Data in Brief 67 (2026) 113009

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

The SP-MILK dataset: A comprehensive dataset of speckle pattern images of pure and adulterated milk samples

Irene Bassi [a], Cristina Nuzzi [b,*], Simone Pasinetti [b], Valentina Bello [a]

[a] University of Pavia, Department of Electrical, Computer and Biomedical Engineering, Via Adolfo Ferrata 5, Pavia, 27100, Italy
[b] University of Brescia, Department of Mechanical and Industrial Engineering, Via Branze 38, 25123 Brescia, Italy

ARTICLE INFO

Article history:
Received 1 April 2026
Revised 11 June 2026
Accepted 20 June 2026
Available online 22 June 2026

Dataset link: SP-MILK: a dataset of speckle pattern images of pure and adulterated animal milk (Original data)

Keywords:
Speckle pattern
Milk adulteration
Scattering
Optoelectronic measurements
Turbid fluids
Laser
Camera
Food fraud detection

ABSTRACT

Speckle pattern (SP) imaging is an optoelectronic technique that can be used to analyse suspensions and opaque liquid foods, such as milk. When coherent light interacts with a diffusive sample, it is back-scattered and it generates an interference pattern, called speckle pattern, that contains information about the physical and compositional properties of the material. Images of the produced SP can be collected with a camera and analysed. In recent years, studies in the scientific literature have started demonstrating how the analysis of SP images can help in distinguishing milk samples with different nutritional characteristics and detect possible adulteration. However, the availability of public datasets of SP images is still very limited, and the datasets currently available mainly include measurements of unadulterated milk samples only. To address this gap, we present the SP-MILK dataset, which contains images of speckle pattern acquired from both pure and milk samples intentionally adulterated and prepared in the laboratory. Images were obtained from cow milk and goat milk in the unadulterated case. For the adulterated case, we considered dilutions of whole cow milk with water and with water–glucose solutions in different proportions. The data were collected using an experimental setup in which a coherent light source illuminates the samples and a CMOS camera records the resulting SPs. The SP-

* Corresponding author.
E-mail address: cristina.nuzzi@unibs.it (C. Nuzzi).

https://doi.org/10.1016/j.dib.2026.113009

MILK dataset is intended to support the development and evaluation of computational methods for milk characterization and adulteration detection, and to provide a reference dataset for future studies based on SP data.

Specifications Table

Subject	Engineering & Materials science, Biology
Specific subject area	Optoelectronic measurements on milk samples carried out in the field of turbid fluid analysis and food fraud detection
Type of data	Raw images (.png) Annotations (.csv)
Data collection	SP images of unadulterated and adulterated milk samples were acquired during experimental campaigns using an optoelectronic setup consisting of a diode laser illuminating a milk sample contained in a cuvette. The laser is temperature-stabilized and driven by a bench power supply providing continuous current. A CMOS camera (without lenses) is positioned in front of the cuvette (on the same side as the laser) to capture backscattered light. Image acquisition was performed using the camera proprietary software installed on a laptop connected to the camera. The dataset consists of raw frames, and no normalization or additional processing was applied. Since the acquisition frame rate is much lower than the decorrelation time of the milk particles, subsequent SP frames are uncorrelated.
Data source location	The data were collected at the Laboratory of ElectroOptics of the University of Pavia, Department of Electrical, Computer and Biomedical Engineering, Via Adolfo Ferrata 5, Pavia, 27,100, Italy. Latitude: 45.202839, Longitude: 9.136811. They are stored in hard disks located in the same laboratory.
Data accessibility	Repository name: Zenodo Data identification number: 10.5281/zenodo.18699723 Direct URL to data: https://zenodo.org/records/20642469 Instructions for accessing these data: download the archive .zip from the repository and decompress it to get full access to the images.
Related research article	None

1. Value of the Data

- The dataset [1] contains raw speckle pattern images of milk samples obtained from experimental campaigns conducted using an optoelectronic setup employing a laser with wavelength of 785 nm and a CMOS camera. For the “Unadulterated” sub-dataset, the tested samples were 6: whole cow milk, partially skimmed cow milk, skimmed cow milk, goat milk, partially skimmed goat milk, and a reference solid sample (powdered dextrose). For the “Adulterated” sub-dataset, the tested samples were 6: whole cow milk and 4 adulterated samples obtained by mixing the whole cow milk with water and a solution of water and glucose at different concentrations, plus one soldi reference sample (powdered dextrose). These adulterations were chosen since they are the most common adulterants used in milk counterfeiting. Data were acquired during 10 acquisition campaigns for each sub-dataset, and for each campaign 2 fresh cuvettes were used to contain the sample. For each cuvette 50 frames were acquired, resulting in a total of 10 campaigns x 6 samples x 2 cuvettes x 50 frames = 6000 frames per sub-dataset (12,000 frames overall). At the present time, this kind of public dataset was never published before.
- The dataset is highly valuable for researchers interested in measurement science and, in particular, in speckle pattern imaging applied to diffusing fluids. The dataset is also of interest for the research community working in the field of milk analysis and food fraud detection,

and for the fields of computer science, artificial intelligence, data analysis, biology, food and nutrition science.

- The raw speckle pattern images can be easily integrated in data analysis pipelines by other scientists by applying feature extraction or image standardization techniques without further processing.
- This dataset is also useful for practitioners working in the field of food fraud detection, especially for chemical laboratories working with local authorities to identify food frauds in the dairy industry, and also to companies working toward the development and validation of new compact and easy to use measurement systems for milk adulteration detection.

2. Background

Recent advancements in optoelectronic technologies for the investigation of liquid foods, such as milk, highlight the potential of the SP imaging to distinguish samples based on their composition. SPs are random interference patterns generated when coherent light is scattered by a sample: the resulting wavefronts interfere on the sensor surface of a camera, and the intensity distribution can be recorded as an image. Despite the increasing attention on food safety, adulteration of liquid foods is a growing concern in many countries. State-of-the-art techniques to detect adulterants require specialized laboratory and expensive equipment. In contrast, SP imaging has proven to be a cost-effective technique. Several works were published proving its applicability to successfully distinguish milk samples according to their nutritional values [2–4], and to detect adulterations [5–8]. However, the only public dataset available [4] just contains SP videos and frames of commercial unadulterated milk samples with different fat content. At the present time, no public datasets of adulterated milk samples are available. Therefore, the SP-MILK dataset was assembled to fill this gap, providing the research community with a suitable number of SP data of both unadulterated and adulterated milk samples

3. Data Description

The dataset published in [1] is structured as depicted in Fig. 1. The root folder contains two sub-folders named “Unadulterated” and “Adulterated” with the same structure, both containing 6000 SP images each saved in .png format. The two folders refer to SP images collected on unadulterated and adulterated samples, respectively, prepared as described in the following Section (see Table 1 and Table 3). A supplementary file named “filenames_and_labels.csv” is also provided containing the filenames of all SP contained in the dataset and their folder structure, as shown in Fig. 2. This file can be used by other scientists to build the full paths to the images and link them to the corresponding class (shown in the third column “Sample”). Please note that, due to the overall size of the dataset, each campaign folder was zipped and uploaded to the Zotero repository.

