

## Research paper

# A white-box machine learning model for predicting compressive strength of oilfield geopolymer cements

Okorie Ekwe Agwu<sup>a,\*</sup>, Saad Alatefi<sup>b</sup>, Muhammad Aslam Md Yusof<sup>a</sup>, Nwode Agwu<sup>c</sup>

<sup>a</sup> Department of Petroleum Engineering, Universiti Teknologi PETRONAS, Seri Iskandar, Perak, 32610, Malaysia

<sup>b</sup> Department of Petroleum Engineering Technology, College of Technological Studies, PAAET, Kuwait City, 70654, Kuwait

<sup>c</sup> College of Physical Sciences and Engineering, Cardiff University, Wales, CF24 3AA, UK

## ARTICLE INFO

## Keywords:

Geopolymer cement  
White-box modelling  
Compressive strength  
Oil well

## ABSTRACT

As the world transitions toward carbon neutrality, low-carbon cement alternatives are increasingly important. Geopolymer cements, synthesized from aluminosilicate precursors activated by alkaline solutions, offer a promising substitute for oilfield well cementing. However, accurate and interpretable prediction of compressive strength is essential for safe field deployment. This study develops a white-box, explainable machine learning model based on a Bayesian regularized neural network (BRNN) to estimate the compressive strength of geopolymer cements. The model targets high-performance, eco-friendly alternatives to Ordinary Portland Cement, particularly for harsh downhole environments. It balances predictive accuracy with interpretability by providing an explicit mathematical relationship between key parameters—alkaline solution concentration, mole ratio, liquid-to-fly ash ratio, curing temperature, curing time and compressive strength. This formulation enables rapid estimation without retraining or computational iteration, supporting practical engineering use. The results obtained shows that the model demonstrated strong predictive capability achieving an  $R^2$  of 0.9787, MSE of 8.33, RMSE of 2.86 MPa and MAE of 2.05 MPa. Sensitivity analysis shows curing temperature ( $r = +0.43$ ), curing time ( $r = +0.417$ ), and alkaline concentration ( $r = +0.411$ ) as dominant positive factors, while mole ratio ( $-0.374$ ) and liquid-to-fly ash ratio ( $-0.295$ ) exhibit negative effects. Trend analysis confirms consistency with established geopolymerization mechanisms. Leverage analysis indicates that 98.5% of data points fall within the model's applicability domain, confirming robustness. Overall, this work demonstrates that transparent, physics-consistent models can reliably support geopolymer cement optimization and sustainable drilling operations.

## 1. Introduction

In recent years, oil and gas discoveries have increasingly shifted toward challenging terrains such as ultra deep waters, high-pressure and high-temperature (HPHT) environments, depleted sands etc. Given the peculiar nature of these terrains, it is difficult to envisage a well drilled in such environments without proper casing and cementing operations. Cementing remains one of the most critical phases of well construction because it ensures zonal isolation, provides structural support to the casing, protects casing strings from corrosive fluids such as  $\text{CO}_2$  and  $\text{H}_2\text{S}$ , and contributes significantly to long-term well integrity [1,2]. Traditionally, the cement material predominantly used for oilfield applications is Ordinary Portland Cement (OPC), particularly API Class G and Class H cements [3], due to their predictable performance and established industry standards.

Failure of cement to effectively perform these functions can result in severe technical, environmental, and economic consequences. These include contamination of freshwater aquifers [4], casing collapse due to loss of structural support, sustained casing pressure, gas migration, and in extreme situations, catastrophic blowouts [5]. Because such failures may lead to human casualties, environmental damage, and significant financial losses, the integrity of wellbore cement is considered a high-priority issue in drilling and completion operations [6]. However, while OPC remains the dominant cementing material due to its reliability, its production process has attracted serious criticism because of its substantial contribution to global carbon emissions and climate change (Shivaprasad et al., [7]; Aceituno et al., [8]).

Traditionally, the process of producing cement begins with crushing limestone and then calcining the ground limestone at high temperature. This process leads to the evolution of  $\text{CO}_2$  which is a major driver of climate change [9]. Studies put the record that  $\text{CO}_2$  produced from this

\* Corresponding author.

E-mail address: [okorie.agwu@utp.edu.my](mailto:okorie.agwu@utp.edu.my) (O.E. Agwu).

<https://doi.org/10.1016/j.rineng.2026.113016>

Received 20 March 2026; Received in revised form 2 May 2026; Accepted 14 September 2026

Available online 15 September 2026

2590-1230/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

### Symbols & Nomenclature

|         |                                                     |
|---------|-----------------------------------------------------|
| GGBFS   | Ground-granulated blast-furnace slag                |
| KAN     | Kolmogorov-Arnold networks                          |
| LPH     | Liquid potassium hydroxide                          |
| LPS     | Liquid potassium silicate                           |
| LSBoost | Least squares boosting ensemble                     |
| LSH     | Liquid sodium hydroxide                             |
| LSS     | Liquid sodium silicate                              |
| LSTM    | Long short-term memory                              |
| RBF-PSO | Radial basis function – particle swarm optimization |
| SPS     | Solid potassium silicate                            |
| SSS     | Solid sodium silicate                               |
| SVR     | Support vector regression                           |

process accounts for 7–8% of the global CO<sub>2</sub> sources [10–12]. Given the destructive nature of this gas, attention has been shifted to sourcing for ecofriendly binder materials [13]. The results of this search has produced viable alternatives such as geopolymers, self-healing cements, slag-based blends, resin systems have gained increasing attention [14].

Among these alternatives, geopolymer cement has emerged as one of the most promising sustainable substitutes for Portland cement in wellbore applications and is the main focus of this study. Generally, synthesizing geopolymer cements involves the use of liquid alkaline activators (e.g. NaOH, KOH or sodium silicate solutions) to dissolve aluminosilicate precursors (fly ash or metakaolin) [15,16]. This process is depicted in Fig. 1.

In the construction industry, geopolymer binders have been extensively studied and deployed because of their excellent chemical resistance, superior mechanical strength, thermal stability, low permeability, and reduced shrinkage ([17,18]; Mohammed et al., [19–22]). More importantly, geopolymer production can reduce CO<sub>2</sub> emissions by more than 80% compared to conventional Portland cement production (Duxson et al., [23]; Juenger et al., [24]; McLellan et al., [25,26]), making them highly attractive from both engineering and environmental perspectives.

From a mechanical performance standpoint, geopolymer cement also demonstrates significant advantages over OPC particularly in early-age strength development [27]. For instance, cement based binders typically achieve about 80% of its design compressive strength within the first 7 days and reaches full structural acceptance at approximately 28 days [28]. In contrast, geopolymer cement, especially fly ash- and slag-based systems, can attain 70% of its ultimate compressive strength within the

first 3–4 hours at curing temperature between 30°C and 85°C due to rapid geopolymerization and accelerated aluminosilicate gel formation [29]. This faster strength development is particularly advantageous in oil well cementing where reduced waiting-on-cement (WOC) time directly lowers rig time and operational costs, especially in arctic drilling or deep-water seabeds. Despite these advantages, geopolymers—although widely adopted in civil construction remain limited in wellbore drilling applications as reported by Rahman et al. ([30]a) and Abdelaal [25].

To improve confidence in oilfield deployment, a number of experimental evaluations have been conducted and results so far indicate its suitability as a viable replacement for the conventional Portland cement ([31,32]; Ahdaya and Imgam, [33]; Ridha [34]; Suppiah et al., [35]; Rahman, et al., [36]b; [37–40]; da Silva et al., [41]). Many of these experimental investigations focus on the compressive strength of geopolymer cement, as it governs the binder's ability to develop effective bonding between the casing and the formation.

To establish compressive strength, traditionally, a destructive test where the cement sample is crushed under axial load is conducted [42]. According to Nyakilla et al. [43], this experimental protocol is characterized by extended waiting periods and there is limited scope for field-based testing. Similar observations were made by Chong et al. [44] and Harooni et al. [45]. They reported that laboratory experiments involving binders are rigorous. This is because they require numerous trials due to the number of influencing factors. These trials are needed to evaluate different mix designs with the goal of isolating the mix design that suits specific wellbore conditions. To illustrate, samples from various mix designs are typically allowed to set for 8 hours in the first instance and 16 or 24 hours as required before the test can be conducted. Meanwhile the idea of destroying the samples during the testing process also results in financial loss since cement materials are essentially wasted. Non-destructive tests equally abound such as the use of ultrasonic pulse velocity tests, use of rebound hammer and pullout tests. While these methods have their advantages, Hidayat et al. [46] assert that they have accuracy issues.

These limitations create a strong need for reliable predictive models capable of estimating compressive strength rapidly and accurately without repeated laboratory trials (Abid et al. 2023). Several models have been developed ranging from numerical, empirical and regression based as well as the widely adopted machine learning models. A review of the literature on this subject indicates that most studies on the prediction of cement compressive strength are predominantly concentrated within the field of concrete, while their application to oilfield cement remains limited and relatively underexplored [47]. The numbers documented in respect of geopolymer cements are even far scarcer than the

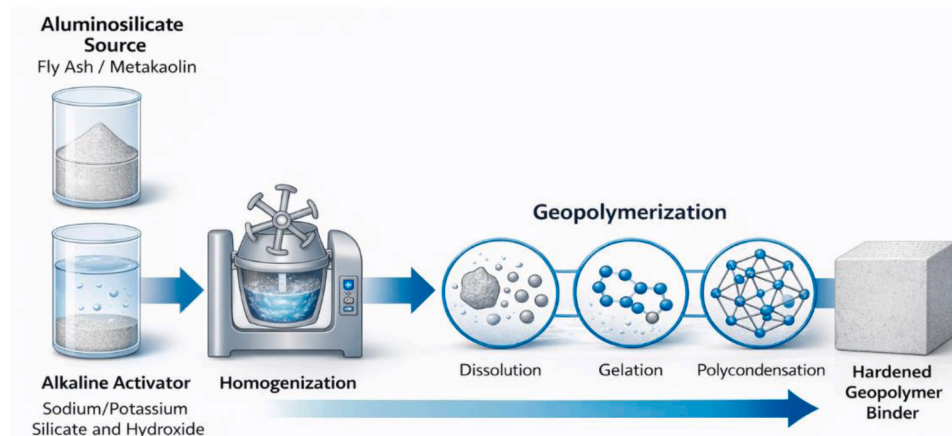


Fig. 1. Process of synthesizing geopolymer cement.

conventional Portland cements. The limited number of studies in this area is the main motivation for this study. Furthermore, the existing ML models for geopolymer cement compressive strength are largely not explainable, not explicitly presented in a usable mathematical equation and are not physically validated.

Therefore, this study focuses on developing an explainable predictive framework for estimating the compressive strength of geopolymer cement for oilfield wellbore applications, thereby improving design efficiency, reducing operational cost, and supporting the broader adoption of sustainable low-carbon cementing systems in the petroleum industry.

## 2. Literature appraisal

This section of the study is designed to showcase previous research efforts in the area of oilfield wellbore cementing especially as it relates to the use of geopolymer cements. Fundamental research often begins with experimental studies. In many cases, these studies provide the data required for modelling. Therefore, this review begins with a brief highlight of existing experimental works on the use of geopolymer cements. It then delves deeper into studies that focus on modelling the compressive strength of oilfield cements. In order to make the review focused, the models to be reviewed would be regression-based models and machine learning based models.

### 2.1. Experimental studies on geopolymer cements

Given the fact that the use of geopolymers cements have not yet gained much ground in the wellbore construction industry as it has in the building and road construction space, it is important to anchor the efforts of predictive modelling within experimentally validated tests. An overview of some of these experimental tests especially those that have a slant towards showcasing geopolymer cement additives, the design mix proportion of these additives provide the fundamental basis for this work and is presented in Table 1. The summary provided in Table 1 shows the specific tests carried out, the design mixture and the findings by each researcher. As this study would later show, the materials used by these authors and their mix ratios in Table 1 is evidence that the proposed model is not just data driven but also materially consistent. The following can be observed in the experiments: First, more than virtually all of the studies conducted compressive strength tests pointing to its crucial nature in oilwell cementing operations. Second, while the additives may vary, however, the main additives are the activators (alkali) and the solid precursors. Third, the findings suggest that geopolymer cements possess requisite properties that make them a viable alternative to Portland cement.

### 2.2. Models developed for compressive strength of oilfield cements

#### 2.2.1. Regression based models

In general, regression models express compressive strength as an explicit mathematical relationship between strength and a set of influencing variables, which can be written in the generic form as shown in Eq. (1).

$$y = a_0 + \sum_{j=1}^m a_j X_j \quad (1)$$

Where  $y$  is the dependent variable,  $X_j$  are the independent variables or the inputs while  $a_0, a_1, \dots, a_m$  are the coefficients or parameters. These regression models have the attraction that once fitted they can be used to perform predictions much more quickly than other modelling techniques and are correspondingly simpler to implement in software [52]. In oilfield cementing applications, these models are particularly valuable because they provide explicit equations that allow direct calculation of strength for operational decision-making such as determining WOC time.

**Table 1**

Summary of experimental studies on geopolymer cements.

| Reference(s)          | Tests                                                          | Inputs mix ratio & test conditions                                                                                                                                                                                                                                | Findings                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------|----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Xian [48]             | Rheology, filtration loss, uniaxial compressive strength (UCS) | <i>Inputs:</i> 70 – 100% fly ash, NaOH, Na <sub>2</sub> SiO <sub>3</sub> , distilled H <sub>2</sub> O, 10 – 30% silica fume                                                                                                                                       | Silica fume at 30% yields the optimum performance in the geopolymer cement mix.                                                                                                                                                                                                                                                                             |
| Paiva et al. [17]     | UCS                                                            | <i>Inputs:</i> Metakaolin, micro silica, mineral fibre, KOH, potassium silicate, propylene glycol based defoamer<br><i>Curing temperature</i> = 49°C<br><i>Curing time</i> = 14 days                                                                              | A higher UCS and lower stiffness was observed with respect to geopolymer cement compared to Portland cement.                                                                                                                                                                                                                                                |
| Suppiah et al. [35]   | UCS                                                            | <i>Inputs:</i> ASTM Class 'F' fly ash, NaOH & Na <sub>2</sub> SiO <sub>3</sub> solution, NaOH pellets<br><i>Curing temperature</i> = 60°C & 90°C. <i>Curing time</i> = 24 h                                                                                       | Curing the samples at higher temperature of 90°C rather than at 60°C imparted higher UCS to the sample. This is attributed to the fact that to stimulate the polymeric reaction, heat is essential.                                                                                                                                                         |
| Ahdaya et al. [49]    | Density, rheology, UCS                                         | <i>Inputs:</i> Class 'C' fly ash, NaOH molarity (5, 10, 15), Na <sub>2</sub> SiO <sub>3</sub> : NaOH (0.25, 0.5, 1, 2), tap water<br><i>Curing temperature</i> = 24°C<br><i>Curing time</i> = 24 h                                                                | When the concentration of NaOH is increased two effects occur: UCS increases while viscosity decreases slightly. However, a rise in the ratio of the alkaline activator to fly ash triggered a rise in the viscosity value which ultimately limits slurry pumpability. Fluid loss was limited when the proportion of sodium silicate to NaOH was increased. |
| Denduluri et al. [37] | Viscosity, porosity, pump time, UCS, tensile & bond strength   | <i>Inputs:</i> Class F fly ash, metakaolin (90%), slag (10%), activators (LPH, LPS, SPS, LSH, LSS, SSS), NaOH or KOH pellets/silicate solution, deionized water<br><i>Conditions:</i> static temperature = 315°C; circulating temperature = 232°C; BHP = 3000 psi | Geopolymers synthesized from K-based activators yielded higher UCS compared to Na-based activators. Furthermore, bond strength of K-based activated geopolymer cements were found to be 3-7 times that of Portland cement.                                                                                                                                  |
| Gomado et al. [50]    | UCS, shear bond strength, hydraulic sealability                | <i>Inputs:</i> Solid precursors (340g granite, 330g GGBFS, 30g silica flour, 30g microsilica); liquid activators (281g of 4M KOH soln, 47g water); retarders (2.5g KNO <sub>3</sub> & 4g Zn(NO <sub>3</sub> ) <sub>2</sub> ); rheology modifier (3.5g of Na       | The use of glass powder to play the role of an expansive agent controller helps to stabilize the shear bond strength and enhances the geopolymer's                                                                                                                                                                                                          |

