INTERPRETABLE MICROTEXTURE FEATURE EXTRACTION FOR FABRIC CLASSIFICATION IN TEXTILE RECYCLING SYSTEMS

MOLCHANOVA, MARYNA*; MAZURETS, OLEKSANDR; ZALUTSKA, OLHA; ZAKHARKEVICH, OKSANA; SOBKO, OLENA; BUKHANTSOVA, LIUDMYLA; ORTYNSKYI, VADYM; KOSHEVKO, YULIIA; LOVINSKYI, MAKSYM AND KOZHEVNIKOV, SERHIY

Khmelnytskyi National University, 11, Instytutska str., Khmelnytskyi, 29016, Ukraine

ABSTRACT

In textile waste management and the circular economy, reliable discrimination between natural and synthetic fabrics can support sorting and recycling. Manual inspection is labor-intensive and subjective, while many automated approaches rely on opaque image-based models. This paper presents a reproducible and interpretable method for binary natural-versus-synthetic fabric discrimination from microimages acquired under controlled conditions. The proposed approach represents each image by a fixed numerical profile of 111 interpretable microtexture features rather than raw pixels. Each input image is resized to 256×256 pixels, converted to grayscale, and represented at three automatically defined centered fields of view (96×96, 160×160, and 256×256). The profile includes GLCM descriptors, local binary pattern features, photometric intensity characteristics, Shannon entropy, and gradient-based edge measures. Statistical tests and effect sizes are used for post-hoc characterization and ranking of interpretable indicators; they are not used to select the input variables of the classifier. The source dataset contains 3,107 images (1,547 natural and 1,560 synthetic). The experiment reported in this paper uses a reproducible balanced subset of 400 images, with 200 images per class selected using a fixed random seed (42). The subset is divided by stratified image-level splitting into 280 training images and a held-out test set of 120 images (60 per class). The predictive pipeline is fitted using the training partition only; S_{63}/C_{15} statistics are reported separately as post-hoc descriptive characterization and do not determine the MLP inputs or the reported test metrics. Of the 111 predefined features, 63 show significant class-distribution differences after FDR correction in this descriptive analysis. A multilayer perceptron trained on the complete 111-feature profile achieves Accuracy = 0.9917 and Macro-F_1 = 0.9917 on the held-out 120-image test set. Median imputation and feature standardization are fitted within the training pipeline only. The resulting confusion matrix contains 59 correctly classified natural images, one natural image classified as synthetic, and all 60 synthetic images correctly classified. The threshold-based E_{14} module is a separate post-hoc interpretation layer; its class ranges and thresholds are calculated from training data and do not constitute or modify the MLP decision boundary. The results demonstrate strong image-level discrimination within the studied acquisition protocol, but they should not be interpreted as specimen-level validation or as a complete industrial textile-identification system. The published image metadata do not preserve physical-specimen identifiers, so specimen-level independence between partitions cannot be verified retrospectively. In addition, photometric features are highly discriminative and may partly encode illumination/reflectance differences. The current task is binary and does not address blends, fibre-content estimation, or multi-class sorting. External validation across independent specimens and imaging conditions is therefore required.

KEYWORDS

Textile recycling; Fabric classification; Microtexture analysis; Interpretable machine learning; Textile waste sorting.

INTRODUCTION

The transition of the textile sector towards circular economy models requires effective strategies for valorising waste through recycling, reuse, and repurposing. It is essential to understand the various post-industrial textile waste streams to identify suitable methods for them. Different types of textile residues vary in fiber composition, structure, and finishing treatments [1]. Consequently, accurately identifying these characteristics is crucial for selecting appropriate recovery technologies and ensuring the quality of recycled materials.

* **Corresponding author:** Bukhantsova L., e-mail: *bukhantsovalv@khmnu.edu.ua*

Received July 6, 2026; accepted September 21, 2026

Textile waste represents a significant resource for promoting circularity in the textile and fashion industry. Diverting them from landfills not only reduces environmental burdens but also conserves virgin raw materials, thereby alleviating pressure on finite natural resources. Furthermore, using recycled materials can provide economic benefits by lowering production costs compared to sourcing virgin fibers [2] [3].

The diversity of textile waste streams also creates opportunities for innovation in material development, production technologies, and product design, fostering new applications and higher-value uses. By recognising textile waste as a resource rather than merely a disposal issue, the industry can embrace circular economy principles and turn environmental challenges into both ecological and economic advantages [4].

Additionally, the increased use of recycled materials reduces energy consumption and greenhouse gas emissions, significantly lowering the environmental impact of textile production. However, the successful utilisation of post-industrial textile waste depends on accurately identifying its material composition and structural characteristics, which determine suitable recycling pathways and the quality of secondary raw materials [2].

Automated textile sorting in recycling systems depends on accurate identification of fibre types and fabric structures from image data. The accuracy of this recognition influences the choice of processing methods, the consistency of recycled material properties, and quality control throughout the sorting process [5] [6]. A major challenge is the complex and variable nature of textile microtextures, which are affected by imaging scale, lighting conditions, sample orientation, deformation, and batch differences [7]. In practice, textile sorting is a multi-stage process, often relying on manual or semi-automated steps based on labels and operator experience, which limits throughput, hampers standardisation, and raises error rates, particularly for blended materials [8]. Additionally, regulatory measures such as the European Union's requirement to separate textile waste from 2025 onwards highlight the growing importance of high-performance automated sorting systems [9].

State-of-the-art industrial solutions increasingly utilise sensor-based approaches. Systems like FiberSort [10], TOMRA AUTOSORT [11] and Valvan sorting lines [12] use near-infrared spectroscopy, machine vision, and artificial intelligence to identify fiber composition and enable automated classification at industrial scales. While these technologies offer high throughput and improved efficiency, they also have notable limitations. These include reduced accuracy for blended materials, sensitivity to colorants and surface conditions, high investment costs, and limited interpretability of classification results. Additionally, spectral methods primarily capture chemical composition but provide insufficient information about the structural properties that influence recycling behaviour.

In this context, the automated recognition of textile composition and properties becomes a critical component of scalable recycling, particularly for streams containing blended fibers and diverse surface finishes [13].

In contemporary research, neural network models are widely used for fabric classification; they are trained directly on images and are capable of achieving high accuracy given sufficient data [14]. For tasks related to fabric texture and microimages, both deep texture representations and the combination of classical microtexture descriptors with machine learning algorithms have proven effective, providing a controllable and reproducible feature profile.

In parallel with visual approaches, sensor-based methods grounded in spectral measurements are widely used in textile sorting, particularly near-infrared spectroscopy and hyperspectral imaging, which enable identification of the polymer nature of fibers based on their spectral signatures. It has been shown that combining near-infrared spectroscopy with machine learning algorithms and neural network models provides accurate classification of common fibers in post-consumer streams and can be integrated into automated sorting lines [8].

A separate practically important task is the determination of the proportion of synthetic content in blended materials, for which regression and classification models based on spectral features are used, particularly for polyester-containing mixtures [15] [16]. It has also been demonstrated that camera-based solutions in the near-infrared range can provide stable discrimination metrics, while the issue of dark dyes, especially carbon black, is explicitly accounted for in measurement schemes and can be diagnosed from sensor data [17]. For scenarios requiring more detailed identification and composition assessment, hyperspectral approaches and spectral processing are employed, which increase informativeness but require careful calibration and strict control of acquisition conditions and material surface state [18] [19].

At the same time, for practical recycling systems, not only accuracy but also interpretability is important: an engineer needs human-readable information about which measurable texture or photometric properties are associated with a prediction and where potential errors may arise [20]. A separate line of research addresses post-hoc interpretation of neural-network predictions, for example through importance maps of image regions [21]. Such methods identify influential image regions, whereas engineering use may additionally require interpretation in terms of measurable material descriptors.

Within the scope of this study, interpretability is treated as an engineering requirement rather than as a reconstruction of the MLP's internal decision mechanism. Accordingly, expert verification during data preparation and formulation of the post-hoc interpretation rules focused on label plausibility and on whether the selected descriptors correspond to observable structural or photometric fabric characteristics.

Interpretable approaches are traditionally based on texture analysis through statistical characteristics and frequency representations. Statistical measures include contrast, homogeneity, entropy, and correlation properties computed from gray-level spatial co-occurrence patterns [22]. Local texture descriptors include local binary patterns, which provide robustness to monotonic illumination changes and are suitable for constructing multi-scale feature profiles [23]. Frequency-based approaches include the use of Gabor filters and related frequency measures that capture structural anisotropy, dominant spatial frequencies, and the regularity of weave elements. An important practical requirement is multi-scale representation: the same fabric exhibits different discriminative properties at different levels of detail, therefore features must be constructed across multiple scales with clearly defined normalization rules.

