PREDICTION OF COMPRESSIVE STRENGTH OF FLY ASH-BASED GEOPOLYMER CONCRETE USING SUPERVISED MACHINE LEARNING MODELS

Final Year Project

SUPERVISOR

DR. ARSLAN QAYYUM KHAN

PROJECT MEMBERS

MUHAMMAD HUZAIFA NAVEED	**BSCE01193115**
MUHAMMAD DAWOOD RASHEED	**BSCE01193105**
ALI HASNAIN	**BSCE01193138**

DEPARTMENT OF CIVIL ENGINEERING
THE UNIVERSITY OF LAHORE (LAHORE CAMPUS)

AUGUST, 2023

PREDICTION OF COMPRESSIVE STRENGTH OF FLY ASH-BASED GEOPOLYMER CONCRETE USING SUPERVISED MACHINE LEARNING MODELS

Final Year Project

PROJECT MEMBERS

MUHAMMAD HUZAIFA NAVEED	**BSCE01193115**
MUHAMMAD DAWOOD RASHEED	**BSCE01193105**
ALI HASNAIN	**BSCE01193138**

SUPERVISOR

Dr. Arslan Qayyum Khan
Assistant Professor
Department of Civil Engineering

EXTERNAL EXAMINER

Engr. Mehfooz Hayat
Assistant Director
Punjab Irrigation Department

HEAD OF DEPARTMENT

Dr. Muhammad Kaleem-Ullah
Department of Civil Engineering

A final year project submitted in partial fulfilments of the requirements for the award of degree of Bachelors of Science in Civil Engineering.

DEPARTMENT OF CIVIL ENGINEERING
THE UNIVERSITY OF LAHORE (LAHORE CAMPUS)

AUGUST, 2023

DEDICATION

“I dedicated this effort to my lovely parents, who always pray for my success and their love and affection have always been a source of inspiration for me”

DECLARATION

We declare that this dissertation is our original work, based on original research. Other authors, researchers, and institutions' references have been appropriately cited.

MUHAMMAD HUZAIFA NAVEED	**BSCE01193115**
MUHAMMAD DAWOOD RASHEED	**BSCE01193105**
ALI HASNAIN	**BSCE01193138**

ACKNOWLEDGEMENT

In the Name of Allah, the Beneficent, the Merciful

I begin by offering my heartfelt gratitude to the Almighty Allah. It is through His unwavering presence and wisdom that I have navigated every facet of my journey. His grace, strength, and guidance have been my constant companions. Equally, I acknowledge the profound influence of Prophet Muhammad (ﷺ) whose teachings have illuminated my path.

I extend my sincerest appreciation to my esteemed teachers. Their guidance, engaging discussions, and valuable suggestions have shaped my growth and progress. Their mentorship has been instrumental in shaping my identity today. I am also deeply indebted to my parents. Their unyielding support and unwavering encouragement have sustained me through every challenge. My father's unwavering belief in me and my mother's steadfast confidence in my capabilities have instilled in me the courage to venture beyond limitations.

I would like to express my heartfelt gratitude to our supervisor, **Dr. Arslan Qayyum Khan**, whose guidance, thorough review, and helpful feedback were invaluable. His timely and efficient contributions played a crucial role in shaping this work into its final form, and we sincerely appreciate his willingness to assist us in any way we requested. Furthermore, we extend our thanks to the Department of Civil Engineering, University of Lahore, Lahore, its leadership, and the staff for providing us with a strong academic foundation that enabled us to undertake and complete this project.

Lastly, we would like to acknowledge and give special thanks to our family members for their unwavering support and encouragement throughout our project journey, helping us navigate through this phase of life. We are truly indebted to them for everything.

Authors

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Fly ash (FA) based geopolymer concrete has lately gained popularity as a low-carbon and sustainable alternative to Portland cement concrete. However, accurate prediction of its compressive strength is still challenging due to the many chemical and physical interactions involved in the geopolymerization process. In this study, three machine learning models backpropagation neural network (BPNN), random forest regression (RFR), and k-nearest neighbors (KNN) were used to estimate the compressive strength of FA-based geopolymer concrete. The models were trained, validated, and tested using a dataset that considered the chemical composition, mix proportions, and pre-curing conditions of the concrete. The coefficient of determination (R^2), mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) were among the metrics used to evaluate each model's performance. The findings showed that, in comparison to RFR and KNN, which had R^2 values of 0.927 and 0.911, respectively, the BPNN model produced the best results with an R^2 value of 0.948. The coarse aggregate content, SiO_2 content in FA, and NaOH concentration were shown to have the highest influence on the compressive strength of the FA-based geopolymer concrete, according to the permutation feature important (PFI) index.

LIST OF ABBREVIATION

ANN	Artificial Neural Network
ML	Machine Learning
PC	Portland Cement
FA	Fly Ash
CA	Coarse Aggregate
FA	Fine Aggregate
W/C	Water to cement ratio
BPNN	Back-Propagation Neural Network
RFR	Random Forest Regressor
KNN	K-Nearest Neighbor
PFI	Permutation Feature Importance
M	Molar
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
SiO_2	Silicone Dioxide
Al_2O_3	Aluminum Trioxide
NaOH	Sodium Hydroxide

1. INTRODUCTION

Concrete is a widely used construction material known for its durability. However, its production contributes significantly to carbon dioxide emissions, exacerbating global warming. As environmental concerns grow, the construction industry is seeking alternatives that conserve natural resources and minimize environmental impacts. One such promising option is geopolymer concrete (GPC), a sustainable material that effectively utilizes industrial waste, particularly fly ash, and serves as a greener alternative to ordinary Portland cement (OPC) concrete.

Globally, Portland cement is the most widely utilized cementing material in concrete production and is associated with considerable energy utilization and elevated carbon dioxide emissions. (Peng & Unluer, 2022). Annually, the world produces 1.6 billion tons of cement, which results in 7% of global carbon emissions, equivalent to 4 billion tons of CO_2. The generation of emissions in production stems from the utilization of fossil fuels for calcination, production of minerals and transportation, which results in indirect emissions, as well as the direct release of CO_2 during the transformation of calcium carbonate ($CaCO_3$) to calcium oxide (CaO). The proportion of indirect and direct CO_2 emissions depends on the production techniques and location, however they both roughly contribute equally (V. M. Malhotra, 2010). The concrete sector consumes

substantial quantities of freshwater, with the yearly requirement for mixing water alone reaching roughly 1 trillion litres. The slower setting and curing speed of concrete that contains a high volume of a mineral admixture can be partially compensated by reducing the water-cement ratio with the assistance of a superplasticizer (V. Malhotra, 1999). A growing trend in the use of Portland cement blends that include cementitious or pozzolanic by-products, such as ground granulated blast-furnace slag and fly ash, is being observed. (Kumar Mehta, n.d.). Fly ash is a byproduct generated from the burning of coal to produce electricity. In 1984, the U.S. produced around 70 million tons of fly ash, of which approximately 7 million tons were utilized in Portland Cement Concrete (PCC) (Gadja & Vangeem, 2001). The usage of supplementary cementitious materials (SCMs) as a replacement for Portland cement (PC) is often limited by proportions. For instance, while fly ash (FA) exhibits pozzolanic effects during various stages of cement hydration, it doesn't play a significant role in early-stage strength development. (V. Malhotra, 1999). The inclusion of FA leads to a decrease in the early rate of hydration and extends the setting time of composite paste. However, this can restrict its utilization in substantial amounts (X. Han et al., 2019).

1.1.BACKGROUND

The production of OPC, the main component of traditional concrete, is expected to increase substantially in the coming years due to global infrastructure development. This surge in OPC production poses serious environmental and health hazards, as cement production accounts for about 7% of total greenhouse gas emissions. The excessive release of carbon dioxide into the atmosphere may lead to a significant rise in Earth's temperature. Furthermore, the disposal of industrial wastes like fly ash and other by-products creates challenges in terms of land usage and environmental impact.

In response to these challenges, researchers have explored the potential of geopolymer concrete as an eco-friendly alternative to OPC. Geopolymer concrete is a novel material that utilizes alkali activation of aluminosilicate-rich materials, including fly ash, metakaolin, red mud, and ground granulated blast furnace slag. Among these materials, fly ash stands out due to its abundance, low cost, and high potential for making geopolymers.

1.2.SIGNIFICANCE

This research brings significant benefits to both the construction industry and the research community. It introduces a novel method to address challenges in civil engineering, offering a fresh approach to problem-solving. By embracing these innovative methods, the construction industry stands to gain in terms of time and cost savings, which would otherwise be expended using conventional approaches. Moreover, the research community can expand their knowledge and expertise by exploring and refining these new methods, fostering advancements in the field of civil engineering.

The use of geopolymer concrete can significantly reduce CO_2 emissions compared to OPC concrete. In fact, studies suggest that GPC contributes about 20–50% lower CO_2 gas emissions. Notably, the carbon footprint of geopolymer concrete has garnered attention in sustainable construction practices. Geopolymer concrete's capability to reduce CO_2 emissions by approximately 80% compared to conventional cement industries makes it a compelling choice for reducing the environmental impact of construction activities.

The use of fly ash and blast furnace slag in the production of other construction materials would both solve an environmental issue and lead to the creation of new, high-performing materials. The search for more environmentally friendly cement-based

binders through the process of alkali activation is a commonly studied method for completely replacing Portland cement (Puertas et al., n.d.). Fly ash has been utilized as a substitute material for the creation of geopolymer, a novel binding agent with similar appearance, reactivity, and characteristics as hydrated cement.

1.3.SCOPE AND LIMITATION

Geopolymer concrete exhibits high early compressive strength, low shrinkage, significant resistance to creep, and good performance in acidic environments. However, its long-term durability remains an area requiring further research. The service life and durability of concrete structures largely depend on material transport properties, such as permeability, sorptivity, and diffusivity, which require in-depth investigation for geopolymer concrete.

As the world becomes more conscious of the need for sustainable practices in civil engineering, geopolymer concrete has emerged as a promising solution. Its utilization of waste materials and reduced CO_2 emissions make it a sustainable alternative to traditional concrete. However, further research and development are necessary to fully understand its long-term durability and to overcome any challenges associated with its application.

1.4.THESIS ORGANIZATION

The thesis comprises five comprehensive chapters, each delving into distinct aspects of the subject matter, thereby providing a thorough and insightful examination of the topic.

1.4.1. CHAPTER 1 – INTRODUCTION

This pivotal chapter presents the readers with essential background information pertaining to the theme, followed by a clear statement of the problem at hand. Additionally, it highlights the significance of the problem in the context of the

construction industry. The scope and limitations of the study are carefully delineated to provide a concise framework. Lastly, the chapter outlines the organization of the thesis, guiding readers through the subsequent chapters.

1.4.2. CHAPTER 2 - LITERATURE REVIEW

In this chapter, an extensive review of relevant literature on various aspects is provided, including the classification of concrete, factors influencing compressive strength development at different ages, and predictive methods for compressive strength. An in-depth exploration of Artificial Neural System (ANN) and Regression Models, its training procedure is presented, along with a discussion of the ANN model and experimental results, and the compatibility and achievements derived from this research.

1.4.3. CHAPTER 3 - RESEARCH METHODOLOGY

This chapter offers valuable insights into the research approach adopted, the sources and nature of data used, and the utilization of Artificial Neural Network (ANN) and Regression Models as a vital tool in the analysis. Detailed information is provided regarding data processing and analyzing techniques, as well as the step-by-step process of modeling an ANN and Regression for the research.

1.4.4. CHAPTER 4 - RESULTS AND DISCUSSION

Within this chapter, a detailed discussion is provided, primarily focusing on the experimental results acquired from 11 references, alongside the predictions generated by the ANN model for concrete compressive strength. An in-depth analysis of these findings serves to shed light on their implications for the field of civil engineering.

1.4.5. CHAPTER 5 - CONCLUSIONS AND RECOMMENDATIONS

As the final chapter, this section offers a comprehensive conclusion, summarizing the research outcomes gained through ANN modeling. Based on these results, informed and practical recommendations are provided, adding to the existing body of knowledge and offering valuable insights for future research and application.

By adopting a meticulous and structured approach, this thesis contributes significantly to the construction industry and research community alike, paving the way for the advancement of knowledge and innovative problem-solving techniques in civil engineering

2. LITERATURE REVIEW

2.1.FLY ASH

Fly ash is a fine, powdery substance that results from burning pulverized coal in electric power plants. When coal is burned, its mineral impurities, mainly silica, alumina, and iron, react with lime during combustion, forming small spherical particles known as fly ash. These particles are then captured from the flue gas by pollution control devices like electrostatic precipitators or baghouses.

As a pozzolanic material, fly ash exhibits cementitious properties when combined with water and an alkaline activator, such as sodium hydroxide or potassium hydroxide. This characteristic makes fly ash an essential component in the production of cement and concrete. By incorporating fly ash as a partial replacement for cement in concrete, the environmental impact of concrete manufacturing can be reduced. This reduction is achieved by lowering the demand for clinker production, a key ingredient in cement manufacturing and a major source of carbon dioxide emissions. Moreover, the addition of fly ash in concrete can enhance specific properties, including workability, durability, and resistance to certain chemical attacks.

Due to its utilization of waste material from power generation and its contribution to minimizing the consumption of natural resources, fly ash is regarded as an environmentally sustainable material. Its versatility and positive environmental

attributes make it a valuable resource with widespread applications in the construction industry and other related fields. Fly ash has been utilized as a substitute material for the creation of geopolymer, a novel binding agent with similar appearance, reactivity, and characteristics as hydrated cement. Essentially, geopolymer is formed through the activation of aluminosilicate materials with alkalis. Alkali activation is achieved by adding sodium hydroxide (NaOH), potassium hydroxide (KOH), sodium silicate (Na_2SiO_3) or potassium silicate (K_2SiO_3) individually or in combination with fly ash. This process, known as geopolymerization, can take place at room temperature or slightly elevated temperatures (typically below 100°C) and most importantly, with minimal CO_2 emissions. Geopolymers possess advantageous mechanical and durability properties due to their distinct chemical structure, and are eco-friendly as they utilize industrial waste as a binding material in contrast to mixes based on Portland cement (PC). Fly ash geopolymer cement is a cost-effective, low-carbon, low-energy alternative to ordinary Portland cement.

2.2.PROPERTIES OF FLY ASH

Fly ash geopolymers have a better ability to adsorb and stabilize toxic metals compared to ordinary Portland cement (OPC) and can be utilized as a sealant for storing CO_2, however with lower permeability when compared to typical sealants.(Zhuang et al., 2016). Geopolymer mortar and concrete exhibit comparable strength, texture, and appearance, and offer similar mechanical characteristics (Chindaprasirt et al., 2011). The interlinked structure of geopolymers results in high compressive strengths. Due to the microstructure of their reaction products, geopolymers and Alkali-Activated Binders (AAB) exhibit good to excellent resistance to chemical degradation such as sulphate attack, acid attack, and seawater exposure (Soutsos et al., 2016). Fly ash also

demonstrates improved durability, including resistance to high temperatures(Kong & Sanjayan, 2010) and chloride penetration (Chindaprasirt & Chalee, 2014).

2.3.FACTORS AFFECTING PROPERTIES OF FLY ASH

In the geopolymerization process, fly ash plays a vital role as a raw material. However, several factors can influence its behaviour and properties during this process.

2.3.1. CHEMICAL ACTIVATORS

The presence of a chemical activator or alkali activator solution is crucial for initiating the geopolymerization of fly ash. A strong alkaline medium, typically containing NaOH, enhances the surface hydrolysis of the aluminosilicate particles in the fly ash, which is essential for the reaction to occur.

2.3.2. CONCENTRATION OF CHEMICAL ACTIVATORS

The concentration of the chemical activator has a pronounced effect on the mechanical properties of the geopolymers formed. Increasing the concentration of chemical activators, especially NaOH, has been shown in research papers to lead to enhanced compressive strength in geopolymers.

2.3.3. MULTI-COMPOUND CHEMICAL ACTIVATORS

The most effective chemical activators for improving mechanical strength in geopolymers are those containing Na_2-SiO_3/NaOH. Incorporating these multi-compound activators into the geopolymer matrix contributes to the desired strength enhancement.

2.3.4. HEAT CURING

Applying heat curing is crucial to expedite the initial strength growth of fly ash-based geopolymer concrete. Both the duration and temperature of this process play pivotal

roles. Generally, the recommended optimal heat curing conditions encompass a maximum period of 24 hours, coupled with a temperature span ranging from 50°C to 90°C. These conditions have proven advantageous for fostering both short-term and long-term strength enhancement, as well as ensuring the durability and steadfastness of geopolymer concrete.

Liquid Alkaline/Ash Ratio: The ratio of liquid alkaline to fly ash in the mixture holds sway over the workability of fresh geopolymer blends. Elevating the liquid alkaline/ash ratio typically yields improved workability.

2.3.5. Na_2-SiO_3/NaOH RATIO

The ratio of Na_2-SiO_3 to NaOH in the mixture affects the workability of fresh geopolymer mixtures. Higher Na_2-SiO_3/NaOH ratios can reduce workability due to the increased viscosity of Na_2SiO_3.

2.3.6. NAOH CONCENTRATION

Higher concentrations of NaOH in the geopolymer mixture can lead to longer setting times. It is crucial to consider the appropriate concentration to achieve the desired setting time and performance.

2.3.7. SHRINKAGE

The shrinkage of geopolymer concrete is closely related to its corresponding strength. Lower strength development in geopolymers is associated with increased shrinkage.