(continued on next page)

**Table 1** (continued)

| Reference(s)                 | Tests                                                          | Inputs mix ratio & test conditions                                                                                                                                                                                                                                                                                                            | Findings                                                                                                                                                                    |
|------------------------------|----------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| da Silva et al. [41]         | Rheology, consistency & thickening time, UCS                   | superplasticizer); Strength enhancer (2g NaOH), expansive additive (8g waste glass powder)<br><i>Inputs:</i> Brazilian rock, GGBFS, microsilica, solid activator, superplasticizer, activator-to-solid ratio<br><i>Curing conditions:</i> (i) Room temperature, atm. pressure (ii) 70°C, 13.8MPa<br><i>Curing time:</i> 1, 3, 7, 14 & 28 days | hydraulic sealing potential.<br><br>Pressure has a significant impact on the geopolymer's pumping time; an enhancement of this pressure to 13.8 MPa slices the time by 50%. |
| Ogienagbon and Khalifeh [51] | Rheology, Free fluid, density, consistency, UCS, sealing tests | <i>Inputs:</i> Metakaolin, microsilica, ground granulated blast furnace slag. Activator: Potassium silicate solution (modular ratio = 2), superplasticizer (2.5% by wt of solid precursor)                                                                                                                                                    | In terms of UCS, geopolymers show a rise in UCS over time however, their rate of early strength is lower to that of thermosetting resins.                                   |

Table 2 summarizes previously published regression-based models built for estimating the uniaxial compressive strength (UCS) of oilfield cement with particular attention paid to their predictive performance, and the pros and cons of each individual model. The following are some observations from Table 2.

- First, the regression models in the table show explicit mathematical based strength equations which makes them easily explainable and allows for quick field deployment.
- Second, while many of the authors failed to include the size of data they used in building their models, the few that did used relatively small data sizes which limits their applicability to a wide range of conditions. For instance, Falode et al. [53] and Anya et al. [54] used 16 and 35 data points respectively.
- Third, more than 60% of the studies focus on UCS for Portland cements, while about 40% were devoted to geopolymer cements UCS.
- Finally, while some models use operationally accessible inputs such as curing time and temperature, others require parameters that may be difficult to obtain in real time.

#### 2.2.2. Critique of regression-based models for estimating oilfield cement UCS

While it may be said that the models in Table 2 are important in oilfield cement UCS prediction as they offer clear mathematical expressions which can be used for quick computation of UCS, however, the following critical analysis shows why their usability may be limited.

First, the domain of model application is narrow. The reason is tied to the small number of data samples used for building the models. In addition, these models are cement class specific with different additives composition. Therefore, the equations developed tend to fit almost perfectly in the range of data used. This is seen in the high R-square value of > 0.99 recorded for the models by Joel and Ademiluyi [55]; Ridha et al. [58] and Falode et al. [53]. These models may likely yield poor error metrics when subjected to new cement classes and additive formulations and downhole conditions. This lack of generalizability attribute attenuates user confidence for field scale use.

Secondly, regression models are built based on pre-defined

equations. The modeller is required to assume a mathematical relationship (which may be polynomial, linear, exponential, logarithmic) between the inputs and the cement UCS. A peek into the cement hydration and geopolymerization chemistry reveals that the process is an intrinsically non-linear, multiscale process governed by complex and synergistic physicochemical interactions ([59]; Vignesh et al., [60]). Thus, these models would have a hard time trying to capture these interactions and therefore given their inflexible nature, they tend to oversimplify these complex interactions.

#### 2.3. Machine learning (ML) based models for estimating oilfield cement UCS

A critical analysis of prevailing ML models for oilfield cement UCS forecasting is showcased in Table 3. The focus of the critical analysis is on the model DNA namely the size of the data utilized in its development, its explainability, its appropriateness for geopolymer cement and the feasibility of utilizing the model to aid in the adjustment of slurry components to reach a predetermined strength. This adopted style allows for the research gaps to be clearly isolated unlike what is obtainable if a pure narrative approach were chosen.

The following can be gleaned from the table:

- First, 28 ML models have been built and spread out across 13 studies with neural network derived models accounting for approximately 43% closely trailed by ensemble learning paradigms accounting for about 33% of the models.
- Second, the data size used by over 75% of the researchers were under 300 data points except for the 1119 and 1740 data points deployed by Abraham et al. [65], b) and Harooni et al. [45].
- Third, about 77% of the models were built for the conventional Portland cements and only about 23% of the models were built for geopolymer cements namely the models by Ling et al. [62], Tanyildizi [57] and Rahmawati et al. [67].
- Finally, some of the models adopted inputs that could be operationally accessible such as temperature, curing time, and proportion of water to cement, others utilized parameters that require derivation from microstructural characterization such as the inputs in the study by Wen et al. [64].

##### 2.3.1. Critical evaluation of existing oil field cement UCS prediction models

A discussion of the critical evaluation of the models as highlighted in column 6 of Table 3 is presented here. In this discussion, emphasis is placed on four key domains namely the model explainability, data size, the alignment with eco-friendly cement development, and the suitability of the predictor variables for real-time field deployment.

- Model explainability:** A prominent pitfall across the studies reviewed is the dearth of model explainability. By definition, explainability has to do with frameworks that are essentially built to enable the decisions made by the ML algorithm interpretable, thus making it transparent and comprehensible to users [69]. The concept of explainability would be anchored on four domains. First, establishing the input variables importance (global and local) using SHAP, LIME, partial dependency plots can be counted as parts of making the model explainable [70]. Second, for simple models, including the underlying mathematical equation that links the inputs to the output is considered making the model intrinsically interpretable (Kruschel et al., [71]). Third, physical consistency analysis using trend plots to verify that the model's logic follow known physical laws is another aspect of model explainability [72]. According to Arabzadeh and Frank [73], this physical plausibility breeds trust which is one of the cornerstones of explainable artificial intelligence. The fourth factor which is more of a validation tool that partially contributes to model

**Table 2**  
Regression based models for estimating oilfield cement compressive strength.

| Reference               | Model                                                                                | Data size & cement class              | Inputs, model developed, performance, model pros & cons                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-------------------------|--------------------------------------------------------------------------------------|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Joel and Ademiluyi [55] | Semi-empirical nonlinear regression model -inverse power law                         | Portland cement                       | <p><i>Model 1:</i> Input: Temperature (T)</p> $CS = 1397.432 + \frac{1466605}{T^{1.5}} - \frac{1.4 \times 10^{-7}}{T^2}$ <p><i>Model 2:</i> Inputs: Compressive strength time (t), slurry weight (<math>S_w</math>)</p> $CS = a + bt^{0.5} + \frac{c}{t^{0.5}}$ <p>Where: <math>a = 84126.25 - 11167S_w^3</math>; <math>b = -8791.66 + 1200.551S_w^3</math>; <math>c = -188777 + 24656.6S_w^3</math></p> <p><i>Performance:</i> Model 1: <math>R^2 = 0.999</math>; Model 2: <math>R^2 = 0.999</math></p> <p><i>Strength:</i> Input variables are easily obtainable parameters.</p> <p><i>Pitfalls:</i> Limited input variables for model 1.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Ridha et al. (2013)     | Modified Hasselman's equation (Power-law regression model)                           | Class G cement                        | <p><i>Inputs:</i> Strength at zero porosity (<math>S_0</math>), bulk conductivity (<math>\sigma_b</math>), pore solution conductivity (<math>\sigma_0</math>), cementation factor (m).</p> $Model: S = S_0 - b \left( \frac{\sigma_b}{\sigma_0} \right)^{1/m}$ <p><i>Performance:</i> <math>R^2 = 0.90 - 0.95</math></p> <p><i>Strengths:</i> Can predict compressive strength up to 50 days.</p> <p><i>Pitfalls:</i> The model ignores critical factors such as: temperature pressure, curing time, water-to-cement ratio, This makes it unsuitable for multivariate field conditions. The model does not inform how to adjust slurry composition to achieve a target strength.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Falode et al. [53]      | Factorial design (Quadratic polynomial regression)                                   | 16 sets of experimental data          | <p><i>Inputs:</i> Extender (<math>x_1</math>), accelerator (<math>x_2</math>), antifoam (<math>x_3</math>), dispersant (<math>x_4</math>)</p> $CS = 584.453125 + 36.171875x_1 + 1.32815x_2 + 26.859375x_3 + 16.984375x_4 + 6.859375x_1x_2 + 40.015625x_1x_3 + 33.765625x_1x_4 - 3.765625x_2x_3 + 13.734375x_2x_4 + 13.140625x_3x_4 - 16.796875x_1x_2x_3 - 25.671875x_1x_2x_4 - 1.140625x_1x_3x_4 + 37.390625x_2x_3x_4 - 13.078124x_1x_2x_3x_4 + 0.13258$ <p><i>Performance:</i> <math>R^2 = 0.998</math>; MSE = 0.070313; SSE = 1.125; Standard error = 0.1325</p> <p><i>Strengths:</i> Model captures additive interactions since additive interactions strongly influence cement hydration and strength development. Model has interpretable coefficients which has a clear statistical meaning (main effect or interaction). Model enables sensitivity analysis and ranking of additives influencing compressive strength. Models allows identification of optimal additive combinations that maximize compressive strength.</p> <p><i>Pitfalls:</i> Model valid only within the additive amount ranges used in the experimental design. Polynomial terms do not represent hydration reactions, phase formation, or microstructural evolution. Model ignores curing and downhole conditions e.g. temperature, pressure, curing time, and fluid chemistry are absent thus limiting its applicability to field cementing scenarios.</p> |
| Anyia et al. (2019)     | Levenberg-Marquardt                                                                  | 35 data samples of class C & H cement | <p><i>Inputs:</i> Final ultrasonic transit time (T), water to solids ratio (w), cement to solids ratio (r), curing time (<math>t_c</math>)</p> $Model 1: S = 119953.54 \exp^{-0.41525T} \times 0.49231 \times t_c^{0.231}$ $Model 2: S = 0.5 \left[ 520.89 \left( \frac{w}{r} \right)^{-1.59677} + 119999.92 \times \exp^{-0.40975T} \right] \times 0.35671 \times t_c^{0.39157}$ <p><i>Performance:</i> Model 1: <math>R^2 = 0.90671</math>, Mean ratio = 1.00927, Model 2: <math>R^2 = 0.87840</math>, Mean ratio = 0.89205</p> <p><i>Strengths:</i> Model requires only a few input parameters (curing time, mix variable). Model can be applied quickly for field predictions.</p> <p><i>Pitfalls:</i> Valid only for the cement systems, additives and conditions used to derive the model.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Zulkarnain et al. [56]  | Response surface methodology (4 <sup>th</sup> order polynomial regression – Quartic) | Geopolymer cement                     | <p><i>Inputs:</i> Transit time (<math>T_t</math>)</p> $F = 1.8875T_t^4 - 73.8324T_t^3 + 1078.3108T_t^2 - 6975.2615T_t + 16895.4948$ <p>Where: <math>T_t</math> = transit time in <math>\frac{\mu s}{in}</math>; <math>F</math> = compressive strength in MPa</p> <p><i>Performance:</i> <math>R^2 = 0.9929</math>.</p> <p><i>Strengths:</i> For estimating compressive strength of geopolymer cement containing antifoam</p> <p><i>Pitfalls:</i> No geopolymer cement additive utilized as inputs. Limited to a single input. Physically non-interpretable beyond trend fitting</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Tanyildizi [57]         | Multivariate linear regression                                                       | Geopolymer cement                     | <p><i>Model:</i> <math>f_c = 1.7208 - 2.674 \times M - 0.55 \times A - 11.43 \times L - 1.1692 \times T - 0.4033 \times C</math></p> <p>Where <math>f_c</math> = compressive strength, M = mole ratio, A = alkaline solution concentration, L = liquid to fly ash mass ratio, C = curing days, T= curing temperature</p> <p><i>Performance:</i> <math>R^2 = 0.8803</math></p> <p><i>Strengths:</i> Model is simple and transparent with an explicit equation which can be easily deployed in the field. Model has low computational cost</p> <p><i>Limitations:</i> Model assumes linear relationships whereas real cement hydration processes are non-linear, time- and temperature-dependent. Model has limited accuracy. Model lacks interaction terms, thus restricting its ability to represent complex cement behavior. There is possible physical inconsistency in the model since all coefficients are negative. This may indicate dataset bias, limited parameter range.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

explainability by complementing it is the application domain of the model; which essentially outlines the boundary within which the model prediction is considered reliable [74]. For these four domains, it is observed from Table 3 that 92% of the studies did not establish the applicability domain of the models and the need to include the mathematical correlation between the inputs except for the studies by Hussain et al. (2025) and Nyakilla et al.

