

Vision-based rock-type classification of limestone using multi-class support vector machine

Snehamoy Chatterjee

Published online: 12 October 2012
© Springer Science+Business Media New York 2012

Abstract Rock-type classification is a challenging and difficult job due to the heterogeneous properties of rocks. In this paper, an image-based rock-type analysis and classification method is proposed. The study was conducted at a limestone mine in western India using stratified random sampling from a case study mine. The analysis of collected sample images was performed in laboratory. Color, morphology, and textural features were extracted from the captured image and a total of 189 features were recorded. The multi-class support vector machine (SVM) algorithm was then applied for rock-type classification. The hyper-parameters and the number of input features of the SVM model were selected by genetic algorithm. The results revealed that the SVM model performed best when 40 features were selected out of the 189 extracted features. The results demonstrated that the overall accuracy of the proposed technique for rock type classification is 96.2 %. A comparative study shows that the proposed SVM model performed better than a competing neural network model in this case study mine.

Keywords Image classification · Support vector machine · Feature selection · Genetic algorithm

1 Introduction

A geological rock formation typically consists of several rock types. A mining deposit is essentially composed of a main mineral(s) and several other minerals along with their associated rocks. The characterization of various rock-types

and their identifications at different stages of mining operations is essential for effective design of a mine [1]. There are several instances when the knowledge of the rock-type is useful to mining engineers and geologists for the successful exploitation of a mineral deposit [2]. For example, to understand the local geology, hydro-geology and other geological settings, the knowledge of rock-type is essential. The physical-mechanical properties, granularities, bond strength and textural properties of ore drastically vary from one rock-type to another rock-type. Accurate rock-type classification assists in establishing the previously listed properties of rock, which in turn guides decision making regarding the selection of excavating mining equipment, blast design, and associated fragmented rock transportation [3, 4]. The rock-type information is also highly important in grade control. As the grade of an ore material depends upon the constituent rock-type, information about the rock-type provides certain useful inputs during the grade control and grade monitoring of material [5].

Given the importance of rock-type information at different points of mining operations, the potential benefit of proper rock-type classification is easily established. However, information about rock-types is not always available due to a lack of an adequate data gathering system. Usually, rock-type information is only gathered during the exploration stage and is limited to a few core samples taken from strategic locations. Rock-type information is then generated for a mineral deposit in the form of a lithological map using interpolation techniques [6]. However, with the advancement of computer vision technology, rock-type information can be generated by capturing the images of rocks. Even with these advancements, classification of natural rock is a challenging task, as rocks are rarely homogenous. It is often observed that rock images are non-homogeneous in their shape, texture and color. In spite of this inherent prob-

S. Chatterjee (✉)
Dept. of Mining Engineering, National Institute of Technology
Rourkela, Orissa, 769008, India
e-mail: snehamoy@gmail.com

lem, with the use of advanced digital image processing techniques, complex rock images can be analyzed and rock-type classification can be correctly performed.

Although the study of rock images for rock-type classification is limited, some notable findings have been presented by researchers [7–9]. Several studies were reported in literature for rock image processing in the areas of rock size distribution, fragment analysis, and ore texture analysis. For example, Lepisto et al. [10] and Hunter et al. [11] reviewed the rock fragmentation analysis techniques using computer vision tools with an emphasis on the analysis of blast muckpiles. An image-classification algorithm was presented [12] to estimate the characteristics of rock fragments (size distribution and shape), but results were heavily dependent on image quality. Salinas et al. [13] also analyzed rock fragmentation using digital image processing techniques. Most of these investigations on the ore textural analysis using image processing were focused on the estimation of average particle size and distinguished ore types in industrial ore feed systems [14, 15]. However, Lin et al. [16] worked to develop an online particle size analyzer.

Most of the vision-based techniques are limited by the need to extract a large numbers of image features. The extraction of a large numbers of features is a computationally demanding task. Apart from that, the presence of redundant features may sometimes skew model performance [17]. Moreover, as the number of features grows, the number of training samples required for model development grows exponentially [18]. Therefore, reducing dimensionalities of the image features is required for a valid model development. Chatterjee et al. [19, 20] applied principal component analysis (PCA) on extracted features to reduce the dimensionality for quality parameter modeling. The main disadvantage of PCA-based modeling is that it creates new features through linear combination of the original features. Model development time can be reduced by a PCA approach; however, all features must be extracted to get accurate PCA scores for an unknown rock-type image. Another alternative can be to apply the branch and bound [21, 22] algorithm to identify important features among the set of all available features; however a branch and bound approach requires huge computational time. Genetic algorithms (GA), a heuristic approach, have great advantages for efficient feature selection which provides close to an optimum feature subset within reasonable amounts of time [23].

In this paper, a vision-based rock type classification model is developed using a support vector machine algorithm. The GA method is applied for image feature selection as well as hyper-parameter estimation of the SVM model.

The paper is organized as follows. Section 2 presents a brief overview of the methods adopted in this paper, while a case study with a limestone deposit is presented in Sect. 3. Section 4 draws conclusions and summarizes the results.

2 Methods

The main steps involved in the proposed algorithm are image acquisition and segmentation, feature selection, and model development. These steps are described in the following sub-sections.

2.1 Image acquisition

The quality of an image depends solely on its illumination and the atmospheric condition in which the image is captured. Thus, the illumination and atmospheric condition should be the same throughout the experiment. The experimental setup for the present study consists of a wooden box, illumination system, a digital camera and personal computer. Lighting type, location and color quality play an important role in bringing out a clear image of the object. Uniform diffuse lighting was used throughout the experiment. Four fluorescent tubes (150 mm diameter, 23 W circular tubes, Philips, India) were placed inside the experimental box, equidistant from the center of the box's base with approximately a 45 degree deviation from the horizontal. The box was made of wood having dimension 25 cm × 30 cm × 30 cm. The top of the box was a cylindrically arched bowl shape of approximately 40 cm diameter. The four tubes were fitted at the four sides of the bowl in such a fashion that produces minimal shadow. The box was painted inside with white magnesium oxide to provide uniform and diffused illumination and to reduce glare and specular reflection. The box top had an opening for placing the camera for capturing images. A UPS (Uninterrupted Power Supply) was used to maintain a stabilized power supply to the fluorescent tube for producing illumination of nearly constant luminance in order of 3500–4000 lux. A lux-meter was used to check the illumination inside the chamber before capturing the image. A schematic diagram of an image acquisition setup is presented in Fig. 1.

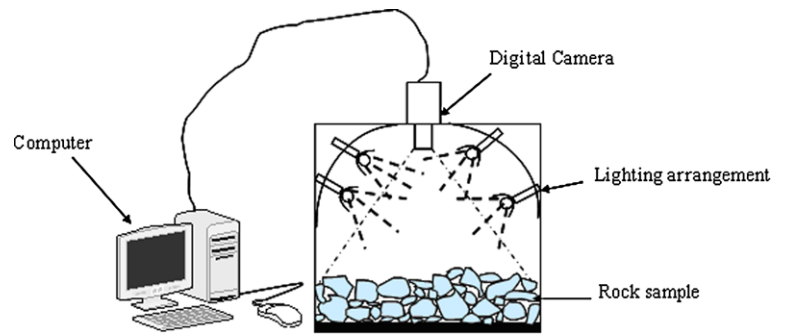
2.2 Image segmentation

After image acquisition, image segmentation is performed to generate a binary image in which each discrete region represents an individual rock sample. The image features are extracted from individual segmented rocks.

