Multidisciplinary Approaches in Artificial Intelligence

Multidisciplinary Approaches in Artificial Intelligence

Edited by

D. S. Kalana Mendis

Cambridge Scholars Publishing

Multidisciplinary Approaches in Artificial Intelligence

Edited by D. S. Kalana Mendis

This book first published 2024

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-0364-1599-0
ISBN (Ebook): 978-1-0364-1600-3

CONTENTS

ACKNOWLEDGEMENTS

The academic editor, Dr D.S. Kalana Mendis, is grateful to all the authors who contributed chapters for the book. Also, my special acknowledgements to Emeritus Professor W.P.S. Dias, who advised me with the manuscript preparation. Further, my acknowledgements go to the following academics and organisations for encouraging me to develop a research culture:

Emeritus Professor Hemaka Sriyananda, Open University of Sri Lanka

Senior Professor Asoka S. Karunananda, University of Moratuwa, Sri Lanka

Senior Professor Sudath R D Kalingamudali, University of Kelaniya, Sri Lanka

Professor Uditha Ratnayake, Open University of Sri Lanka

Professor Thrishantha Nanayakkara, Imperial College, London, U.K.

Sri Lanka Association for Artificial Intelligence

Sri Lanka Association for the Advancement of Science

Royal Asiatic Society of Sri Lanka

Institute of Physics Sri Lanka

SAARC Cultural Centre, Colombo, Sri Lanka

PREFACE

The book "Multidisciplinary Approaches in Artificial Intelligence" contains publications on the theory, applications, and design methods of intelligent systems and intelligent computing. Virtually all disciplines, such as engineering, natural sciences, computer and information science, ICT, e-commerce, environment, healthcare, and life science, are covered. The list of topics spans all the areas of modern intelligent systems and computing such as computational intelligence, soft computing, including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, cognitive science and systems, Perception and Vision, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, intelligent control, robotics and mechatronics, including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence, and multimedia. Multidisciplinary approaches to artificial intelligence open significant opportunities in several areas, such as industry, medicine, energy, security, transportation, and education. This book provides theory and application development using artificial intelligence techniques by organising intelligent systems for many applications for the benefit of humanity. The book covers a range of audiences, from academicians, practitioners, researchers, and students, to stakeholders. It can support graduate students and interns to develop a deep understanding of the latest paradigms in artificial intelligence techniques.

Dr. D.S. Kalana Mendis
Senior Lecturer
Department of Information Technology
Advanced Technological Institute
Dehiwala
Sri Lanka

CHAPTER 1

MACHINE LEARNING APPROACHES TO PREDICT COMPRESSIVE STRENGTH OF RECYCLED AGGREGATE CONCRETE

W.K.V.J.B. KULASOORIYA & W.P.S. DIAS

Abstract

Recycled Aggregate Concrete (RAC) is made of aggregates obtained from construction demolition wastes. RAC is used in the construction industry because it conserves natural resources and lowers the impact on the environment. Mechanical properties of RAC are generally obtained through laboratory experiments, which are time-consuming and costly. Also, the strength characteristics of RAC exhibit non-linear variation when various constituents are involved. At the same time, machine learning has been used to predict the strength characteristics of RAC, since it is time efficient and identifies non-linear patterns in the data. This study used three regression models: Multiple Linear Regression (MLR), Random Forest (RF), and Simple Neural Network (SNN) to predict the compressive strength of RAC by using 118 data samples extracted from a related work. Effective Water Cement Ratio (EWCR), Aggregate Cement Ratio (ACR), Recycled Aggregate Percentage (RA%), Water Absorption (WA), and Los Angeles Abrasion (LAA) were employed as input features; and Compressive Strength (CS) was the output feature. These input features were included in four different subsets, with each subset incorporating different numbers and combinations of features, ranging from combining only three features, i.e. EWCR, ACR, and RA%, to including all five. Subset 4, comprising all the features, yielded the best fits. The comparison revealed that the RF model outperforms the remaining models as the best model with a Mean Absolute Error (MAE) of 5.9 MPa. In contrast, the MAE values for MLR and SNN were 9.4 and 9.6 MPa respectively. The RF model was explained using the Shapley Additive Explanation (SHAP) to interpret the parameter contributions.

EWCR was the most influential parameter. The other parameters of importance were ACR and RA%.

Introduction

Infrastructure development is critical for the economic growth, transportation, and urbanisation of any country. Since the early 1990s, concrete has dominated the market as the most popular material in the construction industry (Walberg 2016, 16-17). The demand for concrete has been rising rapidly for several years. Next to water, it is the most widely consumed resource on Planet Earth (Mehta 2009, 1). Concrete is a composite material made from several components, namely water, aggregates, and cement. Examples of extra elements that can be added to enhance the quality and characteristics of concrete are admixtures and additives.

Concrete production involves a series of steps that transform the raw materials into the finished product. This production process contributes significantly to environmental pollution through the emission of various pollutants such as greenhouse gases; and through natural resource depletion. The primary greenhouse gas emitted during concrete production is carbon dioxide (CO_2). While limestone is being processed to make cement, this gas is emitted into the environment (Adesina 2018, 1-2). Other greenhouse gases, such as methane and nitrous oxide, are also released throughout the manufacture and delivery of raw materials. Apart from that, the energy consumption of concrete production is very high, and such energy consumption itself generates additional CO_2 emissions. Therefore, the inevitable increases in concrete production will lead to environmental contamination as well as the consumption of more resources (Adesina 2018, 1-2).

