Materials Circular Economy (2024) 6:13
https://doi.org/10.1007/s42824-024-00104-1

ORIGINAL PAPER

Raw Cellulosic Fibers: Characterization and Classification by FTIR-ATR Spectroscopy and Multivariate Analysis (PCA and LDA)

Elvis da Cruz Santos[1] · Ana Amelia Benedito Silva[1] · Regis Rossi Alves Faria[1] · Marcia de Almeida Rizzutto[2] · Pedro Henrique Sebe Rodrigues[3] · Julia Baruque-Ramos[1]

Received: 13 December 2023 / Revised: 29 February 2024 / Accepted: 12 March 2024

Abstract

The textile industry is one of the most pollutants in the world. It is estimated that only 20% of the textiles that become solid waste are recycled. In this way, to reduce impacts, the textile industry must include the circular economy in its processing, that is, thinking about textile manufacturing in a closed circuit, which minimizes the consumption of virgin raw materials. The characterization and classification of textile wastes in the industry are done manually, resulting in high costs for the classification of large volumes of textiles. Attenuated total reflected in conjunction with Fourier transformed infrared spectroscopy (ATR-FTIR) can differentiate fibers into plant, animal, and synthetic, being eligible for automated industrial classification. However, this method is limited taking into account the differentiation among cellulosic fibers. This study aimed to evaluate ATR-FTIR spectra obtained from plant origin samples (cotton, kapok, hemp, non-bleached flax, bleached flax, jute, tucum, and tururi) without chemical treatments and observe the potential of multivariate principal component analysis (PCA) and linear discriminant analysis (LDA) on these data. The FTIR "fingerprint" data results values, from 400 to 1800 nm, were employed. To improve the precision of multivariate statistics, these data were previously and individually treated with three types of noise reduction normalization (mean, standardization, and logarithm), and their effects in the final results were analyzed. This database has been normalized with multivariate data analysis PCA (principal component analysis) and LDA (linear discriminant analysis). Employing PCA, tucum, tururi, kapok, jute, and hemp fibers were successfully separated into five different groups, except cotton, non-bleached flax, and bleached flax. For LDA, all fibers were successfully separated, except non-bleached flax and bleached flax. Thus, these results suggest that PCA is a powerful tool in studying textiles with a relatively simple structure, while objects with a more complex or very similar composition, for example, LDA statistics, are more advantageous.

Keywords FTIR · Textiles · Cellulosic fibers · Classification · PCA · LDA

Introduction

The paradigm of the current linear economy concept has been questioned as it is not sustainable, since resources are limited and, on the other hand, demands are increasing (Saccani et al. 2023). Circular economy opposes the current model and proposes to rethink production methods, aiming to mitigate environmental impacts and maximize the useful life of generated products, reinserting material components into the manufacturing process, providing more efficient management of natural resources (Gaustad et al. 2018; de Oliveira Neto et al. 2022; Chowdhury et al. 2023). The industry and different sectors have been making efforts to implement measures that encompass the circular economy, which has been observed in a variety of industrial activities,

✉ Elvis da Cruz Santos
elvis.cruz.santos@usp.br

[1] School of Arts, Sciences and Humanities, University of Sao Paulo, Av. Arlindo Bettio, Sao Paulo, SP 1000, 03828-000, Brazil

[2] University of Sao Paulo, Institute of Physics, R. Do Matao, Sao Paulo, SP 1371, 05508-090, Brazil

[3] University of Sao Paulo, Institute of Mathematics and Statistics, R. Do Matao, Sao Paulo, SP 1010, 05508-090, Brazil

Published online: 27 March 2024

such as packaging, agriculture, and the food industry (Navarro et al. 2018; Haque et al. 2023).

