ORIGINAL ARTICLE

Modelling the edge breakout shear capacity of single anchors using gene expression programming

Oladimeji Benedict Olalusi[1] · Avishkar Durgapershad[1] · Paul Oluwaseun Awoyera[2] · John Temitope Kolawole[3,4]

Abstract
The use of soft computing techniques is becoming more common in providing solutions to complex engineering problems such as the concrete breakout strength of anchor. Available techniques include semi-empirical equations that are known to over or underpredict and some soft computing techniques that is incapable of generating predictive equations. This study proposes a gene expression programming (GEP)-based mathematical model to predict the concrete edge breakout capacity of single anchors loaded in shear. In doing so, an experimental database compiled by the American Concrete Institute (ACI) Committee 355, containing 366 samples, was used for the model training and testing. The independent variables considered in the model development are the edge distance, anchor diameter, embedment depth and concrete strength. Moreover, the predictive performance of the developed model was compared to that of the existing models proposed in ACI 318 and the Eurocode 2 (EC2) design standards. The assessment showed that the proposed GEP-based model provided a much more uniform and accurate prediction of the actual strength than the models in the existing design standards. The proposed mathematical model is simple and robust and is expected to be very useful for evaluating the concrete breakout shear capacity of single anchors in pre-planning and pre-design phases; that is, towards inclusions in design standards.

1 Introduction

The anchorage system or fastening system remains an integral part of a structural system. Some typical applications of anchorages are connections between beam to column, beam to wall and column to foundation. Anchorage systems are also commonly used in non-structural engineering applications such as mechanical and electrical

engineering yet serve the same general purpose—fastening two or more elements together. Therefore, it is imperative to design anchorages in a cost-effective and yet reliable manner. The main action forces experienced by load-bearing anchors are axial tension and shear. Possible failure modes of anchors loaded in shear are failures in the materials used in an anchorage system. Steel related failures may be steel ruptures, prying out failure, bending of the steel anchor. Experimental studies have shown concrete failure modes of anchorage systems under shear to be crushing of concrete at the loading region and concrete edge breakout failure [1]. Concrete failures pose a serious safety issue due to the brittle nature of concrete. This leads to a failure mode which develops abruptly with no preceding signs of failure [2].

Consequently, a predictive model providing a reliable prediction of concrete related failure needs to be used in design. Most of the existing design models available to predict the concrete edge breakout resistance of single anchors loaded in shear are semi-empirical models. This

✉ Oladimeji Benedict Olalusi
olalulsio@ukzn.ac.za

1   Department of Civil Engineering, Structural Engineering & Computational Mechanics Group (SECM), University of KwaZulu-Natal, Durban, South Africa

2   Department of Civil Engineering, Covenant University, Ota, Nigeria

3   Department of Building, Obafemi Awolowo University, Ile-Ife, Nigeria

4   School of Architecture, Building and Civil Engineering, Loughborough University, Leicester, UK

leads to models that sometimes underestimate and, more seriously, overestimates the actual shear resistance [3]. Most of these models have some limitations in their application.

Over the past five decades, advancement in technology has led to the development of various soft-computing (SC) techniques utilising artificial intelligence. These techniques have enabled engineers in various fields to model complicated engineering problems and inspect the effects of variables in real-time analysis. Unlike the conventional method of analysis (such as the semi-empirical models), soft-computing (SC) techniques have become an alternative due to the simplicity of implementation and the probable higher degree of accuracy [4]. Various studies have been conducted in the past on the use of SC techniques to model anchorage related problems [2, 5–10]. The most common techniques employed are artificial neural networks (ANN), support vector machines (SVM) and gene expression programming (GEP).

Alqedra and Ashour [5] used an artificial neural network (ANN) to predict the shear capacity of single anchors near the concrete edge. The developed ANN-based model yielded better accuracy than the conventional concrete capacity design (CCD) method in terms the coefficient of determination and the model error. Sakla and Ashour [7] predicted the tensile capacity of single adhesive anchors using artificial neural networks (ANNs). The authors used a multi-layered feed-forward neural network trained with the back-propagation algorithm and constructed using seven design variables as network inputs and the uniform bond strength of adhesive anchors as the only output. The results indicate that ANNs are a useful technique for predicting the tensile capacity of adhesive anchors.

Ashour and Alqedra [6] proposed a feed-forward neural network model for evaluating the concrete breakout strength of single cast-in and post-installed mechanical anchors in tension. The predictions by use of the ANNs captured the test results with mean ratios between 1.025 to 1.065, standard deviations between 19.9 and 22.6% and coefficients of determination varying between 0.879 and 0.907, and they indicated that the Concrete Capacity Design (CCD) method is a reliable predictor of the ultimate loads. Olalusi and Spyridis [2] and Spyridis and Olalusi [10] used the Gaussian Process regression (GPR) and Support Vector Machine algorithms to model the failure of anchors in shear and tension, respectively. The developed models produced more accurate predictions with lower bias and uncertainties than the existing procedures available in design standards.

One disadvantage of the ANN, GPR, and SVM is the inability to generate mathematical equations that can become predictive models for design purposes. This is a major bottleneck to the wide adoption of the SC methods as opposed to the more traditional analytical predictive models in design codes. Gene expression programming (GEP) is an SC technique that generates predictive models, therefore, overcoming the shortfall of ANN, GPR, and SVM. Currently, there is a scarcity of the literature on the use of GEP for edge breakout shear capacity of singe anchors as opposed to ANN, GPR and SVM (as highlighted in the previous paragraphs). This study, for the first time, is aimed at generating a GEP model as a predictive equation for the edge breakout shear capacity of single anchors. With this sort of equation, the SC model of this study can become generally adopted, used in design standards/codes, and extended for various structural engineering problems.

As a matter of fact, all the SC techniques previously applied to anchorage systems produced more precise predictions than the current models in design standards. A GEP model developed by Gesoğlu et al. [9] for the shear capacity of adhesive anchors post-installed into uncracked concrete yield the highest correlation coefficient and lowest errors when compared to existing standards. All these indicate a lot of potential for a GEP model.

