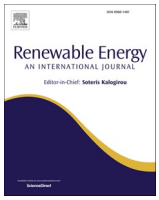
Renewable Energy 257 (2026) 124925

Contents lists available at ScienceDirect

Renewable Energy

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



A novel image processing and machine learning approach for wood pellet size classification using shadow analysis

Thomas Gasperini [a,*], Elena Leoni [a], Nicolò Bartolini [b], Giacomo Ciccone [a], Giuseppe Toscano [a], Carmine De Francesco [a]

[a] Department of Agricultural, Food and Environmental Sciences, Marche Polytechnic University, Via Brecce Bianche, 60131, Ancona, Italy
[b] Department of Information Engineering, Marche Polytechnic University, Via Brecce Bianche, 60131, Ancona, Italy

ARTICLE INFO

Keywords:
Biofuels
Dimensional classification
Image processing
Shadow analysis
Supervised learning

ABSTRACT

The growing importance of wood pellets in renewable energy highlights the need for efficient, scalable methods of quality assessment. Current standards rely on manual caliper measurements, which are time-consuming and poorly capture variability across samples. This study introduces a patent-pending imaging approach that classifies pellet dimensions from shadow features. A prototype system was developed combining controlled lighting, a camera, and computational processing. Shadow characteristics were analyzed statistically and linked to pellet dimensions using machine learning models. Results show that illumination geometry strongly influences classification, with the best performance reaching 71 % accuracy. These findings demonstrate the feasibility of shadow-based imaging as a rapid, non-invasive alternative to manual measurement, with promising applications in pellet production, storage, and combustion systems.

1. Introduction

The growing international and European focus on renewable energy [1,2] has driven to the expansion of the renewable energy and solid biofuel market [3]. Within this context, wood pellets have emerged as a key renewable fuel alternative, supporting fuel switching from fossil sources and contributing directly to decarbonization targets [4]. This trend is reflected in the 80 % rise in wood pellet production across Europe between 2014 and 2023 [5,6].

To ensure sustainable deployment, regulatory standards have been established to define and regulate pellet quality classes [7] based on chemical, dimensional, physical, and energetical properties, along with standardized analytical methods for their assessment [8].

Among these, pellet geometry and especially pellet length, plays a critical role in practical applications. Geometry affects not only handling and storage behavior but also combustion dynamics and emissions. For instance, pellet length influences fuel mass flow in feeding systems, combustion air–fuel ratios, and consequently the emission of pollutants such as carbon monoxide (CO) and particulate matter. Wöhler et al. [9] found a correlation between pellet length and emissions of CO and particulate matter. Likewise, Mack et al. [10] demonstrated that longer pellets reduce fuel mass flow, increase lambda values, and ultimately lower combustion efficiency. Thus, the sustainability of pellets is not only determined by their origin and carbon balance but also by their physical characteristics, which directly influence efficiency, storage, transport, and ultimately emissions performance [11]. From a sustainability perspective, optimizing pellet length is therefore critical, as inefficient combustion or elevated emissions can offset the environmental advantages of renewable fuels [12].

Despite this evidence, geometrical parameters are often overlooked in comparison with more familiar factors such as moisture content or calorific value. Technical standard ISO 17829 [13] provides methodology for assessing pellet geometry, but its reliance on manual caliper measurements from a small sample introduces risks of underrepresenting batch variability. This limitation could undermine the accuracy of quality assessments and, consequently, the reliability of performance predictions. Such restrictions highlight a gap between laboratory testing protocols and the actual operational conditions of pellet-based heating systems, where variability in geometry directly impacts efficiency and emissions. To overcome this limitation, researchers have explored alternative approaches such as image processing-based techniques. These methods enable faster and more

* Corresponding author.
E-mail addresses: t.gasperini@univpm.it (T. Gasperini), e.leoni@univpm.it (E. Leoni), s1118768@studenti.univpm.it (N. Bartolini), s1116352@studenti.univpm.it (G. Ciccone), g.toscano@univpm.it (G. Toscano), c.defrancesco@univpm.it (C. De Francesco).

https://doi.org/10.1016/j.renene.2025.124925
Received 13 March 2025; Received in revised form 25 September 2025; Accepted 1 December 2025
Available online 2 December 2025

representative pellet length determination by measuring a significantly larger number of pellets in a shorter time [14,15]. However, these methods typically depend on segmentation techniques that require individually separated pellets [16] limiting their application to controlled laboratory settings. More advanced strategies, such as the free-fall analysis proposed by Jägers et al. [17], address some of these constraints but still require large vertical setups. Yet, both conventional and alternative methods remain largely unsuitable for real-world industrial and domestic scenarios, such as bulk storage in silos or heating system tanks [18,19]. In such environments, pellets are densely stacked, partially obscured, or variably oriented, complicating measurement and increasing the risk of misclassification (e.g., diameter mistaken for length). Furthermore, variable pellet orientation, such as longitudinal alignment, can cause optical systems to misclassify diameter as length, further distorting dimensional analysis [20]. Precise knowledge of pellet length distributions in these systems is crucial for optimizing combustion. Real-time monitoring would allow automated control units to adjust feed rates and airflow, compensating for pellet size variability to maintain efficiency and reduce emissions [9]. Moreover, considering the increasing role of pellets in district heating networks and industrial-scale applications, reliable dimensional monitoring could also support predictive maintenance and reduce downtime in automated feeding systems [9,10]. Thus, monitoring of pellet dimensions within storage systems could facilitate dynamic optimization of combustion processes [21].

Comparable challenges in stacked material analysis have been addressed in other fields (e.g., aggregates, granules), where imaging, sensor-based, or machine learning techniques are employed [22,23], however, these solutions are either computationally intensive or not adapted to the specific geometric challenges posed by pellets.

The present study presents a novel low-cost image processing method for the classification of objects by size within stacked or piled configurations, with a patent application pending at the time of writing. The proposed approach addresses the limitations of ISO 17829 and conventional image analysis techniques by enabling dimensional assessment without the need for individually separated or stationary pellets. It exploits the correlation between object dimensions and the geometric properties of shadows generated under lateral illumination [21,24]. The system requires only a camera, a light source, and a computing unit, making it adaptable for installation in various settings, including domestic heating system tanks and bulk storage environments. By enabling non-invasive, real-time monitoring of pellet length and other geometric parameters, the method has the potential to support dynamic optimization of combustion systems, ensuring more efficient utilization of pellets as a renewable fuel for decarbonization. This contribution is therefore not only methodological but also strategic, as it addresses a key barrier to scaling up pellet use as a mainstream renewable energy carrier in line with European decarbonization pathways.

2. Materials & methods

The following sections provide a detailed description of the experimental plan. Specifically, sample preparation (Section 2.1), standard dimensional characterization (Section 2.2), shadow data acquisition using a prototypal system (Section 2.3), and data analysis with

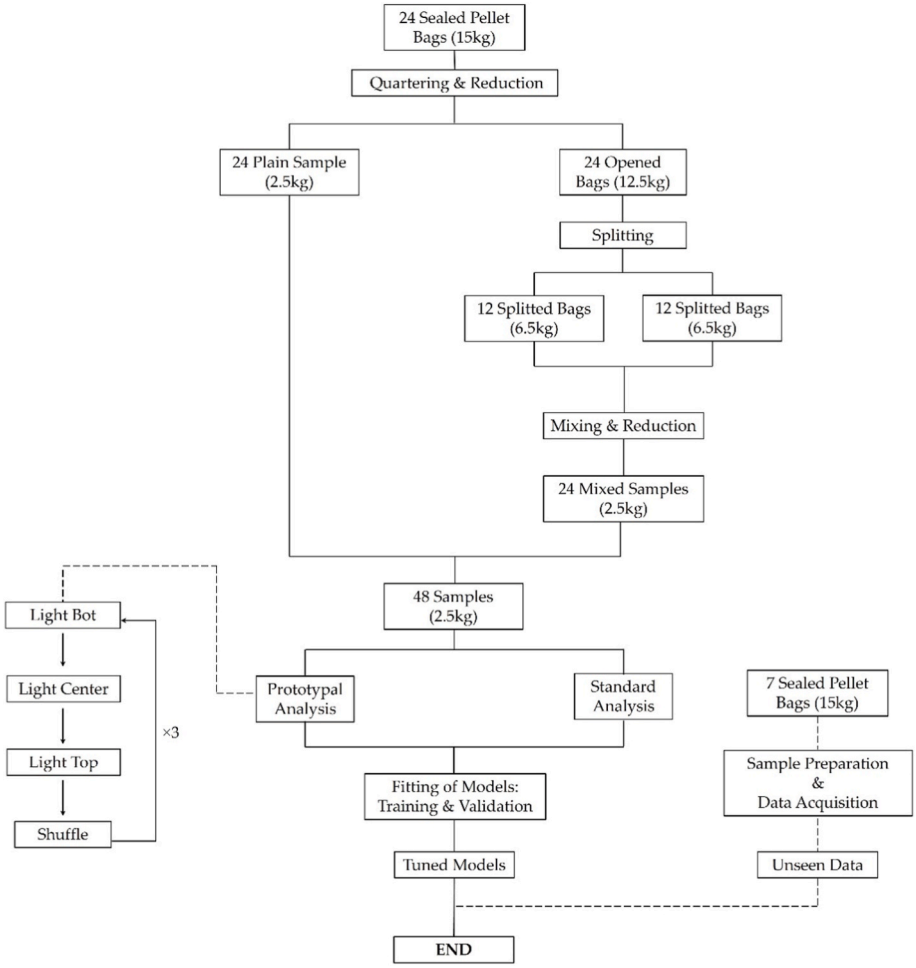


Fig. 1. Overview of the experimental plan.

subsequent model development (Section 2.4) are presented. A schematic representation of the experimental plan is reported in Fig. 1.

2.1. Sample collection and preparation

2.1.1. Training & validation samples

A total of 24 bags of commercial wood pellets, each weighing 15 kg, were acquired from different European producers. From each bag, a representative sample of approximately 2.5 kg was obtained through homogenization and sample reduction, in accordance with the methodology defined by ISO 14780.

To increase the number of available samples, the remaining material from each bag was further divided into two sub-portions. These sub-portions were subsequently mixed to obtain 24 new bags containing mixed pellets. The mixed bags were then homogenized and reduced following the same standardized procedure, resulting in an additional 24 samples of approximately 2.5 kg each. In total, 48 samples of 2.5 kg were obtained. All samples underwent standard dimensional characterization using a caliper, followed by analysis with the prototypal system. The data collected from these samples were used for model training and validation and are hereafter referred to as "training-validation" samples.

