

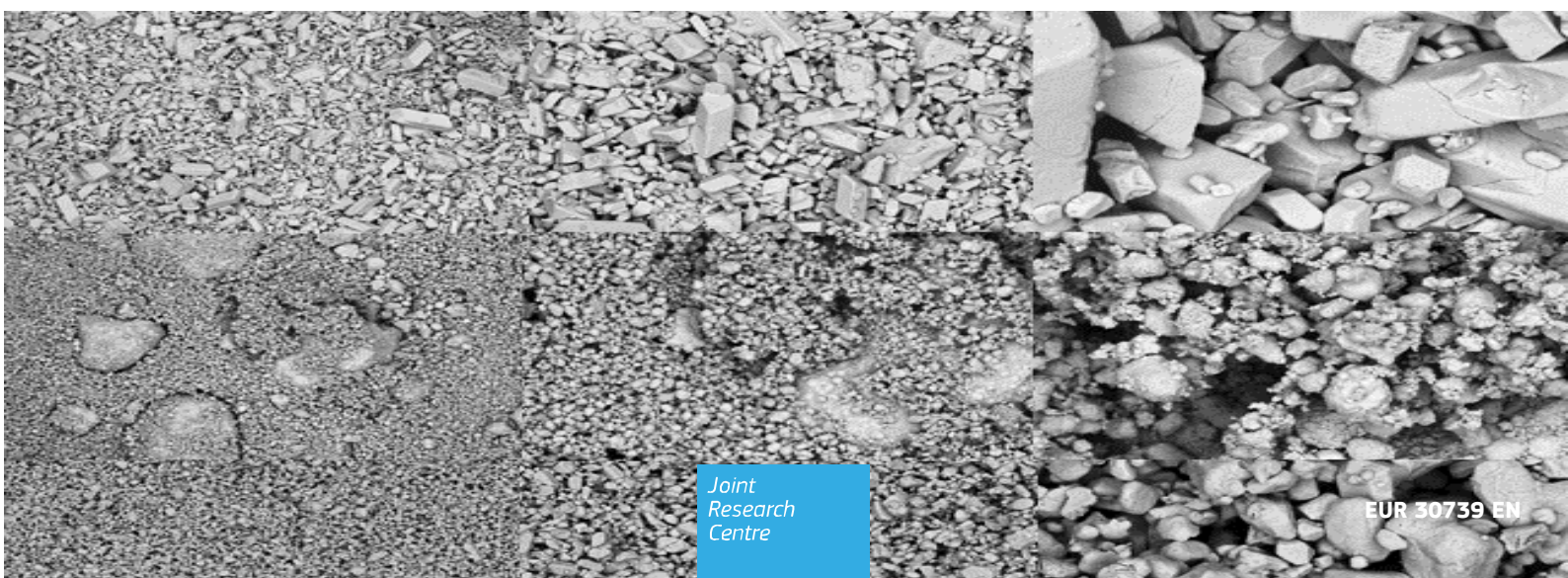
JRC TECHNICAL REPORT

Development of a new approach for a rapid IDENTification and CLASSification of uranium powders using colour, image texture and spectroscopy signatures

A summary of the exploratory research project IDENTICLASS

Fongaro, L., Marchetti, M., Lande I.B., Wallenius, M., Mayer, K.

2021



This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information

Name: Lorenzo Fongaro
Address: JRC-Karlsruhe – Postfach 2340 – 76125 Karlsruhe - Germany
Email: Lorenzo.FONGARO@ec.europa.eu
Tel.: +49 7247 951 593

EU Science Hub

<https://ec.europa.eu/jrc>

JRC124728

EUR 30739 EN

PDF	ISBN 978-92-76-38732-9	ISSN 1831-9424	doi:10.2760/180882
Print	ISBN 978-92-76-38731-2	ISSN 1018-5593	doi:10.2760/28216

Luxembourg: Publications Office of the European Union, 2021

© European Atomic Energy Community, 2021



The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Atomic Energy Community, 2021,

How to cite this report: Fongaro, L., Marchetti, M., Lande, I.B., Wallenius, M., Mayer, K, *Development of a new approach for a rapid IDENTification and CLASSification of uranium powders using colour, image texture and spectroscopy signatures*, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-79-38731-2, doi:10.2760/28216, JRC124728.

Contents

Note on this document.....	1
Acknowledgements.....	2
Abstract.....	3
1 Introduction.....	4
1.1 Project description.....	4
2 Materials and methods.....	7
2.1 Materials.....	7
2.2 Methods.....	8
2.2.1 Colour analysis by spectrophotometer measurements.....	8
2.2.2 Scanning electron microscopy and image texture analysis.....	8
2.2.3 Image texture analysis.....	10
2.2.4 Hyperspectral image analysis.....	10
2.2.5 Models developments by Machine Learning.....	11
3 Results.....	14
3.1 Colour based classification model development.....	15
3.2 Image Texture based classification model development.....	20
3.2.1 Preliminary results.....	20
3.2.2 Development of the final model.....	20
3.3 Spectral based classification model.....	26
4 Conclusions.....	29
5 Further developments and future implementation.....	30
References.....	31
List of abbreviations and definitions.....	33
List of figures.....	34
List of tables.....	36
Annexes.....	37
Annex 1. Description of the texture feature extraction algorithms used for the texture based classification model development.....	37
Annex 2. Pre-processing steps for the near infrared spectra extracted of colour class 3.....	40
Annex 3. Average prediction probability matrices.....	42

Note on this document

This report summarizes the research activities carried out during the two years of the exploratory research project IDENTICLASS. Even if we were not able to collect all the foreseen experimental data, the image datasets acquired and elaborated were enough to be processed by machine learning. The main objectives of the project were achieved and the colour-based, texture-based and spectroscopy-based classification models were developed successfully. Since this document is a technical report, we focused mainly on the big picture of our achievements describing how the models work to identify and classify potential seized UOCs powders using only their colour and morphological and hyperspectral images. For this reason, the discussion on the results are limited on the performance of the final classification models, while the results obtained during the models development procedure are largely omitted.

Acknowledgements

The authors would like to express their gratitude to Professors Cecilia Fursaether, Knut Kvaal and Oliver Tomic from Norwegian University of Life Science, for their precious contribution in particular in the machine learning domain. This project represents the beginning of a fruitful collaboration that, of course, will continue in the future.

A special thanks goes to Luana Cognini who has collaborated in putting the first stones of this project.

Thank you also to the Scientific Committee and the Unit A.5 of JRC that have believed in this project.

Last but not least, I am very grateful to all those colleagues that given their contribution with intellectual advises and practical works during the two years, so many thanks to Antonio Bulgheroni, Markus Ernstberger, Thierry Wiss, Nadya Rauff-Nisthar, Adrian Nicholl, and also to all the administrative staff that help me during the different procurements procedures.

Authors

Lorenzo Fongaro (G.I.6), Mara Marchetti (A.5), Isak Lande (Norwegian University of Life Science, G.II.8), Maria Wallenius (G.II.8), Klaus Mayer (G.II.8)

Abstract

The IDENTICLASS project gave its contribution to the Exploratory Research (ER) Programme in 2018-2020. This ER project was focused on the development of a new analytical approach for the quick characterization and classification of uranium containing powders, primarily UOCs (Uranium Ore Concentrates), to combat their illicit trafficking. The study combines colour analysis by spectrophotometric measurements, different image texture analysis methods (such as AMT, angle measure technique; GLCM, grey level co-occurrence matrix; GLRLM, grey level run length matrix, local binary pattern), with information on the molecular structure (obtained by NIR spectroscopy) and then combine the analytical data for further processing. The expected result of this exploratory research was to obtain a new method, which makes use of a dedicated mathematical model based on data fusion, hence combining colour, image texture features and NIR spectroscopy data. To develop the models, a suite of 79 UOC samples available in the Directorate G.II.8 laboratories have been used as training and test samples. The main objectives were achieved and the classification models were developed successfully. After a final refinement, this approach could be implement as a new methodology to investigate seized uranium ore concentrate powders.

1 Introduction

The legal use of uranium is supposed to be only for energy production, but unfortunately, it was not always like that as history has taught us. The peaceful use of nuclear energy has been promoted in Europe by the EURATOM since 1957. It has been also established that all the Member States that signed the Euratom Treaty, should collaborate in promoting the peaceful use of nuclear energy and in fighting against its illegal use.

Nuclear security remains a priority for the European Union and, in order to avoid unauthorized use of nuclear materials, JRC contributes actively in defining and implementing the EU Nuclear Security Policy in various manners. Nuclear forensics is considered as a key component to respond to a nuclear security event. JRC has been pioneering this area in developing new approaches for the characterization of seized nuclear materials and next to highly sophisticated methods; attention is now given to developing techniques for their rapid identification.

Incidents involving nuclear materials out of regulatory control in the last decades have raised the necessity to include the early stages of the nuclear fuel cycle in nuclear forensics investigations; in this scenario increasing interest has been attributed to the uranium ore concentrates (UOCs) [1, 2].

UOCs are the precursors of nuclear fuel and they are the product resulting from the mining, milling and leaching of the uranium ores in the front-end of the nuclear fuel cycle. The process includes purification by means of solvent extraction or ion-exchange, and precipitation. The UOCs contain about 60-80% of uranium in different chemical composition: e.g. ammonium diuranate, uranyl hydroxide, uranyl peroxide or uranium trioxide when calcination is performed [2]; they are produced and treated in large quantities and, therefore, diversions or thefts can happen.

In the event of a discovery of illicit trafficking of UOCs, questions such as 'what is the material?', 'How it was produced?', 'Where did it come from?' have to be answered as soon as possible [3-6]. This approach is the core of nuclear forensic investigations. Typically, an investigation involves several measurable parameters, also termed as "signatures or fingerprints"; these can be understood as physical, chemical or isotopic characteristics of the nuclear materials that could collectively help to identify its origin [3-6]. Various measurable quantities associated with the UOCs compositions have been reported. These include analysis of major isotopes of uranium, minor isotopes, other minor constituents, non-volatile organics compounds and anionic impurities [7-14].

In the last years the application of different image analysis (IA) techniques have been explored, in order to find a correlation between morphological characteristics of different UOCs and their processing or production history [1, 15-17]. Image analysis methods that need to be further investigated in nuclear forensics domain are image texture analysis and the hyperspectral image analysis, combined with machine learning. Image texture analysis can provide information related on the bulk powder environment [18] (i.e. how the particles are arranged together), while hyperspectral image analysis can provide information related to the chemical composition of the sample analysed because, in each pixel composing the image, the near infrared spectra are also stored [19]. Machine learning, as sub-domain of Artificial Intelligence (AI), allow to find correlation among large amount of data providing, in short time, information about similarities among samples because of the so-called classification models [20].

This ER project aims to further study the methodology proposed by Fongaro et al. [17], developing a new approach for the UOCs characterization combining colour, image texture and spectroscopy analysis, to combat illicit trafficking. In this report the project description, the methodology and the results reached during the two years of the exploratory research, are summarised. Even if not all the goals established during the planning phase have been reached, this research represent a milestone, opening new perspective in the R&D of nuclear forensics.

1.1 Project description

After the publication on the first application of the angle measure technique algorithm, as image texture analysis method to identify and classify uranium powders [17], a further investigation became necessary.

When the development of a new procedure for nuclear forensics purpose is taken into consideration, like in this case to identify and classify seized uranium powders, the time factor plays an important role in the investigative process. Nowadays, the research and development in the frame of nuclear forensic science, is focused on new procedures that are preferably faster or more precise than the other already available, helping in the origin attribution by revealing the processing history.

Our research was oriented not only to extend the application of the image texture analysis to a larger number of uranium ore concentrate samples, and to try other textural feature extraction algorithms; but also to test other selected analytical techniques that can provide, together with image texture analysis, important information in the shortest time possible. In this way, the overall accuracy of the final response in terms of the sample's origin attribution, can be also improved.

In the IDENTICLASS project different analytical methods were considered: colour analysis, image texture analysis applying different algorithms and hyperspectral imaging. The expected result of the development of this new approach was a dedicated mathematical model, developed by machine learning and based on data fusion, combining colour, image texture features and NIR spectroscopy data. This new methodology was developed and tested using a suite of 79 UOC samples that are available in the JRC Karlsruhe.

In particular:

- Colour analysis was carried out via spectrophotometric/colorimetric measurements. This technique allowed the creation of a predefined numbers of colour classes in which seized samples can be attributed factually before being analysed with the other following techniques. It should be noted that the colour of uranium powders is also a reflection of their chemical composition and therefore processing history.
- Image Texture analysis was carried out applying different image texture features extraction algorithms (angle measure technique, AMT, grey level co-occurrence matrix, GLCM, grey level run length matrix, GLRLM, local binary pattern, LBP; etc.). The features extracted with these techniques are related to the morphology characteristics of the powder samples and they are also correlated to the process technology.
- Hyperspectral image analysis was carried out with a near infrared hyperspectral camera in the wavelengths range from 970 nm to 1700 nm. This technique provides information about the chemical composition of the samples analysed. The spatial and spectroscopy information are combined in one single image, where in each pixel composing the image the correspondent infrared spectra is also stored.

The project's work break down structure is shown in Figure 1. It was organized in four phases; in the first 3 phases, a classification model for each methodology was developed using machine learning. In particular, a colour-based, texture-based and spectral-based classification models were successfully developed. The 4th phase focused on the selection of the best data fusion strategy followed by the development of the final classification model; this phase is still on-going. With this final model the chemical composition, and therefore the probably process history of the UOCs should be possible to be narrowed down using only the textural and hyperspectral images.

Each of the phases was organized into specific tasks including:

1. Development of the appropriate analytical procedure to collect the experimental data
2. Data collection
3. Data pre-processing, explorative data analysis and Machine Learning for the supervised classification model development.
4. Validation tests and results assessment

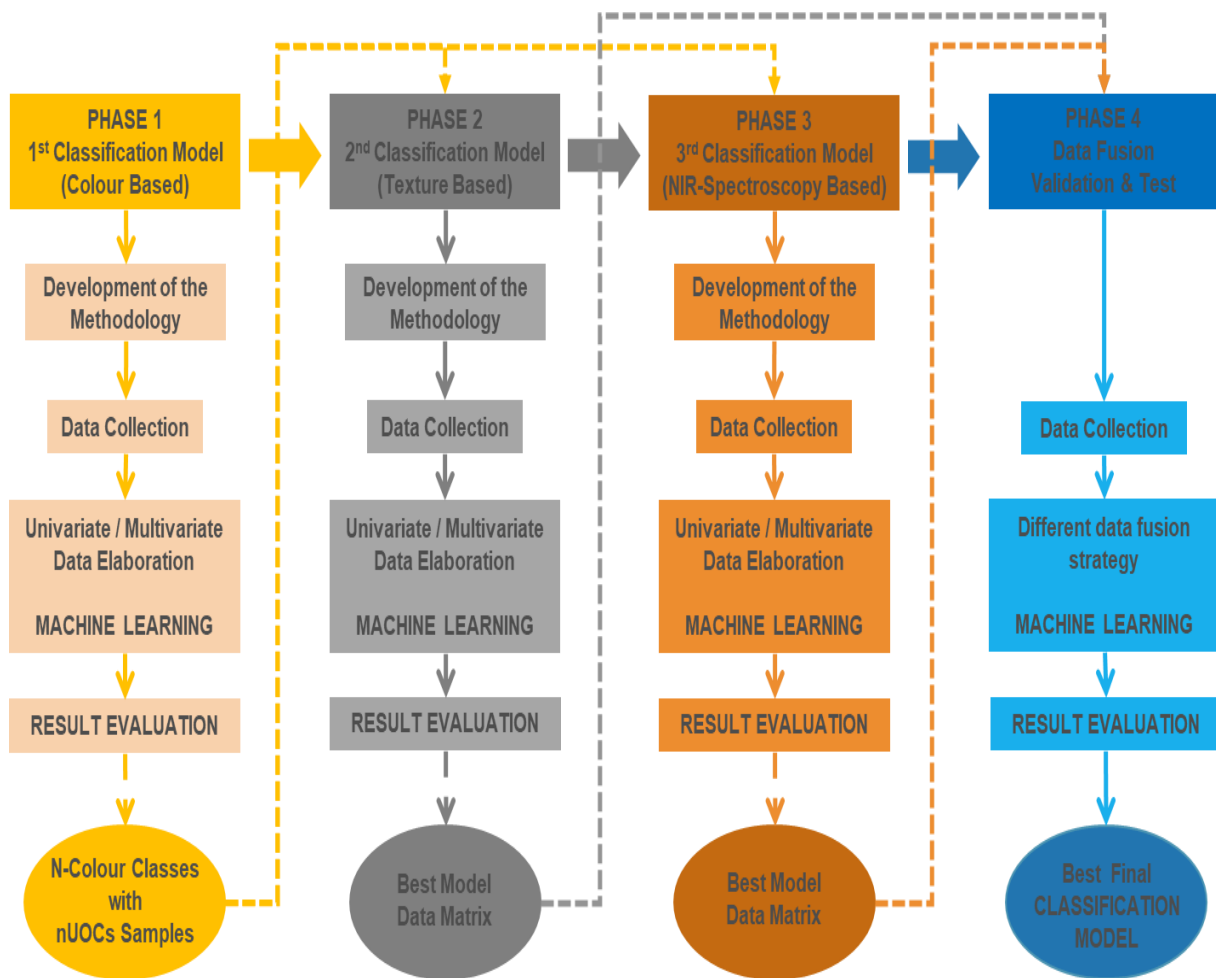


Figure 1. Work breakdown structure of the IDENTICALSS project.

It is important to note that a hierarchical modelling approach was followed allowing to reduce the number of samples to which a seized uranium powder can be attributed. As can be seen from Figure 1, the output of the phase 1 is the input for the following phases 2 and 3. In practice all the colour classes with a specific number of samples, that have been identified and chosen during the colour-based classification model development, are the input for the textural and hyperspectral based classification models development. This means that for each colour class there will be the correspondent textural and hyperspectral classification models. The same approach is applied for the phase 4 where the outputs (the experimental data or the models) of the phases 2 and 3 are used as the input for the phase 4.

2 Materials and methods

2.1 Materials

The samples investigated in the present study are 79 UOC powders which were mostly collected between 1950 and 2000. The powders exhibit colours grading from white/light yellow, yellow and orange to brown and black (Figure 2). A detailed list of the examined samples is shown in Table 1 together with the corresponding major chemical component.

Table 1. List of the analysed UOC samples.

Number	Sample (Country-Facility)	Major Chemical Composition	Number	Sample (Country-Facility)	Major Chemical Composition
1	England Wheal Edward	Unknown	41	Canada North Span	UO ₂ (OH) ₂
2	USA Kerr MgGee	(NH ₄) ₂ U ₂ O ₇	42	USA Utah	(NH ₄) ₂ U ₂ O ₇ + U ₃ O ₈
3	Spain Jen	(NH ₄) ₂ U ₂ O ₇	43	Australia Ranger	Na ₂ U ₂ O ₇
4	Australia Yeelirrie	Mixture	44	Germany Brunhilde	(NH ₄) ₂ U ₂ O ₇
5	S. Africa Rossing	U ₃ O ₈	45	Germany Helweiler	(NH ₄) ₂ U ₂ O ₇
6	S. Africa Mindola	Unknown	46	S. Africa Somair	Na ₂ U ₂ O ₇
7	USA Cotter	Na ₂ U ₂ O ₇	47	Portugal	Unknown
8	S. Africa EFI(Mouand)	(NH ₄) ₂ U ₂ O ₇	48	USA Lucky McGill	(NH ₄) ₂ U ₂ O ₇
9	USA Pathfinder	UO ₂ (OH) ₂	49	USA Everest Black	(NH ₄) ₂ U ₂ O ₇ + mixed oxide
10	Canada Stanrock	(NH ₄) ₂ U ₂ O ₇	50	Canada Rio Algom	(NH ₄) ₂ U ₂ O ₇
11	Holland Delft	Mixture	51	Canada Rabbit Lake	UO _{4.2} H ₂ O + U ₃ O ₈
12	S. Africa Maruzi	Unknown	52	USA Everest Yellow	UO _{4.2} H ₂ O
13	S. Africa Palabora	U ₃ O ₈	53	Sweden Ranstadt	Na ₂ U ₂ O ₇
14	Belgium Belgian Congo	UO ₂ (OH) ₂	54	USA Mesquite	UO _{4.2} H ₂ O ^{El}
15	Brazil Nuclebras	(NH ₄) ₂ U ₂ O ₇	55	USA Union Carbide	Mixture
16	Spain Enusa	(NH ₄) ₂ U ₂ O ₇	56	Canada Denison	(NH ₄) ₂ U ₂ O ₇
17	Australia Queensland	U ₃ O ₈	57	USA Atlas	U ₃ O ₈ + oxide
18	Germany Wismut	(NH ₄) ₂ U ₂ O ₇	58	Australia Mary Kathleen	U ₃ O ₈
19	USA Yankee Yellow	Na ₂ U ₂ O ₇	59	USA United Uranium	UO ₂ (OH) ₂
20	Canada Dyno	UO ₂ (OH) ₂	60	USA South Dakota	Mixed oxide + UO ₂ (OH) ₂
21	Canada Lake Key	U ₃ O ₈	61	Argentina	Na ₂ U ₂ O ₇
22	China Hengyang	U ₃ O ₈ +UO ₂	62	USA Federal American	U ₃ O ₈
23	USA Petromic	Mixed oxide	63	USA Dawn	(NH ₄) ₂ U ₂ O ₇
24	Canada Blind River	UO ₂ (OH) ₂	64	Canada Milliken Lake	(NH ₄) ₂ U ₂ O ₇
25	Canada Gunnair	UO ₂ (OH) ₂	65	Australia Rum Jungle	(NH ₄) ₂ U ₂ O ₇
26	Yugoslavia Spisak Black	UO ₂ (OH) ₂	66	Canada El Dorado	(NH ₄) ₂ U ₂ O ₇
27	Canada Faraday	UO ₂ (OH) ₂	67	Canada Ray Rock	UO ₂ (OH) ₂
28	Australia Olympic Dam	U ₃ O ₈		USA Chevron Hill	(NH ₄) ₂ U ₂ O ₇ + oxide
29	Yugoslavia Spisak Yellow	(NH ₄) ₂ U ₂ O ₇	69	S. Africa Nufcor	U ₃ O ₈
30	USA Mulberry	(NH ₄) ₂ U ₂ O ₇	70	Russia Technab	U ₃ O ₈
31	USA Falls City	Na ₂ U ₂ O ₇ + mixed oxide	71	USA United Nuclear	(NH ₄) ₂ U ₂ O ₇
32	Canada Madawaska	Oxide		Yugoslavia Rudnik	(NH ₄) ₂ U ₂ O ₇
33	USA Irigaray	UO _{4.2} H ₂ O	73	Australia South Alligator	UO ₂ (OH) ₂
34	USA ESI	(NH ₄) ₂ U ₂ O ₇	74	USA (Vermount) Yankee Black	(NH ₄) ₂ U ₂ O ₇ + mixed oxide
35	Canada Macassa	UO ₂ (OH) ₂	75	USA Sesquehann	UO ₂ (OH) ₂ + oxide
36	USA Anaconda	(NH ₄) ₂ U ₂ O ₇	76	Romania	Na ₂ U ₂ O ₇ + oxide
37	USA Shirley Basin	Na ₂ U ₂ O ₇ + oxide	77	USA EFI	(NH ₄) ₂ U ₂ O ₇
38	USA Mobil	UO _{4.2} H ₂ O	78	USA Homestake	(NH ₄) ₂ U ₂ O ₇
39	Canada ESI	(NH ₄) ₂ U ₂ O ₇	79	Canada Stanleigh	(NH ₄) ₂ U ₂ O ₇
40	Australia Radium Hill	(NH ₄) ₂ U ₂ O ₇			

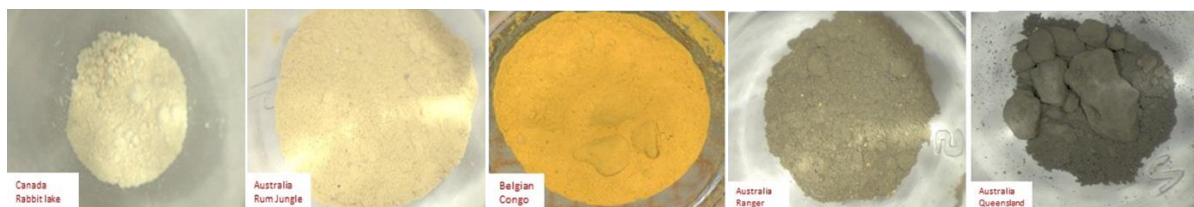


Figure 2. Subset of the samples under investigation. The powders shown are representative of some of the different colours possessed by the UOCs.

