GEOMECHANICS AND GEOENGINEERING
2024, VOL. 19, NO. 4, 421–443
https://doi.org/10.1080/17486025.2023.2288655

Characterization of the shape of aggregates using image analysis and machine learning classification tools

Daniel Niruban Subramaniam[a], Mohan Sajeevan[a], Jeyananthan Pratheeba[b], Sathushka Heshan Bandara Wijekoon[a] and Navaratnarajah Sathiparan[a]

[a]Department of Civil Engineering, University of Jaffna, Jaffna, Sri Lanka; [b]Department of Computer Engineering, University of Jaffna, Jaffna, Sri Lanka

ABSTRACT

Engineering applications including pervious concrete require effective packing of aggregates to optimize strength. Size and shape distribution of aggregates significantly affect the performance. Computational methods numerically represent shape of aggregates, from image analysis, the effectiveness of has not been compared and verified. This study aims to analyse the representability of shape aspects of aggregates by different computational methods. Crushed aggregates were grouped into 5 clusters, and each group was milled in a Los Angeles machine for different degrees (0–2000) to induce morphological changes on the aggregates. Aggregates ranging from 5 to 30mm in diameter were obtained (7191 in total). imageJTM, was used to compute dimensions and shape factors of aggregates from 14 computational methods. Statistical tests, Pearson's Correlations and Principal Components Analysis and machine learning classification tools, Decision-tree, Random-Forest, Naïve-Bayes, Support-Vector-Machines, K-Nearest-Neighbours and Perceptron were employed to assess. In conclusion, no shape factor could be singularly used to numerically represent the morphological changes on aggregate particles but a combination of shape factors is required. Data matrix had three primary dimensions. Combination of Circularity, Kumbrein-Solidity and Barksdale-Shape-Factor yield best representation of aggregate shape. Regression Tree classification method had the highest accuracy (0.9) in classifying milled and unmilled aggregates.

ARTICLE HISTORY
Received 29 May 2023
Accepted 16 November 2023

KEYWORDS
Aggregates; shape factor; morphology; shape distribution; aggregate surface texture

1. Introduction

Aggregate properties define the applicability of aggregate in civil engineering applications in transportation, geotechnical and construction industries. The aggregates would often be arranged and packed in applications, to give a rigid skeleton matrix that in turn contribute to strength in the applications, such as subbase in highways, ballast in railway tracks and in concrete in the construction industry (Evangelista and de Brito 2007, Bhutta *et al.* 2013, Ćosić *et al.* 2015). An aggregate is defined by properties including specific gravity, particle size distribution (conformity and uniformity coefficients), aggregate impact value, aggregate crushing value, Los Angeles Abrasion Value, elongation index and flakiness index (Akçaoğlu, Tokyay *et al.* 2004, Al-Rousan *et al.* 2005, Arasan *et al.* 2011).

Packing of aggregate particles is affected by size and the shape of particles while they are usually assumed as of a single representative size in applications (mono-size particles) and that they are in the shape of a sphere (spherical particles) (Dinger and Funk 2013, Chan and Kwan 2014). Several studies in contemporary literature have attempted to mathematically optimise packing model using dual-size particles, yet, remain highly limited in applications (de Larrard and Sedran 1994, Gan *et al.* 2010, Das *et al.* 2011, Chang and Deng 2020). Naturally, aggregates come in different sizes and are represented by particle size distribution curve, and are numerically represented by the uniformity coefficient, conformity coefficient and at times together with effective diameter which corresponds to D_{10} (where 90% of the particles are bigger than the effective diameter). Although these numerical quotients are used mostly in soil classification as classifying parameters, the efficacy of these numerical values representing the size distribution of particle size distribution in mathematical models remains debatable. Developing a packing optimisation model for a lump of particles with a distribution of sizes may need emphatic numerical representation, perhaps as a single or number of coefficients (Mueller 2010, Ng *et al.* 2016, Pouranian and Haddock 2018, Shen *et al.* 2019).

CONTACT Daniel Niruban Subramaniam Daniel.subramaniam@gmail.com

Natural aggregates, which come in a distribution of sizes, are rarely spherical in shape as assumed in the engineering models. Studies have attempted to numerically represent a shape morphology or aggregates and developed various mathematical computational models to represent the shape of an aggregate (Brzezicki and Kasperkiewicz 1999, Al-Rousan et al. 2005, 2007, Arasan *et al.* 2011, Bangaru and Das 2012, Zhang *et al.* 2012, Ding *et al.* 2017, Xie *et al.* 2017, Ge *et al.* 2018, Xu and Prozzi 2019, Bhattacharya *et al.* 2020, Ghosh *et al.* 2020, Zhao *et al.* 2020). The shape of an aggregate significantly affects packing of particles and hence the performance including the workability, strength and durability of the engineering product. In geotechnical and transportation engineering disciplines, shape of an aggregate is represented by aspect ratio, angularity, flakiness index and elongation index that have standard methods of quantification. Aspect ratio is a measure of the length to width ratio of an aggregate particle where aggregate with a higher aspect ratio tends to have lower packing density. Aggregates with higher aspect ratio tend to have higher packing density, enhanced workability, strength and durability (Zhu *et al.* 2018, Zhang *et al.* 2019). Angularity on the other hand is a measure of the sharpness of the corners and edges of an aggregate particle, that correlates to the strength and frictional properties of materials. Aggregates with higher angularity tend to have a higher interlocking strength and a better resistance to wear and abrasion while lower angularity may result in smoother surface and hence reduced strength of materials (Maerz 2004, Man and van Mier 2011). Flakiness index is a measure of the flatness of the surfaces of an aggregate particle where higher flakiness index tends to have lower packing and workability and thereby reduced strength (Al-Rousan et al. 2005, 2007). Elongation index is a measure of the elongation of an aggregate particle and that higher elongation index indicates lower packing, workability and strength of material (Xie *et al.* 2017). In addition to these shape factors, other parameters that affect the shape of aggregates include texture, roundness, and sphericity of the particles (Wang *et al.* 2005, Su and Yan 2018). Texture refers to the roughness of the particles, while roundness and sphericity refer to the curvature of the particles (Mahmoud *et al.* 2010, Maroof *et al.* 2020). These additional shape factors are hard to be quantified physically and are not often considered in civil engineering applications.

Image analysis is a powerful tool for analysing the shape of aggregates in construction materials and is often used to determine the size, shape, and texture of individual aggregate particles as well as to quantify the distribution of these properties across a sample or an application. Image analysis is using a variety of techniques, including microscopy, digital imaging and scanning electron microscope (Fletcher *et al.* 2003, Bangaru and Das 2012, Gu *et al.* 2014, Bhattacharya, Subedi et al. 2020). In the case of aggregate shape analysis, digital imaging is the most commonly used technique, obtained using high resolution digital cameras and then subsequently analysed using specialised software. The software identifies individual particles and computes their size, shape and texture parameters. While image analysis is advantageous to analyse several individual aggregates in a sample population, subsequent analyses and representing the distribution of shape and size by numerical values are often seen as challenging (Liu *et al.* 2017, Krejsová *et al.* 2018).

Several studies have been conducted to define the shape of aggregates resulting in different computational methods (Mostofinejad and Reisi 2012), more in the context of powder technology (Lin *et al.* 2016) and ballast in geotechnical engineering (Kusumawardani and Wong 2020a, 2020b). While studies assumed spherical particles in modelling packing of particles, several studies assumed the particles of rectangular shape and introduced a flatness ratio and elongation ratio (Gong *et al.* 2016, Ding *et al.* 2017, Altuhafi, O'Sullivan, and Cavarretta 2013). Furthermore, another study attempted to define the shape of aggregates using Digital Image Processing techniques (DIP) with six shape parameters including Flakiness ratio, Elongation ratio, Convexity ratio, Fullness ratio, Shape factor and Sphericity (Kwan *et al.* 2014). Deng *et al.* (2021) introduced the Maximally Random Jammed (MRJ) for ellipsoids, oblate spheroid, and prolate spheroids. In addition to these, another study used sphericity and fullness ratio as parameters to define the shape of aggregates (Zheng *et al.* 2020). Studies have considered both 2-dimensional and 3-dimensional representation of aggregates in quantifying the shape and size of aggregates, while some studies have approximated 3-dimensional representation from 2-dimensional projection of the same (Masad *et al.* 2000, Tafesse *et al.* 2012, Shuguang and Qingbin 2015, Gong *et al.* 2016).

Performance of aggregate particles individually or in an application depends highly on the shape of the aggregates. Prediction of a performance parameter of a model would require a numerical approach with respect to material properties. For example, prediction of a compressive strength would require the mix design and material properties, presented in a numerical form. When the aggregate shape affects the compressive strength, a numerical representation of the shape of aggregates will have to be incorporated in the compressive strength prediction model. This will require numerical representation of the shape of an aggregate, which needs a computational method to quantify a particular morphology that affects the specific performance parameter. In special materials such as pervious concrete, the performance of which solely depends on the shape and size distribution of aggregates used, a more accurate representation of shape and size distribution of aggregates becomes critical. Specifically, the permeability and strength of pervious concrete, which are the prime performance indicators of the concrete, are significantly affected by the characteristics of the pore structure it encompasses, which in turn depends on the shape and size distribution of aggregates. Although several computational methods have been proposed by studies in the contemporary literature to define the shape of an aggregate, they are mostly confined to the context in which the model was either developed or employed. Literature has not established a universal representation of the shape of an aggregate, that could be used in a mathematical performance model, to incorporate the impact of shape of the aggregate on the performance of the material. This study aims to analyse widely used aggregate shape factors and computational models to analyse the efficacy of their representation of the same as a numerical coefficient.

2. Materials and methods

2.1. Aggregates

Crushed aggregates of gneiss origin were obtained from the North-central region of Sri Lanka, varying in size from 5 till 30 mm in diameter. Aggregates were then tested for the physical properties as explained below.

2.1.1. Aggregate preparation for shape analysis

Aggregates were observed to be flaky, elongated and angular in the natural form it was obtained. The measure of Los Angeles Abrasion Value (LAAV) is done using the LAAV instrument that operates similar to a ball mill that by abrasion deforms the aggregates. Deformation of aggregates occurs by abrasion in which the surface of the aggregates would begin to lose roughness as the milling process undergoes with multiple revolutions. As the revolutions continued to increase (milling over a longer period), the aggregate then changes in shape losing the angular corner and becoming more rounded in shape. On further milling, the aggregate would then tend to attain a spherical shape form (with extended amount of milling). Based on this principle, the original aggregates obtained for this study was then milled for different lengths of time in Los Angeles machine, including 500, 1000, 1500 and 2000 revolutions.

Specific amount of aggregate samples was taken from each lump of aggregates obtained from natural collection (that were not milled) and each group from different number of revolutions of milling and the aggregates coloured in black oil paint. They were then allowed to air-dry in room temperature. After drying, those aggregate particles were spread on a white A4 sheet in rows and columns arrangement. An image of arranged aggregates was taken covering the complete sheet. Arranged particles were mixed together and rearranged on the sheet, and an image was taken again. This procedure was repeated 12 times and 12 different images were obtained in order to incorporate all projections of aggregates in the 2-dimensional representations. While taking the image, the background of the sheet was lit with white light, so as to avoid the inclusion of shadow of flash in the image adjoining the aggregates.

The images were obtained as colour images, even though the aggregates were painted in black and laid on a white sheet. The image was then processed using imageJ™, an open source image analysis tool. The image was scaled using the dimensions of the A4 dimensions, and the image was then converted to binary colour (black and white). The pixel with black in colour bore a pixel value of 0, while the pixel with a colour white had a value of 256. Each black zone in the image indicated an aggregate.

2.1.2. Particle size distribution analysis using image analysis

Particle size distribution (PSD) is the representation of the percentage of particles in each size bin (diameter), drawn to a cumulative percentage that passes through particular sieves. In this analysis using image of aggregates, the area of each particle is computed individually using image analysis tool imageJ™, and the area-equivalent circle was drawn, the diameter of which was computed and noted. The analysed aggregate is assumed to be of a particle of the computed diameter.