Due to the massive use of water, electricity, and chemical products, the textile and clothing sector is one of the most polluting and resource-intensive industries, negatively and directly impacting the environment. Brazil produced a total of 5.14 billion clothing units in 2022, with a large part of the textile materials ending up in urban landfills, resulting in a cycle that cannot be sustained (Abit 2023). Due to this non-sustainability of recycling as a result of this chain that does not close, the textile industry is linked to one of the types of consumption that causes the most damage to the environment (Krystofik et al. 2018; Sousa-Zomer et al. 2018; de Oliveira Neto et al. 2022).

As global production of textile fibers has more than doubled since the 2000s, to around 120 million tons in 2018, this topic of textile recycling become timely (Palacios-Mateo et al. 2021). The increase in the production and consumption of textiles is correlated with a reduction in the average time used for clothing; therefore, the amount of textile waste is expected to increase, driven mainly by the fast fashion culture (Saccani et al. 2023; Chowdhury et al. 2023).

Textile recycling is the stage in which textile waste is reprocessed to be used in new products, whether textile or non-textile. Fabric recycling routes can be mechanical (pretreatment), chemical (depolymerization of polymeric fibers or dissolution of natural fibers), or thermal (conversion of PET pellets, chips, or flakes into fibers by melt extrusion) (Sandin and Peters 2018). The recycling of post-consumption clothing, which is subsequently chemically treated and processed into regenerated cellulose, allows the transformation of waste into new fiber products with better mechanical properties, consequently leading to destinations other than landfills. As a result, recycling cellulose fibers makes the industry less dependent on primary fiber production and helps fill the gap created by the growing demand for regenerated textile fibers (Nayak et al. 2012; Hospodarova et al. 2018).

Recycling rates for post-consumption textile waste are currently low. Most materials used, whether natural or synthetic, are discarded as waste rather than being processed for reuse or recycling. The main cause of this is the lack of a specific method of collecting post-consumption textile waste, the difficulty of distinguishing between various discarded textile materials, and the costs associated with sorting significant volumes (Peets et al. 2017; Palacios-Mateo et al. 2021). Textiles are currently classified mainly manually. However, there are disadvantages to the manual sorting style problems including high cost, low speed of operation, and the inability to completely automate the processing of large quantities of materials. Commercially available sorting machines operate on conventional sorting methods and systems and generally cannot sort different textile materials or require the help of skilled operators. These devices are largely time-consuming to operate or very expensive to maintain (Peets et al. 2017, 2019). Attenuated total reflected in conjunction with Fourier transformed infrared spectroscopy (ATR-FTIR) can differentiate fibers into plant, animal, and synthetic.

In industrial-level screening automation, the use of ATR-FTIR characterization can be useful because it works without sample pretreatment, and the implementation of this technique is quick and easy data collection in terms of speed of data acquisition. On the other hand, FTIR analysis produces spectral data that are usually difficult to interpret due to the large amount of information it can provide due to the compounds present in different samples and also due to the very similar chemical composition, which generates very similar FTIR spectra, which does not happen with other polymeric fibers (Riba et al. 2020).

Because of this, multivariate classification algorithms to deal with this considerable number of information are necessary. They can be supervised or unsupervised and help to reduce the dimensionality of the data obtained in the analysis to better concentrate the relevant analytical information from the entire data set into a few latent variables.

This procedure also allows the removal of most of the noise present in the original spectral data. Therefore, multivariate analyses using computational algorithms have currently been used, and consistent results are being obtained (Mäkelä et al. 2021; Quintero Balbas et al. 2022). Despite silk and wool, both fibers of animal origin are successfully differentiated, the distinction between fibers of cellulosic origin (cotton, flax, and viscose) employing ATR-FTIR-spectral data and principal component analysis (PCA) treatment was not always possible (Peets et al. 2017).

The objective of this study is to present the improvement performed in the differentiation and classification of cellulosic textile fibers through the modeling of IR spectrum data employing multivariate methods. The analysis uses an open programming language Python (Python Software Foundation) with the improvement of model robustness when analyzing a large number of fabric samples of plant origin. The novelty of the suggested procedure is the use of ATR-FTIR spectra obtained from samples only from plant origin without chemical treatments and observing the potential of principal component analysis (PCA) and linear discriminant analysis (LDA) on these data.