The image segmentation technique used in this paper is the watershed technique [24, 25] with pre-processing of the gray image [26–28]. The pre-processing steps consist of image thresholding operation [26], image complement operation [27], and image distance transformation [28].

The thresholding operation was performed to obtain a binary image [26]. Thresholding is the process of converting a gray scale image into binary images to distinguish the object from the background. Manual thresholding is done via

Fig. 1 Schematic diagram of image acquisition experimental setup [10, 19]



trial and error by selecting a threshold value, T , using the histogram of the original image. The thresholded image is then compared with the original image and the process is repeated until the entire object is differentiated from the background. In this paper, the Otsu's method [29] was used. To examine the formulation of this histogram based method, one starts by treating the normalising histogram as a discrete probability mass function:

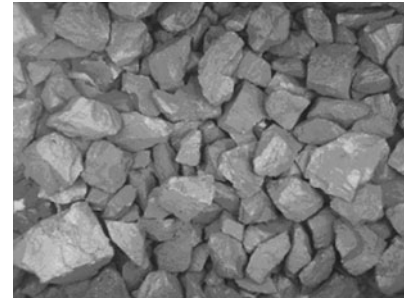
$$p_r(r_q) = \frac{n_q}{n} \quad q = 0, 1, 2, \dots, L - 1 \quad (1)$$

where n is the total number of pixels in the image, n_q is the number of pixels that have intensity level r_q , and L is the total number of possible intensity values.

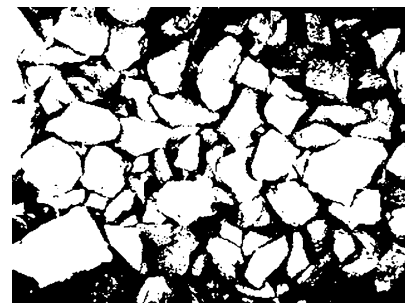
Now suppose the threshold, k , is chosen such that C_0 is the set of pixels with levels $[0, 1, \dots, k - 1]$ and C_1 is the set of pixels with levels $[k, k + 1, \dots, L - 1]$. Otsu's method chooses the threshold value k that maximizes the between-class variance.

Figure 2 shows a gray scale image with its corresponding binary image. The global threshold value was calculated using the Otsu's method. The global threshold value was used to convert the gray scale image to a binary image. This threshold value is a normalized intensity value that lies in the range $[0, 1]$. In this example, the threshold value calculated by the Otsu's method is 0.451.

After thresholding, the image then undergoes a complement operation [27]. Distance transformation is a tool used in conjunction with the watershed transformation. Distance transformation of a binary image is the distance from every pixel to the nearest nonzero valued pixel [28]. This is derived from the Euclidean distance map of the binary image, in which each pixel that forms part of a sample rock is given a value inversely proportional to its distance from the nearest non-rock pixel. The resulting gray scale image may be envisaged as a topographic surface in which rocks are represented by depressions. The 'watersheds' between these depressions were used to segment the binary image. Each depression was gradually 'flooded' until 'water' from one depression overflows into its neighbor. The line along which this occurs is then marked as a watershed and the flooding and marking continues until the image is entirely submerged. The watersheds thus defined are then used to segment the binary



(a)



(b)

Fig. 2 (a) Gray image and (b) its thresholded image using Otsu's method

image [24, 25]. The steps involved in image segmentation process are presented in Fig. 3. To learn more about image segmentation, readers are requested to consult with [30, 31].

2.3 Object identification and feature extraction

The objects in the segmented image are identified using a region labeling algorithm [32]. The features are extracted from each identified rock in an image. The feature extraction process extracts color, morphology and textural features from individual rock samples. The list of the extracted features is presented in Table 1.

The color features extracted were from the seven core components (r, g, b, H, S, I and gray). The type of features included histogram-based features like measure of location (mean, median, and mode), measure of spread (variance, standard deviation, range, mean absolute deviation, inter quartile range), measure of shape (skewness and kurtosis)

Fig. 3 Methodology for image segmentation

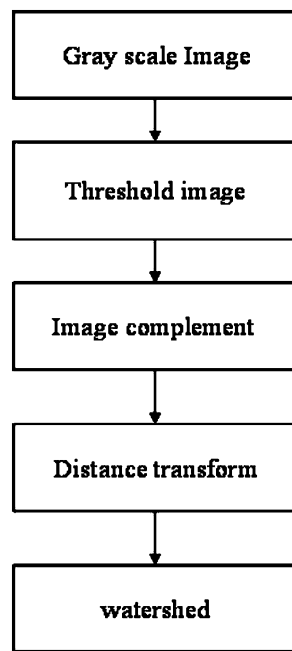


Table 1 Feature extracted from segmented rock sample

Type of feature	Number of feature
Colour feature	112
Measure of location	21
Measure of spread	35
Measure of shape	14
Gray level moment	42
Morphological feature	28
Area	1
Perimeter	1
Major and minor axis length	2
Convex area and FAngle	2
Minimum and maximum radius	2
Derived feature	14
Binary moment	6
Textural feature	49
Statistical feature	4
Co-occurrence matrix	30
Run length matrix	15

and moment (first six moments of histogram) for all seven components [33]. From each component, 16 features were extracted. Therefore, a total number of 112 color features was extracted from all seven components.

The appearance of an object is described by morphological features. The morphological features include area, perimeter, major and minor axis length, convex area, minimum and maximum radius, and F-angle [25]. The area of an object can be defined as the number of pixels contained

within its region. The perimeter of an object is defined as its boundary length, where boundary length is the sum of distances between successive boundary pairs of pixels of an object. Boundary pixels can be identified using 4-neighbour or 8-neighbour connectivity methods. In the 4-neighbour connectivity method the gray level of each pixel relative to its four neighbours is examined. A pixel $X(i, j)$ is considered a boundary pixel if $X(i, j + 1)$ or $X(i, j - 1)$ and $X(i + 1, j)$ or $X(i - 1, j)$ is a background pixel (gray level 0). In the 8-neighbour connectivity method, in addition to the 4 neighbours, the four corner pixels are considered. The perimeter length of objects is determined using the Euclidean distance principle. In the case of the 8-connected method, if the adjacent boundary pixels occur in the horizontal or vertical position, a perimeter length of 1 is added. The perimeter length of 1.414 and 1.207 are added if the neighbouring pixels occur in the diagonal or non-diagonal positions, respectively. Two different shaped objects, for example a circle and square, can have the same number of perimeter pixels. However, perimeter lengths can be used to distinguish these two shapes. The major axis is the length of an object measured through its centroid. The minor axis is the longest length of the object through the centroid that is perpendicular to the major axis. Convex area is the area of the convex hull polygon. Convex hull or convex polygon is calculated from pixel center. This is the smallest convex set containing the object. F-angle is the angle (in degrees) of the major axis with the horizontal. A total of 8 direct measured features were obtained from an image object. A total number of 14 morphological features are derived from measured features:

