Application of percentile color intensities of borehole images for automatic fluorite grade assessment

Enming Li [a], Pablo Segarra [a], José A. Sanchidrián [a,*], Santiago Gómez [a], Alberto Fernández [a], Rafael Navarro [b,c], Maurizio Bernardini [a]

[a] Universidad Politécnica de Madrid – ETSI Minas y Energía, Ríos Rosas 21, 28003 Madrid, Spain
[b] Minera de Órgiva S.L. Mina Carriles. Polígono 13, Parcela 1, 18400 Órgiva, Spain
[c] Universidad de Salamanca – GIR Charrock, Pza. de los Caídos, s.n., 37008 Salamanca, Spain

ABSTRACT

This paper describes a practical application of support vector machine (SVM) for ore grading in an underground fluorite mine. It covers all aspects from the inception of the experiments, data collection, input preparation, model description and results. Forty-eight drilling chips samples are collected while drilling six pseudo-horizontal boreholes at depth intervals of half a meter and their chemical composition determined through X-Ray fluorescence; the response of the drill rig is used to accurately define the depth of each sample along the blasthole. Images of the blasthole walls are collected with an optical televiewer with white and ultraviolet (UV) illumination. The color information of the images is characterized by the cumulative distribution of pixel color intensities of red, green and blue, used as inputs. A well-known metaheuristic algorithm is used to calibrate the SVM hyperparameters. Repeated k-fold cross validation is applied to increase the prediction performance due to the small-size of the dataset. An outlier inspection is made resulting in improved performance. The combination of pixel intensities from white and UV light scans leads to the best prediction of fluorite content (average $R^2 = 0.83$ and RMSE = 3.32 %), while intensities from only white light procures the best classification results (average classification accuracies from 0.77 to 1). These metrics support the utility of the proposed methodology for reducing the amount of lab analysis in ore grade control.

1. Introduction

The accuracy of reserves evaluation and the distribution of ore grades are key aspects in mining economics, planning and design. The prediction and evaluation of mineral grades plays a crucial role in the mining industry in its struggle to stay competitive under volatile prices, variable chemical and mineralogical composition, and declining ore grades. Swift grade determination of the ore that is being mined is instrumental to mining efficiency, hence methods for providing information on ore grade in an inexpensive and efficient way are of great interest for the mining industry. The harsh environment of underground works often limits the applicability of sophisticated and expensive analysis equipment, and the on-site implementation of complex analytical methods.

Two types of direct methods for determining the ore grade are core drilling and drill cuttings analysis. Core drilling is costly resulting in a limited dataset from which an ore grade model is inferred (Starr and Ingleton, 1992). On the contrary, the drill cuttings analysis from ordinary drilling for blasting or roof support allows for a dense sampling net; however, this method often has a limited accuracy as only average values per borehole are determined (Neff, 1987). When the number of sampled holes is large, chemical analysis of drill cuttings can also be very demanding in terms of labor and assaying costs.
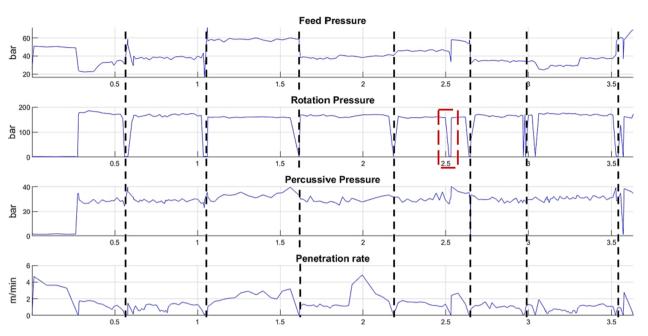
Minerals with different grades and components may present different colors and other optical properties. Some researchers have analyzed the optical properties of minerals at microscopic scale (Donskoi et al., 2013; Donskoi et al., 2015; Lane et al., 2008). Tanaka et al. (2019) reported a method for the recognition of acidic alteration zones in a deposit by distinguishing the intrinsic absorption peaks in the short-wavelength infrared region from various alteration minerals. Donskoi et al. (2007) combined a series of image analysis and mineral measurement techniques to distinguish the minerals with similar composition and texture. Berrezueta et al. (2016) proposed a new method by combining multi-spectral and color image analysis from microscopic observations to

Fig. 1. Drill log of borehole H24. Drilling stops are marked by black dashed lines; a potential rock mass discontinuity is highlighted by a red rectangle.

identify and quantify parameters related with geometallurgical performance such as ore grade, grain size and mineral liberation. Okada et al. (2020) proposed a quick and non-destructive technique to identify mineral types before mineral processing by utilizing RGB (red, blue and green) pixels information, hyperspectral imaging and deep learning techniques. Liu et al. (2019) show that deep learning, transfer learning, clustering algorithms and supervised learning techniques provide a more effective mineral recognition than traditional ones.

Despite rock recognition from spectral information being promising for ore grade assessment, the procurement of these data needs advanced equipment and also benign, friendly working environment. Inexpensive spectrometers are sometimes prone to errors in wavelength shift requiring frequent calibration. In some cases, optimum acquisition of images or segmentation techniques must be applied to guarantee an accurate recognition or classification. In view of this, color parameters of ore images seem to be a good approach to characterize the mineral characteristics due to its accessibility, low cost and convenience (Marschallinger, 1997; Thompson et al., 2001). For instance, Ramil et al. (2018) proposed an automatic identification system of in situ granite minerals based on artificial neural networks and RGB values of pixels of images of small-scale slabs. Li et al. (2017) developed a novel classification method of sandstone microscopic images named Festra based on gray levels. Baykan and Yılmaz (2010) identify minerals with the aid of artificial neural networks using color information such as RGB, hue, saturation and lightness (HSL) of thin sections from a rotating polarizing microscope equipped with a digital camera. Desta and Buxton (2017) acquired in-situ georeferenced RGB images from the mine faces to interpret the distribution of minerals. Unsupervised learning techniques allowed to distinguish five mineral types with an accuracy of nearly 80 %.

In summary, it is apparent that the color properties of mineral images have a great potential to provide reliable information for evaluating mineral grades or for mineral recognition. However, in order to apply this approach for ore grade control in mine planning and production quality assessment, it must be shown that the complex relationship between color parameters and mineral grades generally found from microscopic images at lab scale still applies to macro photography of rock outcrops taken in a production environment.

Recently, machine learning (ML) techniques have been successfully applied to address mineral grade prediction (Dumakor-Dupey and Arya, 2021; Jafrasteh and Fathianpour, 2017; Jooshaki et al., 2021; Kaplan and Topal, 2020; Mery and Marcotte, 2022; Sun et al., 2019). As an example, Patel et al. (2019) apply support vector regression (SVR) of color intensities of images taken over a lab scale conveyor belt to monitor the quality of iron ore; Zhang et al. (2018) use back propagation artificial neural network to offline assess phosphate grade of flotation concentrate samples. RGB, or other color features, seem to be good alternatives to identify mineral grades. Perez et al. (2011) employed principal component analysis to RGB representation to extract color features and combined with texture features from five different rock samples including massive sulfide, disseminated sulfide, "net textured", gabbro and peridotite to recognize the composition. Chatterjee et al. (2010) utilized 189 features extracted from segmented images of a limestone mine and a neural network model to identify the grade attributes of limestone (CaO, Al_2O_3, Fe_2O_3 and SiO_2). Among these 189 features, 112 color features were involved and 42 features were gray level moments.

In the mining site considered in this work, the assaying technique employed is X-Ray Fluorescence (XRF) of pellets prepared from drilling chips. As this procedure is expensive, time consuming and provides the results in a delayed mode, this study proposes an alternative, novel method for estimating the ore grade from images of borehole logs, using the relations between ore grades and RGB pixel intensities of in-borehole images.