[43] respectively. In addition, 83% of the studies did not perform a trend analysis to show physical plausibility of their models and 42% did not show any form of sensitivity analysis. These gaps highlighted essentially limit scientific insight on the subject and constitute part of the motivation for this study.

ii. *Dataset size:* ML modelling requires lots of good quality data to boost its reliability (Priestley et al., [75]). In Table 3, it is

**Table 3**  
Oilwell cement compressive strength estimation using ML algorithms.

| Reference             | Algorithm                                      | Data size & class of cement                                                     | Inputs                                                                                                                                                                                                                                                                              | Performance                                                                                                                                                                                                                           | Model's strengths & weaknesses                                                                                                                                                                               |
|-----------------------|------------------------------------------------|---------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| McElroy et al. [61]   | ANN [5-12-1]                                   | 195 samples of class G & H cement. 70% Training, 15% Testing, 15% Validation    | Amounts of nano-SiO <sub>2</sub> in % BWOC; nano-Al <sub>2</sub> O <sub>3</sub> in % BWOC; nano-TiO <sub>2</sub> in % BWOC, cement density, simulated wellbore curing temperature                                                                                                   | R <sup>2</sup> = 0.8667, MSE = 1.972; MAPE = 4.069%                                                                                                                                                                                   | <i>Strengths:</i> Useful for compressive strength estimation with nanomaterials<br><i>Weaknesses:</i> Dearth of explicit model, trend and sensitivity analysis, No model applicability domain                |
| Ling et al. [62]      | ANN                                            | 273 data points Geopolymer cement                                               | Mole ratio, alkaline solution concentration, liquid to fly ash mass ratio, curing days, curing temperature                                                                                                                                                                          | R <sup>2</sup> = 0.961, RMSE = 543, MAD = 355.3                                                                                                                                                                                       | <i>Strengths:</i> Trend & sensitivity analysis established.<br><i>Weaknesses:</i> Dearth of explicit model, No model applicability domain                                                                    |
| Nyakilla et al. [63]  | Gradient boosted regression trees              | 144 data points. 90% Training, 10% Testing                                      | Cement, water, fly ash, dispersant, curing time, oil well cement                                                                                                                                                                                                                    | R <sup>2</sup> = 0.95, MSE = 0.32, RMSE = 0.8, MAE = 0.349                                                                                                                                                                            | <i>Strengths:</i> Sensitivity analysis available<br><i>Weaknesses:</i> Dearth of explicit model & trend analysis, No model applicability domain                                                              |
| Ouadi et al. [42]     | ANN SVR-GA                                     | 98 data samples 70% Training, 15% Testing, 15% Validation                       | Specific surface (Sb), position parameter, 80% Passing size (P80), uniformity factor, % of particle size in µm (32 µm, 3-16 µm, 16-24µm, >24µm), % Aluminate, % Ferrite, % Alit, % Belite, % Na <sub>2</sub> O, % Sulphur Trioxide, % Lime Saturation Factor.                       | 28 days compressive strength ANN: R <sup>2</sup> = 0.9753, RMSE = 2.581, MAE = 1.82 SVR-GA: R <sup>2</sup> = 0.9757, RMSE = 1.4294, MAE = 0.9869                                                                                      | <i>Strengths:</i> Model is useful for estimating 2, 7 & 28 days compressive strength<br><i>Weaknesses:</i> Dearth of explicit model, trend and sensitivity analysis, No model applicability domain           |
| Nyakilla et al. [43]  | GMDH                                           | 150 data points for class G cement                                              | Cement, water, fly ash, dispersant, curing time, oil well cement                                                                                                                                                                                                                    | R <sup>2</sup> = 0.8611, MSE = 0.304, RMSE = 0.551                                                                                                                                                                                    | <i>Strengths:</i> Explicit model, sensitivity analysis available<br><i>Weaknesses:</i> Dearth of trend analysis, No model applicability domain                                                               |
| Wen et al. [64]       | KAN [4-8-8-1]                                  | Oilwell cement 70% Training, 30% Testing                                        | Gel pore, capillary pore, macropore, porosity                                                                                                                                                                                                                                       | R <sup>2</sup> = 0.86                                                                                                                                                                                                                 | <i>Strengths:</i> Trend analysis available<br><i>Weaknesses:</i> Dearth of sensitivity analysis, no explicit UCS equation, No model applicability domain                                                     |
| Wang et al. [47]      | ANN, SVR ELM, RF GWO-RF PSO-RF GWO-SVR PSO-SVR | 279 data sets 70% Training, 30% Testing                                         | Water/cement ratio, corrosion resistant material content, temperature, silicon content, admixture content, carbonation time (d), total pressure, CO <sub>2</sub> partial pressure (PCO <sub>2</sub> ), and hydrogen sulfide (H <sub>2</sub> S) partial pressure (PH <sub>2</sub> S) | *PSO-RF R <sup>2</sup> = 0.9275, RMSE = 2.6516, MAE = 2.0568, MAPE = 0.1347, MSE = 7.0308                                                                                                                                             | <i>Strengths:</i> Useful for compressive strength estimation under acidic corrosion. Sensitivity analysis done<br><i>Weaknesses:</i> Dearth of explicit model, trend analysis, No model applicability domain |
| Abraham et al. [65]   | Decision tree ANN RF                           | 1119 data points of class C, H & G cements. 70% training, 30% testing           | Sample age, curing temperature, dimensions of the sample, x-sectional area, weight, density, specific gravity, acoustic velocity                                                                                                                                                    | ANN: R <sup>2</sup> = 0.8048<br>Decision tree: R <sup>2</sup> = 0.9345<br>Random Forest: R <sup>2</sup> = 0.9687                                                                                                                      | <i>Strengths:</i> Sensitivity analysis done<br><i>Weaknesses:</i> Dearth of explicit model, trend analysis, No model applicability domain                                                                    |
| Abraham et al. [66]   | SVM                                            | 1119 data points of class C, H & G cement. 70% training, 30% testing            | Age of samples, curing temperature, dimensions of the sample, cross sectional area, weight, density, specific gravity, acoustic velocity (UPV)                                                                                                                                      | Class C: R <sup>2</sup> = 0.9431, MAE = 0.0646, MSE = 0.0061, RMSE = 0.0781<br>Class G: R <sup>2</sup> = 0.9261, MAE = 0.078, MSE = 0.0103, RMSE = 0.1013<br>Class H: R <sup>2</sup> = 0.6, MAE = 0.1341, MSE = 0.0248, RMSE = 0.1576 | <i>Strengths:</i> Sensitivity analysis done<br><i>Weaknesses:</i> Dearth of explicit model, trend analysis, No model applicability domain                                                                    |
| Rahmawati et al. [67] | ANN [2-30-1]                                   | 136 data points of geopolymer cements 70% training, 15% testing, 15% validation | Nanosilica (%), cellulose nanocrystals (%)                                                                                                                                                                                                                                          | R <sup>2</sup> = 0.941, RMSE = 0.192                                                                                                                                                                                                  | <i>Strengths:</i> Good model performance<br><i>Weaknesses:</i> Dearth of explicit model, trend analysis, no sensitivity analysis, limited data size used, No model applicability domain                      |
| Tanyildizi [57]       | LSTM, SVR, LSBoost                             | 273 data points Geopolymer cement                                               | Mole ratio, alkaline solution concentration, liquid to fly ash mass ratio, curing days, curing temperature                                                                                                                                                                          | *LSTM: R <sup>2</sup> = 0.9923, PSNR = 44.5864, MSE = 2.2617, NRMSE = 0.0283, RMSE = 1.5039<br>*GBR: R <sup>2</sup> = 0.961; MAPE = 9.81%, RMSE = 130.96 psi, MSE = 17.15                                                             | <i>Strengths:</i> Sensitivity analysis done<br><i>Weaknesses:</i> Dearth of explicit model, trend analysis. No model applicability domain                                                                    |
| Hussain et al. [68]   | ANN [3-25-5-1], RF, GB, SVR                    | 172 data points of light weight cement                                          | Cenospheres (%), density, temperature                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                       | <i>Strengths:</i> Model application domain established, sensitivity analysis done<br><i>Weaknesses:</i> Dearth of explicit model, trend analysis, limited data size used                                     |
| Harooni et al. [45]   | PSO-RBF                                        | Class G Portland cement 1740 data points. 80% training, 20% testing             | Anti gas migration, time, salt, cement density, silica flour, temperature, densifier, retarder, dispersant, boric acid, baso block                                                                                                                                                  | R <sup>2</sup> = 0.9998; AARD% = 3.38%, RMSE = 7.33                                                                                                                                                                                   | <i>Strength:</i> Trend analysis conducted.<br><i>Weaknesses:</i> Dearth of explicit model & sensitivity analysis. No model applicability domain                                                              |

\*Best performing model

observed that the size of the dataset across the studies is relatively small. For instance, about 75% of the studies used less than 300 data points with the median statistic falling at 172 data points. While it may be argued that there is no distinct definition of what a small dataset size is, however, using a small data size undermines the ability of the model to generalize to new data and increases the chances of overfitting. It is worthy however to remark that the study by Abraham et al. [65], Abraham et al. [66] and Harooni et al. [45] deployed a dataset size that might be considered large (>1000 data points) in comparison with the other studies. In summary, the larger the data size, the higher the potential to yield generalizable models.

- iii. *Consideration of eco-friendly cement systems:* In Table 3, it is noticed that about 77% of the studies focused on Portland cements while 23% focused on the low carbon cements such as geopolymer cements. The gulf in the numbers points to a research gap implying that models/studies that have a slant towards UCS prediction for low carbon cements in the wellbore construction industry is necessary as the clamour for a greener environment continues.
- iv. *Practicality of input variables for real-time UCS forecasting:* From a field perspective, while model inputs vary, however, those related to slurry design/composition of the cement offer some advantages in that they help in quick UCS computations and allow for adjusting component proportions to yield a target strength. This contrasts with the inputs that require parameters that need to be obtained from laboratory work. For instance, the microstructural parameters used as inputs in the model by Wen et al. [64] such as gel pore, capillary pore etc while being vital are challenging to measure and are generally unsuited for real-time UCS estimation.

### 3. Methodology

This section presents the methodology used in the study. The data for this study was obtained from the work by Ling et al. [62]. This data comprises 273 data entries which include five input variables (concentration of alkaline solution, mole ratio, liquid fly ash ratio, curing temperature, curing time) and one output variable (geopolymer cement compressive strength). The statistical analysis of the variables in the database is summarized in Table 4.

#### 3.1. Correlation matrix of input parameters and compressive strength

Fig. 2 is the pictorial representation of the correlation matrix (Pearson) highlighting the interactions among the five inputs and the geopolymer strength. To make the interactions visible at a glance, the direction and strength are represented by distinct colours (red and blue) and size of the bubble (big and small) respectively. Based on this nomenclature, the following can be pointed out:

First, with respect to UCS, (i) mole ratio and liquid fly ash ratio both

have a strong negative correlation of  $-0.77$  and  $-0.66$  respectively, pointing to the fact that increasing these parameters causes strength attenuation (ii) the concentration of alkaline solution, curing time and temperature have moderately positive correlation with values of  $0.42$ ,  $0.41$  and  $0.41$  respectively, pointing to the fact that an increase in their values yields an increase in geopolymer strength.

Second, the correlation among the variables themselves can essentially be described as weak (ranging from  $0.027$  to  $0.36$ ). This shows the variables have low multicollinearity, hence they well suited for use in the model building process.

#### 3.2. Distribution characteristics of input variables

Fig. 3. showcases the data spread for the variables in the database. From the figure, the multiple bulges along the vertical for the alkaline concentration (occurring at 15%, 20% & 25%), mole ratio (occurring at 1, 1.5 & 2), liquid fly ash ratio (occurring at 0.33, 0.4, 0.5 & 0.6) and curing temperature (occurring at  $24^{\circ}\text{C}$  &  $50^{\circ}\text{C}$ ) indicates the data for these parameters are trimodal, trimodal, quadrimodal and bimodal respectively. The narrower section in these plots indicates low data density. The multimodal nature of these variables is traceable to the fact that cement slurry design experiments are conducted at specific design levels. This method essentially simplifies experimentation time and cost and provides the leverage for the experimenter to systematically evaluate the impact of varying factor levels on the compressive strength. The curing time also exhibits multimodality as a result of the discrete experimental testing levels at 1, 3, 7 and 28 days. It has to be remarked that the effect of multimodality can be taken care of by neural networks since they are capable of learning distinct non-linear relationships for each regime. Finally, the strength data exhibits a more continuous spread as a result of being a measured parameter that is impacted by different factors.

#### 3.3. Data preprocessing via data normalization

No matter the capability of a learning paradigm, its effectiveness is diminished in the absence of high-quality data [76]. A veritable means of avoiding this is by subjecting the dataset to a series of independent steps (each of which can be done in isolation) such as cleaning, reduction, transformation (normalization) and integration [77]. These are called preprocessing steps. According to Roy et al. [76], the intervention of domain expert is required in choosing which of the steps are necessary for a given data. Therefore, since the data used in this work is from a secondary source and it had been cleaned, hence only the normalization step is deemed necessary.

Normalization was chosen because the dataset variables had different scales. Normalization entails transforming the data by scaling it so that it approximates a normal distribution. Examples of normalization approaches include log transformation, min-max, z-score, decimal scaling and quantile normalization [78]. The min-max method that scales data in the range of  $-1$  to  $1$  is chosen since it aligns with the zero

**Table 4**  
Database variables statistical analysis.

|                    | Alkaline sol. conc.<br>(%) | Mole ratio | Liquid-fly ash ratio | Curing temperature<br>( $^{\circ}\text{C}$ ) | Curing time<br>(days) | Compressive strength [UCS]<br>(MPa) |
|--------------------|----------------------------|------------|----------------------|----------------------------------------------|-----------------------|-------------------------------------|
| Maximum            | 25                         | 2          | 0.6                  | 50                                           | 28                    | 76                                  |
| Minimum            | 15                         | 1          | 0.33                 | 23                                           | 1                     | 0.3                                 |
| Range              | 10                         | 1          | 0.27                 | 27                                           | 27                    | 75.7                                |
| Average            | 19.96337                   | 1.5        | 0.451136             | 37.24176                                     | 10.21612              | 20.19487                            |
| Median             | 20                         | 1.5        | 0.4                  | 50                                           | 7                     | 15.3                                |
| Standard Deviation | 4.112228                   | 0.411239   | 0.100713             | 13.50436                                     | 10.87568              | 16.98837                            |
| Skewness           | 0.013606                   | 9.05E-17   | 0.259438             | -0.11067                                     | 0.950323              | 1.091281                            |
| Kurtosis           | -1.5219                    | -1.52214   | -1.34053             | -2.00248                                     | -0.93121              | 0.659884                            |

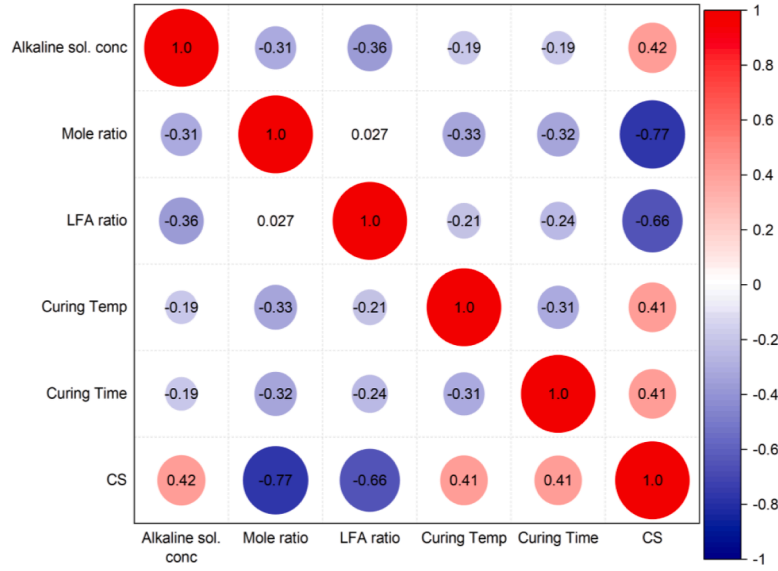


Fig. 2. Correlation matrix for study database.

centered transfer function “tansig” used in MATLAB. This helps in maintaining neuron activations in the high-gradient region and promoting effective learning. The formula for min-max is shown in Eq. (2).

$$UCS_n = 2 \left( \frac{UCS - UCS_{\min}}{UCS_{\max} - UCS_{\min}} \right) - 1 \quad (2)$$

where  $UCS$  is the original value of strength,  $UCS_{\min}$  &  $UCS_{\max}$  are the least and maximum greatest values of the compressive strength, and  $UCS_n$  is the normalized value of the strength. From Table 4, the least and greatest values of each parameter can be used to generate the normalized forms of the variables. This is shown in Equations (3)– (8).

$$A_n = 2 \left( \frac{A - 15}{10} \right) - 1 \quad (3)$$

$$M_n = 2(M - 1) - 1 \quad (4)$$

$$LFA_n = 2 \left( \frac{LFA - 0.33}{0.27} \right) - 1 \quad (5)$$

$$T_n = 2 \left( \frac{T - 23}{27} \right) - 1 \quad (6)$$

$$t_n = 2 \left( \frac{t - 1}{27} \right) - 1 \quad (7)$$

$$CS_n = 2 \left( \frac{CS - 0.3}{75.7} \right) - 1 \quad (8)$$

Since the variables are in normalized forms ranging between  $-1$  and  $1$  and the output would equally be in this range, hence there is need to return the values of the outputs in the original scale to maintain physical interpretability, hence a denormalization needs to be done. This can be achieved by using Eq. (9).

$$CS = 37.85 CS_n + 38.15 \quad (9)$$

### 3.4. Fundamentals of bayesian regularized neural network

Bayesian regularized neural networks (BRNNs) are basically a variety of neural networks that deploy Bayesian inference to the network learning pipeline to boost its generalizability especially when the data size is considered small and noisy [79,80]. The distinction between BRNNs and the conventional neural nets is that BRNNs handle the biases

and weights as arbitrary variables guided/driven by probability distribution rather than as fixed parameters to be optimized [81]. BRNNs focus on pruning down the combined objective function that compensates for both data fitting and weight constraint rather than reducing only the sum of squared errors [82]. The combined objective function is shown in Eq. (10).

$$F = \beta E_D + \alpha E_W \quad (10)$$

In Eq.10,  $E_D$  is data error term while  $E_W$  is weight constraining term while  $\alpha$  &  $\beta$  are regularization coefficients that governs the balance between complexity and accuracy [83]. An added advantage of the weights limiting factor is that only necessary weights are utilized, hence shielding the network from overfitting [84]. During the model learning process, the regularization coefficients ( $\alpha$  &  $\beta$ ) are updated based on the effective number of parameters denoted as  $\gamma$  and mathematically described in Eq. (11).

$$\gamma = N - 2\alpha \text{tr}(H^{-1}) \quad (11)$$

where  $H$  = objective function Hessian matrix &  $N$  = weights (total number). Upon updating,  $\alpha$  &  $\beta$  are recomputed using Eq. (12) and Eq. (13) respectively.

$$\alpha = \frac{\gamma}{2E_W} \quad (12)$$

$$\beta = \frac{n - \gamma}{2E_D} \quad (13)$$

where  $n$  = training samples number. To execute a prediction for parameter  $y$ , BRNN utilizes Eq. (14).

$$y = f(W^{(2)} g(W^{(1)} x + b^{(1)}) + b^{(2)}) \quad (14)$$

In Eq. (14),  $g$  = hidden-layer activation function;  $f$  = output activation function,  $W$  &  $b$  are weights and biases respectively. The entire process described above is summed up in Fig. 4.