$$Thickness\ ratio = \frac{Perimeter^2}{Area}$$

$$Aspect\ ratio = \frac{Major\ axis\ length}{Minor\ axis\ length}$$

$$Circularity = 4 * \pi * Area / Perimeter^2$$

$$Roundness = 4 * Area / (\pi * major\ axis\ length^2)$$

$$Area\ Equivalent\ Diameter = \sqrt{(4\pi) * Area}$$

$$Perimeter\ Equivalent\ Diameter = Area / \pi$$

$$Equivalent\ Ellipse\ Area$$

$$= (\pi * major\ axis\ length * minor\ axis\ length) / 4$$

$$Compactness = \sqrt{(4\pi) * Area} / major\ axis\ length$$

$$Solidity = Area / Convex\ area$$

$$Concavity = Convex\ area - Area$$

$$Convexity = Convex\ hull / Perimeter$$

$$Shape = Perimeter^2 / Area$$

$$RFactor = Convex\ hull / (major\ axis\ length * \pi)$$

$$Sphericity = Minimum\ radius / Maximum\ radius$$

To learn more about those derived features, readers are referred to [34, 35]. Apart from these basic and derived features, moments [33], the statistical representation of a binary object, are also calculated from the segmented image. Six binary moment invariant features were extracted for this study [33]. Therefore, a total number of 28 morphological features were extracted.

The types of textural features extracted in this study includes statistical features, co-occurrence matrix features and run length matrix features. Four statistical features, i.e. smoothness, uniformity, entropy, and maximum probability, [36] are considered as statistical features in this paper. The development of a co-occurrence matrix is an important step to extract the co-occurrence features. For this purpose, the lag distance in the co-occurrence matrix has to be optimally chosen. The optimal parameter lag distance (d) of a co-occurrence matrix depends on the resolution of the texture. Zucker and Terzopoulos [37] developed a method to determine this optimal distance. However, this optimum refers to a distinct texture. In the best case, it can be generalized to the entire texture class. In rock type classification, however, several classes (rock types) are involved. Hence, an optimal distance for all classes cannot be determined. Consequently, the textures were studied according to several distances. The experiments were performed with $d = \{1, 5, 10, 15, \text{ and } 20\}$. The energy, entropy, maximum probability, contrast, correlation, and homogeneity [38] were extracted from co-occurrence matrices developed in all 5 lag distance (d). Therefore, a total of 30 co-occurrence features were extracted. In addition to the co-occurrence matrix, the run length matrix was also calculated from the gray scale image of the segmented rock. Nine features of run length statistics proposed by Galloway [39] were extracted. Chu et al. [40] proposed two new run length features to extract gray level information in the matrix. Dasarathy and Holder [41] described another four feature extraction functions following the idea of a joint statistical measure of gray level and run length. A total number of 15 run length features were extracted in this study. A total number of 49 textural features (4 statistical features, 30 co-occurrence features, and 15 run length features) were extracted for each individual rock sample. Thus, each segmented rock was represented by a vector of 189 features (112 colour features, 28 morphological features, and 49 textural features).

2.4 Support vector machine model for rock type classification

2.4.1 Binary support vector machines

SVM (support vector machine) is a popular technique for data classification [42–44]. The SVM is applied for separating two classes by defining the bounding planes such that the margin between both planes is maximized.

To define SVM for the rock type classification problem presented in this paper, the training data were first prepared. Suppose \mathbf{X} is a $n \times m$ matrix, where n is the number of segmented rock images and m the number of extracted features. In this specific problem, the value of m is 189, since 189 features are extracted from each segmented rock image. Denote x_i as a column vector representing the i th row of \mathbf{X} , i.e., x_i representing all 189 features for a specific segmented image i . Also consider that the classification is performed for assigning a specific rock sample i to either rock type ‘A’ or rock type ‘not A’ and \mathbf{y} is an $n \times 1$ vector with value either +1 or –1 such that

$$y_i = \begin{cases} +1 & \text{if } x_i \text{ sample is rock type 'A'} \\ -1 & \text{if } x_i \text{ sample is not rock type 'A'} \end{cases}$$

The aim of the rock type classification algorithm is to develop a SVM model using the image feature data \mathbf{X} and its corresponding rock type \mathbf{y} . The SVM [40] demonstrated that the mapping can be done by constructing a hyper-plane $\langle w \cdot x \rangle + b = 0$, where $w \in R^m$ represents the normal vector associated with the hyper-plane and b is the bias. The developed hyper-plane maximally separates positive (+1) and negative training classes (–1). The margin corresponds to the distance from the separating hyper-planes to the closest samples of each class. The margin is inversely proportional to $\|w\|$. Therefore, by minimising the Euclidian norm of vector w , the maximal separating hyper-plane can be constructed.

The problem of binary classification can be written as in quadratic programming formulation:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \\ \text{Subject to:} \quad & y_i (w \cdot x_i + b) \geq 1 - \xi_i, \quad \forall x_i \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, n \end{aligned} \quad (2)$$

where C is a trade-off variable that controls the relative importance between training error and classifier complexity and ξ_i is a slack variable that permits dealing with non-linearly separable data. The weight vector w can be obtained by solving the dual problem of Eq. (2) by applying Lagrangian algorithm with w represented as:

$$w = \sum_{i=1}^{n_{sv}} \alpha_i y_i x_i \quad (3)$$

where α_i are Lagrange multipliers, n_{sv} is the number of non-zero α values called as a support vector. The value of b is obtained by replacing the w value with $\sum_{i=1}^{n_{sv}} \alpha_i y_i x_i$ in the constraint function in Eq. (2). The hyper-plane, which will be used for classification of rock type of an unknown rock sample k with feature vector z , can then be represented as $f = ((\sum_{i=1}^{n_{sv}} \alpha_i y_i x_i z) + b)$. It is noted that the size of z is the same 1×189 . Now, if the value of function f is ≥ 0 , then

the rock type of sample k is assigned to 'A' and if value of function f is <0 then the rock type of sample k is assigned to 'not A'.

In the hyper-plane presented, the data is considered to be linearly separable. However, when data are not linearly separable, the non-linear mapping function can be used to map the data in high dimensional space where data can be linearly separable. Instead of mapping each data to the high dimensional space, a kernel function can be used. There are number of kernel functions available in literature [42, 43]. In this paper, Gaussian kernels are used: $K(x, x_i) = \exp(-\|x - x_i\|^2/\sigma^2)$, where σ is the bandwidth of the kernel function. The hyper-plane in high dimensional space can then be represented as $f = (\sum_{i=1}^{n_{sv}} \alpha_i y_i K(z, x_i) + b)$.