Construction and demolition activities generate large amounts of waste concrete annually. The majority of this waste is left in selected areas illegally or dumped as landfill materials. Building construction waste could represent up to around 85% of the entire amount of solid waste (Batayneh, Marie, and Asi 2007, 1870). Around 90% of this waste from the construction industry is disposed of on the land. The Environment Protection Agency estimated that concrete waste accounts for 70% of waste generated during construction and demolition activities (US EPA 2017). Concrete waste refers to any material generated during production, handling, transportation, or use of concrete that is not utilised and discarded. This concrete waste, such as broken concrete pieces, concrete slurry, concrete dust, and concrete blocks, takes up a significant amount of space in landfills and contributes to environmental pollution. The disposal of concrete wastes could also result in greenhouse gas emissions. In order to achieve carbon and climate neutrality by 2050, construction operations, which include the production of concrete and the generation of concrete wastes, are critical (Rosa et al. 2022). The recycling of broken concrete pieces (generally arising out of demolition) as aggregate for fresh concrete production is one way to reduce waste and emissions. Given the overall volume of global concrete production, incorporating recycled aggregate stands out as a key strategy for optimising concrete (Sarshar and Khoury 1993, 194).

Recycled Aggregate Concrete (RAC)

Recycled Aggregates (RA) are materials that have been utilised in construction and demolition operations in the past and processed to be used again. Generally, crushed concrete is separated into three classes. These are coarse recycled aggregates (particle size greater than 4.75mm), fine recycled aggregates (particle size is between 4.75mm and 0.0075mm), and recycled concrete fines (particle size of 0.075mm or less) (Abbas et al. 2006; Behnood, Olek, and Glinicki 2015). The composition of the RA is contingent upon the materials accessible in the surrounding region and the specifications of the intended application. RA is basically produced from concrete rubble after crushing, cleaning, and grading; and eliminating construction debris such as gypsums, glass, plastics, rebars, wood, and paper (Wagih et al. 2013, 194). Wang et al. (2021, 5) state that an RA production process includes nine steps, namely size reduction, manual and mechanical pre-separation, primary crushing, magnetic separation, secondary screening, decontamination, secondary crushing, washing/screening/sifting, and division into size fractions. Normal Aggregates (NA) are materials that are mined or extracted from the earth, such as sand, gravel, and crushed stones. As a result, there are variations across the properties of both NA and RA. The surface of recycled aggregates has a mortar layer which results in low density, high water absorption, and high porosity (Kwan et al. 2011, 566-567; Tam et al. 2020, 2). The water absorption of RA ranges between 3% and 12% for both the fine and coarse fractions (Gomez-Soberon 2002, 1302-1303). The actual amount of water absorption depends on the kind of concrete waste utilised in the production of RA.

The process of making Recycled Aggregates Concrete (RAC) includes substituting either a portion or entire amount of Natural Aggregates (NA) within a concrete blend with Recycled Aggregates (RA). Normal structural concrete can be made using RA, generally with the addition of condensed silica fume, and fly ash (Rao, Jha, and Misra 2007, 72-74). Before replacing the NA with RA, the latter should be tested for quality, particle size distribution, strength, and durability. According to Wang et al. (2021, 171), the mechanical characteristics and long-term outcomes of RAC can be improved in three ways, namely reducing the porosity of RA, reducing the layer of old mortar on the surface of RA, and improving the properties without changing the RA. Wagih et al. (2013) determined how the water-cement ratio and replacement ratio of the RA, as well as the material and moisture of the aggregates, affect the rate of increase or decrease of strength in the RAC. With NA completely replaced in concrete, the RAC has a slightly higher air content (around 4% to 5.5%), and the higher porosity of the RA could cause this increased air content (Katz 2003, 706). Butler, West, and Tighe (2014, 143-144) investigated the connections between aggregate parameters to study how the density of RAC depended on the density of the adhered mortar itself, the amount of adhered mortar, and the density of the original aggregate. The concrete formed with NAC has a bulk density of around 2400kg/m^3, whereas in RAC it is around 2150kg/m^3, irrespective of the cement used in the concrete (Topçu and Günçan 1995, 1388). Akbarnezhad et al. (2011, 3477-3478) determined that the compressive strength of RAC is substantially impacted by the mineral fines, admixture dosage and variety, aggregate type, aggregate grade, substitution rate, sand content, and water cement ratio. In comparison with NAC, RAC can sometimes exhibit a reduction in compressive strength by up to 40% (Evangelista and de Brito 2014, 144-148).

Machine Learning in Concrete Technology

The use of laboratory experiments to evaluate and compare the properties of RAC and NAC are time-consuming, overpriced, and requires specific equipment and skilled technicians. Hence, machine learning (ML) techniques can be used to predict the strength characteristics of RAC, based on a relatively limited set of experimental data. Hybrid Machine Learning (HML) algorithms combine multiple types of ML techniques or models to improve the accuracy and performance of given data. In the early days, Flood and Kartam (1994, 48-131) explained the importance of understanding and potentially using neural networks in Civil Engineering. ML involves training a model or algorithm to recognise patterns and make predictions about new data by using statistical techniques. Dias and Pooliyadda (2001, 371-379) successfully predicted the strength of high-strength concrete using neural networks. Machine Learning (ML) is now becoming increasingly popular in civil engineering as it can lead to reduced experimentation.

