A new versatile algorithm to extract particle's features from FESEM images: method evaluation and a case study on rice kernels

Chemometrics and Intelligent Laboratory Systems 263 (2025) 105415

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

A new versatile algorithm to extract particle's features from FESEM images: method evaluation and a case study on rice kernels

Mattia Sozzi [a], Nicola Cavallini [a,*], Alessandro Chiadò [a], Gentian Gavoci [a], Enrico Cantaluppi [b], Filip Haxhari [b], Francesco Savorani [a]

[a] Department of Applied Science and Technology, Polytechnic of Turin, Corso Duca degli Abruzzi 24, 10129, Turin, Italy
[b] Centro Ricerche sul Riso, Ente Nazionale Risi, 27030, Castello D'Agogna, Italy

ARTICLE INFO

keywords:
Image analysis
Morphological features
FESEM images
Algorithm
Rice kernels
DoE

ABSTRACT

Image analysis approaches allow to quickly extract important information from images of diverse nature. Many techniques produce as a result images that contain regular and irregular objects. The ability of automatically extracting the objects and their related morphological features and properties is becoming fundamental, especially when the number of images to analyse is consistent.

In this context, a new algorithm able to extract a series of morphological features from FESEM images was developed. Starting from a case study on 54 varieties of rice kernels, 220 images were acquired, and the algorithm was coded with the aim of extracting information from the round-shaped starch particles naturally present in rice kernels. The algorithm constitutes of different steps to segment the images and identify the object shapes and boundaries. Once those objects are identified, the algorithm extracts their morphological features, the number of identified objects and the amount of empty spaces among those objects.

The developed algorithm is suitable for a rapid and automated analysis of several images, with the aim of extracting object-related morphological features and information about the general objects space disposition. The use of adaptive thresholds and correction steps allow to analyse images of different natures containing also defective and non-representative objects that will be automatically removed from the features calculation. In addition, to evaluate the algorithm performances, a Design of Experiment approach was developed to determine the effect of the input parameters choice on the algorithm output results, highlighting which parameters show a stronger effect on the output.

1. Introduction

In the past years, the interest towards approaches based on automatic image analysis has been considerably growing. The possibility of rapid automated extraction of qualitative and quantitative information from greyscale and/or coloured images changed the way researchers have been performing their studies and their understanding of image-related scientific issues. Image analysis is generally non-invasive and non-destructive [1], and it provides accurate information by employing methods that can be used in a wide variety of fields such as materials science [2–4], chemistry [5–8], engineering [9–11] and even biomedicine [12–14].

Several analytical techniques generate images with regular shaped objects with a wide variety of dimensions and dispositions [15,16]. Researchers often face the problem of measuring particles and objects present in the images and many approaches, both automatic and manual, have been developed and fruitfully applied [17,18]. Among image analysis and characterization methodologies, Field Emission Scanning Electron Microscopy (FESEM) is gaining more and more momentum. Its application in engineering ranges from empty spaces quantification [19] to particle size distribution measurement [20,21] and morphology exploration [22]. It is also used to characterize grains of different materials (e.g. sand [23]) to evaluate structural morphology [24] and damages [25], and in general also to characterize materials with different aims like carbon materials for energy storage [26] or nanotubes for biomedical approaches [27]. More specifically, in food science, microscopies can play a key role in food components identification, contaminant detection, quality control and texture or structure analysis [28–34].

In the context of image analysis, segmentation [35] is a key step that

* Corresponding author.
E-mail address: nicola.cavallini@polito.it (N. Cavallini).

https://doi.org/10.1016/j.chemolab.2025.105415
Received 7 February 2025; Received in revised form 23 April 2025; Accepted 23 April 2025
Available online 24 April 2025

helps dividing an image into distinct regions to isolate areas of interest. This step, essential for tasks like object recognition and measurement, plays a crucial role and, nowadays, it is largely applied in combination with deep learning techniques to segment and classify macroscopic images [36]. While the principles of segmentation and object recognition are inherently applicable to FESEM images of food samples, the specific challenges lie in adapting algorithms to the unique microstructural features, such as complex textures and the presence of areas differing in brightness and light contrast. AI-driven analysis of macroscopic food images can be found in recent literature, for example for dietary assessment and quality control [37,38]. However, there is a noticeable gap in the direct application of AI-based segmentation and object recognition in FESEM images, especially for food-related topics. Differently from industrial materials, food samples usually show larger heterogeneity with complex structures and object dispositions, leading to different light contrast across the image, which can cause deep learning algorithms to fail. In addition, in the agrifood sector, especially for safety and quality applications, it is crucial to explain how algorithms work and the decisions they make, while deep learning models are often seen as "black boxes,' which can limit their adoption in regulated environments.

Differently from deep learning methods, more "classical" image analysis approaches, based on thresholding and contour-based algorithms, offer greater interpretability, as well as lower computational requirements and, most important, larger control over the extracted features, making them suitable and adjustable to specific applications.

In this context, we developed an algorithm able to perform high-throughput analyses of large sets of FESEM images to automatically extract important morphological features (e.g., area, perimeter, circularity, etc.) of similarly shaped objects. To show the applicability and the effectiveness of this approach, we also present a case study about rice kernels characterization, in which this new algorithm is applied. The aim of the study was to extract a series of morphological features related to the starch particles of rice kernels of different varieties and geographical origin from a set of FESEM images [39]. The use of FESEM in this context allows to easily visualize the starch particle shapes and disposition inside the rice kernel. Each different rice variety is characterized by a different kernel inner composition, resulting also in a different disposition of the starch particles, which are turned to be related with rice biochemical properties.

Furthermore, since the algorithm requires some parameters to be manually set as the model inputs, an experimental design approach was applied to evaluate the effect of these parameters on the morphological features calculation and on the overall algorithm processing runtime.

2. Materials and methods

2.1. Case study on rice samples: sample preparation and FESEM images acquisition

Our case study involved the analysis of 54 rice varieties to evaluate, by means of FESEM imaging, the morphological internal structure of the starch particles of the rice kernels [39]. Three half kernels per variety were analysed and three images for each kernel were acquired at 5000X magnification in the most representative area of the kernel section. Considering only the images obtained for those kernels presenting a non-crystalline starch disposition, a total of 220 images were selected.

In this study, image analysis was employed to investigate the starch particles morphology. In particular, the algorithm was developed to extract features related both to the particle shape (e.g. area, perimeter, etc.) and to the particle disposition (e.g. number of starch particles and empty spaces among them, also defined as "porosity"). Five manual inputs were defined to enhance the adaptability of the algorithm to other applications. These inputs (or parameters) are involved in different steps of the algorithm and were named P1, P2, P3, P4 and P5. Their purpose is explained in detail in the following paragraphs, and an overview is provided in Table 1 [39].

2.2. Algorithm

2.2.1. Algorithm description

The image analysis algorithm was developed under MATLAB environment (version R2021b, Mathworks, Natick, MA, USA). To better explain how the algorithm works, a supporting flowchart containing all the image analysis steps developed for processing the images can be seen in Fig. 1. The code performs fast analysis of large number of images, but, because of the complexity of the task, it is only able to process one image at a time.