2. Data collection and description

The tests were carried out in the Lújar underground mine located in Órgiva (Granada province, Spain). A fluor-lead deposit composed by fluorite, galena and dolomite as gangue is mined. The host rock is mainly dark massive dolomicrite and limestone in which fluorite occurs as dark and white-purple crystals that may develop zebra patterns in some cases. The fact that the ore appears as fault-related veins or as irregular strata bound bodies with typical grades in the order of 15 % in fluorite complicates the in-situ ore recognition (Amor and Navarro, 2016; Ilin et al., 2019).
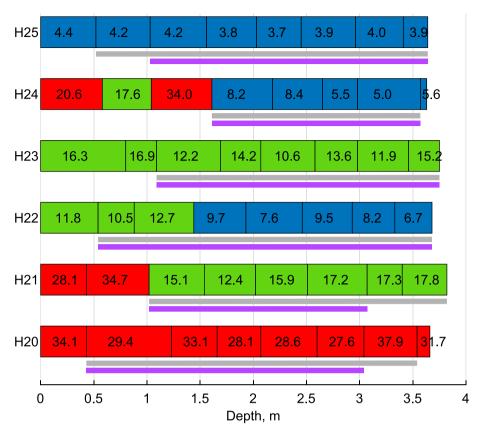
Fig. 2. Ore grades of drilling chip samples. The quantity in each section is the percentage of fluorite content and the color indicates the rock classification: blue for waste, green for low grade ore and red for medium grade ore. The lengths scanned are indicated by grey (white light) and violet (UV light) lines.

Fig. 3. Bottom part of the logging tool (left) and final stages of borehole surveying with the forward centralizer and the glass tube inside the hole (right).

Six pseudo-horizontal boreholes were drilled in the same mine area by an Atlas Copco 282 jumbo equipped with a measurement-while-drilling (MWD) system. The holes had an approximate length of 3.5 m, a diameter of 102 mm and an upward inclination of 5°. This allowed their cleaning by the injection of water in order to improve the quality of the in-hole images. An endoscope inspection was carried out to verify the wall cleanliness and make sure that no faults were crossing the holes. A PVC pipe was used to push the logging tool, with the wireline in its axis, inside the hole. The tool was then pulled back to surface by means of constant-velocity winch.

2.1. Drill chips assaying

Drilling chips were collected by means of a tray placed below the borehole collar. Drilling was stopped at approximate intervals of 0.5 m to collect the detritus and place a new tray. Eight samples were collected for every hole making up 48 samples in total. Drops in percussive pressure and rotation pressure recorded in the drilling logs were used to identify the drilling stops and obtain the corresponding initial and final depths of each sample, see the dashed black lines in the drilling records shown as an example in Fig. 1. The rotation pressure is sensitive to other effects related to the characteristics of the rock mass, such as the presence of structural discontinuities, as is the case of the red dashed rectangle in Fig. 1, but these drops are easily distinguished from the drill stop ones.

The drilling chips from each sample, approximately 1.3 kg, were quartered to 1/8 of the initial mass with a sample splitter with eight slots of 40 mm of aperture. This material was dried at 90 °C during 24 h, and ground in a vibratory disc mill (Restch RM100) to a size below 80 μm. After this, the sample was quartered to 1/16 to obtain a 10 g sample from which powder pellets were prepared. The pellets were analyzed in an X-Ray Fluorescence (XRF) Thermo Scientific ARL OPTIM'X WDXRF 50 kV analyzer, composed by Rhodium anode, crystals LiF200, InSb and AX06, and standard patterns of Thermo Fisher Sci; the software Oxsas 2.2 of Thermo Fisher Sci was used. The amount of compounds, mainly CaF_2, $CaCO_3$, $CaMg(CO_3)_2$, SiO_2, Fe_2O_3 and Al_2O_3, is obtained through stoichiometric balance of the composition provided by the XRF analyzer. Fig. 2 shows for each borehole, the initial and final depth of each drilling chips sample and the corresponding fluorite percentage; the composition of the rock in each of these sections is assumed to be uniform and it is classified as function of the fluorite content as waste (W, $CaF_2 < 10$ %, blue in Fig. 2), low grade ore (LG, $10 \leq CaF_2 < 20$ %, green) and medium grade ore (MG, $20 \leq CaF_2 < 45$ %, red).
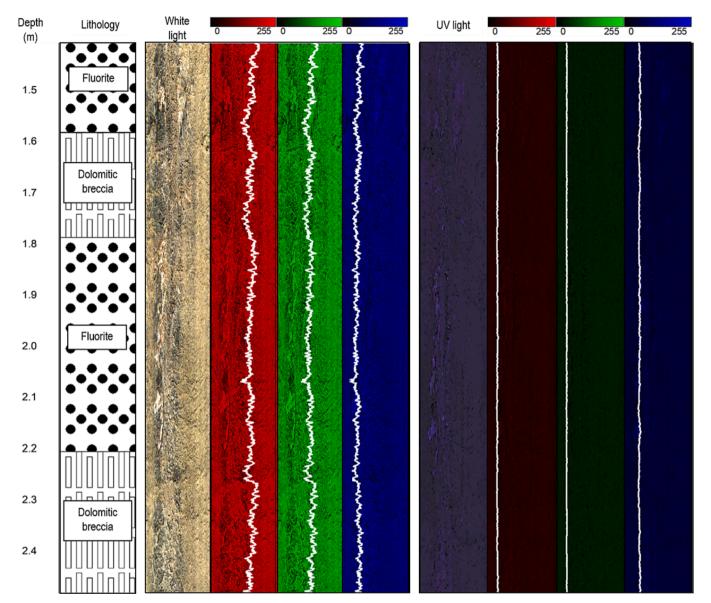
Fig. 4. Section of borehole H20. From left to right: lithology, televiewer processed image and RGB logs with white and UV illumination.

2.2. Borehole logging

Boreholes were logged with an optical televiewer manufactured by ALT composed by a QL40 OBI-2G logging tool of 1.5 m length, a data acquisition system, a mini-winch that pulls the logging tool at constant velocity and a computer to set-up the tool, display, and record the images of the borehole walls. The logging tool has 3-axis accelerometers and magnetometers in its central part to survey the borehole path and a digital image sensor at the bottom, with an active pixel array of 1.2 Mpx and fisheye matching optics (see left image in Fig. 3). It incorporates two LED series for lighting the internal walls: one emits white light and the other UV light. The latter has a wavelength range 340 to 400 nm with the emission peak at 365 nm. Each borehole was logged first with white light to record the natural colors of the rocks and second with ultraviolet light to outline the fluorescence of the main rock types.

The logging tool was centered with respect to the borehole axis with two centralizers mounted at the top (rear) and bottom (forward) parts of the probe; the rear one can be seen outside the borehole in Fig. 3. It was pushed with two plastic rigid pipes until the bottom of the hole though this was not always reached when high resistance was encountered to avoid damage to the optical system. From the end position, the mini-winch pulled the probe outwards while the borehole wall was scanned with an axial and circumferential resolution of 0.36 and 0.33 mm/px, respectively. Since the optical sensor is positioned at the end of the probe it was necessary to manually sustain the probe until the sensor was near the collar (see Fig. 3). Despite this, some 0.5 to 1.5 m of the borehole length in the collar section, depending on the scan, could not be scanned, see Fig. 2. The sections of the borehole scanned with white and UV lights considered for the analysis are coincident with the initial and final depths of the drilling chips samples. The actual lengths covered are shown with grey and violet bars below the fluorite compositions for each borehole in Fig. 2. Borehole sections associated to each drilling sample that are not fully scanned with the televiewer from the initial and final depths of each sample are discarded, as it is unknown whether the scanned part of the section is representative of the chemical composition of the actual drilling chips sample. This reduces the number of sections that can be correlated with rock images at the corresponding depths to 36 for scans with white light and 32 for those with UV light.

Despite that the size of the resulting database is relatively small to develop a model that could be generalized to other geological conditions or operations, it serves the purpose of validating the potential of the methodology proposed, in which the percentile color intensities of
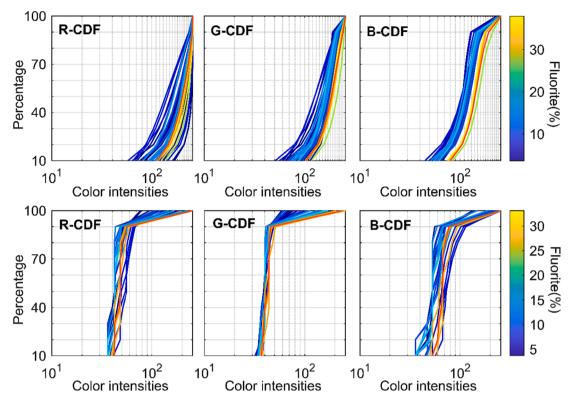
Fig. 5. Cumulative distribution functions of color intensities for white (top graphs) and UV light (bottom graphs) scans; data correspond to all six boreholes.