2.2 Methods

2.2.1 Colour analysis by spectrophotometer measurements

The colour analysis was carried out via spectrophotometric measurements. This technique allowed the creation of predefined colour classes in which seized samples can be attributed objectively before being analysed with the other techniques. The colour spectra have been measured using a diffuse-reflected spectrophotometer, a Konika Minolta CM-700d, equipped with a silicon photodiode array detector, an integrating sphere having a diameter of 40 mm, a xenon lamp with a UV cut filter operating 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). The samples were introduced in quartz vials (75 x 10 mm) or borosilicate glass vials (45 x 14.7 mm) for the colour measurements and two types of data were collected: the reflectance values in the wavelengths range of 360 nm to 740 nm and with a resolution of 10 nm, and the “tristimulus” values in terms of L^* , a^* , b^* in the CIE 1976 $L^*a^*b^*$ colour space [21-22].

At the end 38 variables from the spectral data + 3 variables from the $L^*a^*b^*$ colour space, have been acquired from each sample and used to create the spectrophotometric dataset for the colour based classification model development. Each sample has been measured 5 times and only the average was taken into account [23].

2.2.2 Scanning electron microscopy and image texture analysis

SEM images of the UOC samples were acquired using a FIB/SEM FEI Versa 3D in low-vacuum mode (pressure = 10 Pa) equipped with a concentric backscattered detector for backscattered electrons.

To develop the texture-based classification model, the textural images have been collected (using an in-house produced sample holder in graphite) following the scheme showed in Figure 3. For each specimen, 3 independent sample preparations were done and 5 images were acquired at three different magnifications of 100x, 250x and 500x or 1000x for each sample preparation. Each sample has an image data set of 45 independent images (15 for each magnification), corresponding to a whole image dataset of 3555 images.

As an example, two image data sets (250x and 1000x of magnifications) with all the textural images of the samples belonging to colour class 1, are shown in the Figure 4.

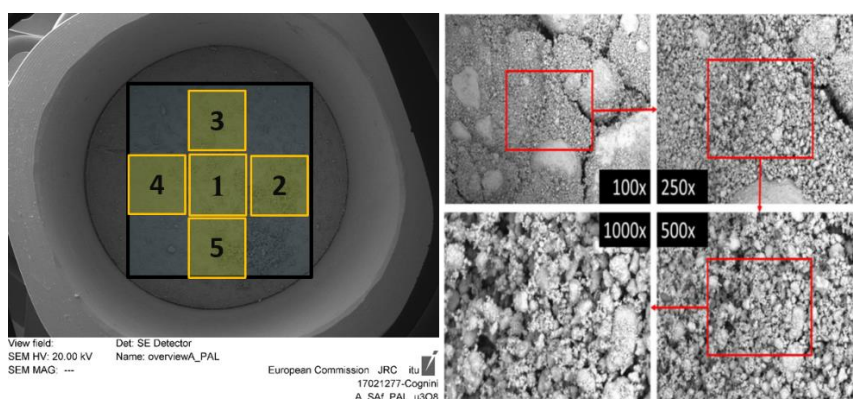


Figure 3. Sample holder with the 5 different regions from which the textural images have been acquired. From each square, the images at 100x, 250x and 500x/1000x of magnification were collected maintaining always the same centre.

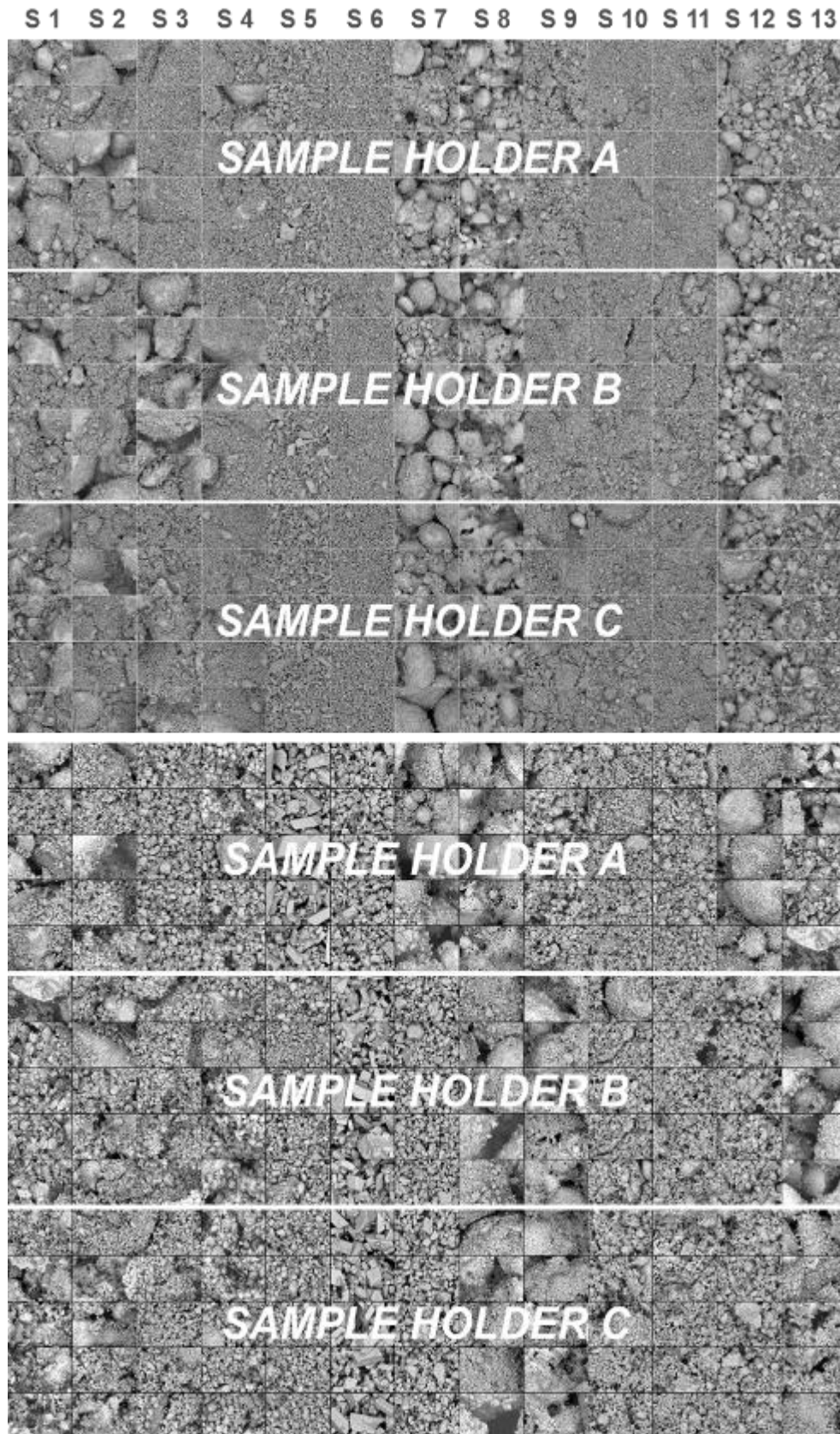


Figure 4. Image texture dataset of samples belonging to the colour class 1. The images represented have been acquired at 250x (top) and 1000x (bottom) of magnification.

2.2.3 Image texture analysis

There are different definitions of image texture, but in general this is an image property highly dependent on the physical surface characteristic of the object analysed (surface topography, morphology, surface texture): it could be defined as the spatial distribution, frequency and grey level value of each pixel composing the image. Different approaches can be used for image texture evaluation; statistical, structural and spectral [17].

To develop the texture-based classification model, different texture features extraction algorithms, belonging to all the three approaches, were used in order to create the textural features datasets; in particular, the texture image datasets were analyzed with the following methods (their description can be found in ANNEX 1):

First Order Statistics (FOS).

Grey Level Co-Occurrence Matrix (GLCM).

Grey Level Run Length Matrix (GLRLM).

Grey Level Size Zone Matrix (GLSZM).

Local Binary Pattern (LBP).

Angle Measure Technique (AMT).

The lists of each of the extracted textural features used for the texture-based classifications models development are listed from Table 10 to Table 12 in ANNEX 1.

2.2.4 Hyperspectral image analysis

The hyperspectral images have been acquired using the SisuCHEMA hyperspectral Chemical Imaging Analyser (SPECIM, Spectral Imaging LTD, Oulu, Finland) system (Figure 5). This system employs a push-broom imaging technology providing several advantages: high speed, less heat load from illumination and flexibility to most sample shapes and sizes. The push-broom technology allows to acquire images one line at time while scanning the sample on a sliding table. Each line could have 320 to 1312 pixels' field of view in function of the hyperspectral camera spatial resolution. In the scanning dimension, the number of lines (refers to the final image height) is dependent on the selected scanning length. The system is composed of:

Scanner table: with a maximum scanning rate of 60 mm/s with a spatial resolution of 600 μm , maximum sample size is 200 x 300 x 45 mm (W x L x T), integrated illumination system with the SPECIM's diffusive line illumination unit.

FX17 hyperspectral camera: with spectral range of 900-1700 nm, spectral resolution of 8 nm, spatial resolution of 640 pixels, spectral sampling/pixel of 3.5 nm, free wavelengths selection from the 224 bands within the camera coverage, built-in image correction.



Figure 5. *SisuCHEMA system installed at the JRC Karlsruhe.*

To acquire the hyperspectral images, the same sample holder was used as to acquire the SEM images. For each sample holder was acquired 1 hyperspectral image. The final hyperspectral images dataset is represented by a total of 237 hyperspectral images.

Figure 6 shows the process to create the final hyperspectral image dataset.

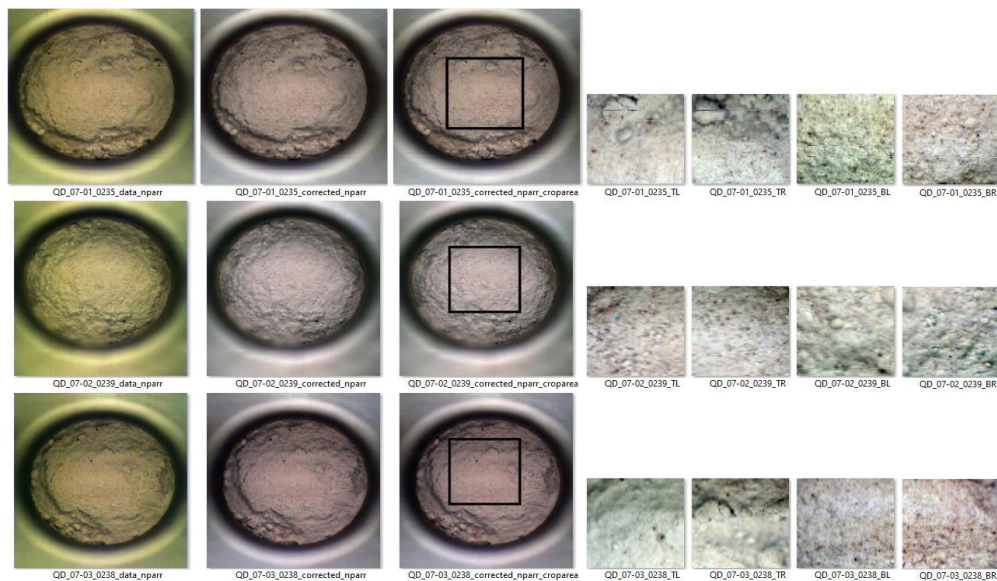


Figure 6. Procedure to create the hyperspectral images necessary to build the corresponding hyperspectral image dataset. The columns show (from left to right) the raw hyperspectral images, the reflectance calibrated images, the reflectance calibrated images with the selected area from which the sub images were cropped, and finally the 4 square sub images cropped from the selected area.

The raw hyperspectral images acquired for each sample were pre-processed before the spectra were extracted and inserted into datasets. The spectra used in this study were first reflectance calibrated and then converting the spectra to absorbance, and baseline corrected [23]. The effect of the reflectance calibration on the appearance of a hyperspectral image can be clearly seen in Figure 7, where the vertical noise appears to be removed.

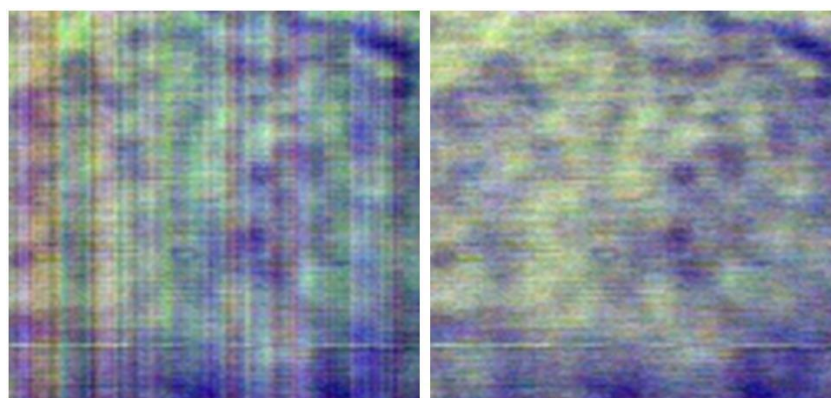


Figure 7. Hyperspectral image before (left) and after (right) the reflectance calibration using the white and dark reference images.

In ANNEX 2 can be found, as an example, the spectra extracted from the color class 3 after each pre-processing steps.

2.2.5 Models developments by Machine Learning

The colour based classification model was developed using PLSToolbox v. 8.7 (Eigenvector Research, Seattle, USA) for Matlab v R2018b (MathWorks Inc, USA). It was carried out, at the beginning, different explorative analysis by PCA; later a hierarchical cluster analysis (HCA) was performed to select the number of different

colour groups in which the UOCs samples should be classified and finally support vector machine (SVM) was used as classifier to build the supervised colour-based classification model.

To build the textural and hyperspectral classification models, different codes were written by Isak Lande in the programming language Python. An extensive description of these codes and their dependencies can be found in [23]. The analysis: image texture features extraction, infrared spectra extraction and pre-processing, machine learning for the models development, was conducted through the integrated development environment (IDE) Spyder (v. 3.3.6). Only the AMT features were extracted using the jAMT Explorer plugin in the open source Java image processing program imageJ [24]. In general, the computational time required from the codes, to elaborate and analyse the data, was rather long reaching a maximum runtime of 12 hours.

The process for finding the most promising classifier for the UOCs classification (based on textural features) was found using the images acquired at 250x that had been attributed by the colour based classification model to the colour class 1 named Black-Dark Brown (see section 3.1).

Several classifiers underwent a screening to find the best one. The screening was done by performing nested cross-validation (n-CV) and this was done on datasets consisting of the different features groups from AMT, GLCM, GLRLM, GLSZM, LBP, all (i.e. containing all the combined group of features). The investigated classifiers were Logistic Regression (LR), Support Vector Machine (SVM), GaussianNB (NB), Linear Discriminant Analysis (LDA), Random Forest Classifier (RF), KNeighbors Classifier (KNN) [23].

The best classifier found on the first dataset was applied also on the other SEM image datasets and on the spectral datasets. Figure 8 illustrate the process of arriving at the best (most promising) classifier.

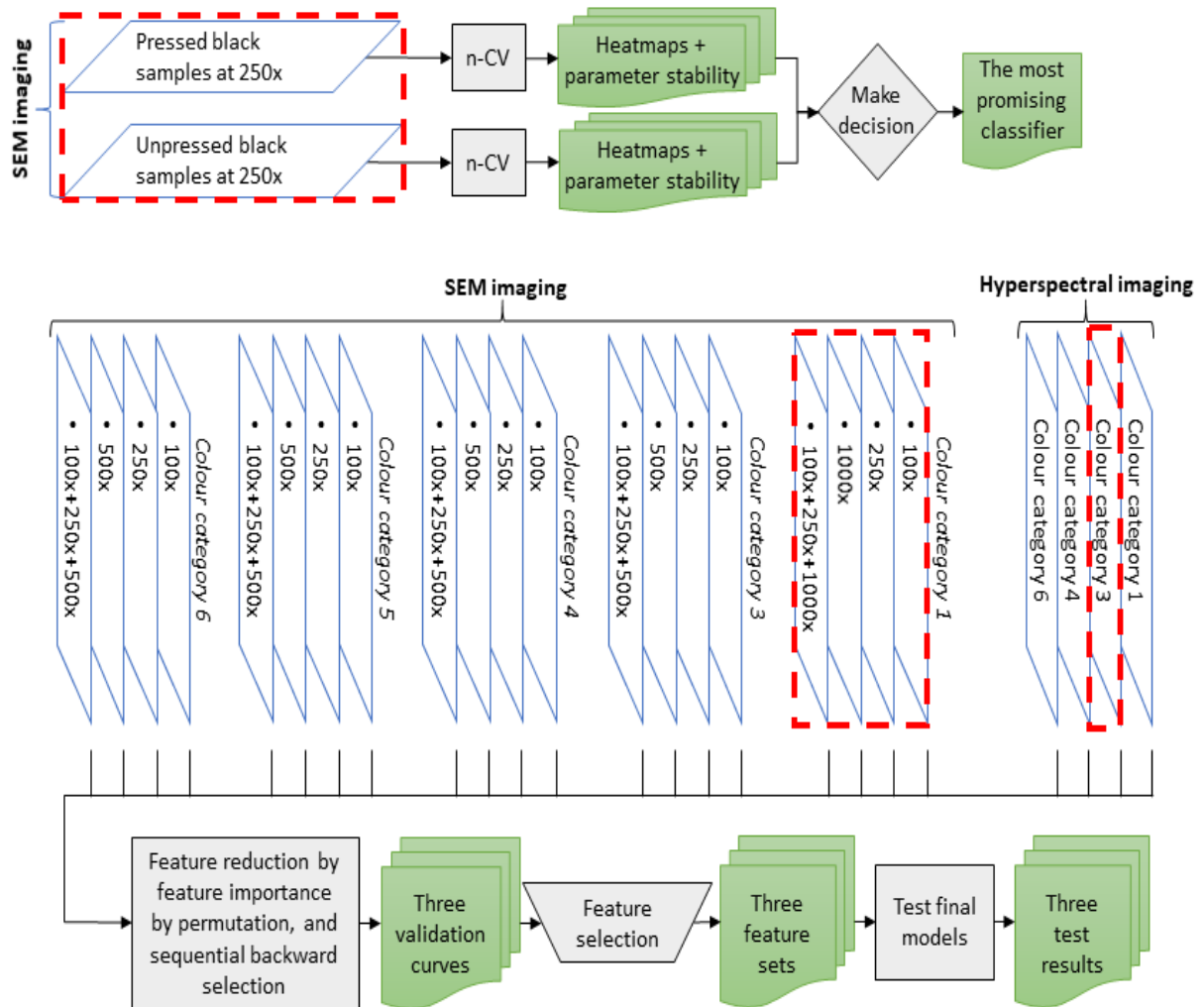


Figure 8. Steps used for finding the best classifier and the dataset used (red dotted squares).

At the end, the classifier was trained on optimized feature sets for each dataset, as a final model for each colour category, and their performance was tested on hold-out test data. As outlined in Figure 9, the dataset

analysed were split three times into a training set and hold-back test set. All samples in the three test sets were unique. The test set consisted of one single sample from each class. For each of these three training sets, an optimised feature set was determined, resulting in three feature sets.

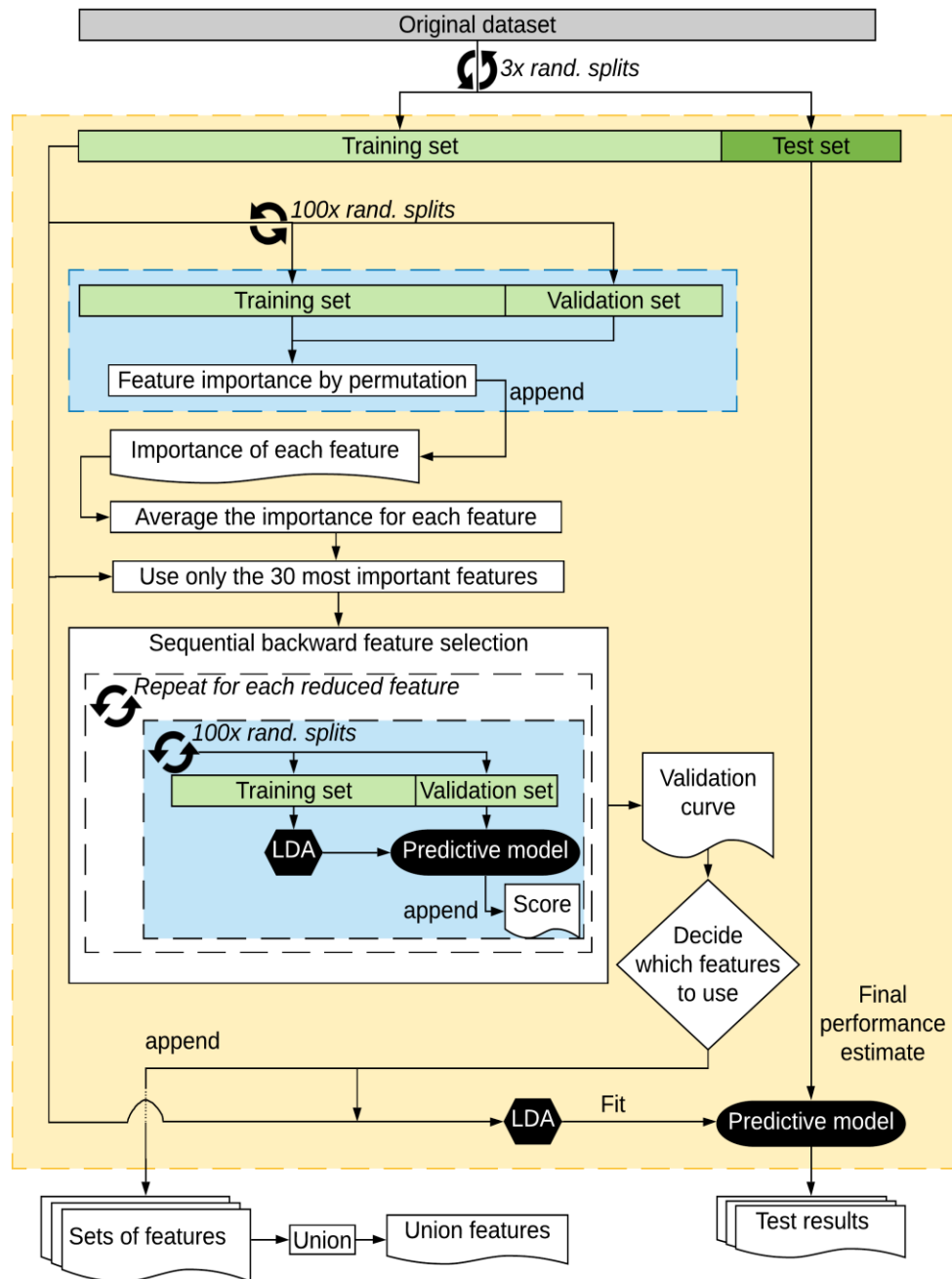


Figure 9. Overview of the process of finding optimized feature sets for a dataset and estimating their performance on hold-out test data.

An additional code was also created to develop a model able to classify “new” (unknown) UOC samples. The information needed to predict these unknown samples were: the colour class in which the unknown sample is attributed and the type of image acquired (SEM with magnification or HSI). The code would then retrieve the features that were optimized during the model development. The feature set would then be extracted from the unknown UOC samples and used as test data by the model to classify the sample. The unknown sample must consist of the same resolution and have underwent the same pre-processing to be applicable for the predictive model.

3 Results

During the first year of the exploratory research, we were able to develop the procedure to measure the colour of the UOC samples, to collect all the data and to build the colour-based classification model. The procedure to acquire the textural images was also developed and they were acquired for most of the samples. During the second year, the SysuCHEMA system was purchased, delivered and installed into the laboratory. The methodology of the hyperspectral image acquisition was also developed and the hyperspectral images were collected. The collection of the textural and hyperspectral image datasets for all samples was not possible for various reasons. First of all, the FIB that was used to collect the textural images was undergoing its “nuclearization” (attached into the glove-box), and it hindered the work considerably. Secondly, the delivery of the SysuCHEMA system was delayed. Anyhow, the image data sets collected were enough to develop the first correspondent textural and hyperspectral classification models. Table 2 shows the samples for which the colour data, textural and hyperspectral images have been collected, processed and analysed.