Lastly, understanding the factors that affect fly ash during the geopolymerization process is vital for optimizing the properties and performance of geopolymers. Careful consideration of chemical activators, heat curing conditions, and mixing ratios can lead to geopolymers with enhanced mechanical strength, workability, and reduced shrinkage.

2.4.NEURAL NETWORKS

Neural networks are a collection of strong machine learning algorithms that have been influenced by the structure and operation of the human brain. These artificial neural networks are made up of linked nodes, or neurons, that are arranged in layers. The input layer receives raw data, which is subsequently passed on to hidden layers, where each neuron conducts mathematical computations on the data. During training, the degree of interconnections between neurons, indicated by weights, is modified to learn patterns and correlations in the data. The last layer, known as the output layer, generates predictions or classification based on the previously acquired patterns. Neural networks have demonstrated extraordinary ability in tasks like as picture and audio identification, processing natural languages, and strategic game play. With advances in machine learning and an abundance of datasets and computational power, neural networks tend to transform different sectors and drive artificial intelligence forward. The most widely used ML methods can be grouped into four main categories: ANN, SVM, decision trees, and evolutionary algorithms. Artificial Neural Networks (ANN) operate by transmitting information through interconnected neurons, where each information is weighted based on its significance.

2.5.ARTIFICIAL NEURAL NETWORK

An Artificial neural Network (ANN) is an algorithmic framework that tries to imitate the structure and operation of biological brain networks. There are two types of ANN applications: classification models and regression models. Researchers have extensively investigated the application of ANN for forecasting concrete compressive strength. In terms of accuracy, they discovered that ANN models trump standard regression models. The multilayer perceptron (MLP) is the most widely utilized of the different ANN models. The MLP model is made up of an input layer with sensory input nodes, one or

more hidden layers for calculation, and an output layer with a single node representing the compressive strength of concrete. The most extensively used and successful learning strategy for training the MLP model is the back-propagation (BP) algorithm. Equations (1) and (2) mathematically explain the activation mechanism for each neuron in the MLP.

$$netk = \sum wkjoj \quad (1)$$

$$yk = f(netk) \quad (2)$$

The neurons in the preceding layer (denoted as j) impact the activity of neuron k (denoted as *netk*) in the context of the neural network model. The weight wkj represents the link between neurons *k* and *j*. Neuron *j's* output is indicated as *oj*, while neuron *k's* ultimate output is commonly computed using sigmoid or logistical transfer functions and is labeled as *yk*. The illustration of ANN is depicted in Fig.1.

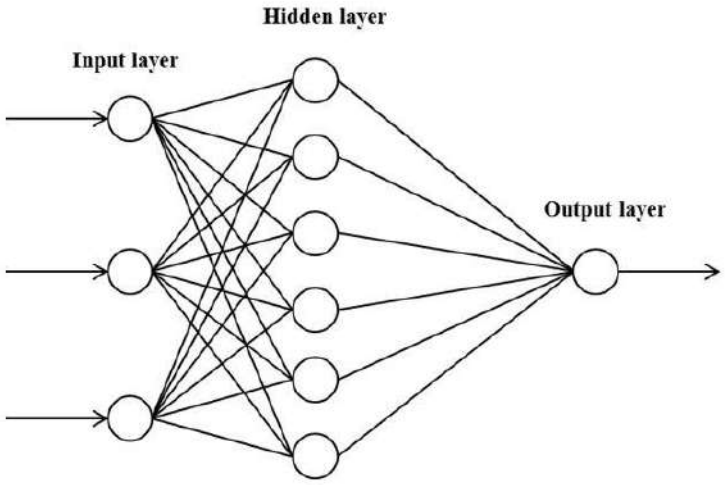


Fig 1: Structure of ANN with one hidden layer (Jin et al., 2016)

2.6.NON-LINEAR REGRESSOR

Non-linear regression is a type of regression analysis used in artificial intelligence methods to describe connections between variables when data does not follow a linear pattern. Non-linear regression, as opposed to linear regression, which assumes a linear

connection between the independent and dependent variables, allows for more complicated and flexible interactions.

A non-linear function represents the connection between the independent factors and the dependent variable in non-linear regression. The function's shape might change based on the task and the type of the data. Polynomial functions, exponential functions, logarithmic functions, and trigonometric functions are common examples of non-linear functions used in regression. Non-linear regression is especially beneficial when working with data that contains complex patterns and connections, or where a linear model is unable to fully reflect the underlying trends. It is extensively used in AI systems for jobs like as demographic prediction.

$$y = f(x, \beta) + \varepsilon \quad (3)$$

Where:

- y is the dependent variable (the predicted output).
- x is the independent variable (the input or predictor variable).
- $f(x, \beta)$ is the non-linear function with parameters β that captures the relationship between x and y.
- ε represents the error term, accounting for the difference between the predicted value and the actual value in the data.

2.7.SUPPORT VECTOR MACHINE

SVM is an algorithm that is used for supervised learning problems. It is most commonly used for classification issues in which the aim is to categorize data into multiple classes or groups. SVM determines the appropriate hyperplane (a decision boundary) for separating data points into their respective classes with the greatest margin. It seeks to

minimize the distance between the closest data points of distinct classes, allowing for greater generalization and performance on fresh data.

The k-nearest neighbour (kNN) algorithm is a traditional non-parametric technique (Cover & Hart, 1967). The kNN algorithm is designed to find k-nearest neighbours of a query and assign a class label to the query through the majority voting rule. Due to its simplicity, effectiveness, and intuitiveness, it is one of the most commonly used algorithms in computer programming today (Jiang et al., 2012; Wu et al., 2007). The kNN algorithm is a non-parametric classification method that does not need a training process. Particularly, it does not require prior knowledge about the statistical properties of the training instances (Li et al., 2008; Pan et al., 2020). The illustration of SVM is depicted in Fig.2.

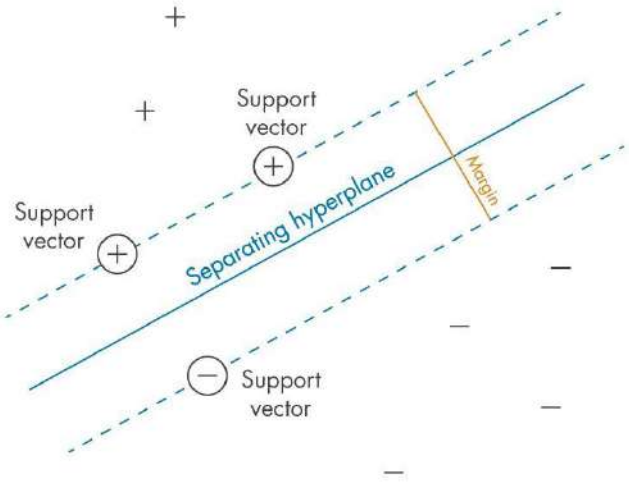


Fig 2: Depiction of SVM Model (Dener et al., 2022)

The functional relationship between one or more independent variables and the response variable is explained by the following model:

$$y(x) = w^T \phi(x) + b \qquad (4)$$

where $x \in R$, $y \in R$, and $\phi(x)$: R^n is the process of mapping to higher dimensional feature space.

2.8.APPLICATION OF NEURAL NETWORKS IN CIVIL ENGINEERING

Over the last five years, there has been a boom in interest among academics in the application of artificial neural networks (ANN) in the subject of Civil Engineering. Because of its ability to handle many issues and complexity in the area, the application of ANN in Civil Engineering has piqued the interest of several academics. ANNs have been investigated for a variety of applications, including structural analysis, material characteristics prediction, construction process optimization, and predictive decision-making in civil engineering projects. The capacity of artificial neural networks (ANNs) to learn from data as well as adapt to complicated patterns has proven useful in tackling real-world engineering challenges. This increased interest in ANN in Civil Engineering is likely to fuel more developments and improvements in the industry, resulting in more efficient solutions.

A number of researchers have applied neural networks to forecast compressive strengths, shear strength of RC Deep Beams, and investigate the impact of various parameters on the outcomes. As a result, this approach has proven to be time-efficient and cost-effective. (H. Zhang et al., 2020) employed two models to forecast the residual compressive strength of geopolymer concrete after being subjected to varying temperatures. These two models were created utilizing a Gaussian-centered mathematical framework, with the coefficients established by analyzing experimental outcomes using MATLAB software. Notably, it was noticed that the actual compressive strengths tended to surpass the predicted strengths(H. Zhang et al., 2020).

The utilization of machine learning models has gained extensive adoption as a potent means to forecast the mechanical characteristics of concrete. These models typically work with sizable datasets, which are commonly partitioned into distinct training, validation, and testing segments. The training subset facilitates model training, whereas

the validation dataset ensures impartial assessment of the model's alignment with the training data. It also safeguards against overfitting by stopping the training process if errors start to rise. Finally, the model is applied to the testing data to evaluate its predictive performance.

In the study by Naderpour et al. (Naderpour et al., 2018), the backpropagation artificial neural network was utilized. The study was conducted to develop an Artificial Neural Network (ANN) for evaluating the strength properties of recycled aggregate concrete. The ANN was based on predetermined input variables that were considered key in determining the strength properties. Later a study conducted by Asteris et al. (Asteris & Kolovos, 2019) used Artificial Neural Network (ANN) to predict the compressive strength of self-compacting concrete (SCC). The model was trained using the Levenberg-Marquardt algorithm and the results showed that Backpropagation Artificial Neural Network (BPNN) was effective in accurately predicting the compressive strength of SCC. The sensitivity analysis conducted further revealed that viscosity-modifying admixtures had the greatest impact on the compressive strength of SCC. Random Forest has been adopted as a forecasting tool in several studies. The technique entails combining multiple decision trees, each generated from a different training set through the bagging method (ben Chaabene et al., 2020). The bagging method, commonly referred to as bootstrap aggregation, is a machine learning technique that involves generating new datasets from the original data by randomly resampling it, and using these datasets to train individual base predictors independently. In the final step, the predictions from each base predictor are averaged to produce the final prediction. This method has been applied by various researchers in predicting the mechanical properties of concrete. In the study by Han et al. (Han et al., 2019), Random Forest (RF) was used to predict the compressive strength of High-Performance Concrete (HPC).

Another study by Mangalathu et al. (Mangalathu & Jeon, 2018) utilized the same RF method to forecast the shear strength of Reinforced Concrete (RC) beam-column joints and obtained good agreement in results. Zhang et al. (J. Zhang et al., 2019a) also used RF, but combined it with a beetle antennae search (BAS) algorithm, to evaluate the uniaxial compressive strength of Self-Compacting Concrete (SCC). The BAS algorithm was based on the behavior of beetles as they search for a location with a higher concentration of odor using their antennae. These studies demonstrate that RF can produce accurate predictions for the mechanical strength of concrete. These approaches make things economic in terms of both cost and time.

As Neural Networks in civil engineering provide efficient solutions for predicting material properties, structural behaviour, and complex interactions. Their adaptability to diverse data sets saves time and costs while enhancing decision-making and improving the overall performance of civil engineering projects.

3. RESEARCH METHODOLOGY

3.1.PREDICTION OF COMPRESSION STRENGTH

The study focuses on predicting the compressive strength of geopolymer concrete using three machine-learning algorithms. It will analyse the chemical composition, mix proportions, and pre-curing conditions of geopolymer concrete to determine the most accurate algorithm for predicting its 28-day compressive strength. The study aims to fill a research gap and provide insights into using machine learning algorithms to predict geopolymer concrete strength.

3.2.MACHINE LEARNING APPROACHES

3.2.1. MACHINE LEARNING MODELS

3.2.1.1. BPNN

Backpropagation Neural Network (BPNN) has gained extensive use as a learning algorithm within multilayer neural networks. Initially introduced by Paul Werbos in 1974 and subsequently rediscovered by Rumelhart and Parker, BPNN stands as the favoured, efficient, and accessible model for intricate, layered networks. Backpropagation is a form of supervised learning rooted in the gradient descent approach. It aims to minimize network errors by descending along the gradient of the error curve. (Hamid et al., 2011; Khalil et al., n.d.). Because of the weight error correct

rules, the supervised learning model is the most popular. BPNN only needs one hidden layer to meet the requirement of prediction accuracy(Peng & Unluer, 2022). The structure of this neural network consists of input sets and one or more layers of neurons. Neurons are connected via layers of parallel neurons except for the special input Xo partially represents each neuron. Each neuron in one layer is linked to all neurons in the following layer. Hidden layers are the last layer that creates the model's output and any layers that come before it. The only role of inputs is to feed input patterns into the rest of the network; they do no computing (Sapna, 2012).

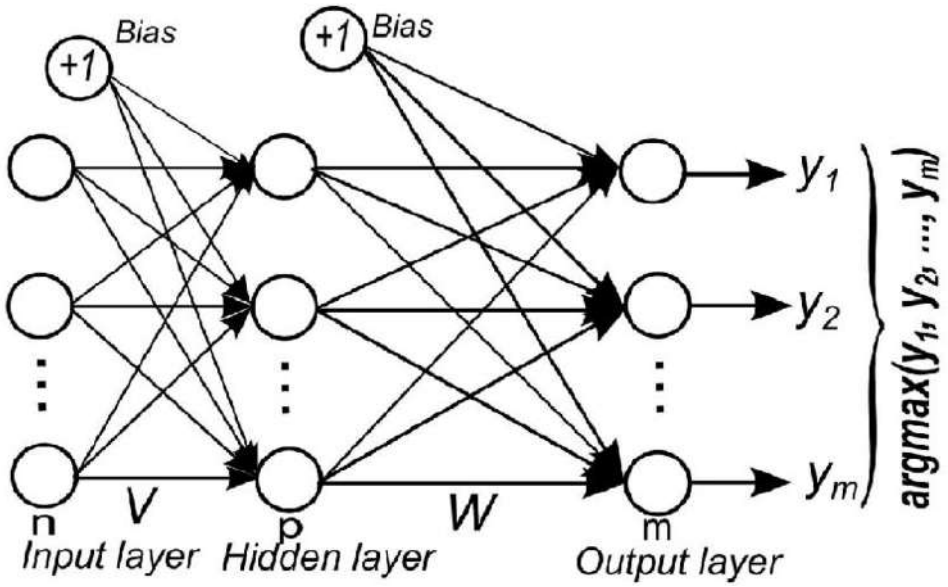


Fig 3: Illustration of BPNN (Yeresime et al., 2014)

3.2.1.2. RFR

RFR was proposed by Breiman. Random forests suggest adding an extra layer of randomization to bagging. Random forests are a useful method for making predictions, and they are less likely to overfit due to the Law of Large Numbers. By incorporating appropriate forms of randomness, they can accurately classify and predict regression outcomes. RFR creates By employing distinct bootstrap data samples for each tree, random forests alter the construction process of classification or regression trees. While

conventional trees divide nodes using the optimal split across all variables, random forests introduce a unique approach. At each node, the division is executed using the superior choice from a subset of predictors selected randomly. This seemingly unconventional technique yields impressive performance, surpassing several other classifiers like discriminant analysis, support vector machines, and neural networks. Furthermore, it exhibits resilience against overfitting as outlined by Breiman in 2001. Moreover, its simplicity shines through, as it only necessitates two parameters (the count of variables in the random subset at each node and the number of trees in the forest) and generally showcases robustness to their values (Liaw & Wiener, 2002). The data are split into several points for each input variable, and the Sum of Square Error (SSE) is computed for both the predicted and actual values at each point. Subsequently, this node's least SSE value is obtained (J. Zhang et al., 2019b).

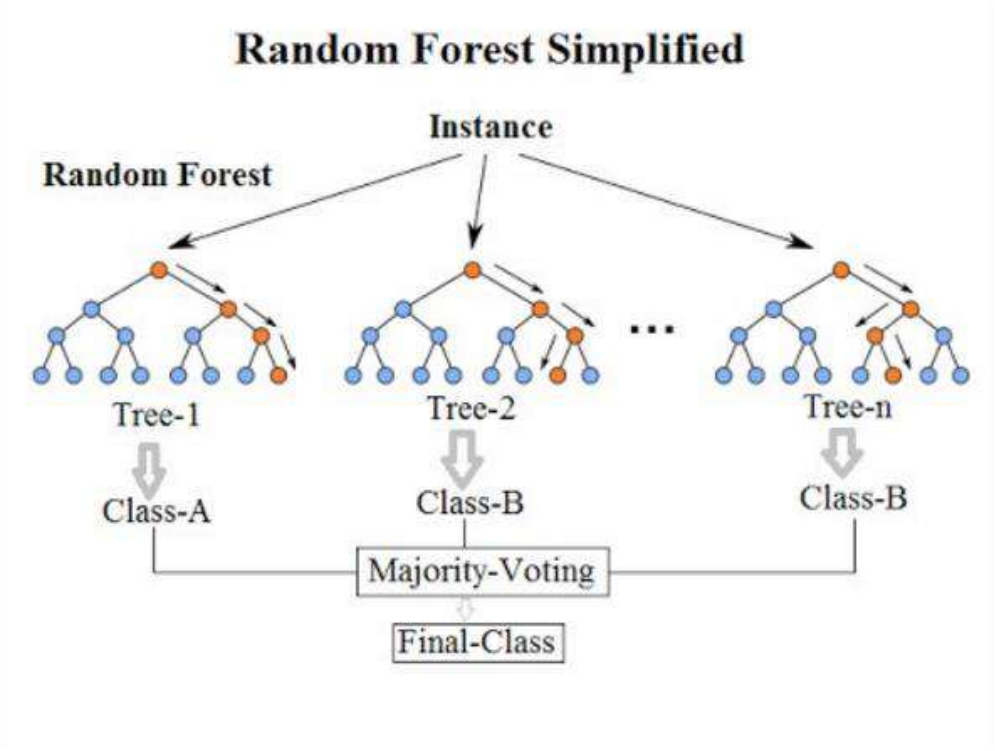


Fig 4: Illustration of RFR (Raja & Fokoué, 2019)

3.2.1.3. KNN

The k-nearest neighbor (kNN) algorithm is a traditional non-parametric technique (Cover & Hart, 1967). The kNN algorithm is formulated to identify the k-closest neighbors of a given query and designate a class label to the query using the principle of majority voting. Its straightforwardness, efficiency, and natural approach contribute to its effectiveness. It is one of the most commonly used algorithms in computer programming today (Jiang et al., 2012; Wu et al., 2007). The kNN algorithm is a non-parametric classification technique that operates without the need for a training phase. Specifically, it doesn't necessitate prior understanding of the statistical characteristics of the training examples (Li et al., 2008; Pan et al., 2020).