Table 1
Summary of scenarios according to input parameters and feature extraction.

Scenario	Input parameters[a]	Feature extraction	No. of training/ testing cases
W_{PCA}	$(W_{R,p}, W_{G,p}, W_{B,p})$	PCA	30/6
W_{CA}		CA	30/6
UV_{PCA}	$(UV_{R,p}, UV_{G,p}, UV_{B,p})$	PCA	26/6
UV_{CA}		CA	26/6
WUV_{PCA}	$(W_{R,p}, W_{G,p}, W_{B,p}, UV_{R,p},$	PCA	26/6
WUV_{CA}	$UV_{G,p}, UV_{B,p})$	CA	26/6

[a] $p = 10, 20, \ldots, 100$.

borehole images are used to automatically assess the ore grade.

2.3. Image processing

Fig. 4 shows a typical example of televiewer logs from scans with white and UV lights at a length interval in which fluorite was defined visually by an experienced geologist; the assay of drilling chips classifies the rock in that section as medium grade ore (see Fig. 2), which is compatible with the presence of dolomitic breccia observed. For each scan, the RGB value of each pixel is calculated by WellCAD (ALT, 2020) and it is represented by a colour palette in the three columns on the right of the corresponding image of the borehole walls in Fig. 4. This gives, for each light type, three 2D matrices for red, green and blue colors with intensities in the range 0–255. The corresponding mean color intensity at each depth is the white curve (three 1D arrays). Note that no apparent difference is observed visually between fluorite and dolomitic breccia.

The image of each section of the televiewer, defined by the initial and final depths of the drill chips samples collected for assaying, is formed by approximately 1500 sets of red, green and blue color intensities. To characterize this color information, percentiles 10, 20, ..., 100 of the distribution of pixel intensities of red, green and blue colors in that section are calculated; this leads to two triplets of percentiles of color intensities (PCI) of pixels, $(W_{R,p}, W_{G,p}, W_{B,p})$ and $(UV_{R,p}, UV_{G,p}, UV_{B,p})$ for each hole section and p percentile for the white and UV light scans ($p = 10, 20, \ldots, 100$), making up 30 PCI for each hole section and type of light scan. The cumulative distribution functions (CDFs) of pixel intensities of red, green and blue from each scan type are shown in Fig. 5. They are colored as function of the fluorite content (hot colors represent medium ore grade; cold ones correspond to waste).

Although some trend between PCI and CaF_2 content can be observed, mainly for green and blue colors from white light scans (see central and right graphs in Fig. 5), these relations are not easily described with simple analytical functions of the PCI that could be applied in a classical multivariant analysis. The resort to advanced machine learning techniques to explore such complex relationship appears to be natural.

3. Data pre-processing

The sets of triplets, $(W_{R,p}, W_{G,p}, W_{B,p})$ and $(UV_{R,p}, UV_{G,p}, UV_{B,p})$, $p = 10, 20, \ldots, 100$, for the sections considered are selected as input parameters; the respective fluorite content and grade classification are taken as the output for the machine learning algorithms.

The combination of the three possible input sets (i.e. PCI from white light, from UV light and both) and two types of feature extraction techniques (i.e. principal component analysis and correlation analysis) to reduce the size of the input dimension are proposed. This makes up six different modeling scenarios that are summarized in Table 1; they are identified with two letters, the first one describes the light source (W for white light scans, UV for UV light scans and WUV for the combination of color characteristics for white and UV light scans) and the second one describes the feature extraction technique (PCA for principal component analysis and CA for correlation analysis).

3.1. Dataset partition

The original dataset needs to be divided into training and testing sets. The first is used for developing the model and the second is used for verifying its generalization and robustness. Generally, the ratio of

Table 2

Percentage of the cumulative total variability in the data explained by each principal component.

Light source	1	2	3	4	5	6	7	8
W[a]	65.1	93.1	97.3	–	–	–	–	–
UV	62.6	73.9	82.3	87.6	90.9	92.9	94.7	95.7

[a] Percentiles with constant intensity are discarded for the analysis.

Table 3

Correlation coefficients between PCI and fluorite content.

Percentile	White light			UV light		
	Red	Green	Blue	Red	Green	Blue
10	0.16	**0.35**	**0.40**	0.17	**0.51**	0.12
20	0.12	**0.33**	**0.39**	0.28	0.25	0.29
30	0.13	**0.37**	**0.41**	0.35	0.25	**0.39**
40	0.12	**0.36**	**0.39**	0.22	**0.42**	0.30
50	0.06	0.33	**0.41**	0.15	**0.45**	0.20
60	−0.01	0.31	**0.45**	0.11	0.30	0.21
70	−0.04	**0.37**	**0.47**	0.18	0.20	0.26
80	−0.10	**0.42**	**0.48**	0.10	0.27	0.31
90	−0.10	**0.40**	**0.45**	0.12	**0.37**	0.27
100	–	–	0.15	**0.62**	**0.53**	0.17

Note: Bold numbers are $|r| \geq 0.3$ and *p*-value ≤ 0.05.

training to testing set cases is 8:2 or 7:3 (Koopialipoor et al., 2019; Li et al., 2021a; Li et al., 2021b), which can be tuned according to the scale of the data. The cases of valid scans are 36 for W light and 32 for UV light, as explained in Section 2.2. They are divided in training/testing 30/6 and 26/6 respectively, as shown in Table 1. For the combined use of the W and UV light, data from both scans must be available, so the number of valid scans is in this case equal to the number of UV valid scans i.e. 32. The cases of waste are 16 and 15 for W and UV light scans, respectively; sections of low grade are 14 and 12 for W and UV light scans, respectively, and sections of medium grade are scarce, 6 and 5 for white and UV light scans. This unbalanced number of cases of the three grades may lead to uncorrelated training and testing data sets, on which a weak generalization of the supervised learning would be obtained. For selecting the training and testing sets, one case was randomly selected from each borehole (hence six cases are selected, see Table 1) and used for developing the testing set and the other cases constituted the training set. Although they don't fully meet the 8:2 rule, they are still reasonable in view of the number of scan data. For eliminating the adverse effects caused by unbalanced data division, a k-fold cross-validation (Anguita et al., 2012; Fushiki, 2011; Rodriguez et al., 2009) is applied. It randomly separates the original training set into *k* equal-size subsets where *k*-1 subsets are used as a new training set and the remaining subset is used for validation. The algorithm searches for a model that leads to the best fitness value for the *k* sets of training samples. According to some authors (Marcot and Hanea, 2021; Yadav and Shukla, 2016), 5 or 10-fold cross-validation works well. Considering the scale of the training set, 5-fold cross-validation has been employed.

The prediction ability of machine learning techniques is often assessed from only one random division of the dataset into training and testing sets. This prediction performance may not properly reflect the overall goodness of the dataset and the prediction ability of the model. The relatively small size of our datasets and their unbalanced nature may cause unstable prediction results as different partitions of the original dataset would lead to different prediction accuracy. In order to account for this, a repeated *k*-fold cross validation technique was employed where thirty random divisions of the dataset are implemented to produce thirty different combinations of training and testing sets. For each training set, the 5-fold cross validation is implemented. The average results from the thirty training/testing combinations provide a more robust evaluation of the prediction ability.

3.2. Feature extraction

For each section, the original inputs from W and UV scans are 30 color intensities respectively which encompass a large input dimension compared with the size of the dataset. To reduce the complexity of calculation and preserve as much statistical information as possible, principal component analysis (PCA) and correlation analysis (CA) are considered.