The particles were arranged in ascending order of corresponding diameter and bins of 1 mm were then devised and the aggregates were categorised. The total area within each bin (summation of the areas of each aggregate in each bin) was computed and recorded as the percentage of the total area (sum of areas of all aggregates). The percentages of each bin was then computed as the cumulative percentage of aggregate area up to the bin in consideration (from the lowest bin-diameter), and the graph was plotted against the bin-diameter. The plot is the PSD of aggregates in the sample analysed using image analysis tool. The mean

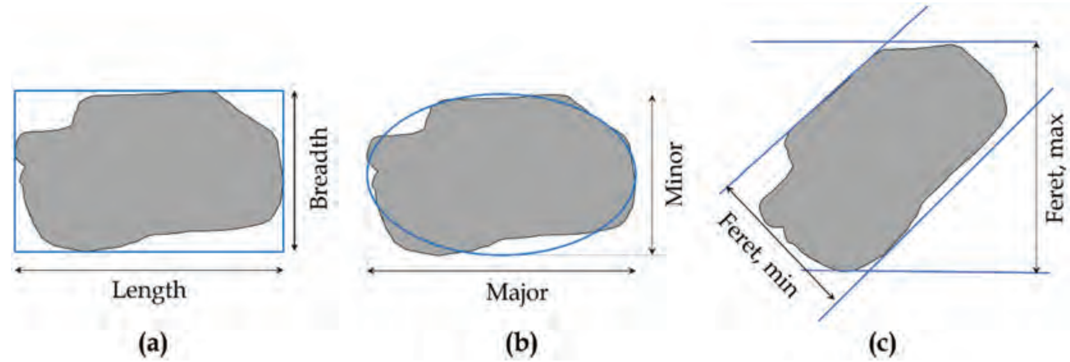

Figure 1. Length and breadth of aggregate particle.

of the percentages across all repetitive samples were obtained to represent the mean PSD of the sample population of aggregates belonging to a particular group.

2.1.3. Computation of shape factors

Computational models to quantify the shape factors require the quantification of the following dimensional parameters.

Bounding rectangle – The smallest rectangle enclosing the selection. This dimension called breadth and height in ImageJ is shown in Figure 1(a).

Fit Ellipse – The best fitting ellipse for this area. For this method, primary axis and secondary axis lengths were need, which are shown as Minor and Major in ImageJ as shown in Figure 1(b).

Feret's Dimensions – There are two Feret's dimensions. They measure the size of an area along a specified direction as shown in Figure 1(c).

Feret – The longest distance between any two points of the section boundary. (Maximum Caliper)

MinFeret- The smallest distance between any two points of the section boundary. (Minimum Caliper)

The following 14 different computational models employed to quantify the shape of aggregate in contemporary literature are considered for analyses in this study, the details of which are given in Table 1.

2.2. Statistical analyses

Several statistical tools were employed in this study to verify the efficacy of shape factors in characterising the aggregates based on shape that is effected by the milling process.

Pearson Correlation: is a measure of how well two continuous variables are linearly correlated, the correlation coefficient that results from the test range between −1 and + 1. A positive value indicates a directly proportional correlation while a negative value indicates inversely proportional relationships (Nebeokike *et al.* 2020, Ahmed and Arocho 2021, Egbueri and Igwe 2021). In addition, a numerical value of 1 indicates a perfect linear correlation, while a numerical value close to zero indicates the absence of a linear correlation. The test also returns a significant value, whereas a value of less than 0.05 indicates more than 95% confidence of the level of linear correlation indicated by the correlation coefficient (Egbueri 2022, Egbueri *et al.* 2023). The point to consider in this test is that the test only tests if the variables are linearly correlated and to assess non-linear correlations, the variable would need to be correspondingly transformed and then tested again (Egbueri and Agbasi 2022, Unigwe *et al.* 2023). This test is employed in this study to assess if any of the shape factors correlated to the degree of milling (if any of the shape factors directly linearly predict the shape variance), and also to understand if any of these shape factors are linearly correlated among them, so that they can be considered surrogate parameters.

Principal Component Analysis (PCA): is considered a pattern recognition tool and not a definitive statistical test, and is often employed to reduce the dimensions of the data matrix. This test reorganises the variance of data in to several principal components that are orthogonal to each other, and computes how much of a variance is explained by each of the principal component. A rotated matrix about varimax lists the variables enlisted in the matrix that are predominantly aligned to each principal components indicated by a value between −1 and + 1. The positive value indicates a positive impact on the principal component, while a negative value indicates a negative impact on the variance of the matrix. In addition, a value close to 1 indicates strong impact of the variable in the dimension indicated by the corresponding principal component, while a value close to zero indicates the direction of the vector being orthogonal (not impacting the variance in the relevant principal component) (Hameed *et al.* 2021, Kobaka 2021). This test is employed in this analysis to identify how

Table 1. Description of shape factors.

	Shape Factor	Computation	Reference
J.C	ImageJ circularity	$\frac{4\pi\, area}{(perimeter)^2}$	Rueden *et al.* (2017)
J.AR	ImageJ aspect ratio	$\frac{major\ axis}{minor\ axis}$	Rueden *et al.* (2017)
J.R	ImageJ roundness	$\frac{4\pi\, area}{\pi \times major\ axis^2}$	Rueden *et al.* (2017)
J.S	ImageJ solidity	$\frac{area}{convex\ area^2}$	Rueden *et al.* (2017)
KS	Krumbein sphericity	$\sqrt[3]{\frac{breadth}{length^2}}$	Kim *et al.* (2019)
BSF	Barksdale shape factor	$\frac{length}{breadth^2}$	Barksdale *et al.* (1991)
FR	Kwan and Mora fullness ratio	$\sqrt[2]{\frac{area}{convex\ area}}$	Mora and Kwan (2000)
SF.FR	Sneed and Folk flatness ratio	$\frac{Breadth}{Length}$	Lewis and McConchie (1994)
FAR	Feret's aspect ratio	$\frac{Feret}{Min\ Feret}$	Xu and Di Guida (2003)
SF	Shape factor	$\frac{perimeter}{Calculated\ perimeter}$	
PSC	Projected sphericity/circularity	$\frac{D.Calculated}{Feret}$	Cruz-Matías *et al.* (2019)
ICS	Inscribed circle sphericity	$\frac{Min\ Feret}{Feret}$	Zheng and Hryciw Roman (2016)
C	Circularity	$\frac{perimeter^2}{area}$	
PS	Projected sphericity	$\frac{area}{AC}$	

many dimensions would be required to explain the variance in the data matrix and that which of the shape factors dominate each principal component (Calabrese *et al.* 2012). The results of this study together with the Pearson's correlation test would essentially be used to reduce the dimensions of the data matrix, that means, removing the redundant or surrogate shape factors in the data matrix (Calabrese *et al.* 2012, Croce *et al.* 2020).

2.3. Classification using machine learning tools

In addition to conventional classification tools as described under clustering, several advanced classification methods using machine learning tools are employed.

2.3.1. Feature selection/feature extraction

Another important data preparing step before building the model. Real-world data come with a lot of unwanted features for the model. They might worsen the performance of the model as well as increase the execution time of the model. To build an effective and efficient machine learning model, those features should be handled or removed from the input. For this purpose feature selection or extraction is used in the preprocessing. Feature selection and feature extraction are two completely different processes, where we reduce dimensions in feature selection while adding in feature extraction (Lu *et al.* 2020, Li *et al.* 2021). As this data has no unwanted features for the problem, we did not do any feature selection or extraction. There are no highly correlated features among themselves in our data, except the temperatures at different points of the concrete block.

2.3.2. DT – decision tree

A decision tree classifier is a type of supervised machine learning algorithm that creates a tree-like model for classification purposes. Each node in the tree represents a feature or attribute, and each branch represents a criterion for dividing the data into subsets. At the end of each branch, there is a leaf node representing a class label. It recursively partitions the data into smaller subsets based on their features until it reaches a stopping criterion, such as a maximum depth or a minimum number of samples (Ben Chaabene *et al.* 2020).

2.3.3. RF – random forest

Random forest classifier is a machine learning algorithm used for classification tasks. It belongs to the ensemble learning family and is a combination of several decision trees that are individually trained on different subsets of the same training dataset. The random forest algorithm works by building multiple decision trees and aggregating their individual predictions into a final prediction. Each decision tree in the random forest is built using a subset of the features from the training dataset, selected randomly at each node. The final prediction of the random forest is made by averaging the predictions of all the decision trees in the forest (Ben Chaabene *et al.* 2020, Mangalathu *et al.* 2020, Kang *et al.* 2021).

2.3.4. NB – Naïve Bayes

A Naive Bayes classifier is a probabilistic algorithm based on Bayes' theorem that is used for classification tasks. This algorithm works under the assumption that all features are independent of each other, hence 'naive'. This assumption allows the algorithm to compute the probabilities of each feature independently and use them to make a prediction. The algorithm is called 'Bayes' because it uses Bayes' theorem to calculate the probabilities of the different classes that a given observation might belong to (Mangalathu *et al.* 2020).

2.3.5. L-SVM – linear support vector machine

A support vector machine (SVM) is another supervised machine learning algorithm that can be used for classification and regression analysis. It works by finding the best possible decision boundary, known as the hyperplane, which separates data into distinct categories. In the case of a binary classification problem, an SVM finds the hyperplane that maximises the margin, or the distance between the decision boundary and the closest data point from each class (Ben Chaabene *et al.* 2020).

2.3.6. SVM-RBF – support vector machine – radial basis function (kernel)

Support vector machine with radial basis function kernel (RBF SVM) is a popular type of SVM that uses a non-linear kernel function to classify data. The RBF kernel function maps the input data into a higher dimensional feature space and identifies a decision boundary that maximises the margin between the two classes (Ben Chaabene *et al.* 2020). The RBF kernel function is defined as follows:

$$K(X, X') = \exp\left(-\gamma ||\mathrm{x} - \mathrm{x}'||^2\right)$$

where x and x' are two input vectors, ϒ is a hyperparameter that controls the smoothness of the decision boundary, and ||x – x'||^2 is the Euclidean distance between x and x'.

2.3.7. SVM-P – support vector machine – polynomial

In the case of polynomial SVM, the algorithm uses a polynomial kernel function to transform the input data into a higher dimensional space. In this higher dimensional space, the SVM can more easily find a hyperplane that separates the different classes.

The polynomial kernel function takes the form:

$$K(X, X') = (\gamma * (X, X') + r)^d$$

where ϒ, r, and d are user-defined parameters that control the shape of the hyperplane. The gamma parameter determines the width of the kernel, the r parameter is an offset value and the d parameter controls the degree of the polynomial. A higher value for the d parameter will result in a higher degree polynomial, which can make the hyperplane more complex and better able to separate the classes. However, choosing too high of a value for the d parameter can result in overfitting of the model to the training data.

2.3.8. KNN – K nearest neighbours

K nearest neighbours (k-NN) is a classification algorithm that predicts the class of a new instance by identifying k number of nearest instances in the training set and assigning the most common class among those instances to the new instance. The algorithm works by calculating the distance between the new instance and all instances in the training set based on a set of features. The k instances with the shortest distance to the new instance are selected as the nearest neighbours. The value of k is an input parameter that determines the number of neighbours to be considered, and it should be chosen carefully based on the data set and the problem at hand (Mangalathu *et al.* 2020, Kang *et al.* 2021).

2.3.9. P – perceptron

A perceptron is a type of artificial neural network that is commonly used for classification tasks. It is a single-layer neural network that takes in input values, multiplies them by weights and passes the result through an activation function to produce an output. The perceptron can be used for binary classification, meaning that it can distinguish between two classes based on input features (Kang *et al.* 2021).

3. Results and discussion

The purpose of the study is to evaluate the credibility of the shape factors used, to represent the shape of an aggregate quantitatively and to analyse the sensitivity of the shape parameters used in the contemporary literature. The results and discussion section of this article is therefore separated into five sections:

(1) Statistical inference analyses of the patterns observed among dependent and independent variables. General trends and correlations among the shape factors, principal component analysis and statistical inferences were considered in this section of the article.
(2) Establishing the effectiveness of the procedure followed to effect changes in the shape of aggregates. Mostly supervised classification is used in this study to establish the distinguishability of the shape of aggregates based on ball-milling processing method.
(3) Analysing the features (shape factors) and selecting the features that numerically define the shape of the aggregates. Different feature selection methods including feature importance, mutual information, forward propagation method and backward propagation methods were used, and the outcomes are analysed in this section.
(4) Analysing conventional machine learning models and evaluate accuracies and uncertainties. The conventional methods used in this analysis include linear regression, LASSO, random forest, ridge regression, decision tree and neural network to predict area of the aggregate based on the independent shape factors.

3.1. Visual inspection of aggregates

Figure 2 shows the pictures of aggregates that have undergone different degrees of milling (up to 2000 revolutions), where 0 rev indicates original crushed aggregates. It is evident from the figure that the characteristics of aggregates significantly differ between the degrees of milling, particularly, the change in texture and shape.

When the milling process is more intense (prolonged), the aggregate becomes more rounded (spherical in 3 dimensional) while a lower degree of milling contrived mere smoothing of the surface, essentially changing the texture of the aggregate surface. Shape factors may intend to quantify any or all of the morphological changes in the shape characteristics of the aggregate ranging from texture to macro-shape due to the process of milling. The analysis in this study aims to identify which of the shape factor computational models retrieved from the literature quantify the morphological change visually observed in the aggregates shown in the figure.