### 3.5. Rationale for the selection of the modelling algorithm

The modelling paradigm adopted in this study is the Bayesian Regularized Neural Network (BRNN). The selection of this algorithm is based on three main considerations. First, while ML models especially neural networks typically perform best with large datasets, the dataset used in this study comprises 273 data points, which can be considered

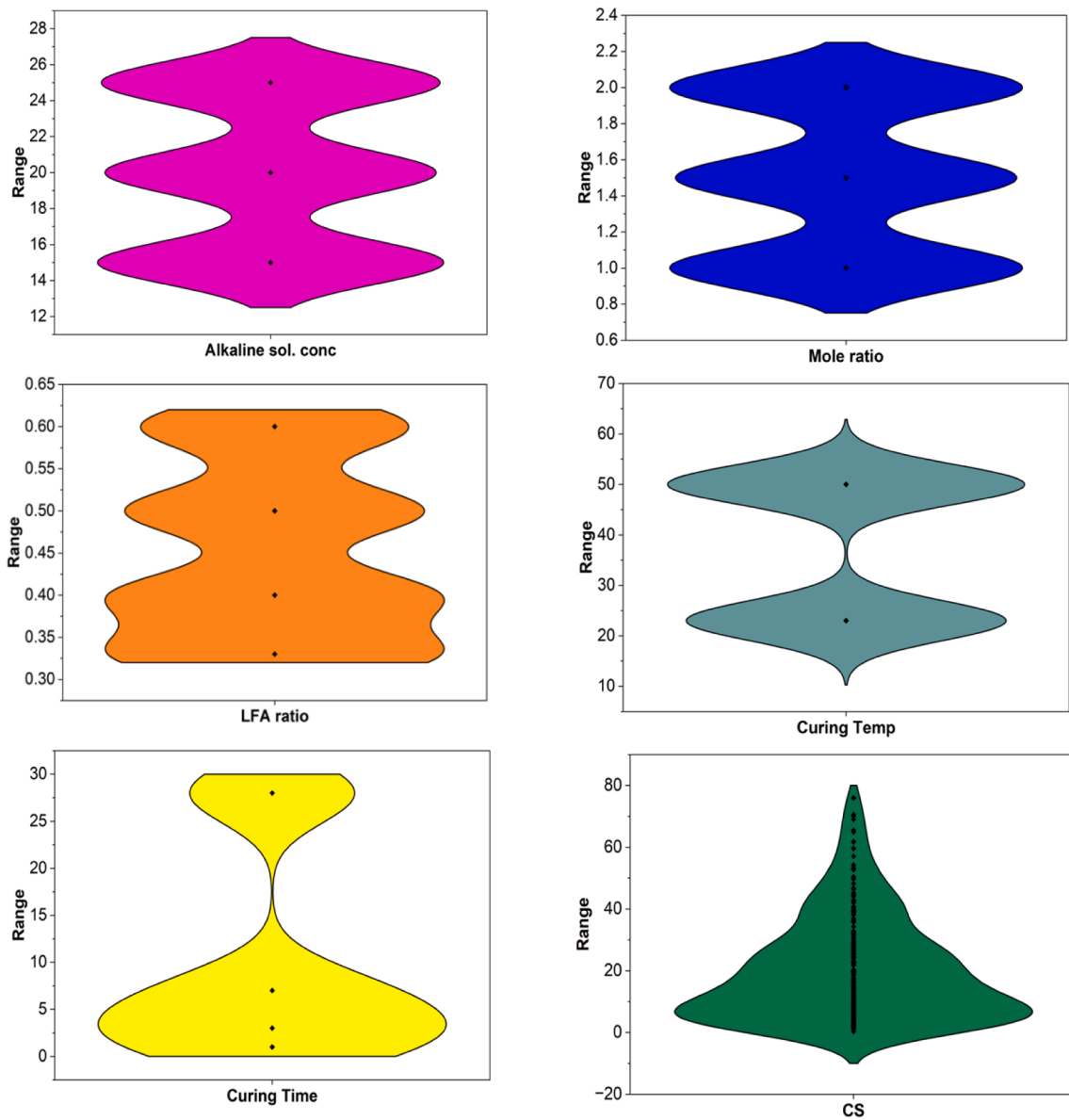


Fig. 3. Violin plots showing variable data spread.

relatively limited. In this case, training algorithms incorporating Bayesian regularization are advantageous, as they are well-suited for small to moderate datasets because it balances model complexity with data fitting using Bayesian inference leading to improved generalization [85]. Second, BRNN models are inherently less prone to overfitting because they determine and utilise an optimal number of effective network weights, thereby suppressing the influence of irrelevant parameters [84]. Finally, the Bayesian regularization framework enhances model robustness and reduces reliance on extensive cross-validation procedures typically required in conventional backpropagation neural networks [84].

### 3.6. Model building platform and hyperparameters tuning

The simulation runs for building a robust model for geopolymers cement strength forecasting were implemented using the neural net toolbox in MATLAB R2023b. The network tool was called up using the “*nftool*” command. The dataset was partitioned into 80% training and 20% testing. The training algorithm deployed is Bayesian regularization (*trainbr*). Before embarking on the training, as already described in

section 3.3, the predictors (input variables) were normalized using the min-max paradigm in order to stabilize gradient-based learning and boost model convergence. Table 5 shows the hyperparameters and their ranges used for model building. For these hyperparameters, the objective is to optimize them in such a manner that a combination of all of them yields the lowest mean square error.

### 3.7. Measures of prediction accuracy

Table 6 is a summary of the benchmarks adopted for evaluating the accuracy of the white box model in this study. In this table four metrics are chosen. The description of each metric with their upper and lower bounds along with their correlations are presented.

## 4. Results and discussion

### 4.1. Network structural design and optimization

The first stage of building a model is to design its architecture. Fig. 5 is the optimized network design for the model. The model architecture

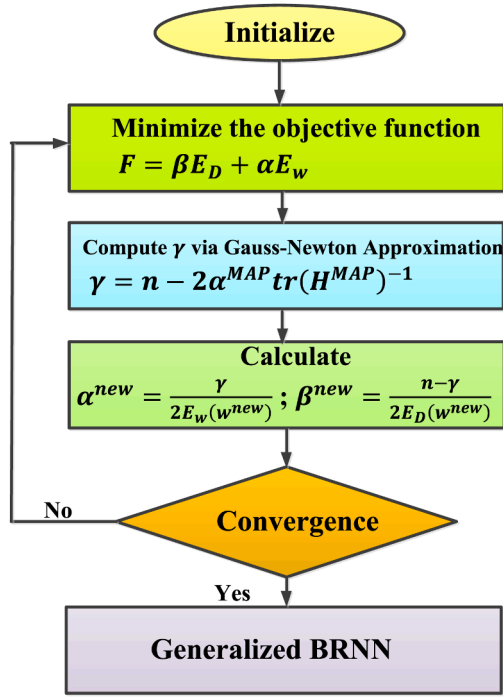


Fig. 4. BRNN algorithm flow protocol.

Source: Adapted from Garg and Mishra [81].

Table 5

Summary of hyperparameters used in model development.

| Hyperparameter                                  | Function                                                                                         | Range                                                           |
|-------------------------------------------------|--------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Regularization Parameters ( $\lambda, \alpha$ ) | Balance model complexity and data fit and to shrink the chances of overfitting [84]              | $10^{-6} - 10^6$<br>Automatically adjusted (Bayesian inference) |
| Number of Hidden Neurons                        | Determines network architecture                                                                  | 3 – 10                                                          |
| Number of hidden layers                         | Determines network depth                                                                         | 1 – 3                                                           |
| Activation Function                             | Introduces nonlinearity to aid network in capturing complex relationships (Gemechu et al., [86]) | Tan-sigmoid ( <i>tansig</i> )<br>[-1 1]                         |
| Epochs                                          | Training iterations                                                                              | 100 – 1500                                                      |
| Minimum gradient threshold                      | Training stoppage when gradient falls below this threshold                                       | $10^{-10} - 10^{-6}$                                            |
| Performance goal (MSE)                          | Tracks predictive accuracy                                                                       | $10^{-6} - 10^{-2}$                                             |

shows the three layers namely the input layer with five neurons; a single hidden layer having seven neurons with a hyperbolic tangent (*tansig*) transfer function; and one output neuron with a pure linear (*purelin*) activation function. The architecture was chosen after 50 independent training trials with varying hidden layer sizes ranging from 2 to 15 neurons. After evaluating numerous topologies, the 5–7–1 topology yielded the least MSE and the highest  $R^2$ , affirming it as the optimal structure. The explicit identification of input–hidden weights (IW), hidden–output weights (LW), and corresponding biases ( $b_1$  and  $b_2$ ) boosts the interpretability of the model by allowing the evolution of a closed-form mathematical expression.

#### 4.2. White box model for rapid forecasting of geopolymers cement strength

Upon training the network, the optimized weights and biases that yielded the least mean square error are shown in Table 7. The weights essentially typify the relative impact (could be positive or negative) each

Table 6

Performance benchmarks.

| Metric                                 | Description                                                                                                                                                   | Correlation                                                                                        |
|----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| Coefficient of determination ( $R^2$ ) | The fraction of the variation in the output explained by the model. Lower & upper bound limits: $-\infty$ & $+1$ (Chicco et al., [87])                        | $R^2 = 1 - \frac{\sum_{i=1}^n (UCS_i - \widehat{UCS})^2}{\sum_{i=1}^n (UCS_i - \overline{UCS})^2}$ |
| Mean square error (MSE)                | The mean squared deviation between the actual/observed and forecasted values. Best MSE = 0. Worst MSE = $+\infty$ (Tyagi et al., [88])                        | $MSE = \frac{1}{n} \sum_{i=1}^n (UCS_i - \widehat{UCS})^2$                                         |
| Root mean square error (RMSE)          | This is the square root of the MSE [89]. Perfect RMSE = 0. Worst RMSE = $+\infty$                                                                             | $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (UCS_i - \widehat{UCS})^2}$                                 |
| Mean absolute error (MAE)              | The average size of the error between the real and model forecasted values with the direction not accounted for. Perfect MAE = 0. Worst MAE = $+\infty$ [90]. | $MAE = \frac{1}{n} \sum_{i=1}^n  UCS_i - \widehat{UCS} $                                           |

Note:  $UCS_i$  = actual strength value;  $\widehat{UCS}$  = estimated strength;  $\overline{UCS}$  = mean strength;  $n$  = data size

predictor has on the final model prediction. On the other hand, the biases serves to adjust the thresholds of the activation function, thereby allowing it to detect the complex associations between variables that affect the geopolymer strength. By presenting these parameters, the model can be reproduced and by extension imparts interpretability to it it has to be noted that there are 40 weights and 6 biases, and these weights and biases are assigned to each of the 7 hidden neurons deployed in the intermediate layer. The low number of weights and biases is indicative of the fact that the model has low computational burden/costs.

These weights are aggregated at the output layer to produce a closed-form predictive equation for compressive strength. For the model, the equations for each of the 7 hidden neurons are presented in Eq.15 – Eq. (21). This explicit formulation enables rapid hand or spreadsheet-based estimation of geopolymer compressive strength without retraining the model, thereby facilitating practical engineering calculations and parametric sensitivity assessments.

$$H_1 = 1.501 \times \tanh[0.393873A_n - 0.17984M_n - 0.19646LFA_n + 0.26557t_n + 0.336157t_n - 0.55586] \quad (15)$$

$$H_2 = 1.175 \times \tanh[-0.44846A_n + 1.2985M_n - 0.3LFA_n - 0.7648T_n - 0.8088t_n + 1.56188] \quad (16)$$

$$H_3 = -1.531 \times \tanh[-0.514A_n + 1.626M_n - 0.275LFA_n - 0.4932T_n - 0.5775t_n + 1.42258] \quad (17)$$

$$H_4 = -0.923 \times \tanh[0.1739A_n - 0.07059M_n - 0.0509LFA_n + 0.029T_n - 1.41297t_n - 1.90745] \quad (18)$$

$$H_5 = 0.97192 \times \tanh[0.366354A_n - 0.37799M_n - 0.12814LFA_n + 0.07819T_n - 0.77151t_n - 0.27096] \quad (19)$$

$$H_6 = 0.83035 \times \tanh[-0.49603A_n + 0.478879M_n + 0.136215LFA_n - 0.11407T_n + 0.266t_n - 0.13597] \quad (20)$$

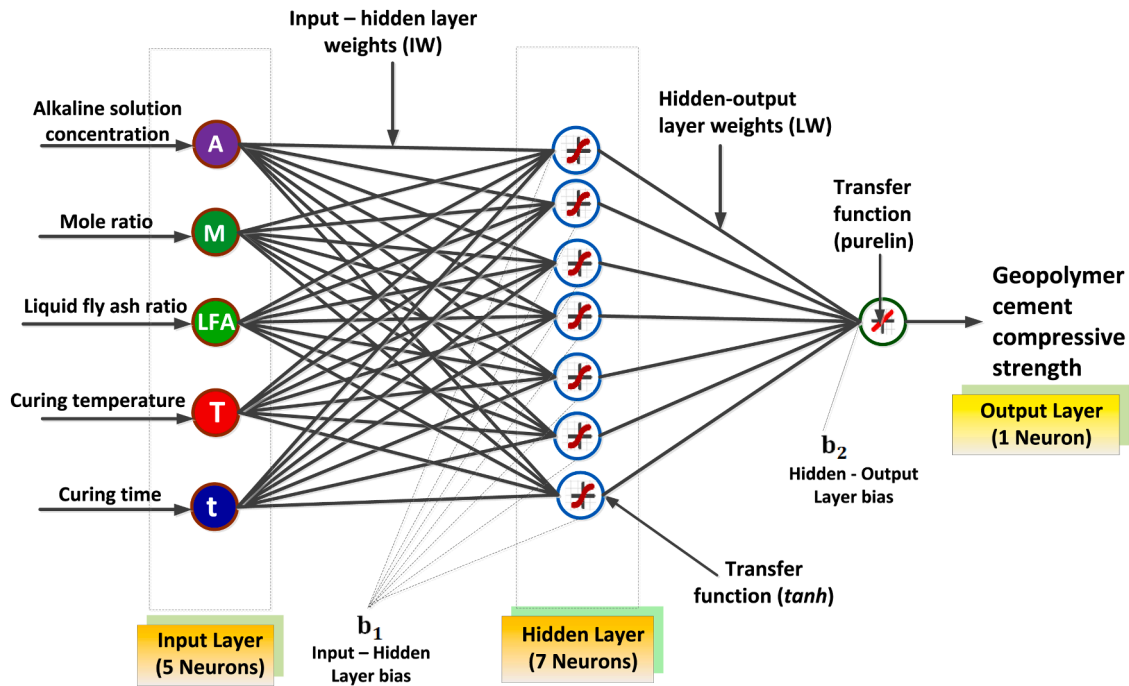


Fig. 5. Topology of contrived network.

