IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Classification of compressive strength grades for lightweight aggregate concrete with palm oil fuel ash (POFA) using k-Nearest Neighbour (k-NN)

To cite this article: Mohamad Hushnie Haron *et al* 2023 *IOP Conf. Ser.: Earth Environ. Sci.* **1238** 012015

Classification of compressive strength grades for lightweight aggregate concrete with palm oil fuel ash (POFA) using k-Nearest Neighbour (k-NN)

Mohamad Hushnie Haron[1], Nur Azzimah Binti Zamri[2], Khairunisa Muthusamy[3]

[1] School of Science and Technology, Wawasan Open University, Georgetown, Pulau Pinang, Malaysia
[2] School of Civil Engineering, Universiti Teknologi MARA Cawangan Pahang Kampus Jengka, Pahang, Malaysia
[3] Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, Persiaran Tun Khalil Yaakob, Kuantan, Malaysia

mhushnieh@wou.edu.my

Abstract. Annually, a massive number of agricultural by-products of the palm oil extraction process including palm oil fuel ash (POFA) were generated which contributes towards ammonia pollution and emission of nitrogen compounds. Fortunately, both by-products can be utilised as mixing additives in lightweight aggregate concrete manufacturing. The utilisation leads to a more sustainable green environment. Traditional methods for classifying concrete grades in civil engineering are difficult due to the non-linear relationship between the composition of concrete and its strength and require a significant amount of time, material resources, and labour. To address these shortcomings, a technique to classify the compressive strength grades for lightweight aggregate concrete containing POFA using a machine learning algorithm has been developed. In terms of method, concrete mixtures consisting of POFA, cement, sand, superplasticizer and water were prepared and tested to determine the compressive strength. The data from this process were first transformed using min-max normalization and then, analysed using exploratory and descriptive analysis to discover patterns between input variables and concrete grades. Next, the grades of concrete were classified using a machine learning algorithm named k-Nearest Neighbour (k-NN). Lastly, a confusion matrix was used to assess the performance of the k-NN classifier. The results showed that k-NN can classify the grades of concrete with accuracies between 71% and 95% using five nearest neighbours. The accuracies are inversely proportional to the number of nearest neighbours. To conclude, the study succeeds in classifying the compressive strength grades for lightweight aggregate concrete with POFA using k-Nearest Neighbour. It can cut down a significant amount of time, material resources, and labour in determining the grades of compressive strength for POFA-based lightweight concrete.

1. Introduction

Concrete grade classification is essential to determine the concrete function such as for load-bearing or non-load-bearing purpose. Concrete grades are specified by their crushing strength at 28 days as measured in standard conditions [1]. To achieve the desired strength and property, the concrete produced must be mixed using the correct mixture consisting of four main ingredients: cement, aggregate, sand and water. Correct concrete proportion is important to meet the durability design of concrete and reduce the cost of construction. With the right and accurate mix design, practitioners do not need to use high-grade concrete for use in non-load-bearing structures. In other words, a quality concrete mix design is crucial for successful construction.

However, the use of conventional concrete in construction causes the depletion of non-renewable materials such as cement and aggregate. Hence, to save cement and other non-renewable materials in concrete, sustainable materials such as palm oil waste which is abundantly produced can be utilized. In Malaysia, the abundance of palm oil wastes namely palm oil clinker (POC) and palm oil fuel ash (POFA) is continuously increasing due to high palm oil demand. These two types of waste are usually dumped at the nearby landfill which causes environmental problems such as soil pollution and erosion and water contamination [2]. Thus, it is seen that the use of these two types of palm oil waste materials in high-strength concrete production would reduce some of the total amounts of waste disposed at landfill. Furthermore, POC used as coarse aggregate in lightweight aggregate concrete would preserve natural resources such as granite and limestone for future generations.

Nevertheless, mix design for normal conventional concrete is not applicable for lightweight aggregate concrete due to the high demand for water as a consequence of the different properties of aggregate used. Lightweight aggregate concrete requires higher water demand due to lightweight aggregates exhibiting significantly higher water absorption than normal-weight aggregates due to their high porosity [3]. Hence, water-reducing admixture needs to be utilized in lightweight aggregate concrete mix design to improve workability [4,5,6]. While the conventional concrete mix may have a lower amount of cement, when it is designed to mix, the cement requirement may be higher for the same grade of lightweight aggregate concrete.