Despite the high utility of spectral methods for identifying polymer composition, practical recycling systems often require additional information about the structural properties of the fabric, including heterogeneity, weave regularity, and local defects that affect material behavior during shredding, opening, and fiber reformation. Visual analysis of microtexture directly reflects the spatial organization of yarns and fibers, but it requires features that are robust to scale, illumination, and orientation, as well as features with clear physical meaning that are suitable for production-level interpretation during quality control.

This paper addresses the extraction and use of interpretable microtexture features for binary fabric classification in recycling-oriented applications. Experimental validation is based on the open “Natural and synthetic fabrics dataset”, created by the authors and published on Kaggle [24]. The source dataset contains 3,107 images, whereas the computational experiment reported here uses a reproducible balanced subset of 400 images (200 per class). The problem is deliberately limited to image-level discrimination between the “natural” and “synthetic” classes under the available controlled acquisition protocol.

The objective of this study is to develop a reproducible method for extracting interpretable microtexture features suitable for neural network–based fabric classification and to experimentally validate the method on the specified dataset. The contributions of the work are defined as follows:

- a binary fabric classification approach is proposed in which the neural network is trained on a fixed 111-dimensional profile of interpretable microtexture features rather than on raw images;
- a separate post-hoc interpretation table is constructed from training data, containing class-specific P10–median–P90 ranges and simple threshold-based consistency indicators that accompany, but do not determine, the MLP prediction;
- an annotated dataset of fabric images with a fixed labeling structure is prepared and published on the Kaggle platform, ensuring reproducibility of experiments;
- the feasibility of post-hoc feature-level interpretation is demonstrated through the developed software, which, alongside the model prediction, reports features that are more or less consistent with the predicted class and helps identify borderline and misclassified cases.

To minimize the risk of assigning technological meaning to purely numerical patterns, expert verification was used at key stages. From the computer-science perspective, the reproducibility of feature computation, the train/test separation, the training-only preprocessing pipeline, and the analysis of borderline and misclassified cases were checked. From the textile-engineering perspective, label plausibility and the correspondence of key features to observable structural or photometric manifestations were verified on control examples. The available published dataset is organized at image level and does not retain physical-specimen identifiers; therefore, the present validation is explicitly treated as image-level internal validation rather than specimen-level validation.

METHODOLOGY

General scheme of the method and principles of interpretability

The study proposes a reproducible method for extracting interpretable microtexture features for neural-network-based binary fabric classification. The input to the classifier is a vector of measurable descriptors with explicit physical or photometric meaning: GLCM texture statistics, LBP micropattern descriptors, photometric intensity characteristics, Shannon entropy, and gradient-based edge indicators. The interpretation layer is deliberately separated from the classifier: it provides post-hoc consistency information in the language of these descriptors and is not intended to reconstruct the internal decision mechanism of the MLP.

The workflow is illustrated in Fig. 1. The predictive branch fixes the 111-feature representation before data splitting, performs a stratified 70/30 image-level split, fits median imputation, standardization, and the MLP using the training partition only, and evaluates the fitted pipeline on the held-out predictive test set. Separately, S_{63}/C_{15} statistics provide post-hoc descriptive characterization of class differences in the complete 400-image experimental subset after the predictive protocol is fixed; they do not select MLP inputs or influence performance calculation. The E_{14} numerical ranges and thresholds applied to held-out examples are estimated from the training partition only. The post-hoc interpretation rules do not feed back into model fitting or alter the predicted class.

The model-development pipeline distinguishes deterministic feature extraction, supervised classifier fitting, and post-hoc interpretation. The classifier receives the complete predefined 111-feature profile, and the feature definitions are fixed independently of class labels. Statistical significance and effect-size analyses are used only for descriptive characterization and ranking after the predictive protocol is fixed; they do not reduce, weight, or tune the MLP input. Thus, held-out labels are not used for any operation that can affect the fitted MLP or the reported test metric. E_{14} numerical references applied to held-out images are derived from the training partition only. The method components and the key parameters fixed in the study are summarized in Table 1.

Step 1. Image standardization is deterministic. Each RGB microimage is resized to a base size of 256×256 pixels using bilinear interpolation and then converted to an 8-bit grayscale representation. No manually selected ROI is used in the computational pipeline reported here.

Step 2. Three levels of spatial context are generated automatically from the standardized 256×256 image as centered square fields of view: *s*0 = 96×96, *s*1 = 160×160, and *s*2 = 256×256. Thus, exactly one centered ROI is obtained for each field of view for every image. The three sizes combine a local field of view (*s*0), an intermediate field (*s*1), and the full standardized field (*s*2), allowing the same descriptor families to be compared across progressively broader spatial context.

Step 3. A fixed set of 37 interpretable indicators is computed for each of the three centered fields of view, giving 111 features per image. The groups comprise 12 aggregated GLCM descriptors, 11 LBP-related indicators, 9 photometric intensity statistics, one Shannon-entropy value, and four gradient/edge indicators. This complete 111-dimensional vector is the input representation of the MLP.

Step 4. The classifier is trained on all 111 predefined features. A scikit-learn Pipeline performs median imputation, standardization, and MLP fitting using the training partition. The MLP has two hidden layers with 64 and 32 neurons, ReLU activation, Adam optimization, *L*2 regularization *alpha* = 1×10^{-4}, *initial learning rate* = 1×10^{-3}, maximum 600 iterations, and

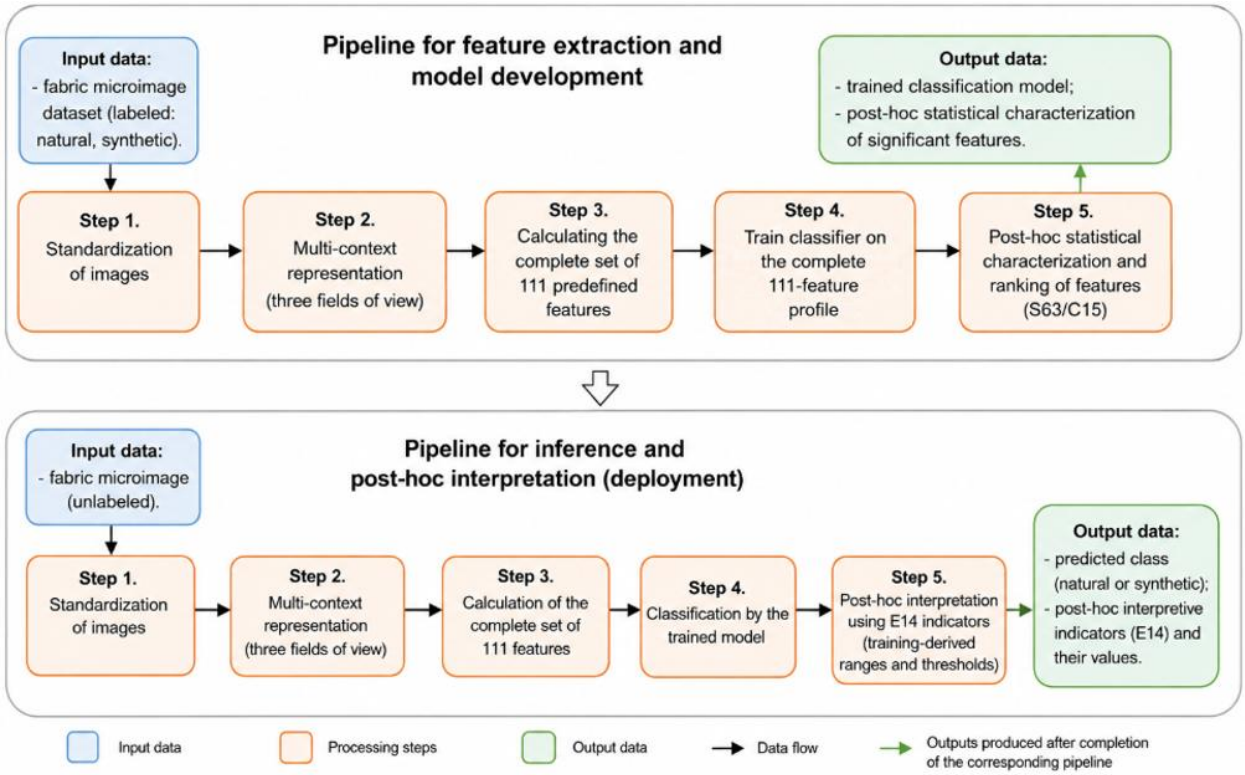


Figure 1. Functional organization of model development, inference, and post-hoc interpretation.

random_state = 42. After model fitting, training-derived percentile ranges and median-midpoint thresholds are stored only for post-hoc interpretation.