2.2.1.1. Scaling factor. To work with the images using the pixels as measure units instead of meter-based units, a scaling factor needs to be defined. To extract such factor directly from the image, a piece of code based on an OCR function able to read the instrumental information printed in the image itself was developed. Pairing this scaling factor with each image becomes fundamental to refer the automatically measured morphological features back to a meter-based reference. The scaling factor is calculated as the ratio between the extracted meter-based unit number (e.g., 10 μm) and the length of the scale reference bar, expressed as the corresponding number of pixels (Fig. S1 – Supplementary Materials).

2.2.1.2. Size and resolution check. After storing the image's instrumental information, the image sizes and resolutions must be checked: to compare different images, sizes and resolutions need to be made consistent. First, a "resize" function was used to ensure that all analysed images would have the same pixel dimensions, thus being directly comparable among each other. This step was coded to work with inputs either fixed, manually selected or automatically derived from an image chosen as a reference.

2.2.1.3. Brightness adjustment. Once all images have been made consistent, the first image editing step entails brightness correction. FESEM images are acquired in grayscale and, depending on the instrumental set-up, they could globally be darker or brighter, also within the same sample scanned area. MATLAB offers a series of functions able to adjust the brightness to facilitate the comparison among images. Simple functions like *imlocalbrighten* just brighten low-light areas, while more complex ones allow to use the greyscale distribution and the brightness of a selected image as a reference or allow manual tuning the light contrast and the brightness [41].

Table 1
List of the factors (parameters) evaluated by DoE. The tested values, the section of the algorithm where they are employed, and a concise description of the parameters function are reported for each factor.

Factor	Levels	Algorithm section (Fig. 1)	Explanation
P1	9, 18, 27	Greyscale threshold	Threshold percentage for adaptive binarization
P2	1, 13	Image segmentation	Threshold to suppress local minima lower than P2
P3	60, 140	Empty spaces definition	Limit to suppress lighter greys
P4	4, 8	Empty spaces definition	Number of thresholds calculated to estimate empty spaces
P5	5, 14	Objects correction	Definition of the step (%) to distinguish fragments from objects

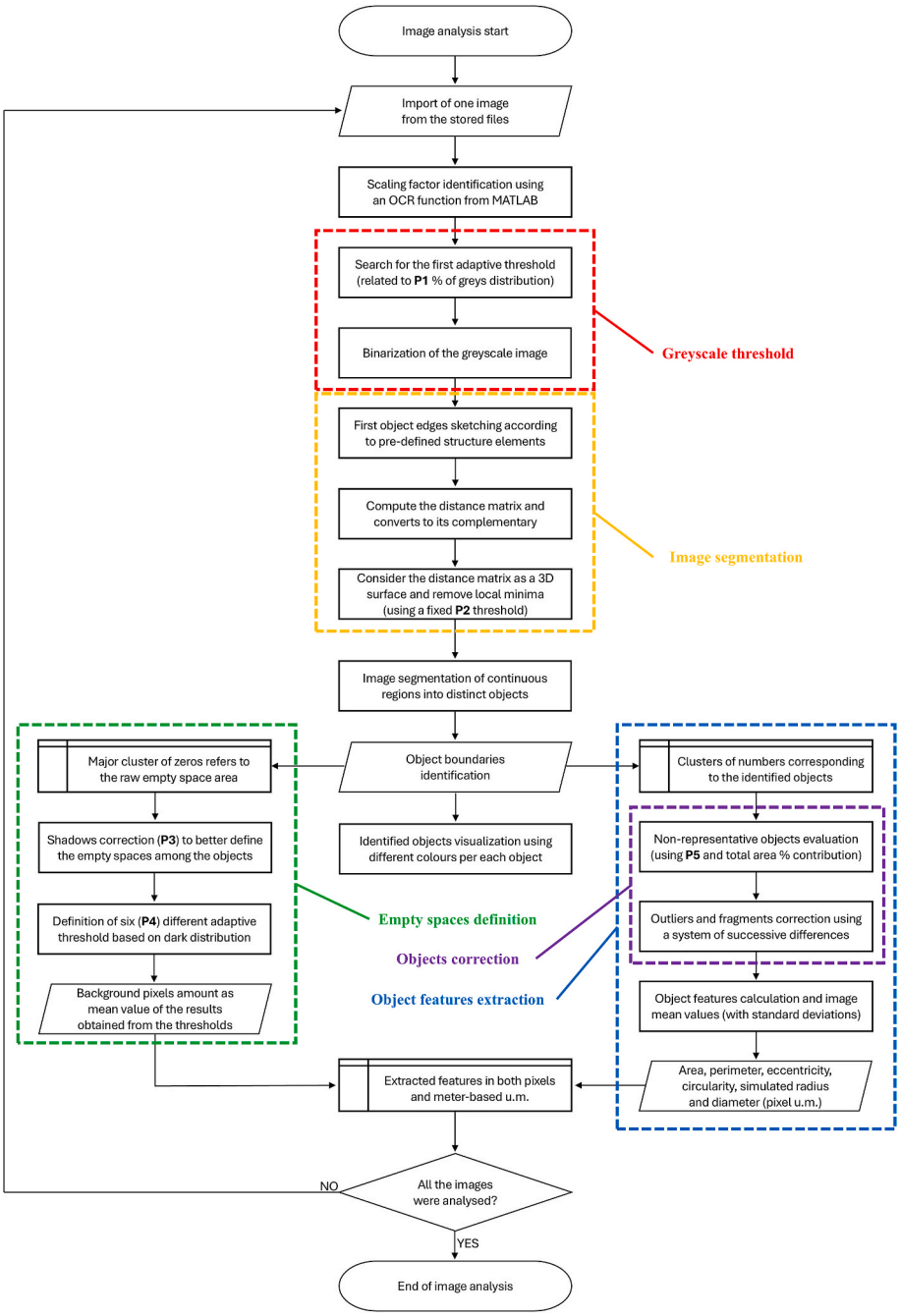

Fig. 1. Flowchart summarizing the algorithm functioning. The red box highlights the definition of the grey distribution threshold. The orange box highlights the image segmentation steps. The green box highlights the definition of the empty spaces. The blue box highlights the object features extraction and contains also the purple box which highlights the optional objects correction step. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2.2.1.4. Image segmentation: the binarization step. Image binarization is a crucial step for both the definition of the empty spaces and the identification of the particles from the image. This step is aimed at dividing the pixels of the image roughly into two sets: one corresponding to the empty spaces and the other corresponding to the particles.

To make the algorithm versatile for a wide variety of applications, an adaptive threshold instead of a fixed one was coded (Fig. 1 – Red box). This threshold is applied to the image grey distribution: all pixels whose value is below the threshold constitute the initial empty space areas of the image (i.e., darker areas not reflecting electrons), while all pixels whose value is above the threshold are identified as the raw areas containing the particles ("raw" because in the next step the measured areas will be better refined through different correction steps). The value of P1 regulates the shade of grey to which the threshold is going to be set. The threshold is adaptive since it is defined as a "percentage of dark grey" and it is applied individually to the grey distributions of each image. In this way, the same percentage threshold value would correspond to a different shade of dark grey from image to image, and therefore potential differences in brightness can be mitigated across the set of images. The threshold computation is based on a cumulative sum over the pixels from the darker end of the distribution, up to the desired threshold value. For our real case of rice images, the threshold value P1 was set to 17 %, but of course for other applications the threshold can be set according to needs.