For each sub-dataset (“Unadulterated” and “Adulterated”), 5 milk jugs per sample were used to produce the test samples. For each milk jug, two acquisition campaigns were conducted (see

Table 1
List of milk samples under test and their nutritional values.

Sample ID	Description	Lipid conc. (g/L)	Protein conc. (g/L)	Sugar conc. (g/L)
S1	Whole cow milk UHT	36	34	51
S2	Partially skimmed cow milk UHT	16	33	50
S3	Skimmed cow milk UHT	< 0.5	34	51
S4	Whole goat milk UHT	35	33	45
S5	Partially skimmed goat milk UHT	16	33	45
R	Powdered dextrose (solid reference)	//	//	//

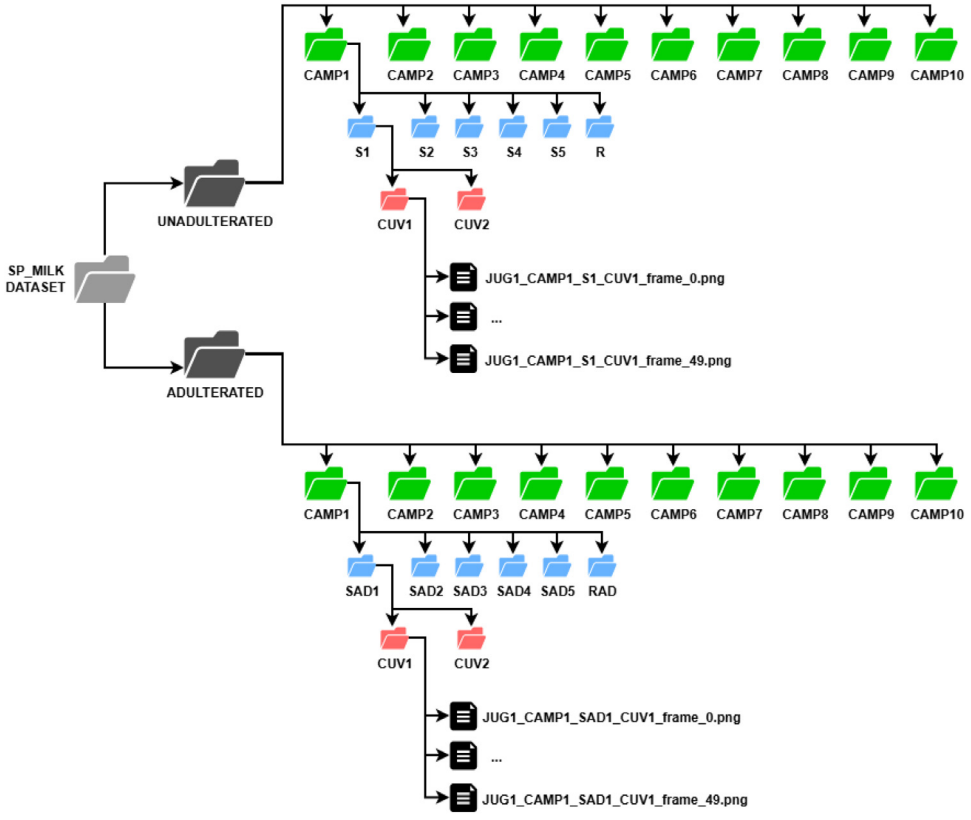


Fig. 1. Folder structure of the SP-MILK dataset after unzipping the files.

Subdataset	Campaign	Sample	Cuvette	Filename
Adulterated	CAMP1	RAD	CUV1	JUG1_CAMP1_RAD_CUV1_frame_0.png
Adulterated	CAMP1	RAD	CUV1	...
Adulterated	CAMP1	RAD	CUV1	JUG1_CAMP1_RAD_CUV1_frame_49.png
Adulterated	CAMP1	RAD	CUV2	JUG1_CAMP1_RAD_CUV2_frame_0.png
Adulterated	CAMP1	RAD	CUV2	...
Adulterated	CAMP1	RAD	CUV2	JUG1_CAMP1_RAD_CUV2_frame_49.png
Adulterated	CAMP1	SAD1	CUV1	JUG1_CAMP1_SAD1_CUV1_frame_0.png
Adulterated	CAMP1	SAD1	CUV1	...
Adulterated	CAMP1	SAD1	CUV1	JUG1_CAMP1_SAD1_CUV1_frame_49.png
Adulterated	CAMP1	SAD1	CUV2	JUG1_CAMP1_SAD1_CUV2_frame_0.png
Adulterated	CAMP1	SAD1	CUV2	...
Adulterated	CAMP1	SAD1	CUV2	JUG1_CAMP1_SAD1_CUV2_frame_49.png

Fig. 2. Graphical example of the structure of the annotations file "filenames_and_labels.csv" provided as supplementary data in the dataset.

Table 2 and Table 4); hence, both folders contain a total of 10 sub-folders named “CAMP[c]”, where “[c]” refers to the progressive number of the acquisition campaign from 1 to 10. In each “CAMP[c]” folder, there are a total of 6 sub-folders for the tested samples. For the “Unadulterated” sub-dataset, 5 sub-folders are named “S[s]” and one is named “R” (containing the reference sample data). For the “Adulterated” sub-dataset, 5 sub-folders are named “SAD[s]” and one

Table 2
List of the measurement campaigns for each milk jug, including acquisition date and time, label campaign and temperature details.

Milk jug	Acquisition date	Campaign	Temperature
1	15/10/2024 12:00 PM	CAMP1	T_{room} = 26.1 °C T_{sam} = 25.5 ~ 25.8 °C
	15/10/2024 3:00 PM	CAMP2	T_{room} = 26.2 °C T_{sam} = 25.7 ~ 26.2 °C
2	24/01/2025 11:00 AM	CAMP3	T_{room} = 24.0 °C T_{sam} = 23.0 ~ 23.2 °C
	24/01/2025 3:00 PM	CAMP4	T_{room} = 24.4 °C T_{sam} = 24.2 ~ 24.7 °C
3	12/11/2025 3:00 PM	CAMP5	T_{room} = 24.5 °C T_{sam} = 23.6 ~ 24.3 °C
	13/11/2025 11:00 AM	CAMP6	T_{room} = 25.3 °C T_{sam} = 24.0 ~ 25.0 °C
4	11/12/2025 2:00 PM	CAMP7	T_{room} = 21.3 °C T_{sam} = 20.9 ~ 21.2 °C
	12/12/2025 12:00 PM	CAMP8	T_{room} = 21.5 °C T_{sam} = 20.7 ~ 21.0 °C
5	09/01/2025 6:00 PM	CAMP9	T_{room} = 22.6 °C T_{sam} = 21.2 ~ 22.3 °C
	10/01/2025 10:00 AM	CAMP10	T_{room} = 22.5 °C T_{sam} = 21.0 ~ 22.3 °C

is named “RAD” (containing the reference sample data). For both datasets, “[s]” refers to the progressive number of the samples acquired, ranging from 1 to 5. In each sample folder there are 2 sub-folders named “CUV[v]”, where “[v]” refers to the progressive number of tested cuvettes from 1 to 2.