3.2.1. Principal component analysis

PCA (Wold et al., 1987) finds new uncorrelated variables, or principal components, that are linear combinations of the original variables that maximize the variance between them. Substituting the original variables by a few principal components reduces the input dimension and simplifies the model fit. The cumulative variance is shown in Table 2; a percent of the total variance higher than 95 % is considered to define the number of components retained, this being 3 principal components for W and 8 principal components for UV datasets.

3.2.2. Correlation analysis

The results of a Spearman correlation analysis between PCI of pixels and the fluorite content are presented for each light source in Table 3; no results are shown when the intensity colors for a given percentile are constant, as occurs for the 100 percentiles from red and green with white light illumination. PCI with significant correlation (coefficient $|r| \geq 0.3$ and *p*-value ≤ 0.05) have been selected as inputs (highlighted in bold in Table 3). For white light, this applies to green percentiles 10–40 and 70–90 and most of the blue percentiles, while no red percentile meets the significance condition; for UV light, these are one percentile for red (100) and blue (30), and five percentiles for green (10, 40, 50, 90 and 100). The different significances of correlations of PCI and fluorite content for different illuminating sources may indicate some differential optical response from the materials.

4. The model

The support vector machine (SVM) will be used in this study as the benchmark tool to predict fluorite grade (Support vector regression, SVR) or classify the rock into waste, low grade ore, and medium grade ore (Support vector classification, SVC), as function of an *n*-dimensional set of input variables (linear combinations of PCI defined from the PCA or the most relevant PCI obtained from CA). SVM was developed initially for tackling classification issues, and it can also be extended to solve regression problems (Quan et al., 2020; Vapnik, 1995). SVM is very appropriate for analyzing small databases with large-dimension input data, as is the case of this work, compared with other classical approaches like artificial neural network and K-nearest neighbors (Qi and Tang, 2018). The main idea of SVC is to find the optimal separating hyperplane that correctly partitions the training dataset with the largest geometric separation, while SVR aims to find a function that deviates from every output by no more than a certain error for each training data point. Details of the SVM optimization can be found in Smola and Schölkopf (2004).

The performance of SVM depends on certain parameters that must be optimized before testing the model. One of them is the penalty factor *C* that weights the cases with an error in excess of a certain value in the training phase; a high penalty factor means a low tolerance for prediction deviations and may induce over-fitting. Another setting is the internal parameter *γ* of the kernel function used (a radial based function, Quan et al., 2020, in the present case), that controls the range of action of each support vector (e.g. small action range of the support vectors induces over-fitting). In order to optimize these parameters, a myriad of metaphor-based optimization algorithms of artificial intelligence-based models is available (Li et al., 2021a; Li et al., 2021b; Zhou et al., 2021a; Biswas et al., 2022. A common practice in these ML-based papers is to rate some of these algorithms according to the resulting performance to

Table 4

Summary of the performance metrics.

Regression	Formula[a]	Classification	Formula[b]
Coefficient of determination	$R^2 = 1 - \dfrac{\sum_{i=1}^{N}(y_i - y'_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$	Total	$AcT = \dfrac{T_W + T_{LG} + T_{MG}}{N_W + N_{LG} + N_{MG}}$
Variance accounted for	$VAF = \left[1 - \dfrac{var(y_i - y'_i)}{var(y_i)}\right]·100$	For waste (W)	$AcW = \dfrac{T_W}{N_W}$
Root mean squared error	$RMSE = \sqrt{\dfrac{1}{N}\sum_{i=1}^{N}(y_i - y'_i)^2}$	For low grade (LG)	$AcLG = \dfrac{T_{LG}}{N_{LG}}$
Mean absolute percentage error	$MAPE = \dfrac{100}{N}\sum_{i=1}^{N}\left|\dfrac{y_i - y'_i}{y_i}\right|·$	For medium grade (MG)	$AcMG = \dfrac{T_{MG}}{N_{MG}}$

[a] y_i is the measured fluorite content, y'_i the corresponding predicted value, $\bar{y}$ the mean of the measured fluorite contents, and N the number of measurements.

[b] Ac means classification accuracy. T_c is the number of true positives of the c category (c is W for waste, LG for low grade and MG for medium grade). N_c is the number of sections of the c category.
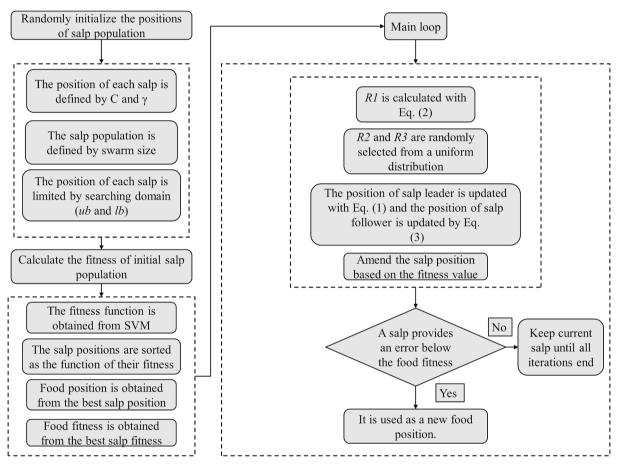

Fig. 6. A general optimization process of the selection of SVM hyper-parameters by SSA.

solve a specific problem; however, the prediction performance by different *meta*-heuristic algorithms is very similar. Regarding this, to address the performance of a bio-inspired *meta*-heuristic algorithm is not the motivation of this study, but to discuss the potential of percentile color intensity of images to assess the rock composition and the ore grade in particular by ML-based models. The salp swarm algorithm (SSA; Mirjalili et al., 2017) that has proved to be effective in solving different optimization problems in various domains (Li et al., 2021a; Li et al., 2021b), it is used here in combination with SVM to select suitable support vector parameter combinations so as to prevent local optima.

Salps are marine animals that form chain-like swarms where there is a leader salp and many follower salps. The leader plays an important role in leading the direction of follower salps and exploring the food source. In the SSA model, a population is generated at first which consists of one leader and some followers. The position of the leader would be updated iteratively to search for the best food source (i.e. fitness function, described by the MSE for the regression and the total accuracy for the classification; both metrics are described in Table 4). The two parameters C and γ of the support vector are connected with the position of the first salp and modified until an optimum condition is met within a maximum number of iterations.

The position of the salp leader is updated within the searching domain through the following function:

$$x_m^1 = \begin{cases} F_m + R1((ub_m - lb_m))R2 + lb_m, & R3 \geq 0 \\ F_m - R1((ub_m - lb_m))R2 + lb_m, & R3 \leq 0 \end{cases} \quad (1)$$

where x_m^1 represents the position of the leader salp (first salp) in the mth dimension of domain (note that for this case, $m = 2$); F_m denotes the

Table 5
Main prediction performance statistics from full-data for white and UV light scans.

Light Source		Training				Testing			
		R^2	VAF	MAPE	RMSE	R^2	VAF	MAPE	RMSE
W_{PCA}	Mean	0.92	92.60	0.21	2.37	0.63	68.82	0.47	5.08
	Min.	0.87	87.94	0.04	0.34	−0.27	−3.91	0.23	2.94
	Max.	1.00	99.86	0.32	3.09	0.87	97.00	0.63	9.06
	Std.	0.03	3.00	0.07	0.74	0.25	23.87	0.11	1.53
W_{CA}	Mean	0.94	94.54	0.16	1.76	0.62	67.61	0.46	4.99
	Min.	0.82	82.29	0.04	0.33	−0.42	−34.89	0.27	2.73
	Max.	1.00	99.87	0.40	3.58	0.92	93.72	0.68	8.97
	Std.	0.05	5.26	0.12	1.17	0.37	34.37	0.09	1.69
UV_{PCA}	Mean	0.94	94.09	0.13	1.67	0.54	57.89	0.45	5.40
	Min	0.84	84.30	0.03	0.29	0.14	14.61	0.23	2.47
	Max	1.00	99.88	0.33	3.32	0.91	92.93	0.79	7.98
	Std.	0.06	5.79	0.09	1.14	0.23	22.82	0.13	1.48
UV_{CA}	Mean	0.80	80.30	0.25	3.62	0.60	66.55	0.38	4.96
	Min	0.71	71.50	0.04	0.80	0.00	0.23	0.15	1.89
	Max	0.99	99.10	0.33	4.41	0.95	95.74	0.81	8.32
	Std.	0.07	6.78	0.07	0.83	0.27	24.84	0.15	1.67

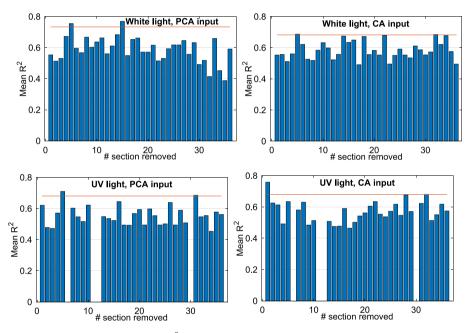Min: minimum; Max: maximum; Std.: standard deviation.

