Challenge Journal of Concrete Research Letters

Research Article

A novel neuro genetic programming framework for modelling compressive strength of recycled aggregate concrete

Preeti Kulkarni [a,*], Shreenivas N. Londhe [a], Pradnya R. Dixit [a]

[a] *Department of Civil Engineering, Vishwakarma Institute of Technology, Pune 411048, India*

ABSTRACT

Recycled Aggregate Concrete (RAC) can represent as a sustainable alternative to conventional concrete; however compressive strength prediction of RAC is difficult due to the complex and nonlinear interactions between the materials in the mix. Researchers have been utilizing soft computing tools like: Genetic Programming (GP), Artificial Neural Network (ANN) etc. for predicting compressive strength, however the performance of these models is satisfactory. To leverage the characteristic of evolutionary optimization in Multi Gene Genetic Programming (MGGP): variant of GP and the effective nonlinear input and output mapping through ANN, the present study introduces a Novel Neuro Genetic Programming framework (NMGGP), an advanced hybrid approach that combines the strengths of MGGP and ANN. The NMGGP framework is a framework devised in three phases: (i) MGGP constructs a multigene model, where each gene represents a weighted mathematical expression derived from input parameters; (ii) High-weighted genes, are selected, and transformed into transformed inputs to enhance the learning process. (iii) These transformed inputs, along with the original output, are fed into an ANN enabling advanced nonlinear mapping and precision in prediction. The proposed model was tested for predicting the 28-day compressive strength of RAC which considered 8 inputs as the kg/m^3 quantity of concrete mix. Further three highly weighted genes were transformed into transformed inputs and predictions were done in phase III. The results displayed an excellent model with r=0.967, RMSE as 4.744 and MAE as 3.944 as compared to the stand-alone models of MGGP and ANN with r=0.930, RMSE more than 5.0 and MAE mode than 5.0. The results also can be said to be robust through a balanced scatter plot and graphical comparison between predicted and observed 28-day strength of RAC.

ARTICLE INFO

Article history:

Received	– April 30, 2025
Revision requested	– August 8, 2025
Revision received	– September 4, 2025
Accepted	– September 15, 2025

Keywords:
Recycled aggregate concrete
Compressive strength
Artificial neural network
Multi gene genetic programming
Neuro gene programming

Citation: Kulkarni P, Londhe SN, Dixit P (2025). A novel neuro genetic programming framework for modelling compressive strength of recycled aggregate concrete. *Challenge Journal of Concrete Research Letters*, 16(4), 203–214.

1. Introduction

Recycled aggregates (RA) and recycled aggregate concrete (RCA) have been used as replacement to certain extent with the original materials and can be said a step towards sustainable approach. Recycled Aggregates and RAC differ from the original materials and normal concrete in their characteristics and thus predicting their strength is a daunting task (Adriana et al. 2013; Duan et al. 2013). The properties of recycled aggregate concrete using the recycled aggregates are highly dependent on the source and quality of the recycled aggregate. Determining strength of recycled aggregate is more critical as the recycled aggregates contain the adhered mortar which can lead to higher porosity and water absorption, potentially lowering the concrete's strength and workability. These properties of the recycled aggregates and their interaction in the resulting strength of recycled ag-

---

* *Corresponding author.* E-mail address: preeti.kulkarni@vit.edu (P. Kulkarni)

ISSN: 2548-0928 / DOI: https://doi.org/10.20528/cjcrl.2025.04.003

gregate concrete are often non-linear and complex (Hamada et al. 2017). Determination of 28-day compressive strength of RAC is very important as it allows the designer to modify the mix proportion accordingly if the concrete fails to reach designed strength reducing the risk of construction failures resulting in less use of raw material and material costs and in a more economical way. In the traditional way, compressive strength of any concrete is determined by testing done in the lab which consumes time and materials which can be minimized by use of Data Driven Techniques (DDTs) like Artificial Neural Network (ANN), Genetic Programming (GP), Model Tree (MT), etc. These DDTs have been studied for prediction of compressive strength of concrete in last two decades with considerable results (Adriana et al. 2013; Duan et al. 2013; Deshpande et al. 2014; Tran et al. 2022).

ANNs are characterized by recognizing intricate non-linear connections between input-output data which prove to be suitable for finding solutions were describing procedures through physical equations presents a challenge (Hsu et al. 1995; Li et al. 2022, ASCE Task committee, 2001). ANNs have been broadly employed in concrete strength modelling and show significant positive results. Deshpande et al. (2014) designed ANN models to predict the strength of RAC at 28 days, incorporating both per cubic meter material proportions and non-dimensional parameters. In a comparative analysis of ANN and GP techniques to predict 28-day compressive strength of RAC, ANN predicts the strength of concrete with good accuracy as compared to GP which can be evident from the higher R values (0.937 and 0.941). Findings of Kulkarni et al. (2019) indicated that ANNs outperformed non-linear regression models, with the inclusion of non-dimensional parameters enhancing predictive accuracy. Similarly, Onyelowe et al. (2022) employed a novel ANN model using a sigmoid function, achieving a correlation coefficient (r) of 0.99 and a mean squared error (MSE) of 28.67% during training, demonstrating high predictive performance. Duan et al. (2020) developed an ANN model to predict the compressive strength of RAC, considering parameters such as water-cement ratio, recycled aggregate replacement ratio, and curing age. Their model achieved high accuracy ($R^2 > 0.95$), demonstrating ANN's effectiveness in strength prediction. The technique of MGGP was used to anticipate modulus of elasticity of concrete. A universal model was suggested for Normal strength concrete and high strength concrete using the 28-day strength data (Bayazidi et al. 2014). MGGP was also used for 28-day compressive strength prediction of RAC which show that ANN outperforms MGGP and Model Tree techniques (Kulkarni et al. 2019). Hybrid models are also been used like ANFIS which leverage the characteristics of ANN and fuzzy logic (Khademi et al. 2016). Pan et al. (2022) introduced a hybrid model combining ANN with Genetic Algorithm (GA) optimization to predict the compressive strength of green concrete, achieving enhanced accuracy over standalone ANN models.

Thus, earlier studies have shown that hybrid models display better accuracy than the standalone techniques in the area of strength determination of concrete. In the present study, the novel Neuro Genetic Programming framework (NMGGP) of modelling is proposed which uses an interaction of MGGP and ANN techniques. The aim of the present study is to develop a novel NMGGP framework by integrating Multi-Gene Genetic Programming (MGGP) and Artificial Neural Networks (ANN) for accurate prediction of the 28-day compressive strength of recycled aggregate concrete (RAC). Specifically, the study seeks to transform conventional mix design parameters into high-performance gene-based features through MGGP, which are then utilized as enhanced inputs for ANN modelling. By combining the strengths of evolutionary feature engineering and neural network prediction, the research aims to achieve higher accuracy and robustness than standalone MGGP or ANN models, thereby contributing to sustainable and reliable strength prediction of RAC. The motivation of this study was derived from a hybrid ANN model for typhoon-rainfall forecasting and brought together different kinds of implicit neural networks (Lin and Wu, 2009). In addition to ensuring forecasting accuracy, this study emphasizes its novelty compared to earlier research. To achieve this objective, the MGGP technique is employed to first transform the observed mix design parameters in kg/m^3 i.e. inputs into high-performance genes. The high weighted genes are selected and used to generate new transformed inputs (termed as transformed inputs in this work) or features as in feature engineering which can now serve as new inputs for predicting strength of RAC via artificial neural networks (ANN). While Artificial Neural Networks (ANN) and Genetic Programming (GP) have been widely applied individually in compressive strength prediction, this study is, to the best of the authors' knowledge, the first to combine MGGP with ANN to develop a novel NMGGP framework. Additionally, it incorporates feature engineering by transforming original inputs into more meaningful transformed inputs parameters to enhance the accuracy of ANN-based strength predictions.