Estimating the compressive strength grade of concrete is often regarded as a difficult and exhausting task due to the non-linear relationship between the set of component materials that will be used in the mix and the final strength. Normal civil engineering methods for determining compressive strength entail creating many concrete samples for each material recipe and then curing and maintaining these samples under standard conditions for 28 days. Only once the appropriate curing period has passed, these samples are subjected to destructive compressive strength testing in the equipped laboratories [7]. Table 1 shows the compressive strength groups and grade designation at 28 days [8].

Table 1. Compressive strength groups and grade designation at 28 days [8]

Group	Grade designation	Characteristics compressive strength of 150 mm cube at 28 days, N/mm^2
Ordinary Concrete	*M10*	*10*
	M15	*15*
	M20	*20*
Standard Concrete	*M25*	*25*
	M30	*30*
	M35	*35*
	M40	*40*
	M45	*45*
	M50	*50*
	M55	*55*
High Strength Concrete	*M60*	*60*
	M65	*65*
	M70	*70*
	M75	75
	M80	80

As a result, creating non-destructive methods using statistics and a machine learning algorithm have been developed for classifying the grade of compressive strength. The methods would result in large savings in time, material resources, and labor, meaning tremendous convenience in civil engineering practices.

The purpose of the study is to classify the compressive strength grades for lightweight aggregate concrete containing POFA using a machine learning algorithm specifically k-Nearest Neighbor (k-NN). The normal civil engineering methods to estimate the compressive strength grade of concrete are difficult and exhausting due to the non-linear relationship between the component materials of concrete and its strength. The methods involve creating many concrete samples for each material recipe and then curing and maintaining these samples under standard conditions for 28 days. Only once the appropriate curing period has passed, these samples are subjected to destructive compressive strength testing in the equipped laboratories [7]. It shows that the conventional methods to determine the grades of concrete require a significant amount of time, material resources, and labor.

2. Method

There are two sub-sections under the Method: Data Acquisition and Data Analysis. Data acquisition contains the process of preparing lightweight concrete for compressive strength testing. In data analysis, data transformation, exploratory data analysis and classification of compressive strength grade were performed.

2.1. Data Acquisition

In this study, six main ingredients that are palm oil fuel ash (POFA), cement, sand, water and superplasticizer have been employed. To obtain these mixes, a series of trials and errors have been carried out [9]. To produce greener lightweight aggregate concrete, palm oil clinker was used as the sole coarse aggregate to replace non-renewable granite aggregate that is commonly used in concrete production. In addition, other green materials namely palm oil fuel ash (POFA) were employed as partial cement replacements. River sand was used as fine aggregates in this experimental work. The water-to-cement ratio used varied from 0.35 to 0.55. Lastly, the amount of superplasticizer used is in the range from 0 to 1.2 % by weight of cement to produce mixes with the best workability.

To find the best workable mix, palm oil clinker lightweight aggregate concrete mixes incorporating 0%, 10%, 20%, 30% and 40% POFA has been cast for this research work. All the samples were cast in the form of 100 mm x 100 mm x 100 mm cubes and then subjected to water curing for 28 days. The specimens were tested for compressive strength test according to BS EN 12390: Part 3 [1] at 28 days.

2.2. Data Analysis

In general, there are four main processes involved in the data analysis as displayed in Figure 1. The first step in data analysis is to acquire the data followed by data transformation using Min. Max Normalization. Next, a box plot was applied to the data for exploratory data analysis. Then, the classification of compressive strength grade using k-Nearest Neighbor (k-NN) was performed. Lastly, a confusion matrix was used to assess the performance of the k-NN classifier.