This study aims to develop an explicit formulation for predicting the shear capacity of single anchors loaded close to the edge with reasonable accuracy. The gene expression programming SC technique was applied for constructing the proposed model. The worldwide experimental database consisting of 366 single anchors loaded in shear, compiled by the ACI Committee 355, was utilised in this study. This database is adopted because of its round-robin nature (results from around the globe) and has been declared accurate/reliable by the America Concrete Institute. Moreover, there is a scarcity of such databases as alternatives.

Predictions from the developed model were compared to the mean predictions from existing formulas given in the new EN 1992–1-4 [11] and ACI 318 [12]. While other standards (such as Prestressed/Precast Concrete Institute handbook [13]) exists for the prediction of shear capacity of single anchors, those of EN 1992–1-4 [11] and ACI 318 [12] are deemed widely accepted and popular because of their conventional approach. Hence, their adoption for this study. Based on a previous study by the author [2] using the ACI Committee 355 database, a comparison was also made with the predictions of other SC techniques (GPR and SVM) (Table 5). This is intended to highlight the efficiencies of the various SC techniques. An advantage here is the monotony of using the same dataset and using similar data mining techniques allowing for equitable comparison of the techniques. That is, how the new GEP technique can possibly affect the result will be showcased.

2 Procedures for calculating the edge breakout shear capacity of a single anchor according to the current design codes

The formulations to estimate the mean concrete edge breakout strength of anchor loaded in shear in uncracked concrete according to the ACI 318 [12] and EN 1992–1-4 [11] are expressed in Eq. 1 and 2, respectively. For anchors loaded in shear resulting in a concrete edge breakout failure, the basic ACI 318 equation is based on the concrete capacity design method (CCD-method) [14–16]. The CCD-method is simplified by the ACI 318 to allow hand calculations to be performed and concrete breakout resistances to be determined easily. The EN 1992–1-4 is valid for a $d_{nom} \leq 60mm$ and embedment depth $h_{ef} \leq 12 d_{nom}$ in case of $d_{nom} \leq 24mm$ and otherwise $\leq \max(8 \max d_{nom}, 300mm)$. The design calculations covered by EN 1992–1-4 are valid only up until $f_{ck} < 60 N/mm^2$. ACI 318 is valid for a concrete compressive cylinder strength of 10,000 psi.

$$V_{ACI-m} = \min\left\{ \left(\frac{l_f}{d_a}\right)^{0.2} \cdot \sqrt{d_a} \sqrt{f_{cm}}(c_1)^{1.5}, 7.1 \cdot \sqrt{f_{cm}}(c_1)^{1.5} \right\}$$
(1)

$$V_{EC2-m} = 3 \cdot d_{nom}^{\alpha} \cdot l_f^{\beta} \cdot \sqrt{f_{cm}} \cdot c_1^{1.5}$$
(2)

where

$$\alpha = 0,1 \cdot \left(\frac{l_f}{c_1}\right)^{0,5}$$

$$\beta = 0,1 \cdot \left(\frac{d_{nom}}{c_1}\right)^{0,2}$$

where d_a is the outside diameter of the anchor. f_c' is the concrete cylinder strength per the ACI acceptance standards. λ_a is the modification factor for applications in lightweight concrete. f_{cm} is the mean concrete cylinder compressive strength.

3 An overview of gene expression programming (GEP)

Gene expression programming uses genetic algorithms which contain populations of individuals, selects them according to their fitness and introduces genetic variations in the form of mutations based on one or more genetic operators. In genetic algorithms, individuals are programmed as linear strings of a fixed length known as chromosomes. Chromosomes are fixed in length and thereafter expressed as nonlinear entities ranging in shapes and sizes [17]. In general, genetic algorithms are used to optimise real-life problems [4]. GEPs seamlessly translate information between chromosomes (genotype) and expression trees (phenotype). This makes modifications in easily manipulated components, such as chromosomes, syntactically translated to expression trees, allowing GEPs to replicate a genotype/phenotype relationship truly. Apart from merely manipulating linear, compact components like chromosomes, an added advantage is that expression trees are based solely on their specific chromosomes and are chosen based on their respective chromosomes. Thus, in the case of selection for fitness, the best-fitted expression trees are chosen for reproduction with modification of their chromosomes, much like the reproduction of humans [17, 18]. Unlike other SC techniques such as ANN, GPR and SVM, gene expression programming (GEP) can generate an explicit equation or expression denoting the relationship between variables. This study uses gene expression programming to derive a new formulation to predict the breakout capacity of single anchors loaded in shear with a reasonable degree of accuracy.
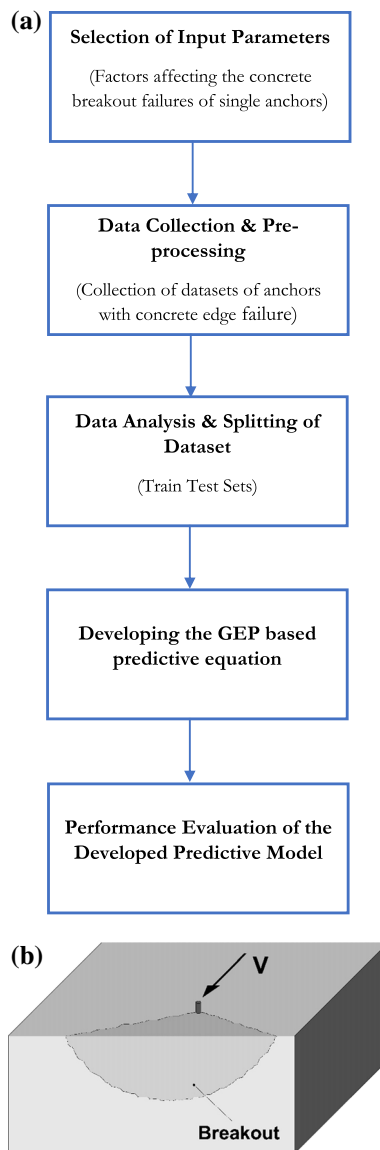
4 Methodology

This section details the methodological approach used to achieve the aims and objectives of this project. The methodological approach describes the input parameters, data collection process, the detailed procedure followed to develop the GEP-based predictive model, and its performance evaluation.