The paper opens with a brief introduction to the techniques used in the study. This is followed by an overview of the study area and the data characteristics. The approach section specifies the parameters that were required to be input and output, the model architecture, and the performance indicators which were utilized to evaluate the model effectiveness. Lastly, the results are analyzed and evaluated, with visual representations such as hydrographs, scatter plots, and a Taylor diagram included toward the conclusion of the paper.

2. Computational Techniques

The techniques used in the present work are Multigene Genetic Programming, Artificial Neural Network and the novel Neuro Genetic Technique.

2.1. Multigene Genetic Programming (MGGP)

Multi-Gene Genetic Programming (MGGP) is a development from classical Genetic Programming (GP), which in turn originated from the evolutionary ideas suggested by Koza in 1992. MGGP adds more expressive power into the traditional GP model by employing multiple low-complexity trees, each of them capturing one weighed linear expression and assembling as one global model. This architecture places MGGP in a position to address non-linear regression problems with a large, complicated set of input data to a single target output. MGGP Unlike classical GP which can result in complex and less interpretable models, MGGP retains a simple, more understandable structure, as it is the product of aggregation of smaller sub-models (genes), and a Population initialization. The final prediction is the linear combination of these genes, leveraging linear regression techniques to estimate weights. The key advantage of MGGP lies in its ability to generate concise, interpretable mathematical expressions that can be readily implemented in common computational tools such as Microsoft Excel. This makes MGGP particularly suitable for practical engineering applications, including tasks like strength prediction etc. where model transparency and ease of deployment are important (Koza 1992; Searson 2015).

2.2. Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are adaptable computational models modelled after how the human brain processes information (ASCE Task committee, 2001). They excel at identifying complex, non-linear relationships within datasets. ANNs predict output values through a training process that involves continuous adjustments of weights and biases. Typically, an ANN is made up of an input layer, one or more hidden layers, and an output layer. These layers are interconnected by a network of weighted links, along with bias terms, and operate using summation and activation functions. Throughout the training process, the discrepancy between the predicted and actual output—known as the error—is computed and minimized through optimization algorithms that adjust the network's weights and biases. The procedure is carried out multiple times, or across several epochs, until the network reaches a satisfactory level of performance (Rumelhart et al. 1986). After training, the artificial neural network with fine-tuned parameters can be utilized to predict outcomes for new datasets. Readers are further referred to ASCE Task Committee (2001) for details. In this study, the ANN model was created and trained using MATLAB 2023 (MathWorks 2023).

2.3. Neuro Genetic Programming (NMGGP)

The Novel Neuro Genetic Programming framework (NMGGP) approach is based on the idea of feature transformation which is an important aspect of feature engineering. In feature transformation the present input variables are restructured to understand the subtle patterns or relationships in the input data that might not be evident in their original form. By creating new transformed parameters or modifying the original ones, the model gains more information on inputs, which can significantly enhance prediction accuracy.

In the present study, the proposed NMGGP model uses the characteristics of traditional Multi-Gene Genetic Programming (MGGP) and Artificial Neural Networks (ANN). Traditional MGGP linearly combines all the genes, while NMGGP selects only the most impactful genes which are identified by their highest weights based on their mathematical expressions. These genes are further transformed to new inputs (termed as transformed inputs) by their mathematical expressions and original input data alongside the original output variable. Further ANN model is developed with these transformed inputs and original output. This approach yields a powerful model, making it highly suitable for predicting tasks like compressive strength determination.

The NMGGP technique is studied in 3 stages: Stage 1: Development of standalone MGGP model: Initially, the original inputs and the output parameter are used to train a standalone MGGP model. MGGP generates several gene expressions, each representing a mathematical combination of input parameters, bias and coefficients. The optimal number of genes (e.g G1, G2, ..., Gn) is determined by balancing forecast accuracy and model complexity, frequently with the help of Pareto-front analysis. Each gene is assigned a weight based on its contribution to the final output. Stage 2: Selection of critical Genes: From the standalone MGGP output i.e the genes, the highest-weighted genes are selected which are the mathematical expressions that most strongly influence the output parameter. For instance, if G2, G3, and G5 are the highly weighted, they are selected for the third of NMGGP. These high-weighted genes display the most critical and informative relationships between the inputs and their effect on prediction of output. Stage 3: Modelling using ANN: The selected gene expressions are evaluated to generate new input parameters termed as transformed inputs (e.g TG2, TG3, TG5). These transformed inputs also linearly consider the weighted coefficient of bias for the selected gene. These transformed inputs, along with the respective original output parameter, are used to develop predictive model using ANN. The ANN recombines these high-level transformed inputs to model complex, non-linear relationships, further enhancing predictive performance. ANN thus acts as a fine-tuning layer that leverages the condensed information captured in the MGGP genes.

This hybrid strategy of NMGGP enables deriving meaningful, high-level inputs from the original data allowing the ANN to learn and predict more accurately. The key advantage lies in selectively using only the most informative genes, which reduces unnecessary complexity while improving model precision.

Fig. 1 shows the complete NMGGP workflow in three stages. The process of optimizing gene selection, transforming inputs, and fine-tuning relationships through ANN integration ultimately leads to more reliable and insightful predictions. The NMGGP technique discussed was tested on predicting 28-day strength of RAC.

3. Experimental Data

The data used in the study was obtained from fresh experimentation done by the authors as well as from the published literature (Deshpande et al. 2014). With a total data of 228 mix design samples were used which mainly contained proportions of materials used for manufacturing concrete with conventional materials and concrete with Recycled Concrete Aggregates (RCA). As per standard mix design procedures followed all over the world following parameters in kg/m^3 are considered as input parameters in concrete mix design (Shetty 2005; Deshpande et al. 2014): cement (C), natural fine aggregate (NFA), natural coarse aggregate–20mm (NC20), natural coarse aggregate–10mm (NC10), recycled coarse aggregate–10mm (RCA10), recycled coarse aggregate–20mm (RCA20), admixture (A), water (W). The 28-day compressive strength (CS) of concrete is considered as the output. Table 1 below shows the characteristic of data adopted in the present study.