Inside each “CUV[v]” folder there are N = 50 frames numbered from 0 to 49 with the following naming pattern:

1. For the “Unadulterated” dataset: **JUG[j]_CAMP[c]_S[s]_CUV[v]_frame_[N]**.png
2. For the “Adulterated” dataset: **JUG[j]_CAMP[c]_SAD[s]_CUV[v]_frame_[N]**.png

For convenience, the data was not divided per jug using another sub-folder structure; instead, the jug numbering “[j]” was kept in the filename. Please note that for the adulterated samples “SAD”, the samples were prepared by mixing the pure milk taken from the corresponding milk jug and the adulteration solution (water or a combination of water and glucose as described in the following Section and Table 3). Therefore, the number of SP images contained in each sub-dataset is the following: 10 campaigns x 6 samples x 2 cuvettes x 50 frames = 6000 frames.

A complete list of the milk jugs, campaigns and samples preparation for each dataset is provided in the following Section. An example of SP image contained in the dataset is shown in Fig. 3.

4. Experimental Design, Materials and Methods

SP imaging is a non-invasive optical technique that exploits the coherent interaction of laser light with complex scattering media. When a coherent light source illuminates a turbid medium, such as milk, photons undergo multiple scattering events caused by suspended particles, including lipid micelles and caseins dispersed in an aqueous matrix of water, lactose, and soluble proteins. The superposition of the back-scattered wavefronts generates the characteristic interference pattern whose statistical properties are directly related to the physical and compositional characteristics of the medium. This pattern can be recorded by an imaging system (without

Table 3
List and description of the adulterated milk samples, with their nutritional values.

Sample ID	Description	Amount of diluent	Amount of pure milk	Lipid conc. (g/L)	Protein conc. (g/L)	Sugar conc. (g/L)
SAD1	Whole cow milk UHT	0 mL	20 ml	36	34	51
SAD2	Whole cow milk UHT diluted with water at 15% proportion	3 mL (water)	17 ml	30.6	28.9	43.35
SAD3	Whole cow milk UHT diluted with water at 30% proportion	6 ml (water)	14 ml	25.2	23.8	35.7
SAD4	Whole cow milk UHT diluted with a solution of water and glucose at 15% proportion	3 ml (water and glucose)	17 ml	30.6	28.9	62.1
SAD5	Whole cow milk UHT diluted with a solution of water and glucose at 30% proportion	6 ml (water and glucose)	14 ml	25.2	23.8	73.2
RAD	Powdered dextrose (solid reference)	//	//	//	//	//

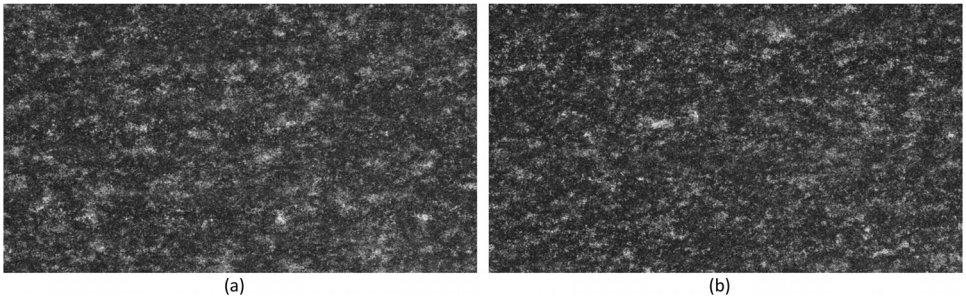


Fig. 3. Examples of raw SPs obtained from (a) a whole milk sample (SAD1) and (b) an adulterated sample diluted with 15% of water (SAD2). For visualization purposes, the overall contrast and illumination of the original images contained in the dataset have been increased.

lenses) and subsequently processed through image analysis algorithms to retrieve quantitative information about the sample under investigation. The following sections describe the experimental setup employed for the acquisition of SP images constituting the datasets presented in this paper, as well as the image acquisition process and the procedures adopted to ensure consistency and reproducibility across measurements.

5. Optoelectronic Setup

The measurements and acquisition of the SP images were performed using the laser–camera configuration shown in Fig. 4 and Fig. 5. All measurements were conducted in a dark room in the Laboratory of ElectroOptics of the University of Pavia. The setup was rigidly mounted on an optical table isolated from external vibrations.

Images of the SPs generated by the scattering samples are acquired using a monochrome CMOS camera (U3–38J0XCP-M-NO, iDS Imaging Development Systems GmbH, Germany) positioned in front of the cuvette. This camera orientation prevents the Snell reflection from the cuvette wall from impinging on and saturating the CMOS sensor. The camera sensor has a size of 5.603 × 3.155 mm, a resolution of 3864 × 2176 pixels, and a pixel size of 1.45 × 1.45 µm. A second linear polarizer is placed in front of the CMOS sensor and it is oriented as the first one. The camera is connected via USB to a laptop, enabling data acquisition through the proprietary software Vimba Viewer (Allied Vision, Germany).

A semiconductor diode laser (L785P090, Thorlabs, NJ, USA) was used as a coherent light source to irradiate the samples. It emits a maximum optical power of 40 mW at a wavelength

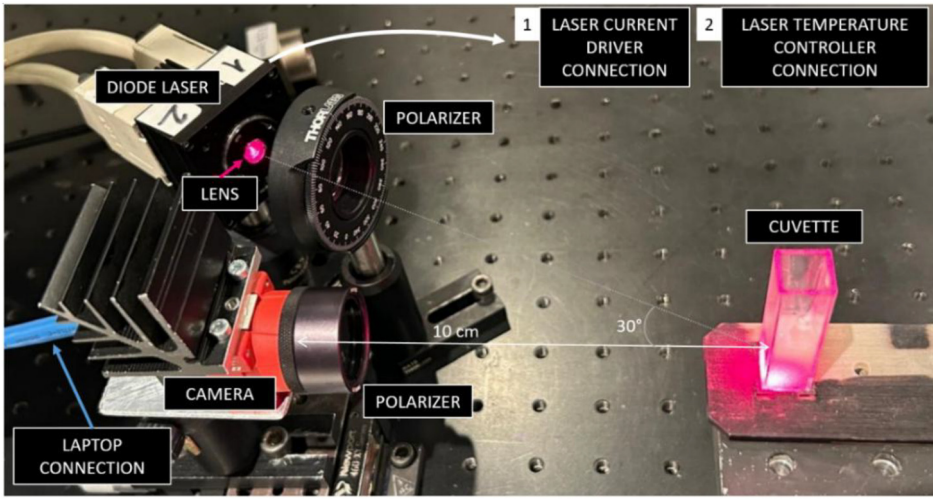


Fig. 4. Picture of the experimental setup showing each mechanical and optoelectronic component.