De-Prado-Gil et al. (2022, 17), using an ML model, evaluated the splitting tensile strength of self-compacting recycled concrete and found correlation coefficients (R) for Scaled Conjugate Gradient Backpropagation (SCGB), Levenberg-Marquardt (LM), and Bayesian Regularization (BR). The best R-value of 0.91 was given by the BR model. Gradient boosting (GB) and Random Forest (RF) are two different ensemble machine-learning models that have been successfully used to evaluate the compressive strength and flexural strength of RAC, yielding better results compared to individual machine-learning methods such as Artificial Neural Networks (ANN) and Support Vector (SV) machine-learning methods. Yuan et al. (2022, 20-21) found an R^2 value of 0.86 for flexural strength and 0.91 for compressive strength prediction for the RF model, and R^2 values of 0.79 for flexural strength and 0.87 for compressive strength for the GB model. In order to forecast the elastic modulus of natural coarse aggregate components that are partially or completely replaced by recycled coarse concrete aggregate, the RF model and the support vector machine (SVM) were combined by Han et al. (2020, 9). Compared to the independent models, ensemble ML models always yield more accurate predictions. According to Nunez, Marani, and Nehdi (2020, 19), GB, Deep Learning (DL), and the Gaussian Process (GP) successfully captured the compressive strength of RAC. Three sophisticated machine-learning models were tuned, trained, and tested using the data set. During the tuning process, K-fold cross-validation was used to make sure the developed models broadly represented the compressive strength of RAC. Mai et al. (2023, 18-21) utilised ML models such as the GB model, Extreme Gradient Boosting (EGB), Light Gradient Boosting (LGB), Stacking, AdaBoost, and Voting to predict the behaviour of brick aggregate concrete. With an R^2 value of 0.95, the results demonstrated that the SM (Stacking Model) had the best prediction capability. Peng and Unluer (2023, 11-12) investigated the complex functional relationship between significant variables like the compressive strength of RAC, RA properties, and mix proportion using two standard algorithms, namely ANN and SV regression, as well as two hybrid models that were optimised – i.e. the grey wolf optimiser-based SVR (GWO-SVR) and the particle swarm optimisation-based SVR (PSO-SVR). The results obtained for R^2 by using the above models of GWO-SVR, PSO-SVR, ANN, and SV were 0.91, 0.89, 0.76, and 0.59, respectively. The determination of the strength of RAC by using DL showed higher efficiency, generalisation ability, and precision compared to the traditional neural networks. Deng et al. (2018, 568-569) used various mix ratios in 74 sets of concrete masonry blocks to forecast the compressive strength of the RAC by using a Convolutional Neural Network (CNN). Ben Chaabene, Flah, and Nehdi (2020, 15) determined elastic modulus, and compressive, tensile, and shear strength by using four major types of ML models known as ANN, decision tree (DT), SVM, and evolutionary algorithms (EA). Using the ANN model, the compressive strength of RFC (Recycled Fibre Concrete), RAC, and NAC was evaluated. The model indicated that the RFC and RAC with the suggested mix design had the highest compressive strength. The coefficient of consistency (COC) was evaluated using SVM and ANN models. Both models performed better than Multiple Linear Regression (MLR), with the SVM having the greatest accuracy.

It is essential to include the properties of both NA and RA when analysing the characteristics of RAC. It is also important to utilise relevant concrete technology ratios such as water/cement ratio and aggregate/cement ratio (which are established domain knowledge parameters) rather than only considering raw inputs. The improved performance obtained by using such ratios has been demonstrated by Dias and Pooliyadda (2001, 371). The study described herein focuses on the use of such ratios rather than raw data and generates models for RAC using different sets of parameters on the one hand, and different ML algorithms on the other.

Methodology

Data Collection

The data for this study were obtained from a study by Yuan et al. (2022, 4), who used 12 parameters and 680 data points to assess the compressive strength and flexural strength of RAC using machine learning algorithms. These parameters were effective water-cement ratio, aggregates-cement ratio, percentage of RA, parent concrete strength, bulk density of RA, bulk density of NA, nominal maximum size of RA, nominal maximum size of NA, water absorption of RA, water absorption of NA, Los Angeles abrasion of RA, and Los Angeles abrasion of NA. From these parameters, the selected key parameters for this study were the Effective Water Cement Ratio (EWCR), Aggregate Cement Ratio (ACR), RA%, Water Absorption (WA), and Los Angeles Abrasion (LAA) for evaluating Compressive Strength (CS). Note that the approach used in our study was targeted at finding the key parameters that affected CS, rather than being overly comprehensive. There were three ways in which this was done. First, the parameters chosen were those that were well established in the literature. Second, they featured two of the three key parameters obtained by Yuan et al. (2022, 19) as well – namely EWCR and RA%. Note that we did not consider parent concrete strength – obtained as influential by Yuan et al.

(2022, 3), because that parameter does not feature very strongly in the concrete technology literature. Third, we used a combined index for water absorption and Los Angeles Abrasion by weighting those values for RA and NA by their relative proportions. Only 118 data points had complete information on all parameters, and hence only such points were used. For modelling purposes, the five selected parameters were categorised into four different subsets, with increasing numbers of input parameters, as given below. EWCR, ACR, and RA% are included in all subsets, based on their importance as reported in the literature. Subset 1 includes only these three parameters, while Subsets 2 and 3 add WA and LAA respectively; and Subset 4 includes all parameters.

(i) Subset 1: EWCR, ACR, RA%
(ii) Subset 2: EWCR, ACR, RA%, WA
(iii) Subset 3: EWCR, ACR, RA%, LAA
(iv) Subset 4: EWCR, ACR, RA%, WA, LAA

Model Selection

The selected machine learning algorithms for this study were Multiple Linear Regression (MLR), Random Forest Regression (RF), and Simple Neural Network (SNN). MLR was selected from regression models, RF was chosen from tree-based models, and SNN was chosen from neural network models. MLR provides interpretability and simplicity as a model, making it suitable for linear relationships. RF, with its ensemble of decision trees, offers non-linear handling, resilience, and flexibility. SNN combines non-linear capabilities with simplicity and is one of the earliest ML applications. This heterogeneous group ensures a thorough investigation of the data, resulting in more reliable conclusions that can be derived across a range of patterns and structures in the dataset. RF is the most complex and likely to yield the most accurate predictions. MLR is the simplest, capable of generating only linear relationships; yet it will readily give a comparison of the relative influence of input parameters through its p-statistics. ANN was expected to hold an intermediate position between simplicity and accuracy. Note that Yuan et al. (2022, 1) used Random Forest (RF) and Gradient Boosting (GB) as their models for this same data.

Model Fitting

All three models were trained to predict the Mean Absolute Error (MAE) of CS predictions. The difference between the Actual CS and the Predicted CS was determined for each data point. After taking the absolute value of this difference, the total absolute differences were summed across all data points, and the MAE was computed by dividing the total sum of absolute differences by the number of data points, which was 118 in this case. Training and testing were conducted using the complete set of data; hence these exercises should more correctly be termed as model fitting. The independent parameters (features) were EWCR, ACR, RA%, WA, and LAA, combined into four separate subsets, and the dependent parameter (target variable) was CS.