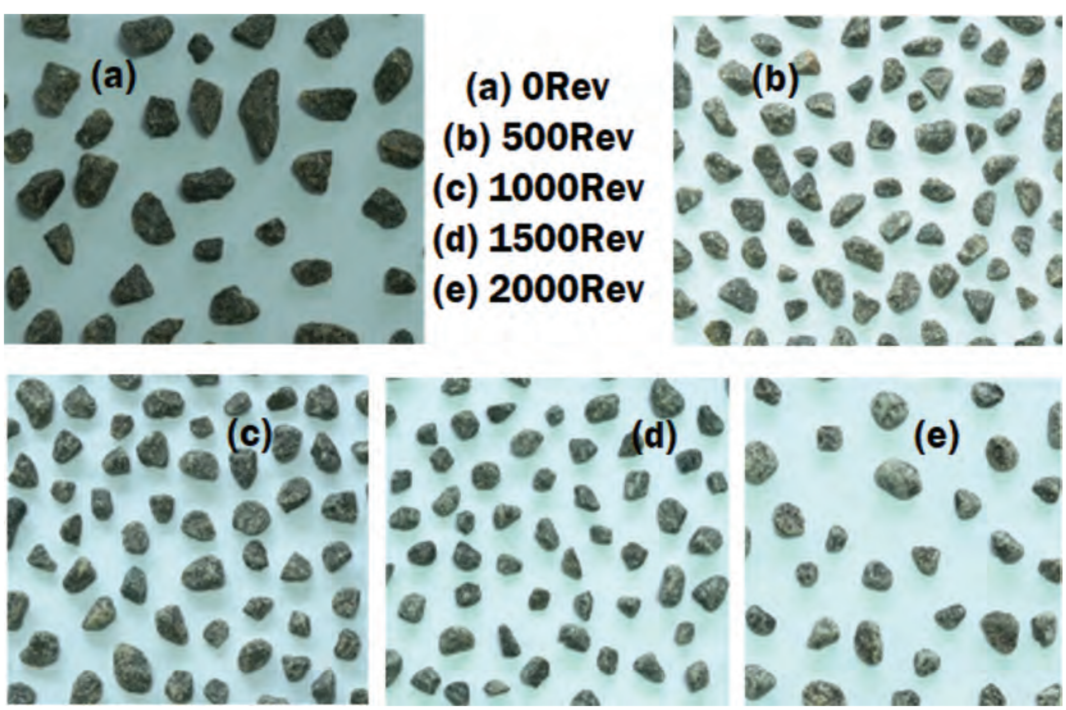

Figure 2. Pictures of aggregates that have undergone varied degrees of milling.

3.2. Descriptive statistics

Table 2 shows details of statistics of aggregate characteristic parameters including projected area, perimeter and various shape factors analysed in this study. The statistics presented here are for a total of 7191 aggregate particles, belonging to different degrees of milling. The area and perimeter have very high standard deviations (more than 50% of the mean) and skewness. Simultaneously, several shape factors indicate a higher variation among aggregate particles analysed. Kurtosis indicates the roundness or sharpness of the distribution at the peaks, which again is observed to vary significantly among shape factors.

The variations of these parameters (aggregate characteristics) within each group of aggregates that have been grouped based on degrees of milling would be essential for further analysis. Figure 3 shows the distribution of J.C KS, J.R and BSF within different groups of aggregates grouped based on degrees of milling. The four shape factors that showed significant variations have been reported here in the figure. It is observed that the shape factors J.C, KS, J.R and BSF have shown an increasing trend with the degree of milling, where a rapid increase was observed between 0 Rev and 500 Rev compared to the difference between a higher degrees of milling. Other than BSF, the other three shape factors tend to saturate with the degree of milling attaining saturation at least by 1500 revolutions.

The shape factors are predominantly based either on the dimensions specifically, length, breadth, perimeter and/or projected area of aggregate particles. The quantified shape factors therefore change based on variations among aggregates based on proportions of dimensions. More specifically, surface texture, angularity, flakiness and elongation, the primary shape aspects of aggregates, change the above-mentioned dimensions. A higher variation within a group of aggregates indicates the presence of highly varying above-mentioned aspects, while a narrower distribution indicates a more uniform presence of the same. The narrower distribution of a parameter indicates that the aggregate of a group has a more uniform distribution of the parameter, while a wider distribution indicates the contrary.

For a given area of an aggregate particle, the more deformed the shape is (due to surface irregularities, angularity, flaky nature or even texture), the perimeter would increase. The proportion of area/perimeter would therefore be lower (increased perimeter for the same area). The process of milling is expected to effectively smoothen the surface and break flaky or sharp corners in lower number of revolutions due to abrasion and impact. Milling further with higher number of revolutions is expected to erode the corners further and make them more rounded, effectively increasing the area/perimeter proportions. This process is further expected to make the particles more uniform in characteristics.

Considering shape factor J.R, the distribution is significantly wider for the zero-Rev group and that the distribution remained rather consistent throughout other groups that had experience milling, irrespective of the degree of milling. This indicates that the process of milling with a degree of fewer than 500 revolutions had instigated a morphological change that removed random irregularities that naturally occurred in the crushed aggregate particles, more particularly the surface texture and angular and flaky corners. Simultaneously, it could be concluded that J.R has significantly quantified such morphological shape aspects of aggregate particles, compared to other shape factors such as KS, that showed rather trivial differences between the distributions among the groups.

3.3. Pattern recognition using graphical representation

As the first step to understand the patterns among the independent variables, Pearson's correlation test was carried out between the 14 shape factors against number of revolutions. The results revealed that none of the shape factors had a significant correlation with revolutions, which indicated that none of the 14 shape factors considered could single handedly represent the morphological changes induced by the process of milling on the aggregate particles. As the next step, the 14 shape factors were analysed for correlation with either perimeter or projected area of aggregate particles. Should a significant linear correlation exist, it would indicate that the shape factor is predominantly affected by the size of the particle and that a significant impact of aggregate size could be expected in its quantification of the shape of the aggregate particles. Figure 4 shows the plots of shape factors (KS, BSF, FAR and SF) with aggregate projected area. Only the shape factors BSF and KS showed a non-linear correlation with the projected area of aggregates, where as all other shape factors showed no significant correlation. In addition, some shape factors, as shown in these plots, revealed an upper/lower boundary with respect to aggregate projected area, while other shape factors had random distribution without such boundaries.

Table 2. Descriptive statistics details of variables of aggregates characteristics analysed in this study.

		P	Area	J.C	J.AR	J.R	J.S	KS	BSF	FR	SF.FR	FAR	SF	PSC	ICS	C	PS
Statistics																	
N	Valid	7191	7191	7191	7191	7191	7191	7191	7191	7191	7191	7191	7191	7191	7191	7191	7191
	Missing	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean		43.12	134.43	0.77	1.35	0.76	0.96	0.41	0.12	0.98	0.85	1.37	1.14	0.83	0.86	16.38	0.69
Median		36.61	81.04	0.78	1.29	0.77	0.97	0.41	0.12	0.98	0.87	1.32	1.13	0.84	0.87	16.07	0.70
Mode		28.93	33.35	0.81	1.21	0.80	0.97	0.48	0.12	0.99	1.00	1.13	1.06	0.54	0.69	14.15	0.30
Std. Deviation		19.09	128.46	0.06	0.25	0.12	0.01	0.06	0.05	0.01	0.11	0.22	0.05	0.06	0.06	1.50	0.10
Variance		364.42	16502.68	0.00	0.06	0.01	0.00	0.00	0.00	0.00	0.01	0.05	0.00	0.00	0.00	2.24	0.01
Skewness		1.37	2.05	−1.54	1.82	−0.53	−1.39	−0.29	0.76	−1.44	−0.80	1.76	2.97	−.74	−0.86	3.90	−.51
Std. Error of Skewness		.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Kurtosis		1.37	4.15	4.86	5.54	0.09	4.49	−0.62	1.42	4.86	0.32	5.10	20.40	0.79	0.89	35.76	0.23
Std. Error of Kurtosis		.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Percentiles	25	29.22	53.23	0.75	1.18	0.69	0.96	0.37	0.08	0.98	0.79	1.22	1.11	0.80	0.83	15.47	0.63
	50	36.61	81.04	0.78	1.29	0.77	0.97	0.41	0.12	0.98	0.87	1.32	1.13	0.84	0.87	16.07	0.70
	75	51.46	159.65	0.81	1.46	0.85	0.97	0.45	0.15	0.99	0.94	1.47	1.16	0.87	0.90	16.87	0.76

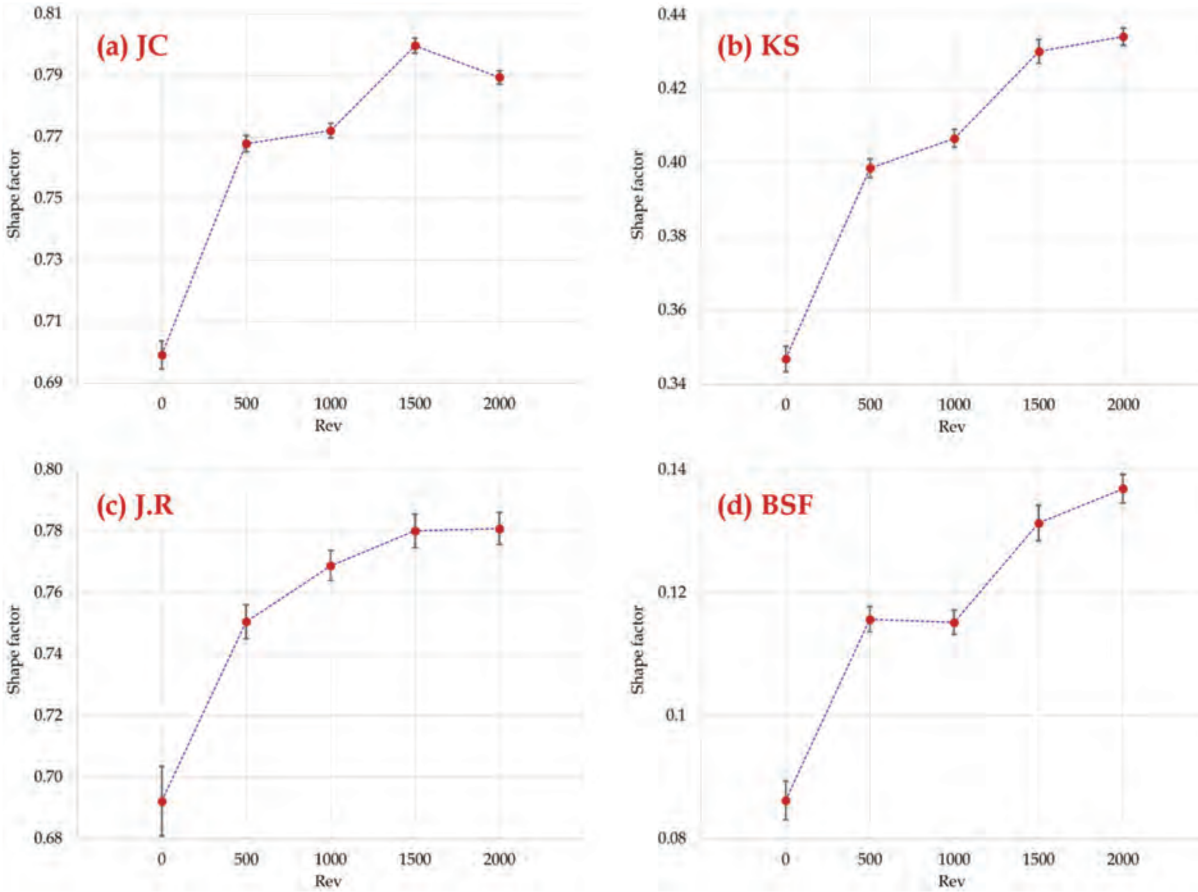

Figure 3. Mean and 95% confidence interval of the distribution of J.C, KS, J.R and BSF within different groups of aggregates.

As shown in Figure 4, KS and BSF have some impact of aggregate projected area for aggregates typically less than 100 mm^2, while the impact was not significant for larger aggregates. A closer inspection further reveals that the impact is more on the boundary of the profile, and not on individual particles, as even in the smaller aggregates, a wider range of variations could be observed in the computed shape factor. In addition, the boundary as observed in the shape factors in the figure means that the method of computation is limited specifically due to the dimensional aspects; for example, the shape of a circle would have the minimum perimeter for shapes of the same area.