Step 5. Statistical analysis characterizes which predefined indicators differ most strongly between classes. Mann–Whitney *U* and Kolmogorov–Smirnov tests, Benjamini–Hochberg FDR correction, and Cohen's *d* are calculated. In the reported S_{63} set, significance is defined by the FDR-adjusted Kolmogorov–Smirnov criterion $q < 0.05$; no non-zero minimum |*d*| threshold is used for inclusion ($d_{min} = 0.0$). Cohen's |*d*| is subsequently used to rank the strongest effects (C_{15}). This statistical analysis is an interpretive/descriptive layer and is not used to select the MLP input features.

For inference, a new or held-out image undergoes the same deterministic image standardization, centered field-of-view construction, and 111-feature extraction. The fitted imputer, scaler, and MLP are then applied to obtain the class prediction. Only after the MLP prediction is produced is the separate E_{14} interpretation layer applied; it compares selected measured features with training-derived class ranges and thresholds. Consequently, threshold checks cannot influence the neural-network prediction. The steps of the deployment pipeline and their key parameters are summarized in Table 2. The interpretation output reports the measured values of selected features, their relation to the training-derived P10–median–P90 ranges, and the corresponding median-midpoint thresholds. This output should be read as an engineering interpretation aid rather than as a representation of the MLP's internal weights or nonlinear decision boundary.

The proposed organization separates prediction from interpretation. The MLP provides the predictive decision, whereas the E_{14} table provides post-hoc, feature-level consistency information. This distinction avoids equating a simple threshold rule with the internal decision mechanism of a nonlinear multilayer perceptron.

Formation of the interpretable feature profile and statistical characterization of indicators

The formation of the feature profile in the proposed method is based on the principle of physical interpretability of each indicator in terms of fabric structure. Unlike latent representations of deep neural networks, where features do not have a direct technological interpretation, this study employs only those microtexture characteristics that can be directly linked to the spatial organization of yarns and fibers, local surface heterogeneity, and photometric properties of the material.

Table 1. Summary of method components and key parameters established in the study.

Method component	Step summary	Key parameters identified in the study
Image standardization	Deterministic image preparation before feature extraction: resize to a common base size, grayscale conversion, and automatic centered fields of view.	Base image: 256×256; no manual ROI selection.
Building centered multi-context representation	Automatic centered fields of view provide local, intermediate, and full standardized spatial context.	*s*0 = 96×96, *s*1 = 160×160, *s*2 = 256×256 centered crops.
Calculation of the full set of interpretable features	Computation of a complete set of indicators for each centered field of view and formation of a multi-context feature profile suitable for further statistical analysis and classifier training	111 fixed features: GLCM 36, LBP 33, photometric intensity 27, Shannon entropy 3, edge/gradient 12.
Training the classifier on the complete feature profile	Stratified 70/30 image-level split; preprocessing and MLP fitted on training data only; test used only for final predictive evaluation.	Input: all 111 features; train/test = 280/120. MLP: (64,32), ReLU, Adam, *alpha*=1e−4, *learning_rate_init*=1e−3, max_iter=600, seed=42.
Post-hoc statistical characterization	Post-hoc characterization and ranking of class differences after the predictive protocol has been fixed; this component does not select, transform, weight, or tune the MLP input variables.	KS and Mann–Whitney *U*; Benjamini–Hochberg FDR; S_{63}: KS $q < 0.05$ with $d_{min} = 0.0$; C_{15} ranked by |Cohen's *d*|.

Table 2. Summary of deployment pipeline steps and key parameters.

Pipeline steps	Step summary	Key parameters identified in the study
Image standardization	Transformation of a new microimage into a unified representation for correct computation of indicators and application of the trained model	Base image 256×256; deterministic preprocessing; no manual ROI.
Building centered multi-context representation	Construction of local, intermediate, and full centered spatial contexts for reproducing feature patterns across the three fields of view.	Centered fields: *s*0 = 96×96, *s*1 = 160×160, *s*2 = 256×256.
Calculation of interpretable features	Compute the same fixed 111-dimensional feature vector used during training.	Identical feature definitions and ordering; no test-time feature selection.
MLP classification followed by post-hoc interpretation	First obtain the MLP class prediction; then compare E_{14} features with training-derived ranges and thresholds.	Threshold rules do not change the MLP prediction and do not represent its nonlinear decision boundary.

For each centered field of view, GLCM descriptors are calculated after quantizing grayscale intensity to 32 levels. Co-occurrence matrices use pixel distances $d = \{1, 2, 3\}$ and four directions (0°, 45°, 90°, 135°), with symmetric = True and normed = True. For contrast, dissimilarity, homogeneity, energy, correlation, and ASM, both the mean and standard deviation over all distance-direction combinations are retained, producing 12 GLCM features per field of view.

Local micropatterns are calculated using uniform LBP with $P = 8$ sampling points and radius $R = 1$; the normalized 10-bin histogram and the stored final-bin ratio form 11 LBP-related indicators per field of view. Photometric descriptors comprise mean, standard deviation, minimum, maximum, P10, median (P50), P90, skewness, and kurtosis. Shannon entropy is calculated separately. Gradient information is derived with the Sobel operator and represented by mean gradient strength, standard deviation, P90, and the proportion of pixels with gradient strength > 0.2.

Such a set of features makes it possible to simultaneously account for weave regularity, local structural variations, and photometric manifestations of the material, forming a multi-context feature profile with clear physical meaning.

For each of the 111 predefined features, class-distribution differences are assessed using the two-sided Mann–Whitney U and Kolmogorov–Smirnov tests. Cohen's d is calculated in parallel to quantify effect magnitude. These statistics characterize interpretable signals and rank the most pronounced differences; they do not determine which variables are passed to the MLP.

Multiple comparisons are controlled using the Benjamini–Hochberg FDR procedure at $\alpha = 0.05$. The S_{63} set is defined using the adjusted Kolmogorov–Smirnov q-value ($q < 0.05$) with $d_{min} = 0.0$, i.e., Cohen's d is not used as a hard inclusion threshold. Mann−Whitney adjusted values are reported as a complementary nonparametric check, and $|d|$ is used to form the C_{15} ranking.

The S_{63} and C_{15} sets are descriptive views of the predefined feature space rather than model-input selection stages. In the Results, these summaries are calculated over the complete 400-image experimental subset as post-hoc characterization after the predictive pipeline and evaluation protocol are fixed. Because S_{63}/C_{15} do not select, transform, or weight the 111 MLP inputs and are not used for model tuning or performance calculation, their use does not affect the held-out Accuracy or Macro-F_1. A separate E_{14} subset is used only for human-readable post-hoc interpretation; the percentile ranges and thresholds applied to held-out test examples are estimated exclusively from the training partition.

Research data and dataset formation

The open "Natural and synthetic fabrics dataset" [24] is used as the source image collection. It is intended for binary natural-versus-synthetic classification and was acquired under controlled microscope conditions. The source collection contains 3,107 images, but the computational experiment reported in this paper deliberately uses a reproducible balanced subset of 400 images to obtain equal class representation and a transparent 280/120 train/test protocol.

The images for the binary dataset were captured using a digital USB microscope "Delta Smart MP5 Pro". During acquisition, a fixed geometry was maintained: the working distance from the lens to the sample was 6.5 cm. Illumination was provided by a built-in system of five LEDs, ensuring stable surface lighting and reducing local brightness variations. The illumination level was maintained within the range of 400–420 lux, which was sufficient for capturing fine microtexture elements. The images were acquired with a spatial resolution of 1024×1024 pixels and stored in ".png" format. To enhance reproducibility of visual characteristics, exposure and white balance were manually adjusted and fixed to identical values across the sample series. An example of dataset samples is shown in Figure 2.

The source dataset contains 3,107 images: 1,547 labeled as natural and 1,560 labeled as synthetic. For the reported experiment, the notebook randomly shuffles the image paths with seed 42 and retains 200 images from each class. The resulting 400-image subset is the dataset on which the 111-feature profile, model fitting, descriptive statistics, and functional evaluation reported below are based [24] [25].

The source collection was created from several dozen physical fabric specimens. Multiple images were acquired from individual specimens, with unequal numbers of images per specimen, including views from different sides and viewing configurations and images acquired after changes in specimen state such as stretching/deformation. Because this first version of the dataset did not retain a specimen identifier for each image, the exact specimen count and the image-to-specimen mapping cannot be verified retrospectively from the published metadata or archived notebook. Consequently, the present split is necessarily image-level, and images originating from the same physical specimen may occur in both partitions. The reported validation is therefore not interpreted as specimen-level validation.

