Data in Brief 59 (2025) 111451

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Dataset for defect detection in textile manufacturing

Beatriz Gil-Arroyo [a,*], Juan Marcos Sanz [b], Ángel Arroyo [a], Daniel Urda [a], Nuño Basurto [a], Álvaro Herrero [a]

[a] Grupo de Inteligencia Computacional Aplicada (GICAP), Departamento de Digitalización, Escuela Politécnica Superior, Universidad de Burgos, Av. Cantabria s/n 09006, Burgos, Spain
[b] Textil Santanderina, Avda. Textil Santanderina S/N, 39500 Cabezón de la Sal (Cantabria), Spain

ARTICLE INFO

Article history:
Received 31 January 2025
Revised 25 February 2025
Accepted 3 March 2025
Available online 7 March 2025

Dataset link: Original and processed dataset of Batavia and Sarga woven fabric images (Original data)

Keywords:
Textile manufacturing
Textile industry
Batavia and Sarga fabric
Defect detection
Image analysis
Artificial Vision
Quality inspection

ABSTRACT

This dataset, collected during November 2022 at Textil Santanderina, a leading textile manufacturer based in Cabezón de la Sal (Cantabria, Spain), comprises high-resolution images of Batavia and Sarga fabrics. The images were captured as part of a project to document and analyze the intricate weaves and patterns of these fabrics. Using a high-resolution camera under controlled lighting conditions, detailed images were obtained to ensure consistent quality and accurate representation of the fabric's texture and colour. The dataset is provided in processed format, where images have been downscaled from 16 bits to 8 bits, cropped, and classified into cases and controls. The primary reuse potential of this dataset lies in its application for Artificial Intelligence (AI) and Machine Learning (ML) models aimed at defect detection in textile manufacturing. By leveraging these high-quality processed images, researchers and developers can train models to identify and classify various types of fabric defects, such as weave inconsistencies, colour variations, and surface irregularities. This can significantly enhance the efficiency and accuracy of quality control processes in textile production. Additionally, the dataset serves as a valuable resource for academic research in textile engineering and material science. It can be used to study the properties and behaviours of Batavia and Sarga weaves under different conditions, con-

* Corresponding author.
E-mail address: bgarroyo@ubu.es (B. Gil-Arroyo).

https://doi.org/10.1016/j.dib.2025.111451

tributing to advancements in fabric design and manufacturing techniques. The detailed visual information provided by the processed images also supports the development of new methodologies for automated textile inspection and quality assurance. By making this dataset available, Textil Santanderina and University of Burgos aim to support innovation and improvement in textile quality control through AI-driven solutions, fostering collaboration and development within the industry.

Specifications Table

Subject	Computer Sciences
Specific subject area	Textil Santanderina: Experts in high-quality fabrics since 1923, committed to global innovation and sustainability.
Type of data	Processed images.
Data collection	The imaging system employs a Basler raL camera with the Awaiba DR-12k-3.5 CMOS sensor, with a resolution from 8 kHz to 12k [1]. Operating within the VIS-NIR bandwidth, with resolution of 10 pixels/mm. A LED array emitting at 850 nm is utilized, covering a field of 15 mm × 1510 mm, situated at a height of 20 cm above the tissue, comprises 360 infrared (IR) LEDs at 850 nm wavelength, each with an incident angle set at 15 degrees. This IR illumination prevents interference with existing standard D65 illumination systems commonly used in visual inspection.
Data source location	The data were collected in: Institution Textil Santanderina City/Town/Region: Cabezón de la Sal, Cantabria Country: Spain Geographical coordinates: 43.31164220450848, -4.225763758747999
Data accessibility	Repository name: Institutional Repository of the University of Burgos Data identification number: *DOI:* 10.36443/10259/9965 Direct URL to data: 10.36443/10259/9965 [2]
Related research article	B. Gil-Arroyo, N. Velasco-Pérez, J.M. Sanz, A. Casas, A. Arroyo, D. Urda, Defect detection in Batavia and Sarga woven fabrics by means of convolutional neural networks, J. Appl. Logics, submitted for publication, 2025 [3].