Multivariate Data Analysis

Supervised mathematical models applied to spectra data analysis are very useful due to the large number of wavenumbers in FTIR spectra. The model can show the differences between materials with very similar chemical composition,

such as plant fibers, with the potential to separate and highlight the differences between molecules, classifying only strongly correlated materials in the same group. Therefore, it is a relatively large database constituted of 3550 wavenumbers (relative to the range from 4000 to 600 cm^{-1}) that constitute the variables measured for each ATR-FTIR spectrum of each textile sample. Considering an industrial approach, it is necessary to use appropriate feature extraction and reduction. In this way, the spectra were reduced to the so-called "fingerprint" region, from 1800 to 600 cm^{-1}, which is effectively correlated to the peaks of target molecules. To improve the precision of multivariate statistics, these data were previously and individually treated with three types of noise reduction normalization (mean, standardization, and logarithm), and their effects in the final results were analyzed. This database has been normalized with multivariate data analysis PCA (principal component analysis) and LDA (linear discriminant analysis). These algorithms are designed to reduce the number of latent variables required and compress the essential discriminant information present in untreated spectra while removing most of the noise present in the raw spectra to maximize discrimination power. One of the great advantages of this proposal is the possibility of developing a fast and accurate method for direct and non-invasive screening and classification of different textile fibers by using FTIR equipment in the ATR.

Mathematic Normalization Approach

Mean Normalization (Feature Scaling)

Standardization by mean scaling is a statistical calculation that aims to correct additive and multiplicative effects, generally caused by radiation scattering and noise. This normalization performs the absolute sum of each observation for a given variable (wavenumber columns) and divides each observation in the column by the absolute sum of the respective observations, generating new data corrected by the normalization factor (Ferreira 2015). This operation can be mathematically represented by Eq. 1:

$$x'_{ij}norm = \frac{x_{ij}}{||k_i||} \tag{1}$$

where x'_{ij} norm is the new corrected observation, x_{ij} is the observation of absorbance for a wavenumber, and $||k_i|| = \sum_{j=1}^{J} x_{ij}$ corresponds to the absolute sum of absorbances related to a wavenumber (Ferreira 2015).

Standardization (Gaussian)

Standardization approach involves subtracting each observation from the average of the observations and dividing this result by the standard deviation (as shown in Eq. 2). In this way, each respective column will have its absorbance values subtracted from the average absorbance of the respective column; finally, the result of the sum is divided by the standard deviation. This procedure is necessary to ensure that all data in a column are compatible with the average value, reducing sample fluctuation (Ferreira 2015).

$$x'_{ij} = \frac{x_{ij-\bar{x}_j}}{s_j} \tag{2}$$

Logarithm Normalization

The logarithm can be applied to correct for non-linear trends in data. The choice of the base of the logarithm is arbitrary, as it will not affect the interpretation of the data. The relationship between intensity and light absorbed by the sample follows the decay law (Eq. 3). For FTIR data generated in transmittance, it is possible to obtain linearization of light intensity and concentration. In the reflectance technique in which the emitted beam is reflected in the crystal instead of passing through the sample, as well as in the transmittance technique, it is possible to obtain linearization of the concentration with the reflectance data.

In this procedure, in each cell x_{ij} of the original database, the log is made in base 10, and a new database is created with these new values:

$$A_\lambda = a_\lambda lc \tag{3}$$

where A_λ is the absorbance for a given wavenumber, a_λ is the respective molar absorption coefficient, l is the optical path, and c is the concentration.