Table 2. List of samples with the corresponding data that have been collected and analysed during the exploratory research. Yes, means that the data have been acquired and analysed; Yes/No, means that the data have been only acquired and need to be still analysed; No, means that the data need to be still acquired and analysed.

ID Number	Sample	C	SE	SP	ID Number	Sample	C	SE	SP
1	England-Wheal Edward	Yes	Yes	No	41	Canada North Span	Yes	No	Yes
2	USA-Kerr McGee	Yes	Yes	No	42	USA Utah	Yes	No	Yes
3	Spain-Gen	Yes	Yes	Yes	43	Australia Ranger	Yes	Yes	Yes
4	Australia- Yeelirre	Yes	Yes	No	44	Germany -Brunhilde	Yes	Yes	No
5	Nambia-Rossing	Yes	Yes	No	45	Germany -Helwiler	Yes	Yes	Yes
6	Zambia-Mindola	Yes	Yes	Yes	46	Niger-Somair	Yes	Yes	No
7	USA Cotter	Yes	Yes	Yes	47	Portugal	Yes	No	No
8	Gabon-EFI(Mouand)	Yes	Yes	No	48	USA-Lucky McGill	Yes	Yes	Yes
9	USA-Pathfinder	Yes	Yes	No	49	USA-Everest-Black	Yes	No	No
10	Canada-Stamrock	Yes	Yes	No	50	Canada-Rio Algom	Yes	No	No
11	Holland Delft	Yes	Yes	Yes	51	Canada-Eldore(Rabbit Lake)	Yes	Yes	No
12	Mozambique-Maruzi	Yes	No	No	52	USA-Everestr-Yellow	Yes	Yes	Yes
13	S.Africa-Palabora	Yes	Yes	Yes	53	Sweden Ranstadt	Yes	Yes	Yes
14	Belgian-Congo	Yes	Yes	No	54	USA-EI Mesquite	Yes	Yes	No
15	Brazil Nuclebras	Yes	Yes	Yes	55	USA-Union Carbide	Yes	No	No
16	Spain-Enusa	Yes	No	No	56	Canada-Denison	Yes	Yes	No
17	Australia-Queensland	Yes	Yes	No	57	USA-Atlas	Yes	Yes	No
18	Germany Wismut	Yes	No	Yes	58	Australia-Mary Kathleen	Yes	Yes	Yes
19	USA Yankee Yellow	Yes	Yes	No	59	USA-United Uranium	Yes	No	No
20	Canada-Dyno	Yes	Yes	Yes	60	USA-South Dakota	Yes	Yes	No
21	Canada-Key Lake	Yes	Yes	Yes	61	Argentina	Yes	No	No
22	China-Hengyang	Yes	Yes	Yes	62	USA-Federal American Partners	Yes	Yes	No
23	USA-Petromic	Yes	No	No	63	USA Dawn	Yes	Yes	No
24	Canada-Blind river	Yes	Yes	No	64	Canada Milliken Lake	Yes	Yes	Yes
25	Canada-Sunnar	Yes	No	Yes	65	Australia Run Jungle	Yes	Yes	Yes
26	Yugoslavia Spisak Black	Yes	Yes	No	66	Canada El Dorado	Yes	Yes	No
27	Canada Faraday	Yes	No	Yes	67	Canada Ray Rock	Yes	No	No
28	Australia Olympic Dam	Yes	Yes	Yes	68	USA Chevron HILP	Yes	No	No
29	Yugoslavia-Spisak-Yellow	Yes	Yes	Yes	69	S.Africa Nufcor	Yes	Yes	No
30	Usa-Mulberry(IMC)	Yes	Yes	Yes	70	Russia-Techsnab	Yes	Yes	Yes
31	USA-Falls City	Yes	Yes	Yes	71	USA-United Nuclear	Yes	No	No
32	Canada-Madawaska	Yes	Yes	No	72	Yugoslavia-Rudnik	Yes	Yes	Yes
33	USA-Irigaray	Yes	Yes	Yes	73	Australia S Alligator	Yes	Yes	Yes
34	USA-ESI	Yes	Yes	No	74	USA Yankee Black	Yes	No	No
35	Canada-Macassa	Yes	Yes	No	75	USA Sesquehan	Yes	No	No
36	USA-Anaconda	Yes	Yes	Yes	76	Rumania	Yes	Yes	Yes
37	USA-Shirley Basin	Yes	No	No	77	USA-Efi(White Messa)	Yes	No	No
38	USA-Mobil	Yes	Yes	Yes	78	USA Homestake	Yes	Yes	Yes
39	Canada-ESI	Yes	Yes	No	79	Canada Stanleigh	Yes	No	No
40	Australia-Radium Hill	Yes	Yes	Yes					

As can be seen from the Table 2, the collection of textural images is still missing for 22 samples, and hyperspectral images for 44 samples, while the textural and spectral features have to be extracted from 30 and 48 samples, respectively.

3.1 Colour based classification model development

In case of discovery of uranium powder outside of the regulatory control, the provenance and the origin of the material should be understood, preferably as soon as possible. After the nuclear forensics analysis and characterisation of the seized sample, the results may be compared with the data available in nuclear forensics databases. Similarly, in the context of the exploratory research, the data of an unknown sample may be compared with the data of the 79 samples used as training samples in this study.

As reported by Klunder et al. [25], the colour of UOCs powders is usually the first and easiest discriminator. The colour is an indicator to be correlated to the processes used for the UOCs production; in particular, in terms of different reagents, separation procedures and drying conditions reflecting the U compound in question.

The idea behind the creation of colour classes, in which the 79 industrial UOCs powder samples can be grouped, is the following: instead to compare an unknown sample to all the samples in the database, so in this case 1 to 79, could be done a comparison between 1 to X number of samples, depending on how many samples there will be in the colour class in which the unknown sample will be attributed by the colour-based classification model. Basically in this way the number of samples to which the unknown sample can be assigned is reduced to the number of samples belonging to each colour class.

It is true that the colour of one sample can be defined by a simple subjective evaluation, as suggested by Klunder et al. [25], but to create colour classes, a more structural and mathematical approach is necessary.

After the creation of the dataset with the spectrophotometric and $L^*a^*b^*$ colour values, the first approach was to carry out a simple explorative analysis by Principal Component Analysis (PCA), in order to understand if there were some similarities among all the sample. It was clear that there was some trend in the sample representation, as can be seen in Figure 10a.

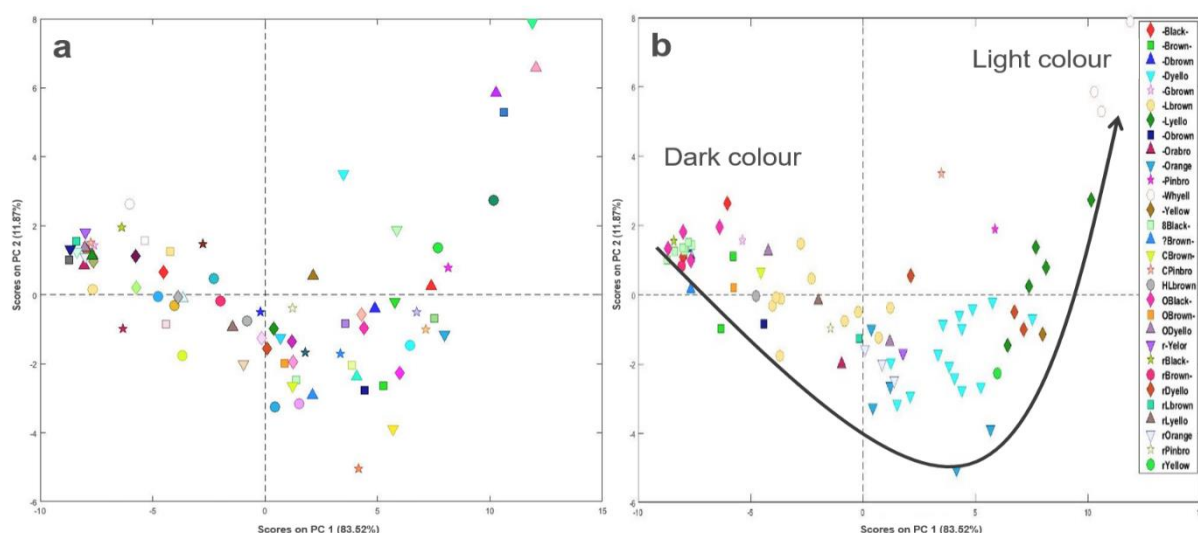


Figure 10. Sample distribution in the score plot before (a) and after (b) a subjective predefined colour labels attribution to all the samples.

In order to better understand the sample distribution in function of their colour (score plot of Figure 10a), a colour label to each sample (based on a visual inspection) have been attributed. These colour labels have been used afterwards as category variables in the data matrix and a PCA was repeated.

Looking the score plot of Figure 10b, the trend is more visible and understandable: the colour changes mainly along the PC1 where on the left side there are sample characterised with dark colours while on the right side there are samples characterised with light colours; and it seems that the samples can be grouped in different clusters.

This was not sufficient to establish in how many colour classes the samples should have been grouped. Thus, in order to understand better how many classes have to be created, a hierarchical cluster analysis (Clustering analysis based Mahalanobis distance) was performed. Figure 11 shows the connection dendrogram. The clusters can be chosen selecting a threshold, by intersecting the horizontal lines with a sliding vertical line; the intersection points defines the number of clusters. In this case, six colour-groups (colour classes) were chosen (from colour class 1 to colour class 6; CC1 - CC6) and they were labelled as: Black/Dark Brown (B-DB), Dark

Brown/Orange (DB-O), Orange/Dark Yellow (O-DY), Dark Yellow/Yellow (DY-Y), Yellow/Light Yellow (Y-LY), Light Yellow/White (LY-W), as can be seen in Table 2. It is worthwhile to highlight that peroxides belong mostly to the CC6 (LY-W), ammonium diuranate, sodium diuranate and hydroxides to CC5, CC4 and CC3 (Y-LY, DY-Y, O-DY), mixed compounds fall in the CC2 (DB-O) whilst the B-DB group (CC1) contains mostly oxides. This categorisation reflects the capability of the cluster analysis to intercept variations in the powders chemical composition that is reflected into their colours appearance. Differences in the colour nuances for powders having identical chemical composition (e.g. yellow and dark yellow) might be the consequence of a specific production process. Figure 12 shows the scattergram with the sample distribution in function of their colour index values (a^* vs L^*): the x-axis corresponds to a^* values, z-axis corresponds to the L^* values and the dimension of each point (sphere diameter) corresponds to the b^* values.

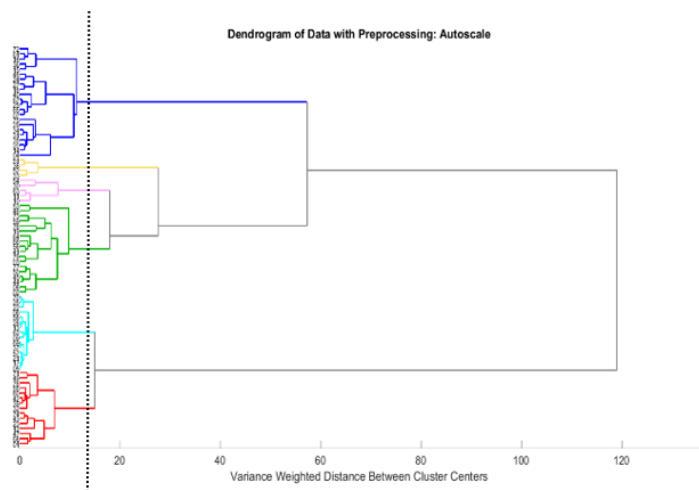


Figure 11. Resulting dendrogram after the Hierarchical cluster analysis of the entire samples collection. Different clusters are possible by selecting different threshold level in the dendrogram. In this work 6 different colour-groups were chosen.

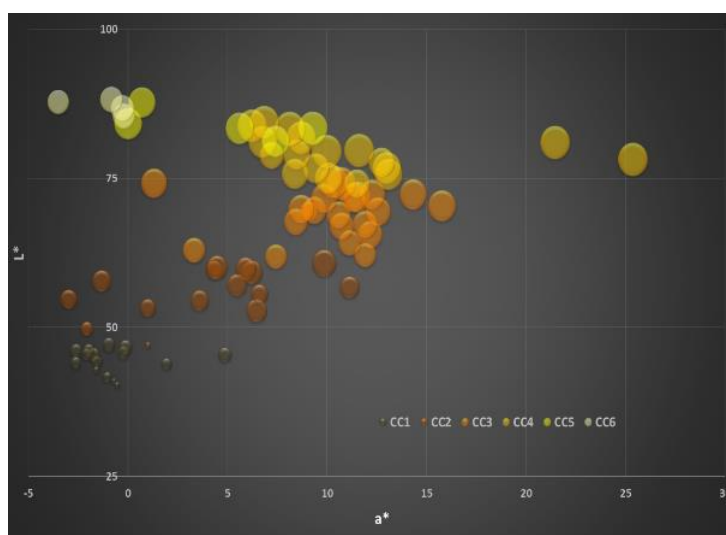


Figure 12. Scattergram representing the sample distribution in function of their a^* and L^* values. The diameter of each sphere correspond to the b^* values.

The label names were chosen in such a way that each one should indicate a transition from one colour class to the next one and not one single colour. This is because few samples are overlapping, in particular as can be seen from Figure 12, for colour class 3, 4, 5 and 6. This suggests that their colour index values are close and the samples could have in some cases a very similar hue. This means that the colour of those samples can be perceived very similar by visual inspection. Table 3 reports details about the 6 colour classes, including the colour labels and the number of UOCs samples that ended up in each class. In addition, the ΔE value between the consecutive colour classes have been also calculated. ΔE is a single number representing the "distance" between two colours and it is tempting to simply compare the Euclidean distance difference between the L^* , a^* and b^* colour values in the $L^*a^*b^*$ colour space [21-22].

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

ΔE was calculated between CC1-CC2, CC2-CC3, CC3-CC4, CC4-CC5 and CC5-CC6 in order to understand if the colours of the neighbour colour classes can be considered different. In fact, generally a result of ΔE value less than 2 indicates that the colours compared can be considered to be perceptually equivalent.

Table 3. Details about the 6 colour classes obtained after the cluster analysis: average values of L^* , a^* and b^* index with the relative standard deviation and the ΔE values, the range and the colour class labels are reported.

Colour Class (CC)	Colour Class Label	N° of Sample	L^*	MIN_{L^*}	MAX_{L^*}	a^*	MIN_{a^*}	MAX_{a^*}	b^*	MIN_{b^*}	MAX_{b^*}	ΔE
CC1	Black / Dark Brown	15	44.4±2.1	40,36	47,09	-0.7±1.9	-2,63	4,86	6.9±3	1,65	11,08	
CC2	Dark Brown / Orange	15	55.9±4	47,06	60,76	4±4.2	-2,98	11,13	20.6±8.1	1,72	35,36	18
CC3	Orange / Dark Yellow	22	69.3±3.9	61,88	74,39	10.3±3.1	1,3	15,78	37.7±7.8	17,68	52,205	22
CC4	Dark Yellow / Yellow	18	79.3±2.9	75,02	84,36	11±5.2	6,24	25,37	48.2±5.9	34,18	55,9	14
CC5	Yellow / Light Yellow	5	84±2.4	81,18	87,71	4.6±4.1	0,01	9,28	46.9±3.5	43,03	52,53	8
CC6	Light Yellow / White	4	87.1±1.3	85,43	88,23	-1.2±1.6	-3,5	-0,22	30.1±2.4	27,26	33,04	18

As can be seen from Table 3, ΔE index suggests that the colours of the 6 colour classes can be considered being different.

Continuing with the explorative analysis, a new PCA was carried out one more time, using the colour class chosen with the cluster analysis as category variables for each sample. As can be seen in Figure 13, now is perfectly clear how the samples are distributed in the score plot.

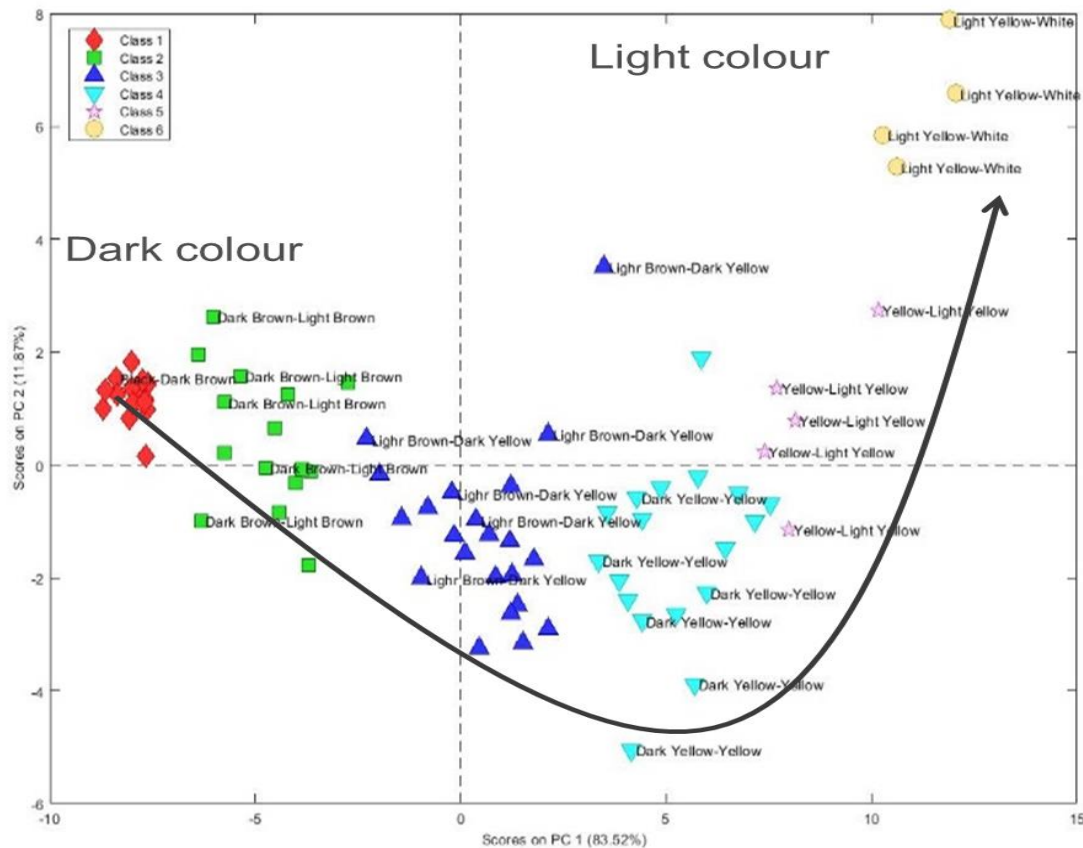


Figure 13. Score plot of colour data using the colour classes, defined by HCA, as category variables.

After the creation of the colour classes, Super Vector Machine Discriminant Analysis (a supervised classification algorithm) was selected as classifier algorithm to build the colour-based classification model. Also in this case, the name of the classes has been used as prior information and as category variables.

The model's performance, obtained in cross-validation, and in terms of sensitivity and specificity values for each colour class, are reported in Table 4 [26]. The model in cross-validation has a Matthews Correlation Coefficient average value of 0.95 which is closed to the maximum possible value of 1 (MCC was used due to its

capability to cope with imbalances in datasets [27–28]). 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 Orange sample is classified as Yellow (see Figure 14). Also in this case the MCC average value is 0.95; this indicates, together with the sensitivity and specificity values in cross-validation and prediction reported in Table 4, that the colour based classification model developed works properly.

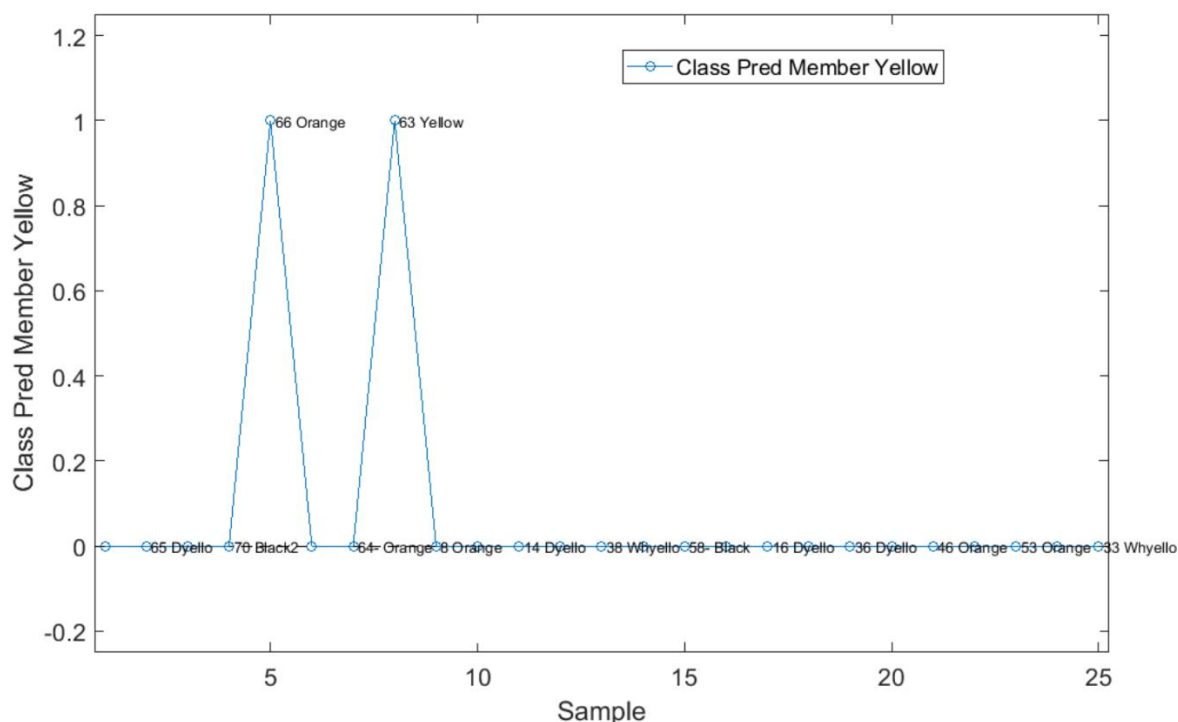


Figure 14. Samples of test-set classified in the Yellow group. The classification prediction membership is measured either by a 0 (non-membership) or by a 1 (membership). Specifically, only the sample 66 is misclassified.

Table 4. Evaluation metric, in cross-validation (CV) and prediction (P), of the colour-based classification model developed with SVM classifier.

COLOUR CLASS	CC1	CC2	CC3	CC4	CC5	CC6
Sensitivity (CV)	1	0,93	1	0,96	0,8	1
Specificity (CV)	1	1	0,97	0,98	1	1
Sensitivity (P)	1	1	0,89	1	1	1
Specificity (P)	1	1	1	1	0,96	1

Table 5 lists the six different colour classes with all the UOCs samples that have been assigned to each class by the colour-based classification model. For each sample it is reported if the SEM and hyperspectral images have been acquired and analysed in order to understand how many samples have been used to develop the relative texture-based and spectral-based classification model for each colour class. When known, information about the different steps: in terms of dissolution, extraction, precipitation and dry/calcination processes, used during the production, are also reported.

Table 5. List of the six different colour classes with all the UOCs samples belonging to each class. Yes, means that the data have been acquired and analysed; Yes/No means that the data have been only acquired and must be analysed; No means that the data must be acquired and analysed. Information about the main steps of the production process have been reported (IX, ion exchange; SX, solvent extraction).