For MLR, the predicted CS was calculated, using Microsoft Excel, as the sum of the intercept coefficient and product of each coefficient of a feature that was given by MLR and its corresponding value in the dataset.

In the RF model, Python was used on Jupyter Notebook for modelling purposes. Jupyter Notebook is an interactive, open-source web application that is commonly used in data analysis, scientific research, machine learning, and software engineering. To develop the RF regression model, the necessary libraries were first imported, and the data file was introduced. Subsequently, the dataset was split into the target variables (Y) and features (X) used for prediction. Out of 118 data recordings, a further internal division of the dataset was made, using 80% of the data for training and 20% for testing. The testing and training dataset consists of 24 and 94 data inputs, respectively. The hyperparameter 'maximum depth' was selected based on minimising the difference between training and testing MAE. The SHAP (Shapley Additive explanations) Global Explanation value was used to quantify the contribution of each parameter to the prediction of CS.

For the SNN, a dataset of 118 data points is generated, with features represented by a random matrix (X) and a target variable (Y). Random noise is introduced to simulate the inherent variability in real-world data. The SNN is composed of an output layer with a linear activation function for regression, two middle layers with 64 and 32 neurons respectively, and an input layer with the number of neurons corresponding to the number of parameters. The Adam optimiser is used to construct the model, and the MAE is selected as the error function.

Results

The results of MLR, RF, and SNN models were analysed to derive MAE values, which serve as indicators of prediction accuracy. The comparison of MAE values allowed for an assessment of the performance of the models and subsets in predicting the compressive strength of Recycled Aggregate Concrete (RAC). By employing various methods in the MLR and RF models, it became possible to identify the influential parameters. This could have been done via sensitivity analysis for the SNN as well (Dias and Pooliyadda 2001, 371-379), but not carried out, since the SNN model turned out to be the least useful.

MLR

The performance evaluation of the MLR model across 4 subsets is presented in Table 1. Subsets 2 and 4 exhibit the highest accuracy in predicting CS, as evidenced by the lower MAE values of 9.44 and 9.48 MPa respectively. If the independent and the dependent variables do not actually relate to one another in MLR, the p-value tends to be higher (typically than 0.05). Among these subsets, EWCR emerges as the most effective parameter, displaying the lowest p-value by far across all subsets; and this is fully consistent with domain knowledge. The intercept also has p-values of a similar order of magnitude. However, the p-values of all other parameters exceed 0.05, suggesting that their influence is negligible. This is not borne out however in the MAE having differing values for the subsets, suggesting that the addition of parameters does in fact improve prediction accuracy. The fact that Subset 2 gives the lowest MAE value indicates that WA could be an important parameter governing CS.

RF

The SHAP (Shapley Additive Explanations) Global Explanation value assesses the contribution of each parameter to a specific prediction instance. SHAP values, whether positive or negative, indicate both the direction and strength of an influence of a parameter on the prediction. According to Figure 1 for Subset 4 (which includes all the parameters), the EWCR plays a significant role, since it is associated with a high SHAP value range. Higher values of EWCR (denoted in red) are associated with a decrease in CS (negative SHAP values), while lower values (denoted in blue) result in positive SHAP values and higher CS. This is well-known domain knowledge and serves to validate the RF outputs. Similarly, higher WA values result in lower CS and vice-versa. This is also to be expected, since greater WA will weaken the zone of interfacial transition between cement paste and the aggregate. The same can be said of LAA as well. However, there is some ambiguity where the ACR and RA% are concerned, as per the RF results.

Table 1 - Summary Results for the MLR

Data Subset	Feature/Parameter	p-value	MAE
Subset 1	EWCR	1×10^{-7}	13.9646
	RA%	0.1508	
	ACR	0.2609	
Subset 2	EWCR	1.17×10^{-7}	**9.4442**
	ACR	0.2689	
	RA %	0.5183	
	WA	0.8429	
Subset 3	EWCR	8.29×10^{-7}	10.7764
	RA %	0.2313	
	ACR	0.2464	
	LAA	0.7629	
Subset 4	EWCR	9.23×10^{-7}	**9.4774**
	ACR	0.2464	
	RA %	0.5769	
	LAA	0.7649	
	WA	0.8456	

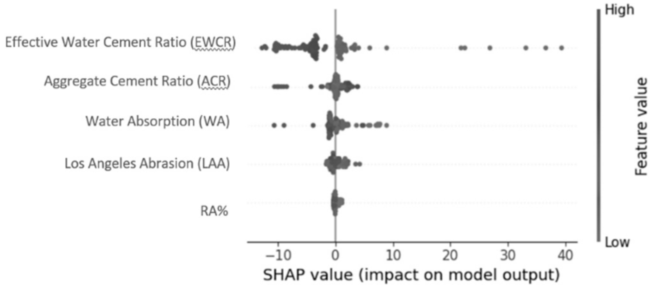

Figure 1 - SHAP Global Explanation on Subset 4

Table 2 - Summary Results for the RF

Subset	Most Effective Parameter on Compressive Strength	Approximate SHAP Value Range	Training MAE	Testing MAE
Subset 1	EWCR	25	6.1424	6.1659
	ACR	20		
	RA %	7		
Subset 2	EWCR	23	6.2120	6.2229
	WA	12		
	ACR	11		
	RA %	8		
Subset 3	EWCR	22	6.5470	5.9577
	ACR	13		
	LAA	12		
	RA %	5		
Subset 4	EWCR	14	**6.0797**	**5.9586**
	ACR	12		
	WA	8		
	LAA	6		
	RA %	2		

The range of SHAP values gives an indication of the relative importance of the parameters, and these are indicated, for all subsets, in Table 2. These also indicate the dominant influence of EWCR (as obtained from MLR), but also suggest that ACR is significant. In addition, WA appears to be more important than LAA. In some ways, both parameters measure, to an extent, the (negative) influence of the original cement mortar stuck on to the RA; and it may be sufficient to use just one of them, which should probably be WA – note also the low MAE value for Subset 2 (containing WA but not LAA) obtained in the MLR exercise. The low range for RA% in Subset 4 (Figure 1) is perhaps misleading. In Subset 2 (see Table 2), it has a comparable value to those for ACR and WA.