It is also relevant to note that several of the shape factors indicated very similar trend with other shape factors, indicating a possibility that they may be quantifying the same morphological aspects of the aggregate particle. Figure 5 shows the Pearson's correlations coefficient between 14 shape factors analysed. It is observed that several shape factors revealed very high linear correlations with others while some shape factors had no correlation to any other factors. FR-J.S, C-SF and PS-PSC showed perfect correlations between them which indicates they had quantified the same aspects of shape of an aggregate, and could be considered as surrogate parameters. Simultaneously, several others reveal very high correlation (proportional and inversely proportional correlations that were indicated by positive and negative coefficients), indicating that they may be quantifying similar aspects of shape of an aggregate, and not the same aspects. Shape factors including SF.FR, BSF and KS showed no correlation to any other shape factors, indicating they quantify some shape aspects of aggregates that are not assessed by the remaining shape factors. This observation vouches further analysis on the dimensionality of the data matrix considered for this study.

In the meantime, even though some shape factors show very high correlation to other shape factors as discussed above, the plots in Figure 5 shows that those shape factors show nonlinear correlation that deviate from each other on the extreme numerical values. The Pearson's correlation coefficients (linear correlations) showed almost close to 1, due to more aggregate particles lay in the middle range contributing to lower deviations.

Figure 6 shows the distribution of five shape factors (J.C, C, J.R, ICS and SF) for aggregate particles milled

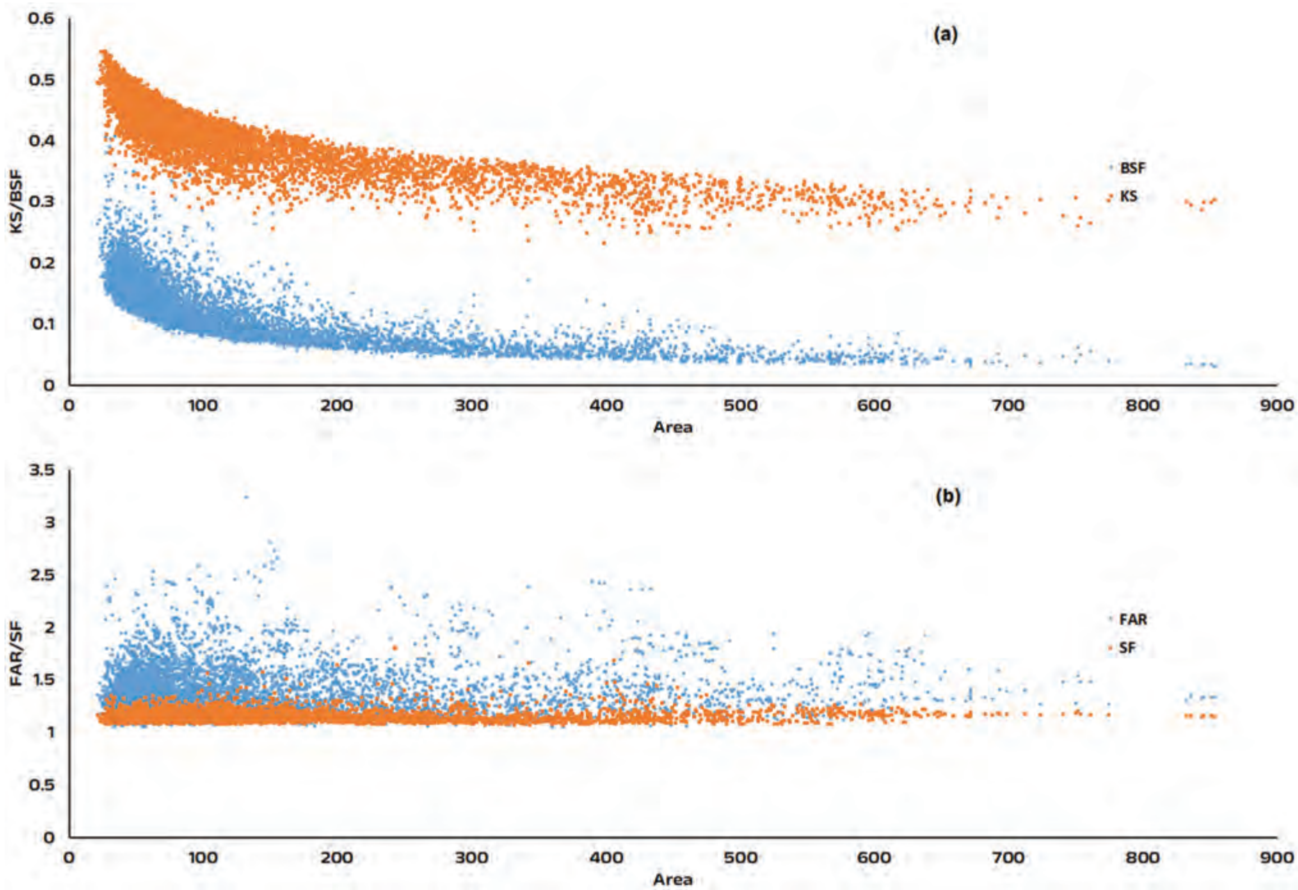

Figure 4. Variation of shape factors with projected area of aggregate particles.

for 1500 revolutions. In the first row (a), a bar in J.C was selected (a group of aggregates that had similar J.C value), and the distribution of that group of aggregates are shown in other shape factors in the same row. It is observed that most of those aggregate particles get grouped in one bar in shape factor C and most of the aggregate particles are grouped in two bars in shape factor SF. However, they are widely distributed in shape factors J.R and ICS. As discussed in correlations, J.C, C and SF have a Pearson's correlations coefficient of almost 1.0 indicating strong correlation. Further analysis on the plots indicate that the bar in shape factor C and SF have most aggregates selected from shape factor J.C, but not alone. They do have more aggregate particles that belonged to other bars in J.C. This is further observable from the plots given in second row (b), where a group of aggregate particles are selected from shape factor C that were distributed in two bars in J.C. Simultaneously, a group of aggregate particles selected from J.R in (c) are grouped in two bars in ICS and are widely distributed in other shape factors. Analysing with the correlations coefficients, J.R and ICS have a strong correlation, although the distribution indicates surrogacy, the deviations are quite significantly different. Therefore, further analysis is required to select shape factors that represent aggregate morphology.

Principal component analysis (PCA) is required to understand the dimensionality of the data matrix with regard to the total variance in the same. A KMO and Bartlett test was carried out which returned a Kaiser-Meyer-Olkin value of 0.697. This indicated that the subsequent PCA would be credible for interpretation. From the PCA, only three PCs had an eigen value more than 1.00, where other PCs were considered insignificant that contained eigen values less than 1. Table 3 shows the percentage of total variance in the data matrix explained by each principal component derived from the PCA. Only three PCs had an eigen value more than 1.00, where other PCs were considered insignificant that contained eigen values less than 1. The first PC explains approximately half of the variance observed in the data matrix (46.505%), while all three PCs together explain more than 90% of the total variance. This indicates that 90% of the variance of the data matrix that had 14 dimensions could be explained with just three dimensions. Considering the noise in the data and that assuming 10% could be attributed to the noise, the total data

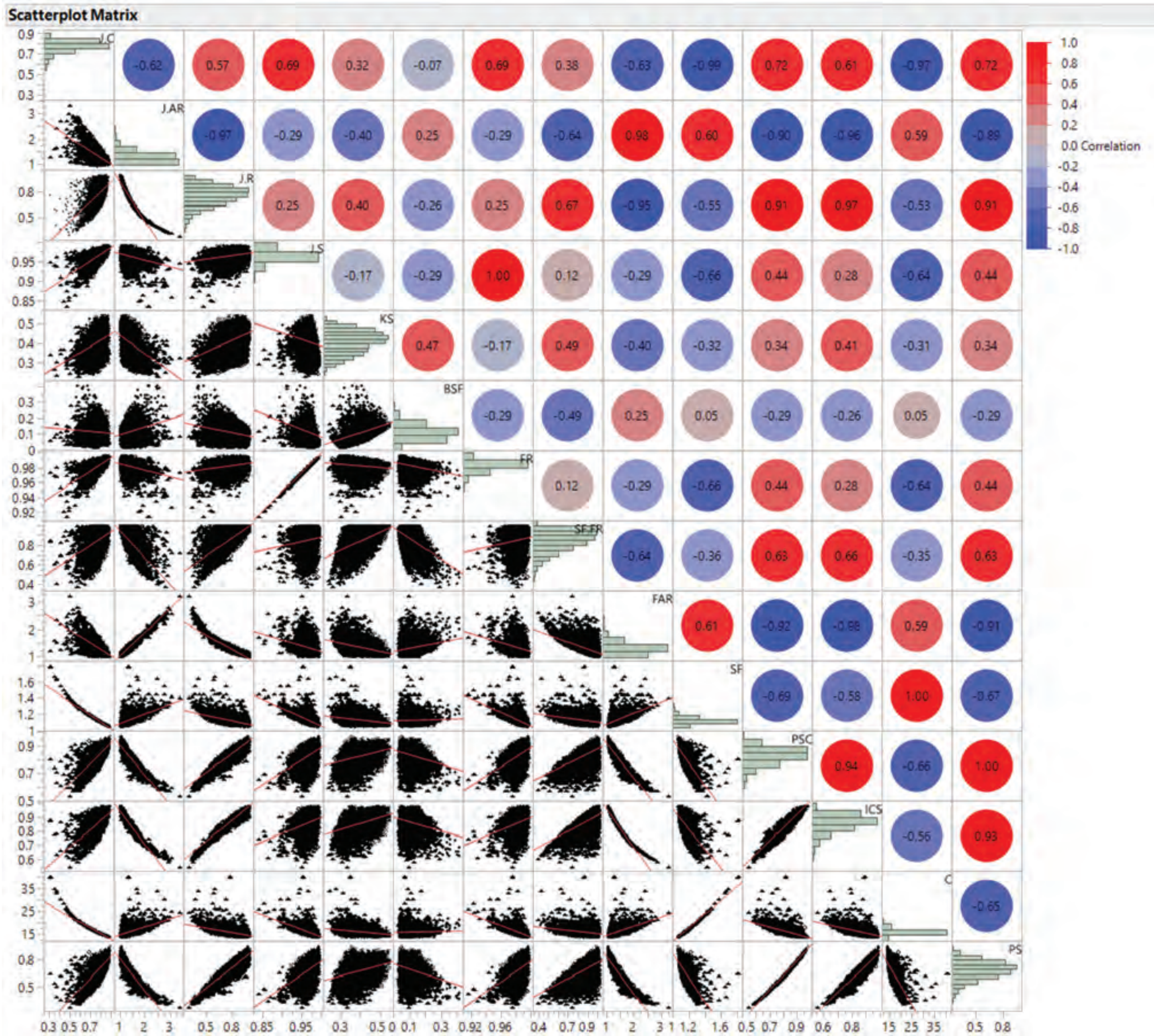

Figure 5. Pearson's correlation coefficients between shape factors.

matrix could be reduced to 3 dimensions from the 14 dimensions that it represents.

Table 4 and Figure 7 show the rotated components (about varimax) from the PCA. First PC is dominated by J.R, J.AR, FAR, SF.FR, PSC, ICS and PS (above 0.85) indicating that these parameters could be considered as a single dimension. However, loading of these variables may be on either side of the PC1, indicating similar but not exact surrogate parameters, which may get an indication from the Pearson's correlation. The second PC is dominated by J.C, J.S, FR, SF and C, where these too could be considered as a single dimension when the total variance of the data matrix is considered. Out of the 14 shape factors analysed, 12 factors have been aligned to two orthogonal principal components, where the remaining two shape factors KS and BSF lay along the third PC. Their loading along the third PC is not as affirmed as several other variables along first two PCs. In addition, KS, SF.FR and BSF have been observed not to correlate linearly with any other shape factors, which may indicate that they could well be individual dimensions. However, considering the total variance observed and that the noise in the data should also be considered, the individual dimensions these shape factors represent may not be very significant in the shape variations observed in the aggregates used in this study.