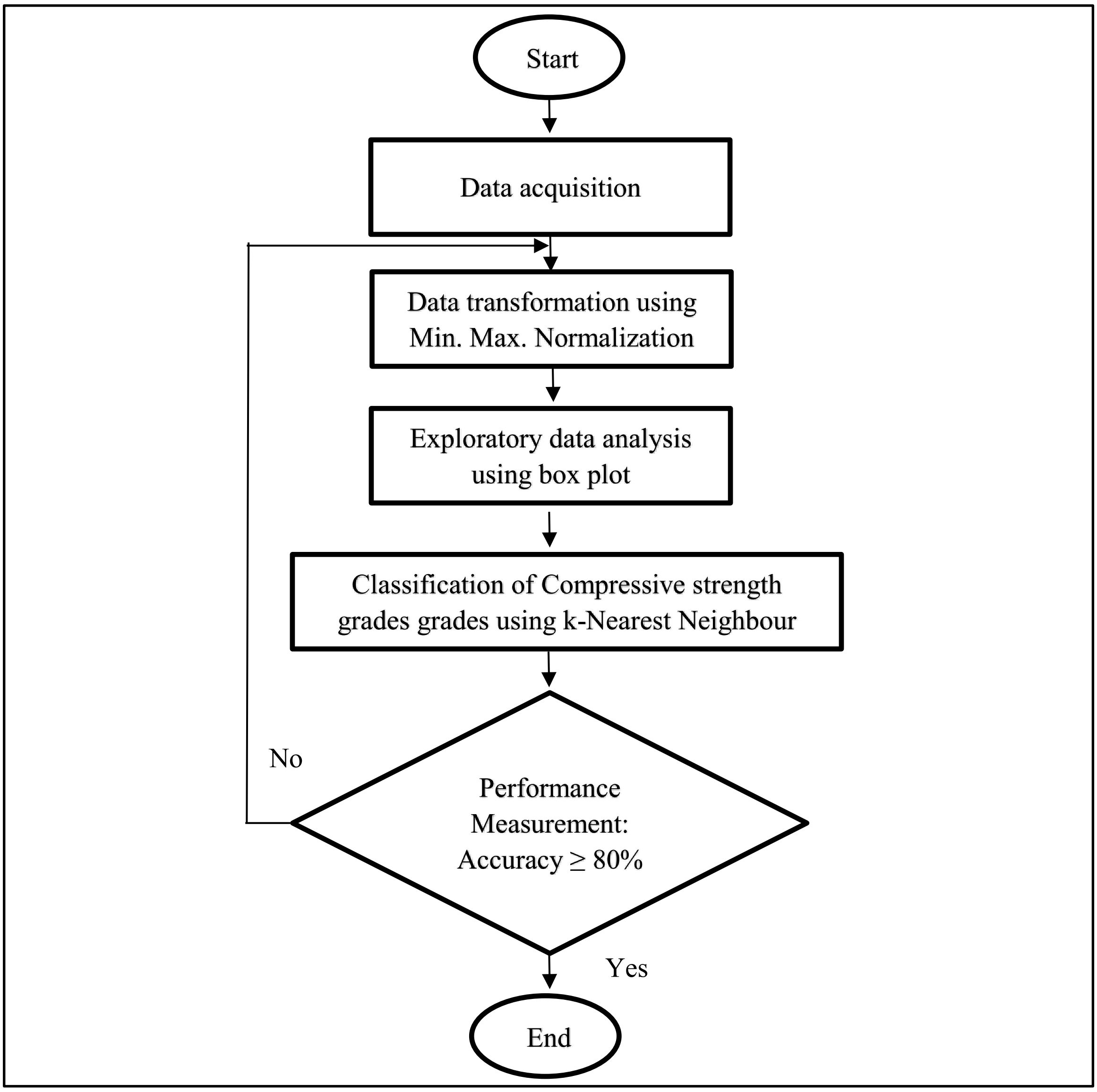


Figure 1. Flow chart of the method

2.2.1 Data Transformation

The data transformation used is the data rescaling technique using minimum-maximum normalization. The data have been rescaled between 0.10 and 0.99. This technique performs a linear transformation to the original data. It has the advantage to preserve features and relationships of original data [10]. The formula for rescaling between 0 and 1 is indicated in equation (1).

$$v' = \frac{(v - minimum\ value\ of\ v)}{(maximum\ value\ of\ v - minimum\ value\ of\ v)} \tag{1}$$

2.2.2 Exploratory Data Analysis

The second step in analysis involves exploratory data analysis using a box plot. The method was chosen due to its ability to provide relevant information on the maximum and minimum value, upper and lower quartile, as well as the median. The advantages of a box plot are the simplicity of its design, the critical information of the dataset is quickly expressed and it also shows the data distribution [11].