Because of the different shapes and the tridimensionality of the sample structure, the set of pixels corresponding to the particles needs to be further refined to clearly distinguish and correctly defining the shapes of the particles (Fig. 1 – Yellow box). First, a smoothing function (*activecontour*) allows to sketch the object boundaries using the active contour algorithm, that superimposes the binarized image over the original one. After that, the first objects identification (Fig. S2 – Supplementary Materials) is performed: using the *strel* function, the script creates a structure element with a user-selected shape and applies it with the *imclose* function. The result of this procedure is a new binarized image where a first correction of the areas recognized as empty spaces and as particles is performed. This correction step prepares the image for the following segmentation step. Depending on the type and nature of the objects described by the images, the user can manually choose the approximate object shape to be used by the *strel* function. For example, in the case of round-shaped objects the best results can be obtained by using the '*disk*' option; instead, in the case of more squared ones, the '*rectangle*' or the '*square*' options prove better suited, whereas for linear and elongated ones the '*line*' option can be chosen.

2.2.1.5. Image segmentation: finding the distinct objects. The binarized image contains an initial rough description of the particles of the image, and it needs to be further refined to identify the distinct objects. To do so, a distance transform matrix (based on Euclidean distance) is computed using the *bwdist* function. This function computes, for each pixel, its distance from the nearest "non-assigned" pixel (i.e., the parts in white of Fig. S2, not recognized as objects). As a result, a numerical matrix ("distance matrix") with the same dimensions of the original image is obtained and can be interpreted as follows: the values that are far from zero correspond to the centre of the objects, while the values become closer to zero moving towards the boundaries of the object, and become equal to zero when pixels represent the empty spaces among the objects (Fig. S3a – Supplementary Materials). To clearly identify the edges of the particles, the complement of the previously obtained "distance matrix" is computed by subtracting the distance value of each pixel from the absolute maximum value of the distance transform matrix to enhance contrast and facilitate the next steps of the algorithm (Fig. S3b – Supplementary Materials). The resulting matrix is then interpreted like a three-dimensional surface where bright pixels represent "surface minima" corresponding to objects, while dark pixels represent background regions (zero values). To refine boundaries, the *imhmin* function is applied to the image. This function suppresses all local minima whose depth is lower than a fixed threshold (P2 input) and allow to enhance object separation. After that, the final image segmentation is obtained using the *watershed transform*, which is used to segment continuous regions of interest into distinct objects, again by treating them like a three-dimensional surface (Fig. S4 – Supplementary Materials).

This approach allows to obtain an image where the identified objects are completely distinguished from the empty spaces among them. In addition, with the *labeloverlay* function, it is also possible to visualize the identified objects by highlighting them with different colours in contrast to the black-coloured empty spaces.

2.2.2. Morphological features calculation

The image-segmentation step produces a list of objects in which each object is described by the pixels belonging to the original image. All objects are from now on treated independently by the algorithm to calculate their morphological features. The script allows the user to choose which features the algorithm will extract from the images and, more importantly, also to select either if the interest is focused only on the objects' features or if also the empty spaces among the objects must be computed.

2.2.2.1. Empty spaces identification. The first morphological feature that the algorithm can compute is the area of the space non-assigned as object (i.e., non-object space), that in our rice case study corresponds to the amount of empty spaces among the starch particles. Since the analysed samples are three-dimensional, the images represent a non-planar surface characterized by holes, cavities and superpositions that, all combined, generate shadows. In greyscale images, the darkest shadows can be easily confused with the dark areas corresponding to the "non-object space". To better estimate the real amount of empty space, an algorithmic step involving a set of adaptive thresholds was developed to process the marginal shadows, with a "soft approach" (Fig. 1 – Green box).

By working on the "non-object" pixels greyscale values, an initial selection is performed by removing all pixels lighter than a chosen adaptive threshold (P3 input): those pixels correspond to "lighter shadows". The remaining darker pixels are then processed by a piece of code which calculates the mean value (M) and the standard deviation (S) of their distribution. Then, six different increasing adaptive thresholds are applied to the distribution (the number of thresholds is the value of P4), by addition to the mean value M. The spacing of these increasing thresholds is regulated by a coefficient K ranging from 1 to its the maximum value (which corresponds to P4 input). The thresholds are calculated with Equation (1):

$$thr_{(K)} = M + ((K-1)*0.1*S) \quad (1)$$

For each threshold value a different estimation of the empty space area is obtained. These areas correspond to the number of pixels with greyscale value lower than the corresponding threshold. The final number of pixels identified as the image "non-object area" is the mean of the areas obtained with the different thresholds. The amount of empty spaces is finally expressed both in terms of area (in pixel) and in terms of percentage with respect to the total image area (Fig. S5 – Supplementary Materials).

2.2.2.2. Object features calculation. Before proceeding with the features calculation (Fig. 1 – Blue box), a step of object quality assessment is performed. The aim is to recognize and remove most of the defective or incomplete objects (e.g., objects cut by image edges or partially hidden because of shadows and overlaps), so that the morphological features can be computed considering only the most representative objects, concerning the real situation captured by the images (Fig. 1 – Purple box). Two consecutive steps are applied to remove the non-representative objects: in the first one the percentage contribution of each object's area to the total area is computed and used to separate larger objects from smaller ones; the second step is operated only on the set of small objects defined in the first step, and it is based on the computation of subsequent differences of the particles' dimensions aimed at finding a significant leap, ideally separating the fragments and the smallest objects. More in detail, the first step starts by calculating the area of all objects found by segmentation; then, all computed areas are sorted from the smallest to the largest and the percentage contribution to the total area is computed. The separation between small and large objects is then performed by selecting, from the smallest value up, all objects until the cumulative percentage contributions sum reaches 8 % (P5 input) of the total area. The second step is then performed by calculating the difference between the areas of two consecutive sorted objects, considering only the previously defined set of smallest objects. This set is finally divided into "small objects" and "fragments" (or "defective" objects) and the threshold between them is fixed at the highest computed difference (i.e. the significant leap). All objects whose area is below this threshold are then excluded from the morphological features calculation (Fig. S6 – Supplementary Materials).

Once this correction is performed, each object can be processed to extract the desired morphological features, the most common being the area (number of pixels composing the object), and the perimeter (number of pixels of the object's boundary). Starting from these two

features the values of radius and diameter can then be derived by assuming a perfect circular shape object.

An additional feature that can be requested by the user is eccentricity, a property commonly used for describing elliptical objects: it describes how different the object's shape is from a perfect circle (0 in the case of a perfect circle, and 1 in the case of a segment). Its mathematical definition is described in Equation (2) and corresponds to the ratio between the distance of the focus from the centre and the length of the semimajor axis a.

$$E = \sqrt{a^2 - b^2} \Big/ a \tag{2}$$

Also circularity can be obtained, again related to the object similarity to a perfect circle, but it is based on the area A and perimeter p values as described by Equation (3):

$$C = 4\pi A / p^2 \tag{3}$$

All requested morphological values are calculated and stored for each object in the image, along with the image mean value of each feature and the corresponding standard deviation, as more general information related to the whole image. The last step of the features extraction process consists in the application of the previously identified scaling factor, to convert the features values from pixels unit to a meter-based unit of measurement.

2.2.3. Algorithm's performance evaluation

The proposed algorithm requires the user to input five parameters. To evaluate the influence of these parameters on the whole image processing and features extraction process, a Design of Experiment (DoE, [40]) approach was developed to systematically explore and model their individual and combined effects. In a DoE approach, the five parameters can be treated as "factors" (from P1 to P5). Due to the computational nature of the problem at hand, and to the processing time required by the algorithm, only a few levels can be tested for each factor. Two values were tested for factors P2, P3, P4 and P5; while for P1, which is the parameter involved in the initial image binarization, three values were tested to better evaluate the effect of the adaptive threshold, which can be considered one of the most critical steps in the algorithm.