COUNTRY / FACILITY	MAJOR COMPOSITION	DISSOLUTION PROCESS	EXTRACTION PROCESS	PRECIPITATION PROCESS	DRY/CALCINATION	COLOUR CLASS	COLOUR NAME	COLOUR CODE	COLOUR	SEM	SPECTRA
S. Africa-Rossing	U ₃ O ₈	Acid Leach	IX, Strip, SX	NH ₃	Calcinated - 500 °C	Class 1	Black-Dark Brown	B-DB	Yes	Yes	No
S. Africa-Mindola	Unknown	Unknown	Unknown	Unknown	Unknown	Class 1	Black-Dark Brown	B-DB	Yes	Yes/No	Yes
USA-Pathfinder	UO ₂ (OH) ₂	Acid Leach	Eluex Process	NH ₃	N/A	Class 1	Black-Dark Brown	B-DB	Yes	Yes	No
S. Africa-Palabora	U ₃ O ₈	Acid Leach	SX, Strip	NH ₃	Calcinated - 700 °C	Class 1	Black-Dark Brown	B-DB	Yes	Yes	Yes/No
Australia-Queensland	U ₃ O ₈	Acid Leach	SX, Strip	NH ₃	Calcinated - 500 °C	Class 1	Black-Dark Brown	B-DB	Yes	Yes	No
Canada-Key Lake	U ₃ O ₈	Acid Leach	SX, Strip	NH ₃	Calcinated - 750 °C	Class 1	Black-Dark Brown	B-DB	Yes	Yes	Yes
China-Hengyang	U ₃ O ₈ + UO ₂	Acid Leach	IX	N/A	N/A	Class 1	Black-Dark Brown	B-DB	Yes	Yes	Yes
Yugoslavia-Spisak Black	UO ₂ (OH) ₂	Unknown	Unknown	Unknown	Unknown	Class 1	Black-Dark Brown	B-DB	Yes	Yes	No
Australia-Olympic Dam	U ₃ O ₈	Acid Leach	SX, Strip	NH ₃	Calcinated - 750 °C	Class 1	Black-Dark Brown	B-DB	Yes	Yes	Yes
USA-Shirley Basin	Na ₂ U ₂ O ₇ + oxide	N/A	N/A	N/A	N/A	Class 1	Black-Dark Brown	B-DB	Yes	No	No
USA-Atlas	U ₃ O ₈ + oxide	Pressure Leach / Acid Leach	SX, Strip	NaOH, H ₂ O ₂ & NH ₃	Calcinated - 650 °C	Class 1	Black-Dark Brown	B-DB	Yes	Yes	No
Australia-Mary Kathleen	U ₃ O ₈	Acid Leach	SX, Strip	MgO, NH ₃	Calcinated - 700 °C	Class 1	Black-Dark Brown	B-DB	Yes	Yes	Yes/No
USA-Federal American Partners	U ₃ O ₈	Acid Leach	Eluex Process	NH ₃	Calcinated - 600 °C	Class 1	Black-Dark Brown	B-DB	Yes	Yes	No
S. Africa-Nulcor	U ₃ O ₈	Acid Leach	IX	NH ₃	Calcinated	Class 1	Black-Dark Brown	B-DB	Yes	Yes	No
Russia-Technab	U ₃ O ₈	N/A	N/A	N/A	N/A	Class 1	Black-Dark Brown	B-DB	Yes	Yes	Yes
England-Wheat Edward	Unknown	N/A	N/A	N/A	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	Yes	No
Mozambique-Maruzi	Unknown	N/A	N/A	N/A	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	No
USA-Petromic	Mixed oxide	Acid Leach	SX, Strip	MgO	Calcinated - 400 °C	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	No
Canada-Madawaska	Oxide	Acid Leach	IX, Strip	MgO	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	Yes	No
Canada-ESI	(NH ₄) ₂ U ₂ O ₇	N/A	N/A	N/A	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	Yes	No
USA-Utah	(NH ₄) ₂ U ₂ O ₇ + U ₃ O ₈	N/A	N/A	NH ₃	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	Yes
Australia-Ranger	Na ₂ U ₂ O ₇	Acid Leach	SX, Strip	NH ₃	(Calcinated)	Class 2	Dark Brown-Light Brown	DB-LB	Yes	Yes	Yes
Portugal-Mine Unknown	Unknown	N/A	N/A	N/A	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	No
USA-Everest-Black	(NH ₄) ₂ U ₂ O ₇ + mixed oxide	N/A	N/A	N/A	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	No
USA-Union Carbide	Mixture	Acid Leach	SX, Strip	UO	UO	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	No
USA-South Dakota	UO ₂ (OH) ₂ + mixed oxide	Acid Leach	IX	MgO	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	Yes	No
USA-Chevron HILL	U ₃ O ₈	Acid Leach	SX	NH ₃	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	No
USA-United Nuclear	(NH ₄) ₂ U ₂ O ₇	Carbonate Leach	N/A	NaOH, NH ₃	(Calcinated)	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	No
USA-Vermont Yankee Black	(NH ₄) ₂ U ₂ O ₇ + mixed oxide	N/A	N/A	N/A	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	No	No
USA-Homestake	(NH ₄) ₂ U ₂ O ₇	N/A	N/A	N/A	N/A	Class 2	Dark Brown-Light Brown	DB-LB	Yes	Yes	Yes
Australia-Yeelirre	Mixture	Alkaline Leach	N/A	NH ₃	Calcinated - 260/650 °C	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	No
USA-Cotter	Na ₂ U ₂ O ₇	Pressure Leach / Acid Leach	SX, Strip	NaOH, H ₂ SO ₄ & H ₂ O ₂	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes
Belgian-Congo	UO ₂ (OH) ₂	N/A	N/A	N/A	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	No
Brazil-Nuclebras	(NH ₄) ₂ U ₂ O ₇	Acid Leach	SX, Strip	Possibly NH ₄ OH	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes
Spain-Enusa	(NH ₄) ₂ U ₂ O ₇	Heap Leach	SX	N/A	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	No	No
Germany-Wismut	(NH ₄) ₂ U ₂ O ₇	N/A	N/A	N/A	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes
Canada-Dyno	UO ₂ (OH) ₂	Acid Leach	SX, Strip	MgO	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes
Canada-Blind river	UO ₂ (OH) ₂	N/A	N/A	N/A	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes/No	No
Canada-Sunnar	UO ₂ (OH) ₂	Acid Leach	IX, Strip	MgO	Dried	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	No	Yes
Canada-Faraday	UO ₂ (OH) ₂	Acid Leach	IX, Strip	MgO	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	No	Yes
Yugoslavia-Spisak-Yellow	(NH ₄) ₂ U ₂ O ₇	Unknown	Unknown	Unknown	Unknown	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes
USA-Falls City	Na ₂ U ₂ O ₇ + mixed oxide	Acid Leach	SX, Strip	NaOH, NH ₃	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes
USA-ESI	(NH ₄) ₂ U ₂ O ₇	Sulphate digestion (75-85 °C)	OPAP extraction, Strip	N/A	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	No
Canada-Macassa	UO ₂ (OH) ₂	Acid Leach	IX	MgO	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	No
USA-Anaconda	(NH ₄) ₂ U ₂ O ₇	Acid Leach	SX, Strip	MgO	Dried - 90-120 °C	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes/No
Australia-Radium Hill	(NH ₄) ₂ U ₂ O ₇	Acid Leach	IX, Strip	MgO	Dried - 320 °C	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes
USA-Lucky McGill	(NH ₄) ₂ U ₂ O ₇	Acid Leach	Eluex Process	NH ₃	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes/No	Yes/No
Argentina-Mine Unknown	Na ₂ U ₂ O ₇	Acid Leach	SX, Strip	NaOH	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	No	No
Australia-Rum Jungle	(NH ₄) ₂ U ₂ O ₇	Acid Leach	SX, Strip	MgO	Calcinated - 800 °C	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes	Yes
Canada-Ray Rock	UO ₂ (OH) ₂	N/A	N/A	N/A	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	No	No
Yugoslavia-Rudnik	(NH ₄) ₂ U ₂ O ₇	Acid Leach	SX	NH ₃	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	Yes/No	Yes
USA-Susquehan	UO ₂ (OH) ₂ + oxide	N/A	N/A	N/A	N/A	Class 3	Light Brown-Dark Yellow	LB-DY	Yes	No	No
USA-Kerr McGee	(NH ₄) ₂ U ₂ O ₇	Acid Leach	SX, Strip	NH ₃	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	No
Spain-Jen	(NH ₄) ₂ U ₂ O ₇	N/A	N/A	NH ₃	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	Yes
Gabon-EFI(Mouand)	(NH ₄) ₂ U ₂ O ₇	N/A	N/A	N/A	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	No
USA-Yankee Yellow	Na ₂ U ₂ O ₇	N/A	N/A	N/A	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes/No	No
Canada-North Span	UO ₂ (OH) ₂	N/A	N/A	N/A	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	No	Yes
Germany-Brunhilde	(NH ₄) ₂ U ₂ O ₇	N/A	N/A	N/A	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes/No	No
Germany-Helwiler	(NH ₄) ₂ U ₂ O ₇	N/A	N/A	N/A	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	Yes
S. Africa-Somair	Na ₂ U ₂ O ₇	Acid Leach	SX, Strip	NaOH	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes/No	No
Canada-Rio Algom	(NH ₄) ₂ U ₂ O ₇	Acid Leach	IX, Strip	MgO	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	No	No
Sweden-Ranstadt	Na ₂ U ₂ O ₇	Acid Leach	SX, Strip	NaOH	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	Yes
Canada-Denison	(NH ₄) ₂ U ₂ O ₇	Acid Leach	IX, Strip	NH ₃	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	No
USA-United Uranium	UO ₂ (OH) ₂	N/A	N/A	N/A	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	No	No
Canada-Milliken Lake	(NH ₄) ₂ U ₂ O ₇	Acid Leach	IX	NH ₃	Dried	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	Yes
Canada-El Dorado	(NH ₄) ₂ U ₂ O ₇	Alkaline Leach	N/A	NaOH	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes/No	No
Australia-S. Alligator	UO ₂ (OH) ₂	Acid Leach	SX, Strip	MgO	Dried - 84 °C	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	Yes
Romania-Mine Unknown	Na ₂ U ₂ O ₇ + oxide	N/A	N/A	N/A	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	Yes	Yes
USA-MESA EF(White Mesa)	(NH ₄) ₂ U ₂ O ₇	Acid Leach	SX, Strip	NH ₃	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	No	No
Canada-Stanleigh	(NH ₄) ₂ U ₂ O ₇	Acid Leach	IX, Strip	MgO	N/A	Class 4	Dark Yellow-Yellow	DY-Y	Yes	No	No
Canada-Stanrock	(NH ₄) ₂ U ₂ O ₇	Heap Leach	IX	MgO	N/A	Class 5	Yellow-Light Yellow	Y-LY	Yes	Yes	No
Holland-Delft	Mixture	N/A	N/A	N/A	N/A	Class 5	Yellow-Light Yellow	Y-LY	Yes	Yes	Yes
USA-Mulberry(IMC)	(NH ₄) ₂ U ₂ O ₇	N/A	N/A	N/A	N/A	Class 5	Yellow-Light Yellow	Y-LY	Yes	Yes/No	Yes
USA-El Mesquite	UO ₂ ·H ₂ O	N/A	N/A	Peroxide	N/A	Class 5	Yellow-Light Yellow	Y-LY	Yes	Yes	No
USA-Dawn	(NH ₄) ₂ U ₂ O ₇	Acid Leach	IX, Strip	NH ₃	N/A	Class 5	Yellow-Light Yellow	Y-LY	Yes	Yes	No
USA-Irizaray	UO ₂ ·H ₂ O	N/A	N/A	Peroxide	N/A	Class 6	Light Yellow-White	LY-W	Yes	Yes	Yes
USA-Mobil	UO ₂ ·H ₂ O	N/A	N/A	N/A	N/A	Class 6	Light Yellow-White	LY-W	Yes	Yes	Yes
Canada-Elidore(Rabbit Lake)	UO ₂ ·H ₂ O + U ₃ O ₈	Acid Leach	SX, Strip	NH ₃ ; MgO and H ₂ O ₂	Calcinated - 650 °C	Class 6	Light Yellow-White	LY-W	Yes	Yes	No
USA-Everest-Yellow	UO ₂ ·H ₂ O	N/A	N/A	N/A	N/A	Class 6	Light Yellow-White	LY-W	Yes	Yes	Yes

3.2 Image Texture based classification model development

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 uranium powder is seized. The samples belonging to each colour class represents the dataset used to build the correspondent texture-based classification models.

3.2.1 Preliminary results

To develop the texture based classification model, a preliminary study was carried out in 2014/2015 [17], where image texture analysis was used for the first time to extract textural features from 26 UOCs powder samples.

The first implementation [26] was focused on the application of the AMT and GLCM to extract the textural features and subsequently, to use these features alone and combined together to develop the texture-based classification model for two colour classes. It is important to note that the extracted textural features are correlated to the morphological particularities of each UOCs powders.

The models were created for CC5 (Y-LY) and CC6 (LY-W) by applying first the SVM algorithm to the GLCM textural features, then to the mean angle spectra and successively by combining all the variables extracted [26].

The obtained results have indicated that the strategy taken into consideration is valid; in fact, all the models developed, in function of the magnification and the feature's dataset used, performed very well. All of them have shown good performance in terms of the diagnostic index values calculated for each model. Furthermore, it was decided to extend the attention to the application of other textural features extraction algorithms and other classifiers; to do that with a consistent and systematic approach, machine learning was employed.

3.2.2 Development of the final model

As mentioned earlier, the discussion on the results are limited on the performance of the final classification models developed for each methodology. Here we focused on the texture-based classifications models. The detailed results obtained during the model development procedure can be found elsewhere [23].

After different test on the black samples datasets (see section 2.5), the Linear Discriminant Analysis was selected for our purpose as the most promising classifier, reaching the highest classification accuracy among all the other tested classifiers [23]. The following steps, as indicated in Figure 9, was to find a feature set for each dataset¹ that provides high accuracy with the smallest number of features. This was done by performing a features reduction.

At the end a model was developed for each of the available datasets; i.e., for each colour class, four texture-based classification models were developed, one for each magnification and one for all magnifications combined together. All the models consisted of the LDA classifier and a unique optimised features set. To build the dataset, labels were given that refer to the country of origin and the name of the production facility. As an example of the models performance, the average prediction matrices^{2,3} of the assigned probabilities for colour class 1 (CC1, B-DB: Black-Dark Brawn) and colour class 3 (CC3, O-DY: Orange-Dark Yellow) are shown in Figures 15 and 16. Each figure has four matrices, one for each magnification (100x, 250x, 500x/1000x) plus one for all the magnification combined together. The first column with the acronym denotes the true class of the samples. The header of the other columns indicates the predicted class of the samples.

¹ One dataset is a data table (matrix) with all the features extracted from the images acquired at one magnification, for all the samples belonging to a specific colour class.

² The prediction has been carried out three times for each dataset, with an independent and unique validation set for each run (see section 2.5); here only the matrix resulted averaging the three matrices of the three validation tests is reported.

³ The figures with the average prediction probabilities matrices for all colour classes, are reported in ANNEX 3.

	True \ Pred.	Au_MAK	Au_OLD	Au_QUE	Ca_KeL	Ch_Hey	Ru_Tec	Saf_Nuf	Saf_Pal	Saf_Ros	USA_Atl	USA_FAP	USA_Pet	Yo_YuSp	
100X	Au_MAK	58,35%	0,02%	0,01%	0,00%	0,00%	0,00%	19,40%	0,20%	16,50%	0,07%	0,43%	4,85%	0,16%	TRUE
	Au_OLD	0,17%	96,91%	0,01%	0,08%	0,00%	0,00%	0,00%	0,00%	1,75%	0,04%	1,02%	0,02%	0,00%	TRUE
	Au_QUE	0,06%	0,00%	41,51%	0,10%	0,00%	33,27%	0,03%	0,00%	2,83%	0,00%	22,14%	0,00%	0,05%	TRUE
	Ca_KeL	0,00%	0,02%	17,02%	81,81%	0,00%	1,03%	0,05%	0,00%	0,01%	0,00%	0,01%	0,00%	0,06%	TRUE
	Ch_Hey	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Ru_Tec	0,00%	0,31%	0,19%	0,43%	0,00%	99,03%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,03%	TRUE
	Saf_Nuf	20,84%	25,70%	0,52%	0,00%	32,90%	0,48%	3,77%	0,03%	14,39%	0,02%	0,07%	1,09%	0,20%	FALSE
	Saf_Pal	5,08%	0,00%	0,00%	0,00%	0,00%	0,00%	0,82%	26,91%	1,08%	0,01%	0,00%	66,10%	0,00%	FALSE
	Saf_Ros	14,19%	0,00%	0,03%	0,00%	0,00%	0,00%	6,95%	0,00%	45,71%	2,13%	22,67%	0,35%	7,97%	TRUE
	USA_Atl	1,21%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	7,19%	87,12%	0,38%	4,09%	0,00%	TRUE
	USA_FAP	0,96%	0,00%	0,00%	0,00%	0,00%	0,00%	0,08%	0,00%	24,25%	1,90%	72,80%	0,00%	0,00%	TRUE
	USA_Pet	1,94%	0,00%	0,00%	0,00%	0,00%	0,00%	0,09%	6,89%	12,60%	20,02%	0,38%	58,08%	0,00%	TRUE
	Yo_YuSp	3,54%	0,03%	0,18%	0,00%	0,00%	0,01%	4,36%	0,00%	1,55%	0,00%	0,03%	0,00%	90,30%	TRUE
250X	Au_MAK	48,29%	0,02%	0,16%	0,00%	0,00%	0,00%	14,92%	0,11%	26,30%	2,56%	0,13%	7,35%	0,17%	TRUE
	Au_OLD	7,75%	49,71%	5,56%	1,09%	0,00%	26,93%	0,05%	0,00%	8,61%	0,00%	0,20%	0,00%	0,11%	TRUE
	Au_QUE	0,61%	24,47%	33,00%	9,18%	0,00%	0,00%	0,10%	0,00%	29,89%	0,00%	1,03%	0,00%	1,72%	TRUE
	Ca_KeL	0,67%	4,47%	42,71%	35,69%	0,00%	0,00%	0,00%	0,00%	4,32%	0,01%	0,10%	0,00%	12,02%	FALSE
	Ch_Hey	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Ru_Tec	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Saf_Nuf	7,73%	0,00%	30,23%	1,29%	0,00%	0,00%	45,55%	8,94%	0,05%	0,00%	0,01%	4,38%	1,82%	TRUE
	Saf_Pal	0,07%	0,00%	0,00%	0,00%	0,00%	0,00%	33,26%	33,81%	0,00%	0,00%	0,00%	32,85%	0,00%	TRUE
	Saf_Ros	18,52%	0,85%	3,03%	0,02%	0,00%	0,00%	0,01%	0,00%	61,86%	0,25%	14,93%	0,01%	0,52%	TRUE
	USA_Atl	0,02%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	2,67%	97,00%	0,30%	0,00%	0,01%	TRUE
	USA_FAP	0,11%	0,00%	0,18%	0,27%	0,00%	0,00%	0,00%	0,00%	24,71%	0,77%	71,66%	0,00%	2,30%	TRUE
	USA_Pet	2,29%	0,00%	0,00%	0,00%	0,00%	0,00%	0,27%	6,11%	0,07%	0,02%	32,95%	58,28%	0,00%	TRUE
	Yo_YuSp	0,37%	0,52%	12,29%	5,22%	0,00%	0,00%	0,01%	0,00%	15,00%	0,14%	1,72%	0,00%	64,73%	TRUE
1000X	Au_MAK	31,76%	0,14%	21,91%	0,01%	0,00%	0,00%	0,00%	0,08%	14,86%	8,21%	0,58%	13,68%	8,78%	TRUE
	Au_OLD	0,16%	60,26%	0,09%	31,63%	0,43%	0,00%	2,16%	2,15%	0,98%	0,00%	1,54%	0,60%	0,00%	TRUE
	Au_QUE	7,14%	0,08%	45,90%	0,06%	0,00%	0,00%	0,00%	0,07%	20,79%	0,03%	10,13%	13,70%	2,10%	TRUE
	Ca_KeL	0,15%	30,09%	0,15%	56,64%	5,25%	0,33%	0,46%	0,05%	0,68%	0,00%	5,77%	0,44%	0,00%	TRUE
	Ch_Hey	0,00%	0,23%	0,00%	0,78%	87,35%	11,53%	0,00%	0,00%	0,00%	0,00%	0,10%	0,00%	0,00%	TRUE
	Ru_Tec	0,01%	0,00%	0,01%	0,01%	4,08%	95,68%	0,00%	0,00%	0,00%	0,13%	0,02%	0,00%	0,06%	TRUE
	Saf_Nuf	0,00%	0,01%	0,00%	0,57%	0,30%	0,00%	85,66%	13,41%	0,00%	0,00%	0,04%	0,00%	0,00%	TRUE
	Saf_Pal	0,07%	0,22%	0,52%	0,53%	0,00%	0,00%	18,13%	69,28%	1,13%	0,00%	1,96%	8,16%	0,00%	TRUE
	Saf_Ros	35,86%	2,41%	15,43%	1,51%	0,00%	0,00%	0,00%	0,05%	19,73%	0,09%	9,25%	15,56%	0,10%	FALSE
	USA_Atl	8,22%	0,01%	0,25%	0,00%	0,05%	23,20%	0,00%	0,00%	0,10%	67,33%	0,02%	0,02%	0,80%	TRUE
	USA_FAP	0,88%	12,95%	10,95%	11,37%	0,17%	0,24%	0,04%	0,74%	10,87%	0,00%	43,14%	8,59%	0,05%	TRUE
	USA_Pet	16,92%	0,63%	13,73%	12,05%	0,00%	0,00%	19,05%	1,85%	17,43%	0,01%	1,59%	16,69%	0,06%	FALSE
	Yo_YuSp	3,26%	0,01%	5,55%	0,01%	0,00%	0,01%	0,00%	0,03%	1,22%	0,64%	0,46%	3,98%	84,84%	TRUE
ALL	Au_MAK	70,81%	0,02%	0,03%	0,00%	0,00%	0,00%	0,00%	0,01%	13,17%	0,25%	0,04%	8,36%	7,32%	TRUE
	Au_OLD	0,01%	33,52%	9,24%	12,28%	0,00%	43,89%	0,01%	0,00%	0,19%	0,00%	0,87%	0,00%	0,00%	FALSE
	Au_QUE	0,08%	13,67%	60,20%	1,60%	0,00%	0,00%	0,00%	0,00%	10,37%	0,00%	7,74%	0,00%	6,34%	TRUE
	Ca_KeL	0,00%	2,31%	15,32%	80,92%	0,00%	0,25%	0,00%	0,00%	0,40%	0,00%	0,46%	0,00%	0,36%	TRUE
	Ch_Hey	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Ru_Tec	0,00%	0,00%	0,00%	0,19%	0,00%	99,81%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,01%	TRUE
	Saf_Nuf	0,01%	0,01%	0,00%	0,00%	0,01%	0,00%	96,43%	0,19%	0,04%	0,00%	0,05%	0,07%	3,19%	TRUE
	Saf_Pal	0,02%	0,00%	0,00%	0,00%	0,00%	0,00%	33,33%	53,01%	0,00%	0,00%	0,00%	13,62%	0,01%	TRUE
	Saf_Ros	3,06%	2,17%	14,11%	0,02%	0,00%	0,00%	0,02%	0,00%	32,42%	1,77%	14,53%	0,01%	31,89%	TRUE
	USA_Atl	1,20%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	4,33%	91,68%	2,70%	0,01%	0,08%	TRUE
	USA_FAP	0,58%	1,70%	2,38%	0,04%	0,00%	0,00%	0,14%	0,00%	18,03%	0,57%	73,41%	0,00%	3,15%	TRUE
	USA_Pet	2,41%	0,00%	0,00%	0,00%	0,00%	0,00%	0,34%	3,55%	0,68%	0,11%	25,83%	67,01%	0,07%	TRUE
	Yo_YuSp	1,86%	0,97%	4,51%	0,03%	0,00%	0,00%	0,00%	0,00%	9,80%	0,01%	0,38%	0,00%	82,45%	TRUE

Figure 15. Average prediction matrices obtained from model's validation test for CC1 at 100x, 250x, 1000x and for all magnifications.