Fig. 7. Average R^2 from 30 divisions for "take one out" models.

location of the food source of the mth dimension; ub_m and lb_m are the searching upper and lower bounds, respectively; $R2$ and $R3$ are random variables uniformly distributed in the interval [0,1], and $R1$ is calculated from the current iteration number (lp) and the total number of iterations (LP) as follows:

$$R1 = 2e^{-\left(\frac{4lp}{LP}\right)^2} \tag{2}$$

The position of the ith follower salp in the mth dimension is updated to search the food source in the local range as follows:

$$x_m^i = \frac{1}{2}\left(x_m^i + x_m^{i-1}\right), i > 1 \tag{3}$$

A general sketch of the optimization process of SSA for C and γ is shown in Fig. 6. The swarm size (i.e. the number of salps) and the maximum number of iterations must be chosen, both being in fact significant optimization parameters, see Section 5.

5. Results and discussion

Regression and classification patterns are developed and evaluated for six scenarios of input parameters (see Table 1). The model was programmed in a Matlab (MATLAB, 2022) environment, with support vector machine code from LIBSVM (Chang and Lin, 2011). Input parameters are the principal components selected in Section 3.2.1 for PCA and in Section 3.2.2 for CA.

For regression, the output is the percentage of fluorite and for classification, it is the fluorite grade class. Both regression and classification models utilize the same SSA parameters (swarm size and iteration number) and inputs. As explained in Section 4, similar to other swarm-based heuristic algorithms, swarm size and maximum iterations have a key impact on speed and prediction performance (Yu et al., 2020; Zhou et al., 2021b; Zhou et al., 2021c). Different iteration numbers and swarm sizes were tested. When swarm size and maximum iterations were small, the prediction performance was generally not stable and over-fitting or under-fitting sometimes occurred. A swarm size of 50 and a maximum iteration number of 200 were found to yield efficient optimization
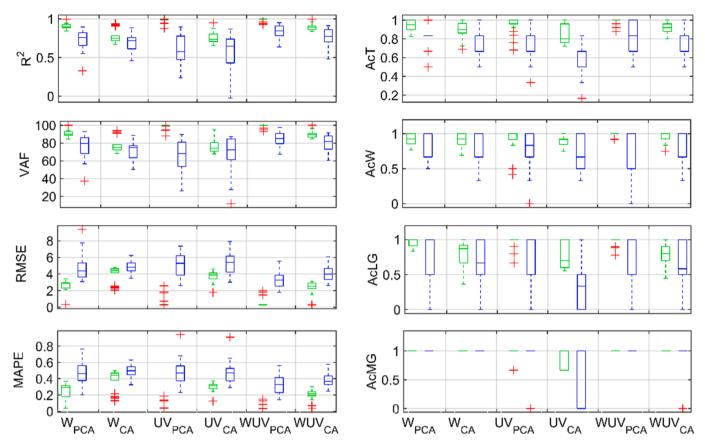
Fig. 8. Summary of the prediction performance of regression (left graphs) and classification (right graphs) for the six scenarios considered after removing section #5; green boxes correspond to training and blue boxes to testing; refer to Table 4 for descriptions of the metrics.

Table A1
PCA results after removing section #5.

Light source	1	2	3	4	5	6	7	8	9
W[a]	63.06	93.26	97.41	–	–	–	–	–	–
UV	60.41	71.19	81.10	86.78	90.63	92.84	94.55	95.62	–
WUV[a]	37.15	67.29	83.16	86.95	90.43	92.19	93.74	94.98	96.02

[a] Percentiles with constant intensity are discarded for the analysis.

Table A2
Correlation coefficients between PCI and fluorite content after removing section #5.

Percentile	White light			UV light		
	Red	Green	Blue	Red	Green	Blue
10	0.12	0.30	**0.35**	0.13	**0.48**	0.07
20	0.08	0.28	**0.34**	0.23	0.19	0.24
30	0.09	0.32	**0.36**	0.30	0.20	0.34
40	0.08	0.31	**0.34**	0.16	**0.37**	0.24
50	0.02	0.27	**0.37**	0.09	**0.38**	0.15
60	−0.05	0.26	**0.41**	0.07	0.23	0.16
70	−0.09	0.32	**0.43**	0.14	0.14	0.22
80	−0.14	**0.37**	**0.44**	0.05	0.20	0.27
90	−0.13	**0.35**	**0.40**	0.08	0.31	0.23
100	–	–	0.15	**0.59**	**0.50**	0.13

Note: Bold numbers are $|r| \geq 0.3$ and p-value ≤ 0.05.

results and, when increased, the prediction performance did not significantly improve while the computational time increased. Those values were thus adopted in all models.

Despite that the model hyperparameters are optimized on the MSE (regression) and the total accuracy (classification), other metrics are considered (see Table 4) to describe the performance of the models. Table 5 shows the statistics of these metrics for the 30 training/testing datasets considered.

For regression, four classical indicators have been employed: coefficient of determination (R^2), variance accounted for (VAF), root mean squared error (RMSE) and mean absolute percentage error (MAPE). For classification, the classification accuracy for total samples (AcT), waste (AcW), low grade (AcLG) and medium grade (AcMG) have been used. As mentioned before, 30 random divisions of training and testing sets are carried out and for each of them, a prediction model is built; the same divisions are used for regression and classification for all scenarios. A summary of the main statistics of the metrics for regression scenarios W_{PCA}, W_{CA}, UV_{PCA} and UV_{CA} can be seen in Table 5 for training (TR) and testing (TS) sets. Taking the R^2 as a specific research objective, the mean value for the training sets is excellent while it is lower for the testing sets.

Some low, or even negative R^2 occur (see the minimum values in Table 5) which could be due to outliers in the data. For detecting such outliers, one section is removed from the dataset and the remaining sections are used for developing the regression models. This is repeated until all sections have been individually removed. We can assume that if one section is an outlier, then the determination coefficient will increase significantly when it is not included in the calculation. In order to assess

Table A3
Regression results.

Regression		Training set				Testing set			
		R^2	VAF	MAPE	RMSE	R^2	VAF	MAPE	RMSE
W_{PCA}	Mean	0.90	90.30	0.26	2.68	0.73	76.93	0.47	4.65
	Min	0.84	84.28	0.04	0.34	0.33	37.39	0.20	3.08
	Max	1.00	99.83	0.37	3.40	0.90	92.78	0.76	9.38
	Std.	0.03	3.21	0.08	0.59	0.13	11.78	0.13	1.42
W_{CA}	Mean	0.78	78.01	0.39	4.01	0.70	71.87	0.50	4.82
	Min	0.67	68.25	0.12	2.07	0.46	50.51	0.33	3.49
	Max	0.94	94.24	0.50	4.83	0.88	88.68	0.63	6.24
	Std.	0.09	8.47	0.13	0.92	0.10	9.31	0.08	0.67
UV_{PCA}	Mean	0.99	99.08	0.05	0.48	0.61	65.01	0.48	5.15
	Min	0.88	87.92	0.03	0.28	0.24	26.45	0.23	2.60
	Max	1.00	99.87	0.19	2.58	0.90	89.60	0.94	7.36
	Std.	0.03	2.51	0.04	0.55	0.19	17.99	0.15	1.32
UV_{CA}	Mean	0.75	76.04	0.30	3.77	0.60	66.90	0.47	5.30
	Min	0.65	67.76	0.12	1.78	−0.03	11.88	0.29	3.02
	Max	0.95	94.96	0.37	4.60	0.87	87.26	0.91	7.93
	Std.	0.07	6.47	0.05	0.60	0.22	20.59	0.13	1.37
WUV_{PCA}	Mean	0.99	99.35	0.04	0.44	0.83	84.72	0.33	3.32
	Min	0.93	93.02	0.03	0.27	0.64	67.34	0.14	1.80
	Max	1.00	99.88	0.15	1.99	0.95	97.55	0.56	5.54
	Std.	0.02	1.60	0.03	0.45	0.09	7.71	0.11	0.90
WUV_{CA}	Mean	0.89	89.41	0.21	2.48	0.77	79.81	0.39	4.03
	Min	0.84	84.33	0.03	0.29	0.48	60.49	0.25	2.63
	Max	1.00	99.85	0.30	3.12	0.91	91.59	0.58	6.07
	Std.	0.04	3.46	0.06	0.55	0.10	9.52	0.09	0.85

Table A4
Classification results.

