Applied Geochemistry 146 (2022) 105443

Contents lists available at ScienceDirect

Applied Geochemistry

journal homepage: www.elsevier.com/locate/apgeochem

Classification of uranium ore concentrates applying support vector machine to spectrophotometric and textural features☆

M. Marchetti [a,c], L. Fongaro [b,*], A. Bulgheroni [b], M. Wallenius [b], K. Mayer [b]

[a] European Commission, Joint Research Centre (JRC), Ispra, Italy
[b] European Commission, Joint Research Centre (JRC), Karlsruhe, Germany
[c] Karlsruhe Institute of Technology, Institute for Nuclear Waste Disposal, P.O. Box 3640, 76021, Karlsruhe, Germany

ARTICLE INFO

Editorial handling by Dr T J. Gallegos

Keywords:
Nuclear forensics
Uranium ore concentrates
Colour analysis
Image texture analysis
Angle measure technique
Grey-level Co-Occurrence matrix
Cluster analysis
Support vector machine

ABSTRACT

Uranium ore concentrates (UOCs) are produced in the early stages of the nuclear fuel cycle, prior to conversion to uranium hexafluoride. Because of their high uranium content and the large-scale production, UOCs diversion from civilian use and proliferation are potential risks. This implies the necessity to develop methods able to recognise characteristic parameters correlating each UOC powder to its history and origin. Here, a novel methodology is proposed: first the reflectance spectra of 79 commercial UOCs are acquired and clustered by means of Ward's clustering analysis, then classified by Support Vector Machine (SVM). Second, SVM classification is applied to the image textural features extracted with the Grey Level Co-occurrence Matrix (GLCM) and the Angle Measure Technique (AMT) algorithms for powders in two different colour groups. The developed SVM models present good classification quality: a Matthews correlation coefficient (MCC) of 0.95 is obtained for the classification based on colours while macro-F1 is generally greater than 0.81 (MCC larger than 0.75) for the texture-based classification. These results reveal the potentiality of the present automated classification for the scopes of nuclear forensics in the identification of an unknown uranium ore concentrate sample.

1. Introduction

The term uranium ore concentrates (UOCs) refers to the product resulting from the mining, milling and leaching of the uranium ores in the front-end of the nuclear fuel cycle. The manufacturing process considers precipitation after purification by means of solvent extraction or ion-exchange; different types of precipitants can be used, the most common are gaseous ammonia, ammonium hydroxide or hydrogen peroxide (Ho, 2015). The UOC powders contain about 60–80 % uranium in different chemical compositions (e.g. ammonium diuranate, uranyl hydroxide, uranyl peroxide, or uranium oxide when calcination is performed) (Kristo et al., 2016) and they are produced in large amounts, from which stems a risk of illicit appropriation. The illicit trafficking of nuclear material represents a potential threat to the population, in particular from the proliferation viewpoint. Hence, it is of paramount importance to ensure, in addition to a proper management and control, effective detection of nuclear material smuggling.

Upon detection of illegal possession or trafficking of UOCs, the understanding of the origin and the history of the material must be promptly achieved to ensure legal procedures and strengthen controls. This tracing process belongs to the nuclear forensics investigations, which are based on the study of characteristic parameters having a robust relationship with the material history. These parameters are known as nuclear "fingerprints" or "signatures" and are generally the result of a particular production process or distinctive of a uranium mine. Nuclear forensics applies the experience developed in other fields (e.g. radiochemistry, nuclear physics, material science) to identify these signatures; concerning UOCs, examples are represented by the ratios $^{234}U/^{238}U$ and $^{234}U/^{235}U$ (Keegan et al., 2008; Brennecka et al., 2010; Spano et al., 2017), strontium and lead or neodymium isotope ratio (Varga et al., 2009; Krajkó et al., 2014), the content and variation over time of ^{232}Th and its daughter ^{228}Th (Varga et al., 2011a) or the content of impurities (Keegan et al., 2014; Kennedy et al., 2013).

However, nuclear forensics investigations are generally time-consuming and a long time is required to collect and process the data. Thus, the combination of techniques capable to extract signatures

☆ This document is the results of the research project funded by the European Commission, Joint Research Centre under the Exploratory Research Programme.
* Corresponding author.
E-mail addresses: mara.marchetti@kit.edu (M. Marchetti), lorenzo.fongaro@ec.europa.eu (L. Fongaro), antonio.bulgheroni@ec.europa.eu (A. Bulgheroni), maria-s.wallenius@ec.europa.eu (M. Wallenius), klaus.mayer@ec.europa.eu (K. Mayer).

https://doi.org/10.1016/j.apgeochem.2022.105443
Received 15 February 2022; Received in revised form 11 August 2022; Accepted 29 August 2022
Available online 7 September 2022

reducing the processing time, in conjunction with automated classification, would significantly shorten the treatment delay and allow for faster decision-making.

The discriminating power of the UOCs colour has not found a systematic application until recently. Some studies considered the colour as a simple indicator of the process root (i.e., type of reagent, separation procedure or drying conditions) (Klunder et al., 2013); while others restricted its use to a general description (e.g., the "green powder" reported in (Keegan et al., 2014)). One exception is the study of Thompson et al. (2021) where they proposed a methodology to objectively assess the colour of uranium dioxide powder from digital images and discussed its application in the nuclear forensics field. On the other hand, the morphological peculiarities originated by a specific UOC production route (e.g., precipitating reagents and conditions) have found more consideration so far. Manna and co-workers in (Manna et al., 2012a) reported the different particle size distributions and microstructures of ammonium diuranate and $UO_3 + U_3O_8$ powders at different stages of precipitation, highlighting also the influence of gaseous versus aqueous ammonia used for precipitation. The formation of pores on the primary platelet of uranium oxide calcined at temperatures greater than 550 °C was also observed by Manna and co-workers in (Manna et al., 2012b). In the field of nuclear forensics, morphological examinations were used by Keegan et al. in (Keegan et al., 2014) to investigate the origin of a confiscated uranium ore concentrate sample in Australia.

In this context, image analysis (IA) was demonstrated to be a suitable technique for morphological characterisation of nuclear materials (Tamasi et al., 2016a, 2016b; Olsen et al., 2017; Hanson et al., 2019); however, the application of the proposed methods to UOCs requires laborious and time-consuming steps both for sample preparation and evaluation. Hence, strategies to overcome this drawback for powders characterisation were explored. In particular, Fongaro et al. (2016) and Ho Mer Lin (Ho, 2015) tested the potentiality of the image texture analysis, specifically the AMT algorithm combined with chemometrics, for the classification of some SEM images of UOCs: this technique can provide information related to the bulk powder environment, namely how the particles are arranged together and their size distribution.

Due to the high number of commercial UOCs, the necessity to combine different analytical techniques was perceived and a novel strategy was implemented and preliminarily explored. In particular, spectrophotometry and image texture analysis were used to assess the existence of defined patterns in the extracted data, which could be used to categorise the UOCs with proper analysis (Marchetti et al., 2020).

The present work focuses on the results obtained by applying machine learning (specifically, SVM) to classify a set of 79 industrial UOCs. A first classification, performed with spectrophotometric variables, enables the attribution of an unknown powder to a predefined colour group. A second classification is successively tested for the powders belonging to two colour groups; in this case, image textural features extracted with AMT and GLCM algorithms are employed.

2. Materials

The specimens investigated in the present experimental campaign are 79 UOC commercial samples mostly collected between 1950 and 2000. The powders exhibit colours grading from light yellow (white-yellow), yellow and orange to brown and black (see Fig. 1). A detailed list of the examined samples is available in Table 1, together with their chemical composition.

The reported composition of the samples was confirmed by infrared (IR) spectroscopy in (Varga et al., 2011b) and successively in 2014 by Raman and IR spectroscopy (Ho, 2015). In both cases, oxidation was not observed. Moreover, in (Ho, 2015), UOCs of different chemical compositions were synthesized and IR and Raman spectra of these compounds were acquired. These spectra compared well to the spectra of corresponding industrial samples, i.e., the latter did not show significant signs of oxidation or hydrolysis. For spectrophotometric investigations, the samples were introduced in quartz vials (75 × 10 mm) or borosilicate glass vials (45 × 14.7 mm); the powders in the vials were analysed

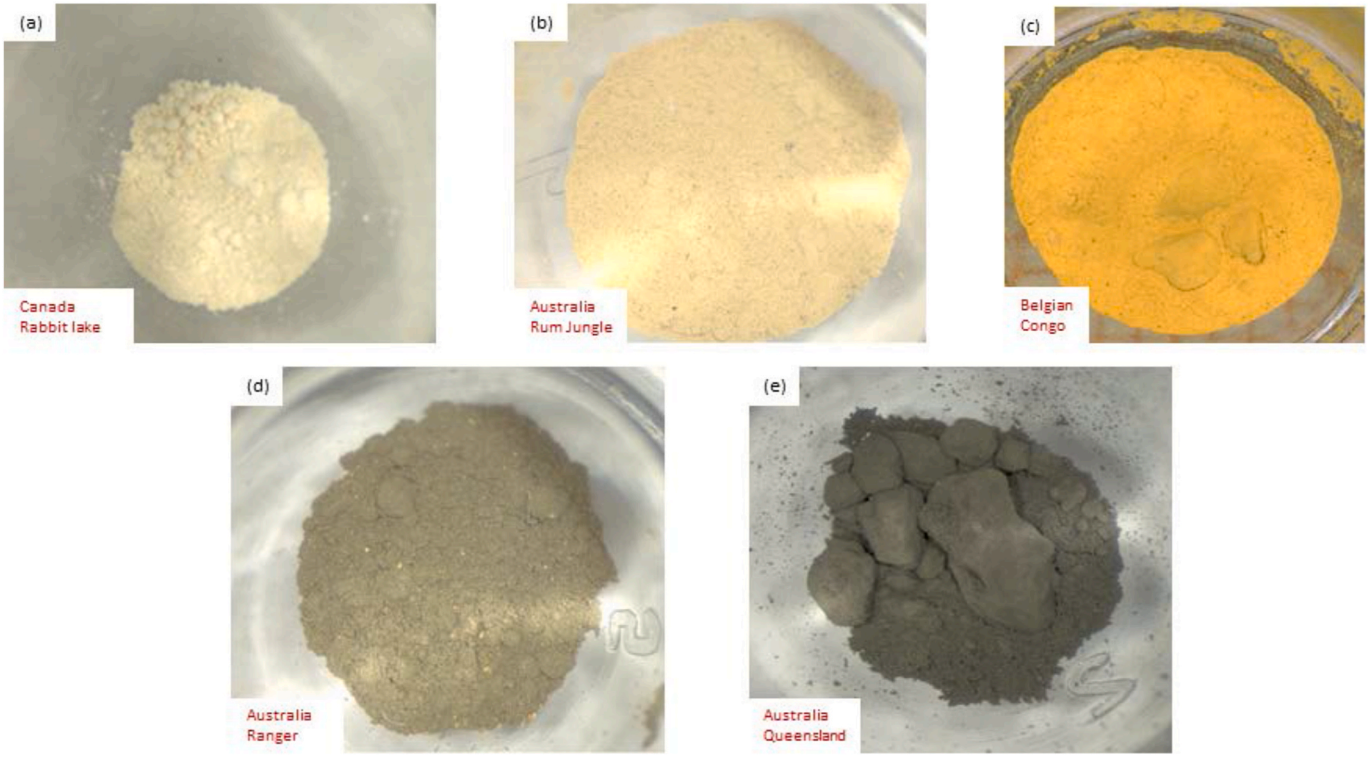


Fig. 1. Subset of the samples under investigation. The powders shown are representative of some of the different colours possessed by the UOCs. (a) Canada Rabbit lake, (white-yellow) (b) Australia Rum Jungle (yellow) (c) Belgian Congo (orange), (d) Australia Ranger (brown) and (e) Australia Queensland (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
UOCs measured in the present experimental campaign. The asterisk (*) indicates uncertainty concerning the reported chemical composition.