Data partitioning, model configuration, and reproducibility of experiments

The 400-image balanced subset is split before model fitting using sklearn.model_selection.train_test_split with test_size = 0.30, random_state = 42, and stratify = y. This produces 280 training images (140 per class) and 120 held-out test images (60 per class). The 111 feature definitions are fixed a priori and extracted

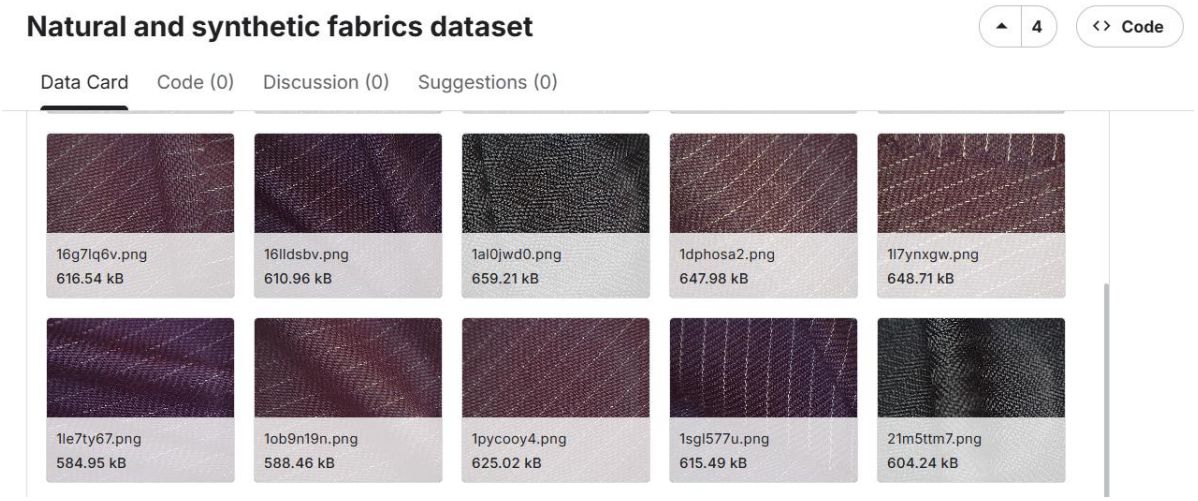


Figure 2. Example view of the "Natural and Synthetic Fabrics Dataset" published on the Kaggle platform.

deterministically; no label-dependent feature-selection step supplies inputs to the MLP. Median imputation, StandardScaler parameter estimation, and MLP fitting are encapsulated in a scikit-learn Pipeline and are fitted on the training partition only. Within the predictive evaluation pipeline, the held-out test images are used only for final prediction and performance calculation. After this protocol is fixed, their labels are also included in the separately reported S_{63}/C_{15} descriptive summaries of the complete 400-image subset; those summaries do not feed back into preprocessing, model fitting, E_{14} threshold estimation, or performance calculation.

The MLP is implemented with sklearn.neural_network.MLPClassifier using hidden_layer_sizes = (64, 32), activation = 'relu', solver = 'adam', *alpha* = 1e−4, *learning_rate_init* = 1e−3, *max_iter* = 600, and *random_state* = 42. Early stopping is not enabled in the reported notebook; *max_iter* = 600 is the upper iteration limit, while the standard scikit-learn convergence criterion is otherwise left at its default setting. The E_{14} P10–median–P90 ranges and midpoint thresholds used to interpret held-out predictions are recalculated from the training partition after splitting. Accuracy confidence intervals are reported using the Wilson binomial interval; the Macro-F_1 interval is estimated by 20,000 bootstrap resamples of paired test predictions with seed 42. The source notebook and fixed random seed make this pre-specified image-level experiment reproducible.

All experiments were implemented as a reproducible Python notebook in the Google Colab environment using a standard GPU runtime. Image processing, feature profile construction, statistical analysis, and classifier training were performed within a single computational pipeline without manual intervention in the data between stages.

NumPy [26] and pandas [27] were used for array and table operations; statistical testing was implemented via scipy.stats [28]; feature processing (standardization and proper handling of missing values) was carried out using scikit-learn [29]. The multilayer perceptron classification model was implemented based on "sklearn.neural_network.MLPClassifier" [30]. All fixed parameters (scales, feature set, selection criteria, model structure) were explicitly defined in the code and remained unchanged across runs.

RESULTS AND DISCUSSION

Feature inventory and statistical characterization criteria

For the binary task "natural" vs "synthetic," a complete set of 111 interpretable microtexture features was constructed across three levels of detail: *s*0, *s*1, and *s*2. The results are organized into three complementary analysis layers. None of these layers is used to reduce the input dimensionality of the MLP; the classifier uses the complete 111-feature profile.

- S_{63} (significant set) – a post-hoc descriptive set of 63 features satisfying the FDR-adjusted Kolmogorov–Smirnov criterion $q < 0.05$;
- C_{15} (core-15) – the 15 strongest statistically significant indicators ranked by absolute Cohen's $|d|$;
- E_{14} (interpretation rules) – a compact set of physically interpretable indicators used after prediction for training-derived threshold checks and engineering consistency control.

Statistically significant features S_{63}: structure by groups and fields of view

Across the complete 400-image experimental subset, the post-hoc descriptive analysis identifies 63 of 111 predefined indicators with FDR-adjusted Kolmogorov–Smirnov $q < 0.05$. This analysis is performed after the predictive representation and evaluation protocol are fixed. The S_{63} results characterize dataset-level class differences; they are

not used as a feature-selection gate or model-tuning criterion for the MLP and therefore do not affect the held-out classification metric.

As shown in Table 3, across all three fields of view, the largest proportion of statistically significant indicators is formed by photometric intensity characteristics (26 out of 27), demonstrating a strong systematic brightness shift between the two image classes in this dataset. GLCM, LBP, entropy, and gradient descriptors also contribute significant structural information, but the predominance of intensity features requires cautious interpretation.

Because brightness P90, mean, and median are among the strongest effects, the present classifier may distinguish classes partly through systematic reflectance/brightness differences that are stable in the controlled dataset. The current experiment does not disentangle this photometric contribution from microstructural contribution. Accordingly, the reported performance is specific to the studied imaging protocol and should not be extrapolated to different illumination, exposure, dyes, finishes, or devices without recalibration and external validation.

Effect size and class separability C_{15} (core-15)

For quantitative assessment of the practical magnitude of differences, the effect size Cohen's *d* was additionally used, with the sign aligned to indicate the direction of dominance ("synthetic" > "natural" or "natural" > "synthetic"). Among the features with the largest absolute values of *d*, GLCM correlation indicators and robust photometric intensity characteristics (in particular P90, mean, and median) are dominant. This indicates that, for the dataset as a whole, the difference between classes is reflected not only statistically but also in the magnitude of distribution shifts. The consistency between effect size magnitude and distribution separability is further illustrated by a volcano plot, which combines Cohen's *d* with the significance of the Kolmogorov–Smirnov test. The ranking of features by effect size is shown in Fig. 3.

As shown in Fig. 3, the largest |*d*| values are associated with GLCM correlation (both directional mean and variability across directions) and photometric intensity characteristics (P90, mean, median), with these effects appearing across multiple centered fields of view. The sign of *d* reflects the direction of the difference: for features where "synthetic" > "natural," *d* is negative, whereas for "natural" > "synthetic," it is positive. The recurrence of the same types of indicators among the top features across different fields of view supports the conclusion that the observed effect is not confined to a single spatial context.

Notably, the C_{15} set is formed by five families of indicators repeated across three centered fields of view (5 × 3 = 15), namely:

- glcm_correlation_mean (regularity, repeatability of local transitions),
- glcm_correlation_std (directional variability, anisotropy),
- int_p90 (upper brightness level),
- int_mean (average photometric level of the ROI),
- int_p50 (typical brightness level).

The influence of effect size on the separability of distributions is further illustrated by the volcano plot in Fig. 4, where Cohen's *d* is combined with the statistical significance of the Kolmogorov–Smirnov test.

In Fig. 4, the *X*-axis represents Cohen's *d*, while the *Y*-axis shows −log10(*p*) of the Kolmogorov–Smirnov test. Accordingly, points with large |*d*| values and high −log10(*p*) correspond to features for which both a substantial shift between classes and a statistically significant difference in distributions are observed. The concentration of points in regions of large negative and positive *d* confirms the presence of a group of indicators with strongly shifted distributions, which provide the main contribution to distinguishing between "natural" and "synthetic" within the studied dataset.

Key interpretable features for the *s*1 field of view

For a detailed comparison of feature distributions between classes, *s*1 (the centered 160×160 field of view) was selected as the reference spatial context because it lies between the local *s*0 field and the full standardized *s*2 field. To maintain consistency with section Effect size and class separability, this subsection focuses on the five strongest families of indicators that form the top-15 features by absolute effect size due to their recurrence across the three centered fields of view. For *s*1, these indicators are represented by the following features:

Table 3. Number of significant features by groups and centered fields of view.