4.1 Selection of the input parameters

Visually depicted by a half-cone shaped failure as shown in Fig. 1, a concrete breakout occurs in anchorage systems where the anchor is located near a free edge of the concrete block. A shearing load applied perpendicular cause failure in the concrete provided the steel's tensile strength is not exceeded [3]. Bede et al. [19] reported that when an anchor is exposed to an arbitrary shearing force, a concrete failure would likely occur first instead of a prying out failure or steel rupture failure, provided the location requirements of the embedded anchor are met. This failure is due to the relatively low edge distance and the required force to cause a concrete breakout instead of the force required for a steel tensile failure. Consequently, this makes concrete breakout failures the critical controlling factor during anchors' design [19]. From above, it is observed that the angle of failure is approximately $35°$ to $45°$ based on the standard used with a length of about three times that of the distance from the edge [11, 12].

Due to the critical nature of concrete edge breakout failures, it is vital to consider all factors affecting a

(a)

(b)

Fig. 1 a. Schematic representation of the methodology and workflow. **b** Sketch of concrete edge breakout for an anchor under shear with a load towards the edge and dimensions of idealised failure bod [10]

concrete breakout failure. Several studies in the literature [1, 2, 14, 19, 20] discusses the range of factors affecting the concrete breakout failures of single anchors located near the edge subjected to shear in uncracked concrete. The major factors are the edge distance c_1, anchor diameters d_{nom}, embedment depth l_{ef} and concrete strength f_c. As shown in Eqs. 1 and 2, these are also the factors considered by the ACI 318 [12] and EN 1992–1-4 [11]. Hence, the factors ($d_{nom}, l_{ef}, c_1, f_c$) are considered as the input parameters for the developed GEP based predictive model discussed in the next sections.

4.2 Data collection and pre-processing

To develop the GEP-based model, a database containing experimentally determined values is required. This provides a source of knowledge for the soft computing technique used (in this study—gene expression programming) to learn from and test against. This study uses a worldwide database compiled and provided by the American Concrete Institute Scientific Committee 355—Anchorage to Concrete. The database consists of 366 experimentally measured failure load for single anchors loaded in shear. The dataset was cleaned to detect and correct inaccurate records. Each sample within the database consists of the following measurements:

- d_{nom} Diameter of embedded anchor bar.
- l_{ef} Effective load-bearing embedded length of the anchor.
- c_1 Distance from the center of an anchor shaft to the edge of concrete in the direction of the shear force.
- f_c Compressive strength of concrete.
- V_{exp} Concrete breakout strength for anchors loaded in shear in uncracked concrete.

4.3 Data analysis & splitting of dataset

Since the experimental database used in this investigation consists of various configurations of anchors loaded in shear, the samples were configured in an orderly manner. The sample configurations range from smallest to largest edge distance c_1, smallest to the largest diameter d_{nom}, lowest to highest concrete strength f_c and shallow to deep embedment depths l_{ef}. The statistical characteristics of the experimental database are summarised in Table 1. An elaborate discussion on the database can be found in [2]. Of the 366 experimental observations in the dataset, 244 samples were used to train the GEP model, while the remaining 122 samples were used to test the predictive model. As noted earlier, the dataset is a form of round robin results, that is, they are independent samples (and results)

Table 1 Statistical Characteristics of Experimental Database

	d_{nom}(mm)	l_{ef}(mm)	c_1(mm)	f_c(MPa)	V_{exp}(kN)
Minimum	8	50	45	18.36	4.72
1st quartile	16	100	70	22.9	21.1025
Median	20	135.5	100	25.4	38.905
3rd quartile	25.4	300	200	28.475	80.755
Maximum	88.9	762	508	85.38	518.62

from around the globe. Hence, they qualify for data splitting for cross-validation of the generated model.

4.4 Developing the GEP based predictive equation

The GeneXproTools version 5.0 programme [21] was adopted in this study to develop a new mathematical model for predicting the concrete breakout capacity of single anchors loaded in shear. GeneXproTools is a powerful tool with multiple modelling capacity. The simple user interface allows users to seamlessly integrate models and provides various controls, thus allowing for a more controlled user experience [17]. The choice of a GEP-enabled data modelling software is authors' preference, and this is believed not to influence the outcomes in any way; other platforms (such as r, python) are also available for such purposes. To avoid overfitting, the programme randomly splits the experimental database into the training datasets (244 samples [66.66%]) and testing datasets (122 samples [33.33%]). The shear parameters (d_{nom}, l_{ef}, c_1, f_c) for both the training and testing datasets have even statistical properties such as mean, standard deviation and coefficient of variation.

The GEP based model was developed using four influencing input parameters, as expressed in Eq. 3. As noted earlier, these are the germane influencing factors known to influence the shear breakout capacity. They are also the measurements available in the database.

$$V = f\left(d_{nom}, l_{ef}, c_1, f_c\right) \tag{3}$$

In developing the predictive model, the selection of various GEP parameters and settings is required. The parameter selection will have a significant impact on the generalisation capability and the predictive performance of the developed model. The selected parameters are listed in Table 2, which were tested to be adequate for the prediction; the addition of more parameters made no significant effects. The values of some of the parameters were selected based on a trial-and-error approach, and some previously suggested values [22–24]. It should be noted that the trial-and-error method for the values is usually the starting point for repetitive algorithms until the convergence of the solution. The programme was run until there was no longer significant improvement in the performance of the models. The best GEP model was selected based on the model's simplicity and the best fitness value on the training and testing dataset.