The selection of the values to test was based on our experience with respect to the algorithm functioning and the real-case image dataset at hand, and at the same time by taking care of setting reasonable values for all the parameters: such considerations are the basis for deploying the tools of Experimental Design. Following the principles of the "full factorial" design, all the possible combinations of the levels tested for the parameters were considered and calculated. This means that all features related to the 220 images of the case study were computed for each experimental point (i.e., the specific combination of five parameters) of the experimental plan (i.e., the whole list of combinations to test).

For each combination different responses were calculated and subsequently modelled. In particular, the algorithm processing time and the identified number of particles were calculated for each of the 48 "experiments" (i.e., the different combinations of factors and levels, represented as "DoE Matrix" in Fig. 2) and for each image. Instead, for the morphological features (empty spaces, area, circularity, eccentricity and perimeter, "Responses" in Fig. 2), the mean values and the corresponding standard deviations were calculated as response variables for each experiment and for each of the 220 images. In addition, to estimate the influence of the "object quality assessment" step described in Section 2.2.2, all the morphological features were calculated before and after this step regulated by parameter P5.

In Table 1 is reported a description of the 5 parameters, with the tested levels and the section of the algorithm where they are applied. The three equally spaced levels tested for the P1 factor were 9 %, 18 % and 27 %. Factor P2 is involved in the image segmentation and two different levels were set: 1 and 13. For factor P3 the two considered levels were 60 and 140. For factor P4 the two levels selected were 4 and 8. Finally, for factor P5 the two considered levels were 5 % and 14 %.

2.3. Data analysis

2.3.1. Principal component analysis (PCA)

An exploratory analysis of the results obtained running the algorithm according to the experimental design (48 combinations of the five factors) was performed by principal component analysis (PCA, [41,42]), using the PLS_Toolbox (version 8.9.2, Eigenvector Research Inc., Manson, WA, USA) software package.

The calculated morphological features (obtained before and after the correction step), the empty spaces values (expressed as a percentage with respect to the total space) and the number of identified objects obtained from the 220 images analysed with all the 48 combinations of factors were inspected using PCA. The data were pre-processed with autoscaling.

2.3.2. ANOVA Simultaneous Component Analysis (ASCA)

ANOVA Simultaneous Component Analysis (ASCA, [43]) was also explored using the PLS_Toolbox (version 8.9.2, Eigenvector Research Inc., Manson, WA, USA) software package. ASCA is a statistical method that can be applied to complex datasets with multiple factors (the parameters, in our case) possibly influencing the variables. It can be used to understand how the different factors contribute to the variance observed in the data and to identify patterns or interactions among these factors.

To better understand and quantify the effect of the parameters on the algorithm performances, ASCA was carried out using a coded DoE matrix (built associating −1 to the lowest value and +1 to the highest

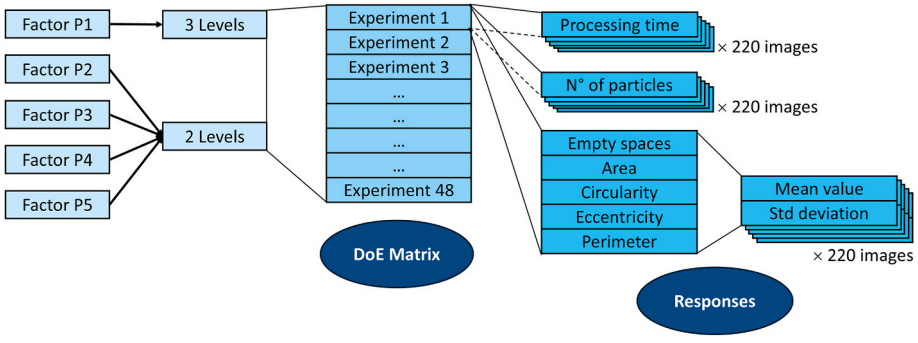

Fig. 2. Experimental design developed to evaluate the 5 manual parameters (from P1 to P5) of the image analysis algorithm. The figure contains a schematic representation of the combinations of factors and levels that generates 48 "experiments" (i.e., the different combination of factors and levels). The different response variables calculated for each experiment are reported on the right side of the figure. For the morphological features, mean values and corresponding standard deviations were calculated for each image.

value) to represent all the possible parameters combinations and using as a response the matrix containing the 11 extracted features (Section 3.2.2). The ASCA results are reported as a table containing the number of principal components used to describe each factor (or the interaction between two factors), the overall effect of each term, and a p-value associated with the significance level of the corresponding result. The number of permutations was set to 100 to calculate the p-value.

2.3.3. Multi linear regression (MLR)

Following the typical approach of Design of Experiments, the combinations of tested parameters were also analysed with multilinear regression (MLR, [44]) to understand the contribution of each factor on the overall processing time of the algorithm. This approach allows to model an experimental domain (i.e., the set of tested combinations of factors) by defining a mathematical function to describe the experimental data. The strong point of MLR is the possibility of describing both the possible effects of the factors and their interactions: this is done by including different terms in the model's mathematical equation. Each term describes a contribution to the modelled response, as related to a factor or an interaction between factors. By inspecting how the response behaves across the experimental domain it is possible to identify interesting sets of experimental conditions, especially in the perspective of response optimization (i.e., maximize or minimize the response). One of the most important outputs to be inspected in an MLR model are the regression coefficients, which are generally displayed as a bar plot in which each bar clearly describes the magnitude and the sign of the different modelled terms (corresponding to individual factors or to pairwise factors' interactions). The coefficients are strictly linked to modelled response, and their global and additive relationship with the response is to be inspected with the response surface plots, which allow to visualize the response relative to the experimental domain.

We decided to model the processing time of the algorithm as a response, to better understanding the bottlenecks of the algorithm workflow.

3. Results and discussion

3.1. Case study: rice starch particles

According to an initial exploration and visualization of the acquired images, the objects to be investigated (starch particles in this case study) were round-shaped, hence the *strel* function (used in the Red Section of the algorithm) was set to '*disk*'. The starch particles were generally clearly visible and distinguishable, so the algorithm was set to calculate all the computable features.

To visually monitor the effectiveness of the different steps of the algorithm, the options were set to show the processed images for visualizing the detected empty spaces (coloured in yellow in Fig. 3a). In addition, to inspect the ability of the algorithm in recognizing the starch particles, the identified objects were marked with different colours (Fig. 3b).

3.2. Evaluation of the algorithm's performance via design of experiment approach

3.2.1. Exploratory analysis with PCA

To explore the performances of the developed algorithms a PCA model was computed on a dataset consisting of the calculated features (for the empty spaces, corresponding in this case study to the measured porosity [39], only the value expressed in percentage was considered) and the number of identified objects. It was decided to exclude the runtime measurements from this exploratory analysis to better focus on the time effect in a second moment. A three components PCA model (cumulative explained variance of 85.35 %) was built on the autoscaled data. The levels of the inspected factors were used to colour the scores plots, thus allowing for direct interpretation of possible trends in the samples' space.