Number	Sample (Country-Facility/Mine)	Chemical composition	Number	Sample (Country-Facility/Mine)	Chemical composition
1	England-Wheal Edward	Unknown	41	Canada-North Span	$UO_2(OH)_2$
2	USA-Kerr MgGee	$(NH_4)_2U_2O_7$	42	USA-Utah	$(NH_4)_2U_2O_7$ + U_3O_8
3	Spain-Jen	$(NH_4)_2U_2O_7$	43	Australia-Ranger*	$Na_2U_2O_7$
4	Australia-Yeelirrie*	Mixture	44	Germany-Brunhilde	$(NH_4)_2U_2O_7$
5	South Africa-Rossing	U_3O_8	45	Germany-Ellweiler	$(NH_4)_2U_2O_7$
6	South Africa-Mindola	Unknown	46	Niger-Somair	$Na_2U_2O_7$
7	USA-Cotter	$Na_2U_2O_7$	47	Portugal	Unknown
8	South Africa-EFI(Mouand)	$(NH_4)_2U_2O_7$	48	USA-Lucky McGill	$(NH_4)_2U_2O_7$
9	USA-Pathfinder	$UO_2(OH)_2$	49	USA-Everest Black	$(NH_4)_2U_2O_7$ + Mixed Oxide
10	Canada-Stanrock	$(NH_4)_2U_2O_7$	50	Canada-Rio Algom	$(NH_4)_2U_2O_7$
11	Holland-Delft	Mixture	51	Canada-Rabbit Lake	$UO_4.nH_2O$ + U_3O_8
12	Mozambique-Mavuzi	Unknown	52	USA-Everest Yellow	$UO_4.nH_2O$
13	South Africa-Palabora	U_3O_8	53	Sweden-Ranstadt	$Na_2U_2O_7$
14	Belgium-Belgian Congo*	$UO_2(OH)_2$	54	USA-El Mesquite	$UO_4.2H_2O$
15	Brazil-Nuclebras	$(NH_4)_2U_2O_7$	55	USA-Union Carbide	Mixed Oxide
16	Spain-Enusa	$(NH_4)_2U_2O_7$ + Oxide	56	Canada-Denison	$Na_2U_2O_7$
17	Australia-Queensland	U_3O_8	57	USA-Atlas	U_3O_8 + oxide
18	Germany-Wismut	$(NH_4)_2U_2O_7$	58	Australia-Mary Kathleen	U_3O_8
19	USA-Yankee Yellow	$Na_2U_2O_7$	59	USA-United Uranium	AU
20	Canada-Dyno	AU	60	USA-South Dakota	Mixed Oxide + AU
21	Canada-Key Lake	U_3O_8	61	Argentina	$Na_2U_2O_7$
22	China-Hengyang	U_3O_8+UO_2	62	USA-Federal American Partners	U_3O_8
23	USA-Petromic	Mixed Oxide	63	USA-Dawn	$(NH_4)_2U_2O_7$
24	Canada-Blind River	$Na_2U_2O_7$	64	Canada-Milliken Lake	$(NH_4)_2U_2O_7$
25	Canada-Sunnar	AU	65	Australia-Rum Jungle*	$UO_2(OH)_2$
26	Yugoslavia-Spisak Black	Mixed Oxide	66	Canada-El Dorado	$(NH_4)_2U_2O_7$
27	Canada-Faraday	AU	67	Canada-Ray Rock	AU
28	Australia-Olympic Dam	U_3O_8	68	USA-Chevron Hill	$(NH_4)_2U_2O_7$ + oxide
29	Yugoslavia-Spisak Yellow	$(NH_4)_2U_2O_7$	69	S.Africa-Nufcor	U_3O_8
30	USA-Mulberry	Oxide + $UO_2(OH)_2$	70	Russia-Techsnab	U_3O_8
31	USA-Falls City	$(Na)_2U_2O_7$ + Mixed Oxide	71	USA-United Nuclear	$(NH_4)_2U_2O_7$
32	Canada-Madawaska	$(Na)_2U_2O_7$	72	Yugoslavia-Rudnik	$(NH_4)_2U_2O_7$
33	USA-Irigaray	$UO_4.nH_2O$	73	Australia-South Alligator	$UO_2(OH)_2$
34	USA-ESI	$(NH_4)_2U_2O_7$	74	USA-Vermont Yankee Black	Mixed Oxide
35	Canada-Macassa	AU	75	USA-Sesquehanna	$UO_2(OH)_2$ + Oxide
36	USA-Anaconda	$(NH_4)_2U_2O_7$	76	Romania	$Na_2U_2O_7$ + Oxide
37	USA-Shirley Basin	U_3O_8	77	USA- EFI	$(NH_4)_2U_2O_7$
38	USA-Mobil	$UO_4.nH_2O$	78	USA-Homestake	$(NH_4)_2U_2O_7$
39	Canada-ESI	$UO_2(OH)_2$	79	Canada-Stanleigh	$(NH_4)_2U_2O_7$
40	Australia-Radium Hill	$(NH_4)_2U_2O_7$			

without further manipulation. This was considered the best option to reflect the UOCs original state.

For scanning electron microscopy (SEM) image acquisition, powders were poured into a 1-cm diameter graphite container (Figure S1 in Supplemental Material) and gently pressed with a weight of 1 g, to flatten the powder surface, thus obtaining a more uniform focusing. Each sample was prepared in three replicas, using three different sample holders, in order to take into consideration the effect of the sample preparation in the models' development. The amount used in the sample holder depended on the sample availability; however, attention was paid that the carbon stub at the bottom of the cup was fully covered.

3. Experimental

3.1. Spectrophotometry

The measurement of the specimens' colour was achieved by means of a Konika Minolta CM-700d sphere-type spectrophotometer equipped with a silicon photodiode array detector, an integrating sphere having a diameter of 40 mm and a xenon lamp with a UV cut filter. The instrument operates in the visible range of the electromagnetic spectrum (360–740 nm).

The light reflected by the sample, collected in the integrating sphere, is normalized to the zero reflection condition and to a pure white standard (100 % reflection). In the present experimental campaign two types of data were collected: the reflectance values as a function of the wavelength and L^*, a^*, b^* in the CIE 1976 L*a*b* colour space (Schanda, 2007).

L* is the lightness index and goes from 0 to 100, a* varies from green to red index and b* from blue to yellow. In the system used in the present experimental campaign, a* and b* are bound in the range [−60, 60].

3.2. Scanning electron microscopy

The scanning electron microscope enables images to be acquired by raster scanning a surface with a focused electron beam. The SEM generally offers a larger depth of field with respect to optical microscopy, a higher resolution resulting from the shorter electron wavelength and can achieve much higher magnifications. Considering the first two advantages, SEM images of the UOC samples were acquired by using a FIB/SEM FEI Versa 3D in low-vacuum mode (pressure = 10 Pa) equipped with a concentric backscattered detector for backscattered electrons.

As the absence of conductive coating caused powder charging during

the imaging phase, the accelerating voltage was reduced to 5 kV, while three different magnifications of 100×, 250 x and 500 x, with a dimension of 3072 × 2048 pixels, were acquired. These magnifications were chosen by considering the good classification performance achieved by using image texture analysis in the pilot experiment (Fongaro et al., 2016).

To further reduce artefacts on the images, integration mode (16 frames to obtain an image), with a dwell time of 300 ns was chosen. For each repetition five subregions were selected (See Figure S1 in Supplemental Material) resulting in a total of 15 independent images per magnification.

4. Image texture analysis

The topography of a surface, in particular its roughness/smoothness, or the regularity in the repetition of a feature within a physical object (e.g white and black squares on a chessboard), is converted into specific colour or grey-level distributions and patterns when captured by an image. The relationships existing among the intensities, frequencies and spatial distributions of the pixels in the image are encompassed by the concept of Image texture (Fongaro et al., 2016; Russ, 1999). The extraction of the image properties can be achieved by applying different textural operators mostly classifiable in statistical, structural, model-based or signal-processing based operators (Hung et al., 2019). In the present work, both a statistical and a signal processing approach were used, respectively the Grey Level Co-occurrence Matrix (GLCM) and the Angle Measure Technique (AMT), separately or combined.

4.1. Grey level Co-occurrence matrix

The GLCM is one of the first methods of the image texture analysis and was introduced by Haralick and co-workers in 1973 (Haralick et al., 1973). A grey level co-occurrence matrix contains frequency information about the grey levels of pixel-pair (i,j). A short description of the GLCM, together with the descriptors used in the present work, is offered in the Supplemental Material; however, a short explanation follows for easier interpretation of the results.

The homogeneity of the image is described by the *Angular Second Moment*, *Homogeneity* and the *Inverse Difference Moment* which are characterised by large values for uniform images with low grey-level difference between adjacent pixels. Inversely, *Contrast* increases when the difference between the intensities of neighbouring pixels is large. A measure of the dispersion from the average value of cell values within the GLCM is given by the *Variance*. This descriptor increases when the grey levels differ from their means and measures the heterogeneity in a manner similar to *Contrast* (Yang et al., 2012). *Cluster Shade* and *Cluster prominence* are respectively measures of the skewness and asymmetry of the GLCM (Yang et al., 2012). *Correlation* defines the linear correlation existing among neighbouring pixels and it is in the interval [−1,1] where the extreme values represent a perfectly negative (−1) or positive (+1) correlation and 0 indicates lack of correlation (Pantic et al., 2013). Finally, *Entropy* quantifies the level of disorder in the image.

Here, GLCM was calculated at a distance $\delta = 1$ and at 0°, 45°, 90° by means of the "GLCM Texture Too" v.0.009 plugin for ImageJ 1.51j8 (Schneider et al., 2012), which represents the updated version of "GLCM Texture" v.0.4, created by Julio E. Cabrera (Sivchenko et al., 2016).

4.2. Angle Measure Technique

The second algorithm used is the AMT, which is a powerful tool for the analysis of single and two-dimensional signals and is capable of describing the complexity of the signal simultaneously on all the existing scales through the Mean Angle (MA) spectrum (Kucheryavski et al., 2008a).

The AMT was introduced by Andrle (1994) for geomorphic coastlines, but it finds useful application in the description of grey-level images which are mostly isotropic, i.e, do not present a preferential direction in the appearance of the characteristic features. The algorithm works in three main steps: the first one regards the unfolding process; the second regards the MA calculation and its representation in a graph (AMT spectra); in the third step chemometrics techniques must be applied to elaborate the mean angle spectral data. Detailed descriptions of the AMT algorithm are available in (Fongaro et al., 2016; Andrle, 1994; Fongaro and Kvaal, 2013; Huang and Esbensen, 2000; Esbensen et al., 1996; Kvaal et al., 2008; Halstensen et al., 2019). In this work a maximum scale of 1000 pixels and a sampling of 1000 points were used as AMT algorithm set up. The image treatment was performed by using the jAMTexplorer plug-in for ImageJ (AMT plugin for ImageJ, 2014).

5. Cluster analysis and support vector machine

In the present work Cluster Analysis and Support Vector Machine were performed by using PLS Toolbox version 8.6.1 (Eigenvectors Research, Inc., USA) for Matlab 2017a (The Matworks Inc., Natick, MA, USA). In addition to the above-mentioned methods which were directly used to develop the model, exploratory data analysis through principal components analysis (PCA) was also performed. PCA performs linear combinations of the observed variables which are defined as "principal components" (PCs). These PCs explain most of the variance in the original variables. The coefficients of the linear combination are the loadings (weights) corresponding to each extracted principal component (Denis, 2015).