The derived GEP-based formulation for the concrete breakout strength prediction of single anchors loaded close

Table 2 GEP parameters used for the proposed model

Parameters	Settings
General	
Chromosomes	30
Genes	3
Head size	8
Mutation rate	0.00138
Inversion rate	0.00546
Linking function	Addition
Function set	$+$, $-$, $*$, $\div$, Pow, Exp, x^2, x^3, x^4, $\sqrt[3]{\ }$
Genetic operators	
Mutation rate	0.00138
Inversion rate	0.00546
Numerical constants	
Constants per gene	
Data type Integer	Floating point
Lower bound	-10
Upper bound	10

to the edge is expressed in Eq. 4. In comparison, the existing formulas in EN 1992–1-4 [11] and ACI 318 [12] are given in Eqs. 1 and 2. The GEP generated model can be said to be more robust and simpler enough for implementation into codes and standards.

$$V_{GEP} = \left(\frac{c_1 + f_c}{25.8}\right)^2 + \frac{f_c + 25.8}{10.44} + \frac{149 - l_{ef}}{-3.2f_c + 3.4 + d_{nom}} + \sqrt[3]{\frac{c_1 - d_{nom}^4 + c_1^3}{l_{ef}^2}} \tag{4}$$

A common form of representing formulations is the use of a binary expression tree diagram. Expression trees represent the variables and operations on leaf nodes indicating the structure of the formulation. Expression trees are usually read from bottom to top and left to right. The corresponding expression tree of the derived model is shown in Fig. 2.where,
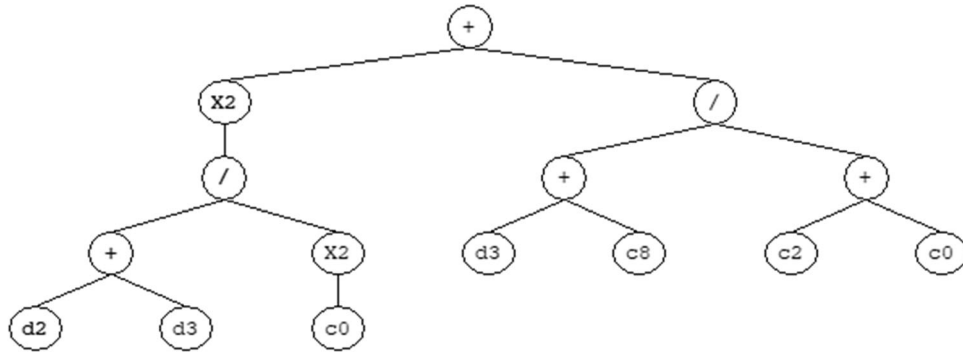
d_0 is the anchor diameter, d_{nom} in mm.

d_1 is the effective load-bearing length of the anchor, l_{ef} in mm.

d_2 is the distance from the centre of an anchor shaft to the edge of the concrete, c_1 in mm.
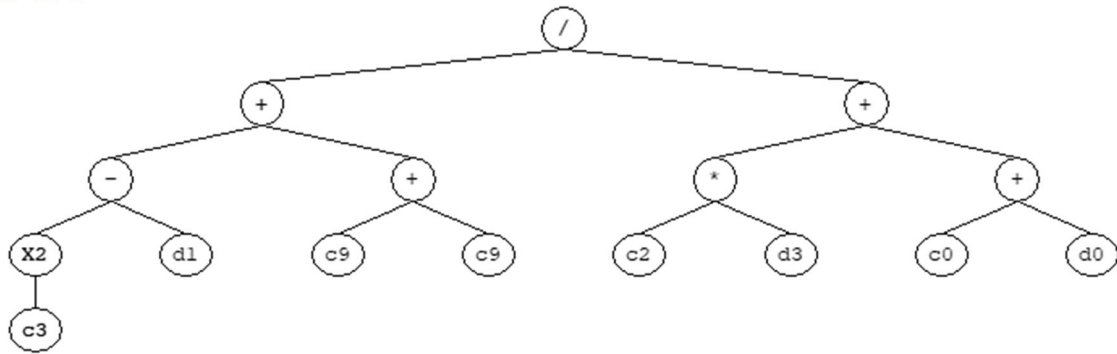
d_3 is the specified compressive strength of concrete f_c, N/mm.

and constants in

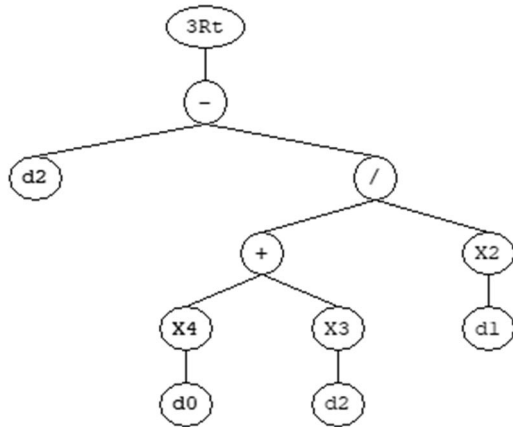**Fig. 2** Demonstrates the expression tree for the terms used in the formulation of the GEP model

$Sub-ET1$:

$c0 = 5.08$

$c2 = 5.34$

$c8 = 25.8$

$c9 = 1.89$

$Sub-ET2$:

$c0 = 3.4$

$c2 = -3.2$

$c3 = 10.6$

4.5 Performance evaluation of the developed predictive model

Several statistical measures were adopted to assess the predictive performance of the developed model. The statistical methods used in this investigation utilised the experimental failure load from the database (V_{exp}), and the predicted capacity from the developed GEP model (V_{GEP}). Whilst various statistical measures can be used to assess a

model's predictive capability, a few of the critical measures are considered in this study. These measures are the coefficient of determination (R^2), Root-Mean-Squared Error (RMSE), Mean Absolute Error (MAE), the mean and standard deviation of the Model error. These four methods are generally acceptable to be adequate for evaluating model's predictive capability.

4.5.1 Coefficient of determination (R^2).

The R^2 of a model is a relative value which denotes the proportion of variance depicted by the model. The R^2 value is only to be compared to models tested with the same dataset [25]. For example, since this study uses the same dataset for the testing of the GEP model, ACI 318 model and EC2 model, a valid comparison between the models using the R^2 value can be made. The R^2 of a model can be calculated according to Eq. 5.

$$R^2 = 1 - \frac{\sum (V_{\exp} - V_{\text{pred}})^2}{\sum (V_{\exp} - V_{\exp,m})^2} \tag{5}$$

where,

$V_{\exp}$ is the experimental shear strength.