3.3.1. Feature selection using machine learning tools

The aggregates are grouped into eleven classes, based on number of revolutions the aggregates were milled, 0,

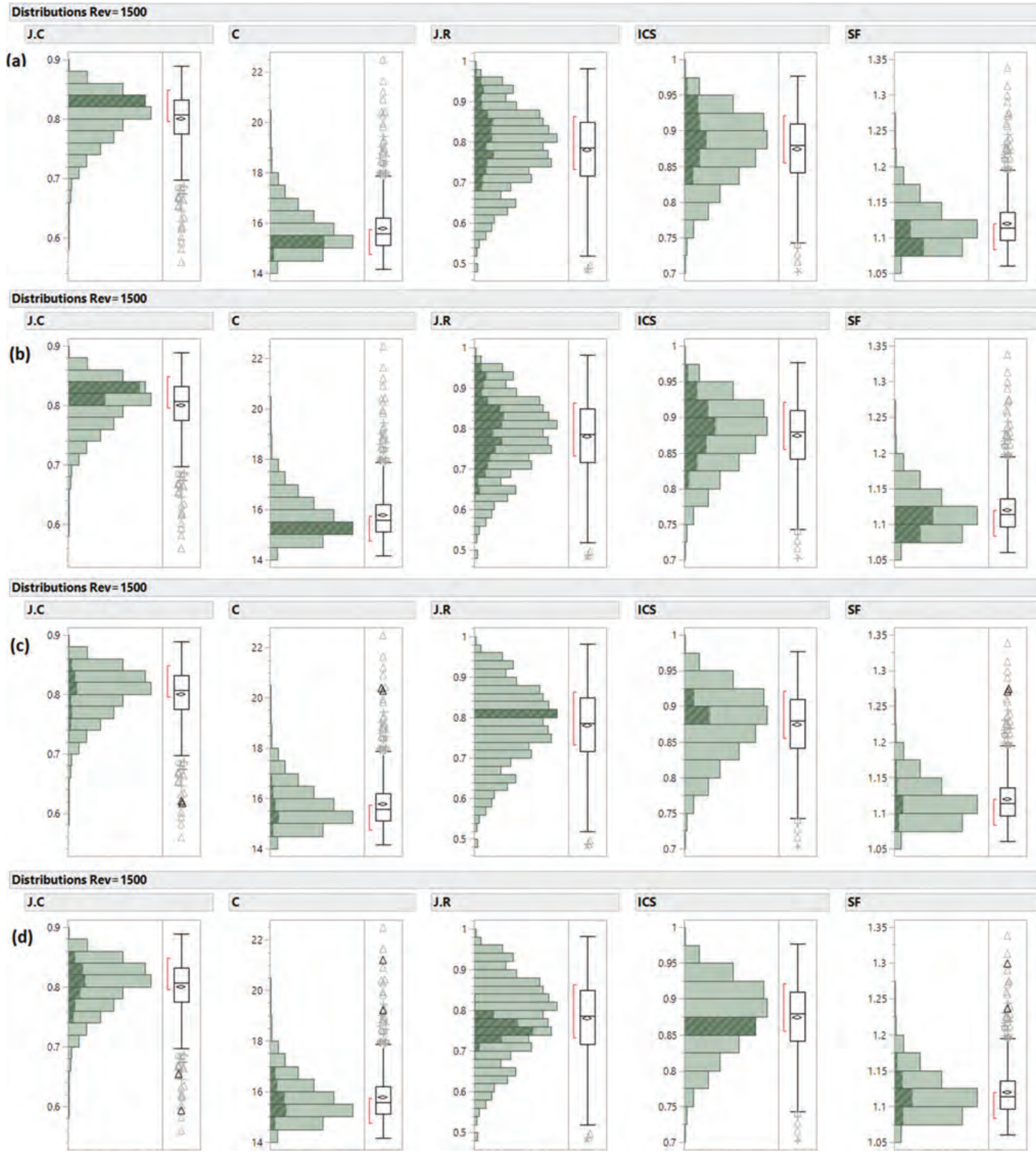

Figure 6. Distribution of aggregate particles across different shape factors (a) a bar in J.C, (b) a bar in C, (c) a bar in J.R and (d) a bar in ICS.

Table 3. Variation explained by each PC from PCA.

PC	% of Variance	Cumulative %
1	46.505	46.505
2	31.059	77.564
3	12.49	90.054

500, 1000, 1500 and 2000. Analyses were done in the following cases, shown in Table 5.

Two feature selection methods, Feature Importance (FI) and Mutual Information (MI), were employed to

Table 4. Rotated (varimax) components and variables.

	PC1	PC2	PC3
J.C	0.437	0.850	0.197
J.AR	−0.924	−0.262	−0.060
J.R	0.945	0.202	0.038
J.S	0.072	0.900	−0.317
KS	0.446	−0.025	0.796
BSF	−0.360	−0.038	0.869
FR	0.072	0.900	−0.315
SF.FR	0.796	0.013	−0.080
FAR	−0.930	−0.266	−0.061
SF	−0.410	−0.847	−0.223
PSC	0.871	0.419	−0.001
ICS	0.948	0.239	0.045
C	−0.394	−0.838	−0.231
PS	0.868	0.411	−0.009

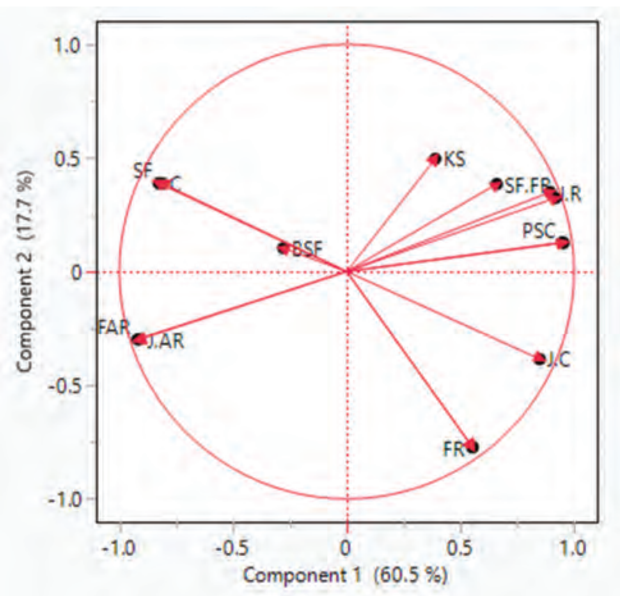

Figure 7. Loadings of rotated matrix of variables about the first and second principal components.

identify the critical shape factors (features). Figure 8 shows the Feature importance (FI) and Mutual Information (MI) for different shape factors (methods). It is evident from the figure that both feature selection methods are congruent and that five shape factors are dominant in the classification of aggregate particles in all the cases analysed. They are namely BSF, KS, C, J.C and SF. Further analysis into the contribution of each shape factor in the classification according to FI and MI shows that there are significant differences in the contribution between the cases analysed. In the figure, each ring corresponds to a case analysed; the inner ring corresponding to C01 and the outer ring corresponding to C08. It could be observed that although there is

Table 5. Details of cases analysed for group of aggregates.

Case	label	Groups Analysed
Case 01	C01	0 Rev – 500 Revs
Case 02	C02	0 Rev – 1000 Revs
Case 03	C03	0 Rev – 2000 Revs
Case 04	C04	500 Revs – 1000 Revs
Case 05	C05	500 Revs – 1500 Revs
Case 06	C06	1000 Revs – 1500 Revs
Case 07	C07	1000 Revs – 2000 Revs
Case 08	C08	1500 Revs – 2000 Revs

significant difference between the contributions of each shape factor, the five factors remain dominant, with a shift in the ranking. Contribution of a shape factor indicates how much it influences in classifying aggregate particles based on the degree of milling. Therefore, change in the contribution of a shape factor in the process of classification indicates a shift in dominance of a shape factor in representing the aggregate morphological changes induced by degree of milling and that the degree of milling in the meantime causes different aspects of changes in the aggregate morphology. Therefore, dominance of a shape factor in the process of classification of different degree of milling indicates that it represents the specific morphological changes that occurred in the degree of milling. The five shape factors dominating generally in all cases indicate that either the process of milling induces a specific morphological aspect that are represented by the five shape factors or the morphological changes induced by the milling process is dominant in defining the shape of the aggregate. Effectiveness of the classification needs to be analysed to obtain the combinations that yield best classification in the machine learning methods analysed.

3.3.2. Classification using machine learning tools

The following machine learning algorithms (Table 6) were employed to classify aggregate particles into degree of milling using combination of 14 shape factors.

Figure 9 shows the accuracy of various machine learning models in classifying images into 5 classes (labelled data) that correspond to 0, 500, 1000, 1500 and 2000 revolutions of milling of aggregates. The figure also shows the accuracy when different number of features (shape factors) used, as shown in the legend. The features of subsequent classification were based on the feature method, ‘mutual information’. It is observed that when all shape factors are used, the accuracy is the highest that is achieved for all models. However, it is

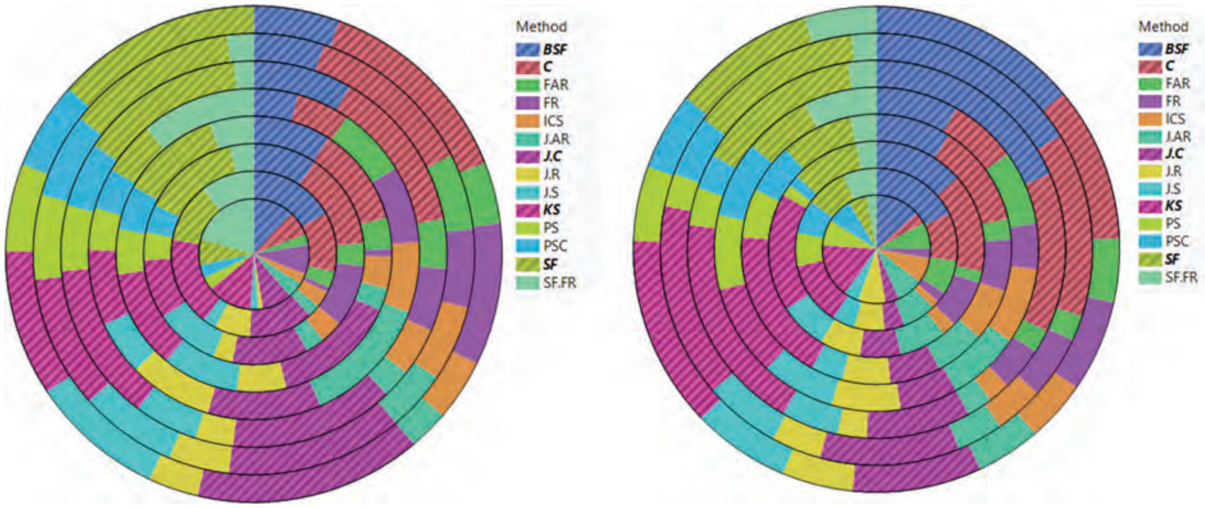

Figure 8. Feature selection (FI and MI) for different shape factors (methods).

Table 6. List of machine learning algorithms used in classification of aggregate particles.

Abbreviation	Machine Learning Algorithm
RF	Random Forest
NB	Naïve Bayes
L-SVM	Linear Support Vector Machines
SVM-RBF	Support vector machine – radial basis function (kernel)
SVM-P	Support vector machine – Polynomial
DT	Decision Tree
KNN	K Nearest Neighbours
P	Perceptron

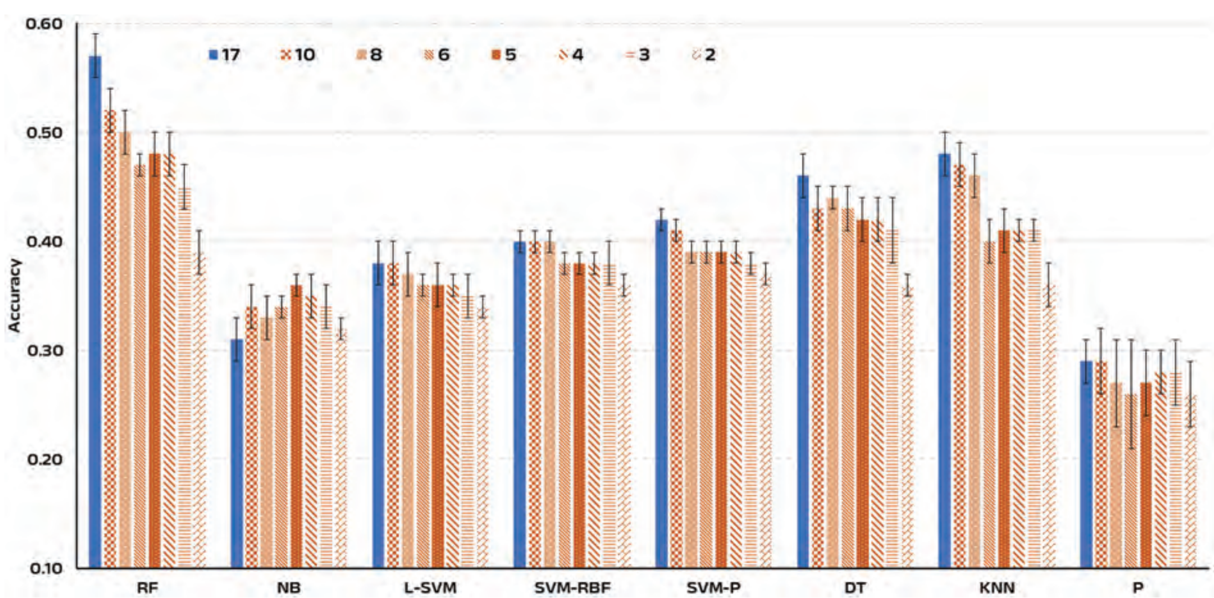

Figure 9. Accuracy of clustering of 5 clusters using different methods of machine learning for different number of features.

worthy to note that for several models other than RF, the variation in the accuracy has not changed significantly, as compared to the observation on RF model. Yet, RF model gives the highest accuracy for classification of 0.57 ± 0.02 for inclusion of all factors. It could therefore be established that there is significant difference between the classes of aggregates that have undergone different degrees of milling. This difference corresponds to the shape of the aggregates that has undergone deformation during the process of milling.