1. Value of the Data

- This dataset facilitates the integration of artificial intelligence and computer vision in the textile industry, serving as a valuable resource to train models for real-time defect detection in textile patterns. By incorporating machine learning models trained on this dataset into production lines, it enhances the efficiency and objectivity of defect detection, improving the ability to identify issues before final products leave the factory.
- The dataset provides a solid foundation for developing and training more advanced machine learning models such as categorizing defects by typology or classifying different twill weave patterns.
- The dataset could be valuable in training and coaching new employees in the textile industry.
- Production process experts can identify patterns and areas for improvement in machinery and product quality.
- Researchers exploring applications beyond the textile industry, such as image recognition and pattern analysis, can benefit from the diverse and well-labeled twill fabric dataset.

2. Background

In Spain, the textile industry plays a crucial role in both the economy and cultural heritage [4]. As a significant contributor to employment and Gross Domestic Product (GDP), it fosters innovation and sustainability in fabric production [5]. Machine learning (ML) is revolutionizing this sector by enhancing quality control and fault detection processes [6,7,8,9]. Algorithms can analyze vast datasets to predict and prevent defects in textiles, ensuring higher product standards and reducing waste. Moreover, ML applications extend beyond quality assurance; they optimize production processes by predicting demand, managing inventory efficiently, and streamlining supply chains. This predictive capability not only boosts productivity but also minimizes costs and environmental impact through better resource utilization. Looking ahead, advancements in ML-driven automation promise further improvements in textile manufacturing, from automated looms to smart factories. By integrating these technologies, Spain's textile industry can maintain its competitive edge globally, embracing sustainability and innovation while meeting evolving consumer demands. This convergence of tradition and technology underscores the pivotal role of textiles in Spain's industrial landscape, driving economic growth and sustainable development into the future.

3. Data Description

The dataset has two different weaves (Batavia and Sarga). These textiles were selected due to their distinctive characteristics and relevance in various textile applications.

Batavia fabric, a type of twill weave within Sarga fabrics, blends 85% cotton and 15% linen. In November 2022, 2755 high-resolution grayscale images (16-bit, 2048 $\times$ 696 pixels) were collected, focusing on a neutral greige fabric.

Sarga fabric, known for its diagonal weave pattern, offers durability and versatility for various textile applications. In November 2022, 1548 high-resolution grayscale images (16-bit, 2048 $\times$ 696 pixels) were gathered, featuring a neutral greige fabric to minimize colour variability.

The dataset is organized by two types of woven fabrics: Batavia and Sarga, each containing images in PNG format. Within each fabric category, there are specific directories:

1. Batavia:
   - Originals: Contains 2755 8-bit PNG images of size 2048 $\times$ 696 pixels.
   - Patches:
     - Cases: Includes 8782 cropped PNG images (365 $\times$ 365 pixels) with defects.
     - Controls: Includes 19,911 cropped PNG images (365 $\times$ 365 pixels) without defects.
     - info_patches.csv: A CSV file with columns for image and patch names, operator's and ground truth labels (0 for no defect, 1 for defect).
     - Note: In Batavia dataset, the total number of patches does not equal the number of originals multiplied by 12 due to a review by Textil Santanderina. During this process, where 4367 ambiguous patches were excluded as their labels (case/control) were unclear.
2. Sarga:
   - Originals: Contains 1548 8-bit PNG images of size 2048 $\times$ 696 pixels.
   - Patches:
     - Cases: Includes 173 cropped PNG images (365 $\times$ 365 pixels) with defects.
     - Controls: Includes 18,403 cropped PNG images (365 $\times$ 365 pixels) without defects.
     - info_patches.csv: A CSV file structured similarly to Batavia's file, providing image and patch details along with defect labels.
     - Note: In Sarga dataset, the total number of patches equals the number of originals multiplied by 12.