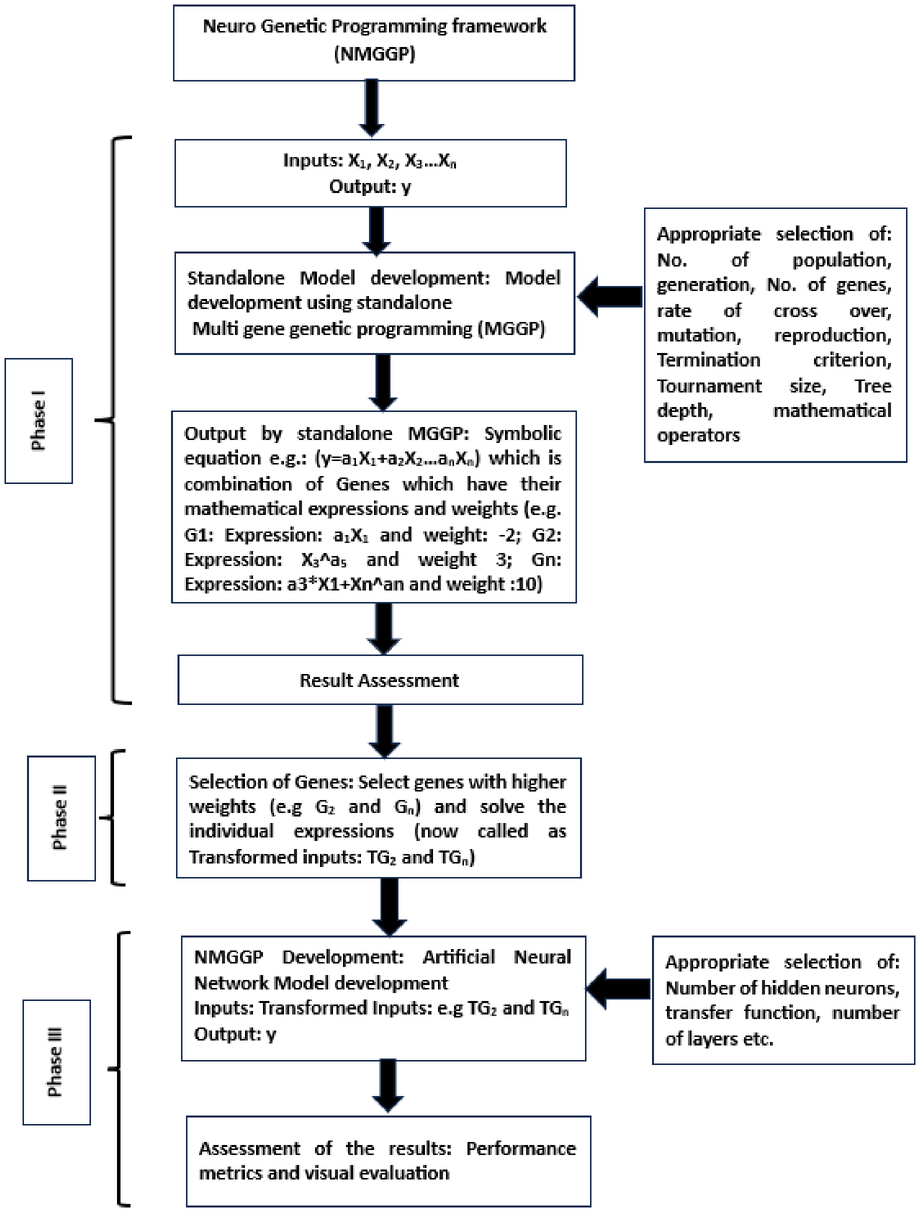

Fig. 1. Flowchart of NMGGP model.

Table 1. Characteristics of experimental data.

Input parameters	C (kg/m^3)	NFA (kg/m^3)	NC-20 (kg/m^3)	NC-10 (kg/m^3)	RC-20 (kg/m^3)	RC-10 (kg/m^3)	A (kg/m^3)	W (kg/m^3)	CS (MPa)
Maximum	645.000	1,050.000	1,508.640	553.000	1,508.640	840.000	41.600	271.000	100.500
Minimum	235.000	217.000	0.000	0.000	0.000	0.000	0.000	120.000	10.319
Average	381.788	707.263	435.828	121.208	420.379	102.599	3.080	187.394	42.757
Standard deviation	67.662	118.063	391.357	177.671	399.802	187.003	5.731	25.108	15.799
Skewness	1.116	–0.472	0.386	1.137	0.653	2.154	2.984	0.170	0.767
Kurtosis	3.928	3.918	-0.825	–0.132	-0.493	4.670	11.756	0.460	0.897

The statistical analysis of the input and output data reveals that most features exhibit non-normal distributions, with admixture showing extreme right skewness (skewness = 2.984) and sharp kurtosis (11.756), indicating the presence of outliers. Cement and water content display more stable distributions, suggesting their reliability as predictors. The target variable, compressive strength, shows moderate variability (mean = 42.76 N/mm^2, standard deviation = 15.80) and a right-skewed distribution, implying the presence of high-strength concrete samples in the dataset. These characteristics highlight the need for data preprocessing and feature scaling prior to modeling.

4. Methodology

The aim of the present study is to check the usability of the novel NMGGP technique to predict the 28-day strength of RAC. Any model development demands inputs and output parameters and thus in the present study the input parameters (in kg/m^3) are: cement (C), natural fine aggregate (NFA), natural coarse aggregate–20mm (NC20), natural coarse aggregate–10mm (NC10), recycled coarse aggregate–20mm (RCA20), recycled coarse aggregate–10mm (RCA10), admixture (A), water (W) (Shetty 2005; Deshpande et al. 2014). The 28-day compressive strength of concrete was considered as the output. Using the relevant output and input parameters, models were constructed employing standalone ANN, MGGP, and the novel NMGGP, with a comparative analysis of their performance.

NMGGP is a framework in which the highly weighted genes from standalone MGGP are selected. Each gene is a mathematical equation which displays the contribution of each input parameter and/or the relation between the parameters. Further the calculated value of mathematical expressions emerging out of the selected genes from MGGP make a new input data set which is the transformed data from the input and is now termed as the transformed inputs. These transformed inputs are used as the input parameters in further stage in which the output is 28-day compressive strength of RAC. The output parameter i.e. 28 compressive strength of RAC in the standalone MGGP, ANN and NMGGP remain the same.

MATLAB 2023 platform was used for developing all the models along with GPTIPS2: the open-source GP toolbox to attain the best multigene model for MGGP and NMGGP (Searson 2015). GPTIPS is a flexible script designed for developing MGGP regression models which integrates various genes in a least squares way with a constant value. For standalone ANN, MGGP, and NMGGP models, parameters as population, generations, number of hidden neurons etc. were chosen to optimize performance and achieve the best results. The Feed forward back propagation ANN training was done till the decided number of epochs for ANN and number of generations for MGGP were reached. The allowable number of genes and tree depth in standalone MGGP and NMGGP were optimized to maintain a balance between model complexity and efficiency of the solutions generated. The best models from standalone MGGP, ANN, and hybrid NMGGP were selected based on their fitness value on the validation and testing data. The parameters for the standalone MGGP, ANN and NMGGP are detailed in Tables 2 and 3 respectively. The parameters like population size, generation size, tournament size, crossover rate, termination criteria, number of genes and tree depth etc. for MGGP and NMGGP were selected by combining the analysis of literature review and trial and error basis on the data. Similarly in ANN, the number of hidden neurons, transfer function, training function etc. were selected. Thus, in both the techniques, parameters were decided depending on the fitness and performance of the models along with its computational efficiency.

Table 2. Parameter settings for standalone MGGP.

