K. & MASLEHUDDIN, M. 2017. Effect of alkaline activators and binder content on the properties of natural pozzolan-based alkali activated concrete. *Construction and Building Materials,* 147**,** 648-660.

IVANOV, Y. V., KARTASHOV, A. V., IVANOVA, A. I., IVANOV, V. P., MARCHENKO, S. I., NARTOV, D. I. & KUZNETSOV, V. V. 2018. Long-term impact of cement plant emissions on the elemental composition of both soils and pine stands and on the formation of Scots pine seeds. *Environmental Pollution,* 243**,** 1383-1393.

JAKKULA, V. 2006. Tutorial on support vector machine (svm). *School of EECS, Washington State University,* 37**,** 3.

JOLLIFFE, I. T. & CADIMA, J. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 374**,** 20150202.

JOSEPH, B. & MATHEW, G. 2012. Influence of aggregate content on the behavior of fly ash based geopolymer concrete. *Scientia Iranica,* 19**,** 1188-1194.

KARADUMPA, C. S. & PANCHARATHI, R. K. 2021. Developing a novel mix design methodology for slow hardening composite cement concretes through packing density approach. *Construction and Building Materials,* 303**,** 124391.

KHAN, A. Q., NAVEED, M. H., RASHEED, M. D. & MIAO, P. 2024. Prediction of Compressive Strength of Fly Ash-Based Geopolymer Concrete Using Supervised Machine Learning Methods. *Arabian Journal for Science and Engineering,* 49**,** 4889-4904.

KHAN, M. A., ZAFAR, A., FAROOQ, F., JAVED, M. F., ALYOUSEF, R., ALABDULJABBAR, H. & KHAN, M. I. 2021. Geopolymer concrete compressive strength via artificial neural network, adaptive neuro fuzzy interface system, and gene expression programming with K-fold cross validation. *Frontiers in Materials,* 8**,** 621163.

KHAYAT, K., HU, C. & LAYE, J. Influence of aggregate grain-size distribution on workability of self-consolidating concrete. International Symposium on High Performance Concrete-Workability, Strength and Durability, 2000. 641-658.

KONG, Y. & KURUMISAWA, K. 2023. Application of machine learning in predicting workability for alkali-activated materials. *Case Studies in Construction Materials,* 18**,** e02173.

KRAMER, O. 2013. K-Nearest Neighbors. *In:* KRAMER, O. (ed.) *Dimensionality Reduction with Unsupervised Nearest Neighbors*. Berlin, Heidelberg: Springer Berlin Heidelberg.

KUMAR, S., KUMAR, R. & MEHROTRA, S. 2010. Influence of granulated blast furnace slag on the reaction, structure and properties of fly ash based geopolymer. *Journal of materials science,* 45**,** 607-615.

KUMAR, S. V. & SANTHANAM, M. 2003. Particle packing theories and their application in concrete mixture proportioning: A review. *Indian Concrete Journal,* 77**,** 1324-1331.

KWAN, A. & MORA, C. 2001. Effects of various shape parameters on packing of aggregate particles. *Magazine of concrete Research,* 53**,** 91-100.

LE, Q.-H., NGUYEN, D.-H., SANG-TO, T., KHATIR, S., LE-MINH, H., GANDOMI, A. H. & CUONG-LE, T. 2024. Machine learning based models for predicting compressive strength of geopolymer concrete. *Frontiers of Structural and Civil Engineering,* 18**,** 1028-1049.

LECUN, Y., BENGIO, Y. & HINTON, G. 2015. Deep learning. *Nature,* 521**,** 436-444.

LEE, N. K. & LEE, H. K. 2013. Setting and mechanical properties of alkali-activated fly ash/slag concrete manufactured at room temperature. *Construction and Building Materials,* 47**,** 1201-1209.

LI, L., LIN, C., CHEN, G., KWAN, A. & JIANG, T. 2017. Effects of packing on compressive behaviour of recycled aggregate concrete. *Construction and Building Materials,* 157**,** 757-777.

LI, Q.-F. & SONG, Z.-M. 2022. High-performance concrete strength prediction based on ensemble learning. *Construction and Building Materials,* 324**,** 126694.

LI, Y., SHEN, J., LIN, H. & LI, Y. 2023. Optimization design for alkali-activated slag-fly ash geopolymer concrete based on artificial intelligence considering compressive strength, cost, and carbon emission. *Journal of Building Engineering,* 75**,** 106929.