2.1.2. Testing samples

Additionally, seven bags of pellets, each weighing 15 kg, were acquired and processed using the same procedures applied to the training-validation samples. The bags were homogenized and reduced to obtain seven samples of approximately 2.5 kg each. These samples were subsequently measured using a caliper and analyzed with the prototypal system. The data obtained from these samples constituted the "Unseen" dataset, which was used to evaluate the accuracy of the fitted models.

2.2. Standard dimensional analysis

Prior to dimensional analysis, each sample was sieved in accordance with ISO 18846 to remove the fine fraction and any material not meeting the pellet definition specified in ISO 17225–2. From each sieved sample, a random 500 g subset of pellets was selected and measured using a Borletti digital caliper with a resolution of 0.01 mm. Following measurement, each pellet was returned to its respective sample. Descriptive statistical analysis was then carried out on the collected data to derive a set of length-based features (please refer to the Supplementary Material for the table listing the variables and their descriptions).

2.3. Prototypal system

The prototypal system is based on a patent-pending principle that hypothesizes that objects of varying sizes, when piled or stacked, generate distinct shadow patterns when exposed to incident or lateral light sources. The system consists of a darkroom, three lamps, a sample tray, and an acquisition unit. The acquisition unit includes a Raspberry Pi computer, a camera, and image processing software developed in Python. The use of multiple lamps is intended to assess whether variations in illumination angle influence the results [24,25]. The image processing software automatically binarizes the captured images, removing the foreground (pellets) and isolating the shadows. Upon completion of this process, a set of features describing the shadows (hereafter referred to as "shadow-based" features) is extracted, and a data file is generated automatically. An in-depth description of the prototype and its functioning is provided in the following paragraphs.

2.3.1. Dark room, lamps and tray

The darkroom (Fig. 2, point A) was constructed using a cardboard box with dimensions of 59 × 41 × 41.5 cm. The bottom panel of the box was removed to allow easy positioning and movement of the sample tray, which also serves as the base of the prototype. To minimize light scattering from the lamps, all interior surfaces of the box, including the removed bottom panel, were coated with acrylic absolute-black paint.

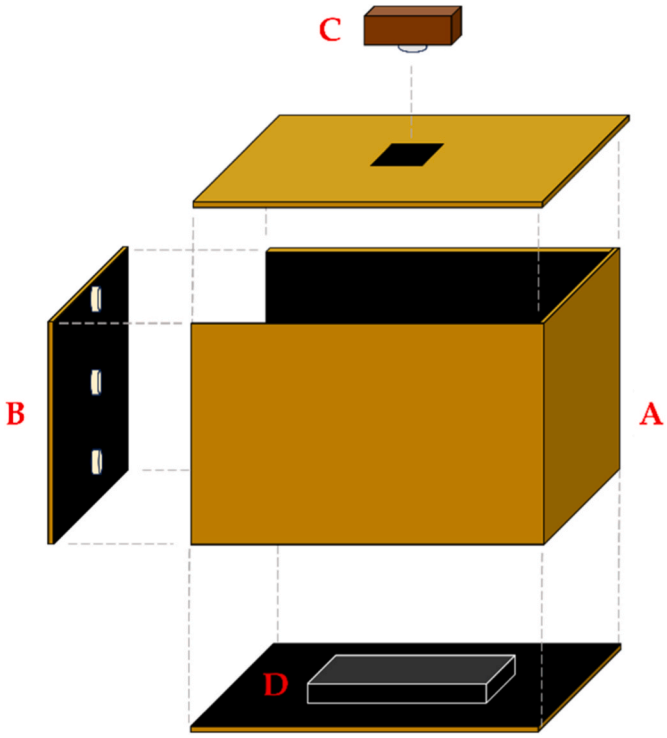


Fig. 2. Exploded diagram of the prototypal system. A) Darkroom; B) Upper, Central, Lower lamps; C) Acquisition system; D) Sample placement tray on the base of the prototype.

Three low-cost commercial LED lamps were installed on the left interior wall of the darkroom (Fig. 2, point B), positioned just outside the field of view of the camera. Each lamp measures 3 × 3 cm, with a luminous flux of 74 lumens and a color temperature of 2700 K. The lamps were vertically aligned along the wall, starting 9 cm above the base, with a spacing of 9 cm between each unit. This configuration was established through preliminary testing aimed at maximizing differences in shadow formation [21].

A 2 × 2 cm aperture was cut in the center of the upper side of the box to accommodate the camera-equipped Raspberry Pi system (Fig. 2, point C). The opening was designed to provide an unobstructed field of view while minimizing intrusion of ambient light.

The sample tray (31 × 22.5 × 3 cm) was fabricated from cardboard and coated with absolute-black paint (Fig. 2, point D). Its dimensions allow for sufficient pellets to form a multi-layer pile, which is then leveled to create a flat surface perpendicular to the camera across all points.

2.3.2. Acquisition system

The acquisition system consists of both hardware and software components. The hardware setup includes a Raspberry Pi 4B with 8 GB of RAM as the computational unit, a Raspberry Pi Camera Module 3 for imaging, an Osoyoo 3.5" touchscreen for user interaction, and a 5V Fan for cooling. To enhance portability and ease of handling, the entire system has been assembled inside a 3D-printed chassis.

After preliminary testing [21], the following camera settings were selected to ensure the acquisition of the sharpest possible images: saturation set to 1, contrast set to 1, sharpness set to 2, brightness set to 0, noise reduction is turned off, gain set at 2.50, red gain set at 1.27, blue gain set at 3.37, and exposure set to 5074 μs.

The software was developed in Python 3.11 utilizing the OpenCV library for image processing [26]. It performs image binarization to highlight, count, and measure shadows resulting from the interaction

between light and pellets. The binarization threshold was determined using Otsu's method [27] and set at 35.

2.3.3. Acquisition procedure

The analytical procedure begins by filling the tray with the sample, which is then leveled to minimize shadow formation due to variations in the height of the stacked material. The tray is subsequently positioned at the center of the base, after which the darkroom is placed over it, effectively enclosing the tray within the system (Fig. 3).

The acquisition system is then placed upon the darkroom and the camera is aligned with the 2 × 2 hole. The dimensions of the darkroom and the positioning of the various components ensure that the field of view of the camera is limited to the sample in the tray.

The imaging process begins with the activation of the lower lamp, under which the first image is captured. The lamp is then switched off, and the central lamp is turned on to acquire a second image. Subsequently, the central lamp is turned off and the upper lamp is activated to capture the final image. After the upper lamp is switched off, the darkroom is lifted, the sample is shuffled and leveled again, and the entire imaging sequence is repeated under the three lighting conditions.

This sequence continues until three replicates are completed for each lighting condition. A total of 144 images (48 samples × 3 replicates) have been acquired for the training-validation samples under each light condition, while 21 images (7 samples × 3 replicates) were obtained for the Unseen (testing) samples.

2.3.4. Acquired data

Following image acquisition and binarization, the software automatically generated a dataset containing information of the detected shadows. Each dataset included a set of shadow-based features (please refer to the Supplementary Material for the table listing the variables and their descriptions). Data obtained under different lighting conditions were stored and processed independently.

2.4. Data analysis and machine learning

2.4.1. Explorative analyses and feature extraction

The training-validation dataset, which includes both length-based and shadow-based features, have been merged into three comprehensive datasets, each corresponding to one of the three lighting conditions used during shadow acquisition. Hereafter, these datasets will be referred to as Top_{Data} (upper lighting), $Center_{Data}$ (central lighting), and Bot_{Data} (bottom lighting). Each dataset contains 144 rows (each corresponding to an image acquired during the prototypal analysis) and 24 columns (the full range of shadow-based and length-based features).

The datasets have been assessed through descriptive statistics, Pearson analysis, and Principal Component Analysis (PCA). These analyses facilitated the identification of the most significant features. Following feature selection, PCA was performed again to evaluate whether the removal of redundant variables could enhance the clarity of relationships among the remaining features. The selected features were subsequently analyzed using Pair Plot matrices, with data points in the subplots being colored according to clusters identified on L_{AVG} through K-means. This approach enabled a clear visualization of the relationships among the selected features while emphasizing the distinctions between clusters and their interaction with the length-based features of the samples.

2.4.2. Model fitting on training-validation data

The classification models assign each sample to a predefined dimensional class (referred to as 'label') using the extracted shadow-based features as input variables. The labels were defined based on L_{AVG} ranges, with a spacing of 5 mm between classes. Specifically, the classes were defined as follows: S ($L_{AVG} < 14.9$ mm), M ($14.9 \leq L_{AVG} \leq 20$ mm), L ($20 < L_{AVG} \leq 25$ mm), and XL ($L_{AVG} > 25$ mm). These classes were chosen to reflect the most commonly observed pellet dimensions available on the market, ensuring practical relevance and consistency with real-world production trends.

Three classification algorithms were evaluated: K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Logistic Regression (LR). The training phase was carried out on the normalized training-validation dataset, employing grid search combined with 10-fold cross-validation for hyperparameter optimization (Please refer to the Supplementary Material for the table listing the hyperparameter grid). A pipeline was implemented to ensure that data normalization was performed independently within each fold of the cross-validation, thereby preventing information leakage and preserving model integrity.

Once the best-performing configurations were identified for each algorithm, the corresponding models were retrained on the entire training-validation dataset to leverage the full sample size and ensure that the final models were trained on all available data. All procedures were implemented using Python 3.11 and the following libraries: Pandas [28], Scikit-Learn [29], Matplotlib [30], Seaborn [31], and Numpy [32].