In the scores plot of PC1 (64.22 %, Fig. 4a), coloured according to the two levels of factor P2, a clear separation can be observed. As seen in the loadings plot of Fig. 4b, when P2 is set to its high value (=13, in green) the values of area, eccentricity and perimeter tend to be higher, while circularity and the number of detected particles tend to be lower. The empty space area does not contribute to this separation, confirming that PC1 mostly refers to the "particle shape-related" features.

By colouring the scores of PC2 (12.40 %, Fig. 4c) according to the three levels of factor P1, two combined tendencies become visible. First, the samples seem to be organized into three blocks corresponding to the levels of P1 from low (=9) to high (=27), and each block, on average, shows increasing PC2 values. By exploring the corresponding loadings plot in Fig. 4d, the samples with high P1 value show higher values for eccentricity and empty spaces, and lower values for area and circularity. The second trend is "nested" within each "P1" block, and it becomes more pronounced with the second and third block (Fig. 4c, green and blue samples): within each block four sub-blocks can be spotted.

This peculiar behaviour is potentially connected to another trend that emerges in the scores of PC3 (8.73 %) when coloured according to the two levels of P3 (Fig. 4e). An increasing separation tendency between these two levels can be spotted horizontally, i.e., following the three increasing levels of P1. By looking at the loadings plot (Fig. 4f), the empty space value is the main responsible for this slight separation, while samples with a higher value of eccentricity (corrected) and number of objects (corrected) seem to be responsible for some samples showing particularly low PC3 scores.

The most relevant outcome of this PCA is that the choice of P1 seems to have an influence on the choice of the other parameters. This behaviour can be associated to the fact that P1 is related with one of the first step of the image analysis algorithm (the binarization step). Interestingly, with low P1 value, the level choices regarding the following

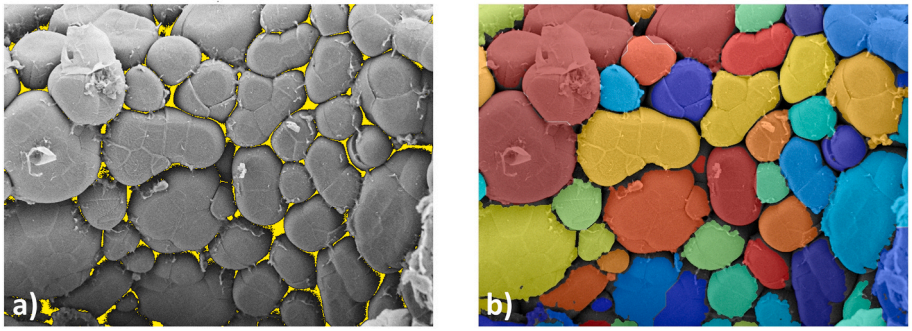

Fig. 3. Images produced by the algorithm during the feature extraction process: (a) a representation of the empty spaces (in yellow) identified within the image, which are interpreted as "sample porosity"; (b) a representation of the starch particles identified by the algorithm and coloured differently. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

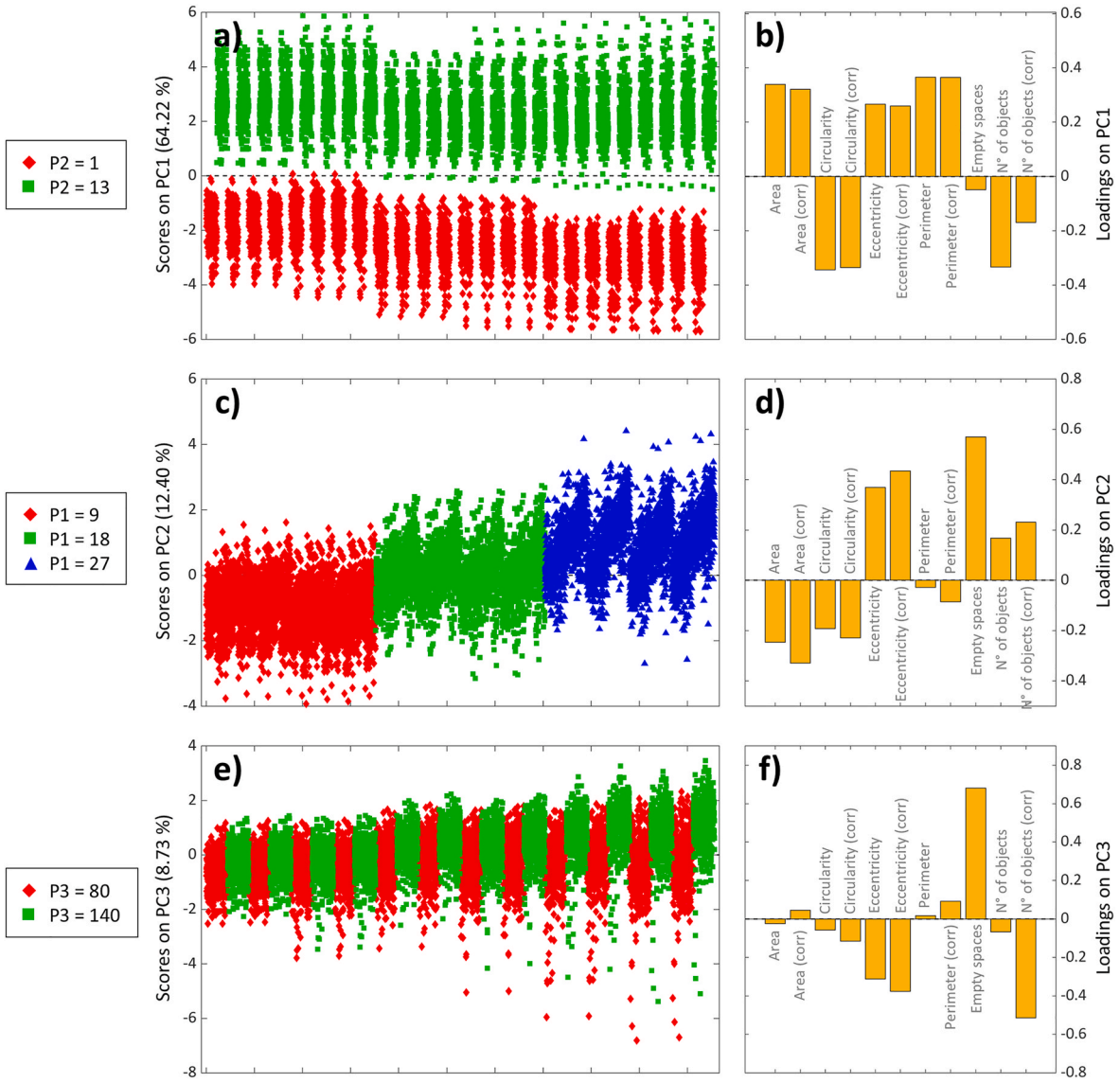

Fig. 4. PCA results. Left column, the scores plots: (a) PC1 coloured according to the two levels of P2 factor, (c) PC2 coloured according to the three levels of P1 factor, and (e) PC3 coloured according to the two levels of P3 factor. Right column, the loadings plots of (b) PC1, (d) PC2, and (f) PC3.

parameters seem to have lower effects on the scores of PC2 and PC3. On the other hand, a higher P1 value results in a larger dispersion of the scores value of PC2 and PC3. To better study the origin of the separations spotted in PCA the ASCA method was applied, as reported in Section 3.2.2.

3.2.2. *Extracted features: evaluation of the parameters' effect with ASCA*

According to the ASCA results reported in Table 2 (and ASCA scores plots reported in Fig. S7 – Supplementary Materials), the largest effect on the features calculation is related to P2 (suppression threshold of local minima in the image segmentation step).