2.4.2 Multi-class support vector machines

Since the rock type classifications presented in this paper have multiple classes, the multi-class support vector model is used. It is clear from the previous section that the standard SVM is designed only for two-class problems. A standard way to solve multi-class SVM problems is to consider them as a collection of binary sub-problems and then combine their solutions. Two approaches are most commonly employed: the one-versus-all (OVA) and the one-versus-one (OVO) [45]. In this paper an OVA approach is used. If the numbers of rock types are J , then the OVA method constructs J number of SVM models. The j th SVM is developed with all m number of segmented rock images. If a segmented rock image i belongs to rock type j , then the value of y_i is assigned to +1 and if the rock sample i is not belongs to rock type j , then the value of y_i is assigned to -1. After developing J number of SVM models, they are combined to use for multi-class classification. For predicting the rock type of an unknown rock, all 189 image features are extracted from the image of the unknown rock and J number of SVM models are run to get the output of function f . The final rock type of the unknown image is the class that corresponds to the SVM with the highest output. For example, if there are 3 rock types (A, B, and C) and the SVM models (3 models) output for an unknown rock image using function f are 0.2, 0.6, and 0.3, respectively, then the rock type of that particular image is B.

2.5 Genetic algorithms for feature and hyper-parameters selection

To classify rock-types from extracted image features, the support vector machine is used to train the model using available data. It can be seen from Sect. 2.3 that 189 features were extracted from each image and that dealing with a large number of features is computationally demanding.

Therefore, some important features have to be selected from the set of all available features for SVM modeling. It is demonstrated in literature that careful selection of features can improve the performance of the classification algorithm [46, 47]. It is also observed from Sect. 2.4 that two parameters C and σ have to be suitably selected for good SVM modeling. To select image features and SVM parameters, the genetic algorithm (GA) is applied in this paper.

A GA is an optimization technique inspired by the process of evolution [48]. A GA uses a population of random solutions known as chromosomes to solve a problem. Each chromosome corresponds to an encoded possible solution to the problem. A reproduction-based mechanism is applied to the population to generate a new population. The population progresses through several generations until a suitable solution is reached or a predefined generation limit is reached. The initial population is evaluated and fitness values are calculated for each solution. The fitness value of a solution is a probability of the survival of that solution in the next generation. The GA implemented in this paper has seven different features.

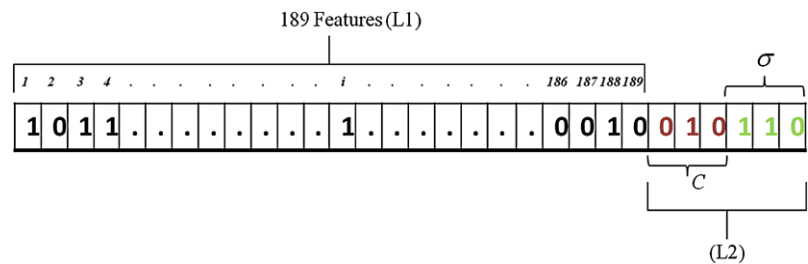
2.5.1 Representation

The chromosomes represented in this paper have two parts. The first part encodes the set of image features. The length of the first part ($L1$) is 189, which is the total number of extracted features from an image. A 1 at the i th position of the chromosome indicates the i th feature is selected in GA-based feature selection algorithm and 0 at the i th position indicates that the i th feature has not been selected. The fitness of an individual chromosome is determined by evaluating the SVM using a training data set whose patterns are represented using only the selected subset of image features. If an individual chromosome has n bits turned on (value 1), the corresponding SVM has n input nodes. The original calculated value of features as obtained from Sect. 2.3 of n selected features will be taken as input for SVM model development.

The second part encodes the values of the SVM parameters C and σ . Figure 4 represents a chromosomal representation of a single solution for the feature selection and SVM parameter selection. Suppose n number of features are selected from 189 number of extracted features, where $n < 189$, then only n bit values will be 1 and rest of the $(189 - n)$ bit values will be zero.

The value of C and σ are represented by three bits each. Therefore, the length of the second part of the chromosome ($L2$) is 6. Three bits for the parameter C only can code the numbers 0 to 7. If this representation is shifted by -3 and interpreted as powers of ten, then one gets a coding for the possible C values: 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000. The parameter σ of the Gaussian kernel is found by shifting the resulting number by -6 and raising the number that

Fig. 4 Schematic representation of a chromosome for this study



results as power of two. Thus, one gets the possible values $2^{-6}, 2^{-5}, 2^{-4}, \dots, 2$ for the parameter σ . It is noted that C and σ can take any real values, however, to search a wide range of values within a reasonable amount of time, these values are considered.

2.5.2 Initialization

Before initializing the chromosome representation for $L1$, the number of features to be included in the model should be decided. For example, if 20 features are to be selected from the available 189 features, the 20 random places in the $L1$ vector will be assigned a value of 1 and the remaining 169 number of places will be assigned a value of 0. The second part ($L2$) of the chromosome was uniformly initialized. The number of chromosomes in the population (population size, P) is selected as 50 in this paper.

2.5.3 Selection

A probabilistic selection is used based on the individual chromosomes' fitness such that the fittest individuals have a higher chance of being selected. In this paper, the normalized geometric ranking scheme $p_i = q^r(1 - q)^{r-1}$ is applied, where p_i represents the probability of the i th individual being selected, q is the probability of selecting the best individual, r is the rank of the individual.

2.5.4 Crossover

The crossover operation is performed in each generation to generate a better solution from available solutions. This is performed by interchanging genetic material of chromosomes in order to create individuals that can benefit from their parents' fitness. In this paper, a uniform crossover with probability rate 0.1 has been used.

2.5.5 Mutation

Mutation is the genetic operator responsible for maintaining diversity in the population. Mutation operates by randomly "flipping" bits of the chromosome, based on some probability. The mutation probability used in this paper is $1/p$, where p is the length of each of the two parts of the chromosomes.

2.5.6 Random immigrant

Random immigrant introduces some diverse solutions in the population which minimizes the risk of premature convergence [49]. Some individuals having low fitness value are deleted from the population and replaced by the same number of recently initialized random individuals. The number of individuals deleted and the number of new individuals initialized in each generation is 5 in this paper.

2.5.7 Fitness

To develop a good SVM model for classification with feature selection, one should concentrate on subsets of features that minimize an estimate of generalization error of the classifier. The fitness function should be chosen in such a way that it minimises the error of misclassification of unforeseen data. A validation data set is used where the fitness function is measured.

The Sensitivity and Specificity are measures for individual rock types. The sensitivity measures the proportion of rock type A being correctly classified as rock type A and the Specificity measures the proportion of the rock type $Non-A$ being classified as rock type $Non-A$. As an example, let the total validation data for SVM classification testing be N and the number of rock type A and Rock type $Non-A$ in the test data be denoted by NO and NW respectively. If truly classified, A and $Non-A$ are denoted by TO and TW , and false classified A and $Non-A$ are denoted by FO and FW , then the sensitivity and the specificity are presented by the following form:

$$Sensitivity_A = \frac{TO}{TO + FW} \quad (4)$$

$$Specificity_A = \frac{TW}{TW + FO}$$

The values of the specificity and the sensitivity are complementary. The purpose of any model is to maximize both these values. However, it has been experimentally and theoretically established that an increase in one value leads to the decreasing of the other [50]. We can take care of this

problem by introducing two indices for measuring the classification accuracy. These are presented by:

$$P_A = \frac{TO + TW}{N} \quad (5)$$

$$RI_A = \frac{|Sensitivity - Specificity|}{|Sensitivity + Specificity|}$$

where P measures the percentage of samples correctly classified and N is total number of segmented rock images. The term (P) measures the overall accuracy of the SVM classification system irrespective of the rock type A and $Non-A$. However, it is often the case that P can be maximized by increasing the value of the more dominant of the two values. This shortcoming can be improved by using the relationship index (RI). In RI , low value of the numerator and the high value of the denominator are most desirable because these help to maximize both the sensitivity and the specificity. Therefore, the fitness function used in this paper consists of these two terms (P and RI) in the form of a weighted sum for all classes:

$$fitness = \frac{1}{CL} \sum_{i=1}^{CL} (w_1 P_i + w_2 (1 - RI_i)) \quad (6)$$

where CL is number of classes, w_1 and w_2 are weights for P and RI , and $w_1 + w_2 = 1$. Generally, equal weights are assigned to w_1 and w_2 . In each iteration, the selection is made based on the fitness function value of the validation data set.