Table 7

Weights and biases of neural network model.

| A        | M        | LFA      | T        | t        | b1       | LW       | B2       |
|----------|----------|----------|----------|----------|----------|----------|----------|
| 0.393873 | -0.17984 | -0.19646 | 0.265521 | 0.336157 | -0.55586 | 1.501066 | -0.68984 |
| -0.44846 | 1.298587 | -0.3     | -0.76483 | -0.80889 | 1.561886 | 1.175332 |          |
| -0.51426 | 1.62632  | -0.27504 | -0.49321 | -0.5775  | 1.422588 | -1.5311  |          |
| 0.173934 | -0.07059 | -0.05093 | 0.029045 | -1.41297 | -1.90745 | -0.92311 |          |
| 0.366354 | -0.37799 | -0.12814 | 0.078193 | -0.77151 | -0.27096 | 0.971916 |          |
| -0.49603 | 0.478879 | 0.136215 | -0.11407 | 0.266006 | -0.13597 | 0.83035  |          |
| 0.654648 | -1.4266  | 0.370802 | 0.10222  | -0.27923 | -1.83996 | -0.69672 |          |

$$H_7 = -0.696 \times \tanh[0.6546A_n - 1.4266M_n + 0.3708LFA_n + 0.10222T_n - 0.27923t_n - 1.83996] \quad (21)$$

The aggregation of Eq. (15) – Eq. (21) yields Eq. (22)

$$CS_n = \sum_{i=1}^7 H_i - 0.68984 \quad (22)$$

Since the parameters in Eq. (15) – Eq. (21) are in their normalized forms, the final denormalized form is shown in Eq. (23).

$$CS = 37.85CS_n + 38.15 \quad (23)$$

Where A = alkaline solution concentration, M = mole ratio, LFA = liquid fly ash ratio, T = curing temperature, t = curing time, CS = compressive strength, n = normalized value of each parameter. This normalized value can be obtained from Eq. (3) to Eq. (7) as shown in section 3.

Given that geopolymer cements have the aluminosilicate precursor and alkaline activator as key components, the absence of either of them means that no polymeric binder can evolve, therefore, compressive strength should yield a negligible value or zero. Thus, there is need to constrain the model to obey this condition as shown in Eq. (24).

$$CS(A = 0, LFA = 0, M, T, t) = 0 \quad (24)$$

To achieve this a smooth gate function  $(1 - e^{-kx})$  is introduced. Here x represents LFA and alkaline concentration and k is a constant (scaling

parameter). The value of k was optimized by means of the solver tool in Microsoft excel where a value of 63 was obtained. Multiplying this gate function with Eq. (23), the final model is obtained as shown in Eq. (25).

$$CS = (1 - e^{-63A})(1 - e^{-63LFA})(37.85CS_n + 38.15) \quad (25)$$

Eq. (25) essentially preserves the predictive efficiency of the model while ensuring that the model yields zero compressive strength when the chief reactants (aluminosilicate and alkaline solution) are not present. This physical constraint boosts the chemical consistency of the model as well as its robustness.

#### 4.3. Model performance

The information in Table 8 shows the performance of the model built in this work. A look at the metrics indicates the model exhibits a high predictive accuracy. This is so because the R-square values of 0.977 and 0.978 recorded for the training and testing sets is an indication that nearly 98% of the variation in geopolymer strength was accounted for by the model. This is corroborated by Chicco et al. [87] who asserted

Table 8

Model performance.

|          | R <sup>2</sup> | MSE     | RMSE   | MAE    |
|----------|----------------|---------|--------|--------|
| Training | 0.9774         | 6.10439 | 2.4707 | 1.8501 |
| Testing  | 0.9787         | 8.2235  | 2.8676 | 2.0486 |
| All data | 0.9774         | 6.5313  | 2.5556 | 1.8901 |

that R-square values greater than 0.8 lie within the very good performance model category. Another point to note regarding the R-square values is that the close proximity in values for both the training and testing sets shows that little or no overfitting occurred. For the other metrics (MSE, RMSE & MAE), it is noticed that there are only marginal rises for the testing data compared to the training set values with MSE increasing by 36.4%, RMSE by 16% and MAE by 10%. The overall model effectiveness as shown in the metrics points to model stability and robustness.

Furthermore, the observed variability in the dataset is notably high (coefficient of variation  $\approx 84\%$ ), corresponding to an approximate dispersion of  $\pm 17$  MPa in compressive strength values. This can be attributed to the broad range of mix designs, curing conditions, and strength levels represented in the data. In this context, the model RMSE of 2.87 MPa is substantially lower than the inherent data variability. This indicates that the predictive error is within acceptable uncertainty bounds and supports the reliability of the model.

#### 4.4. Correlation analysis between experimental and model predicted strength

The parity plots in Fig. 6 where the model forecast data is plotted against the actual data plays the role of visual validation, namely, to verify if the data points lie on the 45° line. It is essentially used for over or under model prediction analysis (Adeniyi et al., [91]). The benchmark for interpretation is as follows: (i) ideal forecast: All points lie on the equiangular line (ii) under estimation – the points tend to aggregate over the line (iii) over estimation – the data points aggregate below the line and (iv) high variability – points are spread out far from the line. From Fig. 6, it is observed that the points essentially aggregate on the 45° line with only a few minor exceptions at high values of geopolymer strength. This could be attributed to experimental variability. In addition, the regression slopes obtained from the straight-line equations shown in each plot (0.9758 for training set, 0.9476 for testing set, and for the overall, a value of 0.9689) is close to 1, shows that the model has high accuracy with respect to compressive strength forecasting.

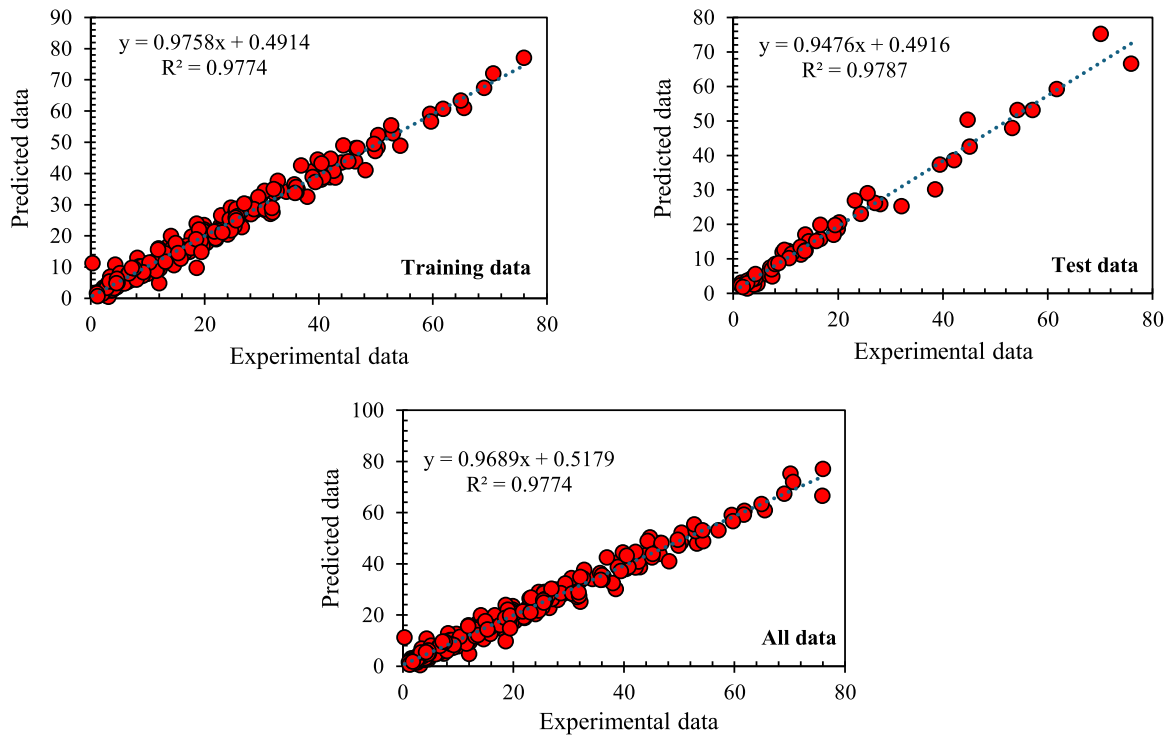


Fig. 6. Parity plots for trained model.

#### 4.5. Residual error analysis and model adequacy

To further establish the quality of the built model, residual plots shown in Fig. 7 were created for the data splits as well as for the entire data. The essence of this undertaking is to check for randomness, equal variance and linearity which are all characteristics of a good fitting model whose residuals are spread randomly around zero [92]. On the basis of these characteristics, the plots in Fig. 7 shows that the points around zero are randomly scattered with no discernible trend showing that the model was able to detect and account for the non-linear association between the predictors and geopolymer cement strength.

#### 4.6. Sensitivity analysis

Another veritable means of whitening the black box of neural nets to enhance its explainability is to establish how each independent variable affects the outcome of the prediction and the manner in which it does so ([93]; Niedbala et al., [94]). This is simply referred to as sensitivity analysis. The sensitivity of the independent variables in this study was established by using the relevancy factor ( $r$ ). The correlation behind this analysis is shown in Eq. (26).

$$r_i = \frac{\sum_{k=1}^D (x_{i,k} - \bar{x}_i) (CS_k - \bar{CS})}{\sqrt{\sum_{k=1}^D (x_{i,k} - \bar{x}_i)^2 \sum_{k=1}^D (CS_k - \bar{CS})^2}} \quad (26)$$

where  $r_i$  = relevancy factor of the  $i$ -th input,  $x_{i,k}$  = value of input variable  $i$  for data point  $k$ ,  $CS_k$  = forecasted compressive strength,  $\bar{x}_i$  &  $\bar{CS}$  = average values,  $D$  = data size. The outcome of the analysis is depicted in Fig. 8. This figure shows the strength (in fraction) and the direction of influence of each parameter on the network forecast. Curing temperature has the most positive relevance with a factor of 0.43 pointing to its influence in establishing geopolymer strength development. This is the same finding by Sajan et al. [95] wherein they

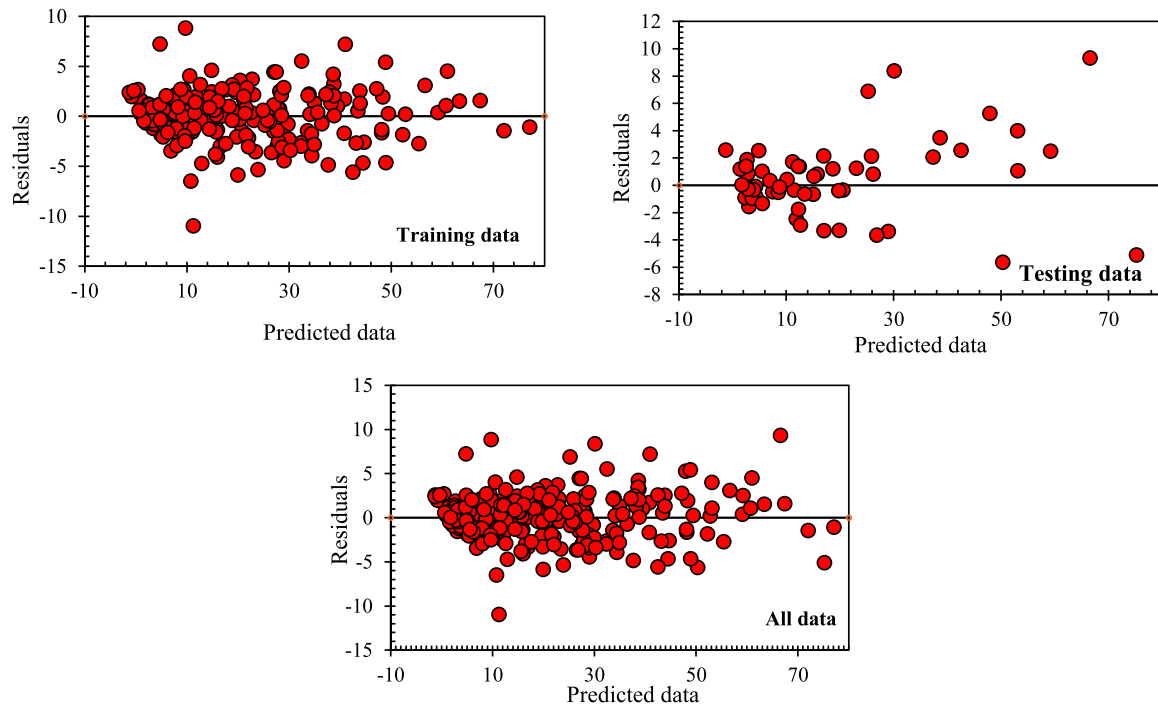


Fig. 7. Residual plots for trained model.

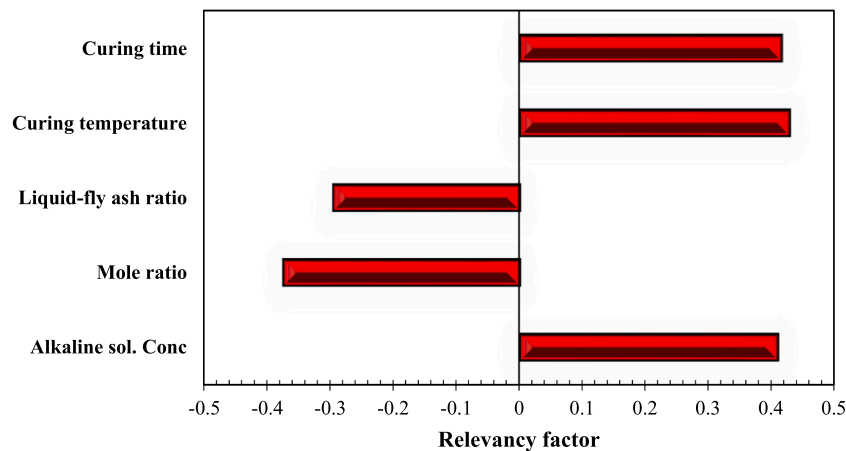


Fig. 8. Relevancy factor for model predictors.

concluded that curing temperature has the most effect on geopolymer strength. The work by Harooni et al. [45] and Siciliano et al. [96] supports that higher temperatures enhance geopolymer strength. Theoretically, the higher the temperature (subject to a threshold), the faster the aluminosilicates dissolve and a faster polymerization reaction [97]. The curing time trails the temperature in the ranking order with a positive factor of relevance of 0.417 implying that longer curing time helps in building strength. Castillo et al. [98] in an extensive review study corroborates this point. Theoretically, extended curing periods lead to continued geopolymerization, maturation of the gel and refinement of the pores which ultimately boosts the geopolymer strength [99]. The third ranking parameter is the alkaline solution concentration with a positive relevancy factor of 0.411. From the direction of its impact, an increase in its concentration yields higher geopolymer strength. El Fadili et al. [100] and Poojalakshmi et al. [101] in their study reported similar findings. In theory, the more the alkaline concentration, the faster the rate of dissolution of reactive Si and Al species, hence, resulting in gel network connectivity boosting and compressive strength [102].