Starting from the large effect of P2 (55.69) the scores plot its ASCA model show a clear separation related to the two levels of the factor (Fig. S7c). This is very coherent with the PCA outcomes reported in Section 3.2.1 (Fig. 4e), where a similar separation was found. Also in this case the separation is related to the particle shape features, while porosity (empty spaces) seems less related to P2, as it can be seen from the loadings plot (Fig. S7d).

The second factor that mostly affects the features calculation is P1 (8.20). Its ASCA scores and loadings plots (Fig. S7a and S7b) make clear that lower P1 values correspond to higher values of area, circularity and

Table 2
Summary of the results obtained using ASCA respect to the extracted features. Five factors and their interaction terms were evaluated. Per each term are reported the number of principal components used, the overall effect (%) and the p-value obtained after a permutation test.

Term	PCs	Effect	P-value
P1	2	8.20	0.01
P2	2	55.69	0.01
P3	1	2.56	0.01
P4	1	0.12	0.01
P5	1	0.44	0.01
P1 × P2	2	1.66	0.01
P1 × P3	1	0.79	0.01
P1 × P4	1	0.02	0.01
P1 × P5	2	0.01	0.01
P2 × P3	1	0.01	0.04
P2 × P4	1	0.00	1.00
P2 × P5	1	0.03	0.01
P3 × P4	1	0.01	0.01
P3 × P5	2	0.00	1.00
P4 × P5	1	0.00	1.00
Residuals	/	30.46	/

perimeter, while, on the other hand, higher P1 values correspond to higher porosity.

ASCA revealed negligible contributions by factors P4 and P5, while a slightly higher effect was found for P3 (2.56). From the loadings plot (Fig. S7f), the only relevant feature related with the separation of the two levels of factor P3 (Fig. S7e) is porosity. Interestingly, the separation between the two levels of P3 is much more pronounced for the samples with the lowest value of P1, while the samples with middle and high P1 values have more overlapped distributions (Fig. S7e).

3.2.3. *Processing time: evaluation of the parameters' effect with MLR*

While performing the tests based on the DoE approach, it was noticed that the total processing time also changed. Therefore, it was decided to explore the effect of the parameters also in relation to the total processing time of the algorithm. In particular, to try to gather more detailed information about the effect's directions of the different parameters, an MLR model was computed using the same DoE coded matrix employed for ASCA, and the overall algorithm processing runtime as a response.

The MLR coefficient plots (Fig. 5a) revealed that the largest contribution on processing time comes from the choice of P1. Interestingly, also P3 and P4 showed significant contributions, as well as the interaction terms P1 × P3 and P1 × P4. In addition, by looking at the bars' orientations in Fig. 5a, it can be noticed that longer processing times generally correspond to higher values of P1, P3 and P4, while P2 has an opposite behaviour but also much reduced relevance. P5 seems to have no effect on the processing time, strengthening our previous findings.

From these general considerations about the regression coefficients, it is also important to consider that different factors interactions and the quadratic term of P1 appear to be relevant, so the response surfaces of the MLR model were explored to globally visualize the contribution of all these coefficients together (Fig. 5). Given that only two factors at a time can be selected for visualizing the response surfaces based on the regression coefficients plot (Fig. 5a) it was decided to evaluate only P1, P3 and P4 while fixing the value of P2 at its high level (=13.0) and the value of P5 at its central level (= average of the two tested values). By exploring the response surfaces of P1 and P3 (Fig. 5b), it becomes apparent that when P1 (i.e., the factor with the strongest effect) is set to a low value, if the value of P3 is set to higher values, shorter processing time can be expected, due to the strong P1 × P3 interaction term. On the contrary, from the response surface of P1 and P4 (Fig. 5c), it is clear that the value set for P4 has a low impact on the processing time, but that setting a low value can be generally preferred. These two pieces of evidence are also confirmed by the response surface of P3 and P4 while setting P1 to its lowest level (Fig. 5d). In this plot, even if the parameters' contribution is small compared to the P1 effect, the lowest processing time corresponds to a high value for P3 and a low value for P4. Interestingly, the second lowest region of the surface corresponds to setting a low value of P3 and a high value of P4, while keeping both P3 and P4 higher or lower results in longer processing times.

To resume, according to the results obtained from ASCA (Section 3.2.2) and MLR (current Section), P1 is the only parameter for which a strong effect both on the features calculation and on the algorithm running time was found. Since P1 is the parameter involved in the binarization step, this first step strongly affects all the following ones. The values chosen for P3 (lighter grey suppression) and P4 (number of thresholds employed for empty spaces definition) seem to slightly affect the values of the calculated features, but they have a remarkable effect on the overall processing time. On the contrary, the choice of P2 has a strong effect on the computed features, but no effect on the total processing time, meaning that the value set for the threshold used to suppress the local minima only affects the features calculation steps without influencing the algorithm running time. Finally, the effect of P5 appears to be negligible both on time and features extraction. This is not unexpected, as P5 regulates the percentage of fragments to be left out while selecting the objects based on their size distribution, which is a simple

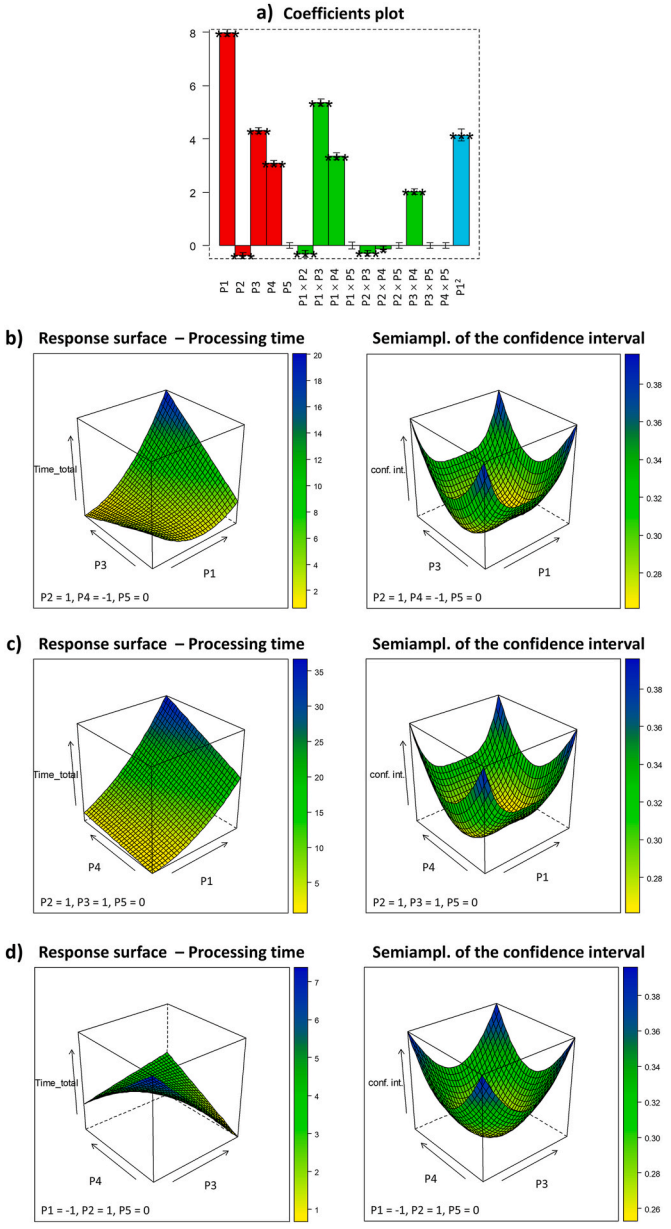

Fig. 5. Results from the MLR model with total processing time as the response. (a) Plot of the regression coefficients with the linear terms in red, the interaction terms in green and the P1 quadratic term in cyan. (b) On the left the response surface of the processing time as related to P1 and P2 and on the right the corresponding confidence interval. (c) On the left the response surface of the processing time as related to P1 and P4 and on the right the corresponding confidence interval. (d) On the left the response surface of the processing time as related to P3 and P4 and on the right the corresponding confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

computation to be performed.