100X	True Pred.	USA_Ana	Be_Cong	Ca_Dyno	USA_Cot	USA_Fall	Bra_Nuc	Au_RadH	Au_RumJ	Yo_SpisYe	USA_ESI	Ge_Wis	Au_Yee	Ca_Mac	
	USA_Ana	81.94%	0.00%	0.00%	16.70%	0.02%	0.13%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	1.19%	TRUE
	Be_Cong	0.00%	79.75%	0.00%	0.00%	0.00%	0.00%	0.07%	20.18%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	Ca_Dyno	0.00%	0.00%	92.64%	0.00%	0.03%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	7.32%	0.00%	TRUE
	USA_Cot	11.40%	0.00%	0.00%	88.36%	0.00%	0.24%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	USA_Fall	0.00%	0.00%	0.01%	0.00%	63.08%	0.00%	0.00%	0.00%	32.85%	0.73%	0.08%	1.27%	1.98%	TRUE
	Bra_Nuc	0.39%	0.00%	0.00%	1.00%	0.00%	98.60%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	Au_RadH	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.82%	0.18%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	Au_RumJ	0.00%	6.47%	0.00%	0.00%	0.00%	0.00%	2.33%	91.20%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	Yo_SpisYe	0.00%	0.00%	3.63%	0.00%	23.01%	0.00%	0.00%	0.00%	54.53%	0.01%	0.05%	16.34%	2.43%	TRUE
	USA_ESI	0.00%	0.00%	0.00%	0.00%	0.44%	0.00%	0.00%	0.00%	0.42%	68.29%	0.06%	0.02%	30.77%	TRUE
	Ge_Wis	0.00%	0.00%	0.00%	0.00%	0.09%	0.00%	0.00%	0.00%	0.20%	0.00%	99.70%	0.01%	0.00%	TRUE
	Au_Yee	0.00%	0.00%	1.26%	0.00%	2.21%	0.00%	0.00%	0.00%	7.09%	0.05%	0.18%	89.14%	0.08%	TRUE
	Ca_Mac	1.90%	0.00%	0.00%	18.51%	0.18%	0.07%	0.00%	0.00%	3.87%	0.03%	0.11%	0.00%	75.32%	TRUE
250X	True Pred.	USA_Ana	Be_Cong	Ca_Dyno	USA_Cot	USA_Fall	Bra_Nuc	Au_RadH	Au_RumJ	Yo_SpisYe	USA_ESI	Ge_Wis	Au_Yee	Ca_Mac	
	USA_Ana	76.65%	0.01%	0.00%	0.69%	10.49%	0.00%	0.00%	0.00%	0.44%	11.46%	0.26%	0.00%	0.00%	TRUE
	Be_Cong	0.00%	76.86%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	23.13%	0.00%	0.00%	0.00%	TRUE
	Ca_Dyno	32.28%	0.00%	66.66%	0.00%	0.00%	0.00%	0.00%	0.00%	0.18%	0.00%	0.88%	0.01%	0.00%	TRUE
	USA_Cot	0.02%	0.00%	0.00%	67.48%	0.01%	0.00%	1.41%	0.00%	6.50%	0.00%	1.07%	0.20%	23.31%	TRUE
	USA_Fall	0.01%	0.00%	0.00%	0.06%	94.70%	0.00%	0.00%	0.00%	0.08%	0.00%	5.14%	0.00%	0.00%	TRUE
	Bra_Nuc	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	Au_RadH	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	Au_RumJ	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	Yo_SpisYe	3.41%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	80.18%	0.00%	1.42%	12.51%	2.46%	TRUE
	USA_ESI	0.00%	5.41%	0.00%	0.09%	30.72%	0.00%	0.02%	0.00%	2.06%	34.30%	27.39%	0.00%	0.00%	TRUE
	Ge_Wis	0.32%	0.00%	0.00%	0.01%	2.10%	0.00%	0.00%	0.00%	0.16%	0.00%	97.41%	0.00%	0.00%	TRUE
	Au_Yee	0.41%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.84%	0.00%	0.00%	83.71%	8.04%	TRUE
	Ca_Mac	0.00%	0.00%	0.00%	0.37%	0.00%	0.00%	0.00%	0.00%	21.06%	0.00%	0.24%	20.67%	57.65%	TRUE
500X	True Pred.	USA_Ana	Be_Cong	Ca_Dyno	USA_Cot	USA_Fall	Bra_Nuc	Au_RadH	Au_RumJ	Yo_SpisYe	USA_ESI	Ge_Wis	Au_Yee	Ca_Mac	
	USA_Ana	97.80%	0.01%	0.00%	0.17%	0.00%	0.00%	0.73%	0.00%	0.00%	0.00%	0.21%	1.07%	0.00%	TRUE
	Be_Cong	25.14%	41.78%	0.00%	0.00%	0.00%	0.00%	0.29%	0.00%	0.00%	32.74%	0.01%	0.04%	0.00%	TRUE
	Ca_Dyno	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	USA_Cot	0.00%	0.00%	0.00%	66.84%	0.00%	0.00%	1.24%	0.01%	0.00%	0.04%	31.86%	0.01%	0.00%	TRUE
	USA_Fall	0.00%	0.00%	0.00%	0.00%	96.36%	0.00%	0.00%	0.00%	0.00%	0.12%	0.18%	0.02%	3.32%	TRUE
	Bra_Nuc	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	Au_RadH	0.00%	0.28%	0.00%	0.08%	0.00%	0.00%	99.52%	0.00%	0.00%	0.00%	0.12%	0.00%	0.00%	TRUE
	Au_RumJ	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.99%	0.00%	0.01%	0.00%	0.00%	0.00%	TRUE
	Yo_SpisYe	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	TRUE
	USA_ESI	0.00%	0.00%	0.00%	0.00%	0.60%	0.00%	0.06%	7.41%	0.00%	41.08%	49.57%	1.03%	0.24%	FALSE
	Ge_Wis	0.00%	0.00%	0.00%	0.00%	0.43%	0.00%	0.00%	0.00%	0.00%	0.00%	98.59%	0.98%	0.00%	TRUE
	Au_Yee	7.62%	0.00%	0.00%	0.00%	0.61%	0.00%	0.00%	0.00%	0.00%	0.01%	0.66%	69.89%	21.20%	TRUE
	Ca_Mac	0.00%	0.00%	0.00%	0.00%	22.23%	0.00%	0.00%	0.00%	0.00%	1.21%	0.50%	22.01%	54.04%	TRUE
ALL	True Pred.	USA_Ana	Be_Cong	Ca_Dyno	USA_Cot	USA_Fall	Bra_Nuc	Au_RadH	Au_RumJ	Yo_SpisYe	USA_ESI	Ge_Wis	Au_Yee	Ca_Mac	
	USA_Ana	73.42%	16.39%	0.00%	10.06%	0.03%	0.07%	0.00%	0.01%	0.00%	0.00%	0.01%	0.01%	0.00%	TRUE
	Be_Cong	1.06%	73.74%	0.00%	0.85%	0.04%	6.02%	0.00%	0.23%	17.91%	0.00%	0.00%	0.05%	0.10%	TRUE
	Ca_Dyno	0.03%	0.00%	37.79%	0.01%	31.60%	0.07%	0.00%	0.00%	0.00%	0.40%	0.00%	0.00%	30.09%	TRUE
	USA_Cot	0.59%	15.36%	0.00%	53.21%	0.03%	0.00%	30.00%	0.03%	0.07%	0.42%	0.00%	0.29%	0.00%	TRUE
	USA_Fall	0.70%	0.00%	0.00%	1.45%	81.99%	5.63%	0.04%	0.28%	0.00%	4.98%	0.19%	1.94%	2.80%	TRUE
	Bra_Nuc	0.00%	0.00%	0.00%	1.83%	0.04%	66.97%	6.23%	4.49%	0.00%	2.74%	0.00%	17.66%	0.03%	TRUE
	Au_RadH	0.01%	0.00%	0.00%	1.45%	0.01%	0.02%	95.47%	0.64%	0.00%	0.38%	0.00%	2.02%	0.00%	TRUE
	Au_RumJ	0.05%	0.11%	0.00%	2.44%	0.05%	2.48%	0.12%	73.64%	11.77%	0.16%	0.00%	9.13%	0.04%	TRUE
	Yo_SpisYe	0.00%	6.81%	0.00%	0.01%	0.00%	0.33%	0.00%	5.87%	86.93%	0.00%	0.00%	0.04%	0.01%	TRUE
	USA_ESI	0.01%	0.00%	1.95%	1.05%	3.61%	4.16%	0.46%	0.96%	0.00%	63.29%	1.52%	8.02%	14.96%	TRUE
	Ge_Wis	0.04%	0.00%	0.00%	0.00%	0.23%	0.01%	0.00%	0.00%	0.00%	0.14%	99.53%	0.04%	0.00%	TRUE
	Au_Yee	0.04%	0.00%	0.01%	4.24%	1.71%	2.85%	1.00%	5.93%	0.01%	2.93%	0.00%	81.10%	0.20%	TRUE
	Ca_Mac	0.03%	0.00%	5.89%	0.39%	11.33%	4.10%	0.05%	0.17%	0.00%	8.56%	0.00%	1.45%	68.03%	TRUE

Figure 16. Average prediction matrices obtained from the model's validation test, for CC3 at 100x, 250x, 1000x and for all magnifications.

Basically, the tables show the average validation test results for the two colour classes. The numbers in the table represent the probabilities of each "unknown" test sample, that in our case correspond to the true class, to be predicted as a specific sample that correspond to that predicted class.

It is important to underline that the green colour of the cells along the diagonal, is proportional to the percentage of the correct prediction probability, while the cell with the different intensity of red colour indicates the probability the “unknown” test sample is predicted as a wrong sample. The last column reports “TRUE” if the highest value of the prediction probability is assigned to the corresponding correct predicted class of the test sample. In other words, the “unknown” test sample is predicted with highest probability as belonging to the correct class. When the last column reports “FALSE”, it means that the “unknown” test sample is predicted with the highest probability to belong to another class.

To explain better how this table can be interpreted, we can focus only on the average prediction matrix of colour class 1 (CC1), of the Black-Dark Brown samples.

Looking in the first matrix for CC1 at 100x of magnification (Figure 15); the model in this case has an average accuracy attribution equal to 85%, i.e. 11 samples over 13 were correctly predicted. Looking the ChHEY sample, as an example, the numbers in the ChHEY’s row represent the probability with which the “unknown” ChHEY test sample is attributed to the correspondent predicted classes. In this case our test sample ChHEY, has been classified and then assigned by the model, to the correct predicted class ChHEY with an average probability of 100%. It is referring to an average probability because one should keep in mind that these are the average prediction matrices; therefore, the probability values in each cell used to describe the model performance, are the average of the three different validation test carried out using different images as test image for each run (as explained in section 2.5). This result indicates that with this model a sample that has similar texture to the ChHEY sample, can be identify with an average probability of 100%, using only the texture signatures. The same assessment can be done with AuMAK sample. The numbers in the AuMAK’s row represent the probability with which the “unknown” AuMAK test sample is assigned to the correspondent predicted classes. In this case, AuMAK test sample has been classified and then attributed from the texture-based classification model to the correct predicted class AuMAK with 58.35% probability, with a probability of 19.40% it was predicted as SaNUF, with the 16.5% as SaROS and with 4.85% as UsPET. In this case, even if the highest value of the prediction probability corresponds the correct predicted class, with only 58.35% probability it is difficult to state that one “unknown” testing sample can be the AuMAK sample. To avoid this problem, the same reasoning used to develop the colour-based classification model can be applied also here, and the number of samples to which the “unknown” AuMAK test sample can be identified, is reduced from 13 to 4. In this example, can be stated that our “unknown” test sample has texture similarities with 4 classes and with a total average probability of 99%, so the “unknown” sample can be only one of the 4 predicted samples: in particular, AuMAK with 58.35% probability, SaNUF, SaROS, UsPET, with 19.4%, 16.5% and 4.85% probability respectively. If the sample has really the same texture properties as sample AuMAK, this can be confirmed also by other analysis like the hyperspectral image analysis.

Considering now the average prediction matrixes at all magnifications for CC1 and CC3, it can be noticed that CC1 has an overall accuracy of 92.3%. This includes also the particular case of the SaROS sample that has been correctly predicted but only with highest prediction probability value of 32.42%. All the other samples shown a prediction probability above 50% and, using this value as a threshold level, the accuracy decreased to 84.62% since now only 11 samples over 13 were correctly predicted. It must be taken into consideration that anyhow the model output is the ranking of the predicted classes in which the unknown sample can be identified, in function of the highest prediction probability values. At the end, the final decision will be taken considering also information collected from other analysis. For CC3 (Figure 16), the overall accuracy is 100%, if the case of the CaDyno sample is taken into account, otherwise the accuracy became 92.31%. The same approach can be used to assess the model performances for each colour class.

Figure 17 shows the three different validation tests (prediction matrixes) carried out using a unique image test set for 100x dataset of colour class 1, from which the correspondent average prediction matrix (showed in Figure 15) have been calculated.

First test files	True Pred.	Au_MAK	Au_OLD	Au_QUE	Ca_KeL	Ch_Hey	Ru_Tec	Saf_Nuf	Saf_Pal	Saf_Ros	USA_Atl	USA_FAP	USA_Pet	Yo_YuSp	
C1AuMAKU3O8100x	Au_MAK	62.68%	0.03%	0.01%	0.00%	0.00%	0.00%	35.05%	0.21%	0.33%	0.00%	0.00%	1.51%	0.18%	TRUE
A3AuOLDU3O8100x	Au_OLD	0.34%	99.40%	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%	0.13%	0.01%	0.00%	0.05%	0.00%	TRUE
C1AuQUEU3O8100x	Au_QUE	0.00%	0.00%	99.75%	0.00%	0.00%	0.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.15%	TRUE
A2CaKELU3O8100x	Ca_KeL	0.01%	0.04%	51.05%	45.46%	0.00%	3.08%	0.16%	0.00%	0.02%	0.00%	0.03%	0.00%	0.17%	FALSE
A5ChHEYUO2_100x	Ch_Hey	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
A2RuTECMix_100x	Ru_Tec	0.00%	0.00%	0.02%	0.00%	0.00%	99.88%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	TRUE
C1SaNUFU3O8100x	Saf_Nuf	45.46%	0.00%	0.00%	0.00%	0.00%	0.00%	7.40%	0.08%	43.11%	0.07%	0.20%	3.24%	0.44%	FALSE
C4SaPALU3O8100x	Saf_Pal	2.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	38.33%	0.04%	0.00%	0.00%	59.38%	0.00%	FALSE
C3SaROSU3O8100x	Saf_Ros	0.19%	0.00%	0.00%	0.01%	0.00%	0.00%	0.01%	0.00%	40.94%	1.11%	57.69%	0.00%	0.03%	FALSE
A3UsATLU3O8100x	USA_Atl	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.95%	0.04%	0.00%	0.00%	TRUE
C2UsFAPU3O8100x	USA_FAP	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.68%	0.01%	99.30%	0.00%	0.00%	TRUE
A1UsPETMix_100x	USA_Pet	0.77%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	37.78%	60.06%	1.13%	0.27%	0.00%	FALSE
A4YuSPBUH_100x	Yo_YuSp	0.40%	0.00%	0.06%	0.00%	0.00%	0.01%	2.16%	0.00%	0.64%	0.00%	0.03%	0.00%	96.70%	TRUE
Second test files	True Pred.	Au_MAK	Au_OLD	Au_QUE	Ca_KeL	Ch_Hey	Ru_Tec	Saf_Nuf	Saf_Pal	Saf_Ros	USA_Atl	USA_FAP	USA_Pet	Yo_YuSp	
A4AuMAKU3O8100x	Au_MAK	60.55%	0.00%	0.00%	0.00%	0.00%	0.00%	20.60%	0.40%	7.80%	0.00%	0.01%	10.63%	0.01%	TRUE
B5AuOLDU3O8100x	Au_OLD	0.03%	99.97%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
A3AuQUEU3O8100x	Au_QUE	0.18%	0.00%	24.55%	0.26%	0.00%	0.00%	0.09%	0.00%	8.49%	0.00%	66.43%	0.00%	0.01%	FALSE
B4CaKELU3O8100x	Ca_KeL	0.00%	0.01%	0.01%	99.98%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
B3ChHEYUO2_100x	Ch_Hey	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
C5RuTECMix_100x	Ru_Tec	0.00%	0.00%	0.21%	0.00%	0.00%	99.79%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
B2SaNUFU3O8100x	Saf_Nuf	0.00%	0.00%	0.06%	0.00%	98.70%	1.24%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	FALSE
A5SaPALU3O8100x	Saf_Pal	10.81%	0.00%	0.00%	0.00%	0.00%	0.00%	2.32%	3.32%	3.21%	0.02%	0.00%	80.33%	0.00%	FALSE
B3SaROSU3O8100x	Saf_Ros	2.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	0.00%	82.22%	5.28%	10.07%	0.08%	0.00%	TRUE
B4UsATLU3O8100x	USA_Atl	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.92%	88.16%	0.69%	0.14%	0.00%	TRUE
B5UsFAPU3O8100x	USA_FAP	2.88%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	0.00%	70.36%	0.19%	26.32%	0.01%	0.00%	FALSE
C3UsPETMix_100x	USA_Pet	1.79%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	2.17%	0.03%	0.00%	0.00%	95.99%	0.00%	TRUE
B4YuSPBUH_100x	Yo_YuSp	10.21%	0.10%	0.05%	0.00%	0.00%	0.00%	10.69%	0.00%	4.00%	0.00%	0.07%	0.01%	74.86%	TRUE
Third test files	True Pred.	Au_MAK	Au_OLD	Au_QUE	Ca_KeL	Ch_Hey	Ru_Tec	Saf_Nuf	Saf_Pal	Saf_Ros	USA_Atl	USA_FAP	USA_Pet	Yo_YuSp	
B5AuMAKU3O8100x	Au_MAK	51.82%	0.03%	0.01%	0.00%	0.00%	0.00%	2.57%	0.00%	41.37%	0.19%	1.28%	2.42%	0.30%	TRUE
C4AuOLDU3O8100x	Au_OLD	0.16%	91.37%	0.02%	0.16%	0.00%	0.00%	0.00%	0.00%	5.11%	0.11%	3.06%	0.00%	0.01%	TRUE
B1AuQUEU3O8100x	Au_QUE	0.00%	0.00%	0.23%	0.06%	0.00%	99.71%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	FALSE
C5CaKELU3O8100x	Ca_KeL	0.00%	0.01%	0.00%	99.99%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
C4ChHEYUO2_100x	Ch_Hey	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
B5RuTECMix_100x	Ru_Tec	0.00%	0.93%	0.35%	1.30%	0.00%	97.42%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	TRUE
A2SaNUFU3O8100x	Saf_Nuf	17.05%	77.10%	1.49%	0.00%	0.00%	0.20%	3.90%	0.00%	0.05%	0.00%	0.00%	0.04%	0.17%	FALSE
B3SaPALU3O8100x	Saf_Pal	2.17%	0.00%	0.00%	0.00%	0.00%	0.00%	0.15%	39.09%	0.01%	0.00%	0.00%	58.59%	0.00%	FALSE
A4SaROSU3O8100x	Saf_Ros	40.11%	0.00%	0.07%	0.00%	0.00%	0.00%	20.75%	0.00%	13.96%	0.01%	0.24%	0.97%	23.89%	FALSE
C2UsATLU3O8100x	USA_Atl	3.54%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	10.66%	73.26%	0.41%	12.12%	0.00%	TRUE
A3UsFAPU3O8100x	USA_FAP	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.71%	5.50%	92.78%	0.00%	0.00%	TRUE
B2UsPETMix_100x	USA_Pet	3.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.26%	18.48%	0.01%	0.00%	0.00%	78.00%	0.00%	TRUE
C4YuSPBUH_100x	Yo_YuSp	0.01%	0.00%	0.43%	0.00%	0.00%	0.01%	0.22%	0.00%	0.00%	0.00%	0.00%	0.00%	99.33%	TRUE

Figure 17. Prediction matrices for the three validation test for CC1 at 100x. UsPET sample has been correctly classified twice, while it was misclassified the first time.

Figure 18 displays the original images for the test samples used for the class UsPET (top row), and one randomly chosen image of both the classes UsATL and SaROS (bottom row). This can help to explain the reason why some images are misclassified by the models. Both C3UsPETMix_100x.tif and B2UsPETMix_100x.tif were correctly classified with respectively assigned probabilities of 96 % and 78 %, whereas A1UsPETMix_100x.tif were only assigned to its true class with 0.3% probability (Figure 17). Instead, this misclassified sample were assigned with 60% and 38% probability, belonging to respectively classes UsATL and SaROS. The two correctly classified UsPET images were both assigned to UsATL and SaROS with 0% probability. These images had large structures, as can be seen in Figure 18, in contrast to the fine and coarse texture in the UsATL and SaROS images respectively. A blend of both fine and coarse texture appears to be present in the misclassified UsPET image, indicating the reason why it was assigned to the wrong classes.

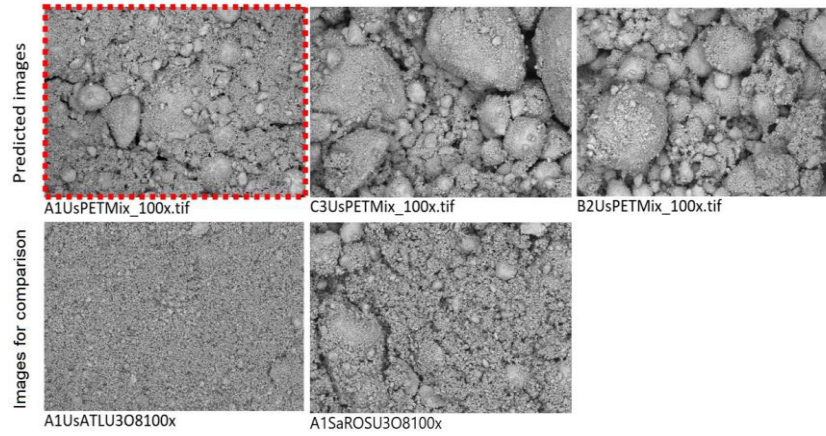


Figure 18. Example of intra- and inter-class differences of images. The top row shows the three images of class UsPET that were classified in the three test runs. The two bottom images belonging to the classes UsATL and SaROS.

In Table 6 the summary of the models' performances of all the datasets is reported. The overall accuracy obtained from each validation test can be compared and the average of the prediction accuracy (the mean among the three validation test) and the average of the mean attribution accuracy (how many samples are correctly classified averaging the three probability prediction matrices) are also given together with the average standard deviation values. The average of the number of textural features used by each classification model is also reported.

Table 6. Summary of models' performances of all the textural features datasets extracted from the SEM images⁴.

Magnification	Description	CC1	CC2	CC3	CC4	CC5	CC6	Mean over all colour categories
100x	First test acc.:	62%		92%	78%	100%	100%	Mean prediction acc.: 88% Mean attribution acc.: 97% Avg. # of feats. in each CC 5,5
	Second test acc.:	69%		77%	100%	100%	100%	
	Third test acc.:	69%		100%	100%	75%	100%	
	Mean prediction acc.:	67%		90%	93%	92%	100%	
	Mean attribution acc.:	85%		100%	100%	100%	100%	
	Mean attribution SD:	29%		14%	11%	14%	0%	
	Avg. # of feats.:	6		4,7	5	6	6	
250x	First test acc.:	85%		69%	100%	100%	100%	Mean prediction acc.: 86% Mean attribution acc.: 98% Avg. # of feats. in each CC 5,4
	Second test acc.:	54%		77%	89%	100%	100%	
	Third test acc.:	62%		92%	89%	75%	100%	
	Mean prediction acc.:	67%		79%	93%	92%	100%	
	Mean attribution acc.:	92%		100%	100%	100%	100%	
	Mean attribution SD:	23%		19%	12%	13%	3%	
	Avg. # of feats.:	7		6	5,7	4,7	3,7	
500x	First test acc.:			85%	78%	50%	75%	Mean prediction acc.: 76% Mean attribution acc.: 86% Avg. # of feats. in each CC 5,8
	Second test acc.:			62%	89%	50%	50%	
	Third test acc.:			92%	78%	100%	100%	
	Mean prediction acc.:			79%	81%	67%	75%	
	Mean attribution acc.:			92%	100%	50%	100%	
	Mean attribution SD:			23%	18%	26%	19%	
	Avg. # of feats.:			6,7	5,3	6,3	5	
1000x	First test acc.:	46%						Mean prediction acc.: 69% Mean attribution acc.: 85% Avg. # of feats. in each CC 4,0
	Second test acc.:	77%						
	Third test acc.:	85%						
	Mean prediction acc.:	69%						
	Mean attribution acc.:	85%						
	Mean attribution SD:	25%						
	Avg. # of feats.:	4						
all magn.	First test acc.:	69%		62%	100%	100%	100%	Mean prediction acc.: 90% Mean attribution acc.: 98% Avg. # of feats. in each CC 4,3
	Second test acc.:	77%		92%	89%	75%	100%	
	Third test acc.:	92%		100%	100%	100%	100%	
	Mean prediction acc.:	79%		85%	96%	92%	100%	
	Mean attribution acc.:	92%		100%	100%	100%	100%	
	Mean attribution SD:	22%		16%	6%	11%	8%	
	Avg. # of feats.:	5		4,3	5,7	3,3	3	

Comparing the performances, CC5 shows the worst mean prediction accuracy values for the 1st and 2nd validation tests at 500x and, as consequence, also the worst mean attribution values with only 50% of accuracy. Apart this case, all the developed models have shown a very good prediction accuracy, going from the lowest value of 67% for CC5 at 500x and CC1 at 100x, up to 100% for almost all the datasets. The best performance was reached by the models for all colour classes developed combining the dataset of all magnification together.