Feature group	*s*0	*s*1	*s*2	Total
GLCM (statistical co-occurrence texture)	7 of 12	8 of 12	6 of 12	21 of 36
LBP (local binary patterns)	3 of 11	3 of 11	3 of 11	9 of 33
Intensity (photometric brightness)	8 of 9	9 of 9	9 of 9	26 of 27
Entropy (Shannon)	0 of 1	0 of 1	1 of 1	1 of 3
Edges (gradient strength)	2 of 4	2 of 4	2 of 4	6 of 12

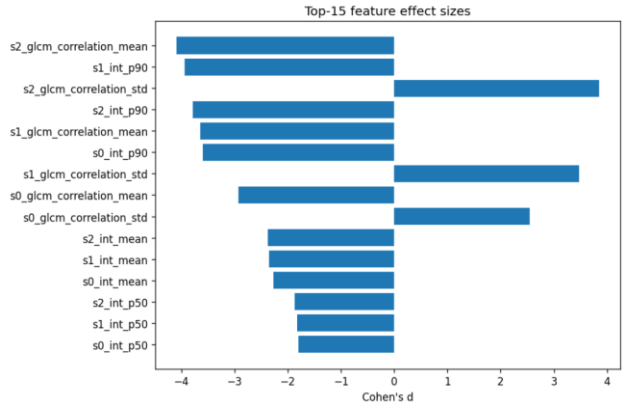


Figure 3. Top 15 features by effect size (Cohen's *d*).

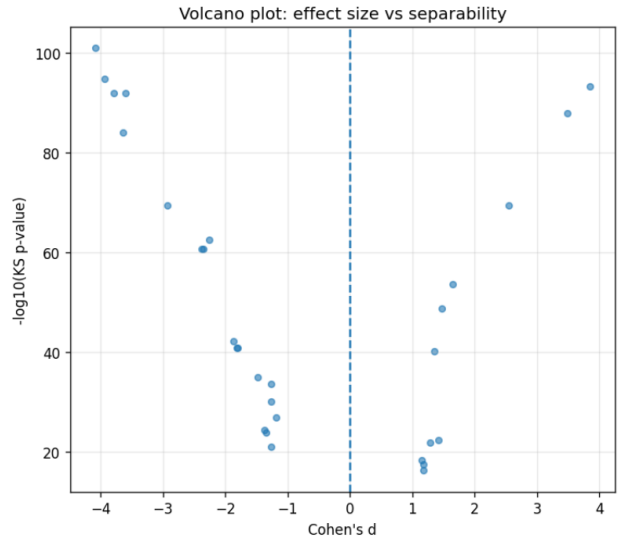


Figure 4. Volcano plot: relationship between effect size (Cohen's *d*) and distribution separability according to the Kolmogorov–Smirnov test.

- GLCM correlation (directional mean);
- GLCM correlation (variability across directions, std);
- brightness P90;
- mean brightness value;
- median brightness (P50).

Table 4 presents the medians and P10–P90 intervals for the "natural" and "synthetic" classes, the direction of the effect, Cohen's *d*, and adjusted *q*-values (MW). A graphical comparison of these indicators is shown in Fig. 5 in the form of boxplot diagrams.

Table 4 provides detailed numerical characterization for the reference *s*1 field of view, whereas Fig. 5 additionally visualizes representative top-ranked features from *s*0, *s*1, and *s*2 to demonstrate their recurrence across spatial contexts.

As shown in Fig. 5, there is a systematic shift of medians and interquartile ranges between classes for the key indicators. For the features GLCM correlation – mean across directions (s1_glcm_correlation_mean) and brightness P90 (s1_int_p90), the distribution of the "synthetic" class is shifted toward higher values, and the overlap of interquartile ranges is limited, which is consistent with the large absolute values of Cohen's *d* reported in Table 4. For the feature GLCM correlation – variability across directions (s1_glcm_correlation_std), the opposite effect is observed: higher values are characteristic of the "natural" class, reflecting increased directional variability of texture. The considered features exhibit isolated outliers; however, these do not alter the overall trend, as the main differences are captured by the medians and interquartile ranges.

Stability of effects across centered fields of view

An important requirement for interpretable microtexture indicators in applied scenarios is the stability of their discriminative behavior across different spatial contexts. In this study, *s*0, *s*1, and *s*2 do not represent different native image resolutions: they are centered fields of view of 96×96, 160×160, and 256×256 pixels extracted from the same standardized 256×256 image. The comparison therefore evaluates whether the direction and magnitude of class differences remain consistent as the analyzed spatial context increases. To this end, Cohen's *d* effect sizes were compared for the same indicators across *s*0, *s*1, and *s*2.

Table 4. Key indicators for the *s*1 field of view: medians, P10–P90 ranges, and effect size.

Group	Feature	«natural»: Med (P10–P90)	«synthetic»: Med (P10–P90)	Direction	Cohen's *d*	*q* (MW)
GLCM	GLCM correlation – mean across directions (s1_glcm_correlation_mean)	0.2657 (0.1086–0.5106)	0.8522 (0.6617–0.9250)	«synthetic» > «natural»	-3.6403	3.89e-60
GLCM	GLCM correlation – variability across directions (s1_glcm_correlation_std)	0.2101 (0.1550–0.2445)	0.0616 (0.0285–0.1177)	«natural» > «synthetic»	3.4796	1.21e-59
Intensity	Brightness – 90th percentile (s1_int_p90)	108.5 (88.9–140.1)	214.0 (185.1–243.1)	«synthetic» > «natural»	-3.9380	1.52e-59
Intensity	Brightness – mean (s1_int_mean)	77.0705 (57.8845–106.8948)	159.2440 (105.2755–194.7710)	«synthetic» > «natural»	-2.3501	1.43e-52
Intensity	Brightness – median (s1_int_p50)	75.0 (51.9–106.0)	157.0 (78.0–200.0)	«synthetic» > «natural»	-1.8242	5.38e-39

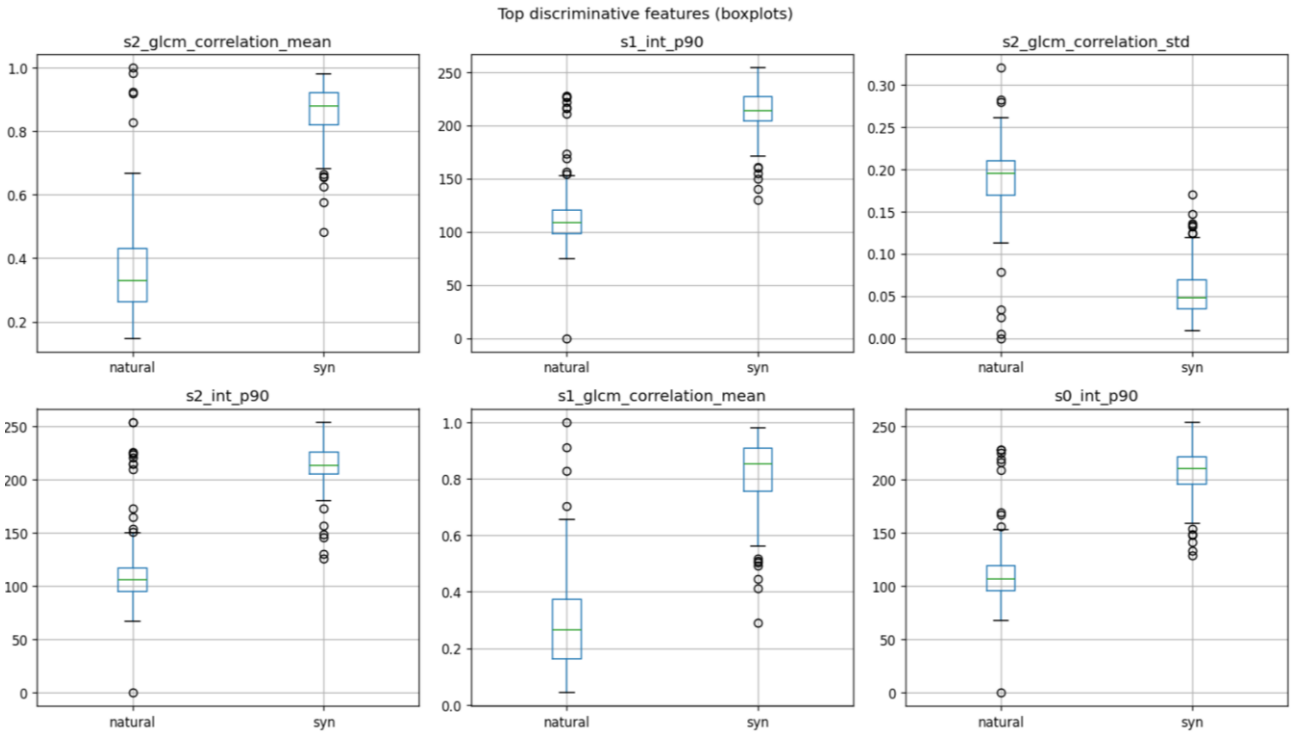


Figure 5. Distributions of representative top-ranked features across the *s*0, *s*1, and *s*2 centered fields of view.