2.2.3 Classification of compressive strength grade

To classify between ordinary, standard and high-grade, the k-Nearest Neighbor (k-NN) technique has been applied. The grades are based on Table 1 where the compressive strength for the ordinary grade is 20 N/mm^2 and below. For standard grade, the compressive strength is between 55 N/mm^2 and 20 N/mm^2. For high grade, the compressive strength is above 55 N/mm^2. The inputs used for k-NN are the density of cement, POFA and water, the percentage of superplasticizer and slump, and the dry density of concrete. This technique is one of the machine learning algorithms and is used widely in pattern recognition and data mining. The broad usage of k-NN is due to its simplicity and easy implementation. In classification using k-NN, the number of nearest neighbors (k) and distance metric play an important role. It has a bigger impact on the performance and behavior of the model. A small k permits a simple implementation and efficient queries. Meanwhile, large k contributes to a smoother model and low noise during the classification process [12]. For this research, k=5 has been chosen. There are many types of distance metrics such as Euclidean, Manhattan, Chebyshev, Mahalanobis, Minkowski and Hamming. Euclidean distance has been chosen as the distance metric for this research classification. It is denoted as the distance between points p1=(x1,y1) and p2(x2,y2) represented in equation (2) [13].

$$\sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2} \quad (2)$$

To assess the performance of the k-NN classifier which is the accuracy, a confusion matrix was employed. Each column in a confusion matrix represents occurrences in a predicted class, while each row represents instances in an actual class. The confusion matrix depicts the various ways in which the classification model becomes "confused" when making predictions. It can provide insight not just into the errors made by a classifier, but also into the sort of errors that result. A multi-class confusion matrix with NxN dimensions for a situation with N different classes includes all conceivable combinations of predicted and actual classes as shown in Figure 2 [14].

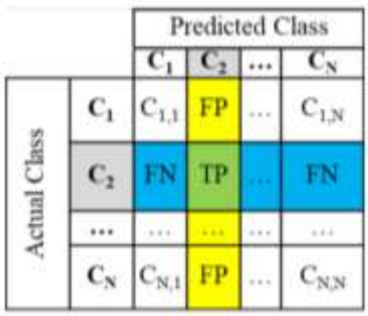


Figure 2. Multi-class confusion matrix [14]

3. Results and Discussion

This section consists of three sub-sections that present the discussion of the results from exploratory analysis using boxplot, correlation using Pearson's correlation and classification using k-Nearest Neighbor (k-NN) with performance measurement using a confusion matrix.

3.1. Exploratory Analysis of Compressive Strength Grades

Figure 3 shows the distribution of variables used to classify grades of compressive strength of lightweight concrete based on a box plot. For cement, POFA, superplasticizer and water, the distribution is symmetric due to the medians that are equidistant from the maximum and minimum values. The distribution is positively skewed for slump because the median to the maximum is greater than the distance from the median to the minimum. Dry density is negatively skewed because the median to minimum is greater than the distance from the median to the maximum. The distribution of the majority of the variables is symmetrical, hence the data do not require any data transformation techniques.