LIU, G. & SUN, B. 2023. Concrete compressive strength prediction using an explainable boosting machine model. *Case Studies in Construction Materials,* 18**,** e01845.

MEHTA, P. K. & MONTEIRO, P. 2006. Concrete: microstructure, properties, and materials. *(No Title)*.

MUCHERINO, A., PAPAJORGJI, P. J. & PARDALOS, P. M. 2009. k-Nearest Neighbor Classification. *In:* MUCHERINO, A., PAPAJORGJI, P. J. & PARDALOS, P. M. (eds.) *Data Mining in Agriculture.* New York, NY: Springer New York.

NANTHAGOPALAN, P., HAIST, M., SANTHANAM, M. & MüLLER, H. S. 2008. Investigation on the influence of granular packing on the flow properties of cementitious suspensions. *Cement and Concrete Composites,* 30**,** 763-768.

NANTHAGOPALAN, P. & SANTHANAM, M. 2012. An empirical approach for the optimisation of aggregate combinations for self-compacting concrete. *Materials and Structures,* 45**,** 1167-1179.

NATH, P. & SARKER, P. K. 2014. Effect of GGBFS on setting, workability and early strength properties of fly ash geopolymer concrete cured in ambient condition. *Construction and Building Materials,* 66**,** 163-171.

NEVILLE, A. M. 1995. *Properties of concrete*, Longman London.

NIYAZUDDIN & B, U. 2023. Mechanical and durability properties of standard and high strength geopolymer concrete using particle packing theory. *Construction and Building Materials,* 400**,** 132722.

OMER, A. 2014. Energy Efficiency Improvement Utilising High Technology: An Assessment of Energy Use in Industry, Buildings Development and Environment. *Buildings Development and Environment (March 1, 2014)*.

PALACIOS, M., GISMERA, S., ALONSO, M. M., D'ESPINOSE DE LACAILLERIE, J. B., LOTHENBACH, B., FAVIER, A., BRUMAUD, C. & PUERTAS, F. 2021. Early reactivity of sodium silicate-activated slag pastes and its impact on rheological properties. *Cement and Concrete Research,* 140**,** 106302.

PALLAPOTHU, S. N. R. G., PANCHARATHI, R. K. & JANIB, R. 2023. Predicting concrete strength through packing density using machine learning models. *Engineering Applications of Artificial Intelligence,* 126**,** 107177.

PAVITHRA, P., SRINIVASULA REDDY, M., DINAKAR, P., HANUMANTHA RAO, B., SATPATHY, B. K. & MOHANTY, A. N. 2016. A mix design procedure for geopolymer concrete with fly ash. *Journal of Cleaner Production,* 133**,** 117-125.

PENG, Y. & UNLUER, C. 2022. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. *Construction and Building Materials,* 316**,** 125785.

PETERSON, L. 2009. K-nearest neighbor. *Scholarpedia,* 4**,** 1883.

POWERS, T. 1968. Properties of Fresh Concrete, John Wiley and Sons. *Inc., New York***,** 301.

PROVIS, J. L. & BERNAL, S. A. 2014. Geopolymers and Related Alkali-Activated Materials. *Annual Review of Materials Research,* 44**,** 299-327.

PROVIS, J. L. & VAN DEVENTER, J. S. 2013. *Alkali activated materials: state-of-the-art report, RILEM TC 224-AAM*, Springer Science & Business Media.

RAFEET, A., VINAI, R., SOUTSOS, M. & SHA, W. 2017. Guidelines for mix proportioning of fly ash/GGBS based alkali activated concretes. *Construction and Building Materials,* 147**,** 130-142.

RAJAKARUNAKARAN, S. A., LOURDU, A. R., MUTHUSAMY, S., PANCHAL, H., JAWAD ALRUBAIE, A., MUSA JABER, M., ALI, M. H., TLILI, I., MASELENO, A., MAJDI, A. & ALI, S. H. M. 2022. Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques. *Advances in Engineering Software,* 173**,** 103267.

RASHAD, A. M. 2013. Properties of alkali-activated fly ash concrete blended with slag. *Iran. J. Mater. Sci. Eng,* 10**,** 57-64.

SHEN, J., LI, Y., LIN, H., LI, H., LV, J., FENG, S. & CI, J. 2022. Prediction of compressive strength of alkali-activated construction demolition waste geopolymers using ensemble machine learning. *Construction and Building Materials,* 360**,** 129600.