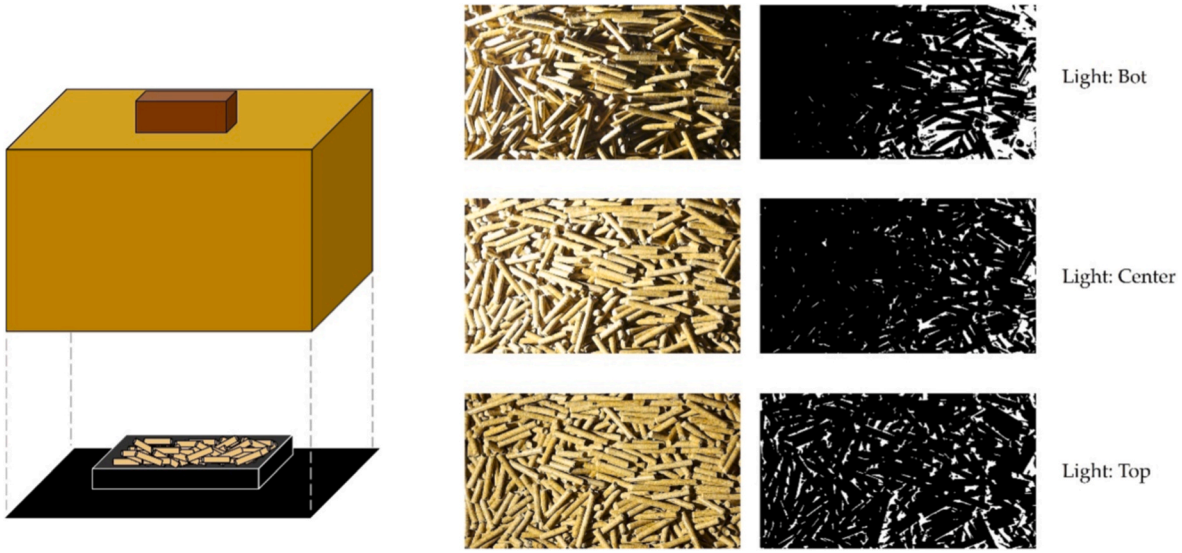


Fig. 3. On the left: acquisition procedure starts with the placement of the prototypal system upon the tray filled with pellets; on the right: captured images (RGB) and binarized images, with shadows highlighted in white.

2.4.3. Model testing on unseen data

The final models were evaluated on the Unseen dataset, with testing conducted on both individual data points and averaged data. Individual data points correspond to the features extracted from each photograph, while the averaged data represent the mean values derived from the three images captured for each sample under different lighting conditions.

For each model, macro performance metrics such as accuracy, precision, and recall were computed for both individual and averaged data. Finally, after identifying the best overall model, class-specific metrics were generated.

3. Results and discussion

This section presents the main results, with a focus on the averaged data obtained from analyses performed using the prototypal system under three different lighting conditions, and their relation with the dimensional measurements.

Data that do not require extensive discussion, including the individual pellet length measurements obtained from standard dimensional analyses, the data from the prototypal system for individual samples, and details of the Unseen samples, are not addressed in dedicated paragraphs.

3.1. Bot data

3.1.1. Explorative analysis: Bot_{Data}

Table 1 presents a statistical summary of the Bot_{Data} resulting from the 144 images.

The gap between the averaged $Area_{PX}$ and $Area_{Q3}$, along with the considerable range between $Area_{max}$ and $Area_{min}$, suggests significant heterogeneity in pellet size. The high value of $Area_{max}$ may be due to the elongation of shadows caused by the low placement of the light source [24,25]. Additionally, such elongation may lead to the merging of adjacent shadows, resulting in lower N_{os} values and higher $Area_{PX}$.

Pearson analysis highlighted that Eccentricity always positively correlates with length-based features, with the highest correlation being with L_{Q2} ($r = 0.75$), and both L_{Q1} and L_{AVG} ($r = 0.74$). On the contrary, Extent consistently exhibits a negative correlation with length-based features, such as with $L_{Q1}$and L_{Q2} ($r = -0.58$), and L_{AVG} ($r = -0.57$). A similar trend was observed for N_{os}, which exhibited a negative correlation with L_{AVG} was $r = -0.53$, and with L_{Q3}, $r = -0.56$.

PCA analysis (Fig. 4) accounted for 69 % of the total variance with the first two principal components (PCs). Specifically, PC1 accounted for 44 % of the total variance, while PC2 for 25 %.

The scatter plot of the first two PCs revealed a pattern in the distribution of the data points. Specifically, it showed a pronounced distribution of points from the negative quadrant of PC1 to the positive quadrant of PC2. Furthermore, the data points clustered more tightly along PC1, while scattered along PC2, indicating that PC1 captured a more consistent pattern within the data.

The analysis of the loading plot, combined with the L_{AVG} based coloring of the scatter plot, revealed insightful relationships within the data. Notably, the distribution of datapoints has been significantly influenced by all the length-based features. Additionally, an evident inverse correlation emerged between these features and both N_{os} and Extent, highlighting a trend in which increasing geometric parameters are associated with a decrease of N_{os} and Extent.

Eccentricity showed a strong inverse relationship with Extent and Solidity. Other parameters, such as $Area_{BBX}$, $Area_{CH}$, $Area_{PX}$, and Entropy, appeared to have a strong influence on the distribution of points in the positive quadrant of PC1 and the negative quadrant of PC2. Meanwhile, the $Area_{Qn}$ features appeared to be responsible for the distribution in the positive quadrant of both PCs.

However, the presence of features carrying similar information may have led to a non-optimal distribution of data points. For this reason, in order to reduce information redundancy, features with unique descriptive information (N_{os}, Extent, Solidity, Eccentricity and Entropy) and features representing averaged data (L_{AVG} and $Area_{PX}$) have been selected and used for subsequent steps.

3.1.2. Extracted features: bot

Feature selection enhanced the explanatory power of the first two PCs (Fig. 5), which combined accounted for 74 % of the total variance. PC1 accounted for 54 % of the total variance, while the PC2 for an additional 20 %.

Feature selction highlighted findings consistent with the PCA performed on all features. Specifically, yet L_{AVG} appeared to be inversely correlated to N_{os}, leading to a shift in the distribution of points in the scatter plot from the negative to the positive quadrant of both PCs. No apparent correlation between L_{AVG} and $Area_{PX}$ has emerged. Furthermore, the inverse relationships between Eccentricity and both Solidity and Extent, have been confirmed. Lastly, a clearer inverse correlation emerged between $Area_{PX}$ and Extent.

The analysis of the Pair Plot matrix (Fig. 6) confirmed the existence of a relationship between specific features and provided additional insight.

For instance, the subplot with L_{AVG} on the x-axis and N_{os} on the y-axis supported the findings from the loading plot (Fig. 5), emphasizing an inverse relationship between these features.

Although no strong correlation between L_{AVG} and $Area_{PX}$ was evident in the Pair Plot, a trend could be observed in which lower L_{AVG} values correspond to a wider range of $Area_{PX}$ compared to higher L_{AVG} values. The Pair Plot also confirmed the absence of a clear relationship between $Area_{PX}$ and N_{os}. In contrast, $Area_{PX}$ displayed a positive correlation with Entropy.

3.2. Center data

3.2.1. Explorative analysis: $Center_{Data}$

Table 2 presents a statistical summary of the $Center_{Data}$ resulting from the 144 images captured.

Compared to Bot_{Data}, shadow sizes appeared more uniform, as indicated by the minimal difference between averaged $Area_{PX}$ and $Area_{Q3}$. Additionally, the decrease in $Area_{max}$ is likely attributable to reduced shadow merging resulting from the higher placement of the light source. This also helps explain the higher number of N_{os} observed relative to Bot_{Data}.

Table 1
Averaged data derived from the 144 images captured under bottom-light illumination.

	N_{os}	$Area_{PX}$	$Area_{std}$	$Area_{min}$	$Area_{Q1}$	$Area_{Q2}$	$Area_{Q3}$	$Area_{max}$	$Area_{bbx}$	$Area_{ch}$	Perim.	Eccentr.	Solid.	Entro.	Extent
Mean	248	1402	11,460	30	61	132	392	168,483	4283	2963	180	0.85	0.82	0.43	0.53
St. Dev.	33	640	9588	0.29	7	24	105	154,051	1685	1329	47	0.02	0.01	0.11	0.02
Median	245	1188	8193	30	60	128	389	110,873	4067	2699	170	0.85	0.82	0.42	0.53
Min	161	232	387	30	49	87	216	3412	598	331	65	0.82	0.78	0.12	0.48
Q1	225	916	3817	30	56	116	325	44,268	2904	1923	142	0.84	0.81	0.34	0.52
Q2	245	1188	8193	30	60	128	389	110,873	4067	2699	170	0.85	0.82	0.42	0.53
Q3	269	1827	16,863	30	64	141	443	257,502	5571	3889	212	0.87	0.82	0.51	0.55
Max	339	3224	40,432	31	80	219	997	603,612	9017	6342	312	0.89	0.86	0.65	0.58

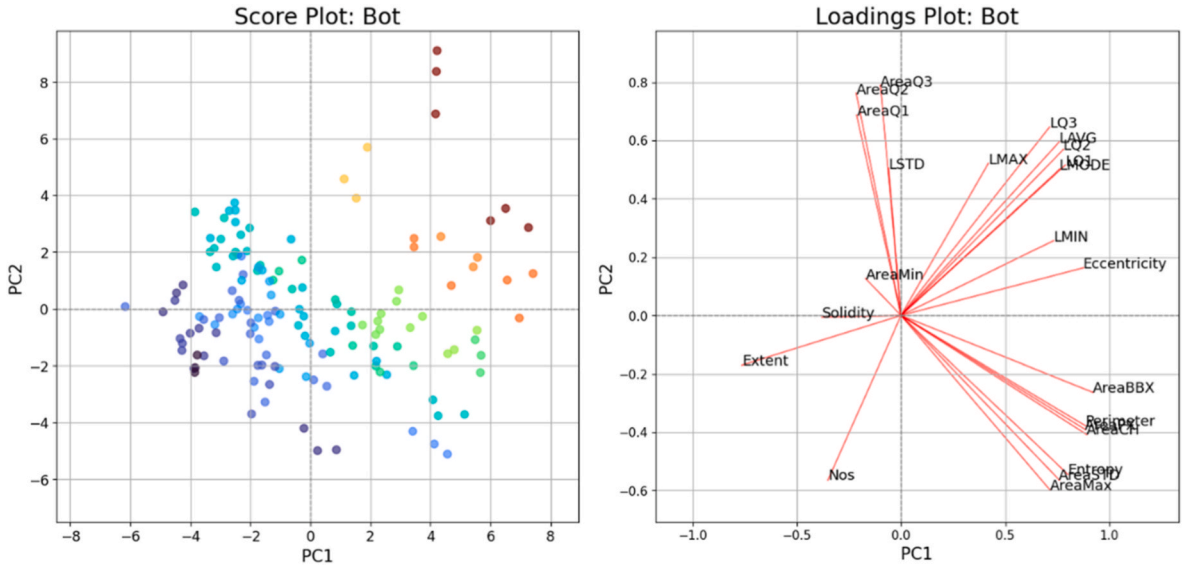


Fig. 4. Left: Score plot colored according to L_{AVG} values. Each point represents a sample (single image paired with dimensional data of the corresponding sample. Right: Loading plot associated to the scatter plot.