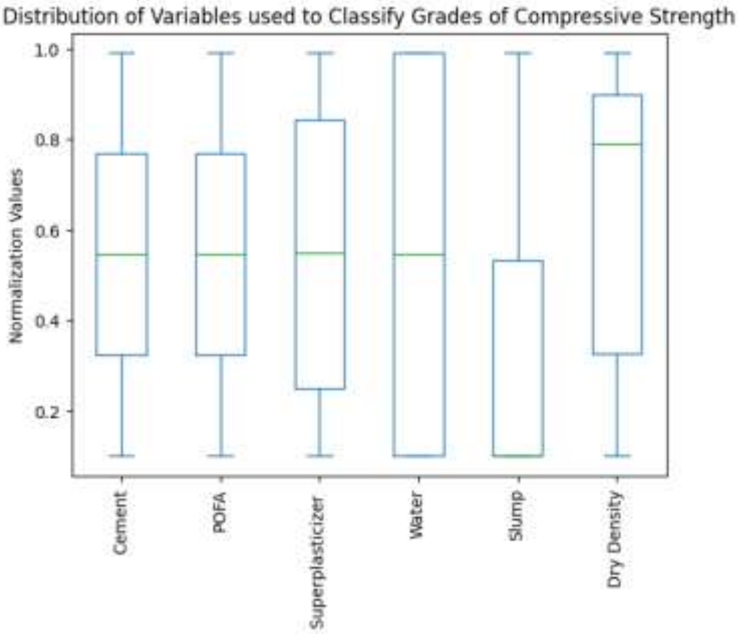
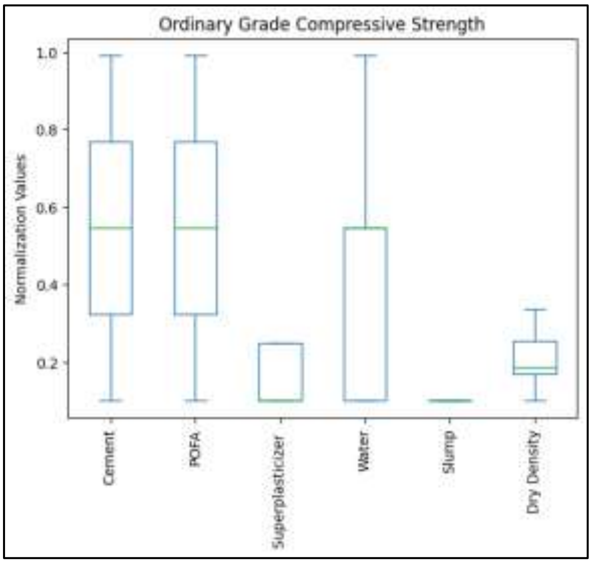
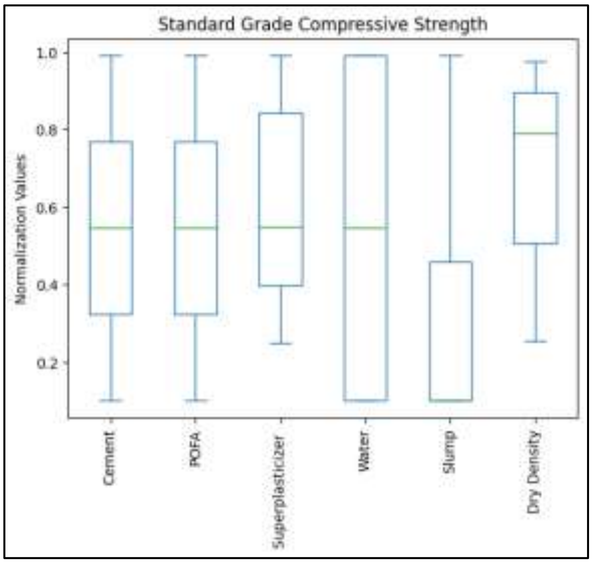


Figure 3. Distribution of variables used to classify the grades of compressive strength

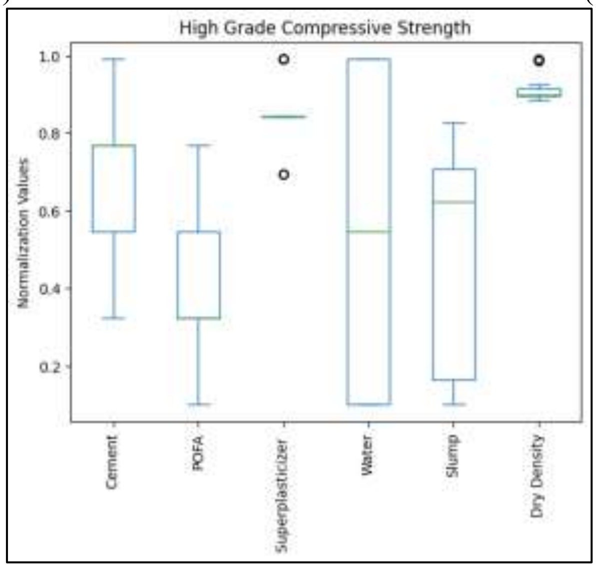
Figure 4 shows the comparison of ordinary, standard and high-grade compressive strength for variables such as cement, POFA, superplasticizer, water, slump and dry density. For cement, there is no difference in terms of median, maximum and minimum values, and interquartile range (IQR) between ordinary and standard-grade compressive strength. But the differences exist for the high-grade compressive strength where the minimum value and median are higher, and the IQR is smaller. For POFA, the interpretation of the box plot is the same as for cement except that in high-grade compressive strength, the maximum value and median are lower, and the IQR is the same. For superplasticizer, the differences are clear. The median and minimum value are low for ordinary-grade compressive strength, high for high-grade compressive strength and in the middle for standard-grade compressive strength. For water, there is no difference between standard and high-grade compressive strength in terms of median, maximum and minimum values, and IQR. But there is a difference in the IQR for ordinary grade compressive strength. It is smaller. For slump, the IQR increases from ordinary to high-grade compressive strength. The median is the same for both ordinary and standard-grade compressive strength. But it is higher for high-grade compressive strength. For dry density, the median and minimum value increase from ordinary to high-grade compressive strength. From the analysis, all the variables have at least a difference either in the median, minimum and maximum values, or IQR. Hence, these variables can be used for the classification of compressive strength grades using k-Nearest Neighbor (k-NN).