PCA and LDA

The most used exploratory analysis technique is principal component analysis (PCA), an unsupervised technique based on data variance. A matrix M containing spectroscopic data from n samples at different λ (wavenumber) variables may contain a large amount of noise and redundant information. This means that the relevant information from the data set that makes up M can be described in a space of reduced dimensionality with v variables, where $v \leq \lambda$. Mathematically, this concept, which is the basis of PCA, can be expressed in the form of the decomposition of the matrix X as exemplified in 4 (Bro and Smilde 2014):

$$\mathrm{M} = \boldsymbol{T}\mathrm{L}^t + \boldsymbol{E} \tag{4}$$

where $\boldsymbol{T}$ is the score matrix and has dimensions $n \times v$, L^t is the orthonormal weight matrix with dimensions $v \times \lambda$, and $\boldsymbol{E}$ is the residual matrix.

The weights are related to the variables that cause greater variability in the data, whereas the PCA score matrix is strongly linked to the concentrations of the substances in the samples. The central objective of this technique is to identify sources of data variation, extracting information about the data set from the number of independent principal components (PCs) that best describe the information, reducing dimensionality without losing relevant information in the data set. However, the spectra of cellulose fabrics are so similar that PCA-based score plots are not effective enough for discrimination (Peets et al. 2017, 2019).

PCA and linear discriminant analysis (LDA) are feature extraction methods. Due to their superior ability to distinguish between objects with very similar chemical compositions, supervised extraction and classification methods like LDA are more advantageous. It behaves similarly to unsupervised PCA in selecting the class allocations of the calibration samples and in allocating each sample to its corresponding class.

LDA is a very well-known and useful pattern recognition method. It produces a linear classification structure that is capable of separating sample series within different classes based on their spectral characteristics, based on the metric of minimizing variations within a class and maximizing the distance between two classes, derived from the common covariance matrix considered for all classes.

LDA finds a model that best discriminates the assigned groups in the original dataset and places samples from the same group as close to each other as possible. Because LDA requires the number of samples to be less than the number of variables, LDA generally follows PCA. The dimensionality matrix is given by Eq. 5 (Ferreira 2015; Quintero Balbas et al. 2022):

$$X = \begin{pmatrix} x1,1 & \cdots & x1,i \\ \vdots & \ddots & \vdots \\ xN,1 & \cdots & xN,i \end{pmatrix} \quad (5)$$

where N represents the total number of reflectance spectra in the matrix (X), and i corresponds to the reflectance values measured in the range of 400 to 1800 cm^{-1}.

Python Open Programming Language and JupyterLab

Python is a high-level, multi-paradigm, general-purpose language, which supports object-oriented, imperative, functional, and procedural paradigms. It has dynamic typing, and one of its main features is that it allows easy reading of the code and requires few lines of code compared to the same program in other languages. Due to its characteristics, it is mainly used for text processing, scientific data, and the creation of CGIs for dynamic web pages. It currently has a community development model, open and managed by the non-profit organization Python Software Foundation.

The statistics calculations were performed in JupyterLab, which is a free software, which provides open standards, and web services for interactive computing across all programming languages.

Methodology (Experimental)

Samples

Ninety-one (91) samples of different plant origin fibers were used to test the viability of ATR-FTIR analysis as described in Table 1. Most fibers are from Brazilian origin, except for hemp and flax (both from European origin), and all of them do not have any treatment except for bleached flax fiber. The samples were inspected to observe and avoid the presence of materials or dirt that could interfere with or lead to incorrect conclusions in the spectra analyses.

Analysis

In the present study, Brucker® model ALPHA device was employed. The ATR (attenuated total reflection) module system used has a Zinc Selenide crystal (ZnSe), and a 4.0 crystal refraction index. The analyses were performed, with 64 scans without sample preparation in the transmission range FTIR, 4000–600 cm^{-1} at a resolution of 4 cm^{-1}.

Table 1 Cellulosic textile fibers analyzed by ATR-FTIR

Fiber	Type fiber	Scientific name	Sample amount
Cotton	seed	*Gossypium herbaceum*	19
Kapok	seed	*Chorisia speciosa*	10
Hemp	bast	*Cannabis sativa*	9
Non-bleached flax	bast	*Linum usitatissimum*	11
Bleached flax	bast	*Linum usitatissimum*	9
Jute	bast	*Corchorus capsularis*	12
Tucum	leaf (Amazon palm)	*Astrocaryum chambira* Burret	10
Tururi	leaf (Amazon palm)	*Manicaria saccifera* Gaertn	11

Subsequently, the spectra were reduced to the fingerprint region from 1800 to 600 cm^{-1} for better characterization. Spectragryph 1.2 software was also used to analyze performance.