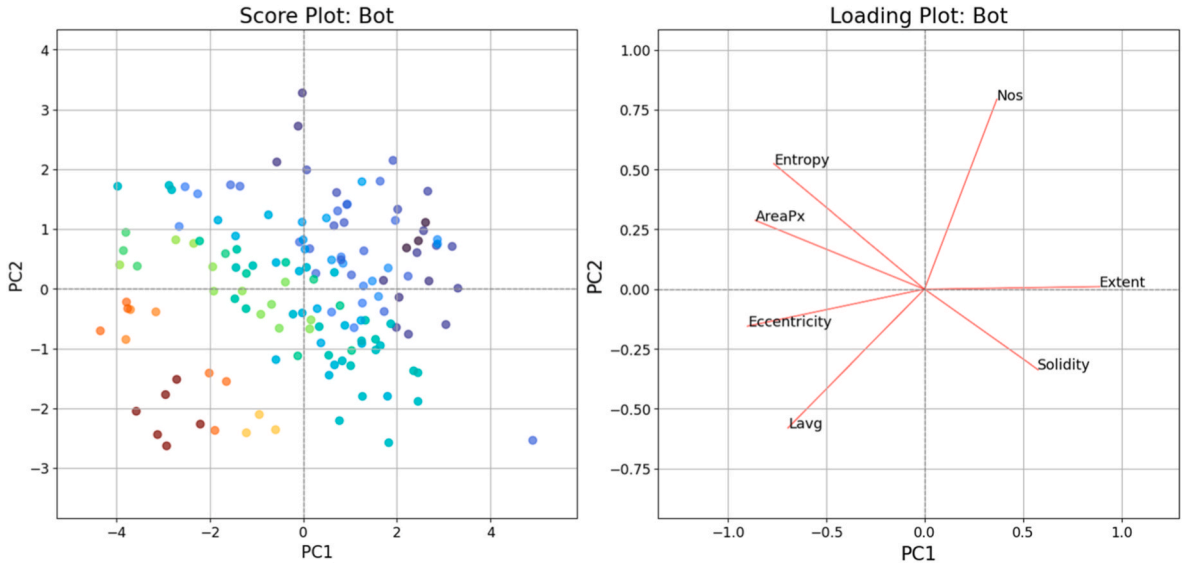


Fig. 5. Left: Scatter plot considering selected features colored according to L_{AVG}. Each point represents a sample (single image paired with dimensional data of the corresponding sample. Right: Loading plot associated to the scatter plot.

Pearson analysis identified Eccentricity as the shadow-based feature with the strongest correlation to sample length, with a positive correlation of 0.72 with both L_{Q2} and L_{mode}, and 0.71 with both L_{Q1} and L_{AVG}. Besides Eccentricity, both $Area_{Q2}$ and $Area_{Q3}$ showed positive correlations with length-based features. For instance, $Area_{Q2}$ correlated with L_{Q1} ($r = 0.60$) while $Area_{Q3}$ with L_{mode} ($r = 0.59$). L_{AVG} exhibited the strongest correlation with Eccentricity ($r = 0.71$), followed by $Area_{Q2}$ ($r = 0.56$). Lastly, N_{os} appeared mostly uncorrelated with length-based features.

PCA (Fig. 7) explained 77 % of the total variance with the first two principal components, where PC1 accounted for 58 % and PC2 for 19 %. This represents an 8 % increase in explained variance compared to the Bot_{Data} dataset.

The scatter plot of the first two principal components displayed a more stratified distribution than Bot_{Data}, with reduced overlap of data points in the negative quadrant.

Considering the distribution of points along the color gradient, both PCs contributed to the overall distribution, from the negative to the positive quadrant for both PCs.

The loading plot highlighted that the elongated distribution of the data points along PC1 was heavily influenced by the inverse relationship between features describing shadows' area and both Solidity and Extent. In contrast, the more dispersed distribution along PC2 seemed to be influenced by the length-based features, area of shadows, and N_{os}.

As for Bot_{Data}, a redundancy of features appeared to be present. As suggested by Pearson correlation analysis, length-based features showed no correlation with N_{os}. Lastly, the available features did not account for the distribution observed in the negative quadrant of PC1 and PC2.

3.2.2. *Extracted features: center*

The reduction of redundant features markedly increased the explained variance of the first two PCs (Fig. 8), capturing 92 % of the total variance. Specifically, PC1 accounted for 75 % of the variance, while PC2 contributed 17 %.

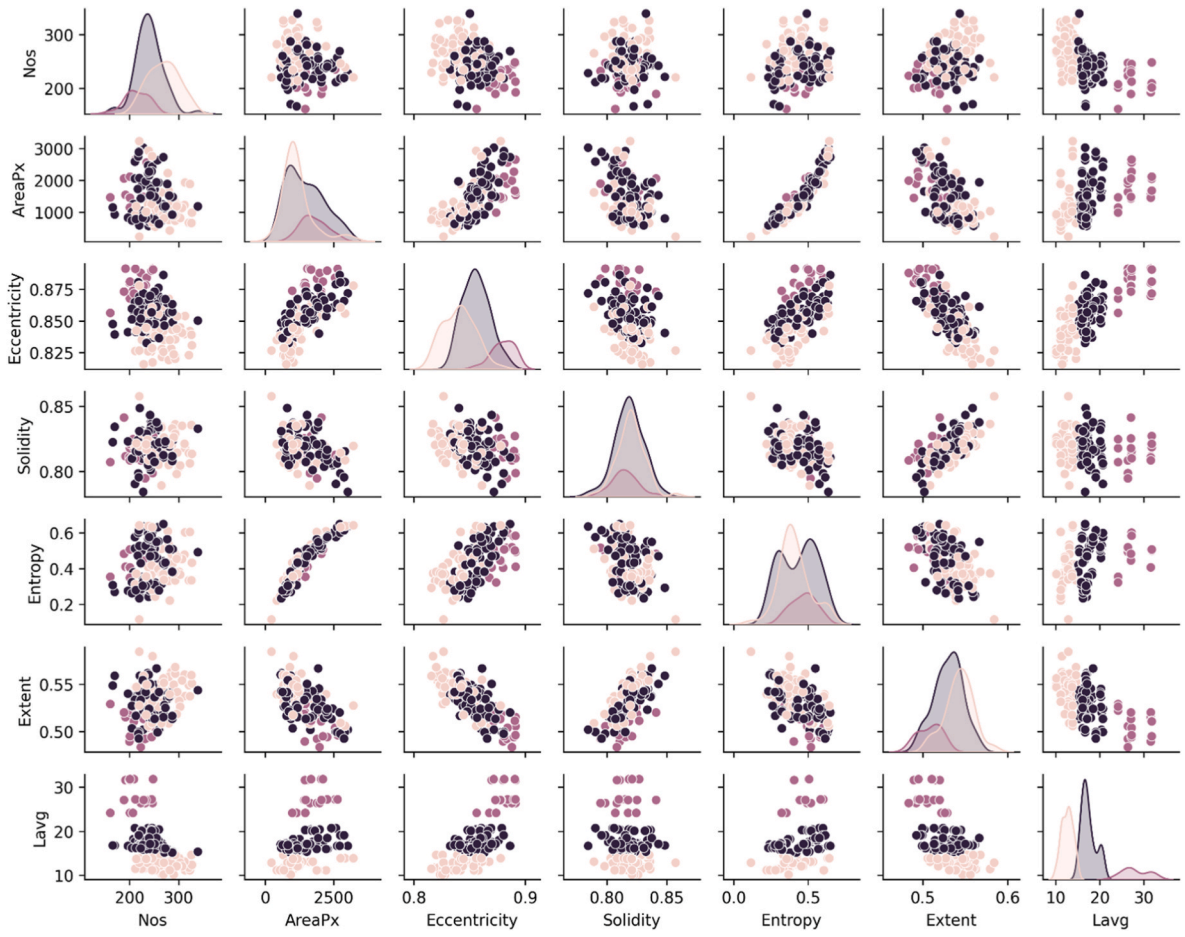


Fig. 6. Pair plot matrix displaying scatter plots of all extracted features. Each point represents a sample (single image paired with dimensional data of the corresponding sample), with colors indicating clusters obtained through K-Means. Purple dots correspond to longer samples, blue dots to medium samples, and yellow dots to shorter samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Averaged data of the 144 images captured with the central lighting.

	N_{os}	$Area_{PX}$	$Area_{std}$	$Area_{min}$	$Area_{Q1}$	$Area_{Q2}$	$Area_{Q3}$	$Area_{max}$	$Area_{bbx}$	$Area_{ch}$	Perim.	Eccentr.	Solid.	Entro.	Extent
Mean	326	406	1044	30	62	130	339	13,092	1299	740	91	0.85	0.83	0.24	0.55
St. Dev.	115	223	1068	0.26	8	31	130	17,530	888	525	32	0.02	0.03	0.13	0.04
Median	349	342	752	30	63	127	323	7847	1013	584	84	0.85	0.83	0.23	0.55
Min	97	114	91	30	43	73	132	535	212	133	42	0.80	0.77	0.03	0.47
Q1	244	223	316	30	56	106	241	2428	543	316	64	0.83	0.81	0.13	0.52
Q2	349	342	752	30	63	127	323	7847	1013	584	84	0.85	0.83	0.23	0.55
Q3	408	586	1354	30	67	150	456	15,696	1859	1062	115	0.86	0.85	0.35	0.57
Max	589	1087	8414	32	85	235	624	123,517	3493	2493	164	0.90	0.90	0.47	0.62

L_{AVG} did not appear to be strongly correlated with N_{os}, while the latter appeared to be positively correlated to $Area_{PX}$, as indicated by their alignment along the same direction in the loading plot. Lastly, both $Area_{PX}$ and N_{os} showed a strong inverse correlation to Solidity and Extent.

Compared to the previous Pair Plots, the Pair Plot matrix (Fig. 9), presented a weaker inverse correlation between L_{AVG} and N_{os}. However, as L_{AVG} increased, the spread of N_{os} values narrowed, indicating reduced variability in N_{os} at higher L_{AVG} values. A similar trend was observed between L_{AVG} and $Area_{PX}$.

The Pair Plot further validated the positive correlation between L_{AVG} and Eccentricity, and a slight negative correlation between L_{AVG} and Extent. Additionally, the Pair Plot highlighted the relationship between $Area_{PX}$ and other shadow-based features, specifically a positive correlation to Entropy and Eccentricity, and a negative correlation to Extent and Solidity.

3.3. Top data

3.3.1. Explorative analysis: Top_{Data}

Top_{Data} (Table 3) suggesting that a higher-positioned light source led to an increase in N_{os} and the formation of more homogeneous shadows.