Where RF is concerned, the optimal model for any subset depends on striking a balance between training and testing performances. Subset 4 displays the lowest training MAE (6.08) and testing MAE (5.96) compared to the other three subsets, as seen in Table 2. These results confirm that

Subset 4, which involves all the parameters, gives the best prediction accuracy, as was the case for MLR modelling.

SNN

In Table 3 for SNN, once again Subset 4 stands out as the most accurate, displaying the lowest MAE of 9.61. Here too the next lowest MAE is given by Subset 2, echoing the MLR results, and confirming that WA is a more important parameter for predicting CS than is LAA.

Table 3 - Summary Results for the SNN

Subset	**MAE**
Subset 1	13.3753
Subset 2	11.1073
Subset 3	12.1636
Subset 4	**9.6089**

Discussion

This work has attempted to compare across (i) modelling techniques; (ii) subsets of parameters and (iii) the relative influence of parameters. The summary of findings where (i) and (ii) above are concerned is given in Table 4. This indicates that RF performs much better with respect to prediction accuracy compared to MLR and SNN. Yuan et al. (2022, 20) also found that RF is better than GB for achieving accuracy. Note that reports favourable to RF have been reported by other authors as well (e.g. Wickramasinghe et al. 2023).

Where subsets are concerned, all three modelling techniques indicate that Subset 4 gives the best prediction accuracy. Although not very clearly seen in the RF results, both the MLR and SNN results indicate that Subset 2 is the next best model. Although accuracy is generally less than for Subset 4, Subset 2 has the merit of using fewer parameters. This is an important contribution from this work. The approach has not been to use as many parameters as available, as for example by Yuan et al. (2022,3), who used 12 input parameters. Rather, the quest for parsimony herein helps to contribute to domain knowledge, and not merely to target prediction accuracy. In this case, the knowledge uncovered is that WA is a more important parameter than LAA for predicting CS.

In fact, when a large number of parameters are used, the modelling sometimes generates results that are at variance with established domain knowledge. For example, when comparing the influence of their 12 parameters, Yuan et al. (2022, 19) identify RA% (Recycled Aggregate Percentage) and EWCR as two of the most important, as obtained in this work too. However, parent concrete strength is also obtained as a key parameter, something that is not reflected in the literature on RAC. In addition, the RA% is found to have a 18.7% contribution to CS, and EWCR one of 14.7% (Yuan et al. 2022, 13). In contrast, our results establish the predominant influence of EWCR, with RA%, ACR, and WA having lower effects. This is much more in consonance with domain knowledge, and has been arrived at by increasing the number of parameters only incrementally in the subsets.

Table 4 - Overall Summary of Accuracies

	MAE		
	MLR	**RF**	**SNN**
Subset 1	13.9646	6.1659	13.3753
Subset 2	9.4442	6.2229	11.1073
Subset 3	10.7764	5.9577	12.1636
Subset 4	9.4774	5.9586	9.6089

Conclusions

1. Among the three modelling techniques explored, Random Forest (RF) consistently demonstrated superior performance to Multiple Linear Regression (MLR) and Simple Neural Networks (SNN) in predicting Compressive Strength (CS). This superiority of RF as a modelling tool has been established by other researchers as well.
2. In general, Subset 4, comprising all the parameters explored, was identified as the best subset for predicting CS. The parameters were Effective Water-Cement Ratio (EWCR), Aggregate-Cement Ratio (ACR), Recycled Aggregate Percentage (RA%), Water Absorption (WA), and Los Angeles Abrasion (LAA). Subset 2 emerged as the next best, comprising the parameters EWCR, ACR, RA%, and WA. The advantage in identifying this smaller subset is that it signals the more important parameters. Also, it was shown that a quest for parsimony in model parameters tends to generate parameter influences that are more consistent with established domain knowledge.

3. The influence of parameters on CS varies, with EWCR being the most important by far, followed perhaps by ACR and RA%. WA was, in general, more influential than LAA. These influences are consistent with domain knowledge. However, the finding that WA is more important than LAA for predicting CS can be seen as a genuinely new contribution to knowledge.

References

Abbas, A., G. Fathifazl, O.B. Isgor, A.G. Razaqpur, B. Fournier, and S. Foo. 2006. ‘Environmental Benefits of Green Concrete’. In 2006 IEEE EIC Climate Change Conference, 1–8. Ottawa, ON, Canada: IEEE. https://doi.org/10.1109/EICCCC.2006.277204.

Adesina, Adeyemi Damilare. 2018. ‘Concrete Sustainability Issues’. In, 38:24–26. London, United Kingdom: Coventry University. https://www.coventry.ac.uk/research/research-directories/research-events/2018/38th-cement-and-concrete-science-conference/.

Akbarnezhad, A., K.C.G. Ong, M.H. Zhang, C.T. Tam, and T.W.J. Foo. 2011. ‘Microwave-Assisted Beneficiation of Recycled Concrete Aggregates’. Construction and Building Materials 25 (8): 3469–79. https://doi.org/10.1016/j.conbuildmat.2011.03.038.

Batayneh, Malek, Iqbal Marie, and Ibrahim Asi. 2007. ‘Use of Selected Waste Materials in Concrete Mixes’. Waste Management 27 (12): 1870–76. https://doi.org/10.1016/j.wasman.2006.07.026.