3 Case study

3.1 Description of mine and data collection

The study was carried out in a limestone mine in India. The area covered by the mine is more than 6 sq. km. Most of the area is covered by soil, except for outcrops of limestone. The mine has nine different lithotypes, namely Pink limestone (PPL), Greenish gray limestone (GGL), Dark gray limestone (DGL), Light gray limestone (LGL), Weathered limestone (WTH), Upper gray limestone (UGL), Shale, Clay, and Overburden soil [51]. Most of the high grade limestone is associated with Pink, Greenish and Upper gray limestone.

Typically, a blasted material of rock sample is heterogeneous with respect to fragment size. Depending on the blast design, the largest sizes could be thrown to the furthest from the blast or they could slump down directly next to the blast. There may be some kind of gravitational segregation, where the fines are covering the larger blocks or alternatively the fines may slip in behind the larger blocks, for example in quarries exposed to wind and rain. If the assumption is made that the exposed surface of the blasted material is representative, sampling can simply be a matter of photographing

the surface. Sampling could also be done during the material handling process, in the haulage trucks, in buckets of loader, or on conveyor belts. In this study, all the image acquisition was carried out in the laboratory with a simulated environment.

The representative samples were collected from the blasted muck of the case study mine while maintaining the proper sampling strategy. The stratified random sampling method was adopted for this study. In this scheme, the samples were collected from different strata which were classified according to the rock types present in the deposit. The rock samples from each stratum were collected randomly. It was also decided to capture an equal number of samples from each stratum. As the mine under investigation was in its initial stage of production, all the lithological units were not exposed at the time of sampling. Therefore, the samples were gathered from 6 rock types (UGL, Clay, DGL, PPL, GGL, WTH) exposed to the working faces. Altogether 120 samples, twenty from each lithology, were collected from the case study mine. The samples weighed approximately 5 kgs and the size range varied from 2 to 8 cm.

3.2 Image acquisition and segmentation

For this study, a digital camera (CX-7300, KODAK; Japan) was used. The camera has an aspect ratio of 4:3. The images collected of the samples were taken in the laboratory set up as described in Sect. 2.1. Ten successive images for each sample were taken by changing the placement and the orientation of the rock samples inside the experimental box. In total, 1200 images were generated collectively from the rock samples. The images taken by the digital camera were then transferred to the personal computer. The size of the imported jpeg images ranged from 520 to 550 KB. The images have a spatial resolution of 0.15 mm/pixel in both horizontal and vertical directions. The images acquired were 2080×1544 pixels in size. Some of the images which were taken during the experiment are shown in Fig. 5.

The generated images were then segmented using the segmentation techniques described in Sect. 2.2. Figure 6 shows the resultant images at each segmentation stage from an example image. Figure 6(a) is an example gray scale image and Fig. 6(b) is the corresponding binary image generated after fixing the thresholding value via Otsu's technique [29]. Image complement operations (Fig. 6(c)) were performed on the threshold image before distance transformation because watershed transformations detect the low intensity part in an image. The distance transformation (Fig. 6(d)) of a binary image is the distance from every pixel to the nearest non-zero valued pixel. After the distance transformation, watershed transformation (Fig. 6(e)) was performed in negative of the distance transformation matrix. It was seen from the resulting image that the large



Fig. 5 Images of different rock types of limestone minerals

size samples are nicely segmented; however the segmentation of small size samples are not satisfactory. In this work we have only dealt with those rocks which have more than 250 pixels. Therefore, most of the smaller-sized rock sample images are not considered for our study.

The segmented images were then processed for identifying and labeling the individual rock present in the segmented parts using regional labeling algorithm. After selecting 250 pixels as the threshold value, only rock more than 250 pixels are identified and labelled for further analysis. Altogether, 5267 distinct rock objects were identified from 1200 images. The features were then extracted from each of the individual rock objects. A total number of 189 features were extracted from each segmented rock.

3.3 Rock type classification using support vector machine

The extracted features from segmented rocks and their corresponding classes are then used for developing the support vector machine model for classification. The OVA multi-class classification was performed as described in Sect. 2.4.2. Before performing SVM modeling, feature normalization and data subdivision were performed.

Before SVM training, the inputs and targets were normalized so that they always fall within a specified range. In this paper, inputs were normalized with mean and standard deviation of the data. It normalizes the inputs so that they will have zero mean and unit standard deviation. The output parameters of the model will be either +1 or -1 based on presence or absence of a particular rock class.

It was mentioned in Sect. 2.5.7 that the fitness function is calculated based on the performance of the SVM model with the validation data set. The use of validation data helps to improve the generalisability of the model. Moreover, a separate data set is required where the developed model will be

tested. To obtain the validation and test data set, the available data is divided into three subsets. The first subset is the training set, which is used for SVM training. The second subset is the validation data set. The fitness values of the validation data set after iteration in GA learning are used for selection process. The test data set is used for testing the developed algorithm. For a reasonable model development, these three data sets should have similar statistical properties. In order to analyze the statistical similarity of the three data sets, an ANOVA test was performed. Out of the 5267 available data points, 2635 (50 %) datum were used for training, 1316 (25 %) datum were used for validation, and the remaining 1316 (25 %) datum were used for testing purposes. The ANOVA F test result showed that they belong to the same population.

For image feature selection and hyper-parameter estimation for the SVM model, genetic algorithms randomly generate an initial population of features from 189 available features. The cross over rate, mutation rate, and the generation gap are kept constant throughout the experiment at 0.9, 0.005, and 0.95, respectively. The one point cross over is applied in this paper. The evolution process is stopped when the best fitness remains unchanged for 50 generations. The best features are selected after a complete run of GA and SVM for a given number of features. For choosing the optimum number of features, the same algorithm was executed by incrementally changing the numbers of randomly selected features until there is significant improvement in the classification accuracy. However, no improvement in the fitness value (classification accuracy) of the validation data set was observed after inclusion of 50 randomly selected features; therefore, the algorithm stopped beyond feature number 50 to save computational time. It is noted that random selection of features incrementally using GA is a computationally expensive job; however, it improves the performance of