Compared to previous results, the narrower range between $Area_{min}$ and $Area_{max}$, along with the reduced interquartile range between $Area_{Q1}$ and $Area_{Q3}$, suggests more uniform shadows. Additionally, the lower $Area_{max}$ value indicates a further reduction in shadow merging.

Pearson analysis revealed that the highest correlation values were between N_{os} and L_{Q3} ($r = -0.69$) and between N_{os} and L_{AVG} ($r = -0.67$). These inverse correlations underline what already emerged in the loading plot of Bot_{Data}, where length-based features were inversely

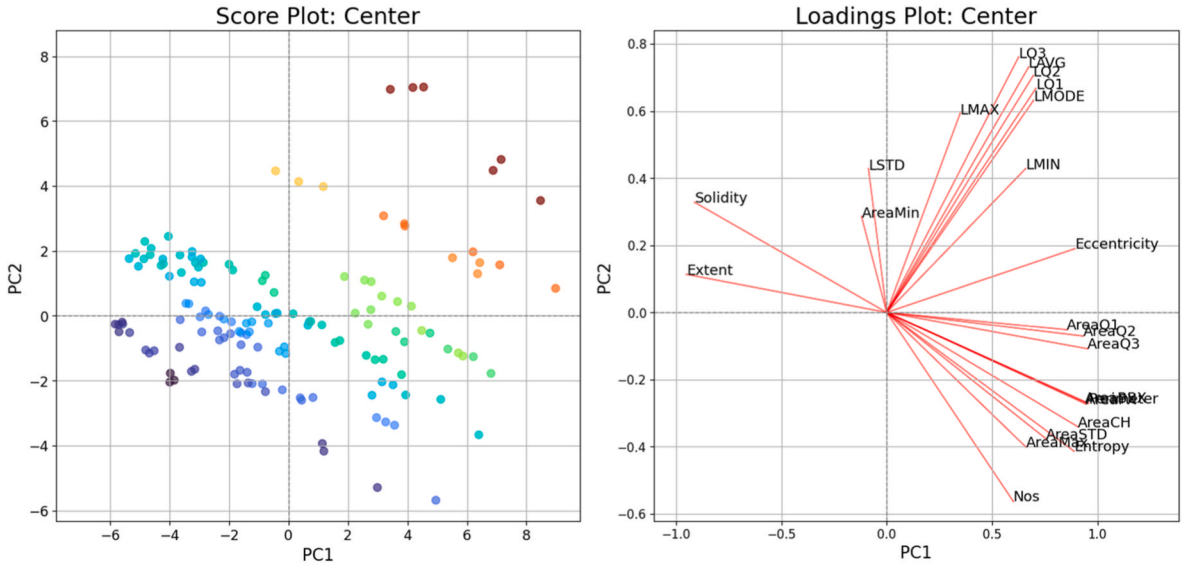


Fig. 7. Left: Score plot colored according to L_{AVG} values. Each point represents a sample (single image paired with dimensional data of the corresponding sample. Right: Loading plot associated to the scatter plot.

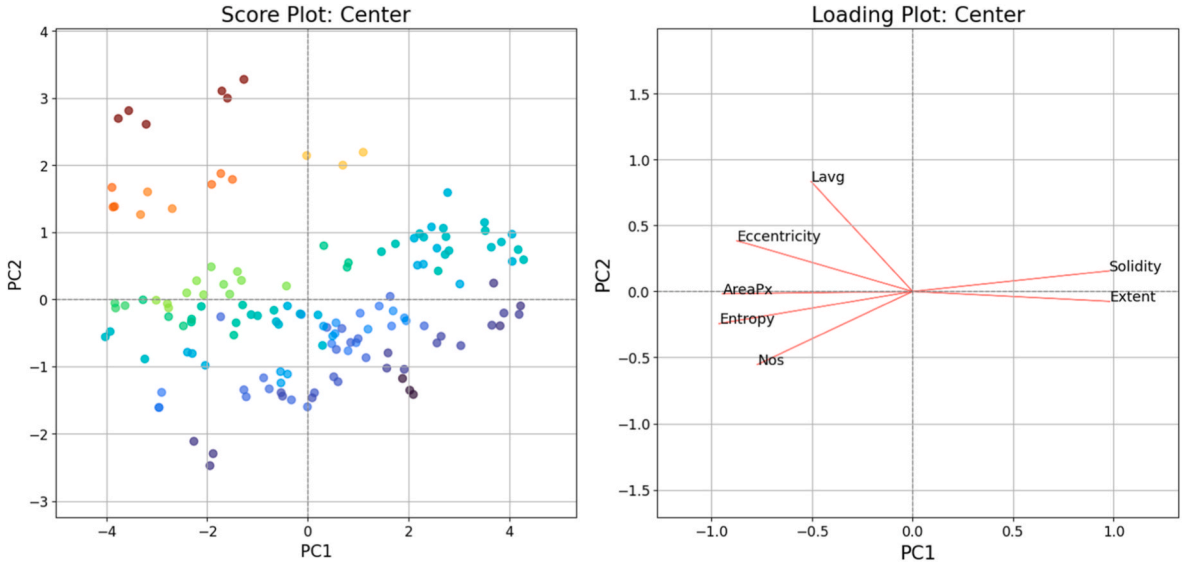


Fig. 8. Left: Score plot colored according to L_{AVG} values. Each point represents a sample (single image paired with dimensional data of the corresponding sample. Right: Loading plot associated to the scatter plot.

correlated with N_{os}. L_{AVG} maintained a consistent, but limited, positive correlation with $Area_{PX}$ (r = 0.54), $Area_{Q1}$ (r = 0.56), $Area_{Q2}$ (r = 0.58), and $Area_{Q3}$ (r = 0.57). Eccentricity also showed positive correlations with length-based features, peaking with L_{Q1} (r = 0.65).

PCA analysis (Fig. 10) accounted for 79 % of the total variance with the first two PCs, with PC1 accounting for 62 % and PC2 for the remaining 17 %. The score plot highlighted a distribution of data points influenced by L_{AVG}, from the negative quadrant of PC1 to the positive quadrant of PC2. In comparison to the previous scatter plots, a clearer separation into three distinct clusters was observed.

Similarly to the Bot_{Data} loading plot, length-based features displayed a strong inverse correlation with N_{os}. The orthogonal positioning between the areas' features and both length-based features and N_{os} suggested no apparent correlation among these variables. Finally, an inverse relationship between Eccentricity and both Solidity and Extent has been confirmed. The latter two features once again appeared to be inversely correlated with features describing shadows' areas and Entropy.

3.3.2. *Extracted features: top*

Feature selection (Fig. 11) increased the explained variance by the first two PCs to 90 %, with PC1 accounting for 70 % and PC2 accounting for 20 % of the total variance. The loading plot highlighted a positive correlation between $Area_{PX}$, Eccentricity, and Entropy, while these variables were inversely correlated with Solidity and Extent. Additionally, L_{AVG} and N_{os} displayed a strong inverse relationship.

The sub-plots (Fig. 12) with L_{AVG} on the x-axis emphasized the relationships initially highlighted in the previous Pair Plot matrices. For instance, it confirmed the inverse relationship between L_{AVG} and N_{os}, and the positive correlation between L_{AVG} and $Area_{PX}$. However, increased variability in $Area_{PX}$ was observed at higher L_{AVG} values, indicating that shadow size became less consistent as pellet length increased, contrasting with previous Pair Plots where longer samples showed less variability in $Area_{PX}$.

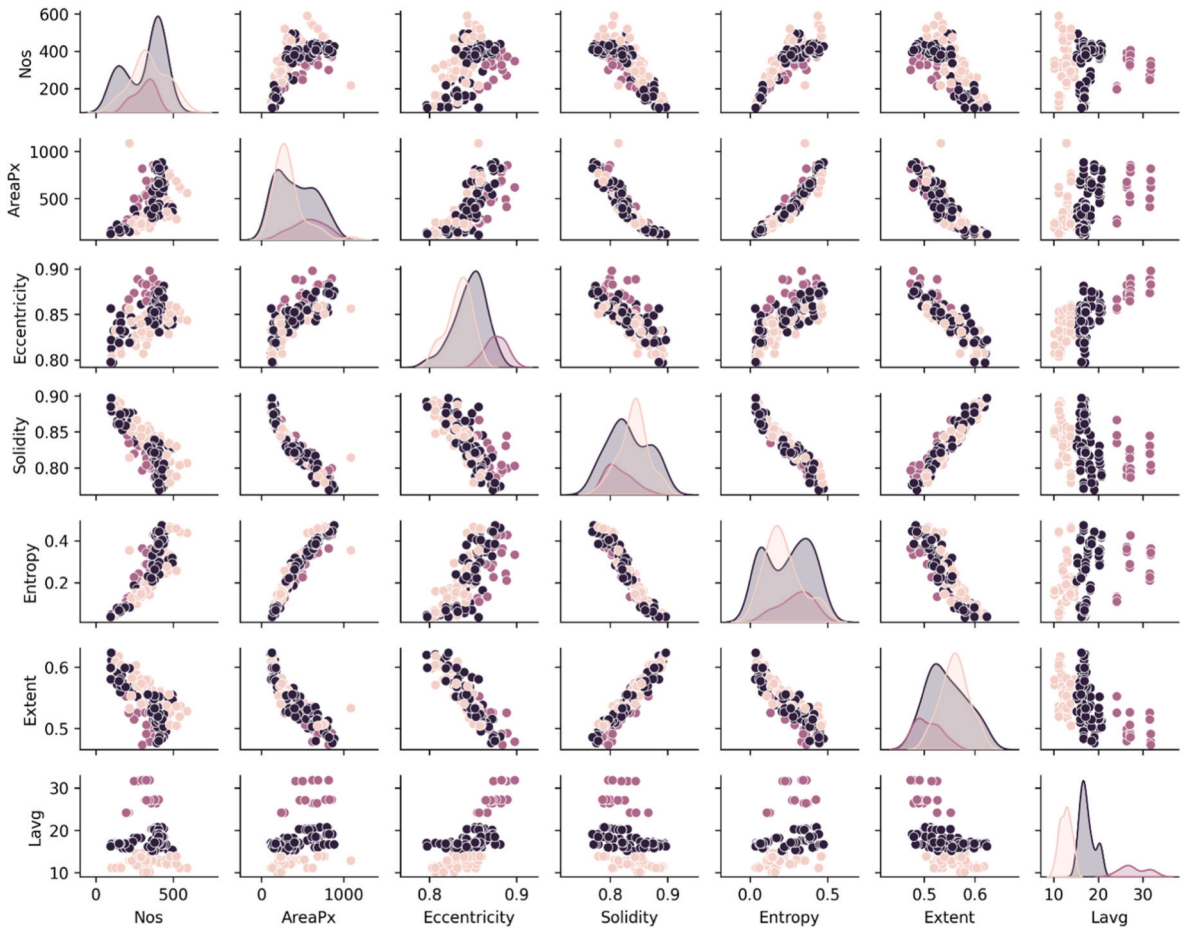


Fig. 9. Pair plot matrix displaying scatter plots of all extracted features. Each point represents a sample (single image paired with dimensional data of the corresponding sample), with colors indicating clusters obtained through K-Means. dots correspond to longer samples, blue dots to medium samples, and yellow dots to shorter samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Averaged data of the 144 images captured with the upper light.