(a)



(b)



(c)

Figure 4. Variable comparison between ordinary, standard and high-grade compressive strength

3.2. Classification of Compressive Strength Grades

Table 1 and 2 show the confusion matrices for the classification of compressive strength grades using k-Nearest Neighbor (k-NN). Table 1 represents the training data. In total, 274 samples were used in classifying the three grades of compressive strength: ordinary, standard, and high. For nearest neighbor 1 (k=1), 61, 159, and 54 samples have been correctly classified into ordinary, standard and high-grade, respectively without any misclassification. For k=2, 61, 156 and 53 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 3 samples from standard grade into ordinary grade and 1 sample from high-grade into standard grade. For k=3, 61, 154 and 53 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 2 and 3 samples from standard grade into ordinary and high grade, respectively and 1 sample from high-grade into standard grade. For k=4, 61, 143 and 49 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 14 and 2 samples from standard grade into ordinary and high grade, respectively and 5 samples from high-grade into standard grade. For k=5, 60, 135 and 49 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 1 sample from ordinary into standard grade, 14 and 10 samples from standard grade into ordinary and high grade, respectively and 5 samples from high-grade into standard grade. In training, the ordinary data can be classified most accurately by k-NN followed by high and standard-grade data.

Table 2. Confusion matrix for training

		Predicted		
		Ordinary	*Standard*	*High*
Actual k=1	*Ordinary*	*61*		
	Standard		*159*	
	High			*54*
Actual k=2	*Ordinary*	*61*		
	Standard	*3*	*156*	
	High		*1*	*53*
Actual k=3	*Ordinary*	*61*		
	Standard	*2*	*154*	*3*
	High		*1*	*53*
Actual k=4	*Ordinary*	*61*		
	Standard	*14*	*143*	*2*
	High		*5*	*49*
Actual k=5	*Ordinary*	*60*	*1*	
	Standard	*14*	*135*	*10*
	High		*5*	*49*

Table 2 represents the testing data. In total, 69 samples were used in classifying the three grades of compressive strength: ordinary, standard, and high. For nearest neighbor 1 (k=1), 11, 41, and 14 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 1 sample from standard grade into ordinary grade and 2 samples from high-grade into standard grade. For k=2, 11, 40 and 14 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 2 samples from standard grade into ordinary grade and 2 samples from high-grade into standard grade. For k=3, 11, 39 and 15 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 1 and 2 samples from standard grade into ordinary and high grade, respectively and 1 sample from high-grade into standard grade. For k=4, 11, 33 and 10 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 7 and 2 samples from standard grade into ordinary and high grade, respectively and 6 samples from high-grade into standard grade. For k=5, 11, 29 and 10 samples have been correctly classified into ordinary, standard and high-grade, respectively but with the misclassification of 7 and 6 samples from standard grade into ordinary and high-grade, respectively and 6 samples from high-grade into standard grade. In testing, the ordinary data can be classified most accurately by k-NN followed by high and standard-grade data.

Table 3. Confusion matrix for testing

		Predicted		
		Ordinary	*Standard*	*High*
Actual k=1	*Ordinary*	*11*		
	Standard	*1*	*41*	
	High		*2*	*14*
Actual k=2	*Ordinary*	*11*		
	Standard	*2*	*40*	
	High		*2*	*14*
Actual k=3	*Ordinary*	*11*		
	Standard	*1*	*39*	*2*
	High		*1*	*15*
Actual k=4	*Ordinary*	*11*		
	Standard	*7*	*33*	*2*
	High		*6*	*10*
Actual k=5	*Ordinary*	*11*		
	Standard	*7*	*29*	*6*
	High		*6*	*10*

Table 3 shows the accuracies for the classification of compressive strength grades using k-Nearest Neighbor (k-NN). Table 3 contains the accuracies for training and testing data of each nearest neighbor (k). For k=1, the training and testing accuracies are 100% and 95.7%, respectively. For k=2, the training and testing accuracies decrease to 98.5% and 94.2%, respectively. For k=3, the training accuracy decreases to 97.8 but the testing accuracy remains at 94.2%. For k=4, the training and testing accuracies are 92.3% and 78.3%, respectively. For k=5, the training and testing accuracies decrease to 89.1% and 71.5%, respectively. The accuracies of training and testing decrease by the number of nearest neighbors. The accuracies are inversely proportional to the number of nearest neighbors. The accuracy of training is higher than the accuracy of testing for all numbers of nearest neighbors.