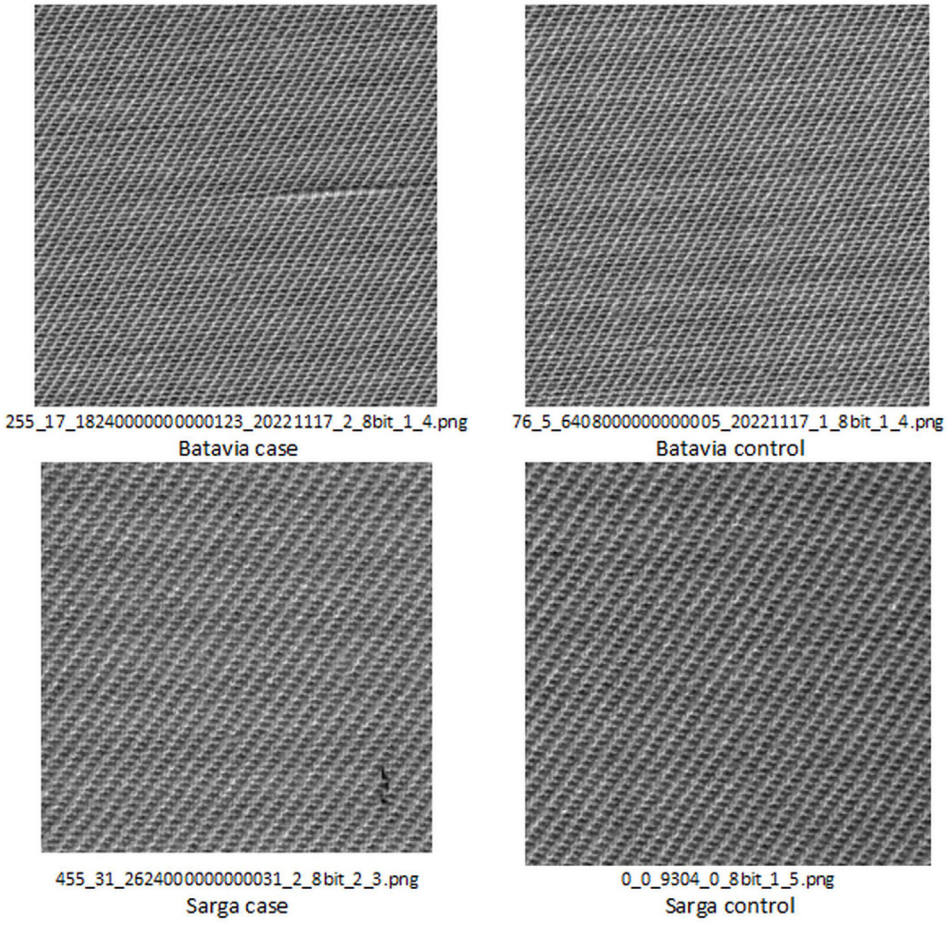

Fig. 1. Examples of patches images.

The directories patches contain the 365 × 365 cropped images with small overlapping areas on both the horizontal and vertical axis.

In Fig. 1, examples of patches can be observed. For each of the two fabrics, there is an example of an image with a defect (case) and another without a defect (control).

The PNG file naming convention follows:

- Original images: Identified by a numeric identifier, the original folder number from Textil Santanderina and followed by the suffix “8bit” For example: 0_0_9304_20221117_0_8bit.png
  - 0_0_9304 is the numeric ID.
  - 20221117 is the acquisition date of the image, in this example November 17, 2022. In Batavia, all the images were taken on November 17, 2022, while in Sarga, this date does not appear in the image name, but they were also taken in November 2022.
  - 0 is the original folder. The range goes from 0 to 2
  - 8bit is the suffix.
  - png is the file extension.
- Patches: Named using the original image name, row, and column indices, referencing the 2 × 6 matrix slicing of the original images into 12 segments of size 365 × 365 pixels, with a small overlap in both the x and y axes. For example: 0_0_9304_20221117_0_8bit_2_1.png:

Fig. 2. 2 × 6 matrix slicing.

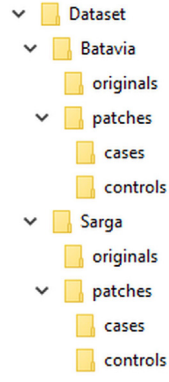

Fig. 3. Directory structure.