	N_{os}	$Area_{PX}$	$Area_{std}$	$Area_{min}$	$Area_{Q1}$	$Area_{Q2}$	$Area_{Q3}$	$Area_{max}$	$Area_{bbx}$	$Area_{ch}$	Perim.	Eccentr.	Solid.	Entro.	Extent
Mean	463	537	812	30	93	246	638	7537	1813	972	114	0.85	0.81	0.34	0.52
St. Dev.	98	349	643	0.36	27	117	399	6640	1500	795	50	0.02	0.04	0.12	0.04
Median	458	391	558	30	88	207	492	5631	1186	634	97	0.85	0.81	0.33	0.52
Min	270	139	124	30	53	97	175	927	277	171	47	0.79	0.74	0.11	0.44
Q1	386	276	345	30	74	153	335	2965	712	402	75	0.84	0.78	0.26	0.49
Q2	458	391	558	30	88	207	492	5631	1186	634	97	0.85	0.81	0.33	0.52
Q3	541	781	1236	30	111	332	889	10,134	2674	1466	150	0.86	0.85	0.43	0.56
Max	633	1670	3278	32	187	646	2099	49,045	7324	3962	255	0.88	0.89	0.54	0.62

When examining the plot of N_{os} on the x-axis and $Area_{PX}$ on the y-axes, an inverse relationship was observed. Small N_{os} values corresponded to either extremely low or extremely high $Area_{PX}$, while higher N_{os} lead to smaller and less variable $Area_{PX}$. Finally, the Pair Plot reinforced the findings from the loading plot, particularly highlighting the relationships between $Area_{PX}$ and image processing features like Eccentricity and Extent.

3.4. Model training and validation

Following exploratory analyses and the elimination of redundant variables during feature extraction, two sets of features were selected for training the dimensional classification models.

The first set, referred to as X_A, includes all extracted variables that provide unique information. In the second set, X_B, $Area_{PX}$ is excluded because its correlation with pellet length is inconsistent and less reliable compared to other variables. especially under bottom and central lighting. This keeps X_B focused on the most stable predictors.

The target variable, L_{AVG}, was divided into the following classification ranges: <14.9, ≤20, ≤25, and >25, with each range reflecting a distinct dimensional class. The resulting models describe the relationship between the predictors and the dimensional classes, as summarized in the following equation (1):

$$Y = f(X_{A,B}) \tag{1}$$

where:
X_A = [N_{os}, $Area_{PX}$, *Eccentricity, Solidity, Entropy, Extent*].
X_B = [N_{os}, *Eccentricity, Solidity, Entropy, Extent*].
and:

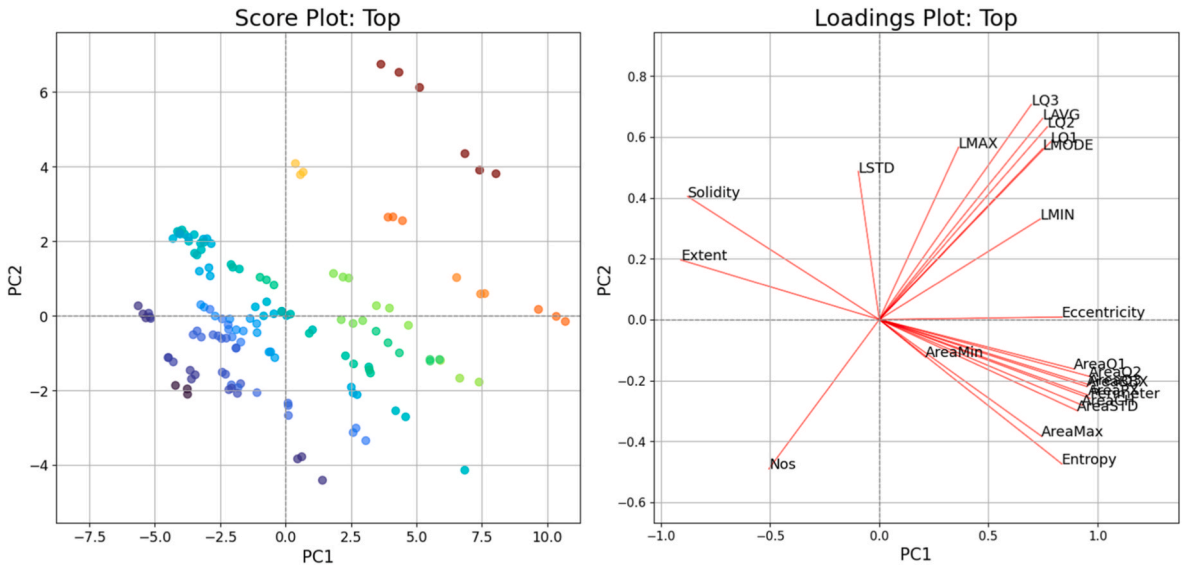


Fig. 10. Left: Score plot colored according to L_{AVG} values. Each point represents a sample (single image paired with dimensional data of the corresponding sample. Right: Loading plot associated to the scatter plot.

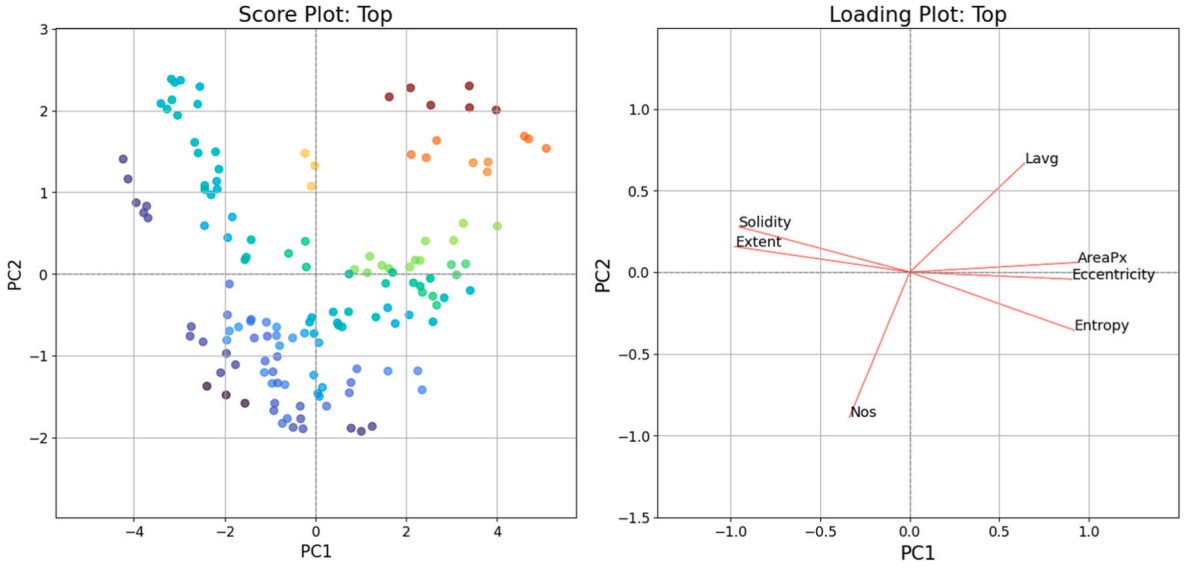


Fig. 11. Left: Score plot colored according to L_{AVG} values. Each point represents a sample (single image paired with dimensional data of the corresponding sample. Right: Loading plot associated to the scatter plot.

$$Y = \begin{cases} S & if \quad L_{avg} < 14.9 \\ M & if \quad 14.9 \geq L_{avg} \leq 20 \\ L & if \quad 20 > L_{avg} \leq 25 \\ XL & if \quad L_{avg} > 25 \end{cases}$$

Commenting in detail on model tuning would be complex and falls outside the scope of this study, which is primarily aimed at evaluating the ability to model shadows and classify dimensional categories. Therefore, the following paragraphs will focus on the influence of the selected predictors on overall accuracy during the training and validation phase.

3.4.1. Bot_{Data}: accuracy of models on training-validation data

Generally, Bot_{Data} models did not achieve high accuracy during training-validation (Table 4). The low performance observed during training and validation may be attributed to the high variability in shadow dimensions caused by the low positioning of the light source. This variability complicates the modelling of shadows and dimensions. The best performing models, regardless of the predictors used, were LR and SVC. Switching from X_A to X_B had a positive effect on the accuracy of LR, increasing it from 0.66 to 0.68, while it had no effect on SVC and KNN.

3.4.2. $Center_{Data}$: accuracy of models on training-validation data

Models trained on the $Center_{Data}$ also yielded poor accuracies (Table 4). Among the various models tested LR and SVM consistently led to the highest performance. Switching from predictors X_A to X_B led to mixed outcomes: while KNN showed a slight improvement in accuracy the accuracies of SVM and LR remained unchanged at 0.64 and 0.66, respectively.

3.4.3. Top_{Data}: accuracy of models on training -validation data

Top_{Data} led to models that consistently outperformed those based on previous datasets (Table 4), with SVC achieving the highest accuracy.