Classification		Training set				Testing set			
		AcT	AcW	AcLG	AcMG	AcT	AcW	AcLG	AcMG
W_{PCA}	Mean	0.95	0.92	0.96	1.00	0.84	0.84	0.77	1.00
	Min	0.83	0.77	0.83	1.00	0.50	0.50	0.00	1.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Std.	0.05	0.07	0.06	0.00	0.12	0.19	0.30	0.00
W_{CA}	Mean	0.88	0.90	0.82	1.00	0.76	0.76	0.67	1.00
	Min	0.69	0.69	0.36	1.00	0.50	0.33	0.00	1.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Std.	0.09	0.09	0.18	0.00	0.11	0.24	0.30	0.00
UV_{PCA}	Mean	0.96	0.94	0.97	0.99	0.71	0.76	0.62	0.93
	Min	0.68	0.42	0.67	0.67	0.33	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Std.	0.08	0.14	0.08	0.06	0.15	0.29	0.33	0.25
UV_{CA}	Mean	0.85	0.90	0.77	0.88	0.58	0.71	0.38	0.73
	Min	0.72	0.75	0.56	0.67	0.17	0.33	0.00	0.00
	Max	1.00	1.00	1.00	1.00	0.83	1.00	1.00	1.00
	Std.	0.11	0.07	0.17	0.16	0.17	0.25	0.36	0.45
WUV_{PCA}	Mean	0.99	0.98	0.99	1.00	0.81	0.79	0.74	1.00
	Min	0.88	0.92	0.78	1.00	0.50	0.00	0.00	1.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Std.	0.03	0.03	0.05	0.00	0.19	0.27	0.31	0.00
WUV_{CA}	Mean	0.91	0.96	0.81	1.00	0.79	0.84	0.66	0.97
	Min	0.80	0.75	0.44	1.00	0.50	0.33	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Std.	0.07	0.07	0.14	0.00	0.12	0.20	0.29	0.18

whether the improvement is significant from a statistical point of view, the 95-percentile of the determination coefficient is employed as criterion. All four scenarios are tested, i.e., W_{PCA}, W_{CA}, UV_{PCA} and UV_{CA}. If the case removed improves the prediction performance in most scenarios, then it can be considered a candidate for outlier. The 'take one out' method involves: i) feature extraction (PCA and CA) from the new dataset; ii) random divisions of the dataset into 30 sets of training and testing data (the testing set is always formed by 6 samples and the training by 29 for white light and 25 for UV light), and iii) train and test the prediction model for each of the random divisions. This operation is repeated 36 times and 32 times for W- and UV-based scenarios, respectively, until all models leaving out one section each are built.

When an outlier is removed from the dataset, significantly better

prediction performance will be procured and this will result in a significantly higher R^2. Fig. 7 shows the mean R^2 of the 30 random divisions for all regression models with one section removed; the numbering of sections is (see Fig. 2) from collar to bottom, hole H20 to H25, so that section 1 is 29.4 % CaF_2 and the last section (#36) is 3.9 % CaF_2. Sections #6, #11, #12 and #30 do not exist in the UV dataset. The horizontal lines in Fig. 6 show the 95 percentiles of R^2. The following sections removal score above this percentile: W_{PCA}: #5, #15; W_{CA}: #5; #32; UV_{PCA}: #5, #31; UV_{CA}: #1, #32.

The removal of section #5 improves outstandingly the results of three out of four scenarios (it scores at percentiles 95.8, 98.6, 98.5 and 88.9) so this section could be an outlier. Section #32 could also be a candidate although its R^2 only scores above 95 % in two cases, while in

the other two is nowhere near that threshold. For this reason, only section #5 (i.e. fifth section of borehole H20, with 27.6 % CaF_2) was removed. We suspect that the reason for its offending behavior is uneven sampling with an uncertain grade rating, probably with lower fluorite percentage than the actual one; higher fluorite percentage has a slight right shift tendency for G- and B-CDF plots as Fig. 5 shows, where the 27.6 % CaF_2 section lies in the rightmost position, though not being the section with the highest-grade. Interestingly, the same procedure was implemented for classification scenarios but the classification results were not noticeably influenced by any single section removal, which can be explained because of the relatively broad classification ranges. For the relevant case of Section 5, this sample has a measured grade of 27.6 %, and is usually wrongly predicted in the regression with a grade much in excess of that value, although still falling in the medium grade ore class, that covers the range $20 \leq CaF_2 < 45$ %, so being correctly classified.

After removal of section #5, the PCA and CA selection is redone with the remaining 35 white and 31 UV light scans and also with combined WUV data. The cumulative total variability explained by each principal component is different but the number of components considered as inputs are the same as in Table 2 for W and UV sets, while one less component is required for WUV. For CA results, significantly correlated PCI are fewer for both white light and UV light scans compared with the results before removing section #5; however, significant correlations are still obtained for most of the blue percentiles from white scans. The detailed results can be seen in Tables A.1 and A.2 in the Appendix. A summary of the main metrics statistics (where 30 new random divisions were employed for each scenario) can be seen in Tables A.3 and A.4 in the Appendix. It appears as if principal component inputs are more sensitive to outliers in the UV scenario, both W and UV light (left graphs in Fig. 6) than the straight variables selected as having a better correlation with grade (right graphs in Fig. 6). In the W_{CA} analysis, section 5 barely exceeds the 95 % threshold, see Fig. 7 upper right plot, while in the UV_{CA} analysis, the removal of section 5 does not relevantly affect R^2. Fig. 8 shows boxplots of the distributions of the metrics for SVR and SVC for the training (green boxes) and testing (blue boxes) sets.

In both regression and classification, PCA provides in general better results than CA in terms of mean and dispersion of the metrics considered. This is more evident for regression when the combination of colors from white and UV lights (WUV) is considered. These results indicate that CA can capture some significant optical information from televiewer scans, but probably ignores some supplementary information that can contribute to characterize the fluorite grade, while the linear combinations of PCI in the principal components provide with a richer information on the pixel properties.

For regression testing sets, UV has worse R^2, VAF and RMSE than W light and slightly better MAPE. The mean and standard deviation of the RMSE for PCA are 5.15 ± 1.32 % (UV_{PCA}) and 4.65 ± 1.42 % (W_{PCA}), with mean R^2 of 0.61 and 0.73, respectively. The best predictions according to the four metrics considered are obtained when both light sources are combined (WUV_{PCA}); the mean RMSE is then 3.32 ± 0.90 %, with a determination coefficient of 0.83 ± 0.09.

The inputs from W_{PCA} provide the best classification accuracies for all rock classes: 0.84 ± 0.12, 0.84 ± 0.19, 0.77 ± 0.30, and 1 for total, waste, low grade and medium grade for testing sets. For the waste, this means that 16 % of the samples are classified wrongly as ore in average terms, while 23 % of low-grade ore sections are misclassified. All medium grade ore sections are classified correctly for the 30 divisions of the dataset. Low grade ore is generally the category with a worse classification accuracy compared with waste and medium grade. These metrics involve that no further chemical assaying is required when a sample is classified as medium grade, while the classification into the other rock classes should be taken as first estimation of the ore grade. Such estimation would need to be confirmed through conventional assaying, especially for low-grade ore, to increase the reliability of the results. Enlarging the database would likely improve the model performance

increasing the significance of the results, allowing a further reduction in the amount of chemical analyses.