V_{pred} is the predicted shear strength from the model under consideration.

4.5.2 Root-Mean-Squared Error (RMSE)

The RMSE is a standard evaluation technique in analysing the predictive capability of a model. The RMSE is known for penalising large errors between the predicted and actual values of a model and having the same units as the outcome values of the dataset. It is highly improbable for a prediction model to predict outcomes with 100% efficiency. Thus, a measure such as the RMSE provides a means of assessing how much values that are not predicted accurately deviate from the experimental values. The RMSE can range from 0 to infinity, with values closer to 0 denoting better predictive capabilities and is calculated according to Eq. 6.

$$\text{RMSE} = \sqrt{\frac{\sum (V_{\exp} - V_{\text{pred}})^2}{n}} \tag{6}$$

where n is the number of samples.

4.5.3 Mean Absolute Error (MAE)

The MAE measures the average magnitude of the errors of a prediction model, without considering the direction in which the error occurs. Thus, the MAE represents the average absolute difference between the experimental

values and the predicted values with each individual difference having an equal bias. The MAE can be calculated according to Eq. 7.

$$\text{MAE} = \frac{\sum |V_{\exp} - V_{\text{pred}}|}{n} \tag{7}$$

4.5.4 Model error

In this study, the model error is characterised as the ratio of experimentally observed to predicted mean shear breakout capacities. ϑ, as shown in Eq. 8 [26]. The model error is calculated for each sample set within the testing database of 122 samples for each model under inspection. The model errors are assessed in terms of their mean value, standard deviation and coefficient of variation.

$$\vartheta_x = \frac{V_{exp,x}}{V_{pred,x}} \tag{8}$$

where, ϑ_x represents the model error for sample x from the testing dataset.

5 Results and discussion

5.1 Comparative assessment of the performance of the GEP, ACI and EC2 models using statistical measures

The statistical metrics for all the models investigated are presented in Table 3. As seen in the table, a strong correlation between the predicted values and experimental values was generally observed with all models producing a correlation coefficient greater than 0.99 (see Table 3). However, upon closer inspection, the GEP model correlates closest to the experimental value when compared to the ACI and EC2 models. This is further reinforced with the RMSE and MAE both scoring the lowest for the GEP model. The RMSE is the dispersion of the unexplained variance or errors which occurs within the model. As such, the lowest RMSE represents the lowest deviation from the

Table 3 Performance assessment of GEP, ACI and EC2 Models

Statistical metric	GEP	ACI	EC2
MAE	8.79	15.90	18.03
RMSE	13.77	25.79	33.17
MSE	189.63	665.62	1100.46
Correlation coefficient r	0.99	0.99	0.99
R^2	0.99	0.98	0.98

mean error for every model tested, thus implying a better predictive model. The MAE denotes the distance of the predicted shear capacity from the experimentally measured shear capacity. Since the absolute value is taken, the direction in which the error occurs becomes irrelevant and only the magnitude of the error is assessed. Overall, the GEP model shows the closest prediction to the actual observed values compared to the ACI and EC2 models. The ACI and EC2 models also yield result closer to the observed experimental values. Figure 3 shows the overlapped plots for all three models under assessment. Since the model is developed using each individual sample set within the testing database, the illustration further reinforces the fact that the database contained a few large (> 200 kN) measurements for breakout capacities. This is proven by the lack of data points which exceed the 200 kN mark (Fig. 3).
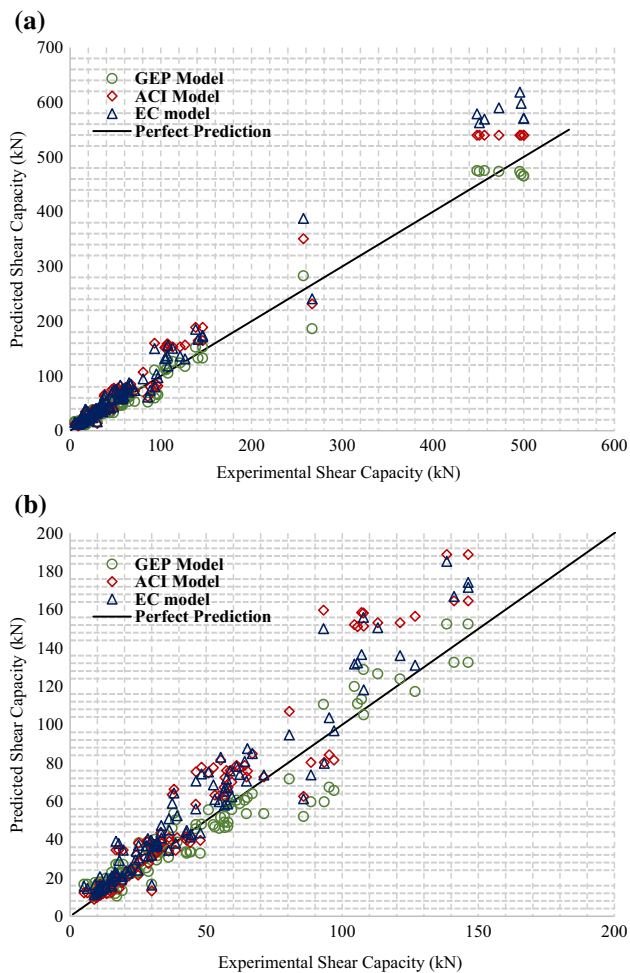
Fig. 3 a Comparison of the experimentally observed and the predicted shear breakout capacity. **b** Comparison of the experimentally observed and the predicted shear breakout capacity for anchors with strength ≤ 200 kN

5.2 Model error assessment

As previously expressed, the model error denoted as ϑ, can be represented using Eq. 8. For this representation, a model with a ratio of 1 represents a perfect relation between the predicted and experimentally observed values for the concrete breakout capacity of a single anchor loaded in shear [26]. Values greater than 1 denotes an underprediction of the model's actual strength, and values less than 1 denotes an overprediction of the actual strength by the model.