SMOLA, A. J. & SCHöLKOPF, B. 2004. A tutorial on support vector regression. *Statistics and computing,* 14**,** 199-222.

SOUTSOS, M., BOYLE, A. P., VINAI, R., HADJIERAKLEOUS, A. & BARNETT, S. J. 2016. Factors influencing the compressive strength of fly ash based geopolymers. *Construction and Building Materials,* 110**,** 355-368.

SUN, B., DING, L., YE, G. & DE SCHUTTER, G. 2023a. Mechanical properties prediction of blast furnace slag and fly ash-based alkali-activated concrete by machine learning methods. *Construction and Building Materials,* 409**,** 133933.

SUN, Y., CHENG, H., ZHANG, S., MOHAN, M. K., YE, G. & DE SCHUTTER, G. 2023b. Prediction & optimization of alkali-activated concrete based on the random forest machine learning algorithm. *Construction and Building Materials,* 385**,** 131519.

TAVARES, C., WANG, X., SAHA, S. & GRASLEY, Z. 2022. Machine learning-based mix design tools to minimize carbon footprint and cost of UHPC. Part 1: Efficient data collection and modeling. *Cleaner Materials,* 4**,** 100082.

TAYLOR, H. F. 1997. *Cement chemistry*, Thomas Telford London.

THOMAS, R. J. & PEETHAMPARAN, S. 2017. Stepwise regression modeling for compressive strength of alkali-activated concrete. *Construction and Building Materials,* 141**,** 315-324.

TOUFAR, W., BORN, M. & KLOSE, E. 1976. Contribution of optimisation of components of different density in polydispersed particles systems. *Freiberger booklet A,* 558**,** 29-44.

TOUFIGH, V. & JAFARI, A. 2021. Developing a comprehensive prediction model for compressive strength of fly ash-based geopolymer concrete (FAGC). *Construction and Building Materials,* 277**,** 122241.

TURNER, L. K. & COLLINS, F. G. 2013. Carbon dioxide equivalent (CO2-e) emissions: A comparison between geopolymer and OPC cement concrete. *Construction and Building Materials,* 43**,** 125-130.

WEIL, M., DOMBROWSKI, K. & BUCHWALD, A. 2009. Life-cycle analysis of geopolymers. *Geopolymers.* Elsevier.

YIP, C. K.-B. 2004. *The role of calcium in geopolymerisation.* University of Melbourne, Department of Chemical and Biomolecular Engineering.

ZHANG, L. V., MARANI, A. & NEHDI, M. L. 2022. Chemistry-informed machine learning prediction of compressive strength for alkali-activated materials. *Construction and Building Materials,* 316**,** 126103.

ZHANG, R., HE, H., SONG, Y., ZHI, X. & FAN, F. 2023. Influence of mix proportioning parameters and curing regimes on the properties of ultra-high strength alkali-activated concrete. *Construction and Building Materials,* 393**,** 132139.

ZHANG, W., WU, C., ZHONG, H., LI, Y. & WANG, L. 2021. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers,* 12**,** 469-477.

ZHU, L., WANG, J.-J., LI, X., ZHAO, G.-Y. & HUO, X.-J. 2020. Experimental and numerical study on creep and shrinkage effects of ultra high-performance concrete beam. *Composites Part B: Engineering,* 184**,** 107713.

SINGH, R. P., REDDY, P. S., VANAPALLI, K. R. & MOHANTY, B. 2023. Influence of binder materials and alkali activator on the strength and durability properties of geopolymer concrete: A review. *Materials Today: Proceedings*.

POWERS, T. 1968. Properties of Fresh Concrete, John Wiley and Sons. *Inc., New York***,** 301.

DE LARRARD, F. 1999. *Concrete mixture proportioning: a scientific approach*, CRC Press.

SOUTSOS, M., BOYLE, A. P., VINAI, R., HADJIERAKLEOUS, A. & BARNETT, S. J. 2016. Factors influencing the compressive strength of fly ash based geopolymers. *Construction and Building Materials,* 110**,** 355-368.

MCLELLAN, B. C., WILLIAMS, R. P., LAY, J., VAN RIESSEN, A. & CORDER, G. D. 2011. Costs and carbon emissions for geopolymer pastes in comparison to ordinary portland cement. *Journal of cleaner production,* 19**,** 1080-1090.

GLUKHOVSKY, V. 1959. *Soil silicates,* Kiev, Gosstroyizdat.