Parameters	Parameter settings	Parameters	Parameter settings
Size of population	100–500	Criteria for termination	Less than 0.01 as fitness value or 500 generations whichever comes first
Generation numbers	150–500	Maximum of genes	4–8
Selection method	Tournament	Maximum number of tree depth	4–8
Tournament size	13–15	Mathematical operations	+, −, x, /, exp, sqrt, {}
Cross-over rate	0.75–0.84	Mutation rate	0.13–0.20

Table 3. Parameter settings for standalone ANN and NMGGP.

Parameters	Parameter settings
No. of hidden neurons	2–35
No. of epochs	500
Transfer functions	Logsigmodial and linear between input and hidden and between hidden layer and output, respectively
No. of hidden layers	1
Training algorithm	Levenberg–Marquardt

About 70% of the data was utilized for training each of the models (standalone MGGP, ANN, and NMGGP) indicated in Table 3, with the remaining 30% being used for validation (15%) and testing (15%) respectively. The effectiveness of the developed models during the testing phase was evaluated by comparing observed strength of RAC with estimated/predicted strength of RAC using performance metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Correlation Coefficient (*r*). Lower MAE and RMSE and higher *r* values indicate improved (Legates and McCabe 1999). A value closer to 1 of correlation coefficient shows a better agreement of estimated/predicted values with the observed discharge values of any model. The visual evaluation of model performance, comparing observed 28-day compressive strength and the predicted strength, was conducted using a scatter plot and a graph of observed compressive strength to predicted compressive strength. Scatter plots exhibit relationship between changes observed in two different sets of variables and graph shows the closeness and trend of compressive strength prediction.

5. Results and Discussion

This present study aims to predict 28-day compressive strength of RAC in MPa using standalone MGGP and ANN techniques and the novel NMGGP technique. In the phase 1 of the work, the models were developed using relevant inputs i.e. mix design parameters in kg/m^3 and output in MPa. The developed models were evaluated on unseen data during the testing/validation phase, and their performance was analyzed both using statistical performance metrics and visual assessment. Table 4 presents the effectiveness of the models during the validation/testing phase, comparing standalone ANN, MGGP, and the innovative NMGGP approaches.

Table 4 shows good performance of standalone ANN and MGGP models with correlation coefficient of 0.927 and 0.933 respectively, lower RMSE 6.134 and 5.898 respectively and lower MAE values. The ANN performed slightly better than MGGP, reflecting its strength in modelling complex nonlinear input–output relationships. However, MGGP provided interpretable mathematical expressions that explicitly revealed the contribution of each input parameter. However, NMGGP outperforms standalone ANN and MGGP with a higher value of correlation coefficient of 0.965 and lower RMSE (4.744) and MAE (3.944). The improvement in the performance of the NMGGP is due to its two-step process: first, MGGP identifies and extracts the most informative gene expressions (G3, G4, G6), which capture hidden linear, nonlinear, and interaction effects among mix parameters; second, ANN leverages these transformed inputs to refine nonlinear mappings. This synergy allowed the model to both preserve interpretability and improve accuracy.

Table 4. Model performance for the compressive strength prediction.

Model	Error		
	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	Correlation Coefficient (*r*)
NMGGP	4.744	3.944	0.965
Standalone ANN	6.134	4.922	0.933
Standalone MGGP	5.898	5.067	0.927

The accuracy of any model depends on understanding the influence of the inputs on the output. In ANN the influences of inputs can be seen through Hinton diagram and in MGGP through frequency analysis chart. A Hinton diagram is a visual representation of the weights and biases matrix in a neural network. It displays the values of the matrix using colored squares, where the size of the square corresponds to the magnitude of the value and the color represents its sign (positive or negative). With red color denoting negative sign and green the positive weights. Hinton diagram shows the connection strengths in neural networks that higher the width and weight of bars in diagram, higher is the influence of the parameter. Frequency analysis in MGGP on the other hand provides insight into the most influential parameters helping to identify which input variables are most relevant to the output through graphical frequency analysis of individual models or a user-defined portion of the population (Searson 2015). As shown in the Figs. 2 and 3, The Hinton diagram shows cement with an indirect influence on the compressive strength of concrete with highest contributions being that of NC20, RC20, water followed by RC10. Increase in cement content increases the strength of concrete however after a certain limit its increase can decrease the strength of concrete owing to increase in the paste content.

On the other hand, increase in the aggregate content (NC20, RC20) shows a positive weighted indicating its increase increases the compressive strength of concrete. Water on the other hand shows a higher influence as specifically in RAC, more water is required due to the presence of RA, which consists of old mortar content on the aggregates which increase the requirement of water (Deshpande et al. 2014; Zongjin 2011). The analysis of variable usage frequencies within evolving models in frequency analysis shows NFA as the most used parameter followed by RCNC10, RC20 and water as influential parameter which also is seen in the fundamentals of concrete technology (Deshpande et al. 2014; Zongjin 2011) (Fig. 3).

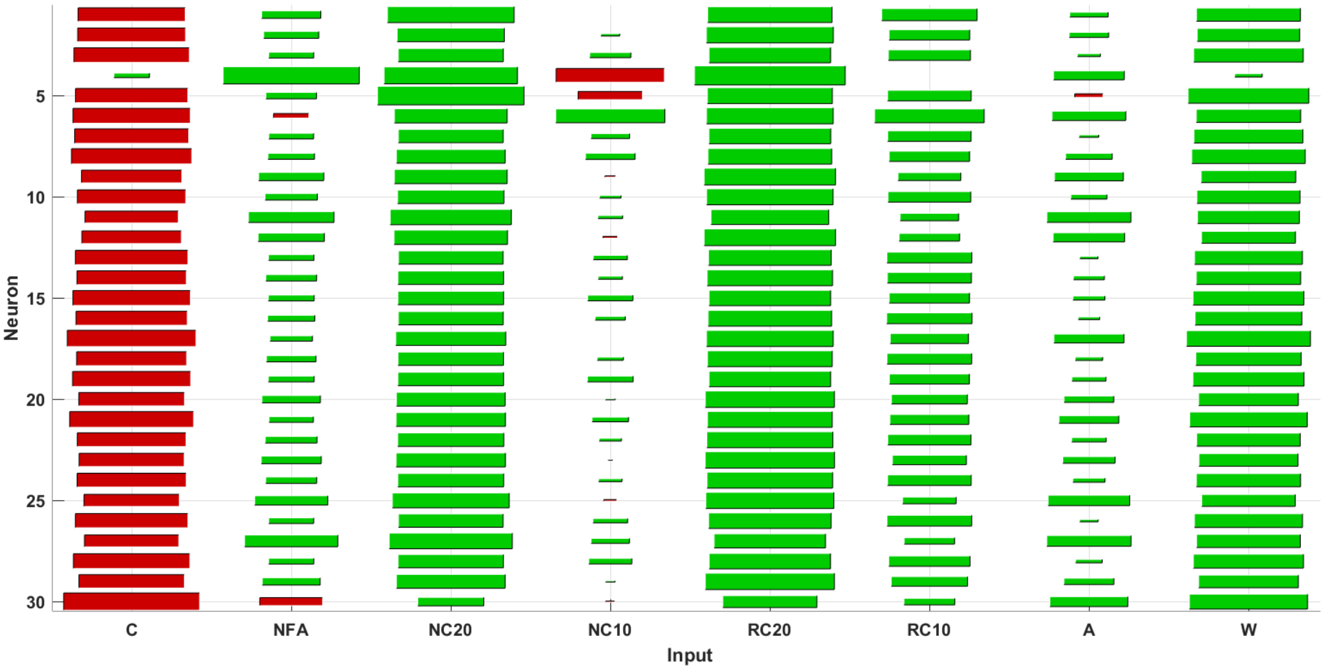

Fig. 2. Hinton diagram with ANN.