The histograms presented in Fig. 4 presents a graphical representation of the estimated model errors for each of the models under investigation. The presented histogram plots the frequency of occurrence of a random model error against the interval in which the error occurs. The width of each interval represents the extents of scattering of the model error.

Figure 4b and c shows that the ACI and EC2 model has a greater frequency of overpredictions than the GEP model. Such correlated with the design principle of accounting for the worst-case scenario and, hence overdesigning. These models, such as the ACI and EC2 models, which have increasing conservative biases provide a greater sense of reliability, however, overdesigning usually increases the construction cost of a project and may lead to unnecessary designs for the intended purpose of the design. The GEP model expressed (Fig. 4a) shows a fairly symmetrical distribution of the frequencies observed about the mean model error of 1.01 (Table 4) in comparison with the ACI and EC2 models. This shows little to no underlying bias present in the GEP model. The scatter plot of V_{exp} vs V_{pred} presented in Fig. 3 serves two functions. Firstly, the plot provides a visual representation of each sample's closeness within the testing database to a 'perfect prediction line' for each model. Secondly, the plot provides a means of assessing potential outliers within the database. The 'perfect prediction line' represented as a black diagonal line in Fig. 3 denotes the point at which the predicted breakout capacities equal the measured breakout capacities. The figure shows the scatter of predictions from the different models around the perfect prediction line. This provides a visual means of understanding the predictive capabilities of each model in comparison with each other.

Outliers are determined as points which significantly deviate from the trend of expected results. Outliers pose a potential threat to a study as it affects the variance of the model. Subsequently, all indicators which incorporate the variance into the calculation are then affected. Outliers can usually be traced back to human error whilst conducting the experiments or faulty instruments used in the experiment. However, outliers may be accurately measured
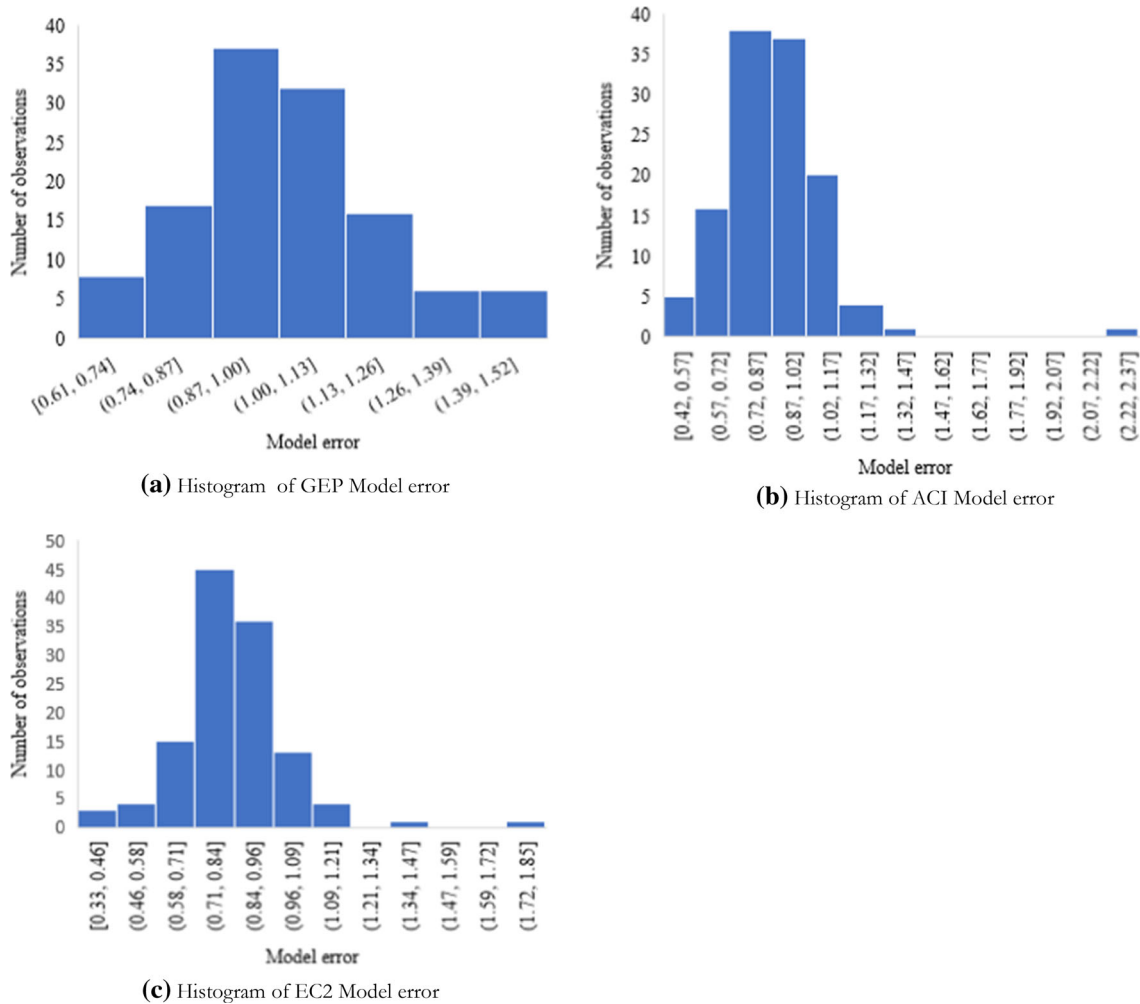
(a) Histogram of GEP Model error

(b) Histogram of ACI Model error

(c) Histogram of EC2 Model error

Fig. 4 Histogram of Model error

Table 4 Statistical properties of the model error

Model error	GEP	ACI	EC2
Mean	1.01	0.89	0.82
Standard deviation	0.18	0.21	0.18
Coefficient of variation	0.18	0.24	0.22
Min.	0.61	0.42	0.33
Max.	1.48	2.28	1.84

outcomes as well [26]. Fortunately, this study does not indicate any potential outlier for the models assessed in this study. Instead, the potential outliers are just poorly predicted outcomes by the prediction models.