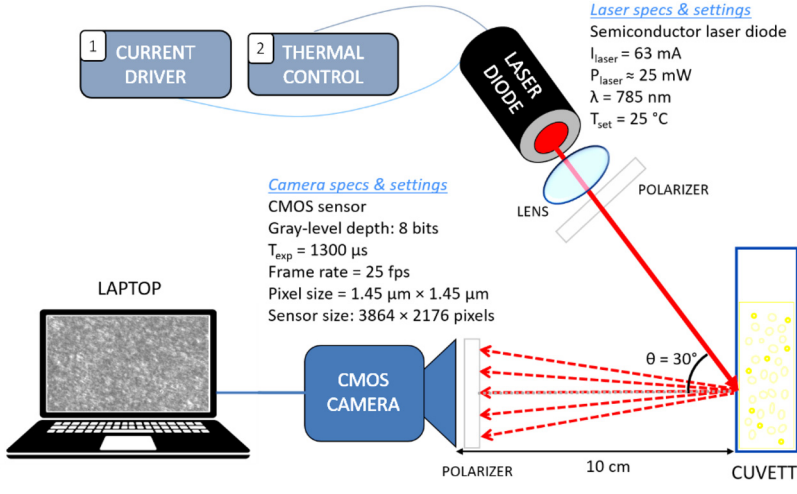


Fig. 5. Scheme of the optoelectronic setup used during the experiments.

of 785 nm. Through a dedicated mechanical mounting, the laser is powered by a current driver (LDC500, Thorlabs, NJ, USA) (via the connection labeled as "1" in Fig. 3), and thermally stabilized by means of a temperature controller (PRO800, Thorlabs, NJ, USA, connected via the connection labeled as "2" in Fig. 3). An aspheric lens (C230260P-B, Thorlabs, NJ, USA) with anti-reflection coating focuses the radiation onto the surface of a transparent plastic cuvette (volume 4.5 mL, dimensions 10 × 10 × 50 mm) containing the sample, at an angle of approximately 30° A linear polarizer (LPVISE100-A, Thorlabs, NJ, USA) is placed after the lens to select the main polarization component.

5.1. Laser and camera calibration strategy

It is worth noting that to obtain a SP the camera must be operated without optical lenses equipped. As a result, camera calibration is not needed. However, it is fundamental to perform

a calibration of the laser driver current so that on the sample the optical power is constant among experiments. At the same time, it is necessary to tune the camera's acquisition parameters so that the resulting SP image exploits all the intensity range preventing the occurrence of overexposed pixels.

The laser driver current was set to 63 mA to obtain a photodiode current of 0.1150 mA from the monitoring photodiode integrated in the laser package. This calibration strategy allowed to always obtain an optical power (measured each time by a power meter) on the sample equal to 25 mW.

The camera's acquisition frame rate was set to 25 fps and the exposure time was set to 1300 µs. Both the camera and laser parameters were kept constant during all measurements reported in this paper.

6. Acquisition Campaign Protocol and Sample Preparation

6.1. Unadulterated sub-dataset

Experimental measurements were carried out on unadulterated milk samples, considering a total of five different samples purchased from local supermarkets, as summarized in Table 1. Three samples of cow milk, pasteurized and sterilized using the UHT (Ultra-High Temperature) process, were analyzed: whole milk, labeled as S1 in Table 1 (produced by Latte Milano – Latteria Soresina, Italy), semi-skimmed milk, labeled as S2 (produced by Sterilgarda Alimenti SpA, Italy), and skimmed milk, labeled as S3 (produced by Sterilgarda Alimenti SpA, Italy). Two additional samples consisted of goat milk: whole goat milk, labeled as S4, and semi-skimmed goat milk, labeled as S5, (both produced by Girau, Italy). Beside the corresponding identification ID label, Table 1 reports the nutritional values of each sample (lipid, protein, and sugar concentrations expressed as g/L). An additional reference sample "R" (powdered dextrose) was included to account for measurement noise and samples variability among campaigns. This material has a granular structure and it is a solid material, thus produces a static SP that can be used as a stable reference condition throughout campaigns.

For each milk type, data acquisition was performed using five different milk jugs. Table 2 reports further details about the milk jugs: the measurement campaigns along with the respective date and time, the room temperature during the experiment (T_{room}) and sample temperature (T_{sam}) recorded after the tests. Each milk jug was measured during the same day or, if necessary, over two consecutive days. The milk jugs were stored in a refrigerator at a temperature of 4 °C and removed a few hours before the experiment to allow the temperature to stabilize close to the room temperature (T_{room}). 2 mL of milk was then taken from the jug with a sterile pipette and moved into the cuvette.

For every campaign and milk sample, a new and clean cuvette was used. The cuvette was first positioned in a 3D-printed mounting (as shown in Fig. 4) and 50 frames were acquired. The resulting SPs were stored in the corresponding folder named "CUV1" (e.g. CAMP1 – S1 – CUV1). Consequently, the cuvette was removed, gently shaken and repositioned in the same location after a few seconds to ensure that the liquid inside the cuvette was stationary. A second acquisition of 50 frames was conducted and the resulting SPs were stored in the corresponding folder named "CUV2" (e.g. CAMP1 – S1 – CUV2). The mounting ensured that the data variability due to the repositioning of the cuvette was kept to a minimum. For both acquisitions, the recording conditions were the same. This procedure was repeated for all liquid samples. For the reference sample, only one cuvette was filled with powdered dextrose and repositioned every time to collect two separate streams of 50 frames each per campaign (CUV1 and CUV2); hence, the reference was the same among all the campaigns. A measurement campaign was considered completed once all samples had been analyzed following this protocol. Each milk jug was tested at two different times, resulting in ten acquisition campaigns (CAMP1–CAMP10). After each measurement session, the milk jugs were returned to the refrigerator until the next campaign. All measurements were performed in dark conditions.

Table 4
List of measurement campaigns for diluted milk samples, including correspondence between the milk jugs from which the five samples were prepared for each campaign, and acquisition details.

Milk jug	Acquisition date	Campaign
1	13/03/2025 12:00 PM	CAMP1
	13/03/2025 4:00 PM	CAMP2
2	19/03/2025 5:00 PM	CAMP3
	20/03/2025 4:00 PM	CAMP4
3	26/03/2025 4:00 PM	CAMP5
	27/03/2025 10:00 AM	CAMP6
4	15/04/2025 12:00 PM	CAMP7
	16/04/2025 3:00 PM	CAMP8
5	08/05/2025 3:00 PM	CAMP9
	08/05/2025 4:00 PM	CAMP10

6.2. Adulterated milk sub-dataset

SP images were also collected by analyzing adulterated milk samples. Specifically, cow milk (produced by Latte Milano – Latteria Soresina, Italy), pasteurized and sterilized using the UHT process was selected and adulterated with different amounts of water or water-glucose solution. Table 3 reports the list of samples used to create this dataset. The first sample corresponds to the pure whole milk (SAD1). The second (SAD2) and the third (SAD3) samples were obtained by diluting the pure milk with 15% and 30% water, respectively, with respect to the total volume. Afterwards, a water–glucose solution was taking into account, dissolving 12.5% glucose in water. The fourth (SAD4) and the fifth (SAD5) were prepared by adding 15% and 30% of a water–glucose solution to the pure milk sample, respectively. As before, an additional reference sample "RAD" (powdered dextrose) was included to account for measurement noise and samples variability among campaigns. This material has a granular structure and it is a solid material, thus produces a static SP that can be used as a stable reference condition throughout campaigns.