5.1. Cluster analysis

Cluster analysis applies a "similarity degree" criterion to group individual objects; in particular it is used in descriptive statistics to observe if subgroups exist within a set of data. The "similarity degree" criterion depends upon the method employed; in general, it is based on the maximisation of the differences among the various groups (Hastie et al., 2009; Härdle and Simar, 2007). Here, the hierarchical Ward's method was applied. This method joins two clusters while minimising the increase in the sum of the square errors. The criteria used by the Ward's method corresponds to the minimisation of the within-cluster distance which is normally calculated as the Euclidean distance between data in the cluster and the centroid (the point whose distance from each other point of the cluster is minimum) (Rencher, 2002; Strauss and von Maltitz, 2017).

5.2. Support vector machine

Support vector machine is a machine learning algorithm initially developed by Vapnik and co-workers (Boser et al., 1992; Vladimir, 1998) and successively implemented/extended to create the so-called *kernel-based methods*. In classification, SVM generates a function, by learning from the training dataset, to find the best hyperplane which maximises the distance among the different classes of data.

In prediction, SVM relies only on a subgroup of the original dataset, whose elements are defined as support *vectors* and are responsible for the margin selection. The separation can be linear or Kernel-based; in the second case, a non-linear mapping to a higher-dimensional space is performed such that a linear separation in this new space is possible. SVM solves the optimisation problem while minimising the error, or a cost function, which depends on the original set of data. In fact, it uses the *structural risk minimisation* (SRM) concept, which consists in a set of models with increasing levels of complexity, balancing between a low complex model (high error) with a possible underfitting condition and a complex model with low error, but consequent overfitting (Awad and Khanna, 2015). The algorithm is commonly known for its high generalisation performance and robustness against non-optimal features (Abe, 2010).

6. Results

6.1. Colour-based classification model

The definition of colour clusters has the advantage of reducing candidates for the morphological examination and categorisation to those within specific colour groups. A possible additional benefit is represented by the simplicity of the discriminating criteria which can provide support for conclusions drawn with different analytical techniques without excessive delays.

The spectrophotometric measurements of the powders enabled 39 spectral reflectance values to be acquired in the wavelength range 360–740 nm, together with their corresponding L^*, a^*, b^* values. Hierarchical cluster analysis was then applied to the obtained data matrix, composed by the 79 samples of Tables 1 and 42 variables.

Fig. 2 shows the connection dendrogram for the Ward's linkage of the data pretreated with the autoscale function. The vertical axis indicates the samples to be clustered, while the horizontal position of the bar displays the sum of the squared errors which, in turn, is related to the between-cluster distance (Rencher, 2002).

The clusters can be chosen by selecting a threshold, by intersecting the horizontal lines with a sliding vertical line; the intersection points define the number of clusters. To reduce the number of very small groups, six colour-classes (from colour class 1 to colour class 6; CC1 ÷ CC6) were chosen and they were labelled as *Black*, *Brown*, *Orange*, *Dark-Yellow*, *Yellow* and *White-Yellow*. It is worthwhile to highlight that peroxides belong mostly to the *White-Yellow* colour group, ammonium diuranate, sodium diuranate and hydroxides to the *Yellow*, *Dark-Yellow* and *Orange* groups, mixed compounds and unknown compositions fall mainly in the *Brown* cluster whilst the *Black* group contains mostly oxides. This categorisation remarks the capability of the cluster analysis to intercept variations in the powders' chemical composition, which is reflected in their colours. Differences in the colour nuances for powders having the identical chemical composition (e.g. yellow and light yellow) might be the consequence of specific production parameters or amount of impurities; however, the process root must be known in detail to perform a classification based on this discriminating factor.

The difference between the clusters was also measured by calculating ΔE, which is defined as the Euclidean distance (equation (1)) between different points in the L*, a*, b* colour space.

$$\Delta E = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \quad (1)$$

The average values were calculated for each colour class, then the difference was measured between *Black-Brown*, *Brown-Orange*, *Orange-Dark-Yellow*, *Dark-Yellow-Yellow* and *Yellow-White-Yellow*. Table 2 reports details about the six colour classes, including the colour labels, the number of UOC samples per colour-class and the range of each colour index; generally a result of ΔE value less than 2 indicates that the compared colours can be considered perceptually equivalent.

The classes selected after the clustering analysis were used as category variables for a PCA. The position of the samples in the score plot is mainly observable along the PC1, that explains the 83.52 % of the variance, while PC2 explains the 11.87 %. In particular, looking at the loadings plot in Fig. 3 (b), the position of each colour class into the scores plot of Fig. 3 (a) can be better explained. In the *Black* colour class (CC1), there are samples characterised by the lowest b* and L* values (the darkest samples). The separation of the colour classes along the PC1 is less explained by the value of a*, whose projection on the PC1 is close to zero. Moving from the negative to the positive values on the PC1 the colour becomes brighter and richer in the yellow component. The separation on the PC2 is explained better, among the other variables, by a*, shifting from a larger red component to a greenish one; this trend is also reproduced by the wavelengths which are longer (maximum = 740 nm) for negative values on the PC2 and shorter (minimum = 460 nm) for the largest values on the same PC.

Afterwards, the six classes were used as category variables for the SVM colour-based classification. SVM implemented with LIBSVM library

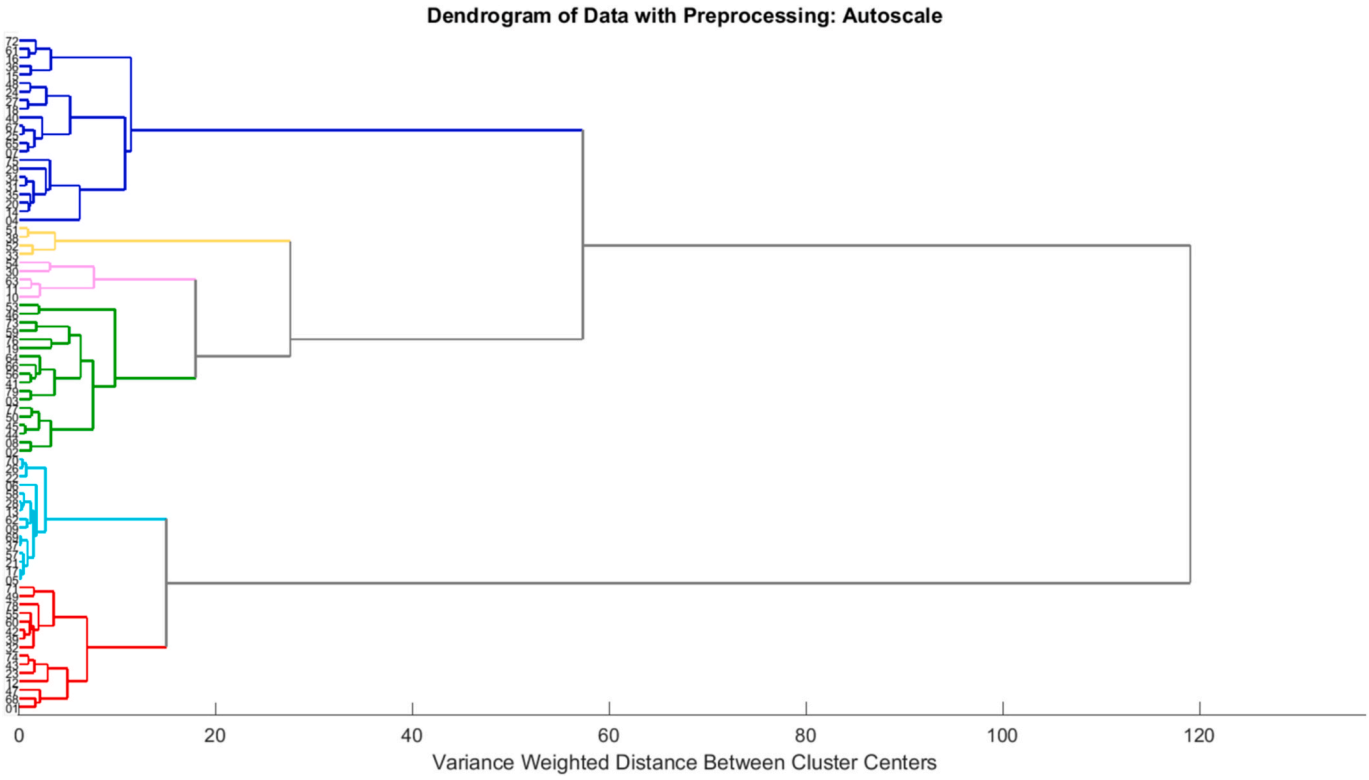


Fig. 2. Hierarchical cluster analysis of the entire samples collection. Different clusters are possible by selecting the position on the dendrogram. In this work six different colour-groups were chosen. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Details about the six colour classes obtained after the cluster analysis: average values of L*, a* and b* indexes with their range and the corresponding standard deviation. ΔE values calculated among neighbouring classes, starting from CC1 are reported together with the colour class labels.

Sample	Class Label	N° of samples	L*	MINL*	MAXL*	a*	MINa*	MAXa*	b*	MINb*	MAXb*	ΔE
(CC1)	*Black*	15	44.4 ± 2.1	40.4	47.1	−0.7 ± 1.9	−2.6	4.9	6.9 ± 3.0	1.7	11.1	-
(CC2)	*Brown*	15	55.9 ± 4.0	47.1	60.8	4.0 ± 4.2	−3.0	11.0	20.6 ± 8.1	1.7	35.4	18
(CC3)	*Orange*	22	69.3 ± 3.9	61.9	74.4	10.3 ± 3.1	1.3	15.8	37.7 ± 7.8	17.7	52.2	22
(CC4)	*Dark-Yellow*	18	79.3 ± 2.9	75.0	84.4	11.0 ± 5.2	6.24	25.4	48.2 ± 5.9	34.2	55.9	14
(CC5)	*Yellow*	5	84.0 ± 2.4	81.2	87.7	4.6 ± 4.1	0.01	9.3	46.9 ± 3.5	43.0	52.5	8
(CC6)	*White-Yellow*	4	87.1 ± 1.3	85.4	88.2	−1.2 ± 1.6	−3.5	−0.2	30.1 ± 2.4	27.3	33.0	18