Figure 10 shows the variation of accuracy of RF model fitted in Figure 9, with the number of features, selected based on 'mutual information' feature selection method. It is observed that there is a significant difference in the accuracy between when all shape factors were used and when only 2 selected factors were used. However, the rate at which the accuracy decreases need to be further investigated to determine the optimum number of features needed to perform the classifying of aggregates, based on their shape. Although there is a linear decrement in accuracy of approximately 0.1 (from 0.57 to 0.49) between the cases when 14 and 8 factors were used, respectively, the accuracy then remains unchanged till about the case when 4 features were used. The graphs then rapidly plunge in accuracy when the number of features used for classification is further decreased. This indicates that the classification could be done satisfactorily based on 4 shape factors. This in other words means that the four shape factors are adequate to define the shape of the aggregates.

In order to investigate if the milling imparted significant deformation throughout the total cycles used in this study, a further analysis was done with only three classes, that correspond to 0, 1000 and 2000 revolutions. Figure 11 shows the variation of accuracy of machine learning models in classifying aggregates into 3 classes and with selection of features based on mutual information feature selection method. The figure is similar to Figure 9, which was analysed based on 5 classes of aggregates. An immediate observation between the two figures is that the accuracy of all models have significantly increased with 3 classes analysis compared to 5 classes analysis. This arises from the fact that the differences between the 3 classes is more pronounced than the difference between the adjacent classes when 5 classes of aggregates were used.

In Figure 11, it is observed that the model RF performs significantly higher than other models, just as in the cases with 5 classes analysis, while several other models including KNN and DT have also performed higher in classifying compared to 5 classes analysis. In addition, given the fact that all models have shown an accuracy above 0.5, it could be established that there is a well-pronounced difference between the shapes of the aggregates. It is also observed that for models other than RF and KNN, the accuracy of other models has not significantly deviated when the number of features used was decreased.

When analysing the two highest performing models, it is observed that the number of features used did have an effect on the accuracy of the model. Figure 12 shows the variation of accuracy of the models RF and KNN, with the number of features used for classifying. It is observed that both the highest performing models resort to the same trend, where the accuracy is somewhat constant until the number of features was reduced to five, and then sharply decreases to significantly low performance of the model in classifying aggregates. In addition, the RF model showed a slightly increased performance when all features were used, which was not significant in the KNN classification method.

From Figure 12 it is also observed that the accuracy was almost same when two or three features were used for classification. It is concluded that an optimum of five features selected from mutual information concept to

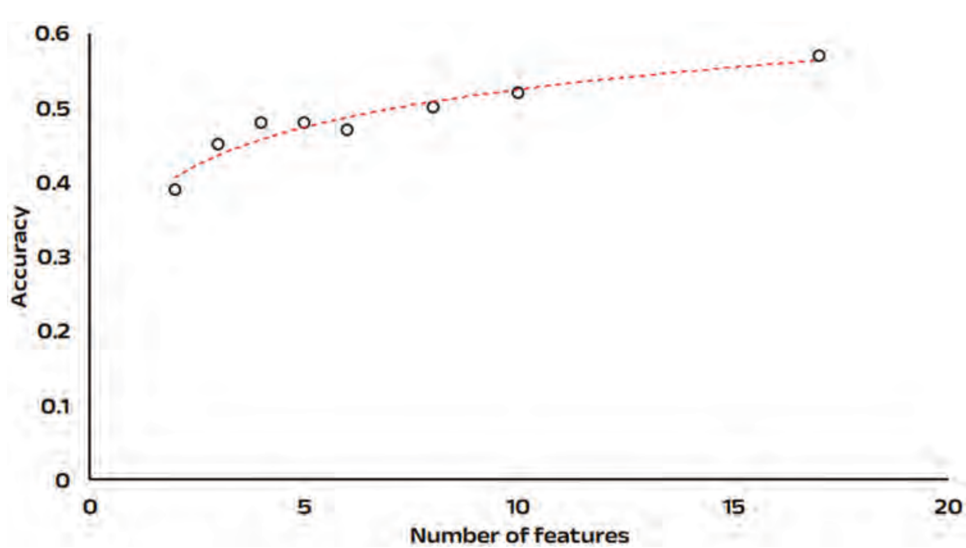

Figure 10. Variation of accuracy of random forest model with number of features used, for the case shown above.

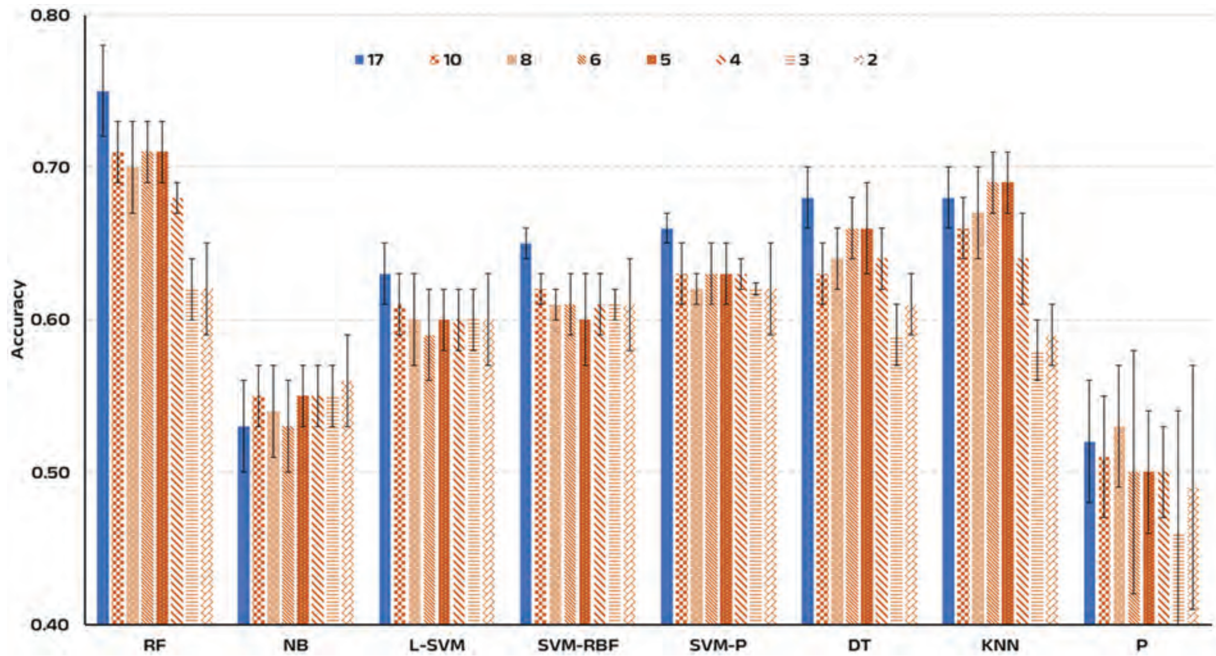

Figure 11. Accuracy of clustering of 3 clusters using different methods of machine learning for different number of features.

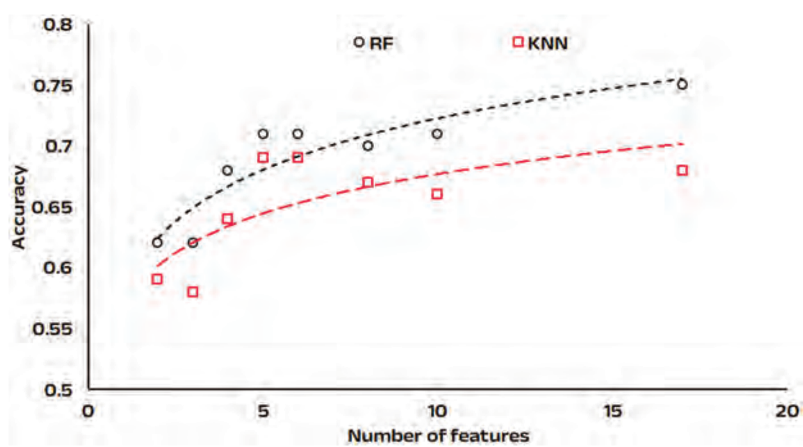

Figure 12. Variation of accuracy of RF and KNN models with number of features used, for the case shown above.

distinguish between classes of aggregates could satisfactorily define their shape.

3.3.3. Classification between 0 and other revolutions (degree of milling)

Figure 13 shows the variation of accuracy in classifying unmilled aggregates (0 revolutions) against different degrees of milling (500, 1000 and 2000 revolutions), for the highest performing two machine learning algorithms, DT and KNN classifiers. It is apparent from the figure that higher the degree of milling, more distinct the aggregates are from unmilled aggregates, indicated by increasing accuracy for both algorithms (it was also observed across other lower performing ML algorithms). It is therefore evident that more morphology aspects changed by the process of milling were almost emphatically identified by the shape factors between 0 and 2000 revolutions indicated by an accuracy of approximately 0.97, while it was high for both 0–500 and 0–1000 classifications as well indicated by accuracies 0.90 and 0.92, respectively. Using 10-fold cross validation, the accuracy is further assured with an uncertainty of less than 0.02, which reinforces the robustness of the classification models. The values mentioned above consider all 14 shape factors for the process of classification. Looking closely at the accuracy values, almost all three classifications have more than 0.9 accuracy, where the reduction from 0.97 to 0.90 between 0–2000 and 0–500 could be mainly due to more aggregates

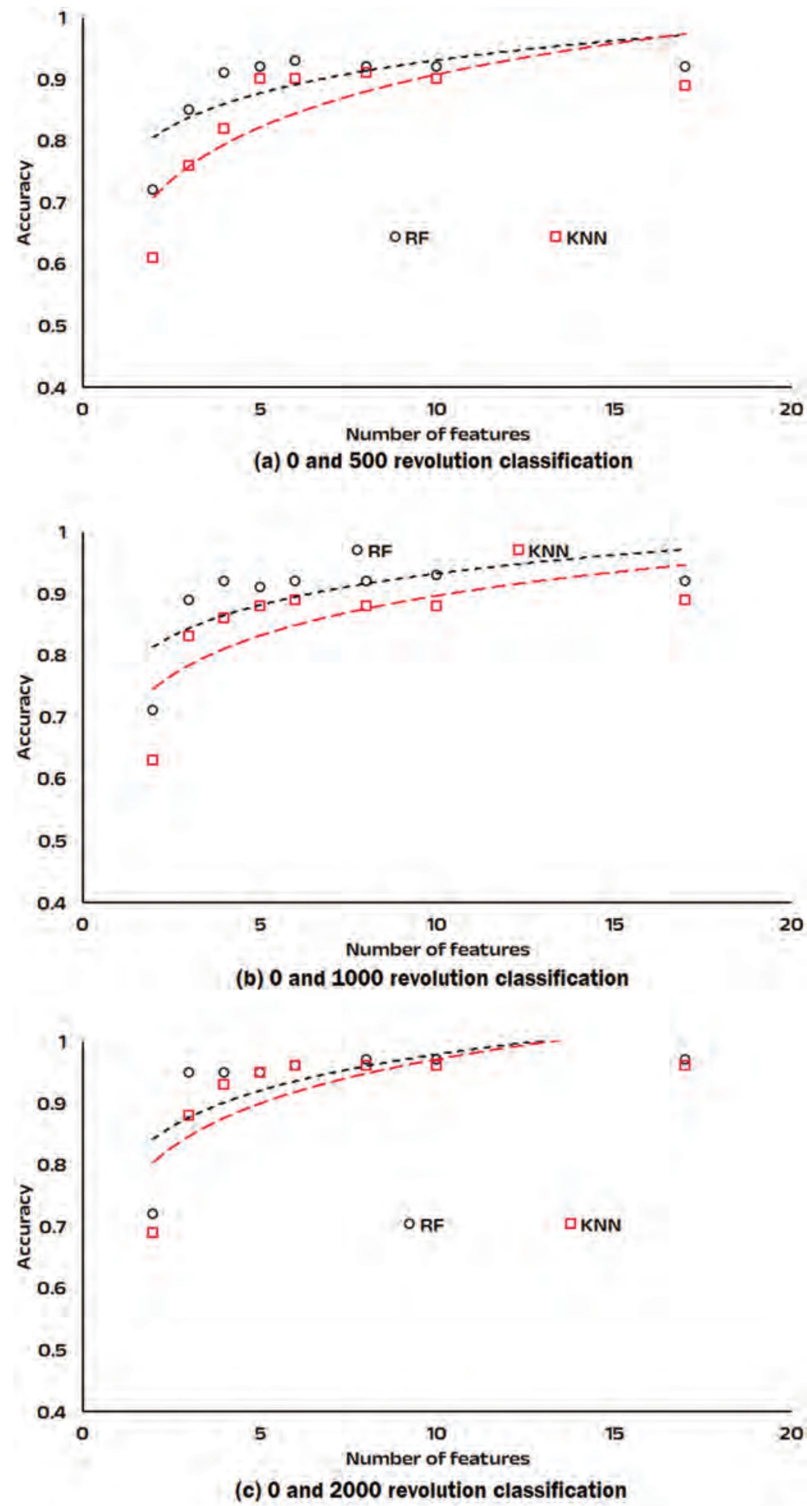

Figure 13. Accuracy of models DT and KNN in classifying aggregates based on degree of milling against unmilled aggregates.

getting equal milling treatment (more uniform distribution of treatment) with increased duration of milling.