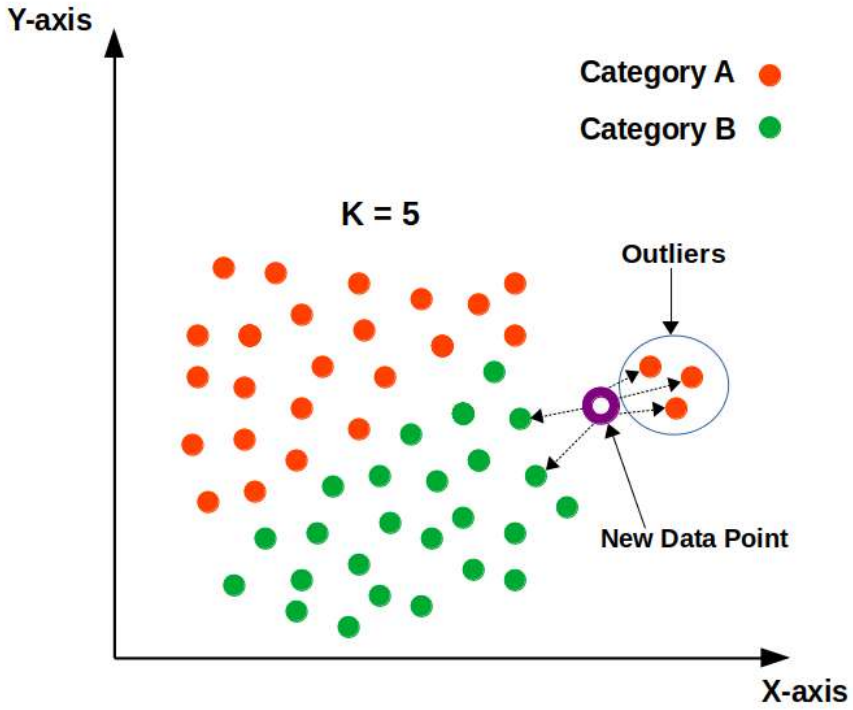


Fig 5: Illustration of KNN (Shivam Sharma, 2021)

3.2.2. TRAINING AND EVALUATING MACHINE LEARNING MODELS

Fig.6 shows the combined calculation process of all three machine learning algorithms. Firstly, the collected data were arranged in random order. BPNN parameters were

established as they control the balance between gradient descent and the LM algorithm. The Levenberg-Marquardt algorithm is a method specifically designed for minimizing functions that consist of sums of squares of nonlinear functions. It is particularly suitable for training neural networks. It is well-suited for training neural networks, and despite requiring many computations, it appears to be the quickest training algorithm for neural networks with moderate numbers of parameters.

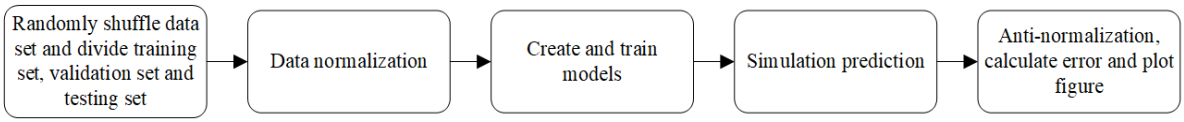


Fig 6: Process of Machine Learning Process

In particular, when it comes to multilayer networks used for function approximation that have up to a few hundred weights and biases, the Levenberg-Marquardt algorithm is generally the fastest training method (Demuth & De Jesús, n.d.). BPNN was trained on MATLAB 2021a while the other two models RFR and KNN were trained on PYTHON. The hyperparameters that were used in training the models are presented in Table *I*. For each of the models, different hyperparameters were used. Each of the models has its parameters. To make sure that the data was effectively divided and analysed, all the algorithms were used to divide the data into three different sets: a training set, a validation set, and a testing set. These sets were made up of 70%, 15%, and 15% of the total data, respectively. This division was carried out to ensure that the data was thoroughly evaluated and that all the models were properly trained and tested. The LM algorithm is a commonly used optimization method for training Backpropagation Neural Networks (BPNN). It is widely used in various fields and is considered a standard technique for non-linear least-squares problems. The LM algorithm can be perceived as a fusion of the steepest descent and Gauss-Newton techniques. It mimics the behavior of a steepest descent method when the current solution is considerably

distant from the correct solution. Conversely, it emulates the Gauss-Newton approach when the current solution is in proximity to the accurate solution. This algorithm has several advantages over other BPNN algorithms, including good convergence, training precision, and efficiency. Additionally, it is the fastest method for training moderate sized feedforward neural networks (Demuth & Beale, 1992; Hagan & Menhaj, 1994; Yue, 2010). Maximum validation failures were set to 6 to make it stop early as early stopping is a well-established technique in machine learning that can save computational resources and time while still achieving optimal performance (Orr & Müller, 1998). The training stopped when the validation error has increased more than maximum validation failures, at that time the number of iterations were 15 and the value of gradient was 1.37. In T The neural network's number of hidden layers was established as 19, based on the fact that it attained the highest degree of goodness of fit and the lowest calculation error. The goodness of fit reached the maximum and the calculation error was the least. The sigmoid function was used as the activation function of BPNN. The sigmoid function is a popular activation function utilized in neural networks because of its key attributes such as differentiability and smoothness, which play a crucial role in facilitating the back-propagation algorithm (Kros et al., 2006). The sigmoid function is shown in Eq.5.

$$s(x) = \frac{1}{1+e^{-x}} \tag{5}$$

The data was scaled using the scikit-learn library in Python for the machine-learning models RFR and KNN. Standard Scaler Standardize features by removing the mean and scaling to unit variance. The standard score of a sample *x* is calculated as:

$$z = \frac{(x-u)}{s} \tag{6}$$

where u is the mean of the training samples or zero. Centring and scaling happen independently on each feature by computing the relevant statistics on the samples in

the training set. In KNN, the weights were set to default as uniform showing all points were weighted equally in the neighborhood. In RFR, Bootstrap is set to True because it is a useful technique for improving performance. Bootstrapping can improve the performance of the model by reducing the variance and overfitting. and robustness of the model, and for making efficient use of the available data (Breiman, 2001; Geurts et al., 2006). To assess the inconsistency between the predicted value and the measured value for a particular sample, error, and error percentage were used as shown in equation Equ.7 and Equ.8, where y_i' is the predicted value and y_i is the actual value.

$$Error = y'_i - y_i \tag{7}$$

$$Error\ Percentage = \frac{y'_i - y_i}{y_i} \tag{8}$$

As the above equation was limited to just one simple sample, to evaluate the entire prediction results, several analytical approaches were used which led to the evaluation of the models. In this evaluation, the coefficient of determination (R^2), mean square error (*MSE*), root mean square error (*RMSE*) and mean absolute error (*MAE*). These parameters can compare the performance of various models, quantify the prediction accuracy of a single algorithm, and determine which model is best suited and most applicable to the given database. The expressions for these parameters are shown in equations Equ.9, 10, 11 and 12.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\overline{y} - yi)^2} \tag{9}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{10}$$

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{12}$$

3.2.2.1. HYPERPARAMETERS OF MACHINE LEARNING MODELS

Table I: Hyperparameters of machine learning models

KNN	N_Neighours	2
	Weights	Uniform
	Metric	Minkowski
BPNN	Max no of epox	1000
	Max validation failures	6
	Min performance gradient	1e -7
	Hidden layer numbers	19
	Algorithm	Levenberg-Marquardt
RFR	Number of trees	100
	Minimum samples split	2
	Minimum sample leaf	1
	Bootstrap	True

3.2.3. PLATOFRMS USED IN TRAINING MACHINE LEARNING MODELS

3.2.3.1. MATLAB

MATLAB is a proprietary multi-paradigm programming language and numeric computing environment developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other MATLAB is a closed-source programming language and numeric computation platform created by MathWorks. It supports multiple programming styles and offers tools for working with matrices, visualizing functions and data, executing algorithms, building user interfaces, and integrating with software written in different programming languages.