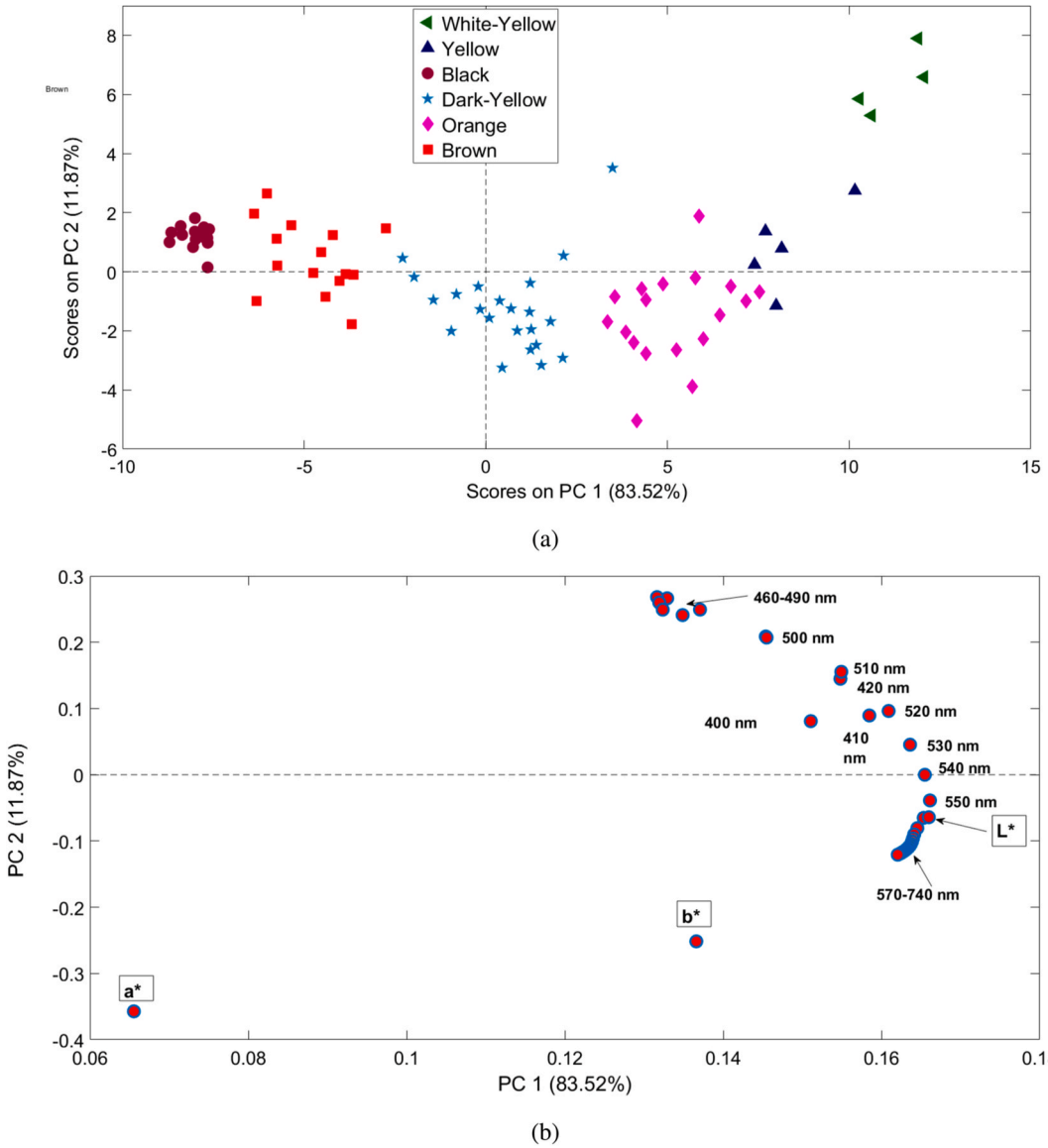


Fig. 3. (a) Principal components analysis score plot for the samples categorised in six colour classes. (b) Loadings plot of the same PCA depicting the weight of the 42 variables on the two principal components. The three arrows indicate, from the top to the bottom, the group of points corresponding to wavelengths in the interval 460–490 nm, the L* index and the points in the 570–740 nm range. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

and the Radial Basis Function Kernel was employed on the data pre-processed with the autoscale function and cross-validated by means of the "Venetian blinds"; a maximum cost of 50 was selected (50 was chosen among the lower costs having the best performance) and $\gamma = 0.1$. Table 3 lists the per-class sensitivity and the specificity values obtained in cross-validation with the present model. From Table 2 and Fig. 2 one can observe the distribution of elements among the six colour classes: two groups, namely *White-Yellow* and *Yellow*, are smaller than the others; hence, the present SVM classification deals with an unbalanced training dataset. The most common metrics used to evaluate the machine learning model (e.g., Accuracy and F1 score) are biased if the distributions among the classes is skewed; this has raised concerns within the Machine Learning community and research on the topic is still ongoing (He and Garcia, 2009; Maratea et al., 2014; Chicco and

Table 3
Evaluation metrics in cross-validation (CV) and prediction (P) for the six colour-group SVM model.

Colour class	*Brown*	*Orange*	*Dark-Yellow*	*Black*	*Yellow*	*White-Yellow*
Sensitivity (CV)	0.93	1	0.96	1	0.80	1
Specificity (CV)	1	0.97	0.98	1	1	1
Sensitivity (P)	1	0.89	1	1	1	1
Specificity (P)	1	1	1	1	0.96	1

Jurman, 2020; Luque et al., 2019). However, several studies have remarked the higher reliability of the Matthews Correlation Coefficient (MCC) in assessing the quality of binary prediction models (Chicco and Jurman, 2020; Luque et al., 2019; Baldi et al., 2000) and this scalar parameter was also used by the US-FDA (MAQC) study for more than 30 000 different models (Shi et al., 2006). MCC possesses the advantage to be an easily interpretable scalar parameter, with natural extension to the multiclass case. The robustness of this metric for multiclass classification was demonstrated by Jurman and co-workers in (Jurman et al., 1371), where they suggest its use as a good compromise in terms of discriminancy, consistency and coherent behaviours also for imbalanced datasets. MCC is in the range [−1,1], where 1 defines the best value, −1 the worst. For the colour-class discrimination model the scikit-learn metrics module in Python (Pedregosa et al., 2011) was used and a MCC of 0.95 for the model in cross-validation computed.

The validation was achieved with a set of new measures for 25/79 UOC powders: in prediction, 24/25 are attributed to the correct class, while one sample in the *Orange* group is classified as *Yellow*. The validation of the model with the blind external dataset results again in a high MCC value of 0.95, which confirms the good quality of the model obtained.

6.2. Texture-based classification model

The categorisation of the UOCs reflectance spectra (colours) allows preliminary discrimination among the 79 UOC samples; this can contribute to reducing the investigation time when an unknown powder is seized. However, the identification process is subordinate to the recognition of other characteristic patterns for the sample under examination. To accomplish this goal, the GLCM and AMT algorithms were

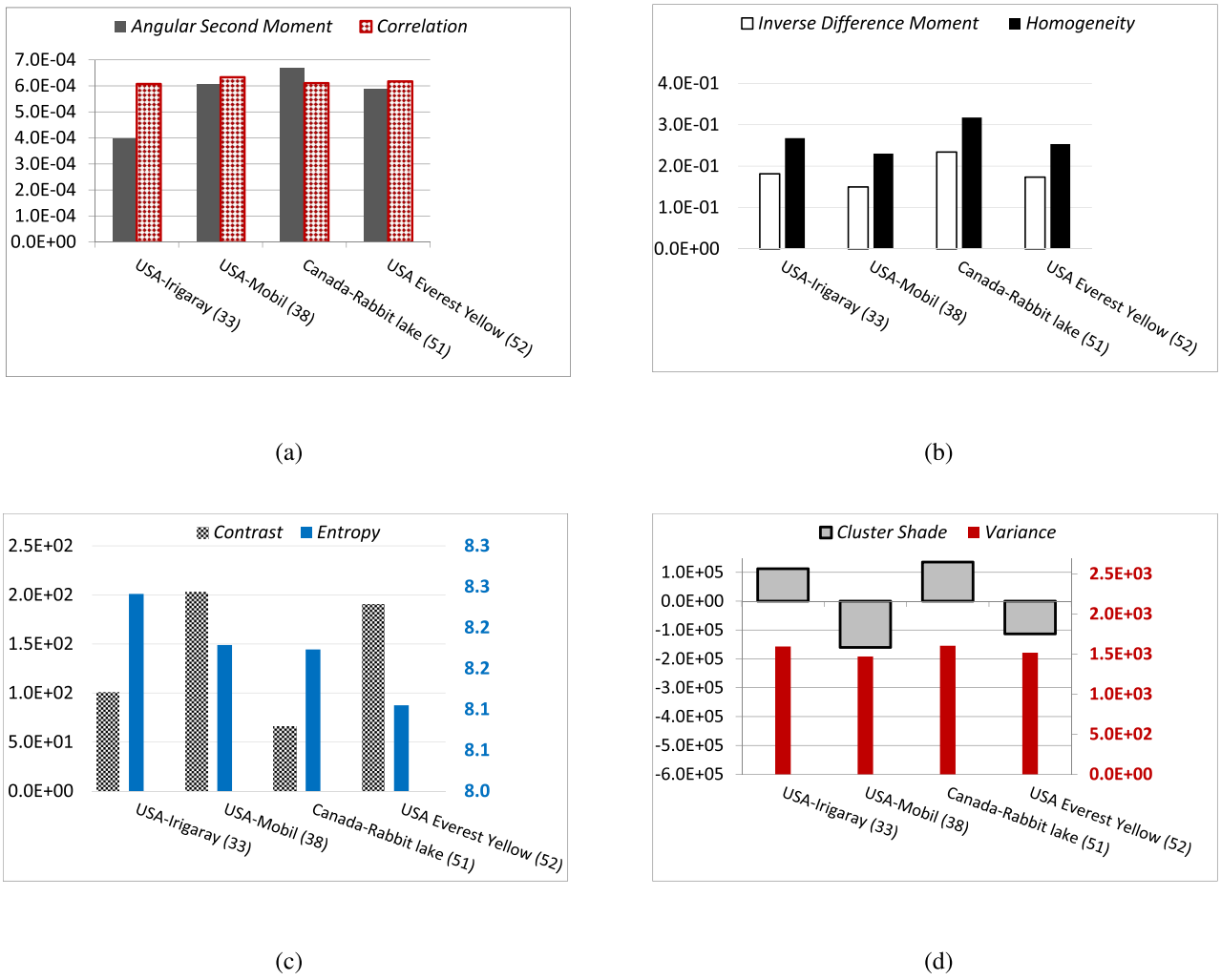


Fig. 4. Average values of the GLCM descriptors for the UOCs images in the *White-Yellow* group at 100 x magnification. The considered angles are $\theta = 0$, 45 and 90. (a) The grey bars define the *Angular Second Moment* values, while the red patterned bars indicate the *Correlation*, (b) The bars with continuous single outline and the black bars refer respectively to *Inverse Difference Moment* and *Homogeneity*, (c) *Entropy* and *Contrast* are defined by the blue bars (right scale) and the patterned black bars, respectively. (d)The red bars (right scale) indicate the *Variance* whilst the light grey outlined bars refer to the *Cluster Shade*. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

employed to extract textural features correlated to the morphological peculiarities of each UOC.

15 images with three magnifications (100×, 250 x and 500 x) and belonging to the *Yellow* (CC5) and *White-Yellow* (CC6) groups of the samples in Table 1 were selected. The images of the samples 33, 38, 51 and 52 (*White-Yellow*) were stacked together and mean-centred to obtain a more uniform contrast prior to the analysis (Fongaro et al., 2016) with "Stack Meancenter4" from NMBU plugin for ImageJ; the same operation was repeated with the samples 10, 11, 54 and 63 (*Yellow*). The training set was selected from 13 images per magnification, while two images (2/15) were used as the external test set.

6.3. Grey level Co-occurrence matrix

Fig. 4 shows the average values of eight GLCM indices calculated for $\theta = 0$, 45 and 90 directions for the images in the *White-Yellow* cluster (100 x). In Fig. 4 (a) and (c), the bars corresponding to the *Angular Second Moment* and *Entropy* would indicate a larger inhomogeneity and randomness in the intensity distribution of the USA-Irigaray (33) images. *Homogeneity* and *Inverse Difference Moment* display an opposite trend with respect to *Contrast*, if the graphs of Fig. 4 (b) and (c) are compared. These results are reasonable if one considers the images in the *White-Yellow* series in Fig. 5: the sample 33, in fact, shows the least monotonous pattern, with grains of different dimensions (the bigger grains have diameter of about 500 μm, while many particles are below 50 μm of size). USA-Mobil (38) displays the lowest *Inverse Difference Moment* and *Homogeneity*, while the *Entropy* level is comparable of that of Canada-Rabbit lake (51): this means that albeit the powder 38 renders heterogeneously distributed grey-levels in the images, the disorder level in their distribution is lower than the level exhibited, for example, by 33. In Fig. 4 (d) *Cluster Shade* is positive for 33 and 51 and negative for 38 and 52: this indicates opposite skewness of the GLCMs for the two pairs of samples. In the same Figure, the distribution of the *Variance* shows the same two pairs of samples as for *Cluster Shade*: 33–51 and 38–52. The couple 33 and 51 has the highest *Variance* values remarking a heterogeneous pattern where larger grains are more interspersed among smaller grains.

The largest *Correlation* is observable in Fig. 4 (a) for the sample 38 meaning that it has a higher linear dependency within the surface in terms of pixel intensities; the *Correlation* is quite comparable for the other three powders.