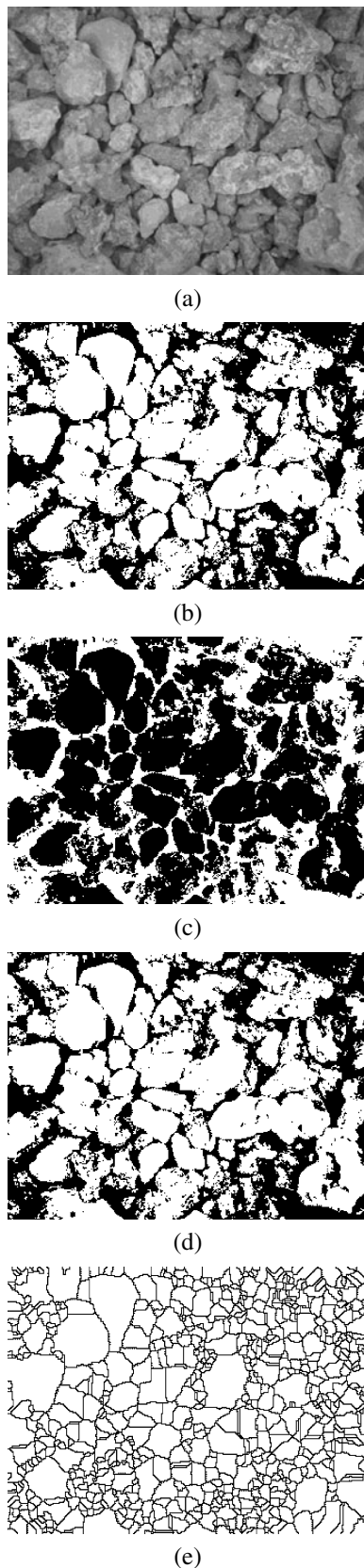


Fig. 6 Images of different stage involved in segmentation technique (a) Gray image (b) Threshold image (c) Complement image (d) Distance transformation (e) Watershed segmentation

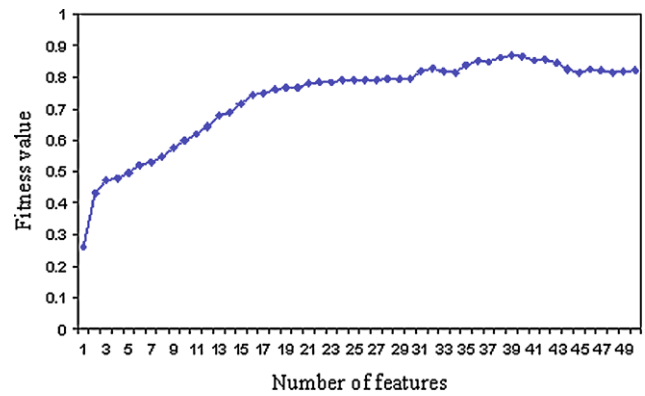


Fig. 7 Fitness value of different number of selected features

the classification results. Therefore, a trade-off analysis between computational time and accuracy is a pre-requisite to applying before applying this algorithm in the case problem; however, this is beyond the scope of the paper. Figure 7 presents the fitness value of the validation data set for different numbers of selected features. It is observed from the figure that the fitness value of the validation data increases up to a certain number of selected image features. The results revealed that with 40 selected features, the maximum fitness value is achieved, the fitness value either remains constant or decreases. Out of 40 selected features, it was observed that 20 color features, 6 morphological features and 14 textural features were selected. It becomes clear from this analysis that the fusion of color, textural and morphological features gives better results than only one type of feature. This result supports that the employment of features from all three domains give better results than features from a single source.

3.4 Model performance

The developed SVM model was then run using the testing data set. Out of 1316 testing data points, UGL, and clay, have 220 data points each and DGL, PPL, GGL, and WTH have 219 each. Table 2 shows the confusion matrix of tested results. The SVM could classify 99 % UGL, 98.6 % clay, 97 % DGL, 94.5 % PPL, 89.5 % GGL and 98.6 % WTH in correct classes. The confusion matrix shows that for the UGL lithology, the misclassification error of 1 % is almost the same as Clay. For UGL, only one sample is misclassified as DGL and one sample is misclassified as GGL. In case of clay, 1 % was misclassified as the WTH. On the other hand, 2 % of DGL was misclassified as the GGL and 1 % was misclassified as the WTH. It was also observed from the results that maximum misclassification occurred with the GGL lithology where 9 % was misclassified as the DGL and 1.5 % was misclassified as the PPL. From the confusion matrices, it was noticed that many GGL samples are misclassified to DGL and vice versa, which is reasonable considering their visual similarity in photographs. Four samples of PPL

Table 2 Confusion matrix of rock-type classification

	UGL	Clay	DGL	PPL	GGL	WTH
UGL	218	0	1	0	1	0
Clay	0	217	0	0	0	3
DGL	0	0	212	0	6	1
PPL	0	0	8	207	4	0
GGL	0	0	19	4	196	0
WTH	0	2	0	0	1	216

Table 3 Percentage of error in classification

Rock-type	Percentage error
UGL	1
Clay	1.4
DGL	3.0
PPL	5.5
GGL	10.5
WTH	1.4

Table 4 Specificity and sensitivity of all rock types for test data set

	UGL	Clay	DGL	PPL	GGL	WTH
Sensitivity	0.99	0.986	0.968	0.945	0.895	0.986
Specificity	1	0.998	0.974	0.996	0.989	0.996

are misclassified to GGL and vice versa. To know the insight of the misclassification of those 8 samples, detailed investigations of the selected input features were carried out. A paired sample *t*-test was performed for all selected features of those two groups of samples. Out of 40 selected features, only 3 features show that the mean of these two groups are significantly different at 95 % confidence level. A non-significant difference of 37 features within these 8 samples may be the reason of their misclassification. A total number of 1266 test data points were correctly classified in their respective classes out of 1316 data. Therefore, overall accuracy of the developed SVM model is 96.2 %. The results of percentage of misclassification of individual lithology are also presented in Table 3.

To verify, the classification accuracy, sensitivity and specificity are calculated. Table 4 shows the value of these two parameters for all six rock types. It is noted that specificity and sensitivity are calculated based on binary classification. To calculate these parameters for a multi-class problem, the binary transformation must be performed. For example, in case of UGL, two classes were considered: UGL and non-UGL. It can be seen from the confusion matrix that out of 220 UGL samples, 218 are truly classified as UGL and 2 are classified as non-UGL (1 DGL, 1 GGL), and out of 1086 non-UGL samples, all are truly classified as non-UGL.

None of the non-UGL is misclassified as UGL. Therefore, the value of the sensitivity is 0.99 (218/220) and the specificity is 1 (1086/1086). It is clearly observed from Table 4 that performance of classification is best for UGL and worst for GGL.

4 Comparative study

The performance of the Support Vector Machines and a Neural Network solution were compared for successful classification of the six rock types. An Artificial Neural Network (ANN), which was a feed forward ANN, consisted of one input layer fed with the set of input variables, one hidden layer and one output layer of one neuron [52–54]. The output layer has six neurons; the maximum activated neuron gives the winner class. The tan-sigmoid and log-sigmoid are used as activation functions in the input and output layer respectively, while the ANN can be trained using a method such as back-propagation [46]. The neural network used one hidden layer with a variable number of neurons, ranging from 2 to 50 and was trained using back-propagation with adaptive learning rate and momentum. The same training, validation and testing data were used for the neural network modelling as used in support vector machine modelling. The optimum number of hidden neurons was selected based on the validation data error.