Furthermore, the liquid – fly ash ratio has a negative relevance factor of -0.295 implying that the higher the ratio, the lower the strength of the geopolymer. This finding is corroborated by the earlier study by Wardhono et al. [103]. As already established in literature, this behaviour is the direct consequence of the liquid diluting the concentration of the activator, reducing the density of the matrix and enhancing capillary porosity. Finally, the mole ratio shows a negative association with the geopolymer strength with a factor of -0.374. This signifies that an increase in the mole ratio attenuates the strength of the geopolymer. Overall, this analysis aligns with the reaction mechanisms of geopolymer cements, thus giving credibility to the model proposed.

#### 4.7. Evaluation of the physics-consistency of the model via trend analysis

According to Osman et al. [104], a crucial factor in validating the physical consistency of a model is to conduct a trend analysis. In this study, the trend plot for each parameter were evolved by changing the values of the parameter while maintaining the others at their average

values. This way, the effect of each parameter on the geopolymer strength is evaluated for possible consistency with the established physicochemical attributes of geopolymer cement. Fig. 9 is the result of the analysis. The trends show that geopolymer strength is enhanced as alkaline solution concentration as well as curing time and temperature increases. On the other hand, the trends also indicate that mole ratio and liquid fly ash ratio increases yield lower geopolymer strength outcomes. The theories about underpinning these has been discussed extensively in the preceding section. These trends confirms that the model is not just governed by data but is also in sync with the fundamental mechanisms driving strength evolution in geopolymer cements.

#### 4.8. Outlier detection and model reliability assessment

To evaluate the validity of the built model, the Leverage plot was deployed. This is a plot of the standardized residuals against the leverage values represented by  $H$ . This is depicted in Fig. 10. This plot serves a

dual purpose: to identify influential points and outliers and the domain of model workability. To compute the  $H$ , Eq. (26) is used.

$$H = X(X^T X)^{-1} X^T \quad (26a)$$

In Eq. (26),  $X$  = input variable matrix. The model workability domain ( $H^*$ ) is computed using Eq. (27).

$$H^* = \frac{3(p+1)}{n} \quad (27)$$

For Eq. (27),  $p$  = input variables number;  $n$  = data size. Eq. (28) shows the correlation for the computation of the standardized residuals.

$$R_i = \frac{e_i}{s\sqrt{1-H_i}} \quad (28)$$

In Eq. (28),  $e_i$  = estimation error;  $s$  = residuals standard deviation.

Upon computing these parameters, Fig. 10 is the pictorial depiction of the Leverage plot. The plot is made of data points in green and some in

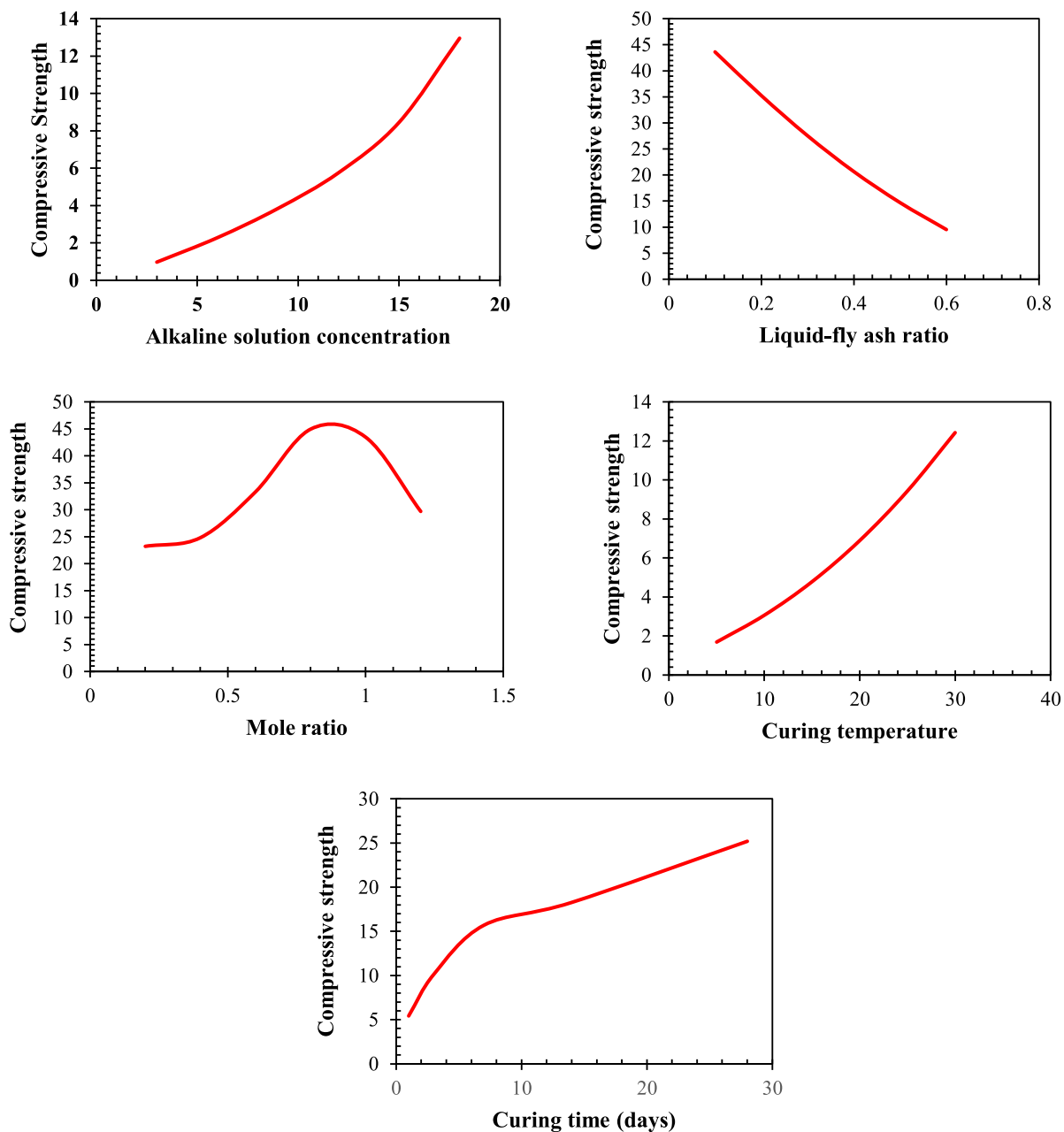


Fig. 9. Model prediction trend analysis.

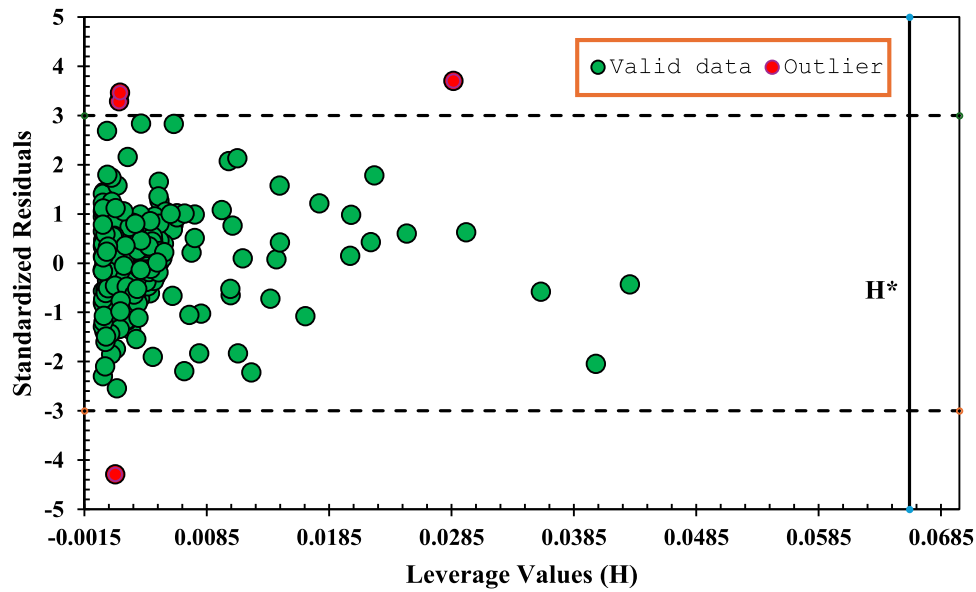


Fig. 10. Leverage plot for model application domain.

Table 9

Model comparison.

| Reference             | Algorithm | Performance                                             | Explicit model? | Sensitivity? | Trend analysis | Applicability domain |
|-----------------------|-----------|---------------------------------------------------------|-----------------|--------------|----------------|----------------------|
| Ling et al. [62]      | ANN       | $R^2 = 0.961$ , RMSE = 543, MAD = 355.3                 | No              | Yes          | Yes            | No                   |
| Rahmawati et al. [67] | ANN       | $R^2 = 0.941$ , RMSE = 0.192                            | No              | No           | No             | No                   |
| Tanyildizi [57]       | LSTM      | $R^2 = 0.992$ , RMSE = 1.5039, MSE = 2.2617             | No              | Yes          | No             | No                   |
| This study 2026)      | ANN       | $R^2 = 0.9787$ , RMSE = 2.867, MSE = 8.223, MAE = 2.048 | Yes             | Yes          | Yes            | Yes                  |

red colour. It can be observed that majority of the data points (green colour) fall within the acceptable residual boundary of  $-3 \leq R_i \leq +3$  while only four data points (red colour) fall outside this boundary. Given that the total number of data points is 273, therefore 269 data points (98.5%) are deemed valid points while only 1.5% of the data are considered outliers. The data that constitutes this 1.5% are likely due to uncertainties in experimentations or measurement noise. It should be noted that no data fell outside the critical leverage boundary  $H^*$  confirming that model forecasts are not dominated by a small subset of influential observations. In conclusion, the fact that valid data amounting to 98.5% of the dataset lies within the  $H_i < H^*$  &  $|R_i| < 3$  boundary is a pointer to the reliability and stability of the model.

#### 4.9. Comparison of model performance

To further validate the proposed model's efficacy, the reported accuracies of previous models built for geopolymers compressive strength is compared with the one obtained in this study. To this end, the performance and attributes of the three models isolated from the literature built specifically for geopolymers cement strength are shown in Table 9. It is observed from the table that earlier studies recorded reasonable model accuracies. Using the  $R^2$  as a yardstick, the ranking of the models (from best to least performing) are as follows: Tanyildizi [57] > This study > Ling et al. [62] > Rahmawati et al. [67]. In terms of model explainability and practical usability, the model developed in this study ranks better as it is explicitly presented, validated for physical consistency, delineated its workability domain and the contribution of each input is clearly outlined. These features make the model in this study a ready and deployable tool for geopolymers cement strength estimation for oilfield cementing purposes.

#### 4.10. Model field deployability pathway

A model is only valuable when it can be effectively deployed in practical applications. In this context, the most reliable pathway for utilizing the developed model is its implementation within a software environment, enabling seamless deployment in real-world field operations. To achieve this, the proposed explicit model can be integrated into existing wellbore cementing simulators by embedding it as a compressive strength prediction subroutine within the cement evaluation module. Accordingly, the model is operationalized through a structured computational workflow where the input parameters are obtained from slurry design data and downhole temperature profiles. The estimation of the geopolymers compressive strength proceeds as follows:

- The simulator computes compressive strength automatically using the explicit equation without requiring iterative numerical solving.
- The estimated compressive strength values are then compared against a predefined minimum safe compressive strength threshold (e.g. 3.45 MPa) required for well integrity before resuming drilling operations. Once the predicted compressive strength equals or exceeds this threshold, the corresponding curing time is identified as the WOC time.
- Sensitivity analysis can be performed by varying liquid-fly ash ratio and alkaline concentration to optimize slurry formulation.
- Finally, validation against field cement bond logs and laboratory strength tests ensure reliable real-time deployment and operational confidence.

## 5. Conclusion

This study was undertaken to aid in building a ready to use tool for estimating geopolymer cement strength for wellbore cementing purposes. To properly isolate the gaps that exist in the literature, a critical evaluation of documented paradigms for oilwell cement compressive strength was embarked upon. As a result, regression and ML based models built for this purpose were identified. The findings indicate that many models (both regression and ML based) exist for oil well cement strength prediction. However, the number of models for geopolymer cement strength estimation are relatively few. A close look at the ML models shows that they are mainly opaque, black boxes and are not explainable. These gaps propelled the motivation to undertake this study. In this regard, an accurate, white box, physics constrained model was built using Bayesian regularized neural network. Based on the findings made, the following are the main conclusions:

- An explicit, ready-to-use model was developed with a physics-based constraint introduced to preserve its physical reliability thereby providing a white-box alternative to existing black-box models.
- The proposed model demonstrates that high predictive performance can be achieved alongside interpretability. This addresses a key limitation in prior studies where accuracy often comes at the expense of transparency.
- The sensitivity evaluation of the inputs reveals the following: strength enhancement variables: curing temperature (+0.43) > curing time (+0.417) > alkaline concentration (+0.411); strength attenuating variables: mole ratio (-0.37) and liquid fly ash ratio (-0.295). This provides actionable engineering insights, unlike many ML models that only give predictions.
- Trend analysis shows that the model preserves the underpinning physicochemical mechanisms of geopolymerization associated with geopolymer cements. This ensures physical reliability, improving confidence in model predictions
- This study establishes a modelling framework that can be extended with larger datasets and advanced techniques. This provides a scalable foundation for future developments in intelligent cementing systems.

## CRedit authorship contribution statement

**Okorie Ekwe Agwu:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Saad Alatefi:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology. **Muhammad Aslam Md Yusof:** Writing – review & editing, Visualization, Validation, Supervision, Formal analysis. **Nwode Agwu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation.

## Declaration of competing interest

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

## Acknowledgements

This research was carried out with funding support from Yayasan Universiti Teknologi Petronas-Fundamental Research Grant (YUTP-FRG) (Grant No. 015LC0-585) and the Postgraduate and Research office of Universiti Teknologi PETRONAS, which is sincerely appreciated. The authors also wish to express their gratitude to the Department of Petroleum Engineering and the Centre for GeoEnergy Resources and

Intelligence (GERI), Universiti Teknologi PETRONAS, Malaysia, for providing the academic and research environment that facilitated this study.

## Data availability

Data will be made available on request.