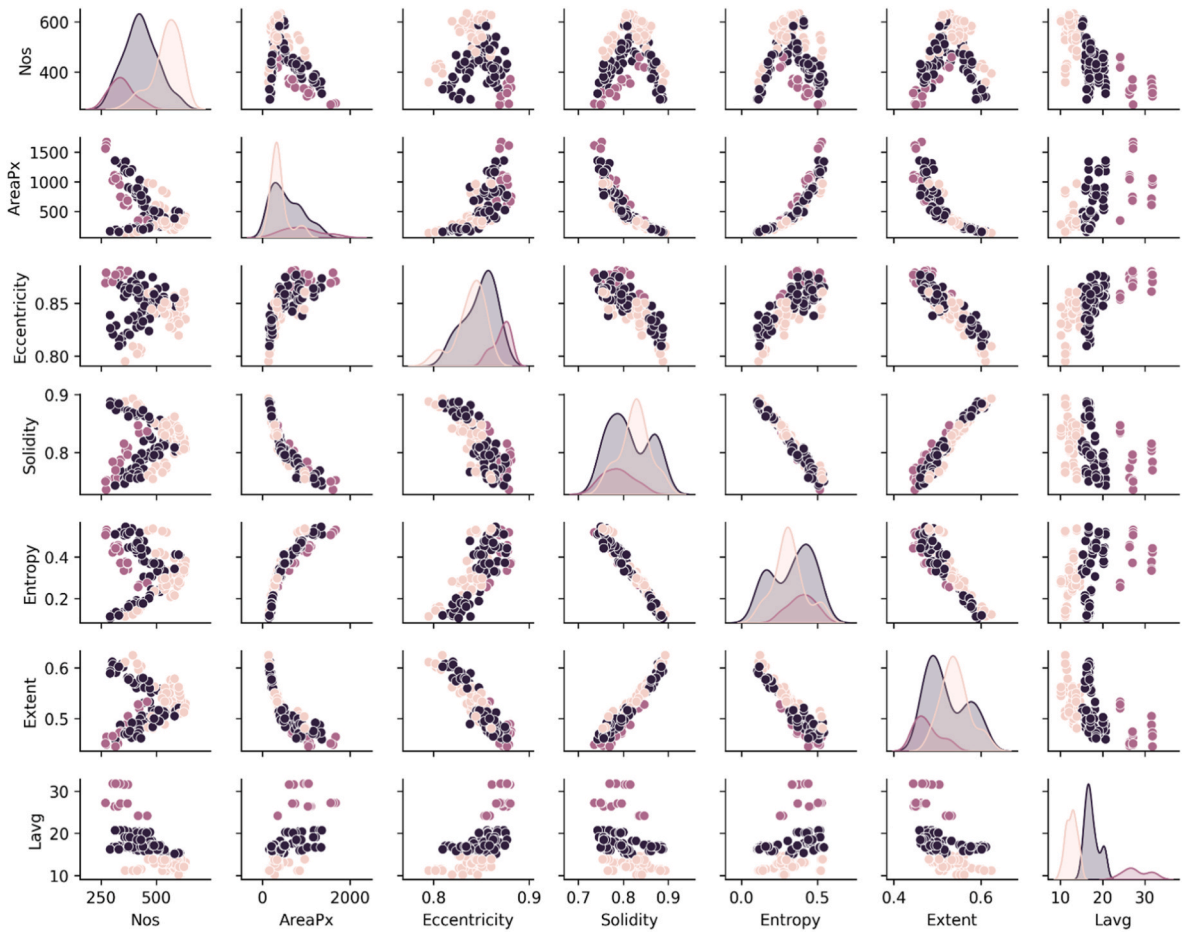


Fig. 12. Pair plot matrix displaying scatter plots of all extracted features. Each point represents a sample (single image paired with dimensional data of the corresponding sample), with colors indicating clusters obtained through K-Means. dots correspond to longer samples, blue dots to medium samples, and yellow dots to shorter samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 4
Hyperparameter tuning of each algorithm that achieved the highest accuracy during training and validation.

Algorithm	Predictors	Hyperparameter	Bot		Center		Top	
			Values	Accuracy	Values	Accuracy	Values	Accuracy
K-Nearest Neighbor		Neighbors	10		3		3	
	X_A	Weights	Distance	0.59	Uniform	0.63	Distance	**0.75**
		Metric	Euclidean		Euclidean		Manhattan	
		Neighbors	7		5		5	
	X_B	Weights	Distance	0.59	Distance	0.64	Distance	0.74
		Metric	Manhattan		Manhattan		Euclidean	
Logistic Regression		C	1		25		1	
	X_A	Penalty	L2	0.66	L2	0.66	L2	0.74
		Solver	Lbfgs		Lbfgs		Lbfgs	
		C	10		10		1	
	X_B	Penalty	L2	0.68	L2	0.66	L2	**0.76**
		Solver	Lbfgs		Lbfgs		Lbfgs	
Support Vector Machine		C	0.1		10		25	
	X_A	Kernel	Linear	0.67	Rbf	0.64	Rbf	0.79
		Gamma	Scale		Scale		Scale	
		C	0.1		10		1	
	X_B	Kernel	Linear	0.67	Rbf	0.64	Rbf	**0.80**
		Gamma	Scale		Scale		Scale	

Where $X_A = [N_{os}, Area_{PX}, Eccentricity, Solidity, Entropy, Extent]$ and $X_B = [N_{os}, Eccentricity, Solidity, Entropy, Extent]$.

Moreover, the accuracy seemed to improve slightly when using the X_B predictors, except for KNN, where it led to a slight decrease in accuracy. The enhanced accuracy of the present dataset relative to previous ones can be explained by the findings from exploratory analyses, which highlighted that a higher light position generates shadows with more uniform characteristics, thereby significantly enhancing the performance of the models.

3.5. Model testing

The test has been carried out on both individual Unseen data (Table 5) and the averaged data from the three images taken for each sample. The performance has been evaluated based on macro-metrics, such as Accuracy, Precision and Recall, which focus on the overall performance, independent of the performance on individual classes. Finally, a more in-depth analysis is presented for the model that achieved the best overall performance.

3.5.1. Bot_{Data}: performance on unseen data

Models trained with X_A predictors exhibited consistently poor performance. When using the X_B predictors, overall performance remained notably low, with a minor improvement observed in the KNN model trained on the individual data-point dataset, reaching 0.29 accuracy. Thus, despite the accuracy achieved during the training-validation phase of the bot models, which suggested some potential for applicability, the macro-metrics highlight the unreliability of their performance.

3.5.2. $Center_{Data}$: performance on unseen data

Even though the accuracy of the models were similar to that of Bot_{Data}, models trained on the $Center_{Data}$ dataset using X_A predictors showed an improvement compared to previous models, particularly for LR and SVC. The better performance of the $Center_{Data}$ models compared to previous models could be explained by the lower presence of "merged" shadows and less data variability.

Specifically for LR, the performance remained unchanged regardless of whether individual or averaged data was used. Conversely, for SVC, accuracy improved from 0.52 to 0.57. Furthermore, for both models, Recall exceeded Precision while maintaining a reasonable balance between the two metrics.

Switching to X_B did not lead to significant changes in LR performance. However, an improvement was observed for KNN and SVC, particularly on average data. In this case as well, SVC demonstrated higher Recall compared to Precision while keeping both metrics in a comparable range.

3.5.3. Top: performance on unseen data

As expected from the accuracies obtained during the training-validation, Top_{Data} models yielded the best performances. When using X_A predictors, all models led to 0.62 accuracy, showing however variations in macro-recall and macro-precision, with the latter being higher in KNN. Testing the models on averaged data improved the performance of SVC and LR, while KNN experienced a slight decline in performance. Switching to X_B predictors had no effect on LR but led to an overall improvement in performance, particularly for KNN.

In summary, the KNN model performed notably well, achieving an accuracy of 0.71. It demonstrated strong macro-recall (0.75), indicating its ability to correctly identify most instances of each class, but had a moderate macro-precision (0.70), suggesting the presence of false positives during classification. Both LR and SVM also achieved an accuracy of 0.71, but their macro-precision (0.56) and macro-recall (0.69) indicated an imbalance, leaving room for improvement in distinguishing among classes.

3.5.4. Better performing model on unseen data

Among all the evaluated models, the best performance was achieved using Top_{Data}, specifically with the KNN model trained on X_B predictors and tested on individual data. This configuration reached an accuracy of 0.71, a macro-precision of 0.70, and a macro-recall of 0.75. The model was tuned using the following hyperparameters: Euclidean distance as the metric, 5 neighbors, and distance-based weighting.

An analysis of class-specific metrics revealed that recall reached a perfect value of 1 for classes S and XL, while it dropped to 0.66 for class M and 0.33 for class L. This indicates that the model reliably identifies all relevant samples for the S and XL classes but struggles with M and, more notably, with L.

In terms of precision, the model achieved a value of 1 for class XL, 0.8 for class M, and 0.5 for both S and L. These results suggest highly accurate predictions for class XL, with no irrelevant samples being misclassified. Predictions for class M were reasonably precise, while lower precision for S and L indicates a greater proportion of misclassified or ambiguous samples within these categories.

Overall, the results suggest that the KNN model performed effectively when classifying samples belonging to the more distinct dimensional classes, particularly S and XL. However, its generalization capacity declined for the more similar or overlapping classes M and L, likely due to feature overlap or insufficient separation in the underlying data.

Other models trained on Top_{Data}, such as Logistic Regression or SVM applied to averaged data and using X_B predictors, achieved accuracy values comparable to that of KNN. Nonetheless, their macro-level performance metrics indicated less desirable overall classification behavior.

Table 5
Classification results on Unseen data. For each lighting condition and both predictor sets, the macro-metrics obtained on both individual and averaged data are reported.

Algorithm	Predictors	Metrics	Bot		Center		Top	
			Individual	Averaged	Individual	Averaged	Individual	Averaged
K-Nearest Neighbor		Accuracy	0.14	0.14	0.38	0.43	0.62	0.57
	X_A	Precision	0.04	0.04	0.41	0.42	0.66	0.58
		Recall	0.25	0.25	0.35	0.38	0.71	0.62
		Accuracy	0.29	0.14	0.52	0.57	**0.71**	0.71
	X_B	Precision	0.39	0.05	0.47	0.44	**0.70**	0.58
		Recall	0.38	0.25	0.48	0.44	**0.75**	0.69
Logistic Regression		Accuracy	0.14	0.14	0.57	0.57	0.62	0.71
	X_A	Precision	0.04	0.04	0.56	0.56	0.52	0.56
		Recall	0.25	0.25	0.62	0.62	0.65	0.69
		Accuracy	0.14	0.14	0.57	0.57	0.62	0.71
	X_B	Precision	0.04	0.04	0.56	0.56	0.52	0.56
		Recall	0.25	0.25	0.62	0.62	0.65	0.69
Support Vector Machine		Accuracy	0.14	0.14	0.52	0.57	0.62	0.71
	X_A	Precision	0.04	0.04	0.52	0.56	0.63	0.56
		Recall	0.25	0.25	0.60	0.62	0.65	0.69
		Accuracy	0.14	0.14	0.57	0.57	0.67	0.71
	X_B	Precision	0.04	0.04	0.53	0.56	0.54	0.56
		Recall	0.25	0.25	0.62	0.62	0.67	0.69

Where $X_A = [N_{os}, AreaPx, Eccentricity, Solidity, Entropy, Extent]$ and $X_B = [N_{os}, Eccentricity, Solidity, Entropy, Extent]$.