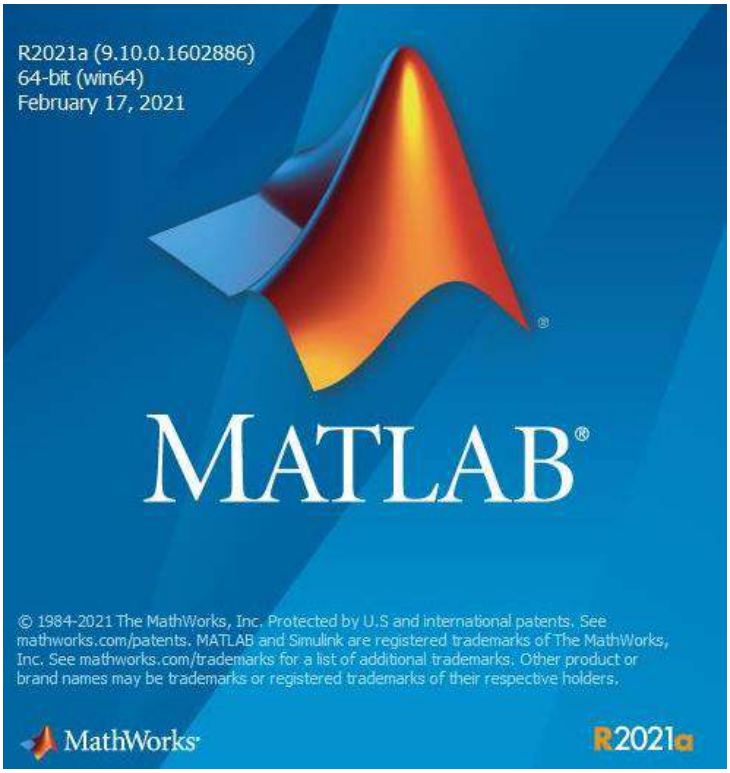


Fig 7: MATLAB LOADING INTERFACE

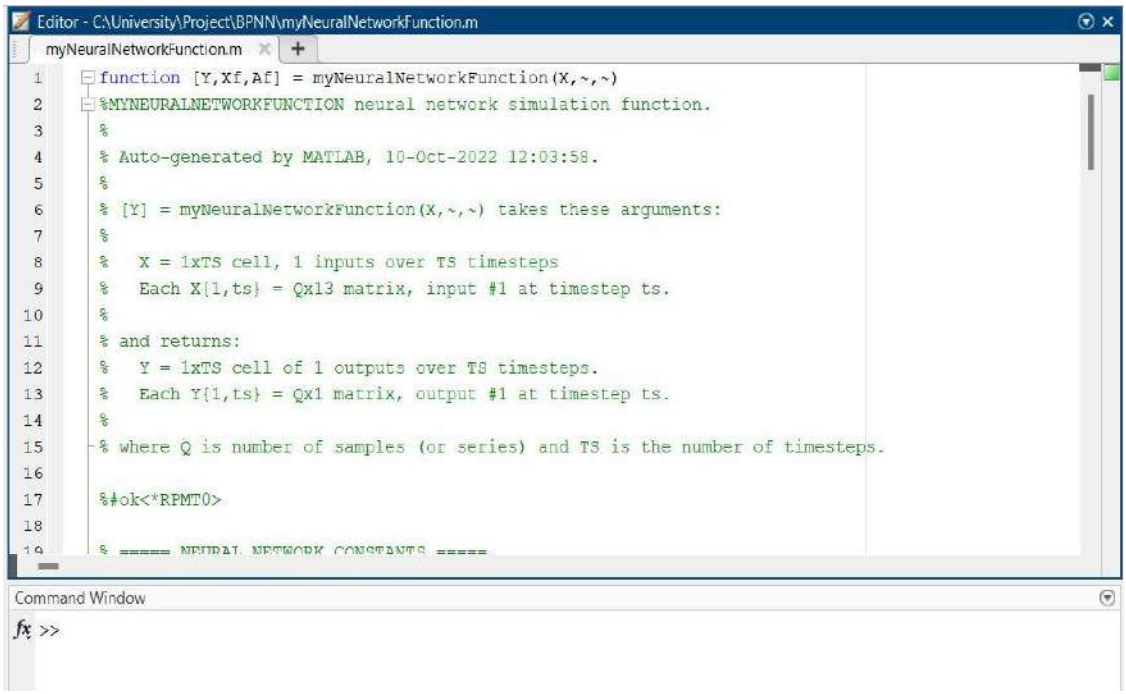


Fig 8: Interface

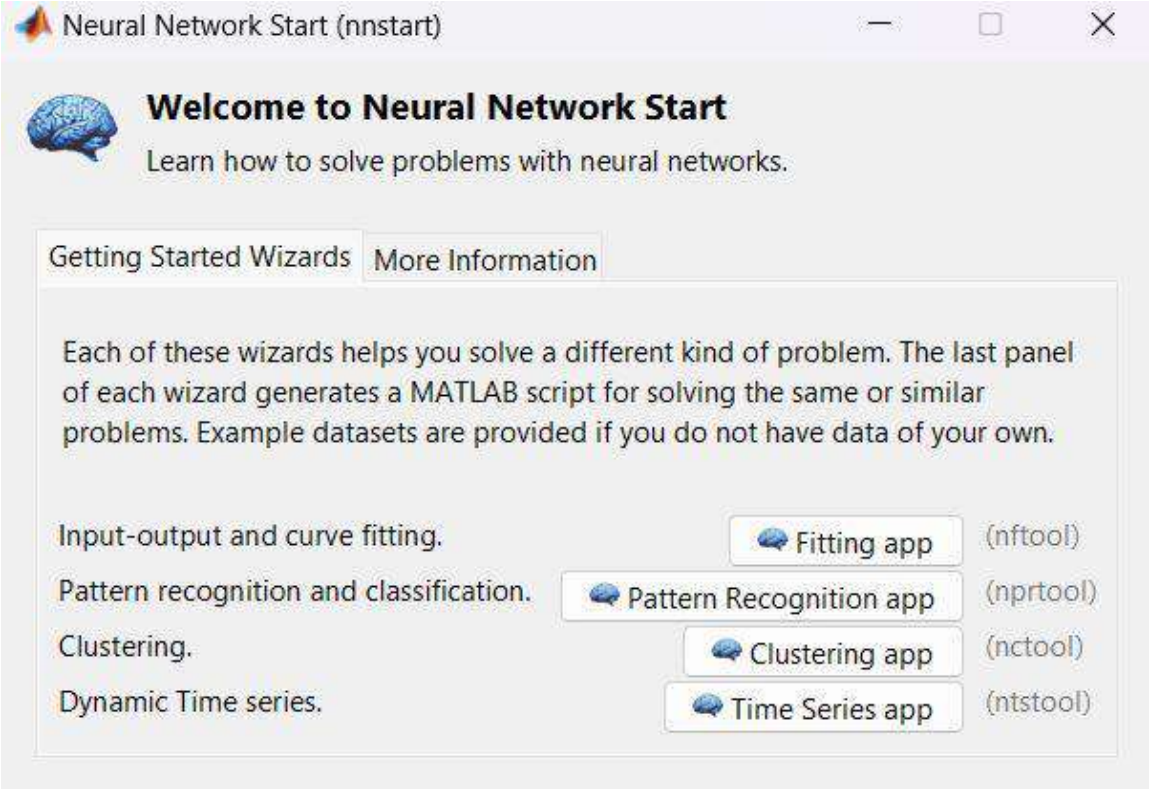


Fig 9: First Step after Interface

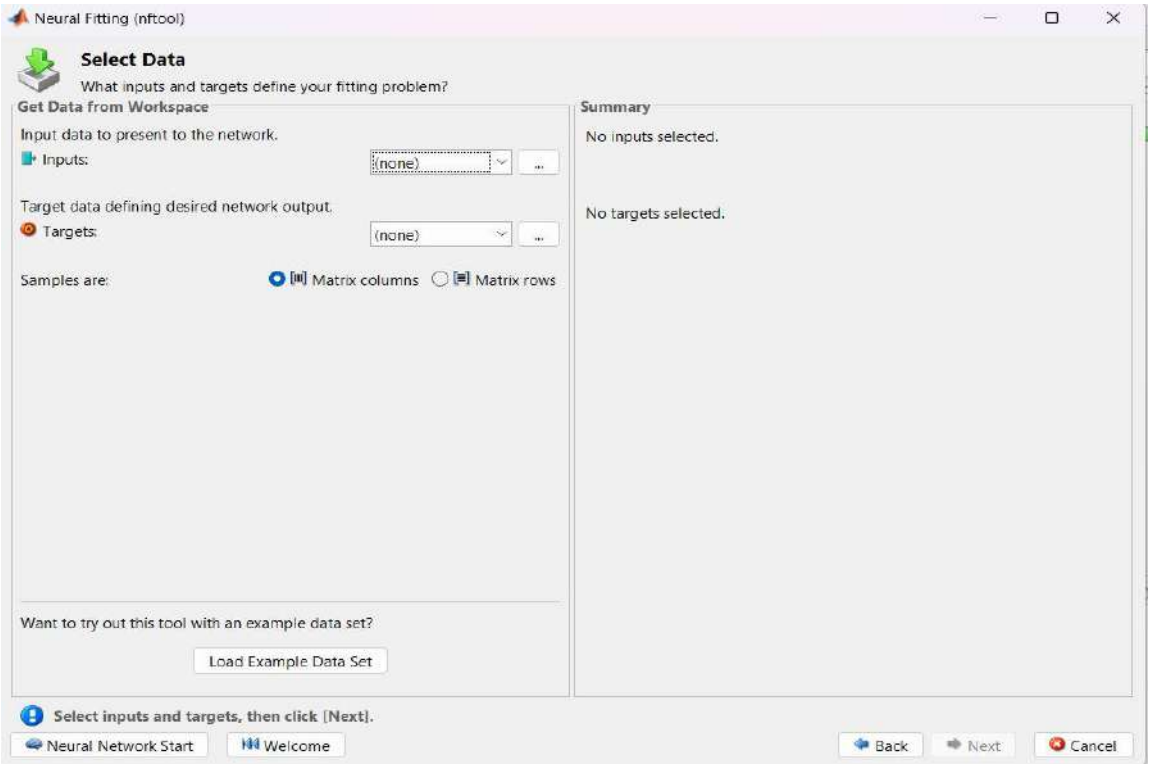


Fig 10: Selection of Inputs & Targets

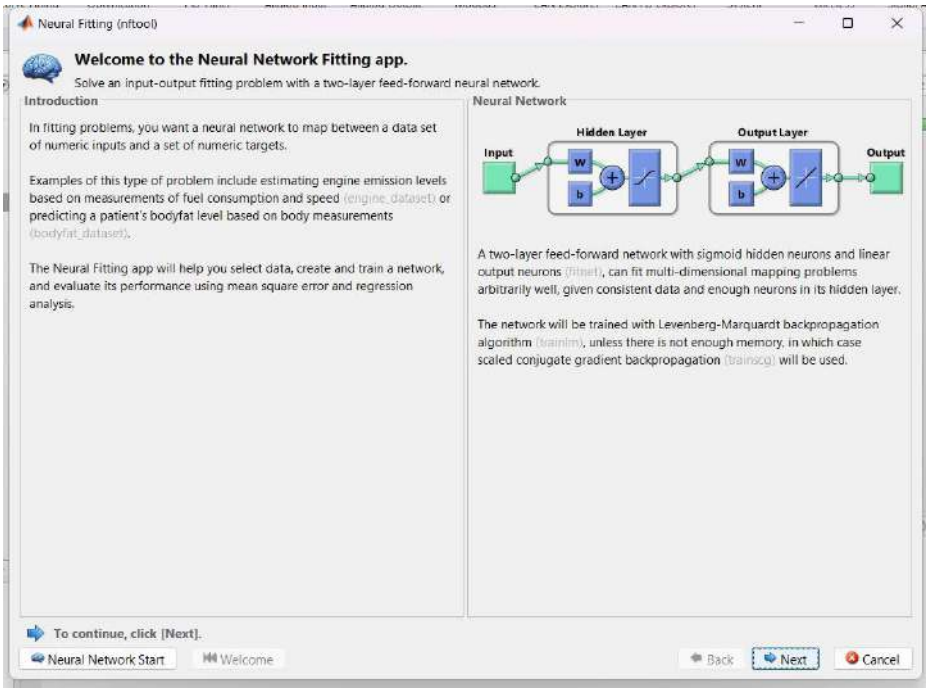


Fig 11: Neural Network Fitting Interface

3.2.3.2. VS CODE

Visual Studio Code, often called VS Code, is a code editor created by Microsoft using the Electron Framework. It is available on Windows, Linux, and macOS. It provides various functionalities such as debugging assistance, syntax highlighting, smart code suggestions, code shortcuts, code restructuring tools, and integrated Git support. Visual Studio Code (VS Code) is a versatile and widely-used code editor, especially favored by Python developers. It offers a powerful development environment that streamlines Python programming tasks. The editor's clean and user-friendly interface allows easy navigation and efficient project management. One of its standout features is the integrated terminal, which allows developers to run Python scripts and execute commands without leaving the editor. The code IntelliSense feature provides intelligent code completion, offering suggestions for Python functions, variable names, and available modules as you type, speeding up coding and reducing errors. Additionally, VS Code provides real-time linting and error checking, which helps identify syntax errors, style issues, and potential bugs in Python code. Its debugging support is another

significant advantage, enabling developers to set breakpoints, inspect variables, and step through code for effective troubleshooting. VS Code's extensibility is a major asset, boasting a vast collection of extensions, including many Python-specific ones, allowing developers to customize and enhance the editor's functionality according to their needs. Furthermore, with built-in Git integration, Python developers can easily manage version control and collaborate seamlessly. The editor also simplifies package management through an integrated terminal, facilitating the installation and management of Python packages. For those working with Jupyter Notebooks, VS Code provides native support, allowing Python developers to execute, edit, and visualize notebook files with ease.

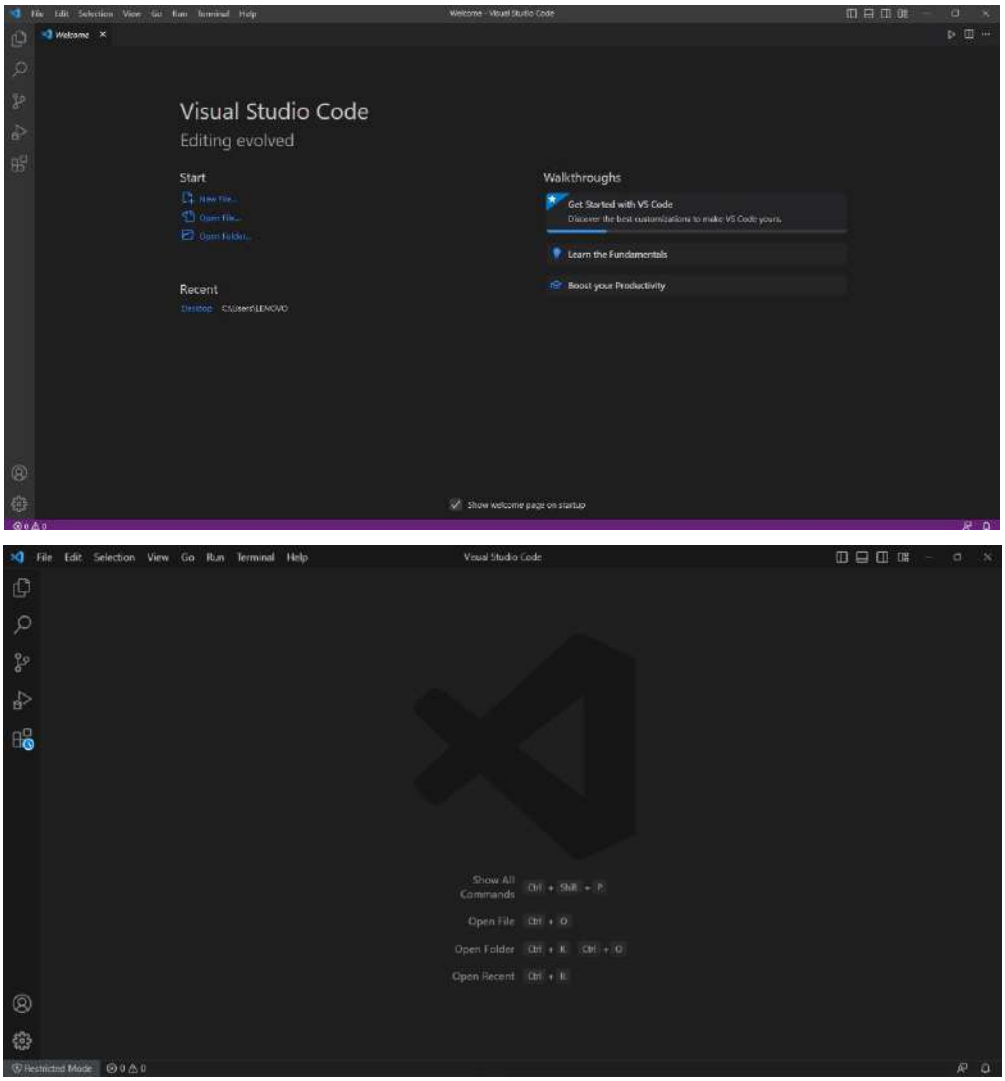


Fig 12: VS CODE Interface

3.3.DATA COLLECTION

The study collected data from 11 references (Aliabdo et al., 2016; Deb, 2013; Deb et al., 2014; Ghafoor et al., 2021a; Gunasekara et al., 2017, 2021; Joseph & Mathew, 2012; Kusbiantoro et al., 2012; Nath & Sarker, 2015; Sarker et al., 2013; Topark-Ngarm et al., 2015) with a total of 149 different mix proportions which consider the influence of Fly Ash, Coarse and Fine Aggregates, Alkaline Activators (NaOH & Na_2SiO_3), and Water is shown in Table *II*.

Table II: Proportion of geopolymer concrete using Fly Ash

FA (kg/m^3)	Coarse aggregate (kg/m^3)	Fine aggregate (kg/m^3)	NaOH solution (kg/m^3)	NaOH (M)	Na_2SiO_3 solution (kg/m^3)	Na_2Sio_3/ NaOH	AA/ FA	Water (kg/m^3)
400	1222	658	40	14	100	2.5	0.35	76.1

The molarity of the Sodium Hydroxide (NaOH) solution used was considered as one of the input parameters, with molarities ranging from 8 to 20. The mass of water was calculated for each mix proportion and added uniformly. The contents of Fly Ash were also considered as input parameters, specifically, the SiO_2 and Al_2O_3 content. The AA/FA ratio and molarity of the alkaline solution were also used as input parameters. Coarse and fine aggregates were also taken as input parameters, with the raw material of each mix proportion and its size range mentioned in Table *III*. However, the fineness modulus of the fine aggregate and size range were not used due to difficulty in quantification. The temperature, which plays an important role in hydration, was also considered. Out of 149 mixes, 70 had high-temperature pre-curing (up to 100°C) and the pre-curing duration varied for each mix. The remaining 79 mixes did not have any pre-curing. A total of 13 inputs were used in the analysis, as listed in Table *IV*. The impact of each input parameter on the output was considered, and every single input from Table *V* was used to determine the applicability of the prediction model.

Table III: Particle size characteristics of aggregates

Reference	Coarse aggregate		Fineness modulus		
	Raw material	Size range (mm)	Raw material	Size range (mm)	Fineness modulus
(Gunasekara et al., 2017)	Crushed basalt	7.0-10.0	River sand		3
(Deb, 2013)	Crushed granite	7.0-20.0	Natural river sand	-	2.67
(Ghafoor et al., 2021a)	-	9.5-12.5	River sand	-	2.35
(Nath & Sarker, 2015)	Crushed granite	7.0-20.0	Natural sand	-	2.64
(Sarker et al., 2013)	Crushed stone	7.0-10.0	River sand	-	-
(Gunasekara et al., 2021)	Crushed granite	10	River sand		2.8
(Aliabdo et al., 2016)	Crushed bluestone gravel	7	Graded sand	<0.4	-
(Deb et al., 2014)	Crushed basalt aggregate	7.0-10.0	Uncrushed river sand	-	3
(Topark-Ngarm et al., 2015)	Limestone	20	River sand	-	2.9
(Joseph & Mathew, 2012)	Crushed granite rock	20	Natural river sand	-	2.64

Table IV: Composition of FA

Reference		SiO_2	Al_2O_3	CaO	SO_3	Fe_2O_3	MgO	LOI
(Gunasekara et al., 2017)	Pt. Augusta	49.37	31.25	4.8	0.24	4.47	1.28	0.51
	Collie	53.82	29.95	1.03	0.34	9.24	0.58	0.63
	Tarong	75.66	19	0.3	0.03	1.38	0	1.16
(Deb, 2013)		53.71	27.2	1.9	0.3	11.17	-	0.68
(Ghafoor et al., 2021a)		71.5	9.2	6.72	2.4	2.37	0.6	3.67
(Nath & Sarker, 2015)		53.71	27.2	1.9	0.3	11.17	-	0.68
(Kusbiantoro et al., 2012)		51.7	29.1	8.84	1.5	4.76	-	-
(Sarker et al., 2013)		50.5	26.57	2.13	0.41	13.77	1.54	0.6
(Gunasekara et al., 2021)		38.7	20.8	26.6	2.1	5.3	1.5	0.1
(Aliabdo et al., 2016)		49	31	5	-	3	3	-
(Deb et al., 2014)	Type-I	47.87	28	3.81	0.27	14.09	0.93	0.43
	Type-II	49.37	31.25	4.8	0.24	4.47	1.28	0.51
	Type-III	53.82	29.95	1.03	0.34	9.24	0.58	0.63
(Topark-Ngarm et al., 2015)		45.23	19.95	15.51	-	13.15	-	-
(Joseph & Mathew, 2012)		59.7	28.36	2.1	0.4	4.57	0.83	1.06
LOI: Loss of ignition								

Table V: Statistics of input/output parameters

Notation	Input parameter	Mean	Standard deviation	Range/categories		
				Minimum	Maximum	Range
X_1	FA (kg/m^3)	443.99	92.70	254.50	600	345.50
X_2	SiO_2 (%)	50.99	14.57	36.20	75.66	39.46
X_3	Al_2O_3 (%)	18.23	6.60	9.20	31.25	22.05
X_4	Coarse aggregate (kg/m^3)	1105.97	231.54	554	1684	1130
X_5	Fine aggregate (kg/m^3)	594.33	45.56	500	706	206
X_6	NaOH (kg/m^3)	74.89	29.43	11.78	198	186.22
X_7	NaOH (M)	11.88	3.01	8	20	12
X_8	Na_2SiO_3 (kg/m^3)	165.18	67.31	29.51	342	312.46
X_9	Na_2Sio_3/NaOH	2.29	0.78	1	8.77	7.77
X_{10}	AA/FA	0.53	0.13	0.09	0.92	0.83
X_{11}	Water (kg/m^3)	143.24	44.41	37.45	206.78	169.33
X_{12}	Temperature (℃)	29.02	32.16	0	100	100
X_{13}	Duration (h)	13.69	17.26	0	72	72

In the collected data from various studies, the mixing methods for water and alkaline activators differed, with some combining them before mixing with the other components and others calculating the water content separately. To standardize the water content across all mix proportions, the total amount of water was determined by adding the mass of water in the NaOH and Na_2SiO_3 solutions and any additional water utilized. Additionally, the variations in the chemical composition of Fly Ash (FA) used in the different studies were also taken into account in the analysis.

The analysis of the collected data considered the role of Fly Ash (FA) in the mix and incorporated the SiO_2 and Al_2O_3 contents in FA as input parameters. This was in line with previous findings that reported the influence of Si and Al contents in FA on the strength development of FA-based mixes. The compressive strength of fly ash-based geopolymer mixes is influenced by several factors including the type and concentration of alkali solutions used, the Si/Al ratio in the fly ash, the calcium content, the curing conditions (such as temperature and duration), and the presence of any additional

additives. A higher Si/Al ratio leads to the formation of more -Si-O-Si- bonds, resulting in a stronger, fully condensed structural matrix of the geopolymer. This is because the -Si-O-Si- bonds are stronger than both the -Si-O-Al- and -Al-O-Al- bonds (Zhuang et al., 2016). In the preparation of fly ash-based geopolymer, the use of Na_2SiO_3 solution in conjunction with NaOH has been shown to enhance compressive strength. The utilization of Na_2SiO_3, due to its high viscosity, promotes the formation of geopolymer gels which results in a compact microstructure of the final fly ash-based geopolymer. Additionally, the method used for activating the fly ash-based geopolymer also affects its compressive strength (Criado et al., 2005). Moreover, the curing duration and temperature play a significant role in determining the compressive strength of fly ash-based geopolymer. Prolonged curing, lasting from 6 hours to 28 days, results in fly ash-based geopolymer with greater compressive strength.

Elevated curing temperatures facilitate the removal of excess water from the fresh geopolymer, leading to the collapse of capillary pores and a denser structure, thus resulting in higher compressive strength (Leung & Pheeraphan, 1995). The higher NaOH concentration results in finer pore structures due to the polycondensation reaction, which leads to reduced chloride penetration. Additionally, geopolymer activated with NaOH has a more stable cross-linked structure, providing improved resistance to sulphates (Zhuang et al., 2016).

In this study, a range of input parameters related to concrete mixtures in civil engineering applications was analysed using statistical analysis. The parameters, such as Fine Aggregate, SiO_2 content, Al_2O_3 content, Coarse Aggregate, Fine Aggregate, NaOH, Na_2SiO_3, Na_2SiO_3/NaOH ratio, AA/FA ratio, water content, temperature, and duration, were examined for their mean, standard deviation, minimum, maximum, and range values. Table *V* shows range of input/output parameters.