## References

- [1] M.F. Fakoya, S.N. Shah, Emergence of nanotechnology in the oil and gas industry: emphasis on the application of silica nanoparticles, *Petrol. Times* 3 (2017) 391–405.
- [2] J.C. Newlove, R.C. Portnoy, D.N. Schulz, K. Kitano, Fluid loss control in oilfield cements, 480, United States Patent, 1984, p. 693, 1984.
- [3] N. Yousuf, O. Olayiwola, B. Guo, N. Liu, A comprehensive review on the loss of wellbore integrity due to cement failure and available remedial methods, *J. Petroleum Sci. Eng.* 207 (2021) 109123, <https://doi.org/10.1016/j.petrol.2021.109123>. December 2021.
- [4] N. Agbasimalo, Experimental Study of the Effect of Drilling Fluid Contamination on the Integrity of Cement-Formation Interface, Department of Pet Eng, Louisiana State University, USA, 2012. MSc Thesis.
- [5] Izon, D. and Mayes, M. (2007) Absence of fatalities in blowouts encouraging in MMS study of OCS incidents 1992–2006 Well Control 84–90.
- [6] C. Vipulanandan, M. Heidari, Q. Qu, H. Farzam, J.M. Pappas, Behaviour of piezoresistive smart cement contaminated with oil-based drilling mud Offshore Technology Conf (2014). OTC 25200-MS1-14.
- [7] K.N. Shivaprasad, H.-M. Yang, J.K. Singh, A path to carbon neutrality in construction: an overview of recent progress in recycled cement usage, *J. CO2 Util.* 83 (2024) 102816, <https://doi.org/10.1016/j.jcou.2024.102816>.
- [8] D. Aceituno, X. Zhang, H. Hao, A comprehensive review on carbon utilization pathways in concrete from conventional to improved strategies, *Carbon Capture Sci. Technol.* 16 (2025) 100467, <https://doi.org/10.1016/j.cst.2025.100467>.
- [9] Yakovlev, A., Lafitte, V., Kapoor, S., Meade, M., Smith, W., Fischer, N., Parton, C., McDonald, V. R., Nenjerama, Y., Wei, X., & Landry, G. (2023, June). An innovative and sustainable Portland cement alternative for oil and gas wells. In Proceedings of the SPE International Conference on Oilfield Chemistry (Paper No. SPE-213805-MS), The Woodlands, TX, United States. <https://doi.org/10.2118/213805-MS>.
- [10] D.N. Huntzinger, T.D. Eatmon, A life-cycle assessment of portland cement manufacturing: comparing the traditional process with alternative technologies, *J. Clean. Prod.* 17 (7) (2009) 668–675, <https://doi.org/10.1016/j.jclepro.2008.04.007>.
- [11] L. Marini, Predicting the compressive strength of ultra-high strength geopolymer concrete using multiple linear regression, *IJRVOCAS* 5 (4) (2026) 14–19, <https://doi.org/10.53893/ijrvocas.v5i4.483>.
- [12] Pearce, F. (2021). Can the world's most polluting heavy industries decarbonize? YaleEnronment360. Yale School of Environment. <https://e360.yale.edu/features/can-the-worlds-most-polluting-heavy-industries-decarbonize> (Date accessed: 10/3/2026).
- [13] F. El Inaty, H. Nasreddine, A. Djerbi, L. Gautron, M. Marchetti, M. Quiertant, O. O. Metalssi, Mechanical and durability performance of metakaolin and fly ash-based geopolymers compared to cement systems, *Results. Eng.* 27 (2025) 105788, <https://doi.org/10.1016/j.rineng.2025.105788>.
- [14] F. Aslani, Y. Zhang, D. Manning, Valdez, L. Carvajal, N. Manning, Additive and alternative materials to cement for well plugging and abandonment: A state-of-the-art review, *J. Petrol. Sci. Eng.*, 215 (2022), <https://doi.org/10.1016/j.petrol.2022.110728> (Part B), Article 110728.
- [15] A. Nazari, J.G. Sanjayan, Modelling of compressive strength of geopolymer paste, mortar and concrete by optimized support vector machine, *Ceram. Int.* 41 (9) (2015) 12164–12177, <https://doi.org/10.1016/j.ceramint.2015.06.037>. Pt. B).
- [16] A.A. Ogiegnagbon, M. Khalifeh, Experimental evaluation of the effect of temperature on the mechanical properties of setting materials for well integrity, *SPE* (5) (2022) 2577–2589, <https://doi.org/10.2118/209794-PA>. J27.
- [17] M.D.M. Paiva, E.C.C.M. Silva, D.M.A. Melo, A.E. Martinelli, J.F. Schneider, A geopolymer cementing system for oil wells subject to steam injection, *J. Petrol. Sci. Eng.* 169 (2018) 748–759, <https://doi.org/10.1016/j.petrol.2018.06.022>.
- [18] J. Provis, J.S.J. Deventer, Geopolymers: Structure, Processing, Properties, and Industrial Applications, CRC Press, 2009. ISBN: 9781845694494.
- [19] A.A. Abdulhameed, A.H. Al-Zuhairi, S.R. Al Zaidee, A.N. Hanoon, A.W. Al Zand, M.M. Hason, H.A. Abdulhameed, The behavior of hybrid fiber-reinforced concrete elements: A new stress-strain model using an evolutionary approach, *Appl. Sci.* 12 (4) (2022) 2245, <https://doi.org/10.3390/app12042245>.
- [20] R.A. Kodah, S.R. Rabab'ah, H.H. Aldeeky, N.D. Lagaros, K.M. Abdalla, Geopolymer stabilization of expansive soils using zeolitic tuff: A sustainable alternative to cement-based stabilization, *Results. Eng.* 26 (2025) 105386, <https://doi.org/10.1016/j.rineng.2025.105386>.
- [21] Z. Mohammed, A.A. Abdulhameed, H.K. Kazim, Effect of alkali - activated natural pozzolan on mechanical properties of geopolymer concrete, *Civil Environ. Eng.* 18 (1) (2022) 312–320, <https://doi.org/10.2478/cee-2022-0029>.
- [22] B. Money, R.H. Modather, D. Abutu, N. Ismail, A. Agi, F. Nyah, S.Q.A. Mahat, A review on advancing clay-based geopolymers for high-temperature oil well

- cements: Mechanisms, durability, and applications, *Petrol. Res.* (2025), <https://doi.org/10.1016/j.ptlrs.2025.08.005>.
- [23] P. Duxson, J.L. Provis, G.C. Lukey, et al., The role of inorganic polymer technology in the development of 'green concrete', *Cem. Concr. Res.* 37 (12) (2007) 1590–1597, <https://doi.org/10.1016/j.cemconres.2007.08.018>.
- [24] M.C.G. Juenger, F. Winnefeld, J.L. Provis, et al., Advances in alternative cementitious binders, *Cem. Concr. Res.* 41 (12) (2011) 1232–1243, <https://doi.org/10.1016/j.cemconres.2010.11.012>.
- [25] Abdelal, A. E. A. I. (2023). Development of geopolymers systems for oil and gas well cementing. Doctoral dissertation, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia. 216 pp.
- [26] B.C. McLellan, R.P. Williams, J. Lay, et al., Costs and carbon emissions for geopolymer pastes in comparison to ordinary portland cement, *J. Clean. Prod.* 19 (9–10) (2011) 1080–1090, <https://doi.org/10.1016/j.jclepro.2011.02.010>.
- [27] Adnan Adnan, M. Anas, Geopolymer concrete as a sustainable alternative to OPC, *J. Umm Al-Qura Univ. Eng. Archit.* 16 (2025) 1223–1245, <https://doi.org/10.1007/s43995-025-00155-8>.
- [28] Y. Zhai, Y. Zhang, M. Zhang, Q. Dong, Q. Li, S. Wei, Compressive strength and intelligent prediction modeling of early-strength cement-based grouting materials considering age, shape, and size effects, *Mater. Today Commun.* 50 (2026) 114634, <https://doi.org/10.1016/j.mtcomm.2026.114634>.
- [29] D. Khale, R. Chaudhary, Mechanism of geopolymerization and factors influencing its development: a review, *J. Mater. Sci.* 42 (2007) 729–746, <https://doi.org/10.1007/S10853-006-0401-4>.
- [30] S.H.A. Rahman, N.N. Zulkarnain, N. Shafiq, Experimental study and design of experiment using statistical analysis for the development of geopolymer matrix for oil-well cementing for enhancing the integrity, *Crystals* 11 (2) (2021) 139, <https://doi.org/10.3390/cryst11020139>.
- [31] Khalifeh, M., Saasen, A., Vrålstad, T., and Hodne, H. (2014). Potential utilization of class C fly ash-based geopolymer in oil well cementing operations. *Cement Concr. Compos.*, 53 10–17, <https://doi.org/10.1016/j.cemconcomp.2014.06.014>.
- [32] S. Salehi, M.J. Khattak, N. Ali, C. Ezeakacha, F.K. Saleh, Study and use of geopolymer mixtures for oil and gas well cementing applications, *J. Energy Resour. Technol.* 140 (1) (2018) 012908, <https://doi.org/10.1115/1.4037713>.
- [33] M. Ahdaya, A. Imqam, Fly ash class C based geopolymer for oil well cementing, *J. Petrol. Sci. Eng.* 179 (2019) 750–757, <https://doi.org/10.1016/j.petrol.2019.04.106>, August.
- [34] S.E. Ridha, Geopolymer cement slurries with enhanced mechanical and physical properties for oilwell cementing, in: Paper presented at the 5th Brunei International Conference on Engineering and Technology, November, 2014, <https://doi.org/10.1049/cp.2014.1095>.
- [35] R.R. Suppiah, S.H.A. Rahman, N. Shafiq, et al., Uniaxial compressive strength of geopolymer cement for oil well cement, *J. Petrol. Explor. Prod. Technol.* 10 (2020) 67–70, <https://doi.org/10.1007/s13202-019-0704-z>.
- [36] S.H. Rahman, A. Medvedev, A. Yakovlev, et al., Development of new geopolymer-based system for challenging well conditions, in: Paper presented at the International Petroleum Technology Conference IPTC-21371-MS, 2021, <https://doi.org/10.2523/IPTC-21371-MS>.
- [37] V.S. Denduluri, G. Il Ulerio, M. Genedy, M. Juenger, E. van Oort, Experimental investigation of geopolymers for application in high-temperature and geothermal well cementing, *SPE Drilling Completion* 38 (3) (2023) 433–443, <https://doi.org/10.2118/212491-PA>.
- [38] Devers, C. (2024). Testing, characterization, and analysis of Class F fly ash geopolymers under oilfield testing conditions. Doctoral dissertation, University of Oklahoma. Norman, OK, United States. 207 p.
- [39] S.H. Hajiabadi, R.I. Korsnes, A.M. Dale, A.P. Apolinario, M. Khalifeh, Applicability of Portland cement standards to granite-based one-part geopolymers: a rock mechanical perspective, *Results. Eng.* 28 (2025) 108187, <https://doi.org/10.1016/j.rineng.2025.108187>.
- [40] S. Pang, M. Li, W. Ni, J. Li, H. Fan, M. Meng, X. Zhang, Y. Gao, Study on spatial deformation mechanism of fiber-EPS optimized geopolymer cemented aeolian sand (GCAS), *Results. Eng.* 27 (2025) 106217, <https://doi.org/10.1016/j.rineng.2025.106217>.
- [41] S.M.S. da Silva, M. Taborda-Barraza, F. de Andrade Silva, et al., Preliminary formulations of geopolymers for oil well cementing application using metakaolin, fly ash and granulated blast furnace slag combinations, *Discov. Sustain.* 7 (2026) 5, <https://doi.org/10.1007/s43621-025-02525-7>, 2026.
- [42] Ouadi, H., Laalam, A., Chemmakh, A., Merzoug, A., Mouedden, N., Rasouli, V., Djezzar, S., Boualam, A. and Hassan, A. (2022). Comparative analysis between different artificial intelligence based models optimized with genetic algorithms for the prediction of oilfield cement compressive strength. Paper ARMA-2022-0279 presented at the 56th U.S. Rock Mechanics/Geomechanics Symposium, Santa Fe, New Mexico, USA, June 2022. <https://doi.org/10.56952/ARMA-2022-0279>.
- [43] E.E. Nyakilla, G. Jun, G. Charles, E.X. Ricky, W. Hussain, S.M. Iqbal, D. C. Kalibwami, A.G. Alareqi, M. Shaame, M.R. Ngata, Application of group method of data handling via a modified levenberg-marquardt algorithm in the prediction of compressive strength of oilwell cement with reinforced fly ash based on experimental data, *SPE Drill Completion* 38 (03) (2023) 452–468, <https://doi.org/10.2118/214679-PA>.
- [44] B.W. Chong, R. Othman, R.P. Jaya, M.R.M. Hasan, A.V. Sandu, M. Nabialek, B. Jež, P. Pietrusiewicz, D. Kwiatkowski, P. Postawa, M.M.A. Abdullah, Design of experiment on concrete mechanical properties prediction: a critical review, *Materials* 14 (8) (2021) 1866, <https://doi.org/10.3390/ma14081866>, 2021.
- [45] Harooni, A.B., Maroof, H., Pourali, O. and Khoshmardan, M.A. (2025). A novel predictive model for the compressive strength development in class G cement slurries at different temperatures, retarder, and salt concentrations. *Case Studies in Construction Materials*, Volume 22, July 2025, e04360. <https://doi.org/10.1016/j.cscm.2025.e04360>.
- [46] I. Hidayat, A.F.S. Rahardhani, G.A. Suhardjo, Utilizing hammer tests and ultrasonic pulse velocity to ascertain the compressive strength of concrete, *IOP Conf. Ser. Earth Environ. Sci.* 1324 (2024), <https://doi.org/10.1088/1755-1315/1324/1/012006>, 7th International Conference on Eco Engineering Development 2023 08/11/2023 - 08/11/2023.
- [47] L. Wang, S. Huang, Z. Li, D. Su, Y. Liu, Y. Shi, Hybrid machine learning model based on GWO and PSO optimization for prediction of oilwell cement compressive strength under acidic corrosion. drilling & completion, *SPE J* 29 (09) (2024) 4684–4695, <https://doi.org/10.2118/221485-PA>.
- [48] T.H. Xian, *Geopolymer-based oil well cementing systems using silica fume* (Undergraduate thesis, Bachelor of Engineering [Hons] in Petroleum Engineering). Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak, Malaysia, 2013.
- [49] Ahdaya, M. S., Imqam, A., Jani, P., Fakher, S. and ElGawady, M. (2019). *New formulation of fly ash Class C-based geopolymer for oil well cementing*. In Proceedings of the International Petroleum Technology Conference (Paper No. IPTC-19393-MS), Beijing, China. <https://doi.org/10.2523/IPTC-19393-MS>.
- [50] F.D. Gomado, M. Khalifeh, M. Rafique, Harnessing waste glass as an expansive agent controller for enhanced Geopolymer performance in oil and gas wells, *Geoenergy Sci. Eng.* 243 (2024) 1–11, <https://doi.org/10.1016/j.geoen.2024.213299>.
- [51] A.A. Ogiegnagbon, M. Khalifeh, Development of geopolymer as treatment fluid for remedial operations—Characterization and comparison with state-of-the-art materials, *SPE J* 30 (9) (2025) 5501–5515, <https://doi.org/10.2118/228407-PA>.
- [52] Abd, A.M. and Abd, S.M. (2017). Modelling the strength of lightweight foamed concrete using support vector machine (SVM). *Case Studies in Construction Materials*, 6:8–15.
- [53] O.A. Falode, K.K. Salam, A.O. Arinkoola, B.M. Ajagbe, Prediction of compressive strength of oil field class G cement slurry using factorial design, *J. Petrol. Explor. Prod. Technol.* 3 (2013) 297–302, <https://doi.org/10.1007/s13202-013-0071-0>.
- [54] A. Anya, H. Emadi, M. Watson, An empirical model for calculating uniaxial compressive strength of oil well cements from ultrasonic pulse transit time measurements, *J. Petrol. Sci. Eng.* 183 (2019) 1–12, <https://doi.org/10.1016/j.petrol.2019.106387>.
- [55] O.F. Joel, F.T. Ademiluyi, Modelling of compressive strength of cement slurry at different slurry weights and temperatures, *Res. J. Chem. Sci.* 1 (2) (2011) 128–134.
- [56] N.N. Zulkarnain, N. Shafiq, S.A. Farhan, S.H.A. Rahman, Modelling the compressive strength development of geopolymer well cement containing antifoam, *J. Hunan Univ. Nat. Sci.* 51 (4) (2024) 1–10, <https://doi.org/10.55463/issn.1674-2974.51.4.1>.
- [57] H. Tanyildizi, Deep learning-based prediction of compressive strength of eco-friendly geopolymer concrete, *Environ. Sci. Pollut. Res.* 31 (2024) 41246–41266, <https://doi.org/10.1007/s11356-024-33853-2>, 2024.
- [58] S. Ridha, S. Irawan, B. Ariwahjoedi, Strength prediction of Class G oilwell cement during early ages by electrical conductivity, *J. Petrol. Explor. Prod. Technol.* 3 (2013) 303–311, <https://doi.org/10.1007/s13202-013-0075-9>.
- [59] D.M. Degefu, Z. Liao, U. Berardi, H. Doan, Salient parameters affecting the performance of foamed geopolymers as sustainable insulating materials, *Construct. Build. Mater.* 313 (2021), <https://doi.org/10.1016/j.conbuildmat.2021.125400>.
- [60] Vignesh, J., Ramesh, B. and Xavier, J.R. (2025). A comprehensive review of materials, processing, and performance of nano-doped engineered geopolymer composites for construction applications. *Case Stud. Constr. Mater.*, Volume 23, December 2025, e05625. <https://doi.org/10.1016/j.cscm.2025.e05625>.
- [61] P.D. McElroy, H. Bibang, H. Emadi, Y. Kocoglu, A. Hussain, M.C. Watson, Artificial neural network (ANN) approach to predict unconfined compressive strength (UCS) of oil and gas well cement reinforced with nanoparticles, *J. Nat. Gas. Sci. Eng.* 88 (2021) 1–12, <https://doi.org/10.1016/j.jngse.2021.103816>.
- [62] Y. Ling, K. Wang, X. Wang, W. Li, Prediction of engineering properties of fly ash-based geopolymer using artificial neural networks, *Neural Comput. Appl.* 33 (2021) 85–105, <https://doi.org/10.1007/s00521-019-04662-3>, 2021.
- [63] E.E. Nyakilla, G. Jun, N.A. Kasimu, E.F. Robert, N. Innocent, T. Mohamedy, M. Shaame, M.R. Ngata, P.E. Mabeyo, Application of machine learning in the prediction of compressive and shear bond strengths from the experimental data in oil well cement at 80°C. Ensemble trees boosting approach, *Constr. Build. Mater.* 317 (2022) 1–15, <https://doi.org/10.1016/j.conbuildmat.2021.125778>.
- [64] Y. Wen, Z. Chen, Y. He, H. Liu, Z. Zhang, L. Liu, R. Meng, Y. Zeng, Predictive methods for the evolution of oil well cement strength based on porosity, *Mater. Struct.* 57 (215) (2024) 1–21, <https://doi.org/10.1617/s11527-024-02493-w>.
- [65] J.J. Abraham, C. Devers, C. Teodoriu, M. Amani, Scaling Field and Experimental Data Using Machine Learning Approaches to Evaluate Oilwell Cement Degradation, Stability and Integrity for CCUS Applications, a, Paper SPE-219115-MS presented at the GOTECH, Dubai, UAE, 2024, <https://doi.org/10.2118/219115-MS>, May 2024.
- [66] Abraham, J.J., Devers, C., Carvero, A., Teodoriu, C. and Amani, M. (2024b). Identifying Trends and Predicting Mechanical Properties in Large Scale Cement Datasets with Machine Learning. Paper SPE-218646-MS presented at the SPE Conference at Oman Petroleum & Energy Show, Muscat, Oman, April 2024. <https://doi.org/10.2118/218646-MS>.
- [67] C. Rahmawati, S. Aisyah, Iqbal Sanusi, M.M. Maulana, Erdiwansyah Erdiwansyah, J. Ahmad, Artificial Intelligence models for predicting