The SVM and ANN classifiers were trained using 189 and 40 features. The numbers of hidden neurons for 189 features and 40 features are 22 and 14, respectively. In Table 5, the sensitivity and specificity are presented for all four models. From Table 5 it can be inferred that the ANN model, when using a reduced feature set, does not perform as well as the SVM model using the same features. This can be expected since the features were selected using an SVM but it serves a reference purpose to compare performance on a common feature set. It is also observed from the table that the value of sensitivity and specificity are increased in both SVM and ANN when selected features are used for model development. Therefore, it can be concluded that the proper subset selection not only reduces the computational time but also improves the performance of the results. It was also observed from the table that the performance of the SVM and ANN models are almost identical when all features were considered for model development. This may be due to the presence of unwanted features, which influences the performance of both the models.

To compare the proposed results with different dimensional reduction algorithm, a comparative study is performed with GA-based feature selection of the ANN model (GA-NN) and dimensional reduction using principal component analysis (PCA) of the ANN model (PCA-NN) applied in [4]. The same GA-based feature selection algorithm

Table 5 Specificity and sensitivity of all rock types for four different models on test data set

		UGL	Clay	DGL	PPL	GGL	WTH
Sensitivity	SVM-40	0.990	0.986	0.968	0.945	0.895	0.986
	SVM-189	0.910	0.935	0.922	0.896	0.862	0.925
	ANN-40	0.930	0.961	0.96	0.938	0.869	0.950
	ANN-189	0.899	0.911	0.908	0.860	0.860	0.917
Specificity	SVM-40	1	0.998	0.974	0.996	0.989	0.996
	SVM-189	0.927	0.965	0.932	0.954	0.944	0.897
	ANN-40	0.980	0.976	0.952	0.968	0.951	0.928
	ANN-189	0.92	0.96	0.93	0.949	0.940	0.887

Table 6 Specificity and sensitivity of all rock types for test data set using GA-NN and PCA-NN [5] models

		UGL	Clay	DGL	PPL	GGL	WTH
Sensitivity	GA-NN	0.960	0.965	0.966	0.941	0.871	0.960
	PCA-NN	0.952	0.951	0.959	0.939	0.88	0.949
Specificity	GA-NN	0.988	0.980	0.955	0.976	0.968	0.962
	PCA-NN	0.978	0.980	0.948	0.963	0.952	0.959

was applied to feature selection of the ANN model. To apply GA for feature selection of the ANN model, the second part of the chromosome encodes the value of hidden neuron number instead of parameters C and σ . Therefore, unlike SVM models, only a single bit is used for representing the hidden node size. The optimum number of features and optimum hidden node number are 45 and 13, respectively. To reduce dimension of neural network inputs, PCA was applied and 40 first principal components (PCs) are selected which cumulatively captured 85 % of the total data variance. The specificity and the sensitivity of both GA-NN and PCA-NN models are presented in Table 6. The results revealed that both of the ANN models performed better than the ANN-40 model. The reason that the GA-NN performed better than the ANN-40 is that the GA-based feature selection algorithm is selecting an optimum combination of features for ANN model which helps to improve the performance of the ANN model. Similarly, PCA-NN reduces the dimension by projecting the data linearly and capturing most of the data variance by few PCs; thus improving the classification accuracy compared to the ANN-40 model. However, results also revealed that the performance of SVM-40 model is comparatively better than both GA-NN and PCA-NN models. This may be due to the learning mechanism of these two algorithms. It is noted that the SVM solves the problem by quadratic optimization that provides an optimum solution if parameters C and σ are selected properly; however, the ANN solves the problem by gradient descent type non-linear optimization algorithm, which has probability to trap at local minima point even if run with an optimum hidden node size.

5 Summary and conclusions

A vision-based rock type classification model using support vector machine is presented by selecting important image features. The images were captured, segmented, and features were extracted from the segmented images. The hybrid segmentation technique is performed by automatic thresholding and watershed segmentation techniques. For conducting this study, samples were collected from a limestone mine. A total number of 189 features were extracted, including 49 textural, 112 color, and 28 morphological features from segmented image.

The effectiveness of the above image-based rock type classification was tested using a testing data set. The results revealed that the rock type classification error using proposed technique is on an average confined within 3.8 %. The misclassification results showed that the UGL, Clay and WTH produced minimum error of misclassification (1 %, 1.4 %, and 1.4 % respectively); whereas, the GGL produced the maximum error of misclassification (10.5 %). This study result indicated the effectiveness of rock image processing technique in rock type classification of the limestone. A comparative study with neural network model reveals that the developed SVM model performed better than the ANN model.

The feature selection approach showed that it is possible to reduce computation time significantly without affecting overall accuracy. As feature characterization and classification techniques advance, proposed overall strategy is expected to provide better results.

The main limitation of this study is that it was conducted at a laboratory scale after collecting data from mine. No field study was conducted for real life mining applications. However, it is assumed that if the same type of image acquisition set up could be developed for actual field implementation, the algorithm will provide equally good results. Moreover, the proposed model is a case-specific model, which is tested by collecting samples from the limestone mine. Therefore, the developed model cannot be directly applied to other types of mineral deposit. However, the methodology is still

valid for other mineral deposits as well. Before applying the model to other deposit, the SVM and GA-based method has to be re-run to select the optimum number of features of the case specific applications.