The average GLCM descriptors values for the images in the *Yellow* cluster are represented in Fig. 6. Here, USA-Dawn (63) has the larger *Angular Second Moment*, *Inverse Difference Moment*, together with the highest *Homogeneity*. This result is probably caused by the presence of very large grains which are often covering the entire field of view (4.14 mm) at the magnification of 100× (see the left-bottom corner of Fig. 7 (d)). The highest dissimilarity is displayed by Canada Stanrock (10) and Holland-Delft (11) respectively in Fig. 6 (a), 6 (b) and 6 (c) (*Angular Second Moment*, *Inverse Difference Moment*, *Homogeneity* and *Contrast*).

The disorder in intensities and the textural irregularities possessed by these samples are also expressed by the high *Entropy* bars in Fig. 6 (c). If one compares the SEM micro-graphs of the four samples of the *Yellow* cluster, can detect smaller grains for the samples 10 and 11 (Fig. 7 (a) and (b)) with respect to 54 and 63 (Fig. 7 (c) and (d)), a fact that can generate a greater variability in terms of grey-levels among contiguous pixels. Furthermore, the *Correlation* index of the powder 54 is the highest, immediately followed by USA-Dawn (63) (Fig. 6 (a)). The GLCM of the sample 11 is strongly negatively skewed (*Cluster Shade*); this descriptor is also negative for 54 and 63, while it is positive for the sample 10. The samples in the *Yellow* group possess comparable *Variance* values.

6.4. Angle Measure Technique

The AMT algorithm was applied to images initially cropped and then reduced in dimension (Kucheryavski et al., 2008b) to 1000 × 600 with a selected sampling set of 2 %, chosen for faster calculation. The interpretation of the mean angle spectra is less straightforward with respect to the GLCM metrics. However, one can generically affirm that at the

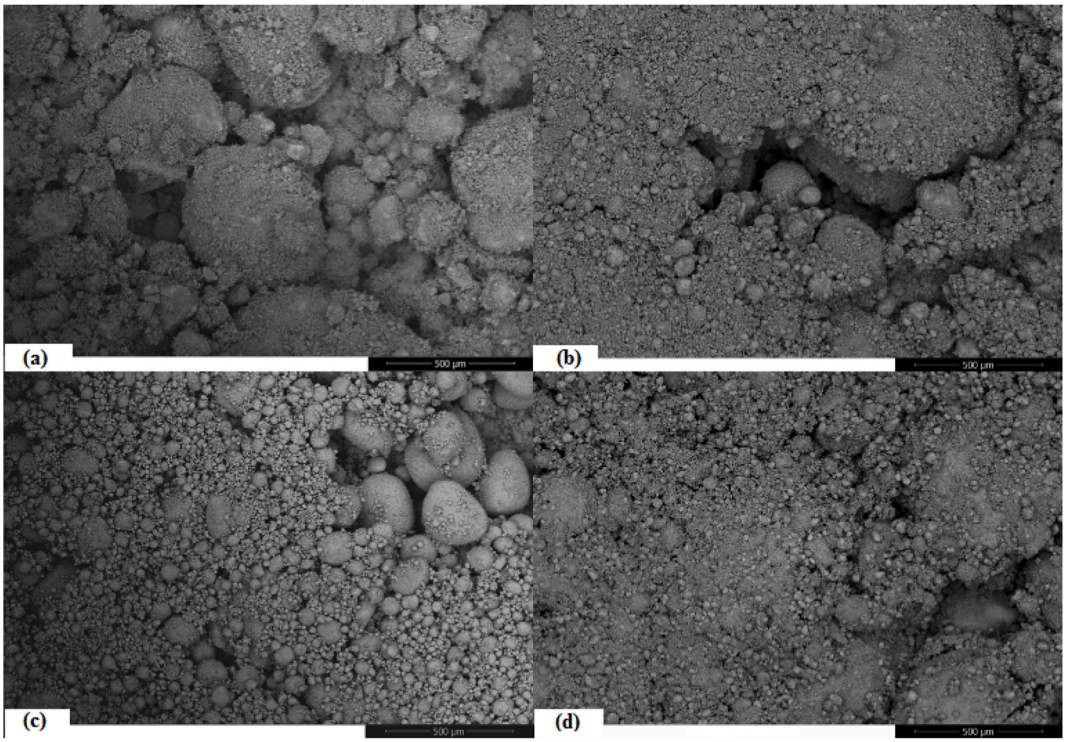
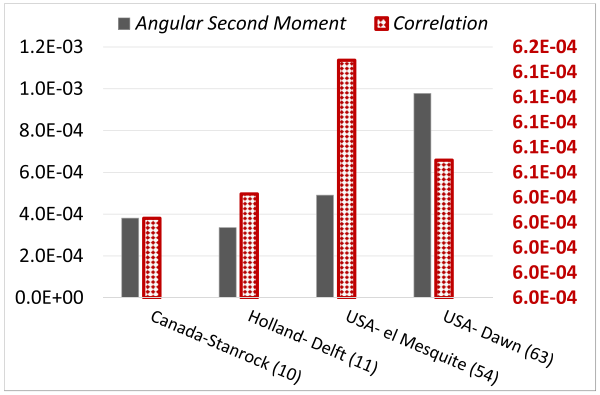
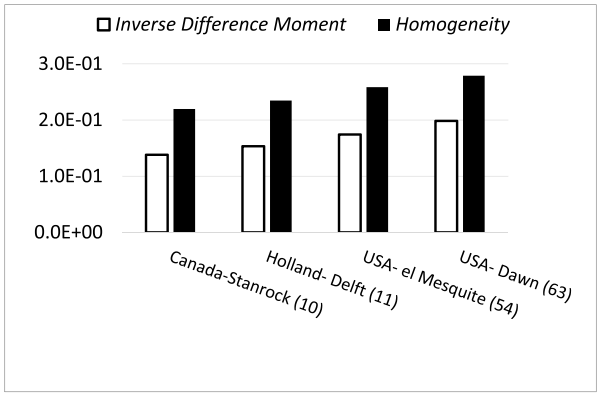


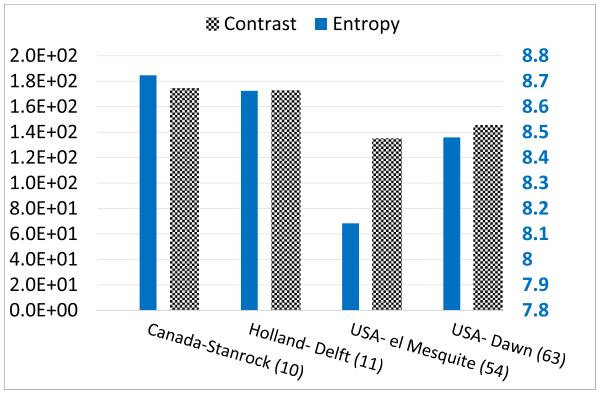
Fig. 5. SEM micro-graphs of UOC samples belonging to the *White-Yellow* group. (a) USA- Irigaray (33), (b) USA-Mobil (38), (c) Canada-Rabbit lake (51) and (d) USA-Everest Yellow (52).



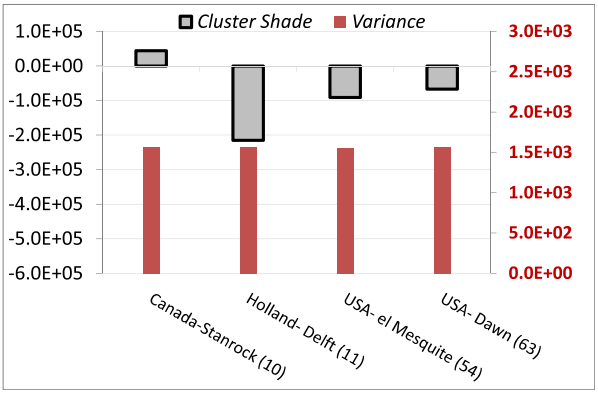
(a)



(b)



(c)



(d)

Fig. 6. Average values ($\theta = 0, 45, 90$) of the GLCM descriptors of the UOCs images in the *Yellow* group at 100 x. (a) The grey bars define the *Angular Second Moment* values, whilst the red patterned columns refer to the *Correlation* (right scale) (b) *Inverse Difference Moment* and *Homogeneity* are respectively represented by the bars with continuous single outline and the black bars (c) The right scale refers to the *Entropy* indices (blue bars) whilst *Contrast* is depicted by the black patterned columns (c) The *Variance* indices are indicated by the red bars and refer to the right scale. *Cluster Shade* indices are represented by the grey outlined bars. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

considered scale, images with local complex/irregular texture have higher mean angles than images with "smooth" local texture (Fongaro et al., 2016; Huang and Esbensen, 2000; Marchetti et al., 2020).

Fig. 8 depicts the average mean angle distributions in the scale range 0–1000, for the *White-Yellow* (a) and *Yellow* (b) images at 100 x. At low scales, the average mean angles of USA Irigaray (33) are higher than the mean angles of Canada-Rabbit lake (51) whilst an inversion is observed after about the fortieth variable (see Fig. 8 (a)). Before this inversion, the highest average mean angles are exhibited by the USA-Mobil (38) followed by Everest-Yellow (52)/USA Irigaray (33) then by Canada- Rabbit lake (51); this decrease seems to reproduce that of the *Contrast* textural descriptor in Fig. 4 (d). An inversion is also observed in the *Yellow* cluster (Fig. 8 (b)) where the mean angle spectrum of Canada Stanrock (10) is above the others at low scales while goes below starting after ~ the sixtieth variable. Also here, the decrease of the average mean angle values at low scales (hence, of textural complexity level) seems to reproduce the trend of the *Contrast* descriptor among the samples (from the highest to the lowest: 10, 11, 63 and 54) which is, in this case, consistent with the *Entropy* indices trend.

6.5. Development of texture-based classification models

The models were created by applying at first the SVM algorithm to the GLCM textural features, then to the mean angle spectra and finally by combining all the variables extracted. The mean angle spectra were manually inspected to discard those variables corresponding to almost overlapping spectra for different samples (large scale values). This last approach can be better visualised on the graph of Fig. 9 where the *Yellow* mean angle distributions, for the images at 100 x, are portrayed together with a green rectangle indicating the cut-off variable.

In the textural classification, the classes are balanced, then widely used parameters such as Accuracy and macro-F1 (Pillai et al., 2017) can be used to evaluate the model performance (F1 as the harmonic mean of precision and sensitivity) (Tharwat, 2210). Here, macro-F1 was selected for decision-making because it is more informative with respect to Accuracy (He and Garcia, 2009); MCC is also calculated for consistency and shown together with the results of the other two metrics at the end. The performance of the textural based classification model obtained by applying SVM to the *White-Yellow* mean angle spectra extracted from the images at 100x is described in Table 4. The macro-F1 in cross-validation (Venetian blinds) and prediction is 0.87.