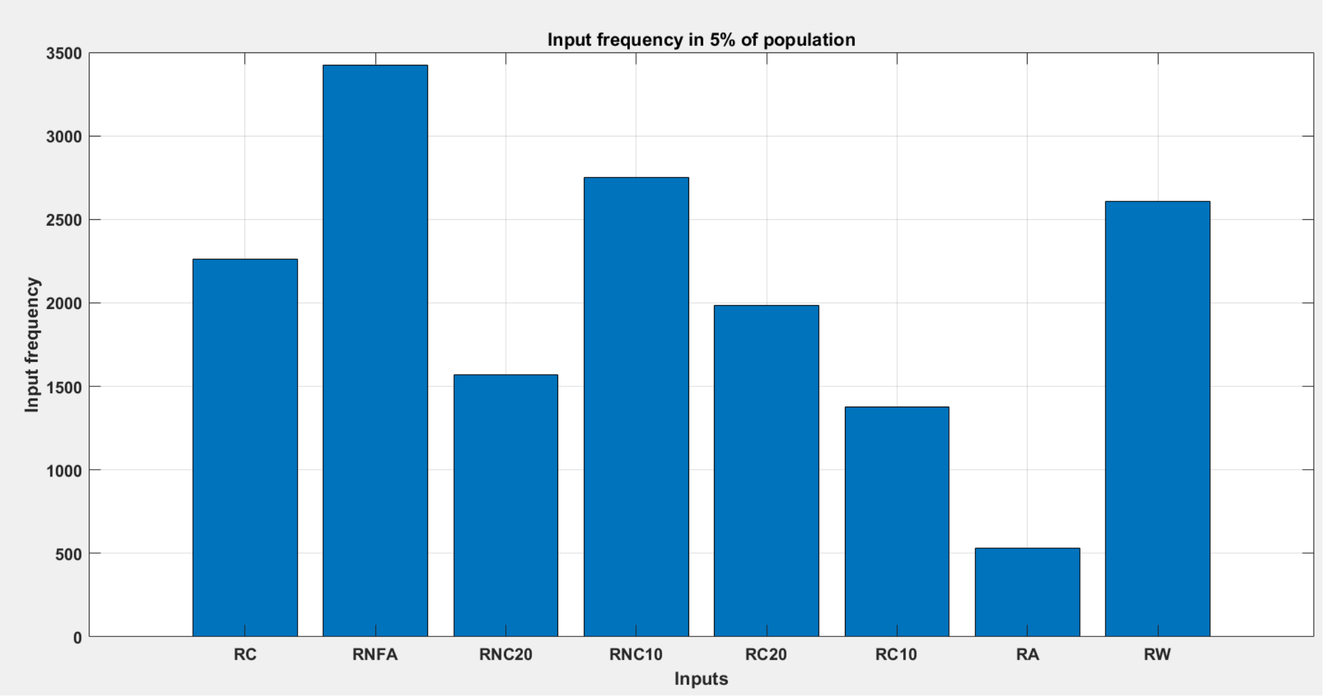

Fig. 3. Frequency analysis with MGGP.

The input parameters, which can effectively influence the desired output can be identified and used for model development. However, with only the input parameters, the interrelation between the input parameters is not captured and thus to enhance the model's performance, these input parameters can undergo transformation, as demonstrated in this study. Transforming input parameters enables the discovery of hidden relationships and patterns within the data that may not be readily apparent in their original form. By generating new features from the input parameters, the transformed inputs reveal critical insights that improve the model’s ability to make accurate and reliable predictions. The first step towards the same in the present study is to use the input parameters in the standalone MGGP which are proportions in kg/m^3 of cement (C), natural fine aggregate (NFA), natural coarse aggregate–20mm (NC20), natural coarse aggregate–10mm (NC10), recycled coarse aggregate–20mm (RCA20), recycled coarse aggregate–10mm (RCA10), admixture (A), water (W) and output as 28 day compressive strength of concrete (in MPa) in standalone MGGP. With the use of this input-output combination, the model's accuracy evolves with each generation during the training phase and is shown in Fig. 4, which is a graphical representation of a GPTIPS iterations seen in terms of predictive performance (RMSE error on training data). The blue line in the first half off the graph shows that the fitness of the 'best' model in population is achieved at 440 generations, after which the fitness curve becomes smoother and further iterations or of MGGP will not show any significant change in the fitness. The orange line in second part of graph which represents the mean performance of the population of models also show the best performance at 440 generations. It indicates that running the genetic programming for more generations does not result in a more favourable outcome. Thus, it suggests that the genetic programming algorithm should have to run for at least 440 generations (Pandey et al. 2015).

The trees developed for standalone MGGP i.e. for the respective genes are shown in Fig. 5 along with the mathematical equations in Table 5. Along with the mathematical expression for each gene, Fig. 6 also shows the weighing coefficient (of gene weights) for each gene. The weighting coefficient determines the importance or contribution of each gene to the final output.