Table 5. Stability of key features across centered fields of view: Cohen's *d* for *s*0, *s*1, and *s*2 and effect direction.

Feature	**Interpretation**	***d* (*s*0)**	***d* (*s*1)**	***d* (*s*2)**	**Consistent direction**
GLCM correlation – mean across directions	Regularity, repeatability of local transitions	−2.9317	−3.6403	−4.0932	«synthetic» > «natural»
GLCM correlation – standard deviation across directions*	Anisotropy, directional dependence of structure	2.5508	3.4796	3.8518	«natural» > «synthetic»
Brightness – 90th percentile (P90)	Upper brightness level (top 10% of pixels)	−3.5979	−3.9380	−3.7900	«synthetic» > «natural»
Brightness – mean value	Average photometric level of ROI	−2.2668	−2.3501	−2.3804	«synthetic» > «natural»
Brightness – median (P50)	Typical brightness level	−1.8007	−1.8242	−1.8687	«synthetic» > «natural»
Edge strength – mean value	Overall density of micro-contours (gradients)	1.1045	1.1741	1.1051	«natural» > «synthetic»
LBP (riu2) – proportion of uniform patterns, bin 02	Frequency of local micropatterns of type "*k*=2"	1.3522	1.4678	1.6424	«natural» > «synthetic»
LBP (riu2) – proportion of uniform patterns, bin 05	Frequency of local micropatterns of type "*k*=5"	−0.6255	−0.6466	−0.6693	«synthetic» > «natural»
LBP (riu2) – proportion of uniform patterns, bin 07	Frequency of local micropatterns of type "*k*=7"	0.5708	0.5834	0.5399	«natural» > «synthetic»
Intensity entropy (Shannon)	Diversity of brightness levels in ROI	−0.2144	−0.4275	−0.9897	effect pronounced at *s*2

**Note: Here, the standard deviation is calculated across GLCM directions and reflects directional variability of the texture, not variability across dataset samples.*

Stability across the three centered fields of view was evaluated according to two criteria:

- consistency of the effect direction ("synthetic" > "natural" or "natural" > "synthetic") across the three fields of view;
- moderate variation in the magnitude of *d* without switching to the opposite trend.

The results (Table 5) show that for most key indicators, the sign of Cohen's *d* remains unchanged across *s*0–*s*2, while |*d*| varies within the same order of magnitude, indicating stability of the observed class differences with respect to the selected field of view.

The most stable indicators are two main types. First, GLCM correlation features: the mean correlation

across directions is consistently higher for the "synthetic" class, while the standard deviation of GLCM correlation across directions is consistently higher for "natural." This combination aligns with the interpretation that, within the studied dataset, synthetic fabrics exhibit a more regular structure with relatively uniform directional properties, whereas natural fabrics demonstrate greater directional variability (anisotropy) of microtexture. Second, photometric intensity indicators (P90, mean, P50) also show high stability, with a consistent dominance of "synthetic" > "natural" across all three fields of view, reflecting a systematic shift in the brightness profile between classes.

Additionally, gradient-based indicators (in particular mean edge strength) and certain LBP features demonstrate a stable effect direction across the three fields of view, although their $|d|$ values are smaller than those of the GLCM correlation and photometric features. They can therefore be treated as supporting interpretable factors that provide additional feature-level consistency information alongside the prediction rather than as explanations of the MLP's internal decision mechanism.

It is also important to highlight intensity entropy (Shannon), for which the effect depends on spatial context: for the smaller and intermediate fields of view ($s0$, $s1$) the difference is weak, whereas for the full 256×256 field ($s2$) the magnitude of d increases substantially. This indicates that entropy is more sensitive to the broader spatial context captured by $s2$ and should not be interpreted as evidence of robustness to changes in native image resolution.

Overall, the obtained results show that the key discriminative patterns between "natural" and "synthetic" are not confined to a single centered field of view but recur across multiple spatial contexts. This consistency supports the use of these descriptors in the post-hoc interpretation layer, where the MLP prediction is accompanied by verifiable feature-level consistency indicators.

Post-hoc prediction interpretation: training-derived percentile profiles and threshold checks

The E_{14} layer translates selected feature values into simple engineering consistency checks after the MLP has produced its prediction. It is not an alternative classifier and is not intended to reproduce the nonlinear internal decision function of the MLP. For the held-out test set, all E_{14} reference statistics are calculated exclusively from the training partition. The E_{14} feature identities were selected as an exploratory post-hoc subset from the complete descriptive analysis, whereas their numerical ranges and thresholds were estimated from the training partition only. Therefore, E_{14} is not treated as an independently validated explanation method and does not affect the MLP performance estimate.

For each E_{14} indicator, the interpretation table specifies:

- working value ranges for the "natural" and "synthetic" classes in the form of "P10–Med–P90";
- a threshold T calculated from the two training-class medians;
- a direction rule indicating which class is supported when the measured feature lies above or below T. A rule may support or contradict the MLP prediction, but it never changes the predicted class.

The threshold is calculated as a robust midpoint between the typical locations of the two training-class distributions:

$$T = (Med_{natural} + Med_{synthetic}) / 2 \quad (1)$$

The arithmetic midpoint of class medians was chosen because it is simple, symmetric, resistant to extreme observations compared with a mean-based midpoint, and easy to communicate in engineering use. It is not optimized for classification accuracy and must not be interpreted as the actual decision boundary of the MLP. The E_{14} indicators retained for this post-hoc layer satisfy three practical criteria:

- have a clear physical interpretation;
- exhibit strong and reproducible effects;
- enable the formulation of concise natural-language interpretation statements.

The resulting E_{14} set is a post-hoc interpretation layer. The training-derived interpretation ranges and thresholds are presented in Table 6.

Table 6 provides training-derived reference ranges for post-hoc interpretation. These rules are intentionally simple: they indicate whether an observed feature value is more consistent with one class profile or the other. They do not represent feature weights, gradients, or the nonlinear decision surface learned by the MLP.

Accordingly, the prediction "natural" or "synthetic" is accompanied by a transparent feature-level consistency profile based on measurable GLCM, photometric, edge, and LBP indicators. This profile can support quality control and flag borderline cases for expert review, but it should not be interpreted as causal evidence for the neural-network decision.

Held-out image-level validation and functional testing

The fitted MLP pipeline was evaluated once on the held-out image-level test set of 120 images (60 per class). Accuracy = 0.9917 with a 95% Wilson confidence interval of 0.9543–0.9985, and Macro-F_1 = 0.9917 with a bootstrap 95% confidence interval of 0.9744–1.0000. The confusion matrix (Table 7) contains 59 correctly classified natural images, one natural image classified as synthetic, no synthetic

Table 6. Post-hoc interpretation rules: training-derived percentile ranges by class and threshold-based consistency indicators.

Feature	Scale	«natural» (P10; Med; P90)	«synthetic» (P10; Med; P90)	Threshold *T*	Interpretation rule
GLCM correlation (mean across directions)	s2	0.1978; 0.3222; 0.5539	0.7555; 0.8895; 0.9394	0.6059	Value > *T* supports "synthetic"
GLCM correlation (variability across directions, std)	s2	0.1488; 0.1957; 0.2243	0.0208; 0.0449; 0.0855	0.1203	Value > *T* supports «natural»
GLCM homogeneity (mean)	s1	0.2340; 0.3079; 0.4412	0.3431; 0.4648; 0.5925	0.3863	Value > *T* supports «synthetic»
GLCM dissimilarity (mean)	s1	1.9007; 2.7527; 3.9384	0.9935; 1.7696; 2.6909	2.2612	Value > *T* supports «natural»
Brightness P90	s1	90.9; 109.0; 140.1	183.9; 214.0; 241.0	161.5	Value > *T* supports «synthetic»
Brightness: mean	s1	59.3437; 76.6330; 106.4549	105.2755; 158.5601; 195.1631	117.5965	Value > *T* supports «synthetic»
Brightness: skewness	s2	0.1836; 1.0833; 2.1436	−1.1476; −0.0918; 1.1211	0.4958	Value > *T* supports «natural»
Brightness: kurtosis	s2	−0.4021; 2.2249; 8.5943	−1.3779; −0.4462; 1.2543	0.8894	Value > *T* supports «natural»
Edge strength: mean	s1	0.0615; 0.0871; 0.1266	0.0327; 0.0572; 0.0865	0.0721	Value > *T* supports «natural»
Edge strength: P90	s1	0.1089; 0.1496; 0.2125	0.0585; 0.1055; 0.1919	0.1275	Value > *T* supports «natural»
LBP riu2: proportion of uniform patterns *k*=2	s2	0.0783; 0.0864; 0.0919	0.0580; 0.0717; 0.0817	0.0790	Value > *T* supports «natural»
LBP riu2: proportion of uniform patterns *k*=5	s2	0.0887; 0.1030; 0.1242	0.0851; 0.1198; 0.1511	0.1114	Value > *T* supports «synthetic»
LBP riu2: proportion of uniform patterns *k*=7	s1	0.0804; 0.0968; 0.1074	0.0663; 0.0851; 0.1043	0.0909	Value > *T* supports «natural»
Intensity entropy (Shannon)	s2	6.1263; 6.4809; 6.7922	6.2838; 6.9991; 7.4125	6.7400	Value > *T* supports «synthetic»

Table 7. Confusion matrix on the held-out 120-image test set (rows = true class; columns = predicted class).