Data acquisition was conducted in the exact same way as the "Unadulterated" sub-datset. For every campaign and milk sample, a new and clean cuvette was used. Two acquisitions were conducted using the same cuvette repositioned in the 3D-printed interface, resulting in the 50 SP saved in folder "CUV1" and the other 50 SP saved in folder "CUV2". For both acquisitions, the recording conditions were the same. This procedure was repeated for all liquid samples. For the reference sample, only one cuvette was filled with powdered dextrose and repositioned every time to collect two separate streams of 50 frames each per campaign (CUV1 and CUV2); hence, the reference was the same among campaigns. Since the acquisition campaigns were performed over several days, the preparation of the adulterated milk samples was repeated five times and each set of samples was measured at two different times, following a similar procedure as the unadulterated samples, resulting in ten acquisition campaigns (CAMP1–CAMP10). Table 4 summarizes the details of each sample preparation, including the corresponding whole milk jug (i.e., a fresh jug of pure whole milk opened for each preparation round). The milk jugs were stored in a refrigerator between acquisitions and removed a few hours before the measurements to allow the temperature to stabilize close to the room temperature. All measurements were performed in dark conditions and the room temperature remained approximately constant across all acquisition sessions.

7. Dataset Validation

In the following Sections, we will refer to "data points" indicating the mean gray level intensity of the pixels computed from each SP image. This value is in range 0–255, for which 0 corresponds to no light hitting the pixel (black color) and 255 to a fully saturated white pixel. This is a standard feature typically used in machine learning applications involving SP, as in

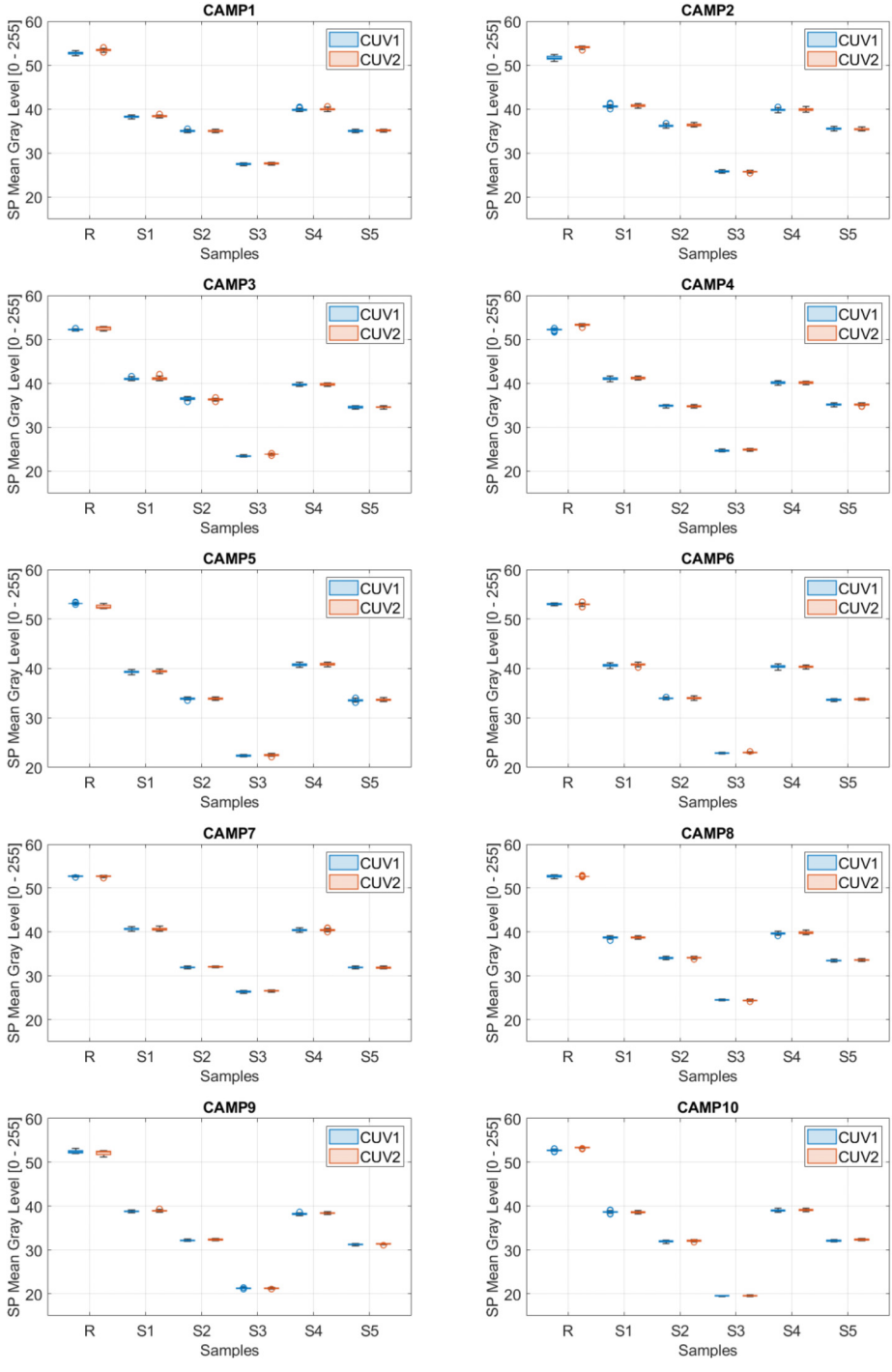


Fig. 6. Box plots of the mean gray level intensity values computed from each SP image of "Unadulterated" sub-dataset. Data is shown grouped per campaign. Blue boxes represent the 50 data points of the first cuvette CUV1, red boxes represent the 50 data points of the second cuvette CUV2.