4. Conclusions

A new versatile method to perform image analysis was developed with the aim of extracting a series of morphological features related to similar-shaped objects from greyscale images. To evaluate the performances of this new method, the developed algorithm was also tested on a real case study consisting of more than 200 greyscale FESEM images acquired on different rice kernels. As a result, the different morphological features (i.e., area, perimeter, eccentricity, circularity, radius,

diameter), related to the starch particles that define the internal kernel structure, were obtained. In addition, also the total area of the empty spaces among the particles was computed. Eventually, to evaluate the effect of the choice of the input parameters on the algorithm result, an experimental design approach was applied, and the resulting outcomes were thoroughly analysed using PCA, ASCA and MLR, all of which yielded consistent results relative to the effects of the five assessed parameters.

The obtained results highlighted the ability of the algorithm to recognize similar-shaped objects (particles) and to automatically calculate their related morphological features. The opportunity of realizing high-throughput image analysis and the possibility of varying the input options depending on the different targets, combined with the adaptability of many crucial steps of the developed algorithm, make this new image analysis method extremely versatile for diverse types of images (potentially not only FESEM acquisitions) and suitable for different research fields.

CRediT authorship contribution statement

Mattia Sozzi: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nicola Cavallini:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alessandro Chiadò:** Investigation. **Gentian Gavoci:** Methodology, Investigation. **Enrico Cantaluppi:** Resources, Project administration, Conceptualization. **Filip Haxhari:** Resources, Project administration, Funding acquisition, Conceptualization. **Francesco Savorani:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve the readability of some sections of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of Competing interest

There is no conflict of interest.

Acknowledgments

Ente Nazionale Risi is acknowledged for providing the rice samples and for starting the project that led to this publication. Francesca Stella is warmly thanked for editing the graphical abstract.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2025.105415.

Data availability

The authors do not have permission to share data.

References

[1] M. Mohd Ali, N. Hashim, S.A. Aziz, O. Lasekan, Emerging non-destructive thermal imaging technique coupled with chemometrics on quality and safety inspection in food and agriculture, Trends Food Sci. Technol. 105 (2020) 176–185, https://doi.org/10.1016/j.tifs.2020.09.003.

[2] L. Zhang, S. Shao, Image-based machine learning for materials science, J. Appl. Phys. 132 (2022), https://doi.org/10.1063/5.0087381.

[3] M. Ragone, R. Shahabazian-Yassar, F. Mashayek, V. Yurkiv, Deep learning modeling in microscopy imaging: a review of materials science applications, Prog. Mater. Sci. 138 (2023), https://doi.org/10.1016/j.pmatsci.2023.101165.

[4] M. Ge, F. Su, Z. Zhao, D. Su, Deep learning analysis on microscopic imaging in materials science, Mater Today Nano 11 (2020), https://doi.org/10.1016/j.mtnano.2020.100087.

[5] J.M. Prats-Montalbán, A. de Juan, A. Ferrer, Multivariate image analysis: a review with applications, Chemometr. Intell. Lab. Syst. 107 (2011) 1–23, https://doi.org/10.1016/j.chemolab.2011.03.002.

[6] L.F. Capitán-Vallvey, N. López-Ruiz, A. Martínez-Olmos, M.M. Erenas, A.J. Palma, Recent developments in computer vision-based analytical chemistry: a tutorial review, Anal. Chim. Acta 899 (2015) 23–56, https://doi.org/10.1016/j.aca.2015.10.009.

[7] M. Rezazadeh, S. Seidi, M. Lid, S. Pedersen-Bjergaard, Y. Yamini, The modern role of smartphones in analytical chemistry, TrAC, Trends Anal. Chem. 118 (2019) 548–555, https://doi.org/10.1016/j.trac.2019.06.019.

[8] H.L. Zhai, B.Q. Li, J. Chen, X. Wang, M.L. Xu, J.J. Liu, S.H. Lu, Chemical image moments and their applications, TrAC, Trends Anal. Chem. 103 (2018) 119–125, https://doi.org/10.1016/j.trac.2018.03.017.

[9] J. Zhang, C. Li, M.M. Rahaman, Y. Yao, P. Ma, J. Zhang, X. Zhao, T. Jiang, M. Grzegorzek, A comprehensive review of image analysis methods for microorganism counting: from classical image processing to deep learning approaches, Artif. Intell. Rev. 55 (2022) 2875–2944, https://doi.org/10.1007/s10462-021-10082-4.

[10] A.A. Appel, M.A. Anastasio, J.C. Larson, E.M. Brey, Imaging challenges in biomaterials and tissue engineering, Biomaterials 34 (2013) 6615–6630, https://doi.org/10.1016/j.biomaterials.2013.05.033.

[11] D. Li, C. Li, Y. Yao, M. Li, L. Liu, Modern imaging techniques in plant nutrition analysis: a review, Comput. Electron. Agric. 174 (2020), https://doi.org/10.1016/j.compag.2020.105459.

[12] S.K. Jung, A review of image analysis in biochemical engineering, Biotechnol. Bioproc. Eng. 24 (2019) 65–75, https://doi.org/10.1007/s12257-018-0372-8.

[13] T.W. Nattkemper, Multivariate image analysis in biomedicine, J Biomed Inform 37 (2004) 380–391, https://doi.org/10.1016/j.jbi.2004.07.010.

[14] Y. Skaf, R. Laubenbacher, Topological data analysis in biomedicine: a review, J Biomed Inform 130 (2022), https://doi.org/10.1016/j.jbi.2022.104082.

[15] R. Murphy, A. Turcott, L. Banuelos, E. Dowey, B. Goodwin, K.O.H. Cardinal, SIMPoly: a matlab-based image analysis tool to measure electrospun polymer scaffold fiber diameter, Tissue Eng Part C Methods 26 (2020) 628–636, https://doi.org/10.1089/ten.tec.2020.0304.

[16] A.Q. Lu, S.Z. Zhang, Q. Tian, Matlab image processing technique and application in pore structure characterization of hardened cement pastes, Adv Mat Res (2013) 1374–1379. https://doi.org/10.4028/www.scientific.net/AMR.785-786.1374.

[17] J.P. Kim, U. Nowostawska, K.A. Hunter, Comparison of particle size spectrum determination from images made using manual and automated image analysis, Environ. Technol. 29 (2008) 1191–1198, https://doi.org/10.1080/09593330802217773.

[18] S. Maaß, J. Rojahn, R. Hänsch, M. Kraume, Automated drop detection using image analysis for online particle size monitoring in multiphase systems, Comput. Chem. Eng. 45 (2012) 27–37, https://doi.org/10.1016/j.compchemeng.2012.05.014.