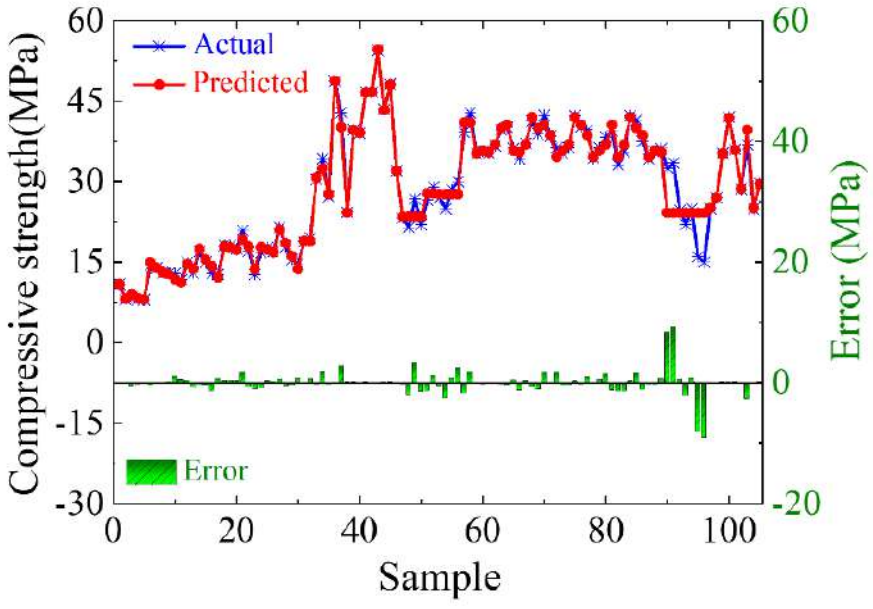
Table VI: Data set

Sr #	**FA**	**SiO2**	**Al2O3**	**Coarse aggregate**	**Fine aggregate**	**NaOH**	**NaOH (M)**	**Na2SiO3**	**Na2SiO3/NaOH**	**AA/FA**	**Water**	**Temperature**	**Duration**	***fcu***
-	kg/m3	kg/m3	kg/m3	kg/m3	kg/m3	kg/m3		kg/m3	-	-	kg/m3	°C	Days	Mpa
1	394	71.5	9.2	1294	554	63	8	94.6	1.5	0.4	117.4	0	0	10.9
2	394	71.5	9.2	1294	554	52.5	8	105.1	2	0.4	115.3	0	0	9.2
3	394	71.5	9.2	1294	554	45	8	112.6	2.5	0.4	113.8	0	0	8
4	368	71.5	9.2	1294	554	73.6	8	110.4	1.5	0.5	132.48	0	0	7.6
5	368	71.5	9.2	1294	554	61.3	8	122.6	2	0.5	129.97	0	0	8.5
6	368	71.5	9.2	1294	554	52.5	8	131.4	2.5	0.5	128.21	0	0	8.5
7	345	71.5	9.2	1294	554	82.8	8	124.2	1.5	0.6	145.59	0	0	7.6
8	345	71.5	9.2	1294	554	69	8	138	2	0.6	142.83	0	0	8
9	345	71.5	9.2	1294	554	59.1	8	147.8	2.5	0.6	140.8	0	0	8
10	394	71.5	9.2	1294	554	63	10	94.6	1.5	0.4	114.25	0	0	14.4
11	394	71.5	9.2	1294	554	52.5	10	105.1	2	0.4	112.68	0	0	13.9
12	394	71.5	9.2	1294	554	45	10	112.6	2.5	0.4	111.55	0	0	13.1
13	368	71.5	9.2	1294	554	73.6	10	110.4	1.5	0.5	128.8	0	0	14.2
14	368	71.5	9.2	1294	554	61.3	10	122.6	2	0.5	126.9	0	0	12.9
15	368	71.5	9.2	1294	554	52.5	10	131.4	2.5	0.5	125.58	0	0	12.9
16	345	71.5	9.2	1294	554	82.8	10	124.2	1.5	0.6	141.45	0	0	13
17	345	71.5	9.2	1294	554	69	10	138	2	0.6	139.38	0	0	11.7
18	345	71.5	9.2	1294	554	59.1	10	147.8	2.5	0.6	137.84	0	0	11.8
19	394	71.5	9.2	1294	554	63	12	94.6	1.5	0.4	111.1	0	0	15
20	394	71.5	9.2	1294	554	52.5	12	105.1	2	0.4	110.05	0	0	14.5
21	394	71.5	9.2	1294	554	45	12	112.6	2.5	0.4	109.3	0	0	13
22	368	71.5	9.2	1294	554	73.6	12	110.4	1.5	0.5	125.12	0	0	17
23	368	71.5	9.2	1294	554	61.3	12	122.6	2	0.5	123.84	0	0	15
24	368	71.5	9.2	1294	554	52.5	12	131.4	2.5	0.5	122.96	0	0	12.9
25	345	71.5	9.2	1294	554	82.8	12	124.2	1.5	0.6	137.31	0	0	12.5
26	345	71.5	9.2	1294	554	69	12	138	2	0.6	135.93	0	0	13
27	345	71.5	9.2	1294	554	59.1	12	147.8	2.5	0.6	134.89	0	0	12.7
28	394	71.5	9.2	1294	554	63	14	94.6	1.5	0.4	108.58	0	0	18.1
29	394	71.5	9.2	1294	554	52.5	14	105.1	2	0.4	107.95	0	0	18
30	394	71.5	9.2	1294	554	45	14	112.6	2.5	0.4	107.5	0	0	17.5
31	368	71.5	9.2	1294	554	73.6	14	110.4	1.5	0.5	122.18	0	0	18.7
32	368	71.5	9.2	1294	554	61.3	14	122.6	2	0.5	121.38	0	0	20.9
33	368	71.5	9.2	1294	554	52.5	14	131.4	2.5	0.5	120.86	0	0	17.1
34	345	71.5	9.2	1294	554	82.8	14	124.2	1.5	0.6	134	0	0	21.5
35	345	71.5	9.2	1294	554	69	14	138	2	0.6	133.17	0	0	16
36	345	71.5	9.2	1294	554	59.1	14	147.8	2.5	0.6	132.52	0	0	12.7
37	394	71.5	9.2	1294	554	63	16	94.6	1.5	0.4	106.06	0	0	16.8
38	394	71.5	9.2	1294	554	52.5	16	105.1	2	0.4	105.85	0	0	17.6
39	394	71.5	9.2	1294	554	45	16	112.6	2.5	0.4	105.7	0	0	17
40	368	71.5	9.2	1294	554	73.6	16	110.4	1.5	0.5	119.23	0	0	21.5
41	368	71.5	9.2	1294	554	61.3	16	122.6	2	0.5	118.93	0	0	17.9
42	368	71.5	9.2	1294	554	52.5	16	131.4	2.5	0.5	118.76	0	0	15.8
43	345	71.5	9.2	1294	554	82.8	16	124.2	1.5	0.6	130.69	0	0	17
44	345	71.5	9.2	1294	554	69	16	138	2	0.6	130.41	0	0	15.5
45	345	71.5	9.2	1294	554	59.1	16	147.8	2.5	0.6	130.16	0	0	14.6
46	400	49	31	1280	547	19.8	14	49.6	2.5	0.17	130.6	0	0	18.8
47	400	49	31	1280	547	19.8	14	49.6	2.5	0.17	130.6	0	0	19.5
48	475	49	31	1253	539	11.781	14	29.512	2.5	0.09	77.71	0	0	30.4
49	475	49	31	1248	535	13.86	14	34.72	2.5	0.1	91.42	0	0	34.3
50	475	49	31	1235	529	16.434	14	41.168	2.5	0.12	108.4	0	0	27.2
51	416	47.87	28	927	699	65	15	292	4.5	0.86	197.23	80	24	48.7
52	416	49.37	31.25	927	699	65	15	292	4.5	0.86	197.23	80	24	42.9
53	420	53.82	29.95	936	706	92	15	241	2.6	0.79	186.52	80	24	24.3
54	414	45.23	19.95	1091	588	104	10	104	1	0.5	116.27	0	0	39.67
55	414	45.23	19.95	1091	588	104	15	104	1	0.5	95.47	0	0	45.34
56	414	45.23	19.95	1091	588	104	20	104	1	0.5	74.67	0	0	37.64
57	414	45.23	19.95	1091	588	69	10	138	2	0.5	112.88	0	0	33.8
58	414	45.23	19.95	1091	588	69	15	138	2	0.5	99.08	0	0	39.02
59	414	45.23	19.95	1091	588	69	20	138	2	0.5	85.28	0	0	46.69
60	414	45.23	19.95	1091	588	104	10	104	1	0.5	116.27	60	24	46.67
61	414	45.23	19.95	1091	588	104	15	104	1	0.5	95.47	60	24	54.4
62	414	45.23	19.95	1091	588	104	20	104	1	0.5	74.67	60	24	43.42
63	414	45.23	19.95	1091	588	69	10	138	2	0.5	112.88	60	24	40.09
64	414	45.23	19.95	1091	588	69	15	138	2	0.5	99.08	60	24	48.18
65	414	45.23	19.95	1091	588	69	20	138	2	0.5	85.28	60	24	49.5
66	408	50.5	26.57	554	647	62	14	93	1.5	0.379902	98.612	60	24	32
67	300	38.7	20.8	1684	681	51.4	14	129	2.50973	0.601333	94.727	60	24	25.8
68	300	38.7	20.8	1684	681	51.4	14	129	2.50973	0.601333	94.727	60	24	23.2
69	300	38.7	20.8	1684	681	51.4	14	129	2.50973	0.601333	94.727	60	24	21.5
70	300	38.7	20.8	1684	681	51.4	14	129	2.50973	0.601333	94.727	60	24	26.8
71	300	38.7	20.8	1684	681	51.4	14	129	2.50973	0.601333	94.727	60	24	20.5
72	300	38.7	20.8	1684	681	51.4	14	129	2.50973	0.601333	94.727	60	24	22
73	600	38.7	20.8	1087	572	89.1	14	223	2.50281	0.520167	163.861	60	24	26.5
74	600	38.7	20.8	1087	572	89.1	14	223	2.50281	0.520167	163.861	60	24	29
75	600	38.7	20.8	1087	572	89.1	8	223	2.50281	0.520167	185.245	60	24	27

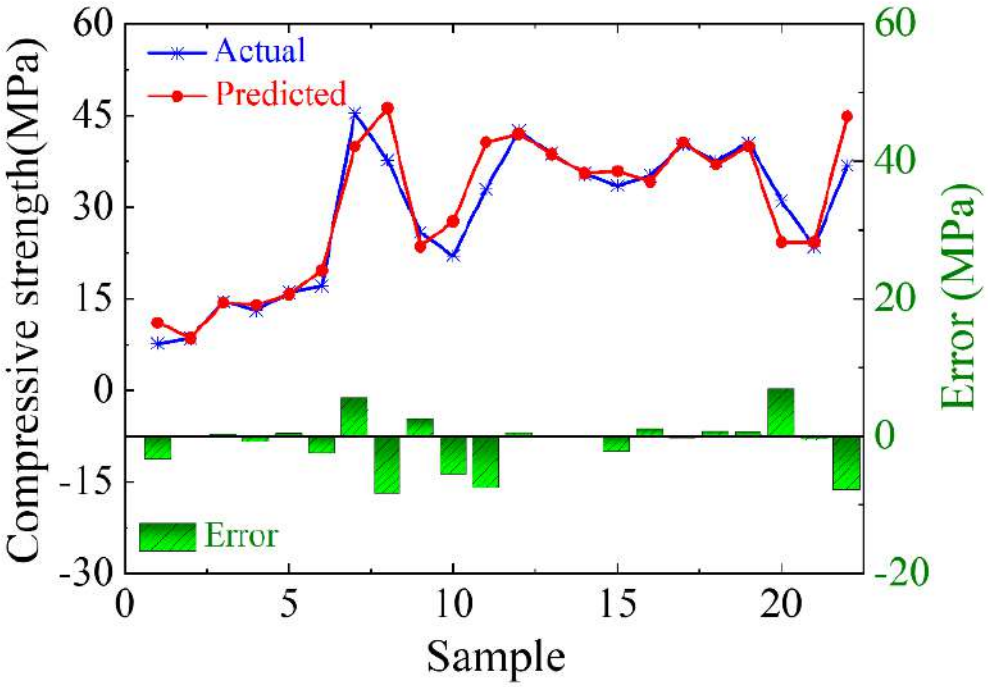
76	600	38.7	20.8	1087	572	89.1	8	223	2.50281	0.520167	185.245	60	24	25
77	600	38.7	20.8	1087	572	89.1	8	223	2.50281	0.520167	185.245	60	24	22.5
78	600	38.7	20.8	1087	572	89.1	8	223	2.50281	0.520167	185.245	60	24	28.5
79	600	38.7	20.8	1087	572	89.1	8	223	2.50281	0.520167	185.245	60	24	22
80	600	38.7	20.8	1087	572	89.1	8	223	2.50281	0.520167	185.245	60	24	30
81	494	38.7	20.8	858	691	198	14	198	1	0.801619	196.02	60	72	39.2
82	494	38.7	20.8	858	691	198	14	198	1	0.801619	196.02	60	72	36.5
83	494	38.7	20.8	858	691	198	14	198	1	0.801619	196.02	60	72	42.8
84	450	36.2	19.9	1150	500	108	12	162	1.5	0.6	149.148	60	48	35.2
85	450	36.2	19.9	1036	500	108	12	162	1.5	0.6	149.148	60	48	32.9
86	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	60	24	33.2
87	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	0	0	35.6
88	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	60	24	35.4
89	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	0	0	36.7
90	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	60	24	42.4
91	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	0	0	39.7
92	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	60	24	40.1
93	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	0	0	38.7
94	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	60	24	34.7
95	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	0	0	36.2
96	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	60	24	34.3
97	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	0	0	37.1
98	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	60	24	41.3
99	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	0	0	38.9
100	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	60	24	42.3
101	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	0	0	38.5
102	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	60	24	36.3
103	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	0	0	35.3
104	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	60	24	36.1
105	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	0	0	36.3
106	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	60	24	42.2
107	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	60	24	40.2
108	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	0	0	39.6
109	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	60	24	34.4
110	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	0	0	36.3
111	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	60	24	35.4
112	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	0	0	38.3
113	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	60	24	39.4
114	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	0	0	38.3
115	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	60	24	33.1
116	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	0	0	33.5
117	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	0	24	35.1
118	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	0	0	35.5
119	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	60	24	42.2
120	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	0	0	41.5
121	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	60	24	40.2
122	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	0	0	37.5
123	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	60	24	34.2
124	550	38.7	20.8	838	600	95	8	239	2.51579	0.607273	203.22	0	0	35.5
125	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	60	24	36.2
126	550	38.7	20.8	838	600	95	10	239	2.51579	0.607273	195.62	0	0	37.5
127	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	60	24	41.4
128	550	38.7	20.8	838	600	95	12	239	2.51579	0.607273	188.02	0	0	40.5
129	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	60	24	40.8
130	550	38.7	20.8	838	600	95	14	239	2.51579	0.607273	180.42	0	0	38.4
131	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	32.5
132	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	33.5
133	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	31
134	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	24.7
135	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	22
136	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	25
137	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	23.5
138	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	16
139	350	38.7	20.8	1250	650	41	8	103	2.5122	0.411429	85.045	60	24	15
140	400	53.71	27.2	1222	658	40	14	100	2.5	0.35	76.1	21.5	72	25
141	400	53.71	27.2	1222	658	56	14	84	1.5	0.35	74.76	21.5	72	27
142	365.2	59.7	28.36	1118	602	34.3	10	73	2.12828	0.293812	56.35	100	24	35.3
143	254.5	59.7	28.36	1290	694.7	22.8	10	48.5	2.12719	0.280157	37.445	100	24	36.8
144	309.9	59.7	28.36	1204	648.4	27.7	10	59	2.12996	0.279768	45.53	100	24	42
145	408	51.7	29.1	1294	554	41	14	103	2.5122	0.352941	121.0086	60	24	36
146	428.6	53.71	27.2	1177	623	68.6	14	102.9	1.5	0.40014	98.7154	21.5	72	28.6
147	416	49.37	31.25	927	699	65	15	292	4.49231	0.858173	197.228	80	24	36.9
148	420	53.82	29.95	936	706	92	15	241	2.61957	0.792857	171.919	80	24	24.9
149	412	75.66	19	918	693	39	15	342	8.76923	0.924757	206.778	80	24	29.6

4. RESULTS AND DISCUSSION

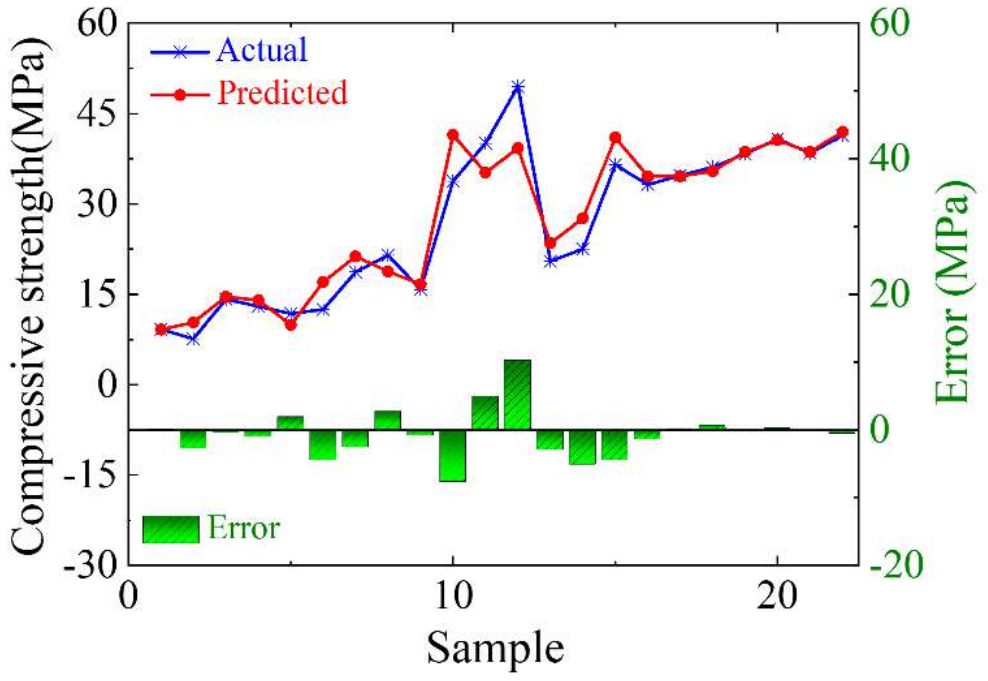
In this chapter the outcomes and results of the whole research are discussed such as machine learning predictions of 28 days compressive strength of fly ash-based geopolymer concrete mixtures using three algorithms BPNN, RFR, and KNN as shown in Fig.13-15. Red points indicate predicted values, whereas blue points reflect measured values. The gray histograms beneath each graphic show the difference between predicted and measured values. The models agree well with the training data, and their predictions for 28-day compressive strength are correct, with errors of less than 10 MPa. Overall, all three models worked admirably, precisely matching the measured compressive strength. Among them, BPNN demonstrated the highest level of accuracy.