Table 1
Slicing details.

Patch	Coordinates (x,y)	Width (pixel)	Height (pixels)
1_1	(0,0)	365	365
1_2	(0, 336)	365	365
1_3	(0, 672)	365	365
1_4	(0, 1008)	365	365
1_5	(0, 1344)	365	365
1_6	(0, 1680)	365	365
2_1	(0, -330)	365	365
2_2	(336, -330)	365	365
2_3	(672, -330)	365	365
2_4	(1008, -330)	365	365
2_5	(1344, -330)	365	365
2_6	(1680, -330)	365	365

- 0_0_9304_20221117_0_8bit is the original image name.
- 2 is the row of the matrix.
- 1 is the column of the matrix.

In the process of slicing the original images, the 2048 × 696 pixel images were divided into 12 segments of 365 × 365 pixels each, following a 2 × 6 matrix with 2 rows and 6 columns. This slicing was carried out to facilitate analysis and experimentation. Fig. 2 illustrates this 2 × 6 matrix slicing of the original 2048 × 696 images into 12 patches of size 365 × 365 pixels each.

Table 1 presents the specific details of this slicing, including the starting coordinates of each patch within the original image. The point (0,0) is considered to be the top-left corner of the

Table 2
Dataset's folder structure and number of images.

Folder	N° images
Batavia\originals	2755
Batavia\patches\cases	8782
Batavia\patches\controls	19,911
Sarga\originals	1548
Sarga\patches\cases	173
Sarga\patches\controls	18,403

2048 × 696 image. The coordinates (x, y) correspond to the top-left corner of each patch. The slicing includes an overlap of 29 pixels in the x-axis and 35 pixels in the y-axis.

In Fig. 3, the directory structure of the dataset is shown. This fig. illustrates the hierarchical organization of directories and files comprising the dataset. The structured design facilitates efficient management and access to the dataset samples, essential for conducting any analysis. The directory structure is organized by fabrics (Batavia and Sarga). Within each fabric, there are original images of size 2,048 × 696 pixels and patches. Inside the 'patches' folder, there are two subfolders: 'cases' and 'controls', and a CSV file named 'info_patches' containing the detailed information mentioned above.

Table 2 presents an overview of the dataset's folder structure and the corresponding number of images included in each category. The dataset is organized into two main types of fabrics, Batavia and Sarga, each containing original images and patches. This systematic categorization is essential for subsequent research and detailed analysis of the dataset, enabling comprehensive exploration and understanding of the dataset's contents and distributions under different experimental conditions.

4. Experimental Design, Materials and Methods

The system was configured to capture images of Batavia and Sarga in November, 2022.

A total of 2755 images of Batavia weave and 1548 of Sarga weave were obtained. All images were 8-bit 2048 × 696 pixels, instead of 16 bits. This reduction was mainly done to decrease the file size; 16-bit images can represent 65,536 different colours, while 8-bit images have a lower colour depth limited to 256 colours. This reduction in the number of colours can affect the visual quality of the images, but if the additional colour information is not critical for the research, this conversion can simplify image processing.

Each original 8-bit 2048 × 696 image was divided into 12 parts of 365 × 365 pixels with a small overlap in the vertical and horizontal axes. Thus, for Batavia tissue, 28,693(the total number of patches does not equal the number of originals multiplied by 12 due to a review by Textil Santanderina) sliced images were obtained, of which 19,911 correspond to controls and 8782 to cases. Accordingly, 18,576 (12 × 1548) Sarga fabrics 365 × 365 patches were acquired, of which 18,403 controls and 173 cases.

We used different python libraries to image pre-processing, so we can obtain 12 patches from each 2048 × 696 8-bit image. The python libraries used have been:

- Os: for manipulating paths and directories [10].
- Glob: for finding file names that match a search pattern [11].
- PIL (Pillow): for opening, manipulating, and saving images [12].
- Shutil: for file and directory manipulation operations, such as copying and moving [13].