On closer observation on the trend of accuracy against the number of features considered for the classification models (DT and KNN), a rather constant accuracy was observed as the number of features (shape factors) considered was reduced up to at least 4 features, for all cases and both ML algorithms. A more distinct deflection is observed at three features in 0–2000 revolution classification for both ML algorithms, where the accuracy was constant during reduction up to three features and

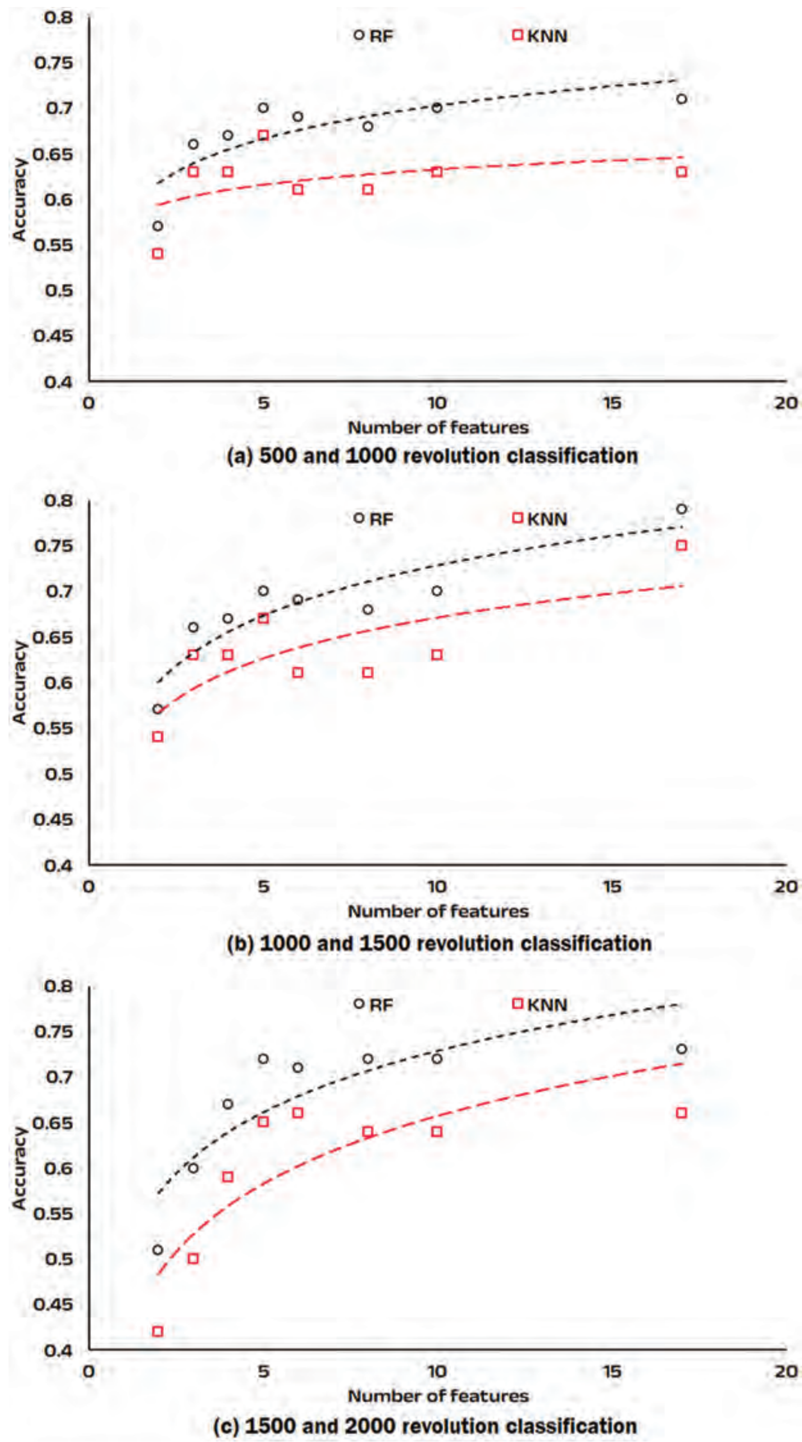

Figure 14. Accuracy of models DT and KNN in classifying aggregates based on degree of milling.

then a rapid drop to 0.72 and 0.69 for DT and KNN, respectively. This indicates that those three features (shape factors), J.C or C, SF and KS, classified the aggregates while other shape factors did not contribute. In the meantime, a similar drop was observed for the other two classifications after reductions to four features. Yet, the drop at three features was more than 0.8 for both algorithms, indicating that three variables were effectively classifying aggregate those milled from unmilled aggregates. Therefore, it could be concluded that J.C/SF/KS combination could effectively quantify the morphological changes on aggregates induced by process of milling. This observation was similar to that was observed in statistical analyses, where the accuracy was observed to be lower.

Observation from Figure 14 shows that the classifications between different degrees of milling (500–1000, 1000–1500 and 1500–2000) indicate a significantly lower accuracy, approximately 0.7. An accuracy of 0.7 indicates that, for example, out of 100 aggregates, the algorithms correctly identified the class of only 70 aggregates. The morphological changes on the aggregates induced by different degrees of milling were not entirely captured by the computational shape factors employed or that the uncertainty on the treatment of milling received by the aggregates are significantly higher. Furthermore, a gradual drop in accuracy was observed beyond five features in all the cases and a rapid drop after three features, when the number of features used to classify was gradually reduced. It is therefore concluded that at least three shape factors, J.C/SF/KS, could numerically represent the shape factor of aggregates.

3.3.4. Difference between statistical clustering and machine learning classification

There is a significant difference between statistical clustering (K-means cluster analysis) and machine learning classification. Statistical clustering (K means) initially assumes aggregates into the number of clusters prescribed by the user and computes the cluster centre and then through iterations alters the membership of aggregates so as to group the aggregates that are close to the cluster centre. The method does not give an indication if a group of aggregate particles would be expected to be grouped together. The cluster and the aggregate-cluster distances are optimised, so as to distinctively group aggregates and the choice of features would be in correspondence to the optimisation of the distances. Since the objective of the study is to analyse the effectiveness of the shape factors in quantifying the morphological changes in aggregates induced by milling, and that the classes to which the aggregates belonged (based on number of revolutions), statistical clustering may not be relevant and thence not employed in this study.

Machine learning classifications, on the other hand, specifically supervised algorithms such as used in this study, are given the class membership defined by the user. The algorithms are expected to optimise the boundary plane or surface between classes, so as to optimise the club membership. The distances between the clusters are vital, yet, it would be optimised to increase the club membership by optimising feature selection and weightages. Decision tree decides on the class membership based on a conditional classification tree that is further believed to accurately determine distinction between the classes, and it is observed to be the best model to classify the aggregates in this study. Decision trees are widely used in material characterisation and property identification, compared to other machine learning algorithms.

4. Conclusions

This study analysed the aggregate shape factors, computed based on dimensions obtained from images of aggregates, in representing the shape aspects of aggregate particles ranging from 5 mm to 30 mm in diameter. The following conclusions are made from the analysis.

- Shape factors had varied distribution patterns among the group of aggregates (degree of milling), that significantly changed across groups. Most shape factors had a significant difference in mean value between 0 Rev group and 500 Rev group, where the texture and angular flaky corners of aggregates were observed to change. Other than BSF and KS, other shape factors did not show a significant difference between the means 1000, 1500 and 2000 Rev aggregates. Histogram of shape factors showed a significant difference in the range and distribution, specifically for J.C and J.R between milled and unmilled aggregates, while KS and BSF showed similar distributions. Meanwhile, the range of 95% confidence was significantly reduced for J.R and J.AR, indicating the effect of milling on the unique texture the crushed aggregates displayed. None of the shape factors showed a linear correlation with revolutions, indicating that none were capable of assessing the shape changes due to milling, as a single factor.
- Shape factors in pairs such as SF/FR, J.R/J.AR, SF/C, J.S/FR, PSC/PS, SF/FR and J.R/ICS had Pearson's correlation coefficients 1 or close to 1. While most shape factors had significant correlations indicated by a value at least greater than 0.7 in Pearson's Correlation test, KS and BSF showed no significant correlation to any other shape factors. A principal components analysis (PCA) suggested the 14-dimensional data matrix be reduced to 3 dimensions explaining more than 90% of the total variance. The shape factors J.R from PC1, J.C and SF from PC2 and KS and BSF from PC3 were chosen as influencing shape factors for combinational analysis.
- Feature selection methods in machine learning tools have indicated dominance of five shape factors, namely BSF, KS, C, J.C and SF. All eight classification methods indicated adequacy of three

or four shape factors in classifying the aggregates based on degree of milling, while the accuracy highly varied. Classifying three clusters of milling degrees (0, 1000 and 2000 Revs) had RF and KNN performing the highest accuracy with above 0.7.

- Classifying two degree of milling exercise revealed very high accuracy (above 0.9) for most classification methods (with RF being the highest). A high classification accuracy is observed between unmilled aggregates and any degree of milling (even with 0 and 500 Rev), while significantly lower accuracy was observed (approximately 0.75) for other degree of milling. It concludes that the dominant morphological change induced by the process of milling was accurately represented by the shape factors, while the morphological changed induced by degree of milling had slightly lower representation with the shape factors considered.
- The accuracy of classifying of pairs 0 and 500 Rev and 0 and 2000 Rev are 0.9 and 0.97 accordingly which conclude that the degree of milling increases the number of equally treated aggregates during milling process.
- Two class classifications indicated that two or three shape factors were efficient in classifying the aggregates. From feature analysis, the shape factors were identified to be BSF, KS and C. It concludes that three shape factors numerically represent the shape of an aggregate in performance prediction models.

This study analyses different computational methods to numerically quantify the shape of an aggregate, obtained within different context. Shape of an aggregate may include different aspects, and the computational methods may not have been effective in quantifying all. In this study, aggregates were altered of their shape using the ball milling technique as discussed in the manuscript. The process of milling may not affect all aspects of the shape. The findings of this study are therefore limited to the morphological features altered by the process of milling.

Acknowledgments

The authors wish to acknowledge the funding agency, National Research Council of Sri Lanka for the grant (19-045) that funded this research study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Research Council of Sri Lanka [19-045].

Author contributions

D N Subramaniam: conceptualisation, experimental design, data analysis and first draft of the manuscript.

P Jeyananthan: machine learning classification and drafting the manuscript

S H B Wijekoon: Data analysis and manuscript revision

N Sathiparan: Manuscript revision

References

Ahmed, S. and Arocho, I., 2021. Analysis of cost comparison and effects of change orders during construction: study of a mass timber and a concrete building project. *Journal of Building Engineering*, 33, 101856.

Akçaoğlu, T., Tokyay, M., and Çelik, T., 2004. Effect of coarse aggregate size and matrix quality on ITZ and failure behavior of concrete under uniaxial compression. *Cement and Concrete Composites*, 26 (6) 633–638.

Al-Rousan, T., *et al.*, 2005. New methodology for shape classification of aggregates. *Transportation Research Record*, 1913 (1), 11–23.

Al-Rousan, T., *et al.*, 2007. Evaluation of image analysis techniques for quantifying aggregate shape characteristics. *Construction and Building Materials*, 21 (5), 978–990.

Altuhafi, F., O'Sullivan, C., and Cavarretta, I., 2013. Analysis of an image-based method to quantify the size and shape of sand particles. *Journal of Geotechnical and Geoenvironmental Engineering*, 139 (8), 1290–1307.

Arasan, S., *et al.*, 2011. Correlation between shape of aggregate and mechanical properties of asphalt concrete. *Road Materials and Pavement Design*, 12 (2), 239–262.

Bangaru, R.S. and Das, A., 2012. Aggregate shape characterization in frequency domain. *Construction and Building Materials*, 34, 554–560.