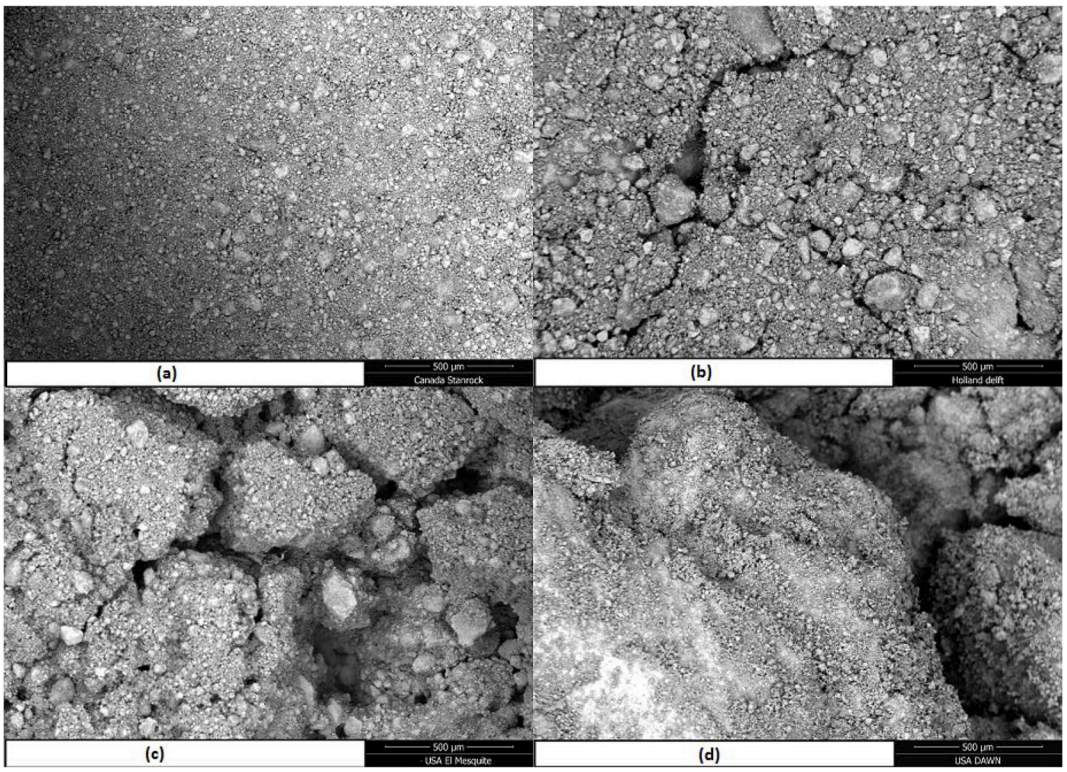


Fig. 7. SEM micro-graphs of UOC samples belonging to the *Yellow* group. (a) Canada- Stanrock (10), (b) Holland-Delft (11), (c) USA-El Mesquite (54) and (d) USA-Dawn (63).

In this case, a better classification was achieved by using only the GLCM descriptors (Table 5): macro-F1 is 0.90 in cross-validation whilst, in prediction, the single-label F1 is always 1. Here, the mean angle variables did not increase the classification performance when merged with the GLCM textural features.

For the magnification at 250×, among the different experimented approaches, Principal Components Analysis (PCA) (Marchetti et al., 2020) of the combined GLCM and AMT variables (maximum scale s_M = 500) was performed before the supervised classification. After inspection of the scores in PCA, one Rabbit Lake sample appeared a strong outlier in terms of Hotelling T^2 (Rencher, 2002). This sample was removed from the dataset and the GLCM textural features merged with the mean angle spectra cut after the variable 500 for 51/52 samples. The model possesses a macro-F1 in cross-validation (Venetian blinds) of 0.83 and 0.87 in prediction.

The entire set of images at 500 x magnification was used to develop the SVM model for the classification of the powders; this time, GLCM textural indexes were coupled with the mean angle spectra having a maximum scale of 300. Macro-F1 of 0.81 and 0.87 were obtained respectively in cross-validation and prediction. The overall accuracies calculated both in cross-validation and in prediction for the samples in the *White-Yellow* group are generally ≥0.93; these values contribute to indicating a good classification performance. The reader can find additional models details in the Supplemental Material. The SVM models for the identification of powders in the *White-Yellow* cluster misclassify one Everest Yellow (52) image as USA-Irigaray (33) at 250 x magnification and one Canada-Rabbit lake (51) image as USA-Mobil (38) at 500 x. Hence, by using different magnifications the correct identity could be attributed to an unknown sample.

The used evaluation metrics highlight the worsening of the SVM classification quality with increasing magnification. The macro-F1 in cross-validation changes from 0.90 achieved with the images at 100 x, to 0.82 for the images at 250 x and it reaches its minimum (0.81) with the highest magnification; a similar decrease of the same parameters is experienced in prediction (Fig. 10). One possible explanation can be the increasing electrostatic charging with magnification observed during images acquisition of the peroxide powders and/or the larger field of view offered by the smaller magnifications which tend to improve the information correlated to the powder bulk properties.

SVM was also employed to classify the samples in the *Yellow* cluster: GLCM features were combined with the mean angle spectra having a maximum scale of 620 for 100 x and 700 for 500 x, while only GLCM features were used for the 250 x magnification, because a good cut-off of the mean angle spectra was not found and generally, the evaluation metric privileged the GLCM textural features. Macro-F1 in cross-validation for 100 x is 0.98, 0.86 for 250 x and 0.81 for 500 x, while in prediction macro-F1 are, respectively, 1, 0.87 and 0.87 (see Fig. 11). The overall accuracies are ≥0.90.

In terms of misclassifications, one of the samples 54 (USA-El Mesquite) is identified as 11 (Holland-Delft) at 250 x while one of the USA-Dawn samples (63) is classified as USA-El Mesquite (54). Figs. 10 and 11 show that Accuracy, F1 and MCC are consistent in describing the classification performance of the model: to a higher accuracy value correspond higher macro-F1 and MCC. However, macro-F1 and MCC have smaller values which would contribute to limit over-optimistic conclusions.

7. Discussion and conclusions

The present work explores the performance of support vector machine for the classification of uranium ore concentrates by means of two combined analytical methods: reflectance spectrophotometry and image texture analysis (i.e Angle Measure Technique and Grey Level Co-occurrence Matrix).

In addition to an intuitive relationship existing between the UOC chemical composition and the powder colour, the spectrophotometric measurements provide evidence for the possible influence of the production process on the UOC colour nuances, hardly perceivable by

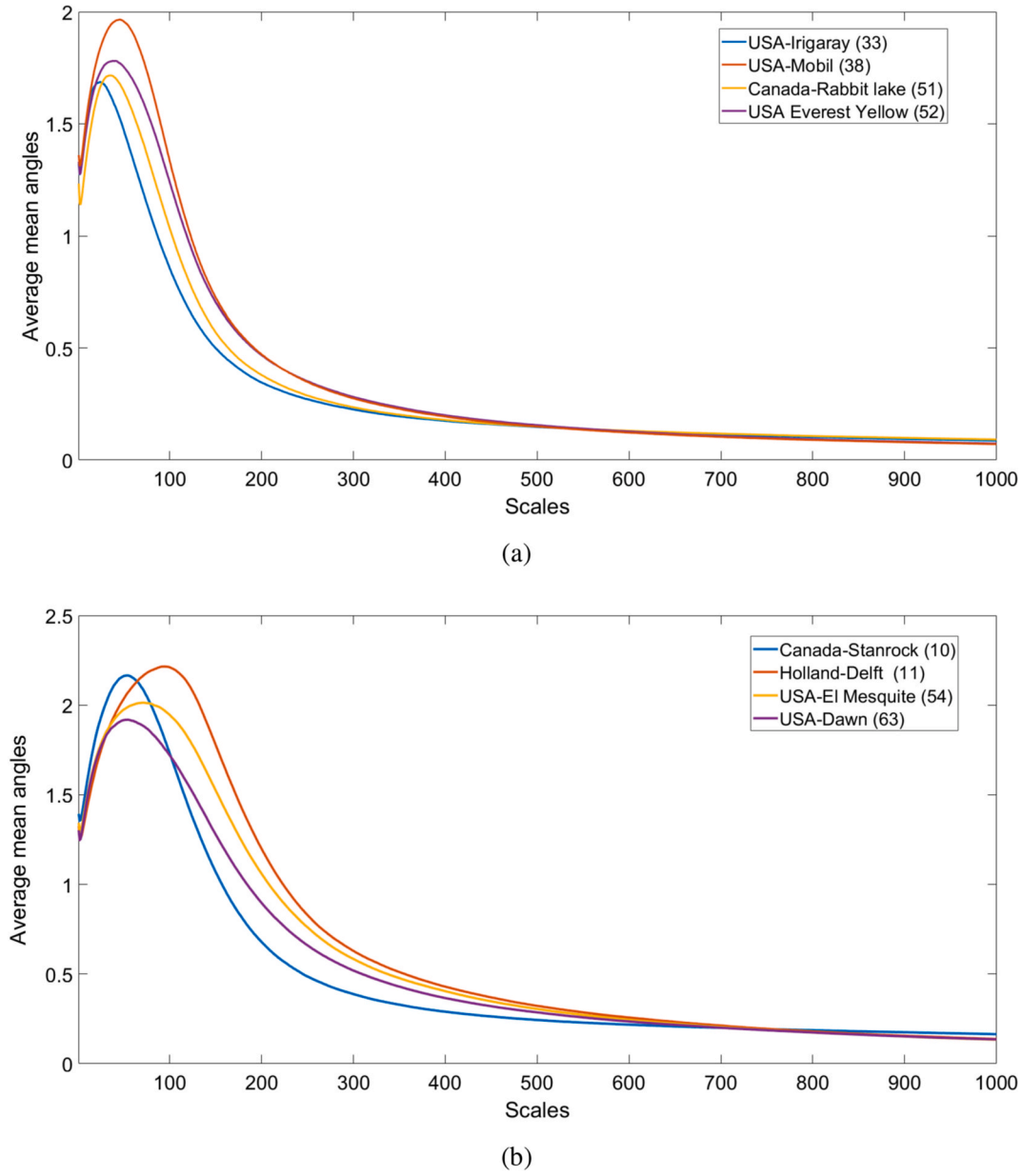


Fig. 8. Average mean angle spectra calculated for the images at 100 x belonging respectively to the (a) *White-Yellow* and (b) *Yellow* clusters. From the top to the bottom, the spectra depict: Usa-Mobil (38), USA-Everest Yellow(52), Canada-Rabbit lake (51) and USA-Irigaray (33) in (a) and Holland-Delft(11), USA-El Mesquite (54), USA-Dawn (63) and Canada-Stanrock (10) in (b). In (b) the reference scales range is 100–500. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

human eyes. Different conditions (e.g., temperature) or simply a non-equal impurity contents are most probably correlated to the variation in the reflectance spectra which enable a first UOCs classification based on colours. In this first part of the study, the 79 commercial UOCs are categorised in six chromatic groups: the built SVM model presents a good classification performance which is described by the large value of the Matthews Correlation Coefficient (MCC = 0.95) and can already contribute to speeding up the recognition process if combined with a traditional forensic method.

The second part of the study presents the results of the textural analysis for the improvement of the identification of an unknown specimen. To test the procedure, SEM images of powders belonging to two chromatic classes (*Yellow* and *White-Yellow*) are analysed using the GLCM and AMT algorithms.

The textural analysis performed with the above-mentioned algorithms highlights various levels of physical textural complexity; e.g, different homogeneities in the images depicting each UOC in the same colour-group, non-uniform disorder-level in the arrangement of the pixels grey values or different Contrasts. In the analysis performed in section 6.3, it appears that among the powders belonging to the *White-Yellow* group, USA-Irigaray (33) displays the most heterogeneous pattern because of the distribution of the powder grains. However, this heterogeneous pattern is revealed by the MA spectra only at low scales, because at intermediate scales, the most complex texture seems to be possessed by USA-Mobil (38). In the *Yellow* group, the textural complexity level revealed by the MA spectra reproduces better the level of dissimilarity, disorder and contrast expressed by the GLCM indices; here, Canada-Stanrock (10) and Holland-Delft (11) display the most heterogeneous textures and the reason may lie in the smaller grains with respect to USA-El Mesquite (54) and USA-Dawn (63).