⁴ The empty cells mean that the classification models for that colour class, at that magnification, was not developed because the corresponding datasets were not created due to the lack of sample images.

Combining all the features extracted from images acquired at all magnifications, the model performance improves comparing with those developed at a single magnification. This is due to the fact that the most important features extracted from each magnification are correlated to some specific morphological characteristics that can be captured at that magnification. In practice, if one sample is difficult to be classified at one magnification [17], maybe this doesn't happen using another one. The textural features are strongly dependent from the magnification, therefore combining all the features together, it is sure that all the important features, captured at different scales, are used by the model to discriminate one sample from another.

3.3 Spectral based classification model

The LDA (Linear Discriminant Analysis) algorithm was selected as best classifier to develop the spectral-based classification model. The hyperspectral images were acquired at one magnification; thus the number of the developed models correspond to the number of the colour classes. The features selection was performed three times, similar as described in section 2.5. The detailed description about the process can be found elsewhere [24]. Four spectral-based classification models were developed due to the fact that only for CC1, CC3, CC4 and CC6 were acquired enough hyperspectral images to be analysed. The relative spectral features were then extracted and used afterwards to build the correspondent spectral-based classification models. In Figure 19 the infrared spectra for all the replicates (12) of each sample, the average spectra for each sample holder (3) and the average spectra for each sample (1), are shown as an example.

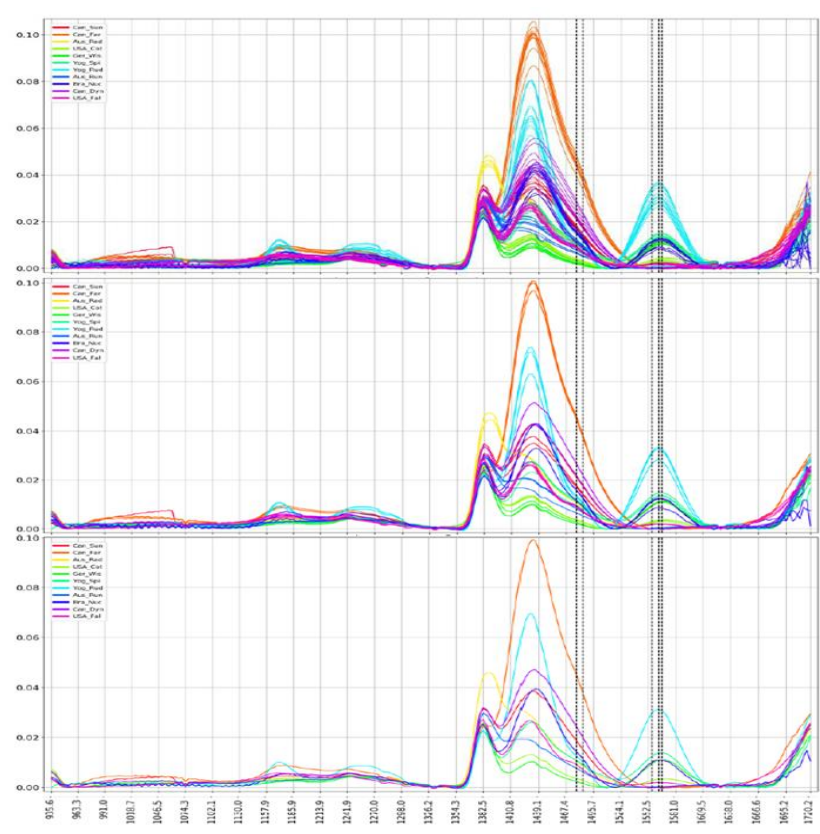


Figure 19. The spectra for all samples replicates (top); average spectra for each sample holder (middle); average spectra for each UOC sample. All spectra are colour coded by UOC sample name (see Table 2). The black vertical dotted lines indicate all selected wavelengths given the union feature set. The wavelengths along the horizontal axis are given in [nm] and the vertical axis is absorbance in arbitrary units.

The performance of the models are shown in Figure 20, where the average prediction matrices⁵ for each colour class are reported.

⁵ See notes 2, at page 23.

COLOUR CLASS 1 (CC1)						
True \ Pred	Aus_Oly	Can_Key	Chi_Hen	Rus_Tec	Zam_Min	
Aus_Oly	100,00%	0,00%	0,00%	0,00%	0,00%	TRUE
Can_Key	0,01%	99,99%	0,00%	0,00%	0,00%	TRUE
Chi_Hen	0,00%	0,00%	100,00%	0,00%	0,00%	TRUE
Rus_Tec	0,00%	0,00%	0,00%	100,00%	0,00%	TRUE
Saf_Min	0,00%	0,00%	0,00%	0,00%	100,00%	TRUE

COLOUR CLASS 3 (CC3)												
True \ Pred	Aus_Rad	Aus_Run	Bra_Nuc	Can_Dyn	Can_Far	Can_Sun	Ger_Wis	USA_Cot	USA_Fal	Yog_Rud	Yog_Spi	
Aus_Rad	80,22%	13,68%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	6,10%	0,00%	0,00%	TRUE
Aus_Run	2,60%	64,77%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	31,78%	0,00%	0,00%	TRUE
Bra_Nuc	0,00%	0,00%	99,90%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,10%	TRUE
Can_Dyn	0,00%	0,00%	0,00%	93,97%	0,00%	6,03%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
Can_Far	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
Can_Sun	0,00%	0,00%	0,00%	4,46%	0,00%	95,54%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
Ger_Wis	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	TRUE
USA_Cot	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	TRUE
USA_Fal	6,10%	24,04%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	69,86%	0,00%	0,00%	TRUE
Yog_Rud	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	TRUE
Yog_Spi	0,00%	0,00%	0,44%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	99,56%	TRUE

COLOUR CLASS 4 (CC4)								
True \ Pred	Aus_S A	Can_Mil	Can_Nor	Ger_Hel	Rum_Rum	Spa_Gen	Swe_Ran	
Aus_S A	99,47%	0,19%	0,00%	0,00%	0,35%	0,00%	0,00%	TRUE
Can_Mil	0,61%	99,15%	0,00%	0,21%	0,04%	0,00%	0,00%	TRUE
Can_Nor	0,00%	0,00%	92,20%	0,00%	0,00%	0,00%	7,80%	TRUE
Ger_Hel	0,00%	2,76%	0,00%	97,24%	0,00%	0,00%	0,00%	TRUE
Rum_Rum	0,00%	0,00%	0,00%	0,00%	99,99%	0,00%	0,00%	TRUE
Spa_Gen	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	TRUE
Swe_Ran	0,00%	0,00%	1,13%	0,00%	0,00%	0,00%	98,87%	TRUE

COLOUR CLASS 6 (CC6)				
True \ Pred	USA_Eve	USA_Iri	USA_Mob	
USA_Eve	79,76%	14,51%	5,72%	TRUE
USA_Iri	0,08%	96,14%	3,78%	TRUE
USA_Mob	24,92%	10,26%	64,82%	TRUE

Figure 20. Average prediction matrices for CC1, CC3, CC4 and CC6 obtained from the spectral-based models validation test.

The interpretation of the results can be made in the same way as have been done for the results obtained with the texture-based classification models.

As can be seen from the Table 7, the model performances are very good; all of them have an average prediction probabilities equal to 100%. There are only four samples, three in colour class 3 and one in colour class 6, that have been assigned to the correct predicted class with a prediction probability less than 90%, but always above 64%. The summary of the model performances of all the datasets are reported in Table 7. Similar to the Table 6, the overall accuracy obtained for each validation test can be compared and the average of the prediction accuracy (the mean among the three validation test) and the average of the mean attribution accuracy (how many samples are correctly classified averaging the three probability prediction matrices) are also given together with the average standard deviation. The average of the number of spectral features used by each classification model is also reported.

Table 7. Summary of model performances of all the textural features datasets extracted from the SEM images.

Description	CC1	CC3	CC4	CC6	Mean over all colour categories	
First test acc.:	100%	100%	100%	67%		
Second test acc.:	100%	100%	100%	100%		
Third test acc.:	100%	100%	100%	100%		
Mean prediction acc.:	100%	100%	100%	89%	Mean prediction acc.:	97%
Mean attribution acc.:	100%	100%	100%	100%	Mean attribution acc.:	100%
Mean attribution SD:	0%	13%	3%	13%	Avg. # of feats. in each CC	2,9
Avg. # of feats.:	2	2	2,7	4,7		

In Table 7 one can see how the models reached 100% of prediction probability in all the validation test for all the colour classes, apart for colour class 6 that reached 67% of accuracy in the first validation test and 89% of average prediction accuracy. One of the most important thing is the average number of features that have been used for each single model to reach that specific accuracy. In particular, the spectral-based classification model for CC1 employed only two features (two wavelengths) as did the model for CC3; the classification model for CC4 employed 2.7 average features while CC6 employed 4.7. The spectral-based classification model for CC6 was the only one that employed more than four average number of features to be able to classify sample in its class. This can be probably explained by the fact that all the samples that have been attributed to CC6 by the colour-based classification model (see Table 5 in section 3.1), and used afterwards as training samples to develop the spectral-based classification model for CC6, are peroxides. This means that they have the same major chemical components, so more features are needed to differentiate one sample to the other.

However, the results clearly indicate how the near infrared spectra, acquired with the hyperspectral camera, can be considered as a fingerprint for the UOC powders.

4 Conclusions

The aim of this project was to develop a new approach for a quick identification and classification of unknown UOC samples. The main objectives were reached and the colour-based, texture-based and spectral-based classification models were successfully developed.

This new approach could be taken into consideration, after the completion of the missing experimental activities and an appropriate validation process, as a new analytical methodology that can be used by the nuclear forensics community to identify and classify interdicted UOC powders. It is important to underline that this approach was not thought to give one single answer, like to have an attribution 1 to 1, but to reduce the number of samples to which an unknown one could be compared. In other words, for an unknown sample, will be given an attribution ranking probability saying to which sample it has the more similarities in terms of morphology characteristics (particle size, particle shape and distribution) and chemical composition, and with which percentage. This could help also in retrieve information related to the production process that could have been used during the production. Should be taken into consideration that, since this approach uses comparative signatures, the results must be also compared with those obtained from other techniques. This can help in taking the final conclusion about the understanding of the process history, the chemical composition and eventually the probably provenience and/or origin (country, place of production/facility) of the interdicted sample.

In this way, this new approach could also give a positive contribution to the investigative process; in fact, reducing from three to five the probably samples to which an unknown UOC powder could be identified, could speed up the starting of the investigation allowing the authorities to take proper decisions in less time.

5 Further developments and future implementation

Even if the results reached were successful, the final objective to have one single model must be reached. To do that, first of all the spectral-based classification model must be upgraded including also the hyperspectral image analysis with the correspondent hyperspectral-based classification model. Then, different strategy to combine all the models together, maybe using a model builder approach, can be explored.

A graphical User Interface must be also developed, in order to allow people involved in nuclear forensic analysis, but who has no experience in machine learning and programming, to use it easily.

Moreover, a validation test will be planned as soon as all the experimental data have been collected and the models have been upgraded with the new training data and a new set of features have been selected. The performance of the final models will be assessed considering also other different metric parameters (sensitivity, specificity, ROC curve, etc.), rather than only the accuracy.

References

1. Doris Mer Lin Ho. Study on the applicability of structural and morphological parameters on selected uranium compounds for nuclear forensic purposes. PhD thesis, 2015.
2. Michael J. Kristo, Amy M. Gaffney, Naomi Marks, Kim Knight, William S. Cassata, and Ian D. Hutcheon. Nuclear forensic science: Analysis of nuclear material out of regulatory control. *Annual Review of Earth and Planetary Sciences*, 44(1):555–579, 2016. doi:10.1146/annurevearth.060115-012309.
3. K. Mayer, M. Wallenius, I. Ray, Nuclear forensics – a methodology providing clues on the origin of illicitly trafficked nuclear materials. *Analyst* 130 (2005) 433–441.
4. M. Wallenius, K. Mayer, I. Ray, Nuclear forensic investigations: Two case studies, *Forensic Science International* 156 (2006) 55–62.
5. K. Mayer, M. Wallenius, T. Fanghänel, Nuclear forensic science – From cradle to maturity, *Journal of Alloys and Compounds* 444–445 (2007) 50–56.
6. M. Wallenius, et al., Nuclear forensic investigations with a focus on plutonium, *Journal of Alloys and Compounds* 444–445 (2007) 57–62.
7. Varga, Z., et al., Analysis of uranium ore concentrates for origin assessment, *Proc. Radiochimica Acta* 1 (2011) 1–4.
8. P.A. Budinger, T.L. Drenski, A.W. Varnes, J.R. Mooney, The Case of the Great Yellow Cake Caper, *Analytical Chemistry* 52 (1980) 942A–948A.
9. E. Keegan et al., Nuclear forensic analysis of an unknown uranium ore concentrate sample seized in a criminal investigation in Australia, *Forensic Science International* 240 (2014) 111–121.
10. E. Keegan et al., The provenance of Australian uranium ore concentrates by elemental and isotopic analysis, *Applied Geochemistry* 23 (2008) 765–777.
11. Z. Varga, M. Wallenius, K. Mayer, E. Hrncsek, Alternative method for the production date determination of impure uranium ore concentrate samples, *Journal of Radioanalytical and Nuclear Chemistry* 290 (2011) 485–492.
12. Z. Varga, M. Wallenius, K. Mayer, E. Keegan, S. Millet, Application of Lead and Strontium Isotope Ratio Measurements for the Origin Assessment of Uranium Ore Concentrates, *Analytical Chemistry* 81 (2009) 8327–8334.
13. Z. Varga, R. Katona, Z. Stefánka, M. Wallenius, K. Mayer, A. Nicholl, Determination of rare-earth elements in uranium-bearing materials by inductively coupled plasma mass spectrometry, *Talanta* 80 (2010) 1744–1749.
14. Z. Varga et al., Characterization and classification of uranium ore concentrates (yellow cakes) using infrared spectrometry, *Radiochimica Acta* 99 (2011) 1–7.
15. Fongaro et al. (2016), Image Analysis in nuclear forensics. EUR 27979 EN; doi: 10.2789/851269.
16. Ho, D.M.L., Manara, D., Varga, Z., Fongaro, L., Nicholl, A., Berlizov, A., Lindqvist-Reis, P., Fanghänel, T., Mayer, K. Exploring spectroscopic and morphological data as new signatures for uranium ore concentrates, Conference Paper, IAEA Nuclear Forensics Conference, Vienna, Austria, 2014.
17. L. Fongaro, Doris Mer Lin Ho, K. Kvaal, K. Mayer, V.V. Rondinella. 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* 152(2016)463–474.
18. J. Huang, K.H. Esbensen, Applications of Angle Measure Technique (AMT) in IA Part I. A new methodology for in situ powder characterization, *Chemometrics and Intelligent Laboratory Systems* 54 (2000) 1–19.
19. M. Vidal, J. M. Amigo. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems* 117 (2012) 138–148.
20. F. Camastra, A. Vinciarelli. Machine Learning for Audio, Image and Video Analysis: Theory and Application. Springer, ISBN: 978-1-84800-006-3.
21. Konica Minolta, Precise colour communication. <https://www5.konicaminolta.eu/en/measuring-instruments/learning-centre/colour-measurement/precise-colour-communication.html>.

22. János Schanda. Colorimetry: understanding the CIE system. John Wiley & Sons, 2007. doi:10.1002/9780470175637.
23. I. Lande. Predictive Machine Learning on SEM and Hyperspectral Images of Uranium Ore Concentrates (UOCs) for Nuclear Forensics. Master Thesis, University of Life Science – Aas (Norway), 2021.
24. AMT plugin for imagej. http://arken.nmbu.no/~kkvaal/eamtexplorer/imagej_plugins.html, 2014. Accessed: 2020-04-21.
25. G. L. Klunder, J. W. Plaué, P. E. Spackman, P. M. Grant, R. E. Lindvall, I. D. Hutcheon. Application of visible-near infrared reflectance spectroscopy to uranium ore concentrates for nuclear forensic analysis and attribution. *Applied Spectroscopy*, *Applied Spectroscopy* 67(9):1049-56
26. Mara Marchetti, Klaus Mayer, Maria Wallenius, Antonio Bulgheroni, Thierry Wiss, Klaus Lützenkirchen, and Lorenzo Fongaro. Image texture analysis and colorimetry for the classification of uranium ore concentrate powders. In *EPJ Web of Conferences*, volume 225, page 07003. EDP Sciences, 2020. doi:<https://doi.org/10.1051/epjconf/202022507003>.
27. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
28. Amalia Luque, Alejandro Carrasco, Alejandro Martín, and Ana de las Heras. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91:216 – 231, 2019. ISSN 0031-3203. doi:<https://doi.org/10.1016/j.patcog.2019.02.023>.

List of abbreviations and definitions

AMT	Angle Measure Technique
B-DB	Black-Dark Brawn
CC1	Colour Class 1
CC2	Colour Class 2
CC3	Colour Class 3
CC4	Colour Class 4
CC5	Colour Class 5
CC6	Colour Class 6
CV	Cross Validation
DB-O	Dark Brawn-Orange
DY-Y	Dark Yellow-Yellow
GLCM	Grey Level Co-occurrence Matrix
GLRLM	Grey Level Run Length Matrix
HCA	Hierarchical Cluster Analysis
LBP	Local Binary Pattern
LY-W	Light Yellow-White
LDA	Linear Discriminant Analysis
MCC	Matthews Correlation Coefficient
O-DY	Orange-Dark Yellow
P	Prediction
PCA	Principal Component Analysis
SVM	Support Vector Machine
UOC	Uranium Ore Concentrate
UOCs	Uranium Ore Concentrates
Y-LY	Yellow-Light Yellow

List of figures

Figure 1. Work breakdown structure of the IDENTICALSS project.	6
Figure 2. Subset of the samples under investigation. The powders shown are representative of some of the different colours possessed by the UOCs.	8
Figure 3. Sample holder with the 5 different regions from which the textural images have been acquired. From each square, the images at 100x, 250x and 500x/1000x of magnification were collected maintaining always the same centre.	8
Figure 4. Image texture dataset of samples belonging to the colour class 1. The images represented have been acquired at 250x (top) and 1000x (bottom) of magnification.	9
Figure 5. SisuCHEMA system installed at the JRC Karlsruhe.	10
Figure 6. Procedure to create the hyperspectral images necessary to build the corresponding hyperspectral image dataset. The columns show (from left to right) the raw hyperspectral images, the reflectance calibrated images, the reflectance calibrated images with the selected area from which the sub images were cropped, and finally the 4 square sub images cropped from the selected area.	11
Figure 7. Hyperspectral image before (left) and after (right) the reflectance calibration using the white and dark reference images.	11
Figure 8. Steps used for finding the best classifier and the dataset used (red dotted squares).	12
Figure 9. Overview of the process of finding optimized feature sets for a dataset and estimating their performance on hold-out test data.	13
Figure 10. Sample distribution in the score plot before (a) and after (b) a subjective predefined colour labels attribution to all the samples.	15
Figure 11. Resulting dendrogram after the Hierarchical cluster analysis of the entire samples collection. Different clusters are possible by selecting different threshold level in the dendrogram. In this work 6 different colour-groups were chosen.	16
Figure 12. Scattergram representing the sample distribution in function of their a^* and L^* values. The diameter of each sphere correspond to the b^* values.	16
Figure 13. Score plot of colour data using the colour classes, defined by HCA, as category variables.	17
Figure 14. Samples of test-set classified in the Yellow group. The classification prediction membership is measured either by a 0 (non-membership) or by a 1 (membership). Specifically, only the sample 66 is misclassified.	18
Figure 15. Average prediction matrices obtained from model 's validation test for CC1 at 100x, 250x, 1000x and for all magnifications.	21
Figure 16. Average prediction matrices obtained from the model 's validation test, for CC3 at 100x, 250x, 1000x and for all magnifications.	22
Figure 17. Prediction matrices for the three validation test for CC1 at 100x. UsPET sample has been correctly classified twice, while it was misclassified the first time.	24
Figure 18. Example of intra- and inter-class differences of images. The top row shows the three images of class UsPET that were classified in the three test runs. The two bottom images belonging to the classes UsATL and SaROS.	25
Figure 19. The spectra for all samples replicates (top); average spectra for each sample holder (middle); average spectra for each UOC sample. All spectra are colour coded by UOC sample name (see Table 2). The black vertical dotted lines indicate all selected wavelengths given the union feature set. The wavelengths along the horizontal axis are given in [nm] and the vertical axis is absorbance in arbitrary units.	26
Figure 20. Average prediction matrices for CC1, CC3, CC4 and CC6 obtained from the spectral-based models validation test.	27

Figure 21. The raw spectra of samples in the colour category 3 dataset. The vertical axis denotes reflectance in arbitrary units and the horizontal axis denotes wavelengths.	40
Figure 22. The reflectance calibrated spectra of samples in the colour category 3 dataset. The vertical axis denotes relative reflectance in arbitrary units and the horizontal axis denotes wavelengths.	40
Figure 23. The absorbance converted spectra of samples in the colour category 3 dataset. The vertical axis denotes reflectance in arbitrary units and the horizontal axis denotes wavelengths.....	41
Figure 24. Baseline correction applied on the absorbance converted spectra of samples in the colour category 3 dataset. The vertical axis denotes absorbance in arbitrary units and the horizontal axis denotes wavelengths.	41
Figure 25. Average prediction matrices obtained from model 's validation test for CC1 at 100x, 250x, 1000x and for all magnifications.	42
Figure 26. Average prediction matrices obtained from model 's validation test for CC3 at 100x, 250x, 500x and for all magnifications.	43
Figure 27. Average prediction matrices obtained from model 's validation test for CC4 at 100x, 250x, 500x and for all magnifications.	44
Figure 28. Average prediction matrices obtained from model 's validation test for CC5 at 100x, 250x, 500x and for all magnifications.	45
Figure 29. Average prediction matrices obtained from model 's validation test for CC6 at 100x, 250x, 500x and for all magnifications.	45

List of tables

Table 1. List of the analysed UOC samples.	7
Table 2. List of samples with the corresponding data that have been collected and analysed during the exploratory research. Yes, means that the data have been acquired and analysed; Yes/No, means that the data have been only acquired and need to be still analysed; No, means that the data need to be still acquired and analysed.	14
Table 3. Details about the 6 colour classes obtained after the cluster analysis: average values of L*, a* and b* index with the relative standard deviation and the ΔE values, the range and the colour class labels are reported.	17
Table 4. Evaluation metric, in cross-validation (CV) and prediction (P), of the colour-based classification model developed with SVM classifier.	18
Table 5. List of the six different colour classes with all the UOCs samples belonging to each class. Yes, means that the data have been acquired and analysed; Yes/No means that the data have been only acquired and must be analysed; No means that the data must be acquired and analysed. Information about the main steps of the production process have been reported (IX, ion exchange; SX, solvent extraction).	19
Table 6. Summary of models' performances of all the textural features datasets extracted from the SEM images.	25
Table 7. Summary of model performances of all the textural features datasets extracted from the SEM images.	27
Table 8. The first order statistics (FOS) features extracted from SEM images.	38
Table 9. The GLCM, GLSZM, GLRLM features extracted from SEM images.	38
Table 10. The LBP features extracted from SEM images.	39

Annexes

Annex 1. Description of the texture feature extraction algorithms used for the texture based classification model development

The principle of the texture feature extraction algorithms used in this research are here described.