(a) Training set



(b) Validation set



(c) Testing set

Fig 13: Actual and predicted compressive strength by BPNN

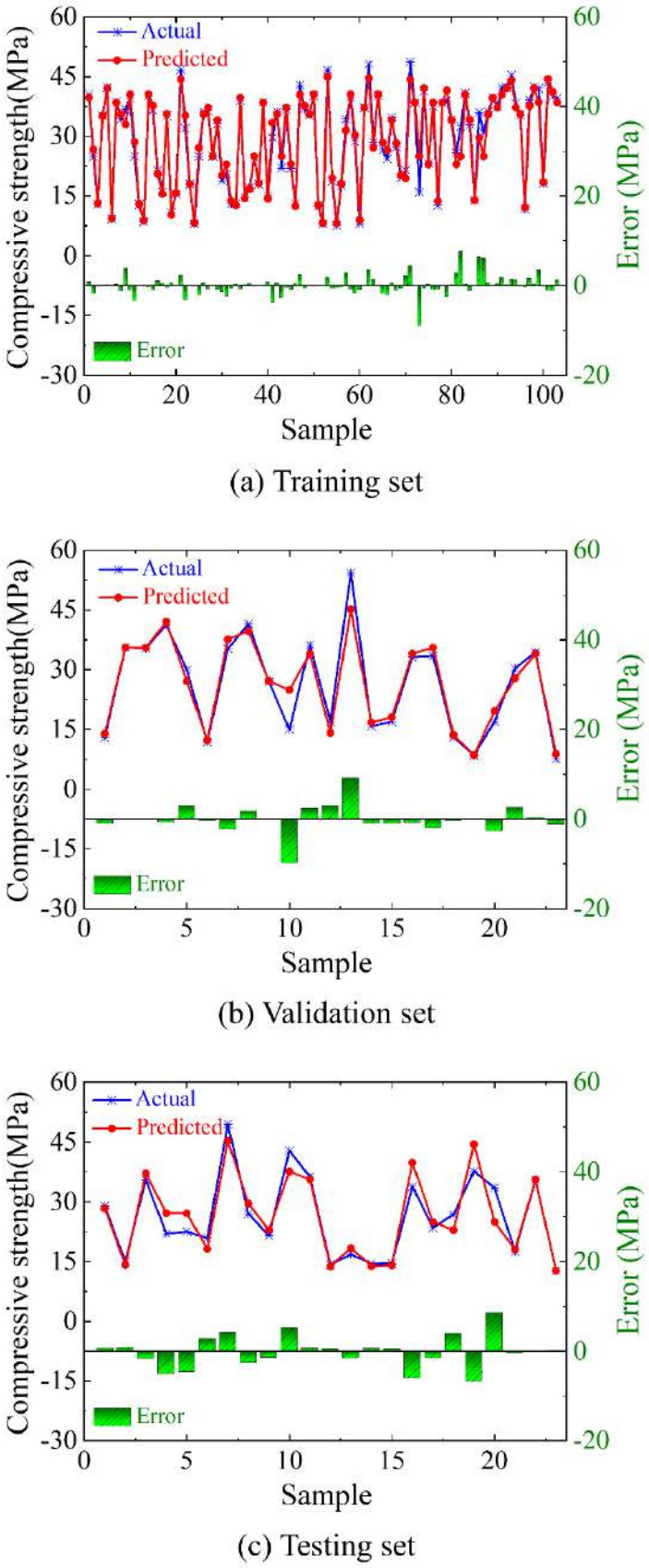


Fig 14: Actual and predicted compressive strength by RFR

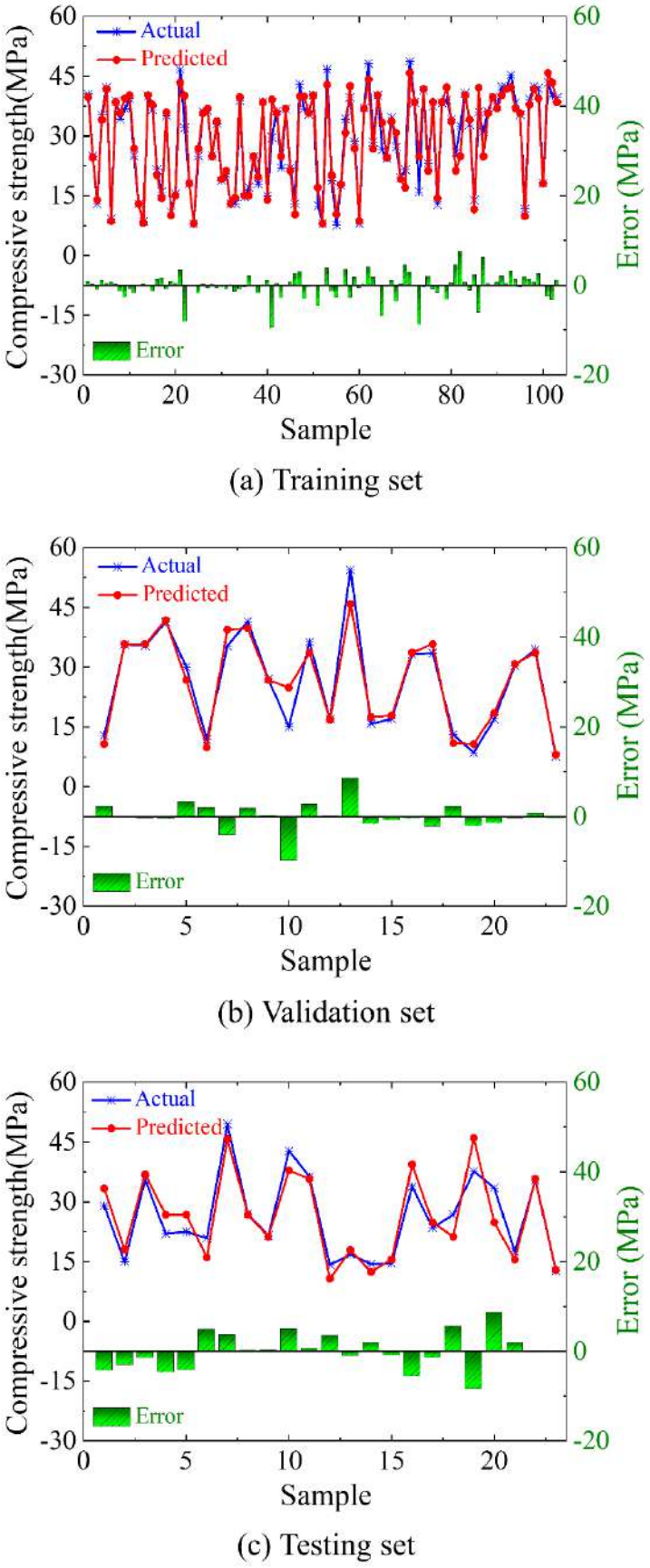


Fig 15: Actual and predicted compressive strength by KNN

4.1.COMPARISON OF NEURAL NETWORK MODELS

Table *VII* present the evaluation results for three models: BPNN, RFR, and KNN, indicating similar performance trends. On average, T BPNN exhibited the highest accuracy, with an impressive R^2 score of above 94% for the entire dataset. All three algorithms showed remarkable accuracy, with results in training, validation, and testing sets closely aligned. However, when comparing R^2 values, BPNN had slightly lower validation R^2 value of 0.897 compared to RFR R^2 value of 0.932 and KNN 0.933. Yet, BPNN performed better in testing R^2 value of 0.9174 compared to KNN 0.852 and RFR 0.878. Overall, the models showed an efficiency of over 90%.

Table VII: Accuracy comparison of three machine learning algorithms

	BPNN				RFR				KNN			
	Training set	Validation set	Testing set	All	Training set	Validation set	Testing set	All	Training set	Validation set	Testing set	All
R^2	0.970	0.897	0.917	0.948	0.971	0.932	0.878	0.927	0.947	0.933	0.852	0.911
MSE	3.90	15.27	13.36	10.84	3.87	10.45	12.31	8.88	6.97	10.23	14.92	10.71
RMSE	1.97	3.91	3.65	3.29	1.97	3.20	3.47	2.95	2.65	3.22	3.97	3.33
MAE	1.01	2.60	2.65	2.09	1.24	1.95	2.61	1.93	1.72	2.05	2.96	2.25

The superiority of BPNN over KNN and RFR, despite lower validation results, can be attributed to BPNN's capability to capture complex patterns in the data better than other models. Neural networks possess unique characteristics such as fault tolerance, non-linearity, self-learning, self-organization, and self-adaptation, which give BPNN an advantage in handling complex relationships.

BPNN is an optimization algorithm specifically designed to minimize the sum of squared errors, making it well-suited for modelling complex, nonlinear relationships between input features and output values. This flexibility allows it to capture the intricacies of the data better than other models like RFR, which rely on predefined

decision rules. As a result, BPNN can often find a better fit to the data and deliver improved predictive accuracy.

Fig.16 illustrates the results of predicted and experimental compressive strength obtained from three models, BPNN, RFR, and KNN. The data points generated by these algorithms generally clustered around the centre line and fell within an error range of 10%. Notably, most of the errors between actual and predicted values were below 10% across all the models. The close alignment of data points around the centre line signifies a high level of accuracy in the predictions made by BPNN, RFR, and KNN for the compressive strength of the geopolymer concrete.

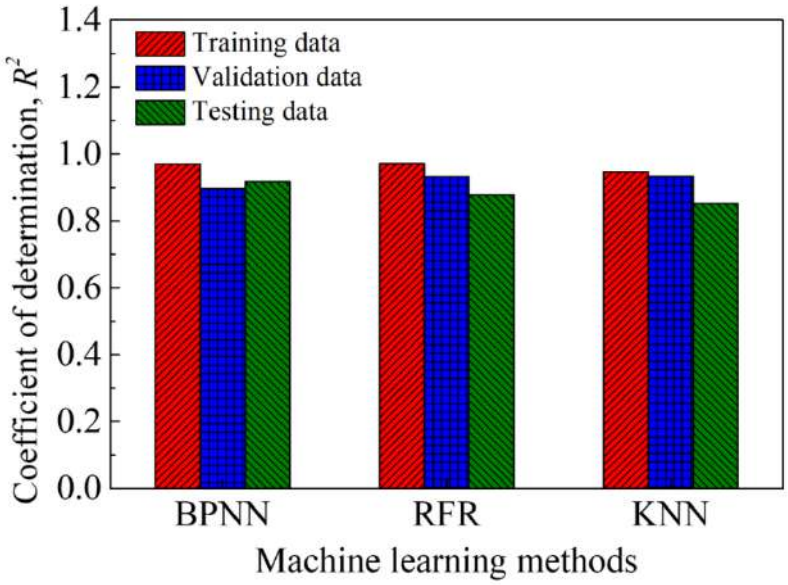
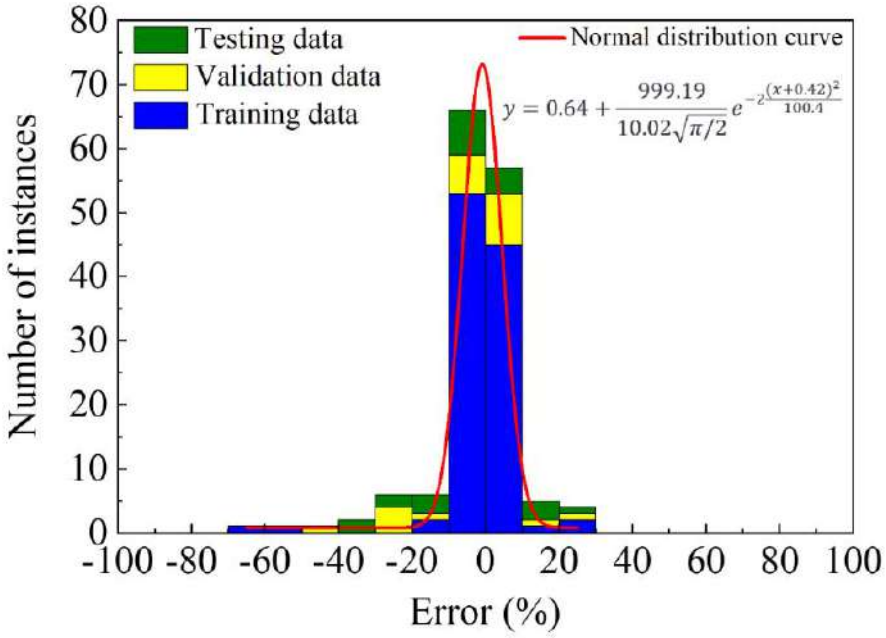


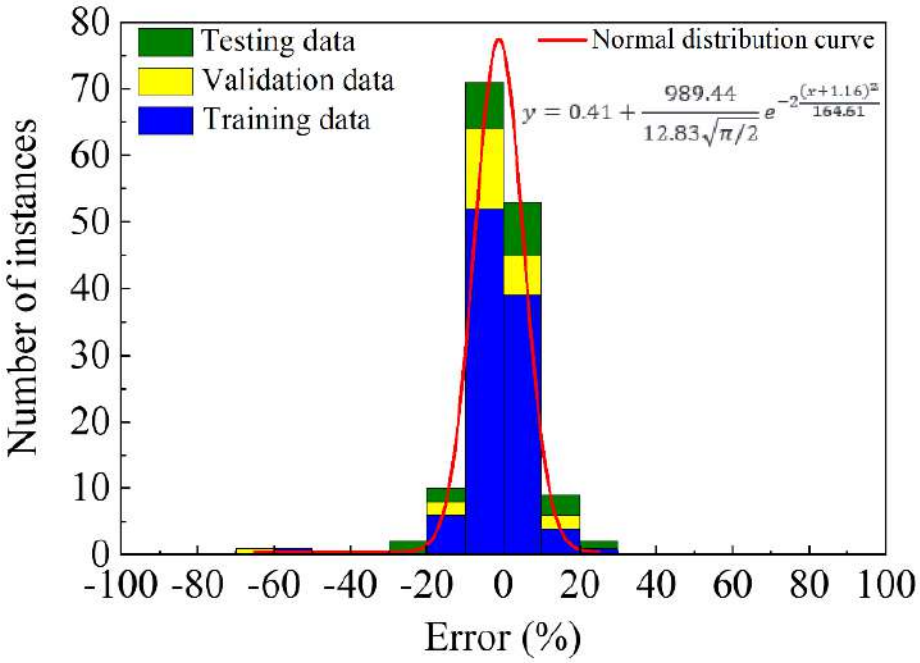
Fig 16: Comparison of coefficient of determination (R^2)

Fig.17 displays the error percentages and their frequency for different machine learning algorithms. To ensure precise measurements, we used error percentage instead of absolute error, and then fit the data to a normal distribution. This allowed us to eliminate the influence of sample size on deviation. The results showed that all three algorithms had an average error percentage close to zero. Among them, RFR had the steepest fitting curve, indicating a more concentrated distribution within the -20% to +20% error range.

However, BPNN produced predictions that were closer to the actual results, even though it had fewer samples in the -20% to +20% error range than RFR. These findings reinforce our earlier conclusion that BPNN showed superior prediction accuracy as illustrated in Fig.18.