For case studies where it is not possible to enlarge the dataset available for the model development, and relatively small databases have to be employed, repeated cross-validation has proved to be a good strategy. Repeated cross-validation can also be combined with other machine learning or deep learning techniques (Phoon and Zhang, 2023), such as convolutional neural networks (Zhang et al., 2021), random forest (Liu et al., 2023; Fernández et al., 2023), or recurrent neural networks (Zhang et al., 2022), all of them applied to improving prediction models in geoscience and geotechnical problems.

6. Conclusions

This paper proposes a new fluorite grade prediction approach based on RGB values obtained from optical televiewer scanning and machine learning techniques in the Lújar underground mine located in Órgiva, Granada, Spain. RGB intensities of the pixels are procured from borehole walls by televiewer scanning with white and UV light illumination. The composition of the rock corresponding to the image logs was determined by chemical analysis of the drill cuttings sampled and used as the output. Percentiles of the color intensities (PCI) of borehole images are used as input parameters for regression and classification issues of fluorite grade. Three types of color information have been tested, comprising PCI from white light scans (W), UV light scans and the combination of both (WUV). Two kinds of feature extraction techniques are employed for input selection: the first one is from the significantly correlated inputs with fluorite components; the second one is from PCA technique.

The support vector machine (SVM) is used to establish the prediction models. The hyperparameters of the SVM (C and γ) are optimized using a salp swarm metaheuristic algorithm. The results of the prediction models are assessed by repeated cross-validation and rated with classical statistical indicators. A "take one out" method is proposed for outlier data detection. One of the data sections was removed with this method, resulting in an improved prediction capacity.

In general, the combination of white light and UV light scans is more effective to predict fluorite grade from regression. If a single light source is used, the white light would be recommended. The average regression results for testing sets are $R^2 = 0.83$ and RMSE = 3.32 % from WUV_{PCA} scenario. For classification, the best result is obtained with white light, W_{PCA}, with average classification accuracies, of 0.84 (total), 0.84 (for waste), 0.77 (for low grade ore) and 1 (for medium-grade ores).

The relatively low-cost and convenience of ore image procurement and processing makes this novel approach robust and easy to implement for fluorite grade prediction. Given the limited errors and acceptable prediction accuracies, the approach described here can be used as a first assessment of the fluorite grade, helping to save a fraction of laboratory analysis work. Additional work is needed in order to investigate the reliability e.g. with other fluorite ores or other minerals. The collection of larger datasets would improve the significance of the results, but, since the models are intrinsically site dependent, they can hardly be generalized to other operations. However, the method proposed can be adapted to other mining sites in order to develop an ore grade prediction model based on RGB intensities of the images of the blasthole walls. Such a methodology will reduce the time offset for grade control, especially when medium ore grade is observed, allowing to detect these areas from the first steps of the drilling and to take prompt decisions on mine development. The availabilty of a system that would automatically log the boreholes would boost this procedure towards a nearly online assessment of the ore grade of the deposit. The potential of UV light scans still needs to be explored, where different wavelengths are likely worthwhile investigating to improve the prediction accuracy of fluorite ore.

Declaration of competing interest

Acknowledgments

Appendix

References

ALT, 2020. WellCAD Software 5.4. Version TK210215. Advanced Logic Technology, Luxemburg.

Amor, C., Navarro, R., 2016. Minería del flúor en Sierra de Lújar. Rocas y Minerales 536 (11), 46–58.

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S., 2012. The 'K'in K-fold cross validation. In: 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), pp. 441–446.

Baykan, N.A., Yılmaz, N., 2010. Mineral identification using color spaces and artificial neural networks. Comput. Geosci. 36 (1), 91–97. https://doi.org/10.1016/j.cageo.2009.04.009.

Berrezueta, E., Ordóñez-Casado, B., Bonilla, W., Banda, R., Castroviejo, R., Carrión, P., Puglla, S., 2016. Ore petrography using optical image analysis: application to Zaruma-Portovelo deposit (Ecuador). Geosciences 6 (2), 30. https://doi.org/10.3390/geosciences6020030.

Biswas, R., Li, E., Zhang, N., Kumar, S., Rai, B., Zhou, J., 2022. Development of hybrid models using metaheuristic optimization techniques to predict the carbonation depth of fly ash concrete. Constr. Build. Mater. 346, 128483 https://doi.org/10.1016/j.conbuildmat.2022.128483.

Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (3), 1–27. https://doi.org/10.1145/1961189.1961199.

Chatterjee, S., Bhattacherjee, A., Samanta, B., Pal, S.K., 2010. Image-based quality monitoring system of limestone ore grades. Comput. Ind. 61 (5), 391–408. https://doi.org/10.1016/j.compind.2009.10.003.

Desta, F.S., Buxton, M.W., 2017. The use of RGB imaging and FTIR sensors for mineral mapping in the Reiche Zeche underground test mine, Freiberg. Proceedings of Real Time Mining - International Raw Materials Extraction Innovation Conference: 10th & 11th, Amsterdam, The Netherlands. https://nbn-resolving.org/urn:nbn:de:bsz:105-qucosa-231302.

Donskoi, E., Suthers, S.P., Fradd, S.B., Young, J.M., Campbell, J.J., Raynlyn, T.D., Clout, J.M.F., 2007. Utilization of optical image analysis and automatic texture classification for iron ore particle characterisation. Miner. Eng. 20 (5), 461–471. https://doi.org/10.1016/j.mineng.2006.12.005.

Donskoi, E., Manuel, J.R., Austin, P., Poliakov, A., Peterson, M.J., Hapugoda, S., 2013. Comparative study of iron ore characterisation using a scanning electron microscope and optical image analysis. Appl. Earth Sci.: Trans. Inst. 122 (4), 217–229. https://doi.org/10.1179/1743275814Y.0000000042.

Donskoi, E., Poliakov, A., Manuel, J.R., Peterson, M., Hapugoda, S., 2015. Novel developments in optical image analysis for iron ore, sinter and coke characterisation. Appl. Earth Sci.: Trans. Inst. 124 (4), 227–244. https://doi.org/10.1179/1743275815Y.0000000013.

Dumakor-Dupey, N.K., Arya, S., 2021. Machine Learning—A review of applications in mineral resource estimation. Energies 14 (14), 4079. https://doi.org/10.3390/en14144079.

Fernández, A., Sanchidrián, J.A., Segarra, P., Gómez, S., Li, E., Navarro, R., 2023. Rock mass structural recognition from drill monitoring technology in underground mining using discontinuity index and machine learning techniques. Int. J. Min. Sci. Technol. 33 (5), 555–571. https://doi.org/10.1016/j.ijmst.2023.02.004.

Fushiki, T., 2011. Estimation of prediction error by using K-fold cross-validation. Stat. Comput. 21 (2), 137–146. https://doi.org/10.1007/s11222-009-9153-8.

Ilin, A., Velasco, F., Navarro, R., Tornos, F., 2019. New data on Alpine type fluorite deposits: Case of Lújar mine in Betic Cordillera (SE Spain). MACLA: Revista Española De La Sociedad De Mineralogía 24 (6), 63–64.

Jafrasteh, B., Fathianpour, N., 2017. A hybrid simultaneous perturbation artificial bee colony and back-propagation algorithm for training a local linear radial basis neural network on ore grade estimation. Neurocomputing 235, 217–227. https://doi.org/10.1016/j.neucom.2017.01.016.

Jooshaki, M., Nad, A., Michaux, S., 2021. A systematic review on the application of machine learning in exploiting mineralogical data in mining and mineral industry. Minerals 11 (8), 816. https://doi.org/10.3390/min11080816.

Kaplan, U.E., Topal, E., 2020. A new ore grade estimation using combine machine learning algorithms. Minerals 10 (10), 847. https://doi.org/10.3390/min10100847.