True / predicted	Natural	Synthetic
Natural	59	1
Synthetic	0	60

images classified as natural, and 60 correctly classified synthetic images. Precision/Recall are 1.00/0.98 for the natural class and 0.98/1.00 for the synthetic class.

The post-hoc E_{14} output for correctly classified examples shows high consistency between the predicted class and the training-derived feature rules, whereas the misclassified natural example produces substantially fewer supporting rules. This behaviour is useful as an engineering quality-control signal, but it is not used to calculate or alter the MLP prediction.

The reported 99.17% result is obtained from one pre-specified stratified image-level split. The confidence intervals above quantify uncertainty for this held-out evaluation but do not establish invariance to alternative train/test partitions. Conventional repeated image-level cross-validation was not used as confirmatory evidence because the dataset contains multiple views and deformed states of the same physical specimens while specimen identifiers are unavailable; correlated images could therefore be distributed across folds repeatedly, producing pseudo-replication rather than an independent generalization estimate. A valid assessment of partition sensitivity and specimen-level generalization requires repeated grouped resampling with explicit specimen identifiers. This is treated as a validation requirement for future data collection rather than inferred from the present image-level dataset.

The misclassified case (GT = natural, Pred = synthetic, P(synthetic) = 0.784) illustrates two limitations of the current protocol. First, its brightness P90 and mean values are shifted toward the synthetic-class range, demonstrating that the model can exploit photometric differences in addition to structural information. Second, its comparatively regular texture raises GLCM homogeneity and reduces dissimilarity/edge indicators, producing a feature profile similar to synthetic images. The E_{14} rules therefore provide a useful warning signal, but they should be interpreted as post-hoc consistency checks rather than an explanation of the MLP's internal computation.

Figures 6 and 7 illustrate the structure of the software output. The authoritative training-derived E_{14} ranges and thresholds used in the reported analysis are presented in Table 6. A low number of E_{14} rules supporting the prediction can be used as a practical signal for additional expert verification of borderline images. However, because the present test split is image-level and the source metadata do not identify

25725111326441.jpg | GT=natural | Pred=natural | P[syn]=0.000 | Rules=13/14

GT=natural | Pred=natural



	Indicator	s0	s1	s2
0	GLCM corr mean	0.241153	0.298398	0.336550
1	Brightness P90	105.000000	111.000000	123.000000
2	Edge mean	0.062890	0.087543	0.112900
3	Entropy	6.305383	6.589905	6.954736

	Feature	Scale	Value	T	Rule	Vote	Supports
0	GLCM corr (mean)	s2	0.336550	0.6059	> T → synthetic	natural	True
1	GLCM corr (std)	s2	0.186124	0.1203	> T → natural	natural	True
2	GLCM homogeneity	s1	0.299560	0.3863	> T → synthetic	natural	True
3	GLCM dissimilarity	s1	2.943002	2.2612	> T → natural	natural	True
4	Brightness P90	s1	111.000000	161.5000	> T → synthetic	natural	True
5	Brightness mean	s1	82.986328	117.5965	> T → synthetic	natural	True
6	Brightness skewness	s2	1.528631	0.4958	> T → natural	natural	True
7	Brightness kurtosis	s2	4.262341	0.8894	> T → natural	natural	True
8	Edge mean	s1	0.087543	0.0721	> T → natural	natural	True
9	Edge P90	s1	0.148674	0.1275	> T → natural	natural	True
10	LBP bin 02	s2	0.082993	0.0790	> T → natural	natural	True
11	LBP bin 05	s2	0.094788	0.1114	> T → synthetic	natural	True
12	LBP bin 07	s1	0.108672	0.0909	> T → natural	natural	True
13	Shannon entropy	s2	6.954736	6.7400	> T → synthetic	synthetic	False

2556141339252.jpg | GT=synthetic | Pred=synthetic | P[syn]=0.999 | Rules=14/14

GT=synthetic | Pred=synthetic



	Indicator	s0	s1	s2
0	GLCM corr mean	0.812015	0.776865	0.729045
1	Brightness P90	218.000000	218.000000	219.000000
2	Edge mean	0.036426	0.045539	0.057519
3	Entropy	6.609130	6.683892	6.790717

	Feature	Scale	Value	T	Rule	Vote	Supports
0	GLCM corr (mean)	s2	0.729045	0.6059	> T → synthetic	synthetic	True
1	GLCM corr (std)	s2	0.079277	0.1203	> T → natural	synthetic	True
2	GLCM homogeneity	s1	0.472298	0.3863	> T → synthetic	synthetic	True
3	GLCM dissimilarity	s1	1.543223	2.2612	> T → natural	synthetic	True
4	Brightness P90	s1	218.000000	161.5000	> T → synthetic	synthetic	True
5	Brightness mean	s1	184.516784	117.5965	> T → synthetic	synthetic	True
6	Brightness skewness	s2	-0.341521	0.4958	> T → natural	synthetic	True
7	Brightness kurtosis	s2	-0.324794	0.8894	> T → natural	synthetic	True
8	Edge mean	s1	0.045539	0.0721	> T → natural	synthetic	True
9	Edge P90	s1	0.084222	0.1275	> T → natural	synthetic	True
10	LBP bin 02	s2	0.078079	0.0790	> T → natural	synthetic	True
11	LBP bin 05	s2	0.141373	0.1114	> T → synthetic	synthetic	True
12	LBP bin 07	s1	0.084609	0.0909	> T → natural	synthetic	True
13	Shannon entropy	s2	6.790717	6.7400	> T → synthetic	synthetic	True

Figure 6. Examples of post-hoc interpretation outputs for neural-network classifications.

25728144537102.jpg | GT=natural | Pred=synthetic | P[syn]=0.784 | Rules=6/14

GT=natural | Pred=synthetic



	Indicator	s0	s1	s2
0	GLCM corr mean	0.506138	0.495946	0.531092
1	Brightness P90	193.000000	195.000000	198.000000
2	Edge mean	0.033136	0.043426	0.049638
3	Entropy	5.822551	5.952681	6.123033

	Feature	Scale	Value	T	Rule	Vote	Supports
0	GLCM corr (mean)	s2	0.531092	0.6059	> T → synthetic	natural	False
1	GLCM corr (std)	s2	0.153432	0.1203	> T → natural	natural	False
2	GLCM homogeneity	s1	0.491225	0.3863	> T → synthetic	synthetic	True
3	GLCM dissimilarity	s1	1.407376	2.2612	> T → natural	synthetic	True
4	Brightness P90	s1	195.000000	161.5000	> T → synthetic	synthetic	True
5	Brightness mean	s1	175.244568	117.5965	> T → synthetic	synthetic	True
6	Brightness skewness	s2	0.683244	0.4958	> T → natural	natural	False
7	Brightness kurtosis	s2	1.119072	0.8894	> T → natural	natural	False
8	Edge mean	s1	0.043426	0.0721	> T → natural	synthetic	True
9	Edge P90	s1	0.080033	0.1275	> T → natural	synthetic	True
10	LBP bin 02	s2	0.084625	0.0790	> T → natural	natural	False
11	LBP bin 05	s2	0.110306	0.1114	> T → synthetic	natural	False
12	LBP bin 07	s1	0.097891	0.0909	> T → natural	natural	False
13	Shannon entropy	s2	6.123033	6.7400	> T → synthetic	natural	False

Figure 7. Examples of misclassification.

physical specimens, this mechanism has not yet been validated for independent unseen specimens.

The error analysis also underscores the sensitivity of the current approach to photometric conditions. Reflections, local overexposure, surface finish, and systematic brightness differences may shift the same indicators that are highly discriminative in the present dataset. The current experiment therefore cannot establish how much of the 99.17% accuracy is attributable to intrinsic material structure versus dataset-specific reflectance and acquisition characteristics. External testing under deliberately varied illumination and on independently identified specimens is required.