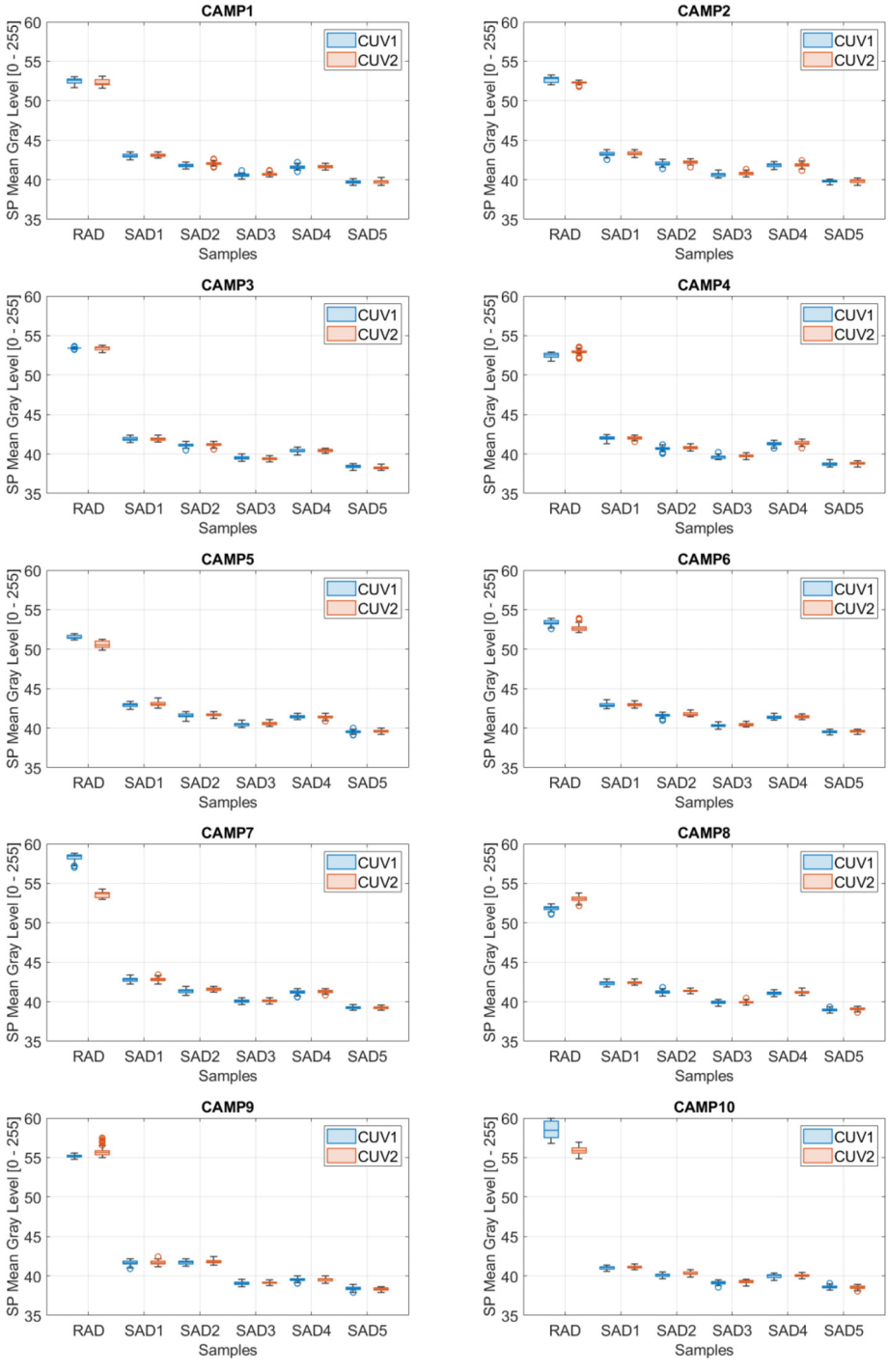
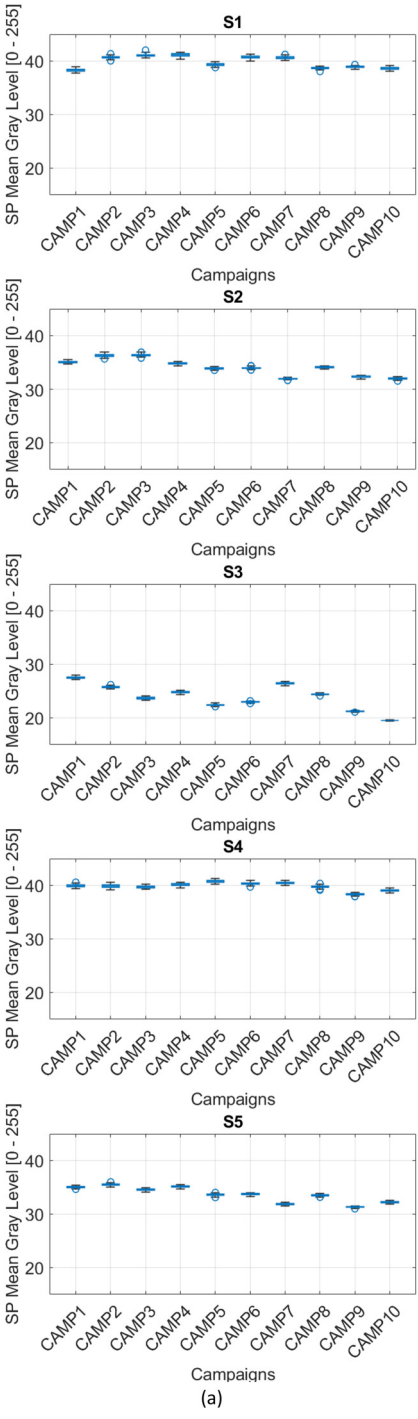


Fig. 7. Box plots of the mean gray level intensity values computed from each SP image of "Adulterated" sub-dataset. Data is shown grouped per campaign. Blue boxes represent the 50 data points of the first cuvette CUV1, red boxes represent the 50 data points of the second cuvette CUV2.

Table 5
For each sample in each campaign of the two sub-datasets, this Table shows the standard deviations of the mean gray level feature computed from the data belonging to the two cuvettes σ_{CUV1} and σ_{CUV2} individually (50 data points each) and the absolute difference among them Δ_{CUVS}.