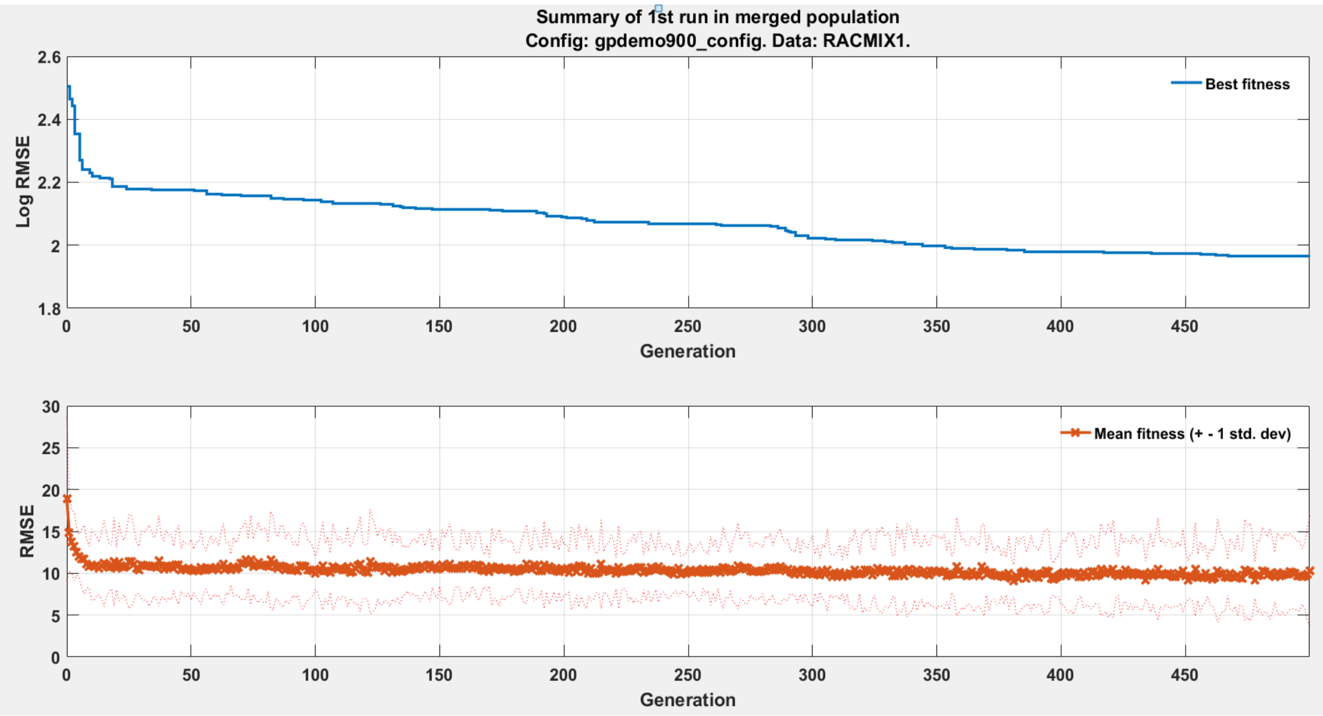

Fig. 4. Summary of MGGP runs for the model.

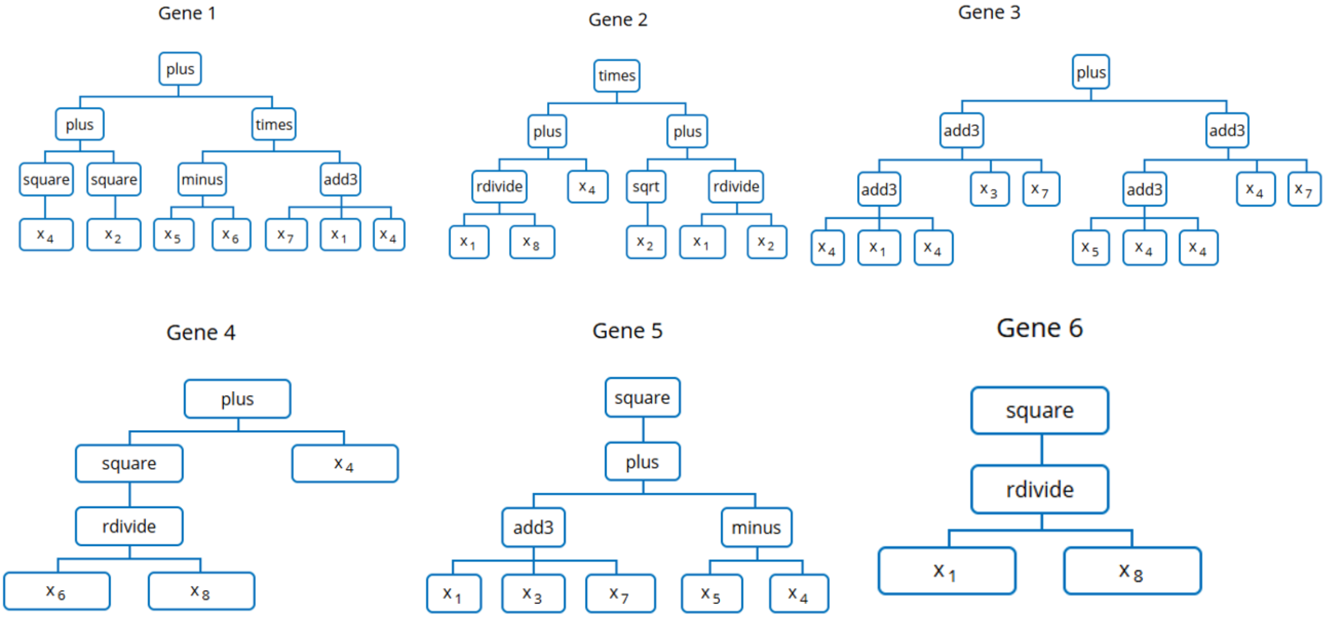

Fig. 5. Genes (trees) generated for the model.

Table 5. Genes and mathematical expression for the model.

Term	Value
Bias	–7.3
Gene 1	$-1.66e-5(RC20-10\cdot RC10)(C+NC10+A)-1.66e-5\cdot NFA^2-1.66e-5\cdot NC10^2$
Gene 2	$0.0257\cdot(NC10+C/W)(C/NFA+NFA^{1/2})$
Gene 3	$0.0861\cdot C+0.0861\cdot NC20+0.431\cdot NC10+0.0861\cdot RC10+0.172\cdot A$
Gene 4	$-(1.18\cdot(NC10\cdot W^2+RC10^2))/W^2$
Gene 5	$-4.1e-5\cdot(C+NC20-1.0\cdot NC10+RC20+A)^2$
Gene 6	$(4.3\cdot C^2)/(W^2)$

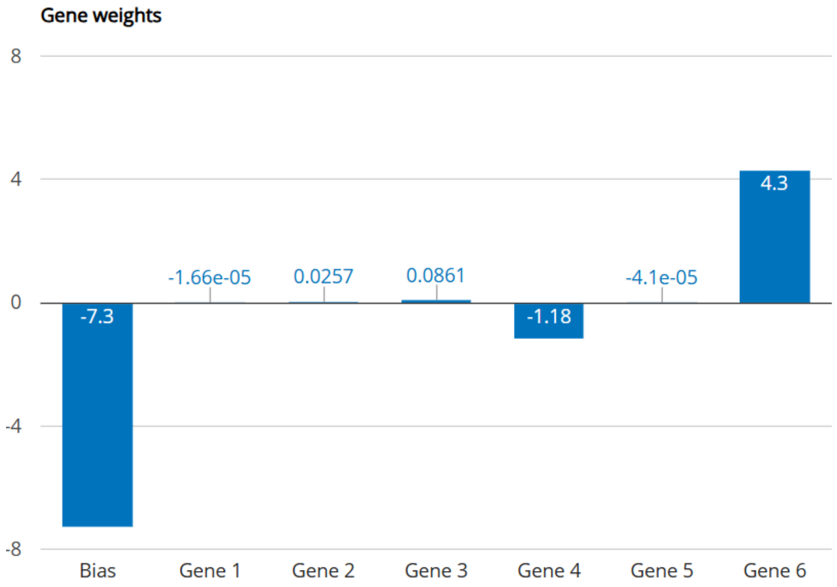

Fig. 6. Weighing coefficient (of gene weights) for each gene.

The mathematical expressions developed for each gene is a linear combination of highly weighted genes and can be further combined as input parameters for the third stage of developing NMGGP. Thus, in the first stage of standalone MGGP, only the original input parameters are considered to predict the output (28-day compressive strength of RAC). However, since these genes have a weighing coefficient which signifies the importance of that gene in prediction of output, an approach of using only the highest weighted genes for prediction of output can be attempted. The mathematical equation of each gene can unveil the interrelation between the original input parameters which is missing in the standalone modelling. Thus, in the stage 2 of the work, the high weighted suitable genes from standalone MGGP model are selected and the remaining genes are neglected. In the present RAC model, the selected genes are Gene 3, Gene 4 and Gene 6 respectively owing to their higher weighing coefficient (weight of G3–0.0861, weight of G4–1.18 and weight of G6–4.3). Further stage 3 includes creating new inputs termed as transformed inputs (TG) from original ones to capture relationships between the input parameters. Since evaluated mathematical expressions of high weighted genes are used as transformed input parameters for ANN in the next phase of NMGGP, the hypothesis of improved accuracy in model can be explained. The transformed inputs i.e. using gene 3 as TG3, using gene 4 as TG4 and using gene 6 as TG6 from the selected genes as mentioned are used for estimation and prediction of output using ANN in the third stage of the model development.