References

- Donskoi E, Suthers S, Fradd S, Young J, Campbell J, Raynlyn T, Clout J (2007) Utilization of optical image analysis and automatic texture classification for iron ore particle characterization. *Miner Eng* 20(5):461–471
- Perez CA, Estévez PA, Vera PA, Castillo LE, Aravena CM, Schulz DA, Medina LE (2011) Ore grade estimation by feature selection and voting using boundary detection in digital image analysis. *Int J Miner Process* 101(1–4):28–36
- Kulatilake PHSW, Hudaverdi T, Wu Q (2012) New prediction models for mean particle size in rock blast fragmentation. *Geotech Geolog Eng* 30(3):665–684
- Singh N, Singh TN, Tiwary A, Sarkar KM (2010) Textural identification of basaltic rock mass using image processing and neural network. *Comput Geosci* 14(2):301–310
- Tessier J, Duchesne C, Bartolacci G (2007) A machine vision approach to on-line estimation of run-of-mine ore composition on conveyor belts. *Miner Eng* 20(12):1129–1144
- Remy N, Boucher A, Wu J (2009) Applied geostatistics with SGeMs: a users's guide. Cambridge University Press, Cambridge
- Perez C, Casali A, Gonzalez G, Vallebuona G, Vargas R (1999) Lithological composition sensor based on digital image feature extraction, genetic selection of features and neural classification. In: International conference on information intelligence and systems (ICIIS'99), p 236
- Casali G, Vallebuona C, Pérez G, Vargas R (2000) Lithological composition and ore grindability sensors, based on image analysis. In: Proc. of the XXI IMPC, Rome, p A1(9-16)
- Singh V, Rao SM (2005) Application of image processing and radial basis neural network techniques for ore sorting and ore classification. *Miner Eng* 18:1412–1420
- Lepisto L, Kunttu I, Visa A (2005) Rock image classification using color features in Gabor space. *J Electron Imaging* 14(4):28–36
- Hunter GC, McDermott C, Miles NJ, Singh A, Scoble MJ (1990) A review of image analysis techniques for measuring blast fragmentation. *Min Sci Technol* 11:19–36
- Wang W (2008) Rock particle image segmentation and systems in pattern recognition. In: Yin P-Y (ed) *Pattern recognition techniques, technology and applications*. InTech, Rijeka, pp 197–226
- Salinas RA, Raff U, Farfan C (2005) Automated estimation of rock fragment distributions using computer vision and its application in mining. *IEE Proc, Vis Image Signal Process* 152(1):1–8
- Petersen K, Aldrich C, Vandeventer JSJ (1998) Analysis of ore particles based on textural pattern recognition. *Miner Eng* 11:959–977
- Kachanubal T, Udomhunsakul S (2008) Rock textures classification based on textural and spectral features. *Int J Comput Intell* 4:240–246
- Lin CL, Yen YK, Miller JD (1993) Evaluation of a pc image-based on-line coarse particle size analyzer. In: *Emerging computer techniques for the minerals industry, society for mining, metallurgy, and exploration, inc*, Littleton, Colorado, pp 201–210
- Steppe JM, Bauer KW, Rogers SK (1996) Integrated feature and architecture selection. *IEEE Trans Neural Netw* 7:1007–1014
- Duda RO, Hart PE (2000) *Pattern classification*. Wiley-Interscience, New York
- Chatterjee S, Bhattacharjee A, Samanta B, Pal SK (2008) Rock-type classification of an iron ore deposit using digital image analysis technique. *Int J Min Miner Eng* 1:22–46
- Chatterjee S, Bhattacharjee A, Samanta B, Pal SK (2009) Image-based quality monitoring system of limestone ore grades. *Comput Ind* 61:391–408
- Chen X (2003) An improved branch and bound algorithm for feature selection. *Pattern Recognit Lett* 24(12):1925–1933
- Somol P, Pudil P, Kittler J (2004) Fast branch & bound algorithms for optimal feature selection. *IEEE Trans Pattern Anal Mach Intell* 26(7):900–912
- Hong J, Cho S (2006) Efficient huge-scale feature selection with speciated genetic algorithm. *Pattern Recognit Lett* 27:143–150
- Osma-Ruiz V, Godino-Llorente JI, Sáenz-Lechón N, Gómez-Vilda P (2007) An improved watershed algorithm based on efficient computation of shortest paths. *Pattern Recognit* 40(3):1078–1090
- Gonzalez RC, Woods RE, Eddins SL (2004) Segmentation using the watershed transform. In: *Digital image processing using MATLAB*. Pearson Prentice-Hall, Upper Saddle River
- Ng HF (2006) Automatic thresholding for defect detection. *Pattern Recognit Lett* 27(14):1644–1649
- Solomon C, Breckon T (2011) *Fundamentals of digital image processing: a practical approach with examples in Matlab*, E-book. Wiley, New York
- Cuisenaire O (2006) Locally adaptable mathematical morphology using distance transformations. *Pattern Recognit* 39(3):405–416
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9:62–66
- Cuevas E, Sención F, Zaldivar D, Pérez-Cisneros M, Sossa H (2012) A multi-threshold segmentation approach based on artificial bee colony optimization. *Appl Intell* 37(3):321–336
- Saha S, Bandyopadhyay S (2011) Automatic MR brain image segmentation using a multiseed based multiobjective clustering approach. *Appl Intell* 35(3):411–427
- Haralick RM, Shapiro LG (1992) *Computer and robot vision 1 & 2*. Addison-Wesley, Reading
- Chatterjee S, Bhattacharjee A (2011) Genetic Algorithms for feature selection of image analysis-based quality monitoring model: an application to an iron mine. *Eng Appl Artif Intell* 24(5):786–795
- Huang YL, Chen DR, Jiang YR, Kuo SJ, Wu HK, Moon WK (2008) Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound. *Ultrasound Obstet Gynecol* 32:565–572
- Lefkaditis D, Tsigiotis G (2009) Morphological feature selection and neural classification for electronic components. *J Eng Sci Technol Rev* 2(1):151–156
- McNitt-Gray MF, Huang HK, Sayre JW (1995) Feature selection in the pattern classification problem of digital chest radiograph segmentation. *IEEE Trans Med Imaging* 14(3):537–547
- Zucker SW, Terzopoulos D (1980) Finding structure in co-occurrence matrices for texture analysis. *Comput Graph Image Process* 12:286–308
- Haralick RM (1979) Statistical and structural approaches to texture. *Proc IEEE* 67(5):786–804
- Galloway MM (1975) Texture analysis using gray level run lengths. *Comput Graph Image Process* 4:172–179
- Chu A, Sehgal CM, Greenleaf JF (1990) Use of gray value distribution of run lengths for texture analysis. *Pattern Recognit Lett* 11(6):415–420
- Dasarathy BR, Holder EB (1991) Image characterizations based on joint gray-level run-length distributions. *Pattern Recognit Lett* 12:497–502
- Vapnik VN (1995) *The nature of statistical learning theory*. Springer, New York
- Dioşan L, Rogozan A, Pecuchet JP (2012) Improving classification performance of Support Vector Machine by genetically optimizing kernel shape and hyper-parameters. *Appl Intell* 36(2):280–294

44. Lee LH, Wan CH, Rajkumar R, Isa D (2012) An enhanced Support Vector Machine classification framework by using Euclidean distance function for text document categorization. *Appl Intell* 37(1):80–99
45. Rifkin R, Mukherjee S, Tamayo P, Ramaswamy S, Yeang CH, Angelo M, Reich M, Poggio T, Lander ES, Golub TR, Mesirov JP (2003) An analytical method for multi-class molecular cancer classification. *SIAM Rev* 45:706–723
46. Li Y, Zeng X (2010) Sequential multi-criteria feature selection algorithm based on agent genetic algorithm. *Appl Intell* 33(2):117–131
47. Vinh LT, Lee S, Park YT, d'Auriol BJ (2012) A novel feature selection method based on normalized mutual information. *Appl Intell* 37(1):100–120
48. Goldberg DE (1989) *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, Reading
49. Congdon CB (1995) A comparison of genetic algorithm and other machine learning systems on a complex classification task from common disease research. PhD thesis, Computer Science and Engineering Division, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA
50. Ranawana R, Palade V (2005) MVGen—ensemble learning for MCS majority voting with a genetic algorithm. Internal Report, Oxford University Computing Laboratory
51. Chatterjee S (2006) Geostatistical and image-based quality control models for indian mineral industry. Unpublished PhD Thesis dissertation, IIT Kharagpur, India
52. Haykin S (1999) *Neural networks: a comprehensive foundation*. Prentice Hall, New York
53. Kim K, Ahn H (2012) Simultaneous optimization of artificial neural networks for financial forecasting. *Appl Intell* 36(4):887–898
54. Nekoukar V, Beheshti MTH (2010) A local linear radial basis function neural network for financial time-series forecasting. *Appl Intell* 33(3):352–356