(a) BPNN



(b) RFR

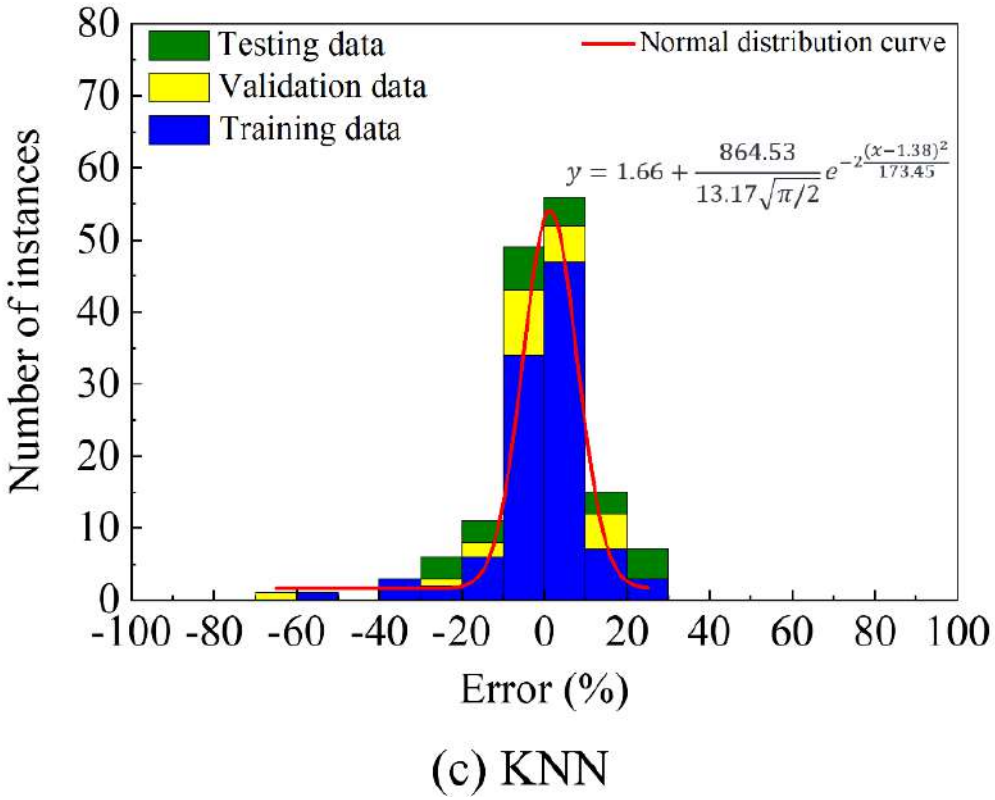
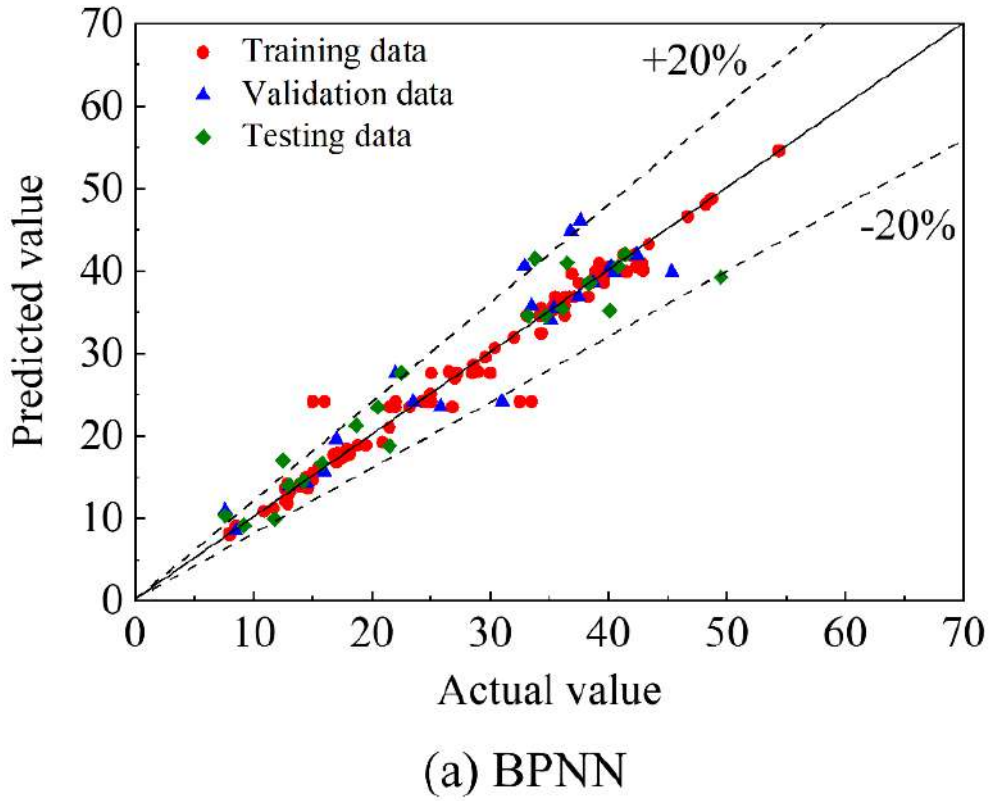
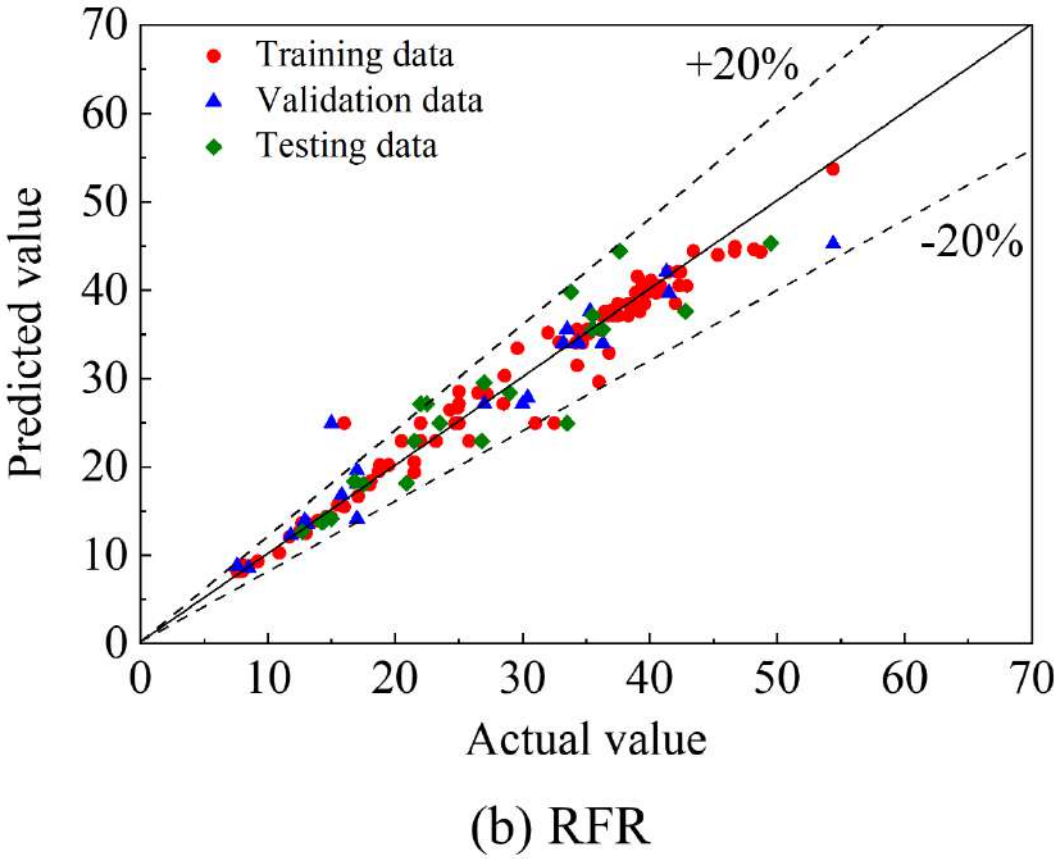
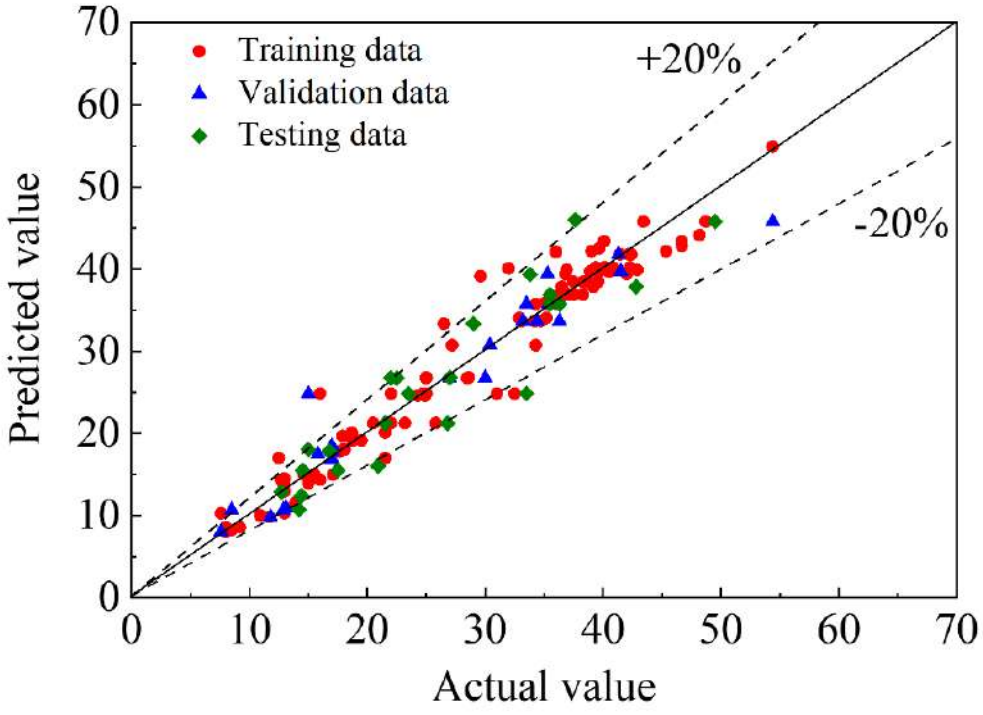


Fig 17: Normal distribution fitting and error percentage distribution of models





(b) RFR



(c) KNN

Fig 18: Comparison of actual and predicted values of compressive strength

4.2.EFFECTS OF INPUT PARAMETERS

The results of our study on the impact of various mix design parameters on the compressive strength of FA-based geopolymer concrete are presented in Fig.19. The findings showed that there was no significant correlation between any of the variables and compressive strength. This result may have been due to the use of a large number of variables with widely varying ranges in our study. This reduction of the impact of any one factor on the outcomes is a common challenge in multi-variable regression analysis. However, by examining the correlation between a few independent variables and several dependent variables, some potential trends might still be found. The results showed that an ideal $SiO_2/Al2O_3$ ratio was revealed by the findings of the SiO_2 and Al_2O_3 contents, which improved the contribution of FA to compressive strength. This highlights the importance of considering the ratio of these two components in the mix design of FA-based geopolymer concrete. Further research on a smaller set of variables could lead to more definitive conclusions about the impact of specific parameters on compressive strength. The compressive strength of FA-based geopolymer concrete was found to be impacted by various factors such as the amount of NaOH and Na_2SiO_3 as well as their molar concentrations. Results showed that an increase in the molar concentration of NaOH solution increased the compressive strength, particularly when the concentration exceeded 12 M. Conversely, a rise in the water content above 100 kg/m^3 led to a decrease in strength.

The proportions of coarse or fine particles, on the other hand, did not appear to have a significant effect on strength. This aligns with previous studies that have revealed that the aggregates' morphology, particle size distribution, and interface transition zone have a greater impact on strength growth compared to their contents. The effect of pre-curing

on compressive strength is detailed in Table *VIII*, which also shows the effect on the minimum, average, and maximum strength values across all data sets.

Table VIII: Impact of pre-curing on the compressive strength

Curing condition	Compressive strength (MPa)		
	Minimum	Maximum	Average
No pre-curing	7.60	46.69	23.63
Pre-curing (up to 100℃ and 72 h)	15	54.40	33.55

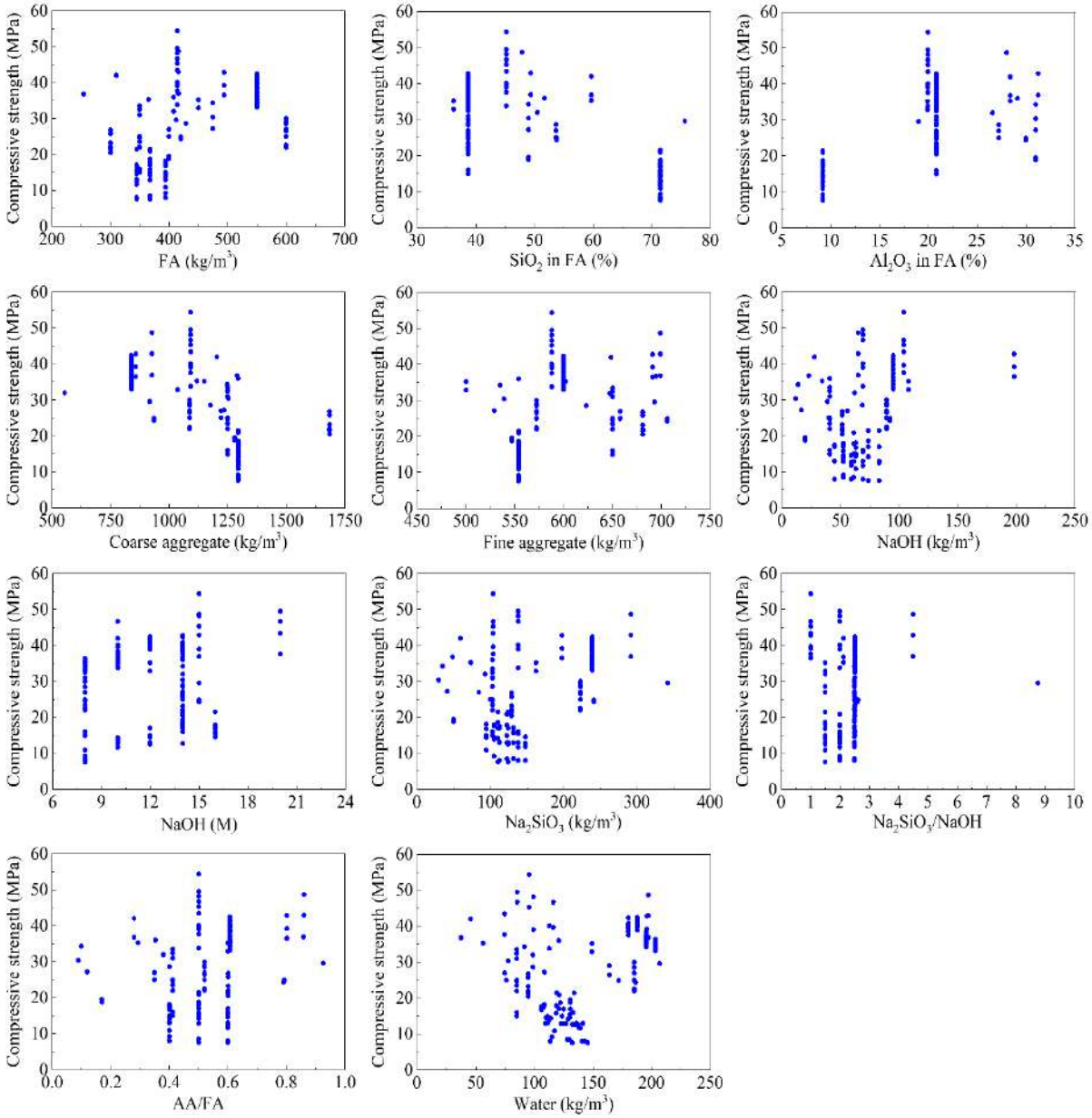


Fig 19: Correlation of different input parameters with Compressive strength

To comprehend the significance of each input parameter concerning prediction outcomes, evaluating the importance of individual parameters is crucial. This is often done through the utilization of permutation feature importance (PFI), a widely used metric for assessing the influence of various features on outcomes. PFI operates on the principle that when an input variable, let's call it *Xi*, significantly affects the outcome, shuffling *Xi* while maintaining the order of other variables will notably decrease predictive performance. This concept is expressed in Equation 13, which calculates the PFI value for a specific variable using mean absolute error (MAE) as the error measurement. Here, MAE_{perm} represents the mean absolute error after randomly rearranging *Xi*, while MAE_{orig} represents the original mean absolute error.

$$PFI=MAE_{perm} - MAE_{orig} \quad (13)$$

By applying this approach, a PFI value approaching zero indicates that modifying a particular feature has minimal impact on the output. Conversely, a substantial PFI value suggests that the feature significantly influences the output. In a recent study, this PFI method was employed to assess the impact of 13 input variables on the compressive strength of geopolymer concrete as shown in Table *IX*.

Table IX: PFI results for different input parameters

Notation	Input parameter	MAE_{perm}	*PFI*
X_1	FA (kg/m^3)	1.77	0.78
X_2	SiO_2 (%)	3.24	2.25
X_3	Al_2O_3 (%)	2.32	1.33
X_4	Coarse Aggregate (kg/m^3)	6.75	5.76
X_5	Fine Aggregate (kg/m^3)	2.52	1.53
X_6	NaOH (kg/m^3)	1.80	0.81
X_7	NaOH (M)	2.94	1.95
X_8	Na_2SiO_3 (kg/m^3)	1.39	0.40
X_9	Na_2Sio_3/NaOH	1.38	0.39
X_{10}	AA/FA	1.24	0.25
X_{11}	Water (kg/m^3)	2.05	1.06
X_{12}	Temperature (℃)	1.42	0.43
X_{13}	Duration (h)	1.31	0.32

The study's findings highlighted that the concrete's compressive strength was notably affected by the quantity of coarse aggregate present in the mix. This is attributed to the substantial volume occupied by coarse aggregates, significantly influencing the overall packing density and void content. Similar observations regarding the influence of coarse aggregates on concrete's compressive strength have been reported in earlier studies (Olivia & Nikraz, 2012),(Malkawi, 2023). Alongside coarse aggregates, the SiO_2 percentage in the fly ash (FA) also demonstrated a high PFI value, indicating its significant effect on concrete's compressive strength. This stems from the fact that the mechanical properties of cement-based mixes depend on hydration pace and extent. The chemical composition of FA plays a crucial role in the strength's development due to its impact on hydration and the formation of strength-contributing products (Moon et al., 2016), (Cho et al., 2019).