1. First Order Statistics (FOS)[xx]. Simple statistic index calculation from the images grey level histograms
2. Grey Level Co-Occurrence Matrix[xx]. The GLCM is one of the first methods developed of for image texture analysis and was introduced by Haralick and co-workers in 1973 [22]. The GLCM calculates how often two pixels, in the matrix element $P_{\delta}(i, j)$, with intensity values i and j at a particular displacement distance δ from along a given direction θ (horizontally, vertically, or diagonally) occurs in the image [23]. If N is the highest number of quantized grey level values in the image, the co-occurrence matrix defined by Haralick is a square matrix G of dimension $N \times N$ (N can have a maximum value of 255 for a 8bit grey scale image), with elements represented by the frequency of occurrences of the pixel with intensity i next to that of intensity j . Haralick et al. [22] introduced 14 descriptors to obtain a quantitative measurement of the image textural features based on the co-occurrence matrix G .
3. Grey Level Run Length Matrix (GLRLM)[xx]. GLRLM method consists in counting the number of pixel segments having the same intensity in a given direction, then representing the results in a matrix. A direction (0° , 45° , 90° or 135°) and a number of gray levels are decided on beforehand. The value contained in the matrix's (l,n) square is equal to the number of segments of length l and gray level n [x].
4. Grey Level Size Zone Matrix (GLSZM) [xx]. GLSZM method is considered the texture in large areas instead of a group of pixels or segments in any given direction. To calculate the matrix is take into account the size of each area with pixels of the same intensity level. It is calculated according to the RLM principle. The resulting SZM matrix (s,n) has a dimension equal to the number of areas of size s and of gray level n .
5. Local Binary Pattern (LBP) [xx]: LBP is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. The most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes caused, for example, by illumination variations. Another important property is its computational simplicity, which makes it possible to analyze images in challenging real-time settings. Given a number of neighbouring points p evenly distributed on a circle of pixels with radius r from a center pixel, the neighbouring pixels are thresholded by the centre pixel so that pixels with an intensity value greater than the center pixel is 1 otherwise 0. Then, for each neighbouring pixel going around the circle (in a consistent way), the number two to the power of the position of a pixel with value 1 are added together. This is the calculate LBP value, which is assigned to the center pixel as a new pixel value. This is repeated for all pixels in the image. Then, a histogram of all these LBP values are created and used for further calculations. [R_16].
6. Angle Measure Technique: which describes the complexity of the signal simultaneously on all the existing scales through the Minimum Angle (MA) spectrum. [27]. The AMT was introduced by Andrieu [28] for geomorphic lines, but it finds useful application in the description of grey-level images which are mostly isotropic: hence do not present a preferential direction in the appearance of the characteristic features. Detailed descriptions of the AMT algorithm are available in [17],[28–32]; here the sequence of operations of the AMT are briefly summarized:
 - a) AMT performs the image unfolding: the image is represented as grey level intensities versus pixels in the unfolded image.
 - b) In each point of the sampling-set AMT centers a circle having $1 \leq \text{radius} \leq n$ (n is, in general, the image width).
 - c) AMT calculates the supplementary of the angle constructed by the intersection of the intensity line and the circle (see Figure 2).
 - d) AMT averages the angle values obtained for the different points of the sampling-set at a fixed radius.

e) AMT repeats the procedure until radius = n

In this study the scales $s=(i+2)$ for all i in $\{i \in \mathbb{Z} | -2 \leq i \leq 66\}$ were used, hence for a total of 35 features; basically all the mean angle values starting from $s=0$ to scale $s= 68$, taken with a step of 2.

Lists of each of the extracted textural features that were used for the texture based classification model development are listed in tables below.

Table 8. The first order statistics (FOS) features extracted from SEM images.

FOS
10Percentile
90Percentile
Energy
Entropy
InterquartileRange
Kurtosis
Maximum
MeanAbsoluteDeviation
Mean
Median
Minimum
Range
RobustMeanAbsoluteDeviation
RootMeanSquared
Skewness
TotalEnergy
Uniformity
Variance

Table 9. The GLCM, GLSZM, GLRLM features extracted from SEM images.

GLCM	GLSZM	GLRLM
Autocorrelation	GrayLevelNonUniformity	GrayLevelNonUniformity
ClusterProminence	GrayLevelNonUniformityNormalized	GrayLevelNonUniformityNormalized
ClusterShade	GrayLevelVariance	GrayLevelVariance
ClusterTendency	HighGrayLevelZoneEmphasis	HighGrayLevelRunEmphasis
Contrast	LargeAreaEmphasis	LongRunEmphasis
Correlation	LargeAreaHighGrayLevelEmphasis	LongRunHighGrayLevelEmphasis
DifferenceAverage	LargeAreaLowGrayLevelEmphasis	LongRunLowGrayLevelEmphasis
DifferenceEntropy	LowGrayLevelZoneEmphasis	LowGrayLevelRunEmphasis
DifferenceVariance	SizeZoneNonUniformity	RunEntropy
Id	SizeZoneNonUniformityNormalized	RunLengthNonUniformity
Idm	SmallAreaEmphasis	RunLengthNonUniformityNormalized
Idmn	SmallAreaHighGrayLevelEmphasis	RunPercentage
Idn	SmallAreaLowGrayLevelEmphasis	RunVariance
Imc1	ZoneEntropy	ShortRunEmphasis
Imc2	ZonePercentage	ShortRunHighGrayLevelEmphasis
InverseVariance	ZoneVariance	ShortRunLowGrayLevelEmphasis
JointAverage		
JointEnergy		
JointEntropy		
MCC		
MaximumProbability		
SumAverage		
SumEntropy		
SumSquares		

Table 10. The LBP features extracted from SEM images.

LBP						
L4_1_1	L16_4_12	L24_6_13	L28_7_28	L36_9_5	L40_10_8	L44_11_7
L4_1_2	L16_4_13	L24_6_14	L28_7_29	L36_9_6	L40_10_9	L44_11_8
L4_1_3	L16_4_14	L24_6_15	L28_7_30	L36_9_7	L40_10_10	L44_11_9
L4_1_4	L16_4_15	L24_6_16	L32_8_1	L36_9_8	L40_10_11	L44_11_10
L4_1_5	L16_4_16	L24_6_17	L32_8_2	L36_9_9	L40_10_12	L44_11_11
L4_1_6	L16_4_17	L24_6_18	L32_8_3	L36_9_10	L40_10_13	L44_11_12
L8_2_1	L16_4_18	L24_6_19	L32_8_4	L36_9_11	L40_10_14	L44_11_13
L8_2_2	L20_5_1	L24_6_20	L32_8_5	L36_9_12	L40_10_15	L44_11_14
L8_2_3	L20_5_2	L24_6_21	L32_8_6	L36_9_13	L40_10_16	L44_11_15
L8_2_4	L20_5_3	L24_6_22	L32_8_7	L36_9_14	L40_10_17	L44_11_16
L8_2_5	L20_5_4	L24_6_23	L32_8_8	L36_9_15	L40_10_18	L44_11_17
L8_2_6	L20_5_5	L24_6_24	L32_8_9	L36_9_16	L40_10_19	L44_11_18
L8_2_7	L20_5_6	L24_6_25	L32_8_10	L36_9_17	L40_10_20	L44_11_19
L8_2_8	L20_5_7	L24_6_26	L32_8_11	L36_9_18	L40_10_21	L44_11_20
L8_2_9	L20_5_8	L28_7_1	L32_8_12	L36_9_19	L40_10_22	L44_11_21
L8_2_10	L20_5_9	L28_7_2	L32_8_13	L36_9_20	L40_10_23	L44_11_22
L12_3_1	L20_5_10	L28_7_3	L32_8_14	L36_9_21	L40_10_24	L44_11_23
L12_3_2	L20_5_11	L28_7_4	L32_8_15	L36_9_22	L40_10_25	L44_11_24
L12_3_3	L20_5_12	L28_7_5	L32_8_16	L36_9_23	L40_10_26	L44_11_25
L12_3_4	L20_5_13	L28_7_6	L32_8_17	L36_9_24	L40_10_27	L44_11_26
L12_3_5	L20_5_14	L28_7_7	L32_8_18	L36_9_25	L40_10_28	L44_11_27
L12_3_6	L20_5_15	L28_7_8	L32_8_19	L36_9_26	L40_10_29	L44_11_28
L12_3_7	L20_5_16	L28_7_9	L32_8_20	L36_9_27	L40_10_30	L44_11_29
L12_3_8	L20_5_17	L28_7_10	L32_8_21	L36_9_28	L40_10_31	L44_11_30
L12_3_9	L20_5_18	L28_7_11	L32_8_22	L36_9_29	L40_10_32	L44_11_31
L12_3_10	L20_5_19	L28_7_12	L32_8_23	L36_9_30	L40_10_33	L44_11_32
L12_3_11	L20_5_20	L28_7_13	L32_8_24	L36_9_31	L40_10_34	L44_11_33
L12_3_12	L20_5_21	L28_7_14	L32_8_25	L36_9_32	L40_10_35	L44_11_34
L12_3_13	L20_5_22	L28_7_15	L32_8_26	L36_9_33	L40_10_36	L44_11_35
L12_3_14	L24_6_1	L28_7_16	L32_8_27	L36_9_34	L40_10_37	L44_11_36
L16_4_1	L24_6_2	L28_7_17	L32_8_28	L36_9_35	L40_10_38	L44_11_37
L16_4_2	L24_6_3	L28_7_18	L32_8_29	L36_9_36	L40_10_39	L44_11_38
L16_4_3	L24_6_4	L28_7_19	L32_8_30	L36_9_37	L40_10_40	L44_11_39
L16_4_4	L24_6_5	L28_7_20	L32_8_31	L36_9_38	L40_10_41	L44_11_40
L16_4_5	L24_6_6	L28_7_21	L32_8_32	L40_10_1	L40_10_42	L44_11_41
L16_4_6	L24_6_7	L28_7_22	L32_8_33	L40_10_2	L44_11_1	L44_11_42
L16_4_7	L24_6_8	L28_7_23	L32_8_34	L40_10_3	L44_11_2	L44_11_43
L16_4_8	L24_6_9	L28_7_24	L36_9_1	L40_10_4	L44_11_3	L44_11_44
L16_4_9	L24_6_10	L28_7_25	L36_9_2	L40_10_5	L44_11_4	L44_11_45
L16_4_10	L24_6_11	L28_7_26	L36_9_3	L40_10_6	L44_11_5	L44_11_46
L16_4_11	L24_6_12	L28_7_27	L36_9_4	L40_10_7	L44_11_6	

Annex 2. Pre-processing steps for the near infrared spectra extracted of colour class 3

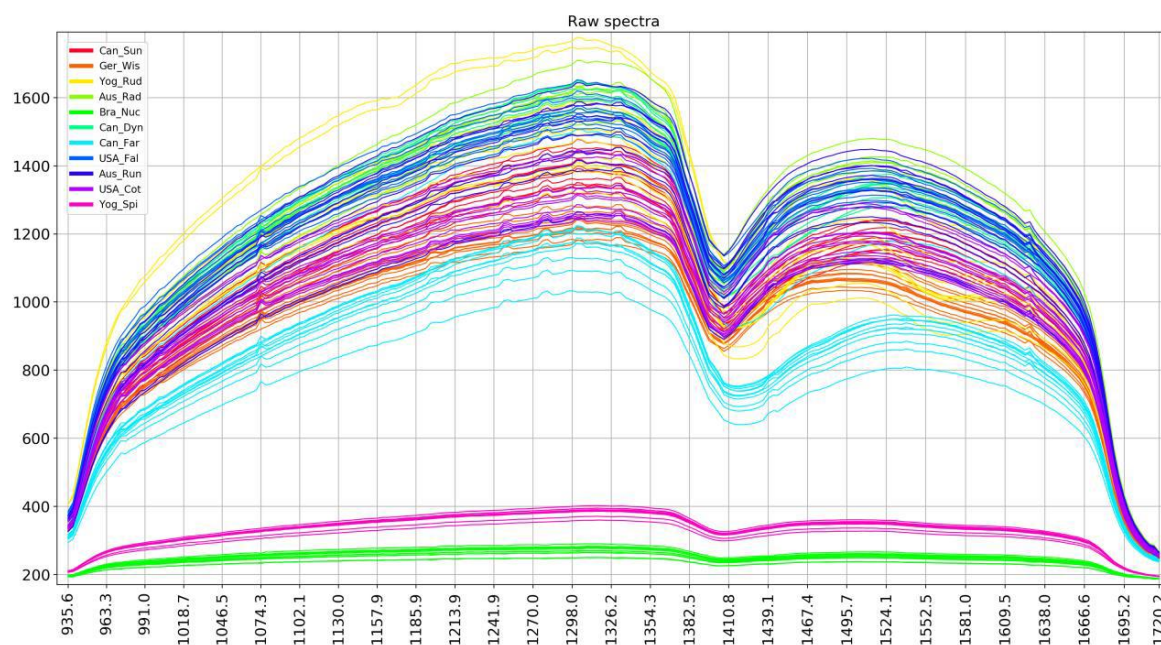


Figure 21. The raw spectra of samples in the colour category 3 dataset. The vertical axis denotes reflectance in arbitrary units and the horizontal axis denotes wavelengths.

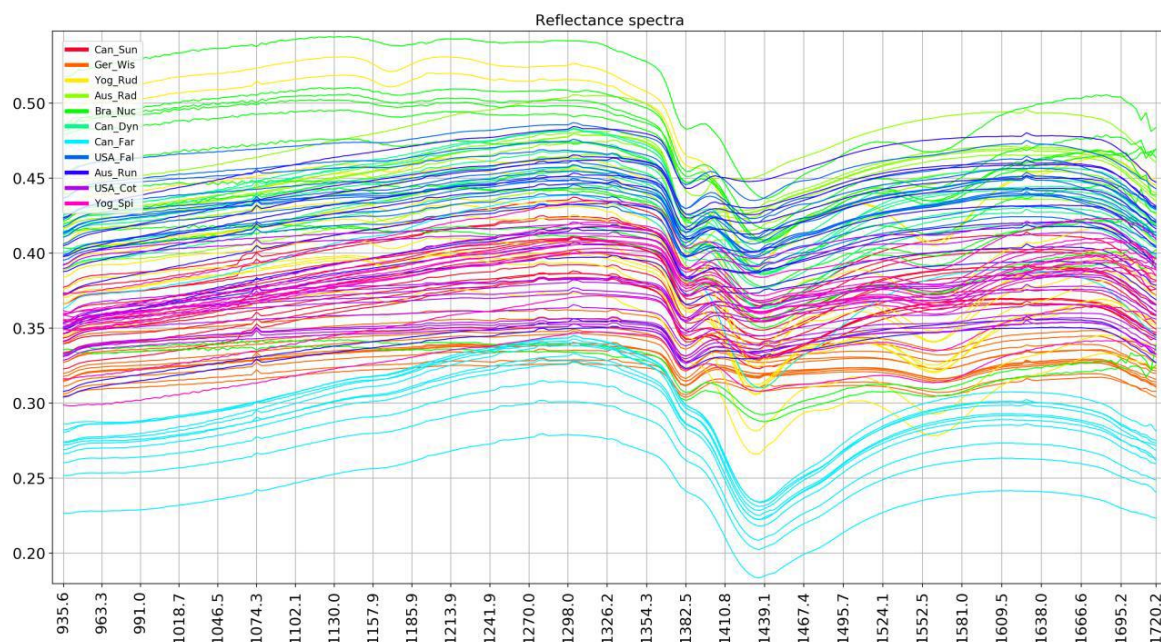


Figure 22. The reflectance calibrated spectra of samples in the colour category 3 dataset. The vertical axis denotes relative reflectance in arbitrary units and the horizontal axis denotes wavelengths.

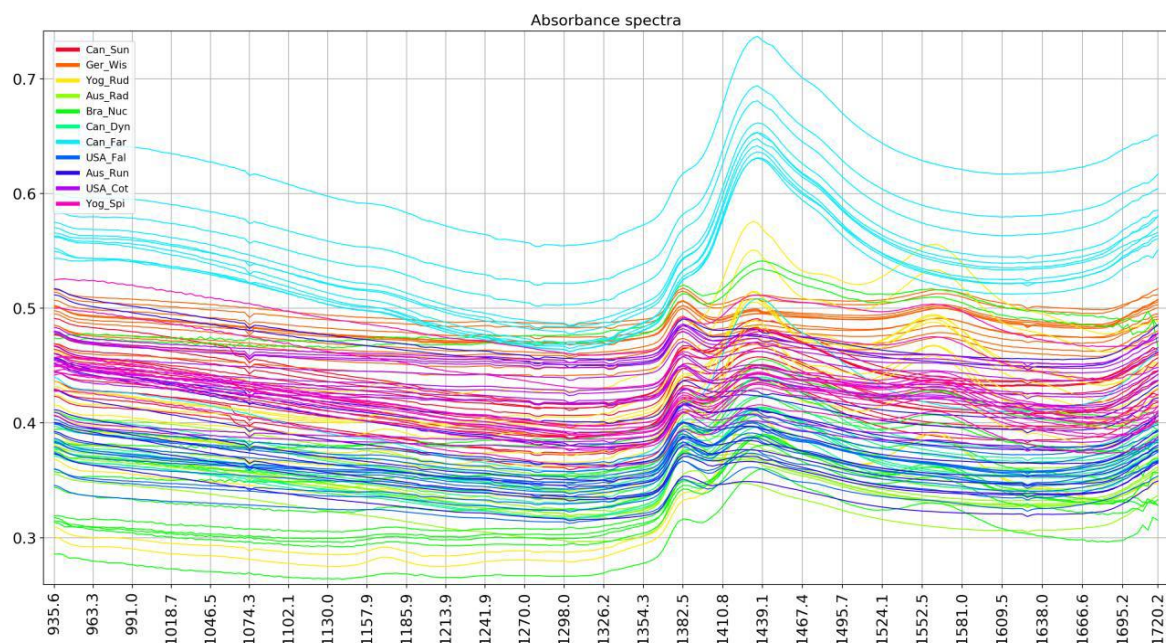


Figure 23. The absorbance converted spectra of samples in the colour category 3 dataset. The vertical axis denotes reflectance in arbitrary units and the horizontal axis denotes wavelengths.

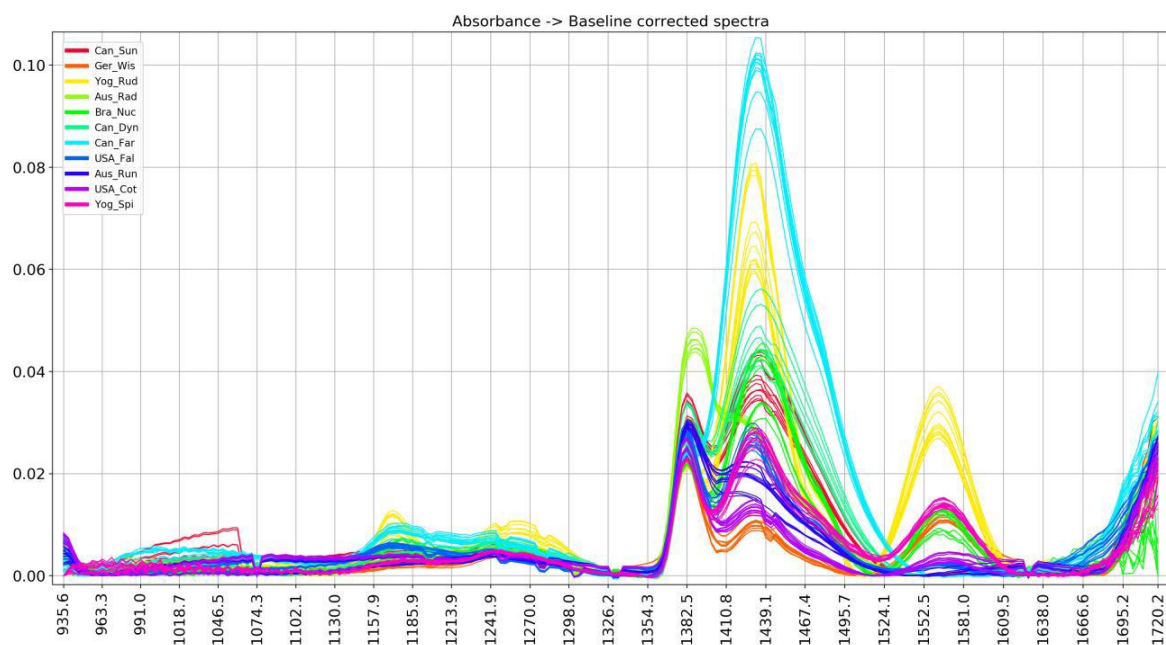


Figure 24. Baseline correction applied on the absorbance converted spectra of samples in the colour category 3 dataset. The vertical axis denotes absorbance in arbitrary units and the horizontal axis denotes wavelengths.