The differences in the textural parameters enable the machine learning algorithm classification/recognition process and the

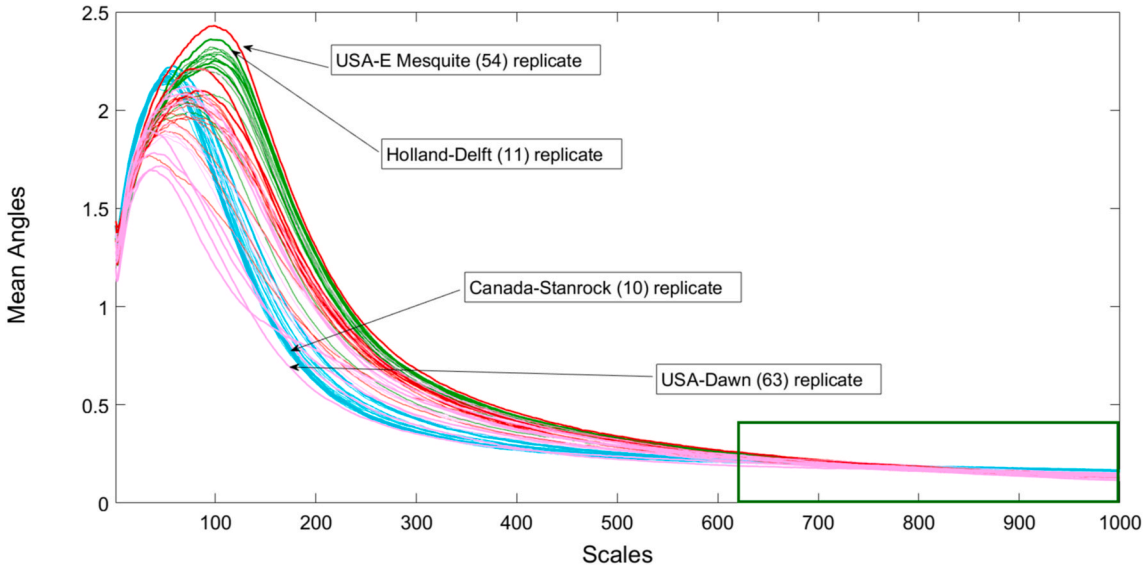


Fig. 9. AMT spectra obtained with images of samples in the *Yellow* group having 100 x of magnification. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 4
Evaluation metrics of the SVM model (cost = 10 and γ = 0.001) in cross-validation (CV) and prediction (P) for the images at 100 x of the *White-Yellow* group by using the AMT algorithm.

Sample Name	USA-Everest Yellow	USA-Irigaray	Canada-Rabbit lake	USA-Mobil
Number	(52)	(33)	(51)	(38)
F1 (CV)	0.80	0.89	0.90	0.87
F1 (P)	0.67	0.80	1	1

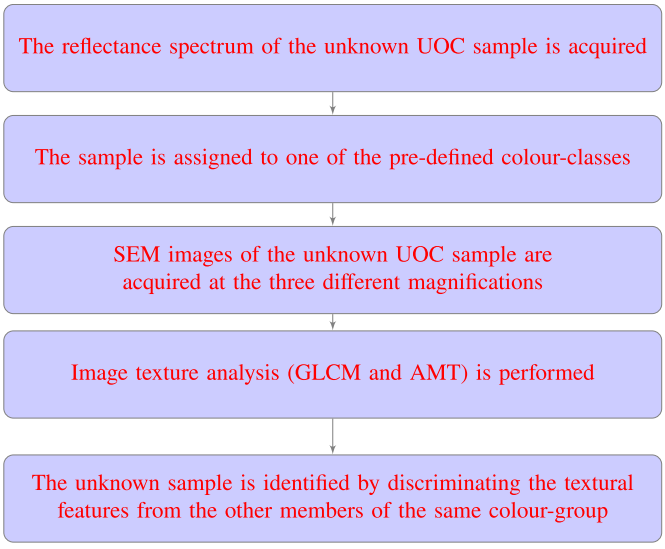
Table 5
SVM classification of *White-Yellow* images at 100 x obtained with GLCM textural features. In bold the parameters improved in respect of the model obtained with AMT. Underlined the decreased parameters. Globally a better performance is observed with the second approach.

Sample Name	USA-Everest Yellow	USA-Irigaray	Canada-Rabbit lake	USA-Mobil
Number	(52)	(33)	(51)	(38)
F1 (CV)	**0.83**	**0.96**	**0.96**	0.85
F1 (P)	**1**	**1**	1	1

development of the models. For this, the variables extracted with the two different methods are treated separately or they are merged when this leads to a better classification in terms of F1 and accuracy.

Generally, the texture-based classification models present a better performance by using an external, blind, dataset, in comparison with the internal cross-validation, although the parameters are not strongly different. In this context, the models obtained with the lowest magnification (100×) perform better in comparison with 250 x and 500 x for both chromatic clusters (*White-Yellow* and *Yellow*). This would suggest that a larger field of view is beneficial for classifications based on textural algorithms; however, the three classifications can be synergically employed to increase the robustness of the identification.

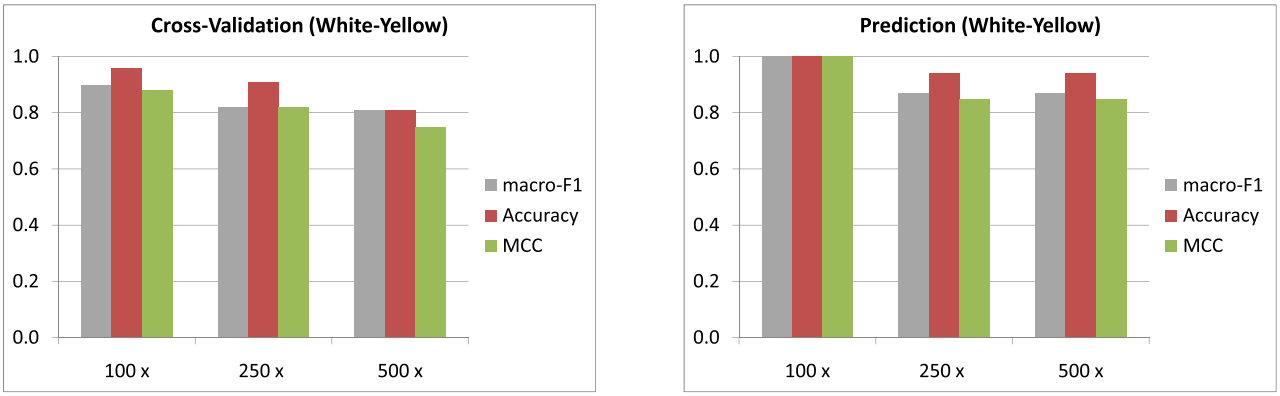
In the case of a real investigation (of an unknown sample), the colour-based and textural-based models can be applied using the following pathway:



Finally, it needs to be understood that it is unlikely to have a "one-to-one match", if the investigated sample has been produced in another facility or in recent times. However, the studied sample set represents a vast amount of facilities and processes still in use for the production of UOCs. Therefore, we believe that there would be substantial similarities between the unknown sample and an UOC from the used sample set, if a similar production process was applied. In addition to that, the models are updated continuously when new UOC samples become available.

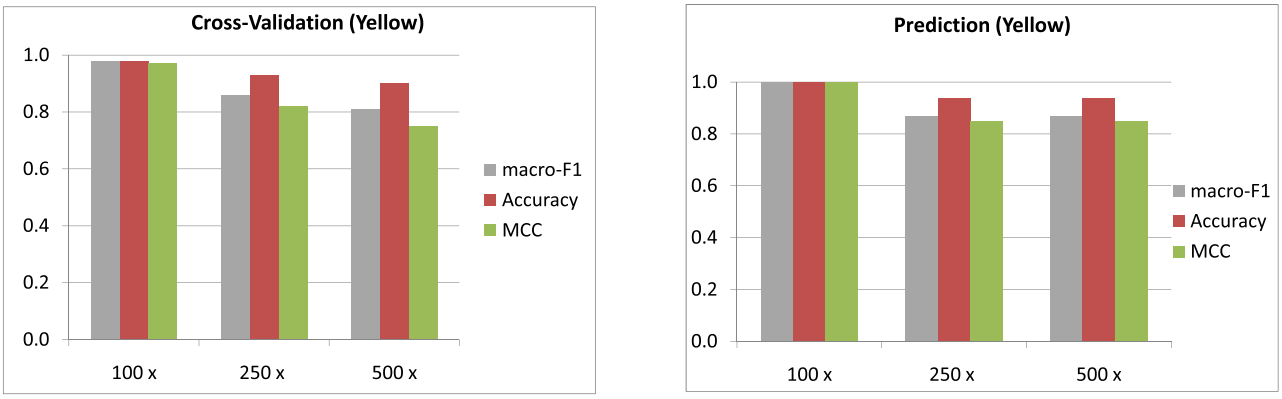
CRediT authors contribution statement

M. Marchetti: Methodology, Investigation, Formal Analysis, Validation, Data Curation, Writing-Original, Editing. L. Fongaro: Conceptualization, Methodology, Funding Acquisition, Project Administration, Supervision,Writing-Review. A. Bulgheroni: Methodology, Investigation, Writing-Review. M. Wallenius: Resources, Writing-Review. K. Mayer: Supervision, Resources, Writing-Review.



(a) (b)

Fig. 10. SVM classification quality for the images in the *White-Yellow* cluster. From left to right: macro-F1, Accuracy and MCC in (a) cross-validation and (b) prediction with the external images set. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



(a) (b)

Fig. 11. SVM classification quality measured by macro-F1, Accuracy and MCC for the images in the *Yellow* cluster. From left to right: macro-F1, Accuracy and MCC in (a) cross-validation (b) prediction with the external images set. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work is part of the IDENTICLASS project, which was supported by the Exploratory Research Programme 2018, of the European Commission, Join Research Centre, Directorate A. The authors convey their gratefulness to A. Benedetti, for the support during the SEM image acquisition and T. Wiss, for the helpful discussion on SEM image acquisition. The authors wish to express their gratitude to Prof. Dr. K. Lützenkirchen for the support received during the project execution and to Adrian Nicholl together with Martin Varga Zuniga for technical support. N. A. DiBlasi is gratefully acknowledged for proofreading the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeochem.2022.105443.

References

Abe, S., 2010. Support Vector Machines for Pattern Classification. Springer London, London, ISBN 978-1-84996-098-4, pp. 1–19. https://doi.org/10.1007/978-1-84996-098-4_1.

AMT Plugin for ImageJ, 2014. http://arken.nmbu.no/kkvaal/eamtexplorer/imagej_plugins.html. (Accessed 21 April 2020).

Andrle, R., 1994. The angle measure technique: a new method for characterizing the complexity of geomorphic lines. Math. Geol. 26 (1), 83–97. https://doi.org/10.1007/BF02065877.

Awad, M., Khanna, R., 2015. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Apress, Berkeley, CA, ISBN 978-1-4302-5990-9, pp. 39–66. https://doi.org/10.1007/978-1-4302-5990-9_3.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., Nielsen, H., 2000. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16 (5), 412–424. https://doi.org/10.1093/bioinformatics/16.5.412.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pp. 144–152. https://doi.org/10.1145/130385.130401.

Brennecka, G.A., Borg, L.E., Hutcheon, I.D., Sharp, M.A., Anbar, A.D., 2010. Natural variations in uranium isotope ratios of uranium ore concentrates: understanding the 238U/235U fractionation mechanism. Earth Planet Sci. Lett. ISSN: 0012-821X 291 (1), 228–233. https://doi.org/10.1016/j.epsl.2010.01.023.

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 21 (1), 1–13. https://doi.org/10.1186/s12864-019-6413-7.

Denis, D.J., 2015. Applied Univariate, Bivariate, and Multivariate Statistics. John Wiley & Sons.

Esbensen, K.H., Hjelmen, K.H., Kvaal, K., 1996. The AMT approach in chemometrics-first forays. J. Chemometr. 10 (5–6), 569–590.

Fongaro, L., Kvaal, K., 2013. Surface texture characterization of an Italian pasta by means of univariate and multivariate feature extraction from their texture images. Food Res. Int. ISSN: 0963-9969 51 (2), 693–705. https://doi.org/10.1016/j.foodres.2013.01.044.