3.6. Conclusions

This study demonstrates the feasibility and potential of a novel, patent-pending imaging methodology for non-invasive classification of wood pellet dimensions in densely stacked configurations. The approach overcomes the limitations of conventional analysis and segmentation-dependent image processing by leveraging the correlation between pellet size and the geometric properties of their shadows under controlled lateral illumination.

The position of the light source was found to be critical for shadow quality. Low-angle illumination produced elongated and often merged shadows, introducing high variability that complicated modeling and reduced classification accuracy. In contrast, higher-angle illumination generated more discrete and uniform shadows, significantly improving the distinguishability between dimensional classes. This highlights the importance of optimizing hardware configuration as a prerequisite for reliable data acquisition.

Application of machine learning models confirmed the predictive value of the extracted shadow-based features. The most effective model, a K-Nearest Neighbors classifier using features from the top-lit scenario and a refined predictor set, achieved 71 % accuracy on unseen data. The model performed best in distinguishing the most distinct size classes (S and XL, with high recall), but struggled with intermediate classes (M and L), indicating overlap in the feature space for pellets within the 15–25 mm range. This suggests that under the current setup, shadow characteristics alone are insufficient for reliable separation of these intermediate sizes, a key area for future refinement.

Methodologically, this shadow-based analysis provides a low-cost, scalable alternative capable of rapidly assessing large numbers of pellets, offering a more representative characterization of batch variability than small-sample manual methods. It also paves the way for real-time, in-situ monitoring of pellet dimensions in industrial storage silos or domestic boiler feeding systems. Real-time data on pellet size distribution could enable automated control systems to adjust feed rates and air-to-fuel ratios proactively, optimizing combustion efficiency and reducing emissions, crucial for maximizing the environmental benefits of pellet-based bioenergy.

Future work should focus on improving system robustness and accuracy, including exploring a wider range of lighting parameters, such as intensity, to produce more distinctive shadow signatures. Validating and refining the system under true industrial conditions, with varying pellet types and environmental factors, will be essential for transitioning this proof-of-concept into a reliable tool for the bioenergy sector. Extending the methodology to other granular biofuels, such as wood chips or agricultural residues, represents another promising avenue for research.

CRediT authorship contribution statement

Thomas Gasperini: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elena Leoni:** Writing – review & editing, Data curation. **Nicolò Bartolini:** Writing – review & editing, Software, Data curation. **Giacomo Ciccone:** Formal analysis. **Giuseppe Toscano:** Supervision, Project administration, Conceptualization. **Carmine De Francesco:** Writing – review & editing, Writing – original draft, Visualization.

Patents

The method for analyzing the dimensional class of objects by examining the shadows cast by objects arranged in groups or in bulk is the subject of a pending patent. The pending patent No. 102024000017878 dated 31/07/2024 has been submitted to the Italian Patent and Trademark Office—UIBM and is titled “Computer implemented method for the analysis of the dimensional class of objects”.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Thomas Gasperini, Giuseppe Toscano has patent #102024000017878 pending to Assignee. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.renene.2025.124925.

References

[1] E. Eckert, O. Kovalevska, Sustainability in the European union: analyzing the discourse of the European green deal, J. Risk Financ. Manag. 14 (2021), https://doi.org/10.3390/jrfm14020080.
[2] D. Gielen, F. Boshell, D. Saygin, M.D. Bazilian, N. Wagner, R. Gorini, The role of renewable energy in the global energy transformation, Energy Strategy Rev. 24 (2019) 38–50, https://doi.org/10.1016/j.esr.2019.01.006.
[3] P. Lamers, M. Junginger, C. Hamelinck, A. Faaij, Developments in international solid biofuel trade - an analysis of volumes, policies, and market factors, Renew. Sustain. Energy Rev. 16 (2012) 3176–3199, https://doi.org/10.1016/j.rser.2012.02.027.
[4] H. Thomson, C. Liddell, The suitability of wood pellet heating for domestic households: a review of literature, Renew. Sustain. Energy Rev. 42 (2015) 1362–1369, https://doi.org/10.1016/j.rser.2014.11.009.
[5] Eurostat, Eurostat Data Browser - Rounwood, Fuelwood and Other Basic Products, (n.d.). ec.europa.eu/eurostat/databrowser (accessed January 1, 2025).
[6] A. Laschi, E. Marchi, S. González-García, Environmental performance of wood pellets' production through life cycle analysis, Energy 103 (2016) 469–480, https://doi.org/10.1016/J.ENERGY.2016.02.165.
[7] International Organization for Standardization, ISO 17225-2:2014 - Solid biofuels-fuel Specifications and classes-part 2: Graded Wood Pellets, 2014. Geneva, Switzerland.
[8] G. Zając, A. Kuranc, J. Szyszlak-Bargłowicz, Monitoring quality parameters of wood pellets available on the Polish market, lecture notes in civil engineering 609 LNCE, 488–496, https://doi.org/10.1007/978-3-031-70955-5_53, 2024.
[9] M. Wöhler, D. Jaeger, G. Reichert, C. Schmidl, S.K. Pelz, Influence of pellet length on performance of pellet room heaters under real life operation conditions, Renew. Energy 105 (2017) 66–75, https://doi.org/10.1016/j.renene.2016.12.047.
[10] R. Mack, C. Schön, D. Kuptz, H. Hartmann, T. Brunner, I. Obernberger, H.M. Behr, Influence of pellet length, content of fines, and moisture content on emission behavior of wood pellets in a residential pellet stove and pellet boiler, Biomass Convers. Biorefinery (2022), https://doi.org/10.1007/s13399-022-03302-6.
[11] R. Labbé, S. Paczkowski, V. Knappe, M. Russ, M. Wöhler, S. Pelz, Effect of feedstock particle size distribution and feedstock moisture content on pellet production efficiency, pellet quality, transport and combustion emissions, Fuel 263 (2020) 116662, https://doi.org/10.1016/J.FUEL.2019.116662.
[12] H.H. Lee, C.M. Hsu, Total quality management enhances wood pellet utilization for sustainable energy, Sustainability 17 (2025) 1562, https://doi.org/10.3390/SU17041562.
[13] International Organization for Standardization, ISO 17829:2015 - solid biofuels-determination of length and diameter of pellets, Geneva, Switzerland, www.iso.org, 2015.
[14] G. Toscano, E. Leoni, C. De Francesco, G. Ciccone, T. Gasperini, The application of image acquisition and processing techniques for the determination of wooden pellet length as an alternative to ISO 17829, Resources 12 (2023), https://doi.org/10.3390/resources12100125.
[15] R. Pierdicca, M. Balestra, G. Micheletti, A. Felicetti, G. Toscano, Semi-automatic detection and segmentation of wooden pellet size exploiting a deep learning approach, Renew. Energy 197 (2022) 406–416, https://doi.org/10.1016/j.renene.2022.07.109.
[16] X. Feng, C. Yu, X. Liu, Y. Chen, H. Zhen, K. Sheng, Y. He, Nondestructive and rapid determination of lignocellulose components of biofuel pellet using online hyperspectral imaging system, Biotechnol. Biofuels 11 (2018) 88, https://doi.org/10.1186/s13068-018-1090-3.
[17] J. Jägers, S. Wirtz, V. Scherer, An automated and continuous method for the optical measurement of wood pellet size distribution and the gravimetric determination of fines, Powder Technol. 367 (2020) 681–688, https://doi.org/10.1016/j.powtec.2020.04.023.
[18] D. Barletta, R.J. Berry, S.H. Larsson, T.A. Lestander, M. Poletto, Á. Ramírez-Gómez, Assessment on bulk solids best practice techniques for flow characterization and storage/handling equipment design for biomass materials of different classes, Fuel Process. Technol. 138 (2015) 540–554, https://doi.org/10.1016/J.FUPROC.2015.06.034.

[19] I. Vitale, R.G. Dondo, M. González, M.E. Cóccola, Modelling and optimization of material flows in the wood pellet supply chain, Appl. Energy 313 (2022), https://doi.org/10.1016/J.APENERGY.2022.118776.
[20] S. Pang, T.H.G. Thio, F.L. Siaw, M. Chen, Y. Xia, Research on improved image segmentation algorithm based on GrabCut, Electronics 13 (2024) 4068, https://doi.org/10.3390/ELECTRONICS13204068.
[21] T. Gasperini, A. Pizzi, L. Olivi, G. Toscano, A. Ilari, D. Duca, Image processing technique for enhanced combustion efficiency of wood pellets, Energies 17 (2024), https://doi.org/10.3390/EN17236144.
[22] Y. Xiao, Y. Peng, M. Wang, Y. Ning, Y. Zhou, K. Kong, Y. Long, A novel method for predicting coarse aggregate particle size distribution based on segment anything model and machine learning, Constr. Build. Mater. 429 (2024) 136429, https://doi.org/10.1016/J.CONBUILDMAT.2024.136429.
[23] O. Péterfi, L. Madarász, M. Ficzere, K. Lestyán-Goda, P. Záhonyi, G. Erdei, E. Sipos, Z.K. Nagy, D.L. Galata, In-line particle size measurement during granule fluidization using convolutional neural network-aided process imaging, Eur. J. Pharmaceut. Sci. 189 (2023) 106563, https://doi.org/10.1016/J.EJPS.2023.106563.
[24] N. Varga, M.J. Mayer, Model-based analysis of shading losses in ground-mounted photovoltaic power plants, Sol. Energy 216 (2021) 428–438, https://doi.org/10.1016/J.SOLENER.2021.01.047.
[25] Y. Wang, Simulation and analysis of the change of shadow length based on MATLAB for a period of time, Adv. Eng. Res. (2016) 604–607, https://doi.org/10.2991/ICMEMTC-16.2016.119.
[26] G. Bradski, The OpenCV library, Dr. Dobb's J. Software Tools (2000).
[27] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1979) 62–66, https://doi.org/10.1109/TSMC.1979.4310076.
[28] W. Mckinney, Data Structures for Statistical Computing in Python, 2010.
[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in python. http://arxiv.org/abs/1201.0490, 2012.
[30] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (2007) 90–95, https://doi.org/10.1109/MCSE.2007.55.
[31] M. Waskom, Seaborn: statistical data visualization, J. Open Source Softw. 6 (2021) 3021, https://doi.org/10.21105/joss.03021.
[32] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant, Array programming with NumPy, Nature 585 (2020) 357–362, https://doi.org/10.1038/s41586-020-2649-2.