Campaigns	Sample	Unadulterated sub-dataset			Sample	Adulterated sub-dataset		
		σ_{CUV1}	σ_{CUV2}	Δ_{CUVS}		σ_{CUV1}	σ_{CUV2}	Δ_{CUVS}
CAMP1	**R**	0.282	0.244	0.038	**RAD**	0.366	0.418	0.052
	S1	0.225	0.198	0.027	**SAD1**	0.247	0.211	0.035
	S2	0.199	0.171	0.028	**SAD2**	0.223	0.222	0.002
	S3	0.152	0.169	0.018	**SAD3**	0.210	0.179	0.032
	S4	0.231	0.235	0.004	**SAD4**	0.238	0.194	0.044
	S5	0.175	0.164	0.012	**SAD5**	0.215	0.206	0.009
CAMP2	**R**	0.436	0.236	0.199	**RAD**	0.348	0.183	0.165
	S1	0.238	0.228	0.010	**SAD1**	0.266	0.249	0.017
	S2	0.238	0.225	0.013	**SAD2**	0.232	0.227	0.006
	S3	0.175	0.143	0.031	**SAD3**	0.238	0.209	0.029
	S4	0.250	0.269	0.018	**SAD4**	0.235	0.256	0.021
	S5	0.212	0.191	0.020	**SAD5**	0.204	0.212	0.009
CAMP3	**R**	0.123	0.336	0.213	**RAD**	0.084	0.223	0.139
	S1	0.224	0.284	0.061	**SAD1**	0.240	0.223	0.017
	S2	0.247	0.190	0.057	**SAD2**	0.218	0.219	0.001
	S3	0.104	0.093	0.011	**SAD3**	0.194	0.160	0.034
	S4	0.220	0.199	0.021	**SAD4**	0.200	0.172	0.028
	S5	0.194	0.153	0.041	**SAD5**	0.204	0.178	0.026
CAMP4	**R**	0.201	0.184	0.017	**RAD**	0.309	0.360	0.051
	S1	0.304	0.234	0.070	**SAD1**	0.231	0.178	0.054
	S2	0.186	0.196	0.010	**SAD2**	0.240	0.207	0.034
	S3	0.152	0.130	0.023	**SAD3**	0.200	0.171	0.029
	S4	0.268	0.220	0.048	**SAD4**	0.236	0.240	0.004
	S5	0.194	0.184	0.010	**SAD5**	0.183	0.156	0.027
CAMP5	**R**	0.094	0.351	0.256	**RAD**	0.227	0.429	0.201
	S1	0.211	0.202	0.009	**SAD1**	0.219	0.255	0.036
	S2	0.145	0.170	0.025	**SAD2**	0.271	0.203	0.068
	S3	0.114	0.117	0.003	**SAD3**	0.205	0.187	0.018
	S4	0.246	0.229	0.016	**SAD4**	0.192	0.224	0.032
	S5	0.177	0.159	0.018	**SAD5**	0.182	0.185	0.003
CAMP6	**R**	0.123	0.210	0.086	**RAD**	0.305	0.526	0.221
	S1	0.259	0.245	0.014	**SAD1**	0.266	0.226	0.040
	S2	0.141	0.190	0.049	**SAD2**	0.220	0.209	0.011
	S3	0.093	0.089	0.004	**SAD3**	0.179	0.185	0.006
	S4	0.252	0.193	0.059	**SAD4**	0.198	0.196	0.002
	S5	0.166	0.129	0.038	**SAD5**	0.164	0.171	0.007
CAMP7	**R**	0.112	0.138	0.025	**RAD**	0.488	0.370	0.118
	S1	0.211	0.241	0.030	**SAD1**	0.232	0.251	0.018
	S2	0.133	0.092	0.041	**SAD2**	0.270	0.188	0.081
	S3	0.142	0.125	0.017	**SAD3**	0.199	0.181	0.019
	S4	0.207	0.206	0.000	**SAD4**	0.230	0.185	0.045
	S5	0.142	0.131	0.012	**SAD5**	0.158	0.160	0.002
CAMP8	**R**	0.215	0.050	0.165	**RAD**	0.309	0.369	0.060
	S1	0.207	0.186	0.021	**SAD1**	0.215	0.192	0.022
	S2	0.158	0.133	0.026	**SAD2**	0.236	0.168	0.068
	S3	0.105	0.114	0.009	**SAD3**	0.187	0.177	0.011
	S4	0.204	0.207	0.003	**SAD4**	0.211	0.206	0.005
	S5	0.152	0.146	0.006	**SAD5**	0.165	0.169	0.005
CAMP9	**R**	0.292	0.411	0.119	**RAD**	0.150	0.675	0.525
	S1	0.169	0.154	0.015	**SAD1**	0.280	0.230	0.050
	S2	0.118	0.124	0.005	**SAD2**	0.211	0.233	0.022
	S3	0.082	0.071	0.011	**SAD3**	0.200	0.144	0.055
	S4	0.152	0.140	0.012	**SAD4**	0.198	0.194	0.004
	S5	0.109	0.086	0.023	**SAD5**	0.202	0.185	0.017
CAMP10	**R**	0.154	0.101	0.053	**RAD**	1.021	0.486	0.534
	S1	0.215	0.187	0.028	**SAD1**	0.192	0.167	0.025
	S2	0.151	0.125	0.027	**SAD2**	0.223	0.218	0.005
	S3	0.070	0.067	0.003	**SAD3**	0.205	0.190	0.015
	S4	0.192	0.175	0.016	**SAD4**	0.210	0.164	0.045
	S5	0.128	0.140	0.012	**SAD5**	0.196	0.199	0.003



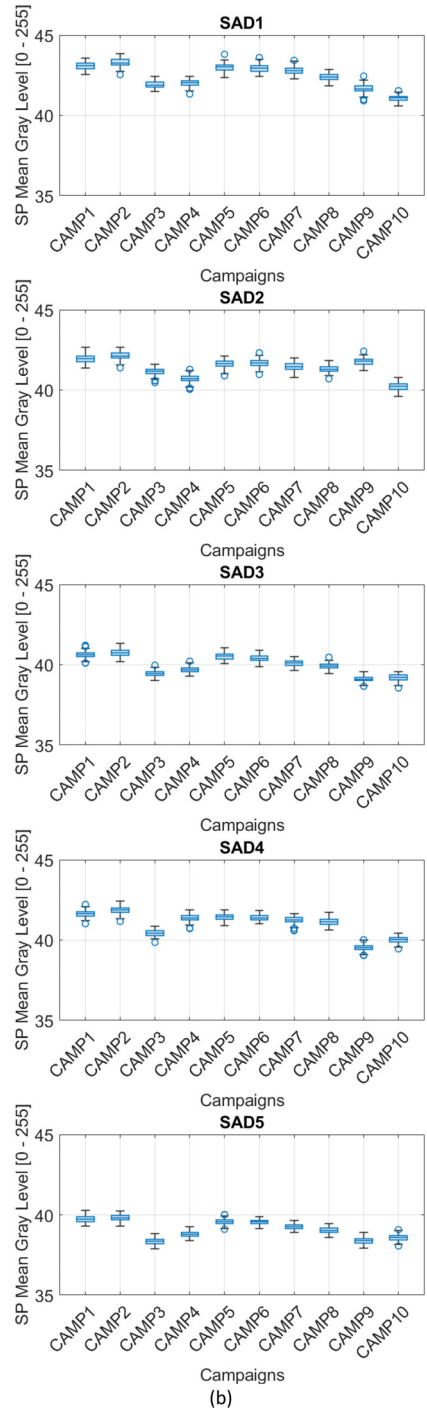


Fig. 8. Box plots of the population of the milk samples grouped per campaign. Each box contains 100 data points. (a) "Unadulterated" sub-dataset, (b) "Adulterated" sub-dataset.

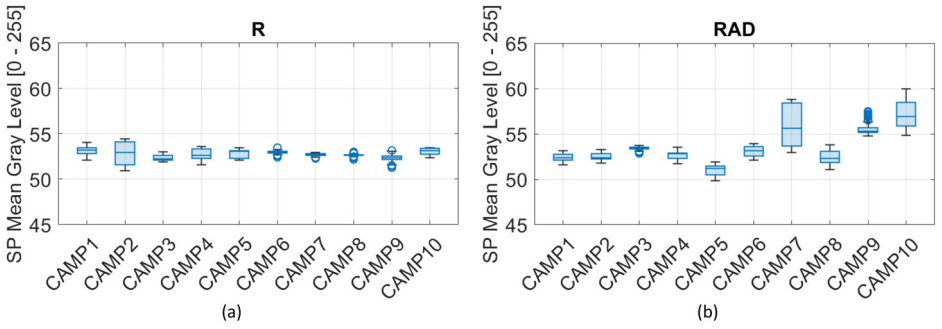


Fig. 9. Box plots of the population of the reference samples grouped per campaign. Each box contains 100 data points. (a) "Unadulterated" sub-dataset, (b) "Adulterated" sub-dataset.

[2,3], that we use to conduct the dataset validation. Researchers can conduct the same tests after computing a different feature of their liking.