The gene 3 with its mathematical expression: $[0.0861\cdot C+0.0861\cdot NC20+0.431\cdot NC10+0.0861\cdot RC10+0.172\cdot A]$; shows a linear combination of the input parameters and each input contributes proportionally to its weight (coefficient). All the inputs show a direct proportionality towards 28-day compressive strength of RAC. An increase in the respective input (according to the coefficient) will show an equivalent increase in the output. NC-10 is most important input parameter with 0.431 as its positive coefficient followed by admixture (A) showing a notable impact (positive coefficient of 0.172). Cement (C) and NC20 have equal but comparatively less impact. Gene 4 with the equation: $[-(1.18\cdot(NC10\cdot W^2+RC10^2))/W^2]$ describes a relationship between variables NC10, RC10, and W, with a positive constant added at the end (weighted coefficient of bias). The NC10 term remains constant, while the RC10 becomes less significant as W increases, due to the inverse square relationship. The equation outputs a value that increases with larger W, depending on how NC10 and RC10 interact. Gene 6 with the equation:

$[((4.3 \cdot \mathrm{C}^2)/(\mathrm{W}^2)) - 9.72]$ shows a relationship between variables C and W, where C is squared and scaled by 4.3, then divided by the square of W. This shows that the term increases as C increases and decreases as W increases. The result of this fraction is then reduced by a constant value of 9.72. Thus, each gene with its mathematical expression shows the inter relationships between the inputs and further they are transformed to transformed inputs. Thus, the value of transformed inputs derived through the respective weighted genes capture the linear, nonlinear and interaction between the original inputs which can be useful to increase the efficiency of the ANN model in next stage of 28-day compressive strength prediction. Thus, it can be said that MGGP also understands the fundamentals of the problem and further is transferred to the NMGGP method as well. Further in the third stage the transformed inputs - TG3, TG4 and TG6 along with the 28-day compressive strength of RAC as output is modelled using ANN. A trial-and-error approach was used to select the number of hidden neurons for development of ANN model with transformed inputs and respective output and the architecture was 3:5:1. The performance of NMGGP shown in table 2 outperforms the standalone MGGP and ANN model. The results show a 21.794% decrease of RMSE than that of ANN and 19.566% decrease from MGGP using NMGGP. Similarly, a 16.423 % and 22.163% decrease in MAE is seen for model developed using NMGGP technique and standalone ANN and standalone MGGP. To compare the statistical significance of the observed values of 28 days compressive strength of RAC with the same predicted by NMGGP, ANN and MGGP, a single-factor ANOVA test was used. The F value, F critical value and p value for NMGGP is 0.255, 0.614 and 3.913 respectively, for ANN is 0.367, 0.83 and 3.913 respectively and for MGGP is 0.066, 0.797 and 3.913 respectively. The values indicate that the F-value was less than the F-critical, and the P value was more than 0.05, indicating that the discrepancy between the expected and actual values was not statistically significant (Parveen et al. 2019).

The framework's synergy between genetic programming and neural networks ensures better non-linear mapping, robustness against overfitting, accelerating efficiency while maintaining interpretability. NMGGP creates a better representation of the data, enhancing the ANN's ability to learn by better capturing patterns in the data through influential transformed genes as inputs. Visual analysis of the model performance in form of scatter plot (Fig. 7) shows a balanced scatter with NMGGP achieving the nearest value for observed compressive strength of RAC for lower and higher values of strength. No obvious under or over-predictions can be seen in the scatter plot in Fig. 7 for lower or higher strength values. It can also be seen that predictions of 28-day strength done, on testing dataset by model 1 i.e. NMGGP shows a higher correlation and lower RMSE with a lower standard deviation of 13.631, 15.399 with standalone MGGP and 15.451 with standalone ANN. The maximum observed compressive strength is 100.5 MPa and that predicted by NMGGP is closer to the observed with 91.917 MPa and that of standalone MGGP with 95.724 MPa and lower with standalone ANN as 90.366 MPa. The prediction of 28-day compressive strength of RAC shows a similar trend as that of observed compressive strength of RAC as shown in Fig. 8.

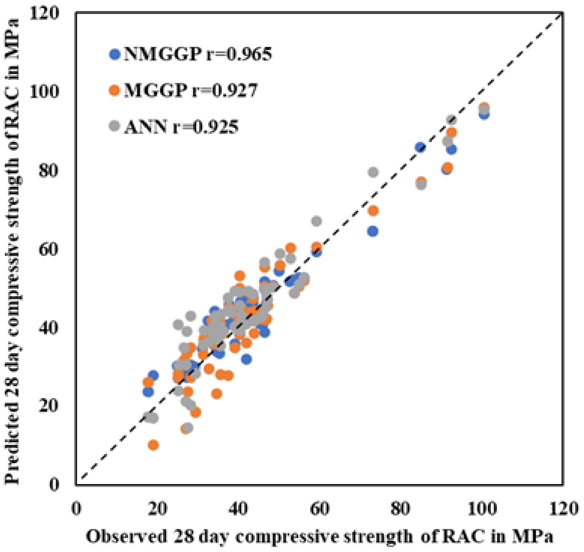

Fig. 7. Scatter plot for the model.

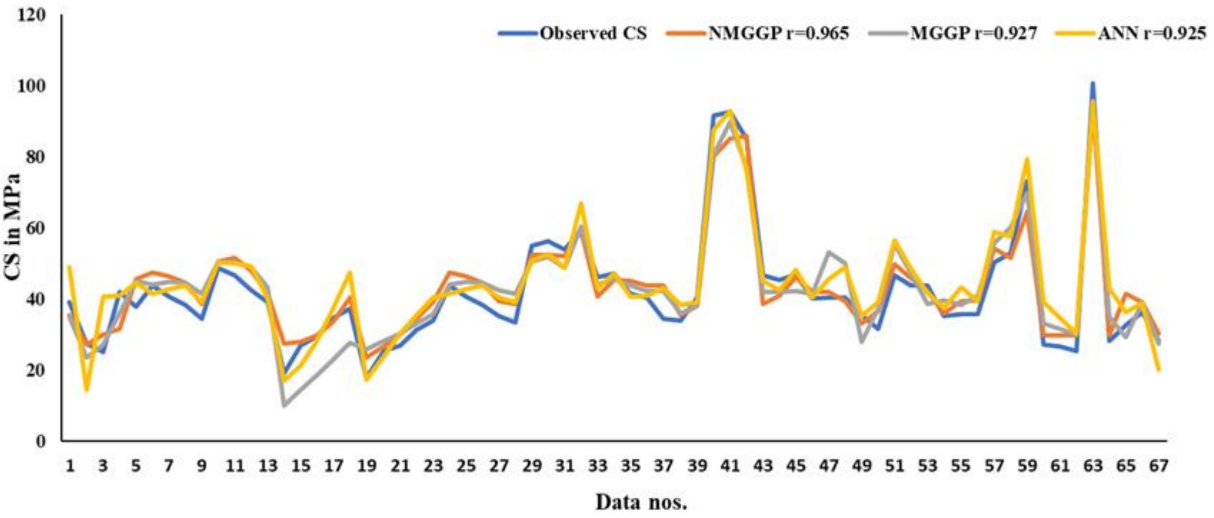

Fig. 8. Variations of observed and predicted compressive strength values.