Table 4 presents a summary of the statistical parameters assessed for the model errors. Here, the main parameters under inspection for the model errors are the mean, standard deviation and coefficient of variation. The best model

is characterised by the model error mean value closest to 1 and the minimum standard deviation. The GEP-based strength model achieved the best ranking with a model error mean value of 1.01 (i.e. 1% bias) and the lowest dispersion of 0.18.

The performance of the developed model is compared to other SC techniques, as presented in Table 5. Olalusi and Spyridis [2] developed models for predicting the concrete breakout strength of single anchors in shear. They assessed the performance of the developed model using the

Table 5 Statistical properties of the GPR and SVM based model [2]

Model error	GPR	SVM
Mean	0.99	1.01
Standard deviation	0.11	0.22
Sample size	110	110

*GPR (Gaussian Process Regression) SVM (Support Vector Machine)

distribution of the model error (obtained as the ratio of experimental breakout strength to predicted breakout strength) in terms of the mean value and dispersion (Table 5). Evidently, the performance of the proposed GEP model aligns with that of the GPR model and outperforms that of the SVM model, as reported in Table 5 [2]. However, this must remain under the perspective that [2] has used a smaller testing dataset compared to what is used in this study.

5.2.1 Sensitivity of the model error to input variables

A perfect predictive model is expected to show no correlation/trend between its input design variables and its model error [26]. According to [27], correlation coefficients ranging between 0 and 0.2 represent very weak correlations, 0.2 and 0.4 represent weak correlations, and 0.4 to 0.6 represent moderate correlations. To further assess the proposed equation's predictive capability, the model error trend is assessed against the main design variable, as presented in Fig. 5. The figure suggests that the GEP based strength model has good accuracy with no significant trend

with respect to the design parameters. The correlation between the model error and the concrete edge distances and anchor diameters for the ACI and EC2 model can be classified as weak correlations. However, the results produced by the building codes are notably scattered.

6 Conclusions

This study developed a mathematical model to predict the concrete edge shear breakout capacity of single anchors using gene expression programming. The choice of GEP emanates from the advantage of generating a mathematical model for prediction as compared to other soft computing techniques. The predictive performance of the developed model is compared to that of the existing standardised models. In comparing the models, a set of experimentally observed shear capacities was used. Each sample within the database contained the edge distance c_1, anchor diameters d_{nom}, embedment depth l_{ef} and concrete strength f_c and the failure load V_u.
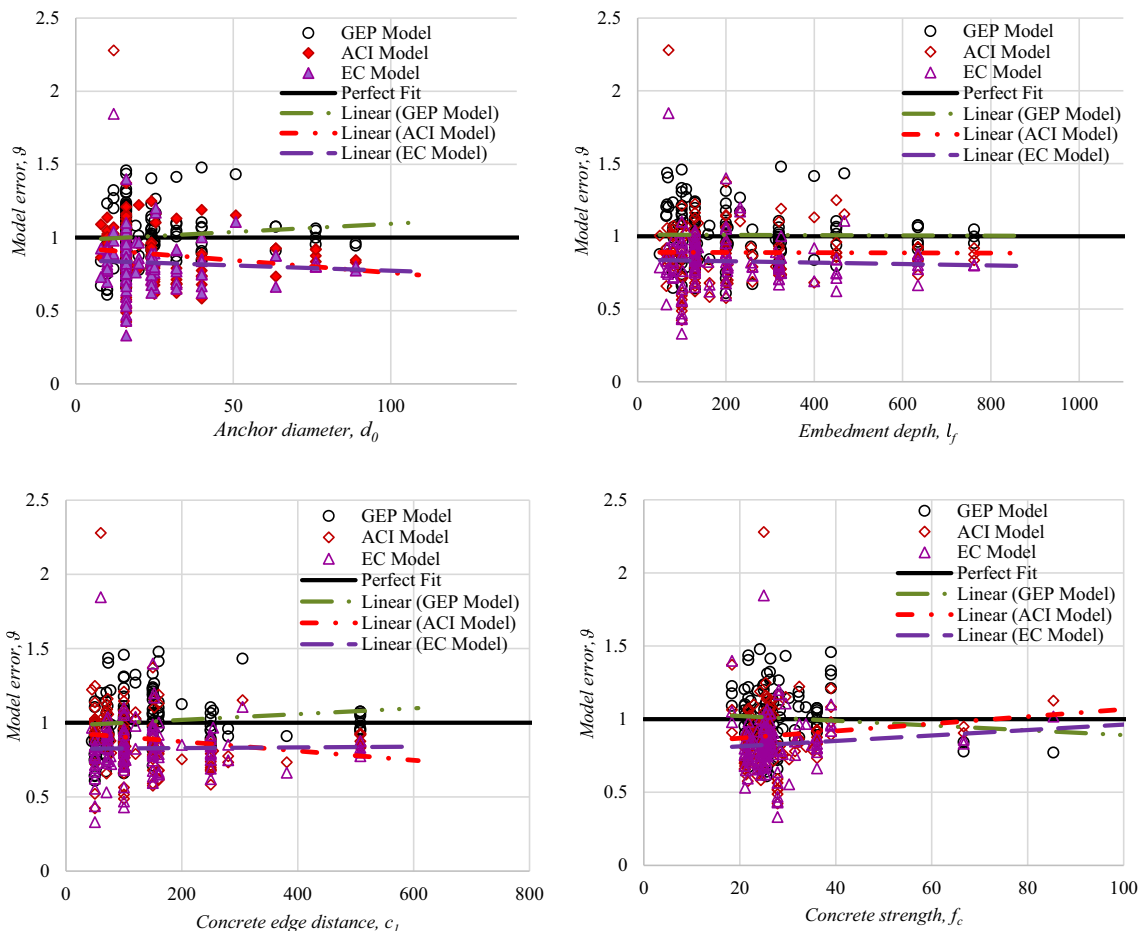