7.1. Between-cuvette variability

To assess the variability among cuvettes, we grouped the 50 data points belonging to the two cuvettes individually. Fig. 6 shows the box plots of the data points in the "Unadulterated" sub-dataset, and Fig. 7 the box plots of the data points in the "Adulterated" sub-dataset. To provide a quantitative number, we calculated the standard deviation (σ) of the 50 data points contained in each "CUV" group. Results are shown in Table 5. The columns named σ_{CUV1} and σ_{CUV2} represent the standard deviation of each cuvette individually, while the column named Δ_{CUVS} shows the absolute difference between the two.

7.2. Between-campaign variability

In this case, for each sample in each campaign, we grouped the 50 frames belonging to the two cuvettes "CUV1" and "CUV2" into a bigger group of 100 values. The box plots in Fig. 8 show the population of the mean gray level intensity values of each milk sample according to the measurement campaign. Fig. 8(a) correspond to the data of the "Unadulterated" sub-dataset, Fig. 8(b) to the data of the "Adulterated" sub-dataset. The same information is shown in Fig. 9(a) and Fig. 9(b) for the reference samples of the "Unadulterated" and "Adulterated" sub-datasets, respectively.

7.3. Within-sample variability

To assess the variability of the samples across campaigns, we grouped all the data points corresponding to the same sample across campaigns. This results in a population of 1000 data points (50 frames x 2 cuvettes x 10 campaigns) per sample for each sub-dataset. The box plots of the populations are shown in Fig. 10(a) for the "Unadulterated" sub-dataset, and in Fig. 10(b) for the "Adulterated" sub-dataset.

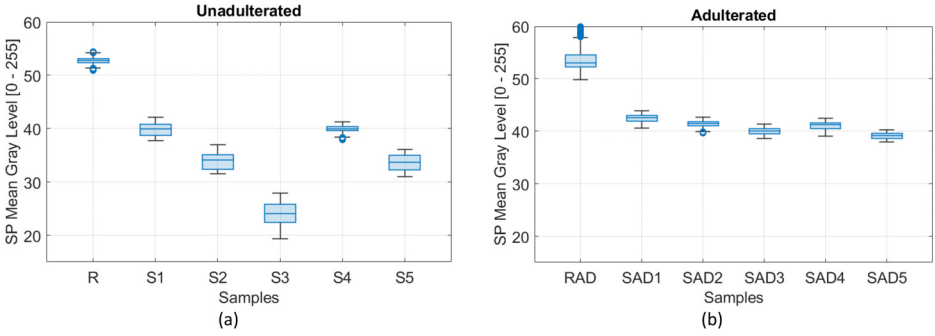


Fig. 10. Box plots of the population of the samples. Each box contains 1000 data points (data from all cuvettes of all campaigns corresponding to the same sample). (a) "Unadulterated" sub-dataset, (b) "Adulterated" sub-dataset.

Limitations

The datasets presented in this article have some limitations that should be considered when interpreting their potential applicability. First, the milk samples were obtained from a limited number of milk jugs purchased at the local supermarkets. As a consequence, the dataset does not capture the full variability of milk properties arising from production by animals raised in different farms and regions, nor does it represent the diversity of brands available on the market. Second, data acquisition was performed over a time span of eight months, which may limit the representation of seasonal variability in milk properties.

Another limitation is related to the origin of the milk. Since the samples were commercially purchased products, no detailed information about the breed of the cows, their diet, or the specific or farming conditions was available. Moreover, the analyzed milk should be considered as an "average" product, likely resulting from the mixing of milk originating from different cows and potentially from different geographical areas.

Finally, for the dataset involving adulterated samples, the temperature of the samples at the time of measurement was not recorded, which may introduce an additional uncontrolled variable.

A final consideration should be made with respect to data standardization. For the purpose of machine learning classification, the features computed from the original SPs must be corrected using the reference SPs to account for differences in the measurement campaigns. These differences are mainly due to instrumentation noise. Readers and users of the datasets are encouraged to adopt specific correction formulas according to their model and data analysis.

Ethics Statement

Authors have read and followed the ethical requirements for publication in Data in Brief. The current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Valentina Bello, Cristina Nuzzi, Simone Pasinetti: Conceptualization. **Valentina Bello, Irene Bassi, Cristina Nuzzi, Simone Pasinetti**: Methodology. **Valentina Bello, Irene Bassi**: Software. **Valentina Bello, Irene Bassi**: Validation. **Valentina Bello**: Formal analysis. **Valentina Bello, Irene Bassi**: Investigation. **Valentina Bello, Irene Bassi**: Resources. **Valentina Bello, Irene Bassi,**

Cristina Nuzzi: Data curation. **Valentina Bello, Irene Bassi, Cristina Nuzzi**: Writing—original draft preparation. **Simone Pasinetti**: Writing—review and editing. **Cristina Nuzzi**: Visualization. **Simone Pasinetti**: Supervision. **Simone Pasinetti**: Project administration. **Simone Pasinetti**: Funding acquisition. All authors have read and agreed to the published version of the manuscript

Data Availability

SP-MILK: a dataset of speckle pattern images of pure and adulterated animal milk (Original data) (Zenodo).

Acknowledgements

This project was supported by the European Union program "Next Generation EU, M4C2 component, investment 1.1", issued as an open call for the national project "PRIN PNRR 2022", grant agreement N° P2022FX228 - MUSKETEER: Milk adUlteration detection using SpecKlE paTtern and machinE lEaRning. The authors wish to thank Prof. Sabina Merlo (University of Pavia) for support and fruitful discussions.

Declaration of Competing Interest

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

References

[1] I. Bassi, V. Bello, C. Nuzzi, S. Pasinetti, SP-MILK: a dataset of speckle pattern images of pure and adulterated animal milk, Zenodo (2026).
[2] C. Nuzzi, S. Pasinetti, I. Bassi, V. Bello, On the applicability of speckle pattern imaging combined with AI for raw milk classification, Measurement 258 (Part C) (2026) 119246.
[3] V. Bello, I. Bassi, M. Fiocchi, C. Nuzzi, P. Bellassi, L. Cattaneo, S. Pasinetti, Milky WAI: unlocking the secrets of raw cow milk through speckle pattern and AI, 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), 2024 St. Albans, United Kingdom21-23 October.
[4] K. Nyandey, D. Jakubczyk, Combining transmission speckle photography and convolutional neural network for determination of fat content in cow milk: an exercise in classification of parameters of a complex suspension, Eur. Phys. J. Plus 139 (123) (2024).
[5] I. Bassi, V. Bello, C. Nuzzi, S. Pasinetti, S. Merlo, The MUSKETEER project: milk adulteration detection using speckle pattern and machine learning, in: Proceedings of SPIE 13527, Optical Sensors 2025, 2025.
[6] I. Bassi, C. Nuzzi, S. Pasinetti, V. Bello, Advancing food safety through AI-driven speckle pattern imaging for milk adulteration detection, in: Proceedings of SPIE, Photonic Technologies in Plant and Agricultural Science III, 2026.
[7] I. Bassi, M. Fiocchi, E. Figus, S. Merlo, V. Bello, Detection of water and glucose adulteration in cow milk by innovative Speckle pattern imaging, IEEE Trans. Instrum. Meas. 74 (2025) 1–9.
[8] A. Mamgain, V. Kumar, S. Dash, Image-based detection of adulterants in milk using convolutional neural network, ACS Omega 9 (25) (2024) 27158–27168.