Koopialipoor, M., Armaghani, D.J., Hedayat, A., Marto, A., Gordan, B., 2019. Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions. Soft Comput. 23 (14), 5913–5929. https://doi.org/10.1007/s00500-018-3253-3.

Lane, G.R., Martin, C., Pirard, E., 2008. Techniques and applications for predictive metallurgy and ore characterization using optical image analysis. Miner. Eng. 21 (7), 568–577. https://doi.org/10.1016/j.mineng.2007.

Li, N., Hao, H., Gu, Q., Wang, D., Hu, X., 2017. A transfer learning method for automatic identification of sandstone microscopic images. Comput. Geosci. 103 (6), 111–121. https://doi.org/10.1016/j.cageo.2017.03.007.

Li, E., Yang, F., Ren, M., Zhang, X., Zhou, J., Khandelwal, M., 2021a. Prediction of blasting mean fragment size using support vector regression combined with five optimization algorithms. J. Rock Mech. Geotech. Eng. 13 (6), 1380–1397. https://doi.org/10.1016/j.jrmge.2021.07.013.

Li, E., Zhou, J., Shi, X., Jahed Armaghani, D., Yu, Z., Chen, X., Huang, P., 2021b. Developing a hybrid model of salp swarm algorithm-based support vector machine to predict the strength of fiber-reinforced cemented paste backfill. Eng. Comput. 37 (4), 3519–3540. https://doi.org/10.1007/s00366-020-01014-x.

Liu, C., Li, M., Zhang, Y., Han, S., Zhu, Y., 2019. An enhanced rock mineral recognition method integrating a deep learning model and clustering algorithm. Minerals 9 (9), 516. https://doi.org/10.3390/min9090516.

Liu, S., Wang, L., Zhang, W., Sun, W., Fu, J., Xiao, T., Dai, Z., 2023. A physics-informed data-driven model for landslide susceptibility assessment in the Three Gorges Reservoir Area. Geosci. Front. 14 (5) https://doi.org/10.1016/j.gsf.2023.101621, 101621.

Marcot, B.G., Hanea, A.M., 2021. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? Comput. Stat. 36 (3), 2009–2031. https://doi.org/10.1007/s00180-020-00999-9.

Marschallinger, R., 1997. Automatic mineral classification in the macroscopic scale. Comput and Geosci. 23 (1), 119–126. https://doi.org/10.1016/s0098-3004(96)00074-x.

MATLAB., 2022. The MathWorks, Natick, MA, USA.

Mery, N., Marcotte, D., 2022. Quantifying mineral resources and their uncertainty using two existing machine learning methods. Math. Geosci. 54 (2), 363–387. https://doi.org/10.1007/s11004-021-09971-9.

Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H., Mirjalili, S.M., 2017. Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. Adv. Eng. Softw. 114 (12), 163–191. https://doi.org/10.1016/j.advengsoft.2017.07.002.

Neff, J.M., 1987. Biological effects of drilling fluids, drill cuttings and produced waters. In: Long-Term Environmental Effects of Offshore Oil and Gas Development. CRC Press, pp. 479–548.

Okada, N., Maekawa, Y., Owada, N., Haga, K., Shibayama, A., Kawamura, Y., 2020. Automated identification of mineral types and grain size using hyperspectral imaging and deep learning for mineral processing. Minerals 10 (9), 809. https://doi.org/10.3390/min10090809.

Patel, A.K., Chatterjee, S., Gorai, A.K., 2019. Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades. Earth Sci. Inform. 12 (2), 197–210. https://doi.org/10.1007/s12145-018-0370-6.

Perez, C.A., Estévez, P.A., Vera, P.A., Castillo, L.E., Aravena, C.M., Schulz, D.A., Medina, L.E., 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. https://doi.org/10.1016/j.minpro.2011.07.008.

Phoon, K. K., Zhang, W., 2023. Future of machine learning in geotechnics. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 17 (1), 7–22. 10.1080/17499518.2022.2087884.

Qi, C., Tang, X., 2018. A hybrid ensemble method for improved prediction of slope stability. Int. J. Numer. Anal. Methods Geomech. 42 (15), 1823–1839. https://doi.org/10.1002/nag.2834.

Quan, Q., Hao, Z., Xifeng, H., Jingchun, L., 2020. Research on water temperature prediction based on improved support vector regression. Neural Comput. Appl. 34 (11), 8501–8510. https://doi.org/10.1007/s00521-020-04836-4.

Ramil, A., López, A.J., Pozo-Antonio, J.S., Rivas, T., 2018. A computer vision system for identification of granite-forming minerals based on RGB data and artificial neural networks. Measurement 117 (5), 90–95. https://doi.org/10.1016/j.measurement.2017.12.006.

Rodriguez, J.D., Perez, A., Lozano, J.A., 2009. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 32 (3), 569–575. https://doi.org/10.1016/j.measurement.2017.12.00.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Starr, R.C., Ingleton, R.A., 1992. A new method for collecting core samples without a drilling rig. Ground Water Monit Remediat. 12 (1), 91–95. https://doi.org/10.1111/j.1745-6592.1992.tb00413.x.

Sun, T., Chen, F., Zhong, L., Liu, W., Wang, Y., 2019. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, Eastern China. Ore Geol. Rev. 109 (6), 26–49. https://doi.org/10.1016/j.oregeorev.2019.04.003.

Tanaka, S., Tsuru, H., Someno, K., Yamaguchi, Y., 2019. Identification of alteration minerals from unstable reflectance spectra using a deep learning method. Geosciences 9 (5), 195. https://doi.org/10.3390/geosciences9050195.

Thompson, S., Fueten, F., Bockus, D., 2001. Mineral identification using artificial neural networks and the rotating polarizer stage. Comput and Geosci. 27 (9), 1081–1089. https://doi.org/10.1016/S0098-3004(00)00153-9.

Vapnik, V., 1995. The nature of statistical learning theory. Springer, Berlin.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2 (1-3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.

Yadav, S., Shukla, S., 2016. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: In 2016 IEEE 6th International Conference on Advanced Computing (IACC). IEEE, pp. 78–83.

Yu, Z., Shi, X., Zhou, J., Chen, X., Miao, X., Teng, B., Ipangelwa, T., 2020. Prediction of blast-induced rock movement during bench blasting: use of gray wolf optimizer and support vector regression. Nat. Resour. Res. 29 (2), 843–865. https://doi.org/10.1007/s11053-019-09593-3.

Zhang, Y., Chen, H., Yang, B., Fu, S., Yu, J., Wang, Z., 2018. Prediction of phosphate concentrate grade based on artificial neural network modeling. Results Phys. 11, 625–628. https://doi.org/10.1016/j.rinp.2018.10.011.

Zhang, W., Li, H., Li, Y., Liu, H., Chen, Y., Ding, X., 2021. Application of deep learning algorithms in geotechnical engineering: a short critical review. Artif. Intell. Rev. 54 (8), 5633–5673. https://doi.org/10.1007/s10462-021-09967-1.

Zhang, W., Gu, X., Tang, L., Yin, Y., Liu, D., Zhang, Y., 2022. Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: Comprehensive review and future challenge. Gondwana Res. 109 (9), 1–17. https://doi.org/10.1016/j.gr.2022.03.015.

Zhou, J., Qiu, Y., Zhu, S., Armaghani, D.J., Li, C., Nguyen, H., Yagiz, S., 2021a. Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. Eng. Appl. Artif. Intell. 97 (1), 104015 https://doi.org/10.1016/j.engappai.2020.104015.

Zhou, J., Qiu, Y., Armaghani, D.J., Zhang, W., Li, C., Zhu, S., Tarinejad, R., 2021b. Predicting TBM penetration rate in hard rock condition: a comparative study among six XGB-based metaheuristic techniques. Geosci. Front. 12 (3), 101091 https://doi.org/10.1016/j.gsf.2020.09.020.

Zhou, J., Shen, X., Qiu, Y., Li, E., Rao, D., Shi, X., 2021c. Improving the efficiency of microseismic source locating using a heuristic algorithm-based virtual field optimization method. Geomech. Geophys. Geo-Energy Geo-Resour. 7 (3), 1–18. https://doi.org/10.1007/s40948-021-00285-y.