The raw data, pre-processed images, and image pre-processing code has been uploaded to GitHub, given that offers a centralized space for collaborative software development, providing essential tools for version control, project management, and collaboration between developers [14].

Limitations

Not applicable.

Ethics Statement

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

CRediT Author Statement

Beatriz Gil-Arroyo: writing - Original Draft, Visualization. **Juan Marcos Sanz**: Validation, Resources. **Ángel Arroyo**: Formal analysis, Investigation, supervising, writing – Review & Editing. **Daniel Urda**: Methodology, Software, Data Curation, Project administration, writing – Review & Editing. **Nuño Basurto**: Review & Editing. **Álvaro Herrero**: Conceptualization, Supervision, Funding acquisition.

Data Availability

Original and processed dataset of Batavia and Sarga woven fabric images (Original data) (University of Burgos. Public Repository).

Acknowledgements

The funding for this project was provided by the DECENT (Deep Learning for automatic Textile Inspection) initiative under the DIH-World 2nd Open Call framework. The authors express their gratitude to INADE for their collaboration in acquiring the images.

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] The Basler racer raL12288-8gm [website], Basler, 2024. https://www.baslerweb.com/en/shop/ral12288-8gm/, 2024 (Accessed 20 February 2025).
[2] B. Gil-Arroyo, N. Velasco-Pérez, N. Basurto, J. M. Sanz, D. Urda, A. Herrero, Original and processed dataset of Batavia and Sarga woven fabric images, , UBU Repository, v1, 2025. https://doi.org/10.36443/10259/9965.
[3] B. Gil-Arroyo, N. Velasco-Pérez, J.M. Sanz, A. Casas, A. Arroyo, D. Urda, Defect detection in Batavia and Sarga woven fabrics by means of convolutional neural networks, J. Appl. Logics (2025) submitted for publication.
[4] H. Oevermann, B.M. Walczak, M. Watson, The Heritage of the Textile Industry, Lodz University of Technology Monographs, 2022, doi:10.34658/9788366741232.
[5] Ministerio de Industria, Comercio y Turismo, Informe sectorial: Textil y confección [website], 2024. https://www.mintur.gob.es/es-es/IndicadoresyEstadisticas/Presentaciones/20sectoriales/02./20Textil/20y/20confeccion.pdf, (Accessed 20 February 2025).
[6] P.R. Jeyaraj, E.R.S. Nadar, Effective textile quality processing and an accurate inspection system using the advanced deep learning technique, Text. Res. J. 90 (9-10) (2020) 971–980, doi:10.1177/0040517519884124.

[7] B. Wei, B. Xu, K. Hao, L. Gao, Textile defect detection using multilevel and attentional deep learning network (MLMA-Net), Text. Res. J. 92 (19-20) (2022) 3462–3477, doi:10.1177/00405175211073773.
[8] I. Koulali, M.T. Eskil, Unsupervised textile defect detection using convolutional neural networks, Appl. Soft Comput. 113 (2021) 107913, doi:10.1016/j.asoc.2021.107913.
[9] H. Zhang, W. Zhang, Y. Wang, S. Lu, L. Yao, X. Chen, Colour-patterned fabric-defect detection using unsupervised and memorial defect-free features, Color. Technol. 138 (6) (2022) 602–620, doi:10.1111/cote.12624.
[10] Python Software Foundation, os — Miscellaneous operating system interfaces [software], 2024. https://docs.python.org/3/library/os.html, (Accessed 20 February 2025).
[11] Python Software Foundation, glob — Unix style pathname pattern expansion [software], 2024. https://docs.python.org/3/library/glob.html, (Accessed 20 February 2025).
[12] Python Software Foundation, Pillow (PIL Fork) documentation [software], 2024. https://pillow.readthedocs.io/en/stable/, (Accessed 20 February 2025).
[13] Python Software Foundation, shutil — High-level file operations [software], 2024. https://docs.python.org/3/library/shutil.html, (Accessed 20 February 2025).
[14] Gicap-UBU, dif_textile [software], 2025. https://github.com/gicap-ubu/dif_textile#contents, (Accessed 20 February 2025).