Behnood, Ali, Jan Olek, and Michal A. Glinicki. 2015. ‘Predicting Modulus Elasticity of Recycled Aggregate Concrete Using M5′ Model Tree Algorithm’. Construction and Building Materials 94 (September): 137–47. https://doi.org/10.1016/j.conbuildmat.2015.06.055.

Ben Chaabene, Wassim, Majdi Flah, and Moncef L. Nehdi. 2020. ‘Machine Learning Prediction of Mechanical Properties of Concrete: Critical Review’. Construction and Building Materials 260 (November): 119889. https://doi.org/10.1016/j.conbuildmat.2020.119889.

Butler, Liam J., Jeffrey S. West, and Susan L. Tighe. 2014. ‘Towards the Classification of Recycled Concrete Aggregates: Influence of Fundamental Aggregate Properties on Recycled Concrete Performance’. Journal of Sustainable Cement-Based Materials 3 (2): 140–63. https://doi.org/10.1080/21650373.2014.909752.

Deng, Fangming, Yigang He, Shuangxi Zhou, Yun Yu, Haigen Cheng, and Xiang Wu. 2018. ‘Compressive Strength Prediction of Recycled Concrete Based on Deep Learning’. Construction and Building Materials 175 (June): 562–69.

https://doi.org/10.1016/j.conbuildmat.2018.04.169.

Dias, W.P.S., and S.P. Pooliyadda. 2001. 'Neural Networks for Predicting Properties of Concretes with Admixtures'. Construction and Building Materials 15 (7): 371–79. https://doi.org/10.1016/S0950-0618(01)00006-X.

Evangelista, L., and J. de Brito. 2014. 'Concrete with Fine Recycled Aggregates: A Review'. European Journal of Environmental and Civil Engineering 18 (2): 129–72. https://doi.org/10.1080/19648189.2013.851038.

Flood, Ian, and Nabil Kartam. 1994. 'Neural Networks in Civil Engineering. I: Principles and Understanding'. Journal of Computing in Civil Engineering 8 (2): 131–48. https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(131).

Gil, Jesús de-Prado-, Osama Zaid, Covadonga Palencia, and Rebeca Martínez-García. 2022. 'Prediction of Splitting Tensile Strength of Self-Compacting Recycled Aggregate Concrete Using Novel Deep Learning Methods'. Mathematics 10 (13): 2245. https://doi.org/10.3390/math10132245.

Gomez-Soberon, Jose M V. 2002. 'Porosity of Recycled Concrete with Substitution of Recycled Concrete Aggregate An Experimental Study'. Cement and Concrete Research.

Han, Taihao, Ashfia Siddique, Kamal Khayat, Jie Huang, and Aditya Kumar. 2020. 'An Ensemble Machine Learning Approach for Prediction and Optimization of Modulus of Elasticity of Recycled Aggregate Concrete'. Construction and Building Materials 244 (May): 118271. https://doi.org/10.1016/j.conbuildmat.2020.118271.

Katz, Amnon. 2003. 'Properties of Concrete Made with Recycled Aggregate from Partially Hydrated Old Concrete'. Cement and Concrete Research 33 (5): 703–11. https://doi.org/10.1016/S0008-8846(02)01033-5.

Kwan, Wai Hoe, Mahyuddin Ramli, Kenn Jhun Kam, and Mohd Zailan Sulieman. 2011. 'Influence of the Amount of Recycled Coarse Aggregate in Concrete Design and Durability Properties'. Construction and Building Materials, July, S0950061811003229. https://doi.org/10.1016/j.conbuildmat.2011.06.059.

Mai, Hai-Van Thi, May Huu Nguyen, Son Hoang Trinh, and Hai-Bang Ly. 2023. 'Toward Improved Prediction of Recycled Brick Aggregate Concrete Compressive Strength by Designing Ensemble Machine Learning Models'. Construction and Building Materials 369 (March): 130613. https://doi.org/10.1016/j.conbuildmat.2023.130613.

Mehta, P. Kumar. 2009. ‘Global Concrete Industry Sustainability’. Concrete International 31 (2): 45–48.

Nunez, Itzel, Afshin Marani, and Moncef L. Nehdi. 2020. ‘Mixture Optimization of Recycled Aggregate Concrete Using Hybrid Machine Learning Model’. Materials 13 (19): 4331. https://doi.org/10.3390/ma13194331.

Peng, Yiming, and Cise Unluer. 2023. ‘Modelling the Mechanical Properties of Recycled Aggregate Concrete Using Hybrid Machine Learning Algorithms’. Resources, Conservation and Recycling 190 (March): 106812. https://doi.org/10.1016/j.resconrec.2022.106812.

Rao, Akash, Kumar N. Jha, and Sudhir Misra. 2007. ‘Use of Aggregates from Recycled Construction and Demolition Waste in Concrete’. Resources, Conservation and Recycling 50 (1): 71–81. https://doi.org/10.1016/j.resconrec.2006.05.010.

Rosa, Lorenzo, Viola Becattini, Paolo Gabrielli, Alessandro Andreotti, and Marco Mazzotti. 2022. ‘Carbon Dioxide Mineralization in Recycled Concrete Aggregates Can Contribute Immediately to Carbon-Neutrality’. Resources, Conservation and Recycling 184 (September): 106436. https://doi.org/10.1016/j.resconrec.2022.106436.

Sarshar, R., and G. A. Khoury. 1993. ‘Material and Environmental Factors Influencing the Compressive Strength of Unsealed Cement Paste and Concrete at High Temperatures’. Magazine of Concrete Research 45 (162): 51–61. https://doi.org/10.1680/macr.1993.45.162.51.

Tam, Vivian W. Y., Anthony Butera, Khoa N. Le, and Wengui Li. 2020. ‘Utilising CO2 Technologies for Recycled Aggregate Concrete: A Critical Review’. Construction and Building Materials 250 (July): 118903. https://doi.org/10.1016/j.conbuildmat.2020.118903.