Table 4. Accuracies of classification of compressive strength grades for training and testing data.

Number of Nearest Neighbors	**Accuracy (%)**	
	Training	*Testing*
k=1	*100.0*	*95.7*
k=2	*98.5*	*94.2*
k=3	*97.8*	*94.2*
k=4	*92.3*	*78.3*
k=5	*89.1*	*71.5*

4. Conclusion

In conclusion, k-Nearest Neighbor classifier (k-NN) can be applied to classify the compressive strength grades of ordinary, standard and high. Variables such as cement, POFA, superplasticizer, water, slump and dry density can be used as a parameter in classifying the compressive strength. Samples from ordinary grade can be classified most accurately by k-NN followed by high and standard grade. The accuracies of compressive strength grades are inversely proportional to the number of nearest neighbors. The k-NN can produce accuracies higher than 90% up to the third nearest neighbor. The study can cut down a significant amount of time, material resources, and labor in determining the grades of compressive strength for POFA-based lightweight concrete.

Acknowledgement
The author gratefully acknowledges the support from the co-authors who have provided valuable insights, theories and data in the field of academic research specifically in civil engineering. The contributions have been essential for furthering the understanding and knowledge involved in this research. The author would also like to express sincere gratitude to the people involved in finances, labor works, equipment and other related supports.

References
[1] Standard B. 2009 Testing hardened concrete *Compressive strength of test specimens BS EN* 12390-3
[2] Obidzinski K, Andriani R, Komarudin H, Andrianto A. 2012 Environmental and social impacts of oil palm plantations and their implications for biofuel production in Indonesia *Ecol. Soc.* **17(1)**
[3] Domagała L. 2015 The effect of lightweight aggregate water absorption on the reduction of water-cement ratio in fresh concrete *Procedia Eng.* **108** 206-13
[4] Muthusamy K, Mirza J, Zamri NA, Hussin MW, Majeed AP, Kusbiantoro A, Budiea AM. 2019 Properties of high strength palm oil clinker lightweight concrete containing palm oil fuel ash in tropical climate *Constr. Build. Mater.* **199** 163-77
[5] Shafigh P, Yahaghi J. 2016 The importance of superplastizer dosage in the mix design of lightweight aggregate concrete reinforced with plypropylene fiber *MATEC Web Conf.* **66** 00020
[6] Hama SM. 2017 Improving mechanical properties of lightweight porcelanite aggregate concrete using different waste material *Int. J. Sustainable Built Environ.* **6(1)** 81-90
[7] Akpinar P, Khashman A. 2017 Intelligent classification system for concrete compressive strength *Procedia Comput. Sci.* **120** 712-8
[8] Purvansh B. S. 2018 Concrete mix design as per is method of mix design *Int. J. Sci. Res.* **7(3)** 1161-64
[9] Shafigh P, Jumaat MZ, Mahmud H. 2010 Mix design and mechanical properties of oil palm shell lightweight aggregate concrete: a review *Int. J. Phys. Sci.* **5(14)** 2127-34
[10] Song Y, Kong X, Zhang C. 2022 A large-scale k-nearest neighbor classification algorithm based on neighbor relationship preservation *Wireless Commun. Mobile Comput.* **2022** 1-1
[11] Hu K. 2020 Become competent within one day in generating boxplots and violin plots for a novice without prior R experience *Methods Protoc.* **3(4)** 64
[12] Ertuğrul ÖF, Tağluk ME. 2017 A novel version of k nearest neighbor: dependent nearest neighbor *Appl. Soft Comput.* **55** 480-90
[13] Amiri IS, Akanbi OA, Fazeldehkordi E. 2014 A machine-learning approach to phishing detection and defense *Syngress*
[14] Markoulidakis I, Kopsiaftis G, Rallis I, Georgoulas I. 2021 Multi-class confusion matrix reduction method and its application on net promoter score classification problem *14th Pervasive Technol. related to Assistive Environ. Conf.* 412-9