[19] R.S. Edwin, M. Mushthofa, E. Gruyaert, N. De Belie, Quantitative analysis on porosity of reactive powder concrete based on automated analysis of back-scattered-electron images, Cem. Concr. Compos. 96 (2019) 1–10, https://doi.org/10.1016/j.cemconcomp.2018.10.019.

[20] R. Ziel, A. Haus, A. Tulke, Quantification of the pore size distribution (porosity profiles) in microfiltration membranes by SEM, TEM and computer image analysis, J Memb Sci 323 (2008) 241–246, https://doi.org/10.1016/j.memsci.2008.05.057.

[21] S. Diamond, M.E. Leeman, PORE SIZE DISTRIBUTIONS IN HARDENED CEMENT PASTE BY SEM IMAGE ANALYSIS, n.d.

[22] H. Kim, J. Han, T.Y.J. Han, Machine vision-driven automatic recognition of particle size and morphology in SEM images, Nanoscale 12 (2020) 19461–19469, https://doi.org/10.1039/d0nr04140h.

[23] N. Prakongkep, A. Suddhiprakarn, I. Kheoruenromne, R.J. Gilkes, SEM image analysis for characterization of sand grains in Thai paddy soils, Geoderma 156 (2010) 20–31, https://doi.org/10.1016/j.geoderma.2010.01.003.

[24] B. Gaël, T. Christelle, E. Gilles, G. Sandrine, S.F. Tristan, Determination of the proportion of anhydrous cement using SEM image analysis, Constr. Build. Mater. 126 (2016) 157–164, https://doi.org/10.1016/j.conbuildmat.2016.09.037.

[25] G.N. Barrera, G. Calderón-Domínguez, J. Chanona-Pérez, G.F. Gutiérrez-López, A. E. León, P.D. Ribotta, Evaluation of the mechanical damage on wheat starch granules by SEM, ESEM, AFM and texture image analysis, Carbohydr. Polym. 98 (2013) 1449–1457, https://doi.org/10.1016/j.carbpol.2013.07.056.

[26] M. Shanmuga Priya, P. Divya, R. Rajalakshmi, A review status on characterization and electrochemical behaviour of biomass derived carbon materials for energy storage supercapacitors, Sustain Chem Pharm 16 (2020), https://doi.org/10.1016/j.scp.2020.100243.

[27] M.K.A. Mohammed, M.R. Mohammad, M.S. Jabir, D.S. Ahmed, Functionalization, characterization, and antibacterial activity of single wall and multi wall carbon nanotubes, in: IOP Conf Ser Mater Sci Eng, Institute of Physics Publishing, 2020, https://doi.org/10.1088/1757-899X/757/1/012028.

[28] L.B. Fiuza de Oliveira, G. Paim de Carvalho, G. Amarante de Almeida Mussi da Silva, A.L.A. Guimarães, Plant anatomy: history and future directions Microscopy in food analysis: a review, Rodriguesia 75 (2024), https://doi.org/10.1590/2175-7860202475030.

[29] J. Tan, B.M. Balasubramanian, Particle size measurements and scanning electron microscopy (SEM) of cocoa particles refined/conched by conical and cylindrical roller stone melangers, J. Food Eng. 212 (2017) 146–153, https://doi.org/10.1016/j.jfoodeng.2017.05.033.
[30] R. Peters, G. ten Dam, H. Bouwmeester, H. Helsper, G. Allmaier, F. vd Kammer, R. Ramsch, C. Solans, M. Tomaniová, J. Hajslova, S. Weigel, Identification and characterization of organic nanoparticles in food, TrAC, Trends Anal. Chem. 30 (2011) 100–112, https://doi.org/10.1016/j.trac.2010.10.004.
[31] V. Sharma, A. Bhardwaj, Scanning electron microscopy (SEM) in food quality evaluation, in: Evaluation Technologies for Food Quality, Elsevier, 2019, pp. 743–761, https://doi.org/10.1016/B978-0-12-814217-2.00029-9.
[32] A. Baiano, Applications of hyperspectral imaging for quality assessment of liquid based and semi-liquid food products: a review, J. Food Eng. 214 (2017) 10–15, https://doi.org/10.1016/j.jfoodeng.2017.06.012.
[33] F. Pieniazek, V. Messina, Texture analysis of freeze dried banana applying scanning electron microscopy combined with image analysis techniques, ETP International Journal of Food Engineering (2018) 127–131, https://doi.org/10.18178/ijfe.4.2.127-131.
[34] J. Tan, H. Zhang, X. Gao, SEM image processing for food structure analysis, J. Texture Stud. 28 (1997) 657–672, https://doi.org/10.1111/j.1745-4603.1997.tb00145.x.
[35] S.A. Mohammed, A.L. Ralescu, Insights into image understanding: segmentation methods for object recognition and scene classification, Algorithms 17 (2024), https://doi.org/10.3390/a17050189.
[36] M. Sirohi, M. Lall, S. Yenishetti, L. Panat, A. Kumar, Development of a Machine learning image segmentation-based algorithm for the determination of the adequacy of Gram-stained sputum smear images, Med. J. Armed Forces India 78 (2022) 339–344, https://doi.org/10.1016/j.mjafi.2021.09.012.
[37] E. Shonkoff, K.C. Cara, X. Pei, M. Chung, S. Kamath, K. Panetta, E. Hennessy, AI-based digital image dietary assessment methods compared to humans and ground truth: a systematic review, Ann. Med. 55 (2023), https://doi.org/10.1080/07853890.2023.2273497.
[38] P. Chotwanvirat, A. Prachansuwan, P. Sridonpai, W. Kriengsinyos, Advancements in using AI for dietary assessment based on food images: scoping review, J. Med. Internet Res. 26 (2024) e51432, https://doi.org/10.2196/51432.
[39] F. Haxhari, F. Savorani, M. Rondanelli, E. Cantaluppi, L. Campanini, E. Magnani, C. Simonelli, G. Gavoci, A. Chiadò, M. Sozzi, N. Cavallini, A. Chiodoni, C. Gasparri, G.C. Barrile, A. Cavioni, F. Mansueto, G. Mazzola, A. Moroni, Z. Patelli, M. Pirola, A. Tartara, D. Guido, S. Perna, R. Magnaghi, Endosperm structure and glycemic index of japonica Italian rice varieties, Front. Plant Sci. 14 (2023), https://doi.org/10.3389/fpls.2023.1303771.
[40] R. Leardi, Experimental design in chemistry: a tutorial, Anal. Chim. Acta 652 (2009) 161–172, https://doi.org/10.1016/j.aca.2009.06.015.
[41] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (2014) 2812–2831, https://doi.org/10.1039/c3ay41907j.
[42] I.T. Jollife, J. Cadima, Principal component analysis: a review and recent developments, Phil. Trans. Math. Phys. Eng. Sci. 374 (2016), https://doi.org/10.1098/rsta.2015.0202.
[43] C. Bertinetto, J. Engel, J. Jansen, ANOVA simultaneous component analysis: a tutorial review, Anal. Chim. Acta X (2020) 6, https://doi.org/10.1016/j.acax.2020.100061.
[44] M. Sergent, D. Mathieu, R. Phan-Tan-Luu, G. Drava, Chemometrics and Intelligent Laboratory Systems Tutorial Correct and Incorrect Use of Multilinear Regression, 1995.