Topçu, ilker Bekir, and Nedim Firat Günçan. 1995. ‘Using Waste Concrete as Aggregate’. Cement and Concrete Research 25 (7): 1385–90. https://doi.org/10.1016/0008-8846(95)00131-U.

US EPA, OLEM. 2017. ‘Construction and Demolition Debris: Material-Specific Data’. Collections and Lists. 12 September 2017. https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/construction-and-demolition-debris-material.

Wagih, Ashraf M., Hossam Z. El-Karmoty, Magda Ebid, and Samir H. Okba. 2013. ‘Recycled Construction and Demolition Concrete Waste as Aggregate for Structural Concrete’. HBRC Journal 9 (3): 193–200. https://doi.org/10.1016/j.hbrcj.2013.08.007.

Walberg, Dietmar. 2016. ‘Solid and Timber Construction in Residential Buildings / Massiv- Und Holzbau Bei Wohngebäuden: Solid and Timber

Construction in Residential Buildings / Massiv- Und Holzbau Bei Wohngebäuden'. Mauerwerk 20 (1): 16–31. https://doi.org/10.1002/dama.201600685.

Wang, Bo, Libo Yan, Qiuni Fu, and Bohumil Kasal. 2021. 'A Comprehensive Review on Recycled Aggregate and Recycled Aggregate Concrete'. Resources, Conservation and Recycling 171 (August): 105565. https://doi.org/10.1016/j.resconrec.2021.105565.

Wickramasinghe, Vajira, W. P. S. Dias, Dilan Robert, and Sujeeva Setunge. 2023. 'The Relative Influence of Environmental Factors Compared to Age on Building Element Degradation'. Journal of Performance of Constructed Facilities 37 (6): 04023053. https://doi.org/10.1061/JPCFEV.CFENG-4521.

Yuan, Xiongzhou, Yuze Tian, Waqas Ahmad, Ayaz Ahmad, Kseniia Iurevna Usanova, Abdeliazim Mustafa Mohamed, and Rana Khallaf. 2022. 'Machine Learning Prediction Models to Evaluate the Strength of Recycled Aggregate Concrete'. Materials 15 (8): 2823. https://doi.org/10.3390/ma15082823.

Author Biographies

W.K.V.J.B. Kulasooriya is a civil engineering graduate from the Sri Lanka Institute of Information Technology, who has published AI related research in Nature Scientific Reports.

W.P.S. Dias is an emeritus professor of Moratuwa University and a professor at the Sri Lanka Institute of Information Technology; he has been placed within the top 2% of cited researchers in the Elsevier-Stanford database for the past 4 years.

CHAPTER 2

MODELLING OF BUILDING ELEMENT DEGRADATION

R.S.S. RANASINGHE & W.P.S. DIAS

Abstract

The modelling of building element degradation is essential for planning maintenance work on building elements and plays a significant role in the management of building assets. These elements deteriorate with age, usage, environmental conditions, and biological damage. Modelling building element deterioration helps determine when maintenance or replacement work should be conducted. This study focuses on degradation modelling using three approaches: a statistical method, namely the Markov Model; and two machine learning models—the Random Forest Regression Model and the Gradient Boosting Regression Model. The time stability of these three models is also assessed. The data used in this study were sourced from a previous study, encompassing the condition states of 12 building elements spanning 10 to 60 years of age. These elements include structural elements (slabs, beams, columns), finishing elements (ceilings, wall paint, wall plaster, rendered cement floors, floor tiles), opening elements (timber windows, timber doors), and service elements (ceiling fans, fan regulators). They are categorised into five conditions ranging from condition one (good) to condition five (bad). The performance of the models is evaluated using Mean Absolute Error (MAE) and the Coefficient of Determination (R^2) based on the predictions obtained. Additionally, a paired t-test is employed to establish the time stability of the models. The study suggests that the Gradient Boosting Regression Model predicts with the best accuracy, compared to the other two models. However, the Markov Model exhibits better time stability compared to the machine learning models.

Introduction

The term "Building Element Degradation" relates to the gradual decay of the physical and mechanical attributes of the elements of a building over time, influenced by several factors, including environmental conditions, chemical reactions, moisture, and usage. This deterioration of building elements is a complex issue that significantly impacts the security and economic stability of a country (Edirisinghe, Setunge, and Zhang 2015,1-2; Gaspar and De Brito 2008,2). The evaluation of building degradation is important in the domains of maintenance and management, directly affecting the capacity for strategic allocation of resources for the continual maintenance of buildings (Wickramasinghe et al. 2022,1). Also, according to Capacci et al. (2022,1-2), it can contribute to risk assessment and mitigation.

Estimating the service life of building elements necessitates a comprehensive understanding of degradation mechanisms impacting their performance over time. Buildings, being intricate infrastructure systems, include a variety of components, including structural, architectural, services, fittings, and fixtures elements. This inherent complexity poses a challenging task for managing building infrastructure while ensuring the expected level of service (Silva 2022,4-5). Deterioration models are key tools in asset management, providing predictive insights into future asset conditions (Beshara and Hasan 2019,41). In response to evolving economic conditions, building owners increasingly opt for optimising the use of existing structures rather than incurring the rising costs of new construction. This underlines the pressing need to efficiently evaluate and manage existing buildings and their elements (Beshara and Hasan 2019,41-42).

Conventional Approaches to Building Degradation Modelling

Infrastructural assets, which encompass roads, bridges, and buildings, play an indispensable role in national economies. However, these critical assets inevitably undergo deterioration due to aging, environmental conditions, and external factors (Hassan et al. 2022,1-4). Managing the complexity of building infrastructure presents several challenges. These include limitations in current methods of condition rating, the inability of existing deterioration prediction methods to address the randomness of degradation, an incomplete understanding of factors influencing degradation, and the necessity for systematic maintenance prioritization and accurate cost forecasting (Silva 2022). There is an ongoing debate between accelerated aging laboratory tests and field data collection. Critics argue that while the former offers controlled conditions, it lacks the intricacies of real-world environmental contexts.