CONCLUSIONS

A reproducible image-level approach to binary natural-versus-synthetic fabric discrimination has been implemented using a fixed 111-dimensional profile of interpretable microtexture features at three centered spatial contexts (fields of view: 96×96, 160×160, and 256×256). The MLP receives the complete predefined feature profile; S_{63}/C_{15} statistical analyses provide post-hoc characterization of the feature space, while the E_{14} module is a separate post-hoc interpretation layer based on training-derived percentile ranges and median-midpoint thresholds.

The source Kaggle dataset contains 3,107 images, while the experiment reported here uses a reproducible balanced subset of 400 images (200 per class) and one pre-specified stratified 70/30 image-level split. The held-out test set contains 120 images. The MLP achieved Accuracy = 0.9917 (95% CI 0.9543–0.9985) and Macro-F_1 = 0.9917 (bootstrap 95% CI 0.9744–1.0000), with a confusion matrix of [[59, 1], [0, 60]] for natural and synthetic classes, respectively.

The principal limitations concern generalizability. The source collection contains multiple images from individual physical fabric specimens, including different sides/viewing configurations and stretched or otherwise deformed states, but specimen identifiers were not retained at image level. Consequently, specimen-level independence between train and test partitions cannot be demonstrated, and the present confidence intervals do not quantify variability across alternative grouped

partitions. Moreover, photometric intensity features are among the strongest discriminators, meaning that the classifier may partly exploit systematic brightness/reflectance differences tied to the acquisition protocol. The reported 99.17% accuracy should therefore be interpreted as internal image-level performance under these controlled conditions rather than evidence of specimen-level generalization, general textile-material identification, or immediate industrial sorting readiness. The current formulation is binary and does not address blended materials or synthetic-content estimation.

Future work should prioritize specimen-level external validation with explicit specimen identifiers, repeated grouped resampling, controlled illumination/exposure perturbation experiments, and testing across different microscopes and textile batches. Extension to blended materials and multi-class classification should follow only after this robustness has been established. From a practical perspective, the current method is best viewed as a decision-support component for controlled binary screening rather than a complete autonomous textile-sorting system.

Acknowledgement: *This research was financially supported by Khmelnytskyi National University under project No. 0126U002064, titled "Neural Network Methods for Textile Waste Classification in Circular Economy Systems." The authors would also like to express their gratitude for the consultations and expert support provided by the European Digital Innovation Hub for Clothing Technology and Textiles, known as "CLOTEX-HUB".*

REFERENCES

1. Bukhantsova L., Zacharkevich O., et al.: Data analysis for the prediction of textile waste recycling in Ukraine, Fibres and Textiles 31(2), 2024, pp. 66-73. https://doi.org/10.15240/tul/008/2024-2-009
2. Das A.K., Hossain M.F., et al.: Circular economy: a sustainable model for waste reduction and wealth creation in the textile supply chain, SPE Polymers 6(1), 2025, e10171. https://doi.org/10.1002/pls2.10171
3. United Nations Environment Programme: Sustainability and circularity in the textile value chain - a global roadmap, Paris, 2023, online: https://circulareconomy.europa.eu/platform/sites/default/files/2023-12/Full%20Report%20-%20UNEP%20Sustainability%20and%20Circularity%20in%20the%20Textile%20Value%20Chain%20A%20Global%20Roadmap_0.pdf [cit. 26.3.2026]
4. Abtew M.A., Atalie D., Dejene B.K.: Recycling of cotton textile waste: technological process, applications, and sustainability within a circular economy, Journal of Industrial Textiles 55, 2025, 15280837251348663. https://doi.org/10.1177/15280837251348663
5. Faghih E., Saki Z., Moore M.: A systematic literature review – AI-enabled textile waste sorting, Sustainability 17(10), 2025, 4264 P. https://doi.org/10.3390/su17104264
6. Halvorsen T.S., Tyapin I., Jha A.: Autonomous textile sorting facility and digital twin utilizing an AI-reinforced collaborative robot, Electronics 14(13), 2025, 2706 P. https://doi.org/10.3390/electronics14132706
7. Tian R., Lv Z., et al.: Qualitative classification of waste garments for textile recycling based on machine vision and attention mechanisms, Waste Management 183, 2024, pp. 74-86. https://doi.org/10.1016/j.wasman.2024.04.040
8. Cura K., Rintala N., et al.: Textile recognition and sorting for recycling at an automated line using near infrared spectroscopy, Recycling 6(1), 2021, 11 P. https://doi.org/10.3390/recycling6010011
9. European Parliament and the Council of the European Union: Directive (EU) 2018/851 of the European Parliament and of the Council of 30 May 2018 amending Directive 2008/98/EC on waste, 2018, online: https://eur-lex.europa.eu/eli/dir/2018/851/oj/eng [cit. 26.3.2026]
10. Fibersort: Fibersort® technology, 2024, online: https://www.fibersort.com/en/technology [cit. 26.3.2026]
11. Tomra: Automated textile sorting systems, 2024, online: https://www.tomra.com/waste-metal-recycling/applications/waste-recycling/textiles [cit. 26.3.2026]
12. Valvan: Textile sorting solutions for the recycling industry, 2024, online: https://www.valvan.com/en/solutions/textile-sorting-recycling [cit. 26.3.2026]
13. Hu Y., Long Z., et al.: Fabric surface characterization: assessment of deep learning-based texture representations using a challenging dataset, The Journal of The Textile Institute 112(2), 2021, pp. 293-305. https://doi.org/10.1080/00405000.2020.1757296
14. Seçkin M., Seçkin A.Ç., et al.: A microscopic image dataset of woven fabrics for predicting texture and weaving parameters through machine learning, Sustainability 15(21), 2023, 15197 P. https://doi.org/10.3390/su152115197
15. Tang K.H.D.: State of the art in textile waste management: a review, Textiles 3(4), 2023, pp. 454-467. https://doi.org/10.3390/textiles3040027
16. Becker A., Datko A., Kroell N., et al.: Near-infrared-based sortability of polyester-containing textile waste, Resources, Conservation and Recycling 206, 2024, 107577 P. https://doi.org/10.1016/j.resconrec.2024.107577
17. Eghtedari F., Pecyna L., et al.: Identification of textile fibres using a near infra-red (NIR) camera, Journal of Imaging 11(4), 2025, 96P. https://doi.org/10.3390/jimaging11040096
18. Li J., Liu Q., et al.: Identification of textile fiber composition in waste textiles using improved CNN and near-infrared spectroscopy, Journal of Industrial Textiles 55, 2025, 15280837251359681. https://doi.org/10.1177/15280837251359681
19. Sormunen T., Mahlamäki E., et al.: Hyperspectral imaging quantifies blend composition change in workwear textiles, Resources, Conservation & Recycling Advances, 2025, 200282 P. https://doi.org/10.1016/j.rcradv.2025.200282
20. Selvaraju R.R., Cogswell M., et al.: Grad-cam: visual explanations from deep networks via gradient-based localization, Proceedings of the IEEE international conference on computer vision, 2017, pp. 618-626. https://doi.org/10.1109/ICCV.2017.74
21. Molchanova M., Didur V., et al.: Method for construction and demolition waste classification using two-factor neural network image analysis, CEUR Workshop Proceedings 3970, 2025, pp. 168-182, online: https://ceur-ws.org/Vol-3970/PAPER14.pdf [cit. 26.3.2026]
22. Haralick R.M., Shanmugam K., Dinstein I.: Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics SMC-3(6), 1973, pp. 610-621. https://doi.org/10.1109/TSMC.1973.4309314
23. Alashour G.H., El Abbadi N.K.: Advances and insights in image texture analysis: a review, Mesopotamian Journal of Big Data, 2025, pp. 108-135. https://doi.org/10.58496/MJBD/2025/008
24. Olyarnncnn: Natural and synthetic fabrics dataset, 2024, online: https://www.kaggle.com/datasets/olyarnncnn/natural-and-synthetic-fabrics-dataset [cit. 26.3.2026]
25. Sobko O., Mazurets O., et al.: Method for analysis and formation of representative text datasets, CEUR Workshop

Proceedings 3899, 2025, pp. 84-98, online: https://ceur-ws.org/Vol-3899/paper9.pdf [cit. 26.3.2026]
26. NumPy: fundamental package for array computing in Python, 2024, online: https://numpy.org [cit. 26.3.2026]
27. Pandas: powerful Python data analysis toolkit, 2024, online: https://pandas.pydata.org [cit. 26.3.2026]
28. SciPy stats: statistical functions for Python, 2024, online: https://docs.scipy.org/doc/scipy/reference/stats.html [cit. 26.3.2026]
29. Scikit-learn: machine learning in Python, 2024, online: https://scikit-learn.org [cit. 26.3.2026]
30. Scikit-learn neural network: MLPClassifier for multi-layer perceptron models, 2024, online: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html [cit. 26.3.2026]