- the compressive strength of geopolymers cements, *Civil Eng. J.* 10 (2024) 37–50, <https://doi.org/10.28991/CEJ-SP2024-010-03>.
- [68] A. Hussain, A. Addo-Yobo, H. Emadi, et al., Compressive strength prediction of ultra lightweight cement using data-driven modeling techniques, *Multiscale and Multidiscip. Model. Exp. and Des.* 8 (2025) 275, <https://doi.org/10.1007/s41939-025-00863-7>, 2025.
- [69] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhamma, J.M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez, F. Herrera, Explainable Artificial Intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence, *Info. Fusion* 99 (2023) 101805, <https://doi.org/10.1016/j.inffus.2023.101805>, November 2023.
- [70] A. Saranya, R. Subhashini, A systematic review of Explainable Artificial Intelligence models and applications: recent developments and future trends, *Decis. Anal. J.* 7 (2023) 100230, <https://doi.org/10.1016/j.dajour.2023.100230>, June 2023.
- [71] S. Kruschel, N. Hambauer, S. Weinzierl, et al., Challenging the performance-interpretability trade-off: an evaluation of interpretable machine learning models, *Bus. Inf. Syst. Eng.* 68 (2026) 159–183, <https://doi.org/10.1007/s12599-024-00922-2>.
- [72] S. Alatefi, O.E. Agwu, M.N. Amar, H. Djema, Explainable artificial intelligence models for estimating the heat capacity of deep eutectic solvents, *Fuel* 394 (2025) 135073, <https://doi.org/10.1016/j.fuel.2025.135073>, 15 August 2025.
- [73] V. Arabzadeh, R. Frank, A four-dimensional analysis of explainable AI in energy forecasting: a domain-specific systematic review, *Machine Learn. Knowl. Extract* 7 (4) (2025) 153, <https://doi.org/10.3390/make7040153>.
- [74] T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D.W. Roberts, T.W. Schultz, D.T. Stanton, J.J.M van de Sandt, W. Tong, G. Veith, C. Yang, Current status of methods for defining the applicability domain of (quantitative) structure activity relationships, *Altern. Lab. Anim.* 33 (2) (2005) 155–173, <https://doi.org/10.1177/026119290503300209>.
- [75] M. Priestley, F. O'donnell, E. Simperl, A Survey of data quality requirements that matter in ML development pipelines, *J. Data Info. Quality* 15 (2) (2023) 139, <https://doi.org/10.1145/3592616>.
- [76] S. Roy, P. Sharma, K. Nath, D.K. Bhattacharyya, J.K. Kalita, Pre-processing: A data preparation step, in: S. Ranganathan, K. Nakai, C. Schönbach (Eds.), *Pre-processing: A data preparation step, Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics 1* (2018) 463–471, <https://doi.org/10.1016/B978-0-12-809633-8.20458-5>, Elsevier.
- [77] B. Calabrese, Data cleaning, in: S. Ranganathan, M. Nakai, C. Schönbach (Eds.), *Data cleaning, Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics 1* (2018) 472–476, <https://doi.org/10.1016/B978-0-12-809633-8.20360-9>, Elsevier.
- [78] V. Cetin, O. Yildiz, A comprehensive review on data preprocessing techniques in data analysis, *Pamukkale Univ. J. Eng. Sci.* 28 (2) (2022) 299–312, <https://doi.org/10.5505/pajes.2021.62687>.
- [79] B. Dharmalingam, S. Annamalai, S. Areeya, K. Rattanaporn, K. Katam, P.L. Show, M. Sriariyanun, Bayesian regularization neural network-based machine learning approach on optimization of CRDI-split injection with waste cooking oil biodiesel to improve diesel engine performance, *Energies* 16 (6) (2023) 2805, <https://doi.org/10.3390/en16062805>, 2023.
- [80] S.T. Nguyen, T.M. Pham, A. Hoang, L.V. Trieu, T.T. Cao, Bayesian inference for regularization and model complexity control of artificial neural networks in classification problems, *Bayesian Inference - Recent Trends*, IntechOpen, 2023, <https://doi.org/10.5772/intechopen.1002652>.
- [81] D. Garg, A. Mishra, Bayesian regularized neural network decision tree ensemble model for genomic data classification, *Appl. Artif. Intell.* 32 (5) (2018) 463–476, <https://doi.org/10.1080/08839514.2018.1483115>.
- [82] Pan, G., Qiao, J., Chai, W. and Dimopoulos, N. (2014). An improved RBM based on Bayesian Regularization. Paper presented at the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 2014, pp. 2935–2939, <https://doi.org/10.1109/IJCNN.2014.6889458>.
- [83] S. Jun, Bayesian neural networks with regularization for sparse zero-inflated data modeling, *Information* 17 (1) (2026) 81, <https://doi.org/10.3390/info17010081>, 2026.
- [84] F. Burden, D. Winkler, Bayesian regularization of neural networks, *Artificial Neural Networks*, in: D.J. Livingstone (Ed.), *Bayesian regularization of neural networks*, *Methods Mol. Biol.* 458 (2008), [https://doi.org/10.1007/978-1-60327-101-1\\_3](https://doi.org/10.1007/978-1-60327-101-1_3).
- [85] V. Vendittoli, W. Polini, M.S.J. Walter, S. Geißelsöder, Using Bayesian regularized artificial neural networks to predict the tensile strength of additively manufactured polylactic acid parts, *Appl. Sci.* 14 (8) (2024) 1–15, <https://doi.org/10.3390/app14083184>.
- [86] W.F. Gemechu, W. Sitek, G.F. Batalha, Improving hardenability modeling: a bayesian optimization approach to tuning hyperparameters for neural network regression, *Appl. Sci.* 14 (6) (2024) 2554, <https://doi.org/10.3390/app14062554>, 2024.
- [87] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, *PeerJ. Comput. Sci.* 7 (2021) e623, <https://doi.org/10.7717/peerj-cs.623>.
- [88] K. Tyagi, C. Rane, Manry, M. Harshvardhan, Regression analysis. Artificial intelligence and machine learning for edge computing, Elsevier, 2022, pp. 53–63, <https://doi.org/10.1016/B978-0-12-824054-0.00007-1>.
- [89] R.J. Hyndman, A.B. Koehler, Another look at measures of forecast accuracy, *Int. J. Forecast.* 22 (4) (2006) 679–688, <https://doi.org/10.1016/j.ijforecast.2006.03.001>.
- [90] H. Belyadi, A. Haghighat, Supervised learning. Machine learning guide for oil and gas using Python: A step-by-step breakdown with data, algorithms, codes, and applications, Elsevier, 2021, pp. 169–295, <https://doi.org/10.1016/B978-0-12-821929-4.00004-4>.
- [91] A.G. Adeniyi, C.A. Igwegbe, J.O. Ighalo, ANN modelling of the adsorption of herbicides and pesticides based on sorbate-sorbent interphase, *Chem. Africa* 4 (2021) 443–449, <https://doi.org/10.1007/s42250-020-00220-w>.
- [92] V. Verma, A comprehensive framework for residual analysis in regression and machine learning, *J. Info. Syst. Eng. Manage.* 10 (31) (2025) 34–46.
- [93] M.J. Ankenbrand, L. Shainberg, M. Hock, D. Lohr, L.M. Schreiber, Sensitivity analysis for interpretation of machine learning based segmentation models in cardiac MRI, *BMC. Med. Imaging* 21 (1) (2021) 27, <https://doi.org/10.1186/s12880-021-00551-1>.
- [94] G. Niedbala, D.K. Popowska, M. Piekutowska, T. Wojciechowski, M. Kwiatek, J. Nawracala, Application of artificial neural network sensitivity analysis to identify key determinants of harvesting date and yield of soybean (*Glycine max* [L.] Merrill) *Cultivar Augusta*, *Agriculture* 12 (6) (2022) 754, <https://doi.org/10.3390/agriculture12060754>, 2022.
- [95] P. Sajan, T. Jiang, C.K. Lau, G. Tan, K. Ng, Combined effect of curing temperature, curing period and alkaline concentration on the mechanical properties of fly ash-based geopolymer, *Clean. Mater.* 1 (2021) 100002, <https://doi.org/10.1016/j.clema.2021.100002>.
- [96] U.C.C.S. Siciliano, J. Zhao, A.C.C. Trindade, M. Liebscher, V. Mechtcherine, F. Silva, A. de, Influence of curing temperature and pressure on the mechanical and microstructural development of metakaolin-based geopolymers, *Constr. Build. Mater.* 424 (2024) 135852, <https://doi.org/10.1016/j.conbuildmat.2024.135852>.
- [97] Y. Yu, F. Xu, S. Wang, L. Fan, J. Zhang, P. Li, L. Yu, Influence mechanism of curing temperature on geopolymerization reaction: a comprehensive review, *J. Build. Eng.* 103 (2025) 112195, <https://doi.org/10.1016/j.job.2025.112195>, 1 June 2025.
- [98] H. Castillo, H. Collado, T. Drogue, S. Sánchez, M. Vesely, P. Garrido, S. Palma, Factors affecting the compressive strength of geopolymers: a review, *Minerals* 11 (12) (2021) 1317, <https://doi.org/10.3390/min11121317>, 2021.
- [99] A.M. Alnadish, Geopolymer concrete: a comprehensive review of current research trends, challenges, and environmental-economic implications, *Arch. Comput. Methods Eng.* (2026), <https://doi.org/10.1007/s11831-026-10515-9>, 2026.
- [100] El Fadili, H., Ait-Khouia, Y., Ouffa, N., Taha, Y., Moukannaa, S., Benzaazoua, M. (2025). Comprehensive review on alkaline dissolution of aluminosilicates and its role in the mechanisms and properties of geopolymers and alkali-activated materials. *Case Studies in Construction Materials*, Volume 23, December 2025, e05330, <https://doi.org/10.1016/j.cscm.2025.e05330>.
- [101] E.S. Poojalakshmi, P. Nagarajan, J. Sudhakumar, B.S. Thomas, Impact of alkaline activator concentration on mechanical properties and microstructure of a ternary blended one-part geopolymer cement, *Sci. Rep.* 15 (1) (2025) 33808, <https://doi.org/10.1038/s41598-025-01610-1>.
- [102] L. Spitzmüller, F. Nitschke, B. Rudolph, J. Berson, T. Schimmel, T. Kohl, Dissolution control and stability improvement of silica nanoparticles in aqueous media, *J. NanoPart Res.* 25 (2023) 1–17, <https://doi.org/10.1007/s11051-023-05688-4>, 2023.
- [103] A. Wardhono, Y. Risdianto, K.D. Handayani, N. Andajani, D.W. Law, The role of water-to-binder ratio on strength of fly ash geopolymer as sustainable green materials in supporting SDG 13, *E3S Web Conf.* 640 (2025) 1–6, <https://doi.org/10.1051/e3sconf/202564001005>.
- [104] Osman, E.A., Ayoub, M. A., Aggour M. A. (2005). An Artificial Neural Network Model for Predicting Bottomhole Flowing Pressure in Vertical Multiphase Flow. Paper SPE-93632-MS presented at the SPE Middle East Oil and Gas Show and Conference, Kingdom of Bahrain, March 2005. <https://doi.org/10.2118/93632-MS>.