Fig. 5 Scatter plots of model error versus shear parameters

Upon assessing the generated mathematical model in comparison with other models in terms of statistical measures such as the MAE, RMSE, R^2 and correlation coefficient r of all three models, evidence suggests that the GEP model produced the closest prediction to the experimentally measured shear breakout capacities. Notably, both the ACI and EC2 models produced fairly accurate predictions with both models having a correlation coefficient of 0.99. The models were further assessed in terms of the statistical properties of their model error which were obtained as the ratio of the experimentally observed shear capacity to the predicted shear capacity. GEP model produced the best performance in terms of mean value and spread of the model error. The sensitivity analysis shows a low correlation between the input variables and the model error for all the models investigated. Whilst a poor correlation usually indicates poor performance, in this case, an ideal model has little or no correlation between the input variables and the model error.

In comparison, the predictions obtained from the proposed GEP model were more uniform and much closer to the actual results than the predictions obtained from the models adopted in the ACI 318 and EC2 design standards. The proposed model is expected to be very useful for evaluating the concrete breakout shear capacity of single anchors in pre-planning and pre-design phases. This study can be used as a reference for code-researchers and the developers of standards in providing information on the performance of the models investigated in this study. Since the proposed GEP model predicts the shear capacity of a single anchor more accurately than existing models, the implementation of the GEP model will produce more cost-effective and reliable designs as the designs will be less susceptible to over-designing or under-designing.

One of the drawbacks of soft computing approach is the need for many large-samples databases. With the availability of more adequate experimental dataset (on the shear capacity of single anchors) and testing, a generalised equation can be developed, based on a GEP computing, that will find its way into design standards and codes.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Eligehausen R, Mallée R, Silva JF (2006) Anchorage in concrete construction. Wiley, New Jersey
2. Olalusi OB, Spyridis P (2020) Machine learning-based models for the concrete breakout capacity prediction of single anchors in shear. Adv Eng Softw 147:102832
3. Grosser P (2012) Load-bearing behavior and design of anchorages subjected to shear and torsion loading in uncracked concrete
4. Ibrahim D (2016) An overview of soft computing. Proc Comput Sci 102:34–38
5. Alqedra MA, Ashour AF (2005) Prediction of shear capacity of single anchors located near a concrete edge using neural networks. Comput Struct 83(28–30):2495–2502
6. Ashour AF, Alqedra MA (2005) Concrete breakout strength of single anchors in tension using neural networks. Adv Eng Softw 36(2):87–97
7. Sakla SS, Ashour AF (2005) Prediction of tensile capacity of single adhesive anchors using neural networks. Comput Struct 83(21–22):1792–1803
8. Güneyisi EM, Gesoğlu M, Güneyisi E, Mermerdaş K (2016) Assessment of shear capacity of adhesive anchors for structures using neural network based model. Mater Struct 49(3):1065–1077
9. Gesoğlu M, Güneyisi EM, Güneyisi E, Yılmaz ME, Mermerdaş K (2014) Modeling and analysis of the shear capacity of adhesive anchors post-installed into uncracked concrete. Compos B Eng 1(60):716–724
10. Spyridis P, Olalusi O (2021) Predictive modelling for concrete failure at anchorages using machine learning techniques. Materials 14:62
11. Technical Committee 250/European Committee for Standardisation (CEN/TC 250). EN 1992–4 Eurocode 2. Design of concrete structures. Design of fastenings for use in concrete, (2018)
12. ACI Committee and International Organization for Standardization. Building code requirements for structural concrete (ACI 318–14) and commentary. American Concrete Institute, (2014)
13. PCI Design Handbook (1998) Precast-prestressed concrete, 5th edn. Precast-Prestressed Concrete Institute, Chicago
14. Fuchs W, Eligehausen R, Breen JE (1995) Concrete capacity design (CCD) approach for fastening to concrete. Struct J 92(1):73–94
15. Lee NH, Park KR, Suh YP (2010) Shear behavior of headed anchors with large diameters and deep embedments. ACI Struct J. https://doi.org/10.14359/51663530
16. Meinheit DF, Anderson NS, Windisch A (2011) One discussion from the March/April 2010 ACI structural journal and two discussions from the May/June 2010 ACI structural journal. Struct J 108(2):248–253
17. Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. Complex Syst 13(2):87–129
18. Zhong J, Ong Y, Cai W (2015) Self-learning gene expression programming. IEEE Trans Evol Comput 20:1–1
19. Bede N, Grosser P, Ožbolt J (2018) Shear breakout capacity of various fastening systems in concrete elements. Građevinar 69:1093–1100
20. Comité euro-international du béton. Fastenings to Concrete and Masonry Structures: State of the Art Report (No. 216). Thomas Telford, (1994)
21. GEPSOFT. GeneXproTools. Version 4.0. Available online; 2006, <http://www.gepsoft.com/>. (2006)
22. Gandomi AH, Babanajad SK, Alavi AH, Farnam Y (2012) A novel approach to strength modeling of concrete under triaxial compression. J Mater Civil Eng ASCE 24(9):1132–1143
23. Gandomi AH, Alavi AH, Kazemi S, Gandomi M (2014) Formulation of shear strength of slender RC beams using gene

expression programming, part I: without shear reinforcement. Automat Constr 42:112–121

24. Baykasoglu H, Gullub H, Canakcı LO (2008) Prediction of compressive and tensile strength of limestone via genetic programming. Expert Syst Appl 35(1–2):111–123

25. Hale J (2020) Which evaluation metric should you use in machine learning regression problems?, Towards Data Science, viewed 26 December 2020, https://towardsdatascience.com/which-evaluation-metric-should-you-use-in-machine-learning-regression-problems-20cdaef258e

26. Olalusi OB, Spyridis P (2020) Uncertainty modelling and analysis of the concrete edge breakout resistance of single anchors in shear. Eng Struct 222:111112

27. Franzblau AN (1958) A primer of statistics for non-statisticians