In contrast, field data collection methods, starting with visual defect identification and classification under actual service conditions, provide a more authentic assessment of degradation (Murali Krishna et al. 2020,1-2). Degradation prediction models significantly contribute to the asset management process by forecasting the future condition of building elements. These models aid in devising effective maintenance and repair strategies, optimizing resources, and extending the life cycle of assets, although they face complexity due to the intricate hierarchy and numerous components involved (Edirisinghe, Setunge, and Zhang 2015,2-3). Traditional methods for estimating the service life of building elements, which rely on accelerated laboratory tests, are challenged by their need for controlled environments; and generic deterioration ratings fail to appreciate the nuances associated with variations across elements. This has led to a growing emphasis on the significance of field data collection methods for a more realistic evaluation (Gaspar and De Brito 2008,2-3; Wickramasinghe et al. 2022,1-2).

Approaches and Models for Building Element Degradation

Where visual inspections are concerned, there is difficulty in monitoring degradation over a significant length of time. However, the use of snapshot data (i.e. surveying buildings and their elements of different ages at a single point in time) provides an efficient way to encompass a broad range of building ages. This innovative methodology not only streamlines data collection but also enables a more comprehensive understanding of degradation patterns over time and has been used by several researchers (Wickramasinghe et al., 2022,1-3; Aisyah et al., 2019,1). The data set can also be customised to specific building types, e.g. Local Authority buildings in Sri Lanka (Wickramasinghe et al., 2022,1-2) and student dormitories at a specific university in India (Aisyah et al., 2019,1-2).

Once the data, as a function of time, has been collected, it is used to construct a time-dependent model of degradation. Here too several approaches can be used; and it is possible to predict the remaining useful life of building components and plan maintenance tasks by utilising models like time series analysis, Markov chains, and artificial intelligence techniques (Kellouche, Ghrici, and Boukhatem 2021,793-796).

A time series model is a statistical method utilised to analyse such time-dependent data, and involves applying a mathematical model to the data to identify patterns and trends, allowing for predictions of future behavior (Gharehbaghi et al. 2020,1-2). The iterative nature of time series analysis involves the repetition of the same process, where output data becomes input for subsequent predictions. An illustration of this approach can be found in the study by Gode (2014,11-13), where time series analysis is employed to predict the remaining lifespan of concrete bridges.

The Gamma process is viewed as an appropriate method for predicting the deterioration of building elements and recognising the temporal variations in degradation. This Gamma deterioration process is defined as a stochastic process featuring independent zero or positive increments, adhering to a gamma distribution with a matching scale parameter (Edirisinghe, Setunge, and Zhang 2015,1-3). Other stochastic reliability-based methods, like the Markov chain, have been extensively studied for assets such as bridges and roads, and used for forecasting the deterioration of buildings (Wickramasinghe, Dias, and Setunge 2021,1-4; Wickramasinghe et al. 2022, 1-4; 2023,1-4).

More recently, Artificial Intelligence (AI) approaches have been adopted for modelling building element degradation. Aisyah et al. (2019,1-3) utilise predictive analysis through Artificial Neural Networks (ANN) with input variables aligned with ISO factors such as the age of building components, while the severity level of their deterioration serves as the output. This method demonstrates how each input variable influences the output. AI tools such as ANN are adept at grasping intricate relationships within the data and making predictions based on learned patterns. Murali Krishna et al. (2020,1-2) proposed an ANN to predict the lifespan of reinforced concrete beams. Authors have also assessed how well an artificial neural network (ANN) model performs by conducting a comparison with results obtained from a traditional ordinary least square (OLS) regression model (Hassan et al. 2022,1-3). Furthermore, the accuracy of the ANN model's predictions is assessed by examining R^2 values across training, cross-validation, and testing sets. Concurrently, the regression model is utilised as a reference point for comparison (Hassan et al. 2022,1-3; Kellouche, Ghrici, and Boukhatem 2021,793-794). AI techniques have now moved beyond ANN, also termed Simple Neural Networks (SNN), to encompass Deep Neural Networks (DNN), Random Forest (RF), and Gradient Boosting (GB) techniques. Wickramasinghe et al. (2023,1-7) found RF to be particularly promising when modelling the degradation of Local Authority building elements using several AI modelling techniques including SNN, DNN, and RF, in addition to Markov models and multiple linear regression (MLR). They were able to incorporate both age as well as environmental factors as inputs, in order to predict the building elements' degradation condition ratings.

Based on the literature review, our research aimed to use relatively cutting-edge AI techniques such as RF and GB to model the age dependence of degradation. At the same time, Markov modelling was used as a reference point for the comparison of model performance. In addition, there is very little in the literature on the time stationarity of the models – i.e. whether a model trained on a particular period of age data can be used to predict degradation over a different (typically later) period of age. Hence, evaluating the time stationarity of models was another objective of this work.

Methodology

Data Collection

The data for this study was obtained from the research carried out by Wickramasinghe et al. (2022). Degradation information was obtained for 12 building components, encompassing structural elements (slabs, beams, columns), finishing elements (ceilings, wall paint, wall plaster, rendered cement floors, floor tiles), opening elements (timber windows, timber doors), and services elements (ceiling fans, fan regulators), from structures owned by seven local councils in Sri Lanka, with ages of up to 60 years. The snapshot data was characterised by age on the one hand and the condition rating on the other. Each element has a specific deficiency-based condition rating. Such deficiency-based ratings have been found to yield greater accuracy in model predictions than generic ratings common to all elements (Wickramasinghe et al. 2022,1-5). The ratings are associated with rather specific deficiencies in some detail, thus making the allocation of ratings to inspected elements much less subjective than having a generic rating. Condition rating 1 represents an "as new" condition, whereas rating 5 typically represents a condition where the element could need replacement. At each age, there would be different numbers of elements that are observed, with these elements being distributed to conditions 1 to 5, depending on condition descriptions such as in Table 1 for wall plaster (as an example). The number of observations in each condition for each year