Barksdale, R.D., *et al.*, 1991. Measurement of aggregate shape, surface area, and roughness. *Transportation Research Record*, 1301, 107–116

Ben Chaabene, W., Flah, M., and Nehdi, M.L., 2020. Machine learning prediction of mechanical properties of concrete: critical review. *Construction and Building Materials*, 260, 119889.

Bhattacharya, S., Subedi, S., Lee, S. J., and Pradhananga, N., 2020. Estimation of 3D Sphericity by Volume Measurement – Application to Coarse Aggregates. *Transportation Geotechnics* ,23, 100344

Bhutta, M.A.R., *et al.*, 2013. Properties of porous concrete from waste crushed concrete (recycled aggregate. *Construction and Building Materials*, 47, 1243–1248.

Brzezicki, J. M., and Kasperkiewicz, J., 1999. Automatic Image Analysis in Evaluation of Aggregate Shape. *Journal of Computing in Civil Engineering*. 13 (2), 123–128.

Calabrese, L., Campanella, G., and Proverbio, E., 2012. Noise removal by cluster analysis after long time AE corrosion monitoring of steel reinforcement in concrete. *Construction and Building Materials*, 34, 362–371.

Chan, K.W. and Kwan, A.K.H., 2014. Evaluation of particle packing models by comparing with published test results. *Particuology*, 16, 108–115.

Chang, C.S. and Deng, Y., 2020. Packing potential index for binary mixtures of granular soil. *Powder Technology*, 372, 148–160.

Ćosić, K., *et al.*, 2015. Influence of aggregate type and size on properties of pervious concrete. *Construction and Building Materials*, 78, 69–76.

Croce, P., Formichi, P., and Landi, F., 2020. Influence of Reinforcing steel Corrosion on life cycle reliability assessment of existing R.C. Buildings. *Buildings*, 10 (6), 99. doi:10.3390/buildings10060099

Cruz-Matías, I., *et al.*, 2019. Sphericity and roundness computation for particles using the extreme vertices model. *Journal of Computational Science*, 30, 28–40.

Das, N., *et al.*, 2011. Modeling granular particle shape using discrete element method. *Geo-Frontiers*, 2011, 4293–4302.

de Larrard, F. and Sedran, T., 1994. Optimization of ultra-high-performance concrete by the use of a packing model. *Cement and Concrete Research*, 24 (6), 997–1009.

Deng, W., *et al.*, 2021. Shape effects on mechanical properties of maximally random jammed packings of intersecting spherocylinders. *Powder Technology*, 383, 443–453.

Ding, X., Ma, T., and Gao, W., 2017. Morphological characterization and mechanical analysis for coarse aggregate skeleton of asphalt mixture based on discrete-element modeling. *Construction and Building Materials*, 154, 1048–1061.

Dinger, D.R. and Funk, J.E., 2013. Particle-packing phenomena and their application in Materials processing. *MRS Bulletin*, 22 (12), 19–23.

Egbueri, J.C., 2022. Soil erosion and landslide susceptibility insights based on hierarchical clustering and multilayer perceptron networks: a Nigerian case study. *International Journal of Environmental Science and Technology*, 20, 10763–10786.

Egbueri, J.C. and Agbasi, J.C., 2022. Combining data-intelligent algorithms for the assessment and predictive modeling of groundwater resources quality in parts of southeastern Nigeria. *Environmental Science and Pollution Research*, 29, 57147–57171.

Egbueri, J.C. and Igwe, O., 2021. Assessing the role of soil engineering properties in gully development and enlargement in southeast Nigeria using geostatistical and novel indexical techniques. *Environmental Earth Sciences*, 81, 7.

Egbueri, J.C., *et al.*, 2023. Development of MLR and variedly optimized ANN models for forecasting the detachability and liquefaction potential index of erodible soils. *Geosystems and Geoenvironment*, 2, 100104.

Evangelista, L. and de Brito, J., 2007. Mechanical behaviour of concrete made with fine recycled concrete aggregates. *Cement and Concrete Composites*, 29 (5), 397–401.

Fletcher, T., *et al.*, 2003. Aggregate imaging System for characterizing the shape of fine and coarse aggregates. *Transportation Research Record*, 1832 (1), 67–77.

Gan, Y., Kamlah, M., and Reimann, J., 2010. Computer simulation of packing structure in pebble beds. *Fusion Engineering and Design*, 85 (10), 1782–1787.

Ge, H., *et al.*, 2018. Three-dimensional characterization of morphology and abrasion decay laws for coarse aggregates. *Construction and Building Materials*, 188, 58–67.

Ghosh, S., Dhang, N., and Deb, A., 2020. Influence of aggregate geometry and material fabric on tensile cracking in concrete. *Engineering Fracture Mechanics*, 239, 107321.

Gong, F., *et al.*, 2016. A New method for Characterizing coarse aggregate morphology through a MATLAB program. *Geo-China*, 2016, 53–60.

Gu, X., Tran, Y., and Hong, L., 2014. Quantification of coarse aggregate shape in concrete. *Frontiers of Structural and Civil Engineering*, 8 (3), 308–321.

Hameed, M.M., *et al.*, 2021. Incorporation of artificial neural network with principal component analysis and cross-validation technique to predict high-performance concrete compressive strength. *Asian Journal of Civil Engineering*, 22, 1019–1031.

Kang, M.-C., Yoo, D.-Y., and Gupta, R., 2021. Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete. *Construction and Building Materials*, 266, 121117.

Kim, Y., Suh, H.S., and Yun, T.S., 2019. Reliability and applicability of the krumbein-sloss chart for estimating geomechanical properties in sands. *Engineering Geology*, 248, 117–123.

Kobaka, J., 2021. Principal component analysis as a statistical tool for concrete mix Design. *Materials*, 14 (10), 2668. doi:10.3390/ma14102668

Krejsová, J., *et al.*, 2018. The influence of different aggregates on the behavior and properties of gypsum mortars. *Cement and Concrete Composites*, 92, 188–197.

Kusumawardani, D.M. and Wong, Y.D., 2020a. Evaluation of aggregate gradation on aggregate packing in porous asphalt mixture (PAM) by 3D numerical modelling and laboratory measurements. *Construction and Building Materials*, 246, 118414.

Kusumawardani, D.M. and Wong, Y.D., 2020b. The influence of aggregate shape properties on aggregate packing in porous asphalt mixture (PAM. *Construction and Building Materials*, 255, 119379.

Kwan, A.K.H., Ng, P.L., and Huen, K.Y., 2014. Effects of fines content on packing density of fine aggregate in concrete. *Construction and Building Materials*, 61, 270–277.

Lewis, D.W. and McConchie, D., 1994. *Textures. Analytical sedimentology*. D. W. Lewis and D. McConchie, Ed. 92–129. Boston, MA: Springer US.

Li, K., *et al.*, 2021. Modeling and Sensitivity analysis of concrete Creep with machine learning methods. *Journal of Materials in Civil Engineering*, 33, 04021206.

Lin, W., *et al.*, 2016. Development of permeability test method for porous concrete block pavement materials considering clogging. *Construction and Building Materials*, 118, 20–26.

Liu, P., *et al.*, 2017. Modelling and evaluation of aggregate morphology on asphalt compression behavior. *Construction and Building Materials*, 133, 196–208.

Lu, S., *et al.*, 2020. A novel feature selection approach based on tree models for evaluating the Punching Shear Capacity of steel fiber-reinforced concrete flat slabs. *Materials*, 13 (17), 3902. doi:10.3390/ma13173902

Maerz, N.H., 2004. Technical and computational aspects of the measurement of aggregate shape by digital image analysis. *Journal of Computing in Civil Engineering*, 18 (1), 10–18.

Mahmoud, E., Masad, E., and Nazarian, S., 2010. Discrete element analysis of the influences of aggregate properties and internal structure on fracture in asphalt mixtures. *Journal of Materials in Civil Engineering*, 22 (1), 10–20.

Man, H.K. and van Mier, J.G.M., 2011. Damage distribution and size effect in numerical concrete from lattice analyses. *Cement and Concrete Composites*, 33 (9), 867–880.

Mangalathu, S., *et al.*, 2020. Data-driven machine-learning-based seismic failure mode identification of reinforced concrete shear walls. *Engineering Structures*, 208, 110331.

Maroof, M.A., Mahboubi, A., and Noorzad, A., 2020. *A new method to determine specific surface area and shape coefficient of a cohesionless granular medium*. Elsevier, Advanced Powder Technology. https://www.sciencedirect.com/science/article/pii/S092188312030251X#:~:text=In%20the%20current%20study%2C%20the,dimensional%20(3D)%20image%20analysis

Masad, E., Button, J.W., and Papagiannakis, T., 2000. Fine-Aggregate Angularity: Automated Image Analysis Approach. *Transportation Research Record*, 1721 (1), 66–72.

Mora, C.F. and Kwan, A.K.H., 2000. Sphericity, shape factor, and convexity measurement of coarse aggregate for concrete using digital image processing. *Cement and Concrete Research*, 30 (3), 351–358.

Mostofinejad, D. and Reisi, M., 2012. A new DEM-based method to predict packing density of coarse aggregates considering their grading and shapes. *Construction and Building Materials*, 35, 414–420.

Mueller, G.E., 2010. Radial porosity in packed beds of spheres. *Powder Technology*, 203 (3), 626–633.

Nebeokike, U.C., *et al.*, 2020. Erodibility characteristics and slope stability analysis of geological units prone to erosion in udi area, southeast Nigeria. *Modeling Earth Systems and Environment*, 6, 1061–1074.

Ng, P.-L., Kwan, A.K.-H., and Li, L.G., 2016. Packing and film thickness theories for the mix design of high-performance concrete. *Journal of Zhejiang University-SCIENCE A*, 17 (10), 759–781.

Pouranian, M.R. and Haddock, J.E., 2018. Determination of voids in the mineral aggregate and aggregate skeleton characteristics of asphalt mixtures using a linear-mixture packing model. *Construction and Building Materials*, 188, 292–304.

Rueden, C.T., *et al.*, 2017. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18 (1), 529.

Shen, P., *et al.*, 2019. The effect of curing regimes on the mechanical properties, nano-mechanical properties and microstructure of ultra-high performance concrete. *Cement and Concrete Research*, 118, 1–13.

Shuguang, L. and Qingbin, L., 2015. Method of meshing ITZ structure in 3D meso-level finite element analysis for concrete. *Finite Elements in Analysis and Design*, 93, 96–106.

Su, D. and Yan, W.M., 2018. Quantification of angularity of general-shape particles by using Fourier series and a gradient-based approach. *Construction and Building Materials*, 161, 547–554.

Tafesse, S., Fernlund, J.M.R., and Bergholm, F., 2012. Digital sieving-Matlab based 3-D image analysis. *Engineering Geology*, 137-138, 74–84.

Unigwe, C.O., *et al.*, 2023. Understanding the influence of geotechnical and geomorphological characteristics on the erosional processes of two geologic units in udi and Aguata, SE Nigeria. *Arabian Journal of Geosciences*, 16, 231.

Wang, L., *et al.*, 2005. Unified method to quantify aggregate shape angularity and texture using Fourier analysis. *Journal of Materials in Civil Engineering*, 17 (5), 498–504.

Xie, X., *et al.*, 2017. Evaluation of morphological characteristics of fine aggregate in asphalt pavement. *Construction and Building Materials*, 139, 1–8.

Xu, H. and Prozzi, J.A., 2019. Correlation of aggregate form properties with Fourier frequencies. *Transportation Research Record*, 2674 (8), 405–419.

Xu, R. and Di Guida, O.A., 2003. Comparison of sizing small particles using different technologies. *Powder Technology*, 132 (2), 145–153.

Zhang, D., Huang, X., and Zhao, Y., 2012. Investigation of the shape, size, angularity and surface texture properties of coarse aggregates. *Construction and Building Materials*, 34, 330–336.

Zhang, S., Li, R., and Pei, J., 2019. Evaluation methods and indexes of morphological characteristics of coarse aggregates for road materials: a comprehensive review. *Journal of Traffic and Transportation Engineering (English Edition)*, 6 (3), 256–272.

Zhao, L., *et al.*, 2020. A digitalized 2D particle database for statistical shape analysis and discrete modeling of rock aggregate. *Construction and Building Materials*, 247, 117906.

Zheng, J. and Hryciw Roman, D., 2016. Roundness and sphericity of soil particles in assemblies by computational geometry. *Journal of Computing in Civil Engineering, American Society of Civil Engineers*, 30 (6), 04016021. doi:10.1061/(ASCE)CP.1943-5487.0000578

Zheng, J., *et al.*, 2020. Three-dimensional particle shape characterizations from half particle geometries. *Powder Technology*, 367, 122–132.

Zhu, Z., Provis, J.L., and Chen, H., 2018. Quantification of the influences of aggregate shape and sampling method on the overestimation of ITZ thickness in cementitious materials. *Powder Technology*, 326, 168–180.