Thus, the novel NMGGP technique integrates the interpretability of MGGP with the nonlinear learning capacity of ANN to provide a more accurate model for predicting the 28-day compressive strength of RAC. The reduction in prediction errors and improved correlation confirm the effectiveness of NMGGP framework and feature transformation. Importantly, this approach not only enhances accuracy but also ensures that the relationships among mix design parameters are better understood.

6. Conclusions

The present study proposes a novel hybrid framework: Neuro Genetic Programming framework (NMGGP), which integrates Multi-Gene Genetic Programming (MGGP) with Artificial Neural Networks (ANN) in a structured three-stage process. Within this framework, the most influential genes, representing optimized mathematical expressions obtained through standalone MGGP, are selected and used as transformed input parameters for the ANN. The ANN subsequently utilizes these inputs to enhance the accuracy of compressive strength prediction. The developed NMGGP model is applied for predicting the 28-day compressive strength of Recycled aggregate concrete (RAC) by using the mix design parameters as input parameters. The study shows that:

- The standalone MGGP and ANN models efficiently predict the 28-day compressive strength of RAC with a correlation coefficient of 0.927 with MGGP and 0.933 with ANN. Lower RMSE (MGGP:6.134 and ANN:6.134) and lower MAE (MGGP:5.067 and ANN:4.922)
- NMGGP outperforms Standalone ANN and MGGP models which can be seen through higher correlation coefficient of 0.965, lower RMSE (4.744) and MAE (3.944) values. NMGGP displays a 21.794% decrease of RMSE than that of standalone ANN and 19.566% decrease from standalone MGGP using NMGGP. Similarly, a 16.423 % and 22.163% decrease in MAE is seen for model developed using NMGGP technique and standalone ANN and standalone MGGP respectively. This highlights the effectiveness of feature transformation and hybridization in enhancing predictive performance.
- NMGGP is characterized by removing the genes with less weight and using only the informative genes termed as transformed inputs in the form of expressions, which add to the performance of the models. These informative transformed inputs with ANN can use the meaningful relations between them enhancing its predicting efficiency.
- The visual performance analysis of models in form of scatter plot and comparative graphs confirms the robustness and reliability of NMGGP by showing a balanced scatter and similar trends of predictions.

Acknowledgements
None declared.

Funding
The authors received no financial support for the research, authorship, and/or publication of this manuscript.

Conflict of Interest
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this manuscript.

Author Contributions
All of the authors made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; were involved in drafting the manuscript or revising it critically for important intellectual content; and gave final approval of the version to be published.

Data Availability
The datasets created and/or analyzed during the current study are not publicly available, but are available from the corresponding author upon reasonable request.

REFERENCES

Adriana TAD, Monica BL, Koji de JN (2013). Prediction of compressive strength of concrete containing construction and demolition waste using artificial neural networks. *Construction and Building Materials*, 38, 717–722.

Bayazidi BM, Wang GG, Amir H, Alavi H, Gandomi AH (2014). Multigene genetic programming for estimation of elastic modulus of concrete. *Mathematical Problems in Engineering*, 2014, 474289.

Deshpande N, Londhe, S (2014). Modelling compressive strength of recycled aggregate concrete using neural networks and regression. *Challenge Journal of Concrete Research Letters*, 5(2), 15–21.

Duan ZH, Kou SC, Poon CS (2013). Prediction of compressive strength of recycled aggregate concrete using artificial neural networks. *Construction and Building Materials*, 40, 1200–1206.

Hamada BS, Dawib AH. (2017). Sustainable normal and high strength recycled aggregate concretes using crushed tested cylinders as coarse aggregates. *Case Studies in Construction Materials*, 7, 228–239.

Hsu KL, Gupta HV, Sorooshian S (1995). Artificial neural network modelling of the rainfall-runoff process. *Water Resource Research*, 31(10), 2517–2530.

Khademi F, Jamal SM, Deshpande N, Londhe S (2016). Predicting strength of recycled aggregate concrete using Artificial Neural Network, Adaptive Neuro-Fuzzy Inference System and Multiple Linear Regression. *International Journal of Sustainable Built Environment*, 5(2), 355–369.

Koza J (1992). Genetic Programming: On the Programming of Computers by Means of Natural Selection. A Bradford Book, MIT Press. London, England, 87–112.

Kulkarni PS, Londhe SN, Dixit PR (2019). A comparative study of concrete strength prediction using artificial neural network, multigene programming, and model tree. *Challenge Journal of Structural Mechanics*, 5(2), 42–61.

Legates DR, McCabe GJ (1999). Evaluating the use of 'goodness of fit' measures in hydrological and hydroclimatic model validation. *Water Resource Research*, 35(1), 233–224.

Li S, Xie E, Yang J (2022). Daily suspended sediment forecast by an integrated dynamic neural network. *Journal of Hydrology*, 604, 127–258.

Lin GF, Wu MC (2009). A hybrid neural network model for typhoon-rainfall forecasting. *Journal of Hydrology*, 375(3–4), 450–458.

MathWorks (2023). MATLAB. https://in.mathworks.com/products/matlab.html [accessed 01-04-2025].

Onyelowe KC, Gnananandara T, Ebid AM, Mahdi HA, Ghadikolaee MR, Al-Ajamee M (2022). Evaluating the compressive strength of recycled aggregate concrete using novel artificial neural network. *Civil Engineering Journal*, 8(8), 1700–1713.

Pan L, Wang Y, Li K, Guo X (2022). Predicting compressive strength of green concrete using hybrid artificial neural network with genetic algorithm. *Structural Concrete*, 24(2), 1980–1996.

Pandey DS, Pan I, Das S, Leahy JJ, Kwapinski W (2015). Multi-gene genetic programming based predictive models for municipal solid waste gasification in a fluidized bed gasifier. *Bioresource Technology*, 179, 524–533.

Rumelhart DE, Hinton GE, Williams RJ (1986). Learning representations by back-propagating errors. *Nature,* 323(6088), 533–536.

Searson DP (2015). GPTIPS 2: an open-source software platform for symbolic data mining. Chapter 22 in Handbook of Genetic Programming Application, New York, 551–573.

Shetty MS (2005). Concrete Technology, 17th ed. S. Chand and Company, New Delhi.

Sihag P, Sahar MK, Angelaki A (2019). Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity. *Applied Water Science*, 9, 129.

Tran QV, Dang QV, Ho LS (2022). Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Construction and Building Materials*, 343, 126578.

Zongjin Li (2011). Advanced Concrete Technology, 1st ed. John Wiley and Sons Inc., New Jersey.