Another influential input variable was the concentration of sodium hydroxide (NaOH) in molarity. This parameter significantly impacted the compressive strength of concrete as well. The NaOH concentration influences the dissolution of geopolymer precursor materials, consequently affecting the uniformity and strength of the geopolymer matrix. Higher NaOH concentrations lead to better dissolution, yielding a more uniform and stronger matrix. While fine aggregates, aluminium oxide (Al_2O_3), and water also influenced compressive strength, their impact was comparatively lower compared to other input variables.

On the other hand, factors like fine aggregate, the ratio of aluminium oxide to fly ash (AA/FA), and pre-curing conditions exhibited relatively lower PFI values, suggesting their limited impact on compressive strength. Nevertheless, it's worth noting that the PFI index primarily indicates the individual influence of each variable and doesn't capture their combined effects (Naseri et al., 2020), (Asteris et al., 2021).

To delve deeper into the impact of changing individual input variables on concrete's compressive strength, a sensitivity analysis was conducted. For instance, raising the SiO_2 content in fly ash beyond a certain threshold negatively impacted compressive strength due to the production of unreacted particles. Similarly, within the studied range, increasing the coarse aggregate content, SiO_2 in fly ash, and NaOH concentration contributed positively to improved compressive strength (Malkawi, 2023), (Moon et al., 2016), (Ghafoor et al., 2021b).

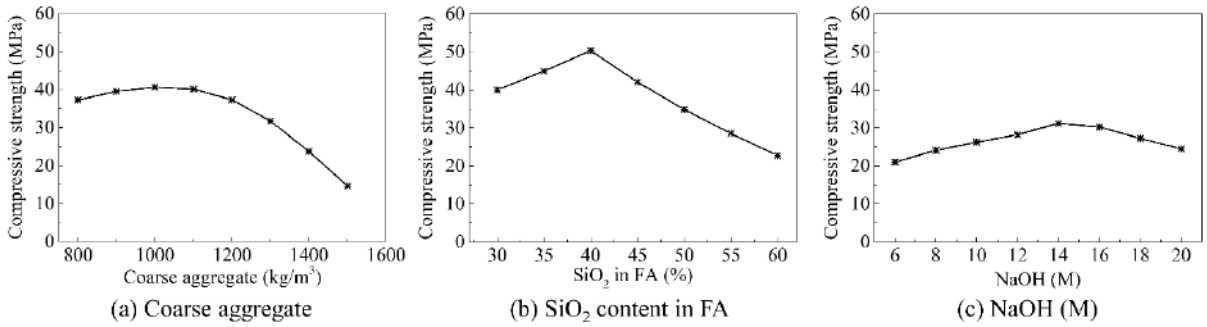


Fig 20: Sensitivity analyses for featured parameters

5. CONCLUSIONS AND RECOMENDATIONS

5.1.CONCLUSION

The objective of this study was to employ three distinct machine learning algorithms—BPNN, RFR, and KNN—to predict the compressive strength of geopolymer concrete based on fly ash (FA). The dataset was compiled from an extensive literature review, comprising 149 sets of mixing proportions. These sets encompassed 13 variables, such as FA composition (SiO_2 and $Al2O_3$), coarse and fine aggregate quantities, NaOH and Na_2SiO_3 contents and ratio, AA/FA ratio, water content, and pre-curing temperature and duration. Model effectiveness was assessed using metrics like R^2, MSE, RMSE, and MAE. The models were then compared, and the influence of input parameters on outcomes was analyzed. The conclusions drawn from the results are as follows:

1. The three machine learning algorithms—BPNN, RFR, and KNN—performed well in predicting the compressive strength of FA-based geopolymer concrete. The predictions closely aligned with actual values, with errors generally staying within 20%.
2. The BPNN model exhibited the best performance, achieving an exceptional R^2 of 0.948, outperforming RFR and KNN with R^2 values of 0.927 and 0.911, respectively. The success of BPNN can be attributed to its capacity to grasp intricate data patterns by approximating non-linear functions and adaptively learning from the data, setting it apart from RFR and KNN.

3. Coarse aggregate content, SiO_2 content in FA, and NaOH concentration emerged as the most influential factors with high PFI values, signifying their significant impact on compressive strength. Fine aggregates, Al_2O_3, and water had a comparatively smaller influence, whereas AA/FA, FA, and pre-curing conditions had lower PFI values, implying a limited effect. Sensitivity analysis indicated that excessive coarse aggregate, SiO_2, and NaOH concentration had adverse effects on compressive strength, while moderate increases within the studied range improved it.

5.2.RECOMMMENDATIONS

Incorporating machine learning techniques for predicting mechanical properties of FA-based geopolymer concrete offers the potential to streamline and enhance conventional empirical methods. This approach presents a more efficient way to assess concrete strength with untested mixes, reducing the time, effort, and resources required for experimentation. Additionally, it offers a cost-effective and environmentally friendly means of optimizing concrete mixes. Integrating machine learning expedites the progress and adoption of environmentally friendly concrete formulations, such as geopolymers, while minimizing costs and environmental impact. This innovative approach paves the way for a more sustainable and efficient construction industry, promoting eco-friendly development and reducing its carbon footprint.

REFERENCES

Aliabdo, A. A., Abd Elmoaty, A. E. M., & Salem, H. A. (2016). Effect of water addition, plasticizer and alkaline solution constitution on fly ash based geopolymer concrete performance. *Construction and Building Materials*, *121*, 694–703. https://doi.org/10.1016/j.conbuildmat.2016.06.062

Asteris, P. G., & Kolovos, K. G. (2019). Self-compacting concrete strength prediction using surrogate models. *Neural Computing and Applications*, *31*, 409–424. https://doi.org/10.1007/s00521-017-3007-7

Asteris, P. G., Skentou, A. D., Bardhan, A., Samui, P., & Pilakoutas, K. (2021). Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *Cement and Concrete Research*. https://doi.org/10.1016/j.cemconres.2021.106449

ben Chaabene, W., Flah, M., & Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. *Construction and Building Materials*, *260*. https://doi.org/10.1016/j.conbuildmat.2020.119889

Breiman, L. (2001). *Random Forests* (Vol. 45).

Chindaprasirt, P., & Chalee, W. (2014). Effect of sodium hydroxide concentration on chloride penetration and steel corrosion of fly ash-based geopolymer concrete under marine site. *Construction and Building Materials*, *63*, 303–310. https://doi.org/10.1016/j.conbuildmat.2014.04.010

Chindaprasirt, P., Chareerat, T., Hatanaka, S., & Cao, T. (2011). High-Strength Geopolymer Using Fine High-Calcium Fly Ash. *Journal of Materials in Civil Engineering*, *23*(3), 264–270. https://doi.org/10.1061/(asce)mt.1943-5533.0000161

Cho, Y. K., Jung, S. H., & Choi, Y. C. (2019). Effects of chemical composition of fly ash on compressive strength of fly ash cement mortar. *Construction and Building Materials*. https://doi.org/10.1016/j.conbuildmat.2019.01.208

Cover, T. M., & Hart, P. E. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Criado, M., Palomo, A., & Fernández-Jiménez, A. (2005). Alkali activation of fly ashes. Part 1: Effect of curing conditions on the carbonation of the reaction products. *Fuel*, *84*(16), 2048–2054. https://doi.org/10.1016/j.fuel.2005.03.030

Deb, P. S. (2013). *School of Civil and Mechanical Engineering Department of Civil Engineering Durability of Fly Ash Based Geopolymer Concrete*.

Deb, P. S., Nath, P., & Sarker, P. K. (2014). The effects of ground granulated blast-furnace slag blending with fly ash and activator content on the workability and strength properties of geopolymer concrete cured at ambient temperature. *Materials and Design*, *62*, 32–39. https://doi.org/10.1016/j.matdes.2014.05.001

Demuth, H., & Beale, M. (1992). *Neural Network Toolbox For Use with MATLAB User's Guide*. www.mathworks.com

Demuth, H., & De Jesús, B. (n.d.). *Neural Network Design 2nd Edition*.

Dener, M., Ok, G., & Orman, A. (2022). Malware Detection Using Memory Analysis Data in Big Data Environment. *Applied Sciences 2022, Vol. 12, Page 8604*, *12*(17), 8604. https://doi.org/10.3390/APP12178604

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Ghafoor, M. T., Khan, Q. S., Qazi, A. U., Sheikh, M. N., & Hadi, M. N. S. (2021a). Influence of alkaline activators on the mechanical properties of fly ash based geopolymer concrete cured at ambient temperature. *Construction and Building Materials*, *273*. https://doi.org/10.1016/j.conbuildmat.2020.121752

Ghafoor, M. T., Khan, Q. S., Qazi, A. U., Sheikh, M. N., & Hadi, M. N. S. (2021b). Influence of alkaline activators on the mechanical properties of fly ash based geopolymer concrete cured at ambient temperature. *Construction and Building Materials*. https://doi.org/10.1016/j.conbuildmat.2020.121752

Gunasekara, C., Atzarakis, P., Lokuge, W., Law, D. W., & Setunge, S. (2021). Novel analytical method for mix design and performance prediction of high calcium fly ash geopolymer concrete. *Polymers*, *13*(6). https://doi.org/10.3390/polym13060900

Gunasekara, C., Setunge, S., & Law, D. W. (2017). Long-term mechanical properties of different fly ash geopolymers. *ACI Structural Journal*, *114*(3), 743–752. https://doi.org/10.14359/51689454

Hagan, M. T., & Menhaj, M. B. (1994). Brief Papers Training Feedforward Networks with the Marquardt Algorithm. In *IEEE TRANSACTIONS ON NEURAL NETWORKS* (Vol. 5, Issue 6).

Hamid, N. A., Nawi, N. M., Ghazali, R., Najib, M., & Salleh, M. (2011). Accelerating Learning Performance of Back Propagation Algorithm by Using Adaptive Gain Together with Adaptive Momentum and Adaptive Learning Rate on Classification Problems. In *CCIS* (Vol. 151).

Han, Q., Gui, C., Xu, J., & Lacidogna, G. (2019). A generalized method to predict the compressive strength of high-performance concrete by improved random forest

algorithm. *Construction and Building Materials*, *226*, 734–742. https://doi.org/10.1016/j.conbuildmat.2019.07.315

Jiang, S., Pang, G., Wu, M., & Kuang, L. (2012). An improved K-nearest-neighbor algorithm for text categorization. *Expert Systems with Applications*, *39*(1), 1503–1509. https://doi.org/10.1016/J.ESWA.2011.08.040

Jin, C., Jang, S., Sun, X., Li, J., & Christenson, R. (2016). Damage detection of a highway bridge under severe temperature changes using extended Kalman filter trained neural network. *Journal of Civil Structural Health Monitoring*, *6*(3), 545–560. https://doi.org/10.1007/S13349-016-0173-8

Joseph, B., & Mathew, G. (2012). Influence of aggregate content on the behavior of fly ash based geopolymer concrete. *Scientia Iranica*, *19*(5), 1188–1194. https://doi.org/10.1016/j.scient.2012.07.006

Khalil, M., Alsmadi, S., bin Omar, K., & Noah, S. A. (n.d.). Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm. In *IJCSNS International Journal of Computer Science and Network Security* (Vol. 9).

Kong, D. L. Y., & Sanjayan, J. G. (2010). Effect of elevated temperatures on geopolymer paste, mortar and concrete. *Cement and Concrete Research*, *40*(2), 334–339. https://doi.org/10.1016/j.cemconres.2009.10.017

Kros, J. F., Lin, M., & Brown, M. L. (2006). Effects of the neural network s-Sigmoid function on KDD in the presence of imprecise data. *Computers and Operations Research*, *33*(11), 3136–3149. https://doi.org/10.1016/j.cor.2005.01.024

Kusbiantoro, A., Nuruddin, M. F., Shafiq, N., & Qazi, S. A. (2012). The effect of microwave incinerated rice husk ash on the compressive and bond strength of fly

ash based geopolymer concrete. *Construction and Building Materials*, *36*, 695–703. https://doi.org/10.1016/j.conbuildmat.2012.06.064

Leung, C. K. Y., & Pheeraphan, T. (1995). *VERY HIGH EARLY STRENGTH OF MICROWAVE CURED CONCRETE* (Vol. 25, Issue 1).

Li, B., Chen, Y. W., & Chen, Y. Q. (2008). The nearest neighbor algorithm of local probability centers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *38*(1), 141–154. https://doi.org/10.1109/TSMCB.2007.908363

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest* (Vol. 2, Issue 3). http://www.stat.berkeley.edu/

Malkawi, A. B. (2023). Effect of Aggregate on the Performance of Fly-Ash-Based Geopolymer Concrete. *Buildings*. https://doi.org/10.3390/buildings13030769

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

Moon, G. D., Oh, S., & Choi, Y. C. (2016). Effects of the physicochemical properties of fly ash on the compressive strength of high-volume fly ash mortar. *Construction and Building Materials*. https://doi.org/10.1016/j.conbuildmat.2016.08.148

Naderpour, H., Rafiean, A. H., & Fakharian, P. (2018). Compressive strength prediction of environmentally friendly concrete using artificial neural networks. *Journal of Building Engineering*, *16*, 213–219. https://doi.org/10.1016/j.jobe.2018.01.007

Naseri, H., Jahanbakhsh, H., Hosseini, P., & Moghadas Nejad, F. (2020). Designing sustainable concrete mixture by developing a new machine learning technique. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2020.120578

Nath, P., & Sarker, P. K. (2015). Use of OPC to improve setting and early strength properties of low calcium fly ash geopolymer concrete cured at room temperature. *Cement and Concrete Composites*, *55*, 205–214. https://doi.org/10.1016/j.cemconcomp.2014.08.008

Olivia, M., & Nikraz, H. (2012). Properties of fly ash geopolymer concrete designed by Taguchi method. *Materials and Design*. https://doi.org/10.1016/j.matdes.2011.10.036

Orr, Genevieve., & Müller, K.-Robert. (1998). *Neural networks : tricks of the trade*. Springer.

Pan, Z., Wang, Y., & Pan, Y. (2020). A new locally adaptive k-nearest neighbor algorithm based on discrimination class. *Knowledge-Based Systems*, *204*. https://doi.org/10.1016/j.knosys.2020.106185

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*. https://doi.org/10.1016/j.conbuildmat.2021.125785

Puertas, F., Martõ Ânez-Ramõ Ârez, S., Alonso, S., & Va Âzquez, T. (n.d.). *Alkali-activated fly ash/slag cement Strength behaviour and hydration products*.

Raja, S., & Fokoué, E. (2019). Multi-Stage Fault Warning for Large Electric Grids Using Anomaly Detection and Machine Learning. *Mathematics for Applications*, *8*(2), 115–130. https://doi.org/10.13164/ma.2019.08

Sapna, S. (2012). *Backpropagation Learning Algorithm Based on Levenberg Marquardt Algorithm*. 393–398. https://doi.org/10.5121/csit.2012.2438

Sarker, P. K., Haque, R., & Ramgolam, K. V. (2013). Fracture behaviour of heat cured fly ash based geopolymer concrete. *Materials and Design*, *44*, 580–586. https://doi.org/10.1016/j.matdes.2012.08.005

Shivam Sharma. (2021). *KNN - The Distance Based Machine Learning Algorithm*. https://www.analyticsvidhya.com/blog/2021/05/knn-the-distance-based-machine-learning-algorithm/

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. https://doi.org/10.1016/j.conbuildmat.2015.11.045

Topark-Ngarm, P., Chindaprasirt, P., & Sata, V. (2015). Setting Time, Strength, and Bond of High-Calcium Fly Ash Geopolymer Concrete. *Journal of Materials in Civil Engineering*, *27*(7). https://doi.org/10.1061/(asce)mt.1943-5533.0001157

Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J., & Steinberg, D. (2007). Top 10 algorithms in data mining. *Knowledge and Information Systems 2007 14:1*, *14*(1), 1–37. https://doi.org/10.1007/S10115-007-0114-2

Yeresime, S., Pati, J., & Rath, S. K. (2014). Review of software quality metrics for Object-Oriented methodology. *Advances in Intelligent Systems and Computing*, *216*, 267–278. https://doi.org/10.1007/978-81-322-1299-7_26

Yue, Shigang. (2010). *ICNC 2010 : proceedings : 2010 Sixth International Conference on Natural Computation : 10-12 August 2010, Yantai, Shandong, China*. IEEE.

Zhang, H., Li, L., Yuan, C., Wang, Q., Sarker, P. K., & Shi, X. (2020). Deterioration of ambient-cured and heat-cured fly ash geopolymer concrete by high temperature

exposure and prediction of its residual compressive strength. *Construction and Building Materials*, *262*. https://doi.org/10.1016/j.conbuildmat.2020.120924

Zhang, J., Ma, G., Huang, Y., sun, J., Aslani, F., & Nener, B. (2019a). Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. *Construction and Building Materials*, *210*, 713–719. https://doi.org/10.1016/j.conbuildmat.2019.03.189

Zhang, J., Ma, G., Huang, Y., sun, J., Aslani, F., & Nener, B. (2019b). Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. *Construction and Building Materials*, *210*, 713–719. https://doi.org/10.1016/J.CONBUILDMAT.2019.03.189

Zhuang, X. Y., Chen, L., Komarneni, S., Zhou, C. H., Tong, D. S., Yang, H. M., Yu, W. H., & Wang, H. (2016). Fly ash-based geopolymer: Clean production, properties and applications. In *Journal of Cleaner Production* (Vol. 125, pp. 253–267). Elsevier Ltd. https://doi.org/10.1016/j.jclepro.2016.03.019