Annex 3. Average prediction probability matrices

	Pred.														
		AuMAK	AuOLD	AuQUE	CaKEL	ChHEY	RuTEC	SaNUF	SaPAL	SaROS	UsATL	UsFAP	UsPET	YuSPB	
100X	AuMAK	58,35%	0,02%	0,01%	0,00%	0,00%	0,00%	19,40%	0,20%	16,50%	0,07%	0,43%	4,85%	0,16%	TRUE
	AuOLD	0,17%	96,91%	0,01%	0,08%	0,00%	0,00%	0,00%	0,00%	1,75%	0,04%	1,02%	0,02%	0,00%	TRUE
	AuQUE	0,06%	0,00%	41,51%	0,10%	0,00%	33,27%	0,03%	0,00%	2,83%	0,00%	22,14%	0,00%	0,05%	TRUE
	CaKEL	0,00%	0,02%	17,02%	81,81%	0,00%	1,03%	0,05%	0,00%	0,01%	0,00%	0,01%	0,00%	0,06%	TRUE
	ChHEY	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	RuTEC	0,00%	0,31%	0,19%	0,43%	0,00%	99,03%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,03%	TRUE
	SaNUF	20,84%	25,70%	0,52%	0,00%	32,90%	0,48%	3,77%	0,03%	14,39%	0,02%	0,07%	1,09%	0,20%	FALSE
	SaPAL	5,08%	0,00%	0,00%	0,00%	0,00%	0,00%	0,82%	26,91%	1,08%	0,01%	0,00%	66,10%	0,00%	FALSE
	SaROS	14,19%	0,00%	0,03%	0,00%	0,00%	0,00%	6,95%	0,00%	45,71%	2,13%	22,67%	0,35%	7,97%	TRUE
	UsATL	1,21%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	7,19%	87,12%	0,38%	4,09%	0,00%	TRUE
	UsFAP	0,96%	0,00%	0,00%	0,00%	0,00%	0,00%	0,08%	0,00%	24,25%	1,90%	72,80%	0,00%	0,00%	TRUE
	UsPET	1,94%	0,00%	0,00%	0,00%	0,00%	0,00%	0,09%	6,89%	12,60%	20,02%	0,38%	58,08%	0,00%	TRUE
	YuSPB	3,54%	0,03%	0,18%	0,00%	0,00%	0,01%	4,36%	0,00%	1,55%	0,00%	0,03%	0,00%	90,30%	TRUE
250X	AuMAK	48,29%	0,02%	0,16%	0,00%	0,00%	0,00%	14,92%	0,11%	26,30%	2,56%	0,13%	7,35%	0,17%	TRUE
	AuOLD	7,75%	49,71%	5,56%	1,09%	0,00%	26,93%	0,05%	0,00%	8,61%	0,00%	0,20%	0,00%	0,11%	TRUE
	AuQUE	0,61%	24,47%	33,00%	9,18%	0,00%	0,00%	0,10%	0,00%	29,89%	0,00%	1,03%	0,00%	1,72%	TRUE
	CaKEL	0,67%	4,47%	42,71%	35,69%	0,00%	0,00%	0,00%	0,00%	4,32%	0,01%	0,10%	0,00%	12,02%	FALSE
	ChHEY	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	RuTEC	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	SaNUF	7,73%	0,00%	30,23%	1,29%	0,00%	0,00%	45,55%	8,94%	0,05%	0,00%	0,01%	4,38%	1,82%	TRUE
	SaPAL	0,07%	0,00%	0,00%	0,00%	0,00%	0,00%	33,26%	33,81%	0,00%	0,00%	0,00%	32,85%	0,00%	TRUE
	SaROS	18,52%	0,85%	3,03%	0,02%	0,00%	0,00%	0,01%	0,00%	61,86%	0,25%	14,93%	0,01%	0,52%	TRUE
	UsATL	0,02%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	2,67%	97,00%	0,30%	0,00%	0,01%	TRUE
	UsFAP	0,11%	0,00%	0,18%	0,27%	0,00%	0,00%	0,00%	0,00%	24,71%	0,77%	71,66%	0,00%	2,30%	TRUE
	UsPET	2,29%	0,00%	0,00%	0,00%	0,00%	0,00%	0,27%	6,11%	0,07%	0,02%	32,95%	58,28%	0,00%	TRUE
	YuSPB	0,37%	0,52%	12,29%	5,22%	0,00%	0,00%	0,01%	0,00%	15,00%	0,14%	1,72%	0,00%	64,73%	TRUE
1000X	AuMAK	31,76%	0,14%	21,91%	0,01%	0,00%	0,00%	0,00%	0,08%	14,86%	8,21%	0,58%	13,68%	8,78%	TRUE
	AuOLD	0,16%	60,26%	0,09%	31,63%	0,43%	0,00%	2,16%	2,15%	0,98%	0,00%	1,54%	0,60%	0,00%	TRUE
	AuQUE	7,14%	0,08%	45,90%	0,06%	0,00%	0,00%	0,00%	0,07%	20,79%	0,03%	10,13%	13,70%	2,10%	TRUE
	CaKEL	0,15%	30,09%	0,15%	56,64%	5,25%	0,33%	0,46%	0,05%	0,68%	0,00%	5,77%	0,44%	0,00%	TRUE
	ChHEY	0,00%	0,23%	0,00%	0,78%	87,35%	11,53%	0,00%	0,00%	0,00%	0,00%	0,10%	0,00%	0,00%	TRUE
	RuTEC	0,01%	0,00%	0,01%	0,01%	4,08%	95,68%	0,00%	0,00%	0,13%	0,02%	0,00%	0,06%	0,00%	TRUE
	SaNUF	0,00%	0,01%	0,00%	0,57%	0,30%	0,00%	85,66%	13,41%	0,00%	0,00%	0,04%	0,00%	0,00%	TRUE
	SaPAL	0,07%	0,22%	0,52%	0,53%	0,00%	0,00%	18,13%	69,28%	1,13%	0,00%	1,96%	8,16%	0,00%	TRUE
	SaROS	35,86%	2,41%	15,43%	1,51%	0,00%	0,00%	0,00%	0,05%	19,73%	0,09%	9,25%	15,56%	0,10%	FALSE
	UsATL	8,22%	0,01%	0,25%	0,00%	0,05%	23,20%	0,00%	0,00%	0,10%	67,33%	0,02%	0,02%	0,80%	TRUE
	UsFAP	0,88%	12,95%	10,95%	11,37%	0,17%	0,24%	0,04%	0,74%	10,87%	0,00%	43,14%	8,59%	0,05%	TRUE
	UsPET	16,92%	0,63%	13,73%	12,05%	0,00%	0,00%	19,05%	1,85%	17,43%	0,01%	1,59%	16,69%	0,06%	FALSE
	YuSPB	3,26%	0,01%	5,55%	0,01%	0,00%	0,01%	0,00%	0,03%	1,22%	0,64%	0,46%	3,98%	84,84%	TRUE
ALL	AuMAK	70,81%	0,02%	0,03%	0,00%	0,00%	0,00%	0,00%	0,01%	13,17%	0,25%	0,04%	8,36%	7,32%	TRUE
	AuOLD	0,01%	33,52%	9,24%	12,28%	0,00%	43,89%	0,01%	0,00%	0,19%	0,00%	0,87%	0,00%	0,00%	FALSE
	AuQUE	0,08%	13,67%	60,20%	1,60%	0,00%	0,00%	0,00%	0,00%	10,37%	0,00%	7,74%	0,00%	6,34%	TRUE
	CaKEL	0,00%	2,31%	15,32%	80,92%	0,00%	0,25%	0,00%	0,00%	0,40%	0,00%	0,46%	0,00%	0,36%	TRUE
	ChHEY	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	RuTEC	0,00%	0,00%	0,00%	0,19%	0,00%	99,81%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,01%	TRUE
	SaNUF	0,01%	0,01%	0,00%	0,00%	0,01%	0,00%	96,43%	0,19%	0,04%	0,00%	0,05%	0,07%	3,19%	TRUE
	SaPAL	0,02%	0,00%	0,00%	0,00%	0,00%	0,00%	33,33%	53,01%	0,00%	0,00%	0,00%	13,62%	0,01%	TRUE
	SaROS	3,06%	2,17%	14,11%	0,02%	0,00%	0,00%	0,02%	0,00%	32,42%	1,77%	14,53%	0,01%	31,89%	TRUE
	UsATL	1,20%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	4,33%	91,68%	2,70%	0,01%	0,08%	TRUE
	UsFAP	0,58%	1,70%	2,38%	0,04%	0,00%	0,00%	0,14%	0,00%	18,03%	0,57%	73,41%	0,00%	3,15%	TRUE
	UsPET	2,41%	0,00%	0,00%	0,00%	0,00%	0,00%	0,34%	3,55%	0,68%	0,11%	25,83%	67,01%	0,07%	TRUE
	YuSPB	1,86%	0,97%	4,51%	0,03%	0,00%	0,00%	0,00%	0,00%	9,80%	0,01%	0,38%	0,00%	82,45%	TRUE

Figure 25. Average prediction matrices obtained from model 's validation test for CC1 at 100x, 250x, 1000x and for all magnifications.

	Pred.	USA_Ana	Be_Cong	Ca_Dyno	USA_Cot	USA_Fall	Bra_Nuc	Au_RadH	Au_RumJ	Yo_SpisYe	USA_ESI	Ge_Wis	Au_Yee	Ca_Mac	
	True														
100X	USA_Ana	81,94%	0,00%	0,00%	16,70%	0,02%	0,13%	0,00%	0,00%	0,02%	0,00%	0,00%	0,00%	1,19%	TRUE
	Be_Cong	0,00%	79,75%	0,00%	0,00%	0,00%	0,00%	0,07%	20,18%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Ca_Dyno	0,00%	0,00%	92,64%	0,00%	0,03%	0,00%	0,00%	0,00%	0,02%	0,00%	0,00%	7,32%	0,00%	TRUE
	USA_Cot	11,40%	0,00%	0,00%	88,36%	0,00%	0,24%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	USA_Fall	0,00%	0,00%	0,01%	0,00%	63,08%	0,00%	0,00%	0,00%	32,85%	0,73%	0,08%	1,27%	1,98%	TRUE
	Bra_Nuc	0,39%	0,00%	0,00%	1,00%	0,00%	98,60%	0,01%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Au_RadH	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	99,82%	0,18%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Au_RumJ	0,00%	6,47%	0,00%	0,00%	0,00%	0,00%	2,33%	91,20%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Yo_SpisYe	0,00%	0,00%	3,63%	0,00%	23,01%	0,00%	0,00%	0,00%	54,53%	0,01%	0,05%	16,34%	2,43%	TRUE
	USA_ESI	0,00%	0,00%	0,00%	0,00%	0,44%	0,00%	0,00%	0,00%	0,42%	68,29%	0,06%	0,02%	30,77%	TRUE
	Ge_Wis	0,00%	0,00%	0,00%	0,00%	0,09%	0,00%	0,00%	0,00%	0,20%	0,00%	99,70%	0,01%	0,00%	TRUE
	Au_Yee	0,00%	0,00%	1,26%	0,00%	2,21%	0,00%	0,00%	0,00%	7,09%	0,05%	0,18%	89,14%	0,08%	TRUE
	Ca_Mac	1,90%	0,00%	0,00%	18,51%	0,18%	0,07%	0,00%	0,00%	3,87%	0,03%	0,11%	0,00%	75,32%	TRUE
250X	True														
	USA_Ana	76,65%	0,01%	0,00%	0,69%	10,49%	0,00%	0,00%	0,00%	0,44%	11,46%	0,26%	0,00%	0,00%	TRUE
	Be_Cong	0,00%	76,86%	0,00%	0,00%	0,00%	0,00%	0,00%	0,01%	0,00%	23,13%	0,00%	0,00%	0,00%	TRUE
	Ca_Dyno	32,28%	0,00%	66,66%	0,00%	0,00%	0,00%	0,00%	0,00%	0,18%	0,00%	0,88%	0,01%	0,00%	TRUE
	USA_Cot	0,02%	0,00%	0,00%	67,48%	0,01%	0,00%	1,41%	0,00%	6,50%	0,00%	1,07%	0,20%	23,31%	TRUE
	USA_Fall	0,01%	0,00%	0,00%	0,06%	94,70%	0,00%	0,00%	0,00%	0,08%	0,00%	5,14%	0,00%	0,00%	TRUE
	Bra_Nuc	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Au_RadH	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Au_RumJ	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Yo_SpisYe	3,41%	0,00%	0,00%	0,00%	0,01%	0,00%	0,00%	0,00%	80,18%	0,00%	1,42%	12,51%	2,46%	TRUE
	USA_ESI	0,00%	5,41%	0,00%	0,09%	30,72%	0,00%	0,02%	0,00%	2,06%	34,30%	27,39%	0,00%	0,00%	TRUE
	Ge_Wis	0,32%	0,00%	0,00%	0,01%	2,10%	0,00%	0,00%	0,00%	0,16%	0,00%	97,41%	0,00%	0,00%	TRUE
	Au_Yee	0,41%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	7,84%	0,00%	0,00%	83,71%	8,04%	TRUE
	Ca_Mac	0,00%	0,00%	0,00%	0,37%	0,00%	0,00%	0,00%	0,00%	21,06%	0,00%	0,24%	20,67%	57,65%	TRUE
500X	True														
	USA_Ana	97,80%	0,01%	0,00%	0,17%	0,00%	0,00%	0,73%	0,00%	0,00%	0,00%	0,21%	1,07%	0,00%	TRUE
	Be_Cong	25,14%	41,78%	0,00%	0,00%	0,00%	0,00%	0,29%	0,00%	0,00%	32,74%	0,01%	0,04%	0,00%	TRUE
	Ca_Dyno	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	USA_Cot	0,00%	0,00%	0,00%	66,84%	0,00%	0,00%	1,24%	0,01%	0,00%	0,04%	31,86%	0,01%	0,00%	TRUE
	USA_Fall	0,00%	0,00%	0,00%	0,00%	96,36%	0,00%	0,00%	0,00%	0,12%	0,00%	0,18%	0,02%	3,32%	TRUE
	Bra_Nuc	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	Au_RadH	0,00%	0,28%	0,00%	0,08%	0,00%	0,00%	99,52%	0,00%	0,00%	0,00%	0,12%	0,00%	0,00%	TRUE
	Au_RumJ	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	99,99%	0,00%	0,01%	0,00%	0,00%	0,00%	TRUE
	Yo_SpisYe	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	USA_ESI	0,00%	0,00%	0,00%	0,00%	0,60%	0,00%	0,06%	7,41%	0,00%	41,08%	49,57%	1,03%	0,24%	FALSE
	Ge_Wis	0,00%	0,00%	0,00%	0,00%	0,43%	0,00%	0,00%	0,00%	0,00%	0,00%	98,59%	0,98%	0,00%	TRUE
	Au_Yee	7,62%	0,00%	0,00%	0,00%	0,61%	0,00%	0,00%	0,00%	0,01%	0,66%	69,89%	21,20%	0,00%	TRUE
	Ca_Mac	0,00%	0,00%	0,00%	0,00%	22,23%	0,00%	0,00%	0,00%	1,21%	0,50%	22,01%	54,04%	0,00%	TRUE
ALL	True														
	USA_Ana	73,42%	16,39%	0,00%	10,06%	0,03%	0,07%	0,00%	0,01%	0,00%	0,00%	0,01%	0,01%	0,00%	TRUE
	Be_Cong	1,06%	73,74%	0,00%	0,85%	0,04%	6,02%	0,00%	0,23%	17,91%	0,00%	0,00%	0,05%	0,10%	TRUE
	Ca_Dyno	0,03%	0,00%	37,79%	0,01%	31,60%	0,07%	0,00%	0,00%	0,00%	0,40%	0,00%	0,00%	30,09%	TRUE
	USA_Cot	0,59%	15,36%	0,00%	53,21%	0,03%	0,00%	30,00%	0,03%	0,07%	0,42%	0,00%	0,29%	0,00%	TRUE
	USA_Fall	0,70%	0,00%	0,00%	1,45%	81,99%	5,63%	0,04%	0,28%	0,00%	4,98%	0,19%	1,94%	2,80%	TRUE
	Bra_Nuc	0,00%	0,00%	0,00%	1,83%	0,04%	66,97%	6,23%	4,49%	0,00%	2,74%	0,00%	17,66%	0,03%	TRUE
	Au_RadH	0,01%	0,00%	0,00%	1,45%	0,01%	0,02%	95,47%	0,64%	0,00%	0,38%	0,00%	2,02%	0,00%	TRUE
	Au_RumJ	0,05%	0,11%	0,00%	2,44%	0,05%	2,48%	0,12%	73,64%	11,77%	0,16%	0,00%	9,13%	0,04%	TRUE
	Yo_SpisYe	0,00%	6,81%	0,00%	0,01%	0,00%	0,33%	0,00%	5,87%	86,93%	0,00%	0,00%	0,04%	0,01%	TRUE
	USA_ESI	0,01%	0,00%	1,95%	1,05%	3,61%	4,16%	0,46%	0,96%	0,00%	63,29%	1,52%	8,02%	14,96%	TRUE
	Ge_Wis	0,04%	0,00%	0,00%	0,00%	0,23%	0,01%	0,00%	0,00%	0,00%	0,14%	99,53%	0,04%	0,00%	TRUE
	Au_Yee	0,04%	0,00%	0,01%	4,24%	1,71%	2,85%	1,00%	5,93%	0,01%	2,93%	0,00%	81,10%	0,20%	TRUE
	Ca_Mac	0,03%	0,00%	5,89%	0,39%	11,33%	4,10%	0,05%	0,17%	0,00%	8,56%	0,00%	1,45%	68,03%	TRUE

Figure 26. Average prediction matrices obtained from model´s validation test for CC3 at 100x, 250x, 500x and for all magnifications.

	Pred.										
		Ca_Den	Ge_Hel	S.Af_EFI	USA_KMcG	Ca_MiLL	Swe_Ran	Rum	Spa_Gen	Au_ShAll	
100x	True										
	Ca_Den	97,46%	0,00%	0,00%	0,00%	0,00%	2,36%	0,15%	0,00%	0,03%	TRUE
	Ge_Hel	0,00%	98,08%	0,00%	0,00%	0,42%	1,50%	0,00%	0,00%	0,00%	TRUE
	S.Af_EFI	0,00%	0,00%	99,87%	0,02%	0,00%	0,00%	0,00%	0,01%	0,10%	TRUE
	USA_KMcG	0,00%	0,00%	0,01%	97,79%	0,00%	0,00%	0,00%	2,20%	0,00%	TRUE
	Ca_MiLL	0,00%	0,26%	0,00%	0,00%	98,49%	0,27%	0,98%	0,00%	0,00%	TRUE
	Swe_Ran	0,04%	0,14%	0,00%	0,00%	0,70%	98,39%	0,73%	0,00%	0,00%	TRUE
	Rum	0,04%	0,38%	0,00%	0,00%	8,45%	24,80%	66,21%	0,00%	0,11%	TRUE
	Spa_Gen	0,00%	0,00%	0,00%	21,15%	0,00%	0,01%	0,00%	78,83%	0,00%	TRUE
	Au_ShAll	0,07%	0,00%	3,62%	0,00%	0,00%	0,00%	0,00%	0,00%	96,31%	TRUE
250x	True										
	Ca_Den	99,95%	0,00%	0,01%	0,02%	0,00%	0,02%	0,00%	0,00%	0,00%	TRUE
	Ge_Hel	0,00%	96,62%	0,00%	0,00%	0,00%	0,00%	0,00%	3,38%	0,00%	TRUE
	S.Af_EFI	0,33%	0,00%	99,26%	0,40%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	USA_KMcG	0,05%	0,00%	0,02%	99,92%	0,00%	0,01%	0,00%	0,00%	0,00%	TRUE
	Ca_MiLL	0,02%	27,45%	0,00%	0,00%	72,47%	0,02%	0,02%	0,00%	0,02%	TRUE
	Swe_Ran	0,08%	0,00%	0,00%	0,04%	0,00%	99,88%	0,00%	0,00%	0,00%	TRUE
	Rum	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	TRUE
	Spa_Gen	0,00%	4,08%	0,00%	0,00%	0,00%	0,00%	0,00%	95,92%	0,00%	TRUE
	Au_ShAll	0,00%	0,00%	0,00%	0,00%	0,02%	0,00%	33,32%	0,00%	66,65%	TRUE
500x	True										
	Ca_Den	66,07%	0,00%	0,00%	0,00%	0,00%	0,00%	6,41%	0,00%	27,52%	TRUE
	Ge_Hel	0,00%	99,56%	0,24%	0,00%	0,00%	0,00%	0,00%	0,20%	0,00%	TRUE
	S.Af_EFI	0,00%	0,89%	91,89%	5,30%	0,00%	0,59%	0,00%	1,33%	0,00%	TRUE
	USA_KMcG	0,00%	0,00%	5,13%	93,34%	0,00%	0,00%	0,00%	1,53%	0,00%	TRUE
	Ca_MiLL	0,00%	0,00%	0,00%	0,00%	99,99%	0,00%	0,01%	0,00%	0,00%	TRUE
	Swe_Ran	0,00%	0,12%	0,18%	0,00%	0,00%	99,69%	0,00%	0,00%	0,01%	TRUE
	Rum	19,08%	0,00%	0,00%	0,00%	21,92%	0,00%	53,48%	0,00%	5,53%	TRUE
	Spa_Gen	0,00%	0,85%	23,28%	13,81%	0,00%	0,00%	0,00%	62,06%	0,00%	TRUE
	Au_ShAll	32,95%	0,00%	0,00%	0,00%	0,00%	0,04%	2,09%	0,00%	64,91%	TRUE
ALL	True										
	Ca_Den	81,86%	0,00%	0,00%	0,00%	17,98%	0,06%	0,00%	0,10%	0,00%	TRUE
	Ge_Hel	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	S.Af_EFI	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	TRUE
	USA_KMcG	0,00%	0,05%	0,01%	97,84%	0,00%	0,00%	0,00%	0,00%	2,09%	TRUE
	Ca_MiLL	0,03%	0,33%	0,00%	0,00%	88,82%	0,03%	10,78%	0,00%	0,00%	TRUE
	Swe_Ran	0,47%	0,00%	0,00%	0,00%	0,00%	99,47%	0,00%	0,06%	0,00%	TRUE
	Rum	0,00%	0,00%	0,00%	0,00%	0,03%	2,10%	97,87%	0,00%	0,00%	TRUE
	Spa_Gen	0,92%	0,00%	0,00%	0,00%	0,00%	0,80%	0,00%	98,27%	0,00%	TRUE
	Au_ShAll	0,00%	0,00%	0,00%	13,61%	0,00%	0,00%	0,00%	0,00%	86,39%	TRUE

Figure 27. Average prediction matrices obtained from model 's validation test for CC4 at 100x, 250x, 500x and for all magnifications.

100x	Pred.	USA_ElMe	Hol_Del	Ca_Stan	USA_Dawn	
	True					
	USA_ElMe	89,09%	0,00%	0,00%	10,91%	TRUE
	Hol_Del	4,54%	95,42%	0,04%	0,00%	TRUE
	Ca_Stan	0,00%	0,00%	100,00%	0,00%	TRUE
250x	Pred.	USA_ElMe	Hol_Del	Ca_Stan	USA_Dawn	
	True					
	USA_ElMe	56,20%	13,32%	7,67%	22,81%	TRUE
	Hol_Del	17,35%	71,57%	10,21%	0,87%	TRUE
	Ca_Stan	0,06%	0,18%	90,94%	8,82%	TRUE
500x	Pred.	USA_ElMe	Hol_Del	Ca_Stan	USA_Dawn	
	True					
	USA_ElMe	22,26%	23,63%	51,17%	2,94%	FALSE
	Hol_Del	62,29%	31,03%	1,76%	4,92%	FALSE
	Ca_Stan	1,63%	1,92%	90,67%	5,77%	TRUE
ALL	Pred.	USA_ElMe	Hol_Del	Ca_Stan	USA_Dawn	
	True					
	USA_ElMe	70,05%	7,50%	0,55%	21,90%	TRUE
	Hol_Del	9,98%	89,64%	0,14%	0,24%	TRUE
	Ca_Stan	0,04%	0,28%	99,54%	0,14%	TRUE
500x	Pred.	USA_ElMe	Hol_Del	Ca_Stan	USA_Dawn	
	True					
	USA_ElMe	76,46%	0,57%	0,96%	22,01%	TRUE
	USA_Iri	7,53%	92,42%	0,01%	0,04%	TRUE
	Ca_RaL	13,85%	0,00%	65,06%	21,10%	TRUE
ALL	Pred.	USA_ElMe	Hol_Del	Ca_Stan	USA_Dawn	
	True					
	USA_ElMe	81,51%	18,49%	0,00%	0,00%	TRUE
	USA_Iri	0,04%	99,96%	0,00%	0,00%	TRUE
	Ca_RaL	0,00%	0,00%	99,96%	0,04%	TRUE
500x	Pred.	USA_ElMe	Hol_Del	Ca_Stan	USA_Dawn	
	True					
	USA_ElMe	76,46%	0,57%	0,96%	22,01%	TRUE
	USA_Iri	7,53%	92,42%	0,01%	0,04%	TRUE
	Ca_RaL	13,85%	0,00%	65,06%	21,10%	TRUE
ALL	Pred.	USA_ElMe	Hol_Del	Ca_Stan	USA_Dawn	
	True					
	USA_ElMe	81,51%	18,49%	0,00%	0,00%	TRUE
	USA_Iri	0,04%	99,96%	0,00%	0,00%	TRUE
	Ca_RaL	0,00%	0,00%	99,96%	0,04%	TRUE

Figure 28. Average prediction matrices obtained from model's validation test for CC5 at 100x, 250x, 500x and for all magnifications.

100x	Pred.	USA_EvYe	USA_Iri	Ca_RaL	USA_Mobil	
	True					
	USA_EvYe	100,00%	0,00%	0,00%	0,00%	TRUE
	USA_Iri	0,00%	99,70%	0,00%	0,30%	TRUE
	Ca_RaL	0,00%	0,06%	99,94%	0,00%	TRUE
250x	Pred.	USA_EvYe	USA_Iri	Ca_RaL	USA_Mobil	
	True					
	EverYe	99,56%	0,39%	0,00%	0,05%	TRUE
	Irigar	0,03%	99,77%	0,00%	0,20%	TRUE
	Rabbla	0,00%	0,00%	99,16%	0,84%	TRUE
500x	Pred.	USA_EvYe	USA_Iri	Ca_RaL	USA_Mobil	
	True					
	USA_EvYe	76,46%	0,57%	0,96%	22,01%	TRUE
	USA_Iri	7,53%	92,42%	0,01%	0,04%	TRUE
	Ca_RaL	13,85%	0,00%	65,06%	21,10%	TRUE
ALL	Pred.	USA_EvYe	USA_Iri	Ca_RaL	USA_Mobil	
	True					
	USA_EvYe	81,51%	18,49%	0,00%	0,00%	TRUE
	USA_Iri	0,04%	99,96%	0,00%	0,00%	TRUE
	Ca_RaL	0,00%	0,00%	99,96%	0,04%	TRUE
500x	Pred.	USA_EvYe	USA_Iri	Ca_RaL	USA_Mobil	
	True					
	USA_EvYe	76,46%	0,57%	0,96%	22,01%	TRUE
	USA_Iri	7,53%	92,42%	0,01%	0,04%	TRUE
	Ca_RaL	13,85%	0,00%	65,06%	21,10%	TRUE
ALL	Pred.	USA_EvYe	USA_Iri	Ca_RaL	USA_Mobil	
	True					
	USA_EvYe	81,51%	18,49%	0,00%	0,00%	TRUE
	USA_Iri	0,04%	99,96%	0,00%	0,00%	TRUE
	Ca_RaL	0,00%	0,00%	99,96%	0,04%	TRUE

Figure 29. Average prediction matrices obtained from model's validation test for CC6 at 100x, 250x, 500x and for all magnifications.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub
ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



EU Science, Research and Innovation



EU Science Hub



Publications Office
of the European Union

doi:10.2760/180882

ISBN 978-92-76-38732-9