Fongaro, L., Ho, D.M.L., Kvaal, K., Mayer, K., Rondinella, V.V., 2016. Application of the angle measure technique as image texture analysis method for the identification of uranium ore concentrate samples: new perspective in nuclear forensics. Talanta. ISSN: 0039-9140 152, 463–474. https://doi.org/10.1016/j.talanta.2016.02.027.

Halstensen, M., Kvaal, K., Esbensen, K.H., 2019. Image analytical sandstone plug poro-perm prediction using angle measure technique (AMT) and chemometrics – afeasibility study. Chemometr. Intell. Lab. Syst. ISSN: 0169-7439 193, 103847 https://doi.org/10.1016/j.chemolab.2019.103847.

Hanson, A.B., Lee, R.N., Vachet, C., Schwerdt, I.J., Tasdizen, T., McDonald IV, L.W., 2019. Quantifying impurity effects on the surface morphology of U_3O_8. Anal. Chem. 91 (15), 10081–10087. https://doi.org/10.1021/acs.analchem.9b02013.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3, pp. 610–621. https://doi.org/10.1109/TSMC.1973.4309314, 6.

Härdle, W., Simar, L., 2007. Applied Multivariate Statistical Analysis, vol. 22007. Springer.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer New York, New York, NY, ISBN 978-0-387-84858-7, pp. 485–585. https://doi.org/10.1007/978-0-387-84858-7_14.

He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21 (9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239.

Ho, D.M.L., 2015. Study on the Applicability of Structural and Morphological Parameters on Selected Uranium Compounds for Nuclear Forensic Purposes. Ph.D. thesis.

Huang, J., Esbensen, K.H., 2000. Applications of Angle Measure Technique (AMT) in image analysis: Part I. A new methodology for in situ powder characterization. Chemometr. Intell. Lab. Syst. ISSN: 0169-7439 54 (1), 1–19. https://doi.org/10.1016/S0169-7439(00)00100-3.

Hung, C.-C., Song, E., Lan, Y., 2019. Image Texture Analysis: Foundations, Models and Algorithms. Springer International Publishing, Cham, pp. 3–20. https://doi.org/10.1007/978-3-030-13773-1_1.

G. Jurman, S. Riccadonna, C. Furlanello, A comparison of MCC and CEN error measures in multi-class prediction, PLoS One 7 (8), doi:https://doi.org/10.1371/journal.pone.0041882.

Keegan, E., Richter, S., Kelly, I., Wong, H., Gadd, P., Kuehn, H., Alonso-Munoz, A., 2008. The provenance of Australian uranium ore concentrates by elemental and isotopic analysis. Appl. Geochem. ISSN: 0883-2927 23 (4), 765–777. https://doi.org/10.1016/j.apgeochem.2007.12.004.

Keegan, E., Kristo, M.J., Colella, M., Robel, M., Williams, R., Lindvall, R., Eppich, G., Roberts, S., Borg, L., Gaffney, A., Plaue, J., Wong, H., Davis, J., Loi, E., Reinhard, M., Hutcheon, I., 2014. Nuclear forensic analysis of an unknown uranium ore concentrate sample seized in a criminal investigation in Australia. Forensic Sci. Int. 240, 111–121. https://doi.org/10.1016/j.forsciint.2014.04.004.

Kennedy, A.K., Bostick, D.A., Hexel, C.R., Smith, R.R., Giaquinto, J.M., 2013. Non-volatile organic analysis of uranium ore concentrates. J. Radioanal. Nucl. Chem. 296 (2), 817–821. https://doi.org/10.1007/s10967-012-2032-3.

Klunder, G.L., Plaue, J.W., Spackman, P.E., Grant, P.M., Lindvall, R.E., Hutcheon, I.D., 2013. Application of visible/near-infrared reflectance spectroscopy to uranium ore concentrates for nuclear forensic analysis and attribution. Appl. Spectrosc. 67 (9), 1049–1056. https://doi.org/10.1366/12-06947.

Krajkó, J., Varga, Z., Yalcintas, E., Wallenius, M., Mayer, K., 2014. Application of neodymium isotope ratio measurements for the origin assessment of uranium ore concentrates. Talanta. ISSN: 0039-9140 129, 499–504. https://doi.org/10.1016/j.talanta.2014.06.022.

Kristo, M.J., Gaffney, A.M., Marks, N., Knight, K., Cassata, W.S., Hutcheon, I.D., 2016. Nuclear forensic science: analysis of nuclear material out of regulatory control. Annu. Rev. Earth Planet Sci. 44 (1), 555–579. https://doi.org/10.1146/annurev-earth-060115-012309.

Kucheryavski, S.V., Kvaal, K., Halstensen, M., Mortensen, P.P., Dahl, C.K., Minkkinen, P., Esbensen, K.H., 2008a. Optimal corrections for digitization and quantification effects in angle measure technique (AMT) texture analysis. J. Chemometr. 22 (11–12), 722–737. https://doi.org/10.1002/cem.1118.

Kucheryavski, S.V., Kvaal, K., Halstensen, M., Mortensen, P.P., Dahl, C.K., Minkkinen, P., Esbensen, K.H., 2008b. Optimal corrections for digitization and quantification effects in angle measure technique AMT texture analysis. J. Chemometr.: J. Chemomet. Soc. 22 (11–12), 722–737.

Kvaal, K., Kucheryavski, S.V., Halstensen, M., Kvaal, S., Fløe, A.S., Minkkinen, P., Esbensen, K.H., 2008. eAMTexplorer: a software package for texture and signal characterization using Angle Measure Technique. J. Chemometr. 22 (11–12), 717–721. https://doi.org/10.1002/cem.1160.

Luque, A., Carrasco, A., Martín, A., de las Heras, A., 2019. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recogn. ISSN: 0031-3203 91, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023.

Manna, S., Roy, S.B., Joshi, J.B., 2012a. Study of crystallization and morphology of ammonium diuranate and uranium oxide. J. Nucl. Mater. ISSN: 0022-3115 424 (1), 94–100. https://doi.org/10.1016/j.jnucmat.2012.02.012.

Manna, S., Karthik, P., Mukherjee, A., Banerjee, J., Roy, S.B., Joshi, J.B., 2012b. Study of calcinations of ammonium diuranate at different temperatures. J. Nucl. Mater. ISSN: 0022-3115 426 (1), 229–232. https://doi.org/10.1016/j.jnucmat.2012.03.035.

Maratea, A., Petrosino, A., Manzo, M., 2014. Adjusted F-measure and kernel scaling for imbalanced data learning. Inf. Sci. ISSN: 0020-0255 257, 331–341. https://doi.org/10.1016/j.ins.2013.04.016.

Marchetti, M., Mayer, K., Wallenius, M., Bulgheroni, A., Wiss, T., Lützenkirchen, K., Fongaro, L., 2020. Image texture analysis and colorimetry for the classification of uranium ore concentrate powders. In: EPJ Web of Conferences, vol. 225. EDP Sciences, 07003. https://doi.org/10.1051/epjconf/202022507003.

Olsen, A.M., Richards, B., Schwerdt, I., Heffernan, S., Lusk, R., Smith, B., Jurrus, E., Ruggiero, C., McDonald IV, L.W., 2017. Quantifying morphological features of α-U_3O_8 with image analysis for nuclear forensics. Anal. Chem. 89 (5), 3177–3183. https://doi.org/10.1021/acs.analchem.6b05020.

Pantic, I., Pantic, S., Paunovic, J., Perovic, M., 2013. Nuclear entropy, angular second moment, variance and texture correlation of thymus cortical and medullar lymphocytes: grey level co-occurrence matrix analysis. An Acad. Bras Ciências 85, 1063–1072. https://doi.org/10.1590/S0001-37652013005000045.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pillai, I., Fumera, G., Roli, F., 2017. Designing multi-label classifiers that maximize F measures: state of the art. Pattern Recogn. ISSN: 0031-3203 61, 394–404. https://doi.org/10.1016/j.patcog.2016.08.008.

Rencher, A.C., 2002. Methods of Multivariate Analysis. John Wiley & Sons.

Russ, J.C., 1999. The Image Processing Handbook. chap. 4, 3 edn. CRC press.

Schanda, J., 2007. Colorimetry: Understanding the CIE System. John Wiley & Sons. https://doi.org/10.1002/9780470175637.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9 (7), 671.

Shi, L., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C., Collins, P.J., De Longueville, F., Kawasaki, E.S., Lee, K.Y., et al., 2006. The MicroArray Quality Control (MAQC) project shows inter-and intraplatform reproducibility of gene expression measurements. Nat. Biotechnol. 24 (9), 1151. https://doi.org/10.1038/nbt1239.

Sivchenko, N., Kvaal, K., Ratnaweera, H., 2016. Evaluation of image texture recognition techniques in application to wastewater coagulation. Cogent Eng. 3 (1), 1206679 https://doi.org/10.1080/23311916.2016.1206679.

Spano, T.L., Simonetti, A., Balboni, E., Dorais, C., Burns, P.C., 2017. Trace element and U isotope analysis of uraninite and ore concentrate: applications for nuclear forensic investigations. Appl. Geochem. ISSN: 0883-2927 84, 277–285. https://doi.org/10.1016/j.apgeochem.2017.07.003.

Strauss, T., von Maltitz, M.J., 2017. Generalising Ward's method for use with Manhattan distances. PLoS One 12 (1), e0168288. https://doi.org/10.1371/journal.pone.0168288.

Tamasi, A.L., Cash, L.J., Eley, C., Porter, R.B., Pugmire, D.L., Ross, A.R., Ruggiero, C.E., Tandon, L., Wagner, G.L., Walensky, J.R., et al., 2016a. A lexicon for consistent description of material images for nuclear forensics. J. Radioanal. Nucl. Chem. 307 (3), 1611–1619. https://doi.org/10.1007/s10967-015-4455-0.

Tamasi, A.L., Cash, L.J., Mullen, W.T., Ross, A.R., Ruggiero, C.E., Scott, B.L., Wagner, G. L., Walensky, J.R., Zerkle, S.A., Wilkerson, M.P., 2016b. Comparison of morphologies of a uranyl peroxide precursor and calcination products. J. Radioanal. Nucl. Chem. 309 (2), 827–832. https://doi.org/10.1007/s10967-016-4692-x.

A. Tharwat, Classification assessment methods, Appl. Comput.Informat. ISSN 2210–8327, doi:https://doi.org/10.1016/j.aci.2018.08.003.

Thompson, N.B., O'Sullivan, S.E., Howell, R.J., Bailey, D.J., Gilbert, M.R., Hyatt, N.C., 2021. Objective colour analysis from digital images as a nuclear forensic tool. Forensic Sci. Int. ISSN: 0379-0738 319, 110678 https://doi.org/10.1016/j.forsciint.2020.110678.

Varga, Z., Wallenius, M., Mayer, K., Keegan, E., Millet, S., 2009. Application of lead and strontium isotope ratio measurements for the origin assessment of uranium ore concentrates. Anal. Chem. 81 (20), 8327–8334. https://doi.org/10.1021/ac901100e.

Varga, Z., Wallenius, M., Mayer, K., Hrnecek, E., 2011a. Alternative method for the production date determination of impure uranium ore concentrate samples. J. Radioanal. Nucl. Chem. 290 (2), 485–492. https://doi.org/10.1007/s10967-011-1233-5.

Varga, Z., Öztürk, B., Meppen, M., Mayer, K., Wallenius, M., Apostolidis, C., 2011b. Characterization and classification of uranium ore concentrates (yellow cakes) using

infrared spectrometry. Radiochim. Acta 99 (12), 807–813. https://doi.org/10.1524/ract.2011.1886.
Vladimir, N.V., 1998. Statistical Learning Theory. John Wiley & Sons, ISBN 978-0-471-03003-4.
Yang, X., Tridandapani, S., Beitler, J.J., Yu, D.S., Yoshida, E.J., Curran, W.J., Liu, T., 2012. Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: an in vivo study of late toxicity. Med. Phys. 39 (9), 5732–5739. https://doi.org/10.1118/1.4747526.