In the ATR-FTIR method within the wavenumber range from 4000 to 400 cm^{-1}, mid-infrared electromagnetic radiation has the energy required to change the rotation and vibration levels of molecular bonds. Due to the exceptional selectivity of molecular bonds for absorption of a certain and characteristic length of mid-infrared radiation, it is then possible to understand the chemical composition of matter through radiation absorption. This area of the spectrum is widely used in quantitative and qualitative analyses. The fingerprint range extending from 1800 to 600 cm^{-1} presents the highest concentration of characteristic peaks of the material. With two media of different refractive indices, measurements in the attenuated total reflectance (ATR) module use the behavior of the infrared radiation beam that undergoes reflections between the two media several times. The infrared beam entering through a transparent crystal in these systems has an angle of incidence greater than the critical angle. Only a small percentage of the beam called the evanescent wave passes through the crystal-sample interface and penetrates slightly into the sample when the beam is almost completely reflected. In areas of the infrared spectrum where the sample absorbs energy, the beam has already been attenuated and can promote the excitation of molecular vibrations. The advantage of operating the equipment in this module is the possibility of analyzing any sample without pre-treatment and in any state of matter (Stuart 2004).

Results and Discussion

ATR-FTIR Spectra

The FTIR spectra for textile fibers described in Table 1 (namely, cotton, kapok, hemp, non-bleached flax, bleached flax, jute, tucum, and tururi) are presented in Fig. 1.

The spectra shown from Fig. 1a and b, as expected, exhibit very similar behaviors and patterns (as evidenced in Fig. 1c) since they are lignocellulosic materials, with very similar chemical composition. The FTIR spectra show characteristic absorption bands according to the functional groups in the molecules of the different types of fibers. The most characteristic bands of the chemical functional groups for plant fibers occur mainly in the fingerprint region (from 1800 to 600 cm^{-1}) as shown in Table 2.

Thus, the infrared spectrum of a molecule is considered to be a hallmark for each being slightly different from one to another, although it may be difficult to distinguish by visual inspection of the spectrum. Most of the molecular excitations for textile fibers, as mentioned, occur in the region called the fingerprint region (1800 to 600 cm^{-1}). In this way, the database was reduced to only include the fingerprint range.

Database

The database generated from FTIR measurements on textile fibers has the configuration as shown in Table 3. The columns contain the wavenumbers and the fibers in rows. Each cell corresponds to an observation of absorbance data for a given sample and its associated wavenumber, generating in the present case 91 rows × 583 columns. On this database in this configuration, standard normalization followed by PCA is subsequently applied.

To apply mean normalization scaling, the transposed matrix of the original data is executed; now, this matrix evolves 583 rows × 91 columns. This procedure is necessary so that the sum of the lines is related to their original wavenumber as observed in Table 4. After applying the matrix transpose, the normalization calculation is made, and the matrix is transposed again to return to the original dimension settings and then the PCA is carried out.

First Approach—PCA Preceded by Mean Normalization

To keep the data on the same magnitude scale, a transformation by a normalization factor was used for each absorbance observation for a given fiber. In this way, a new database was created with new absorbance values standardized by dividing each previous observation by the absolute sum of the values of all absorbances of a given fiber.

This normalization is, in principle, efficient for removing systematic variations that are generally associated with the sampling size or the impossibility of controlling the volume of the fiber that was being exposed to radiation, highlighting only information that qualitatively distinguishes one sample from another (Ferreira 2015).

The spectrum after this normalization (Fig. 2a) shows that this metric was more effective in keeping the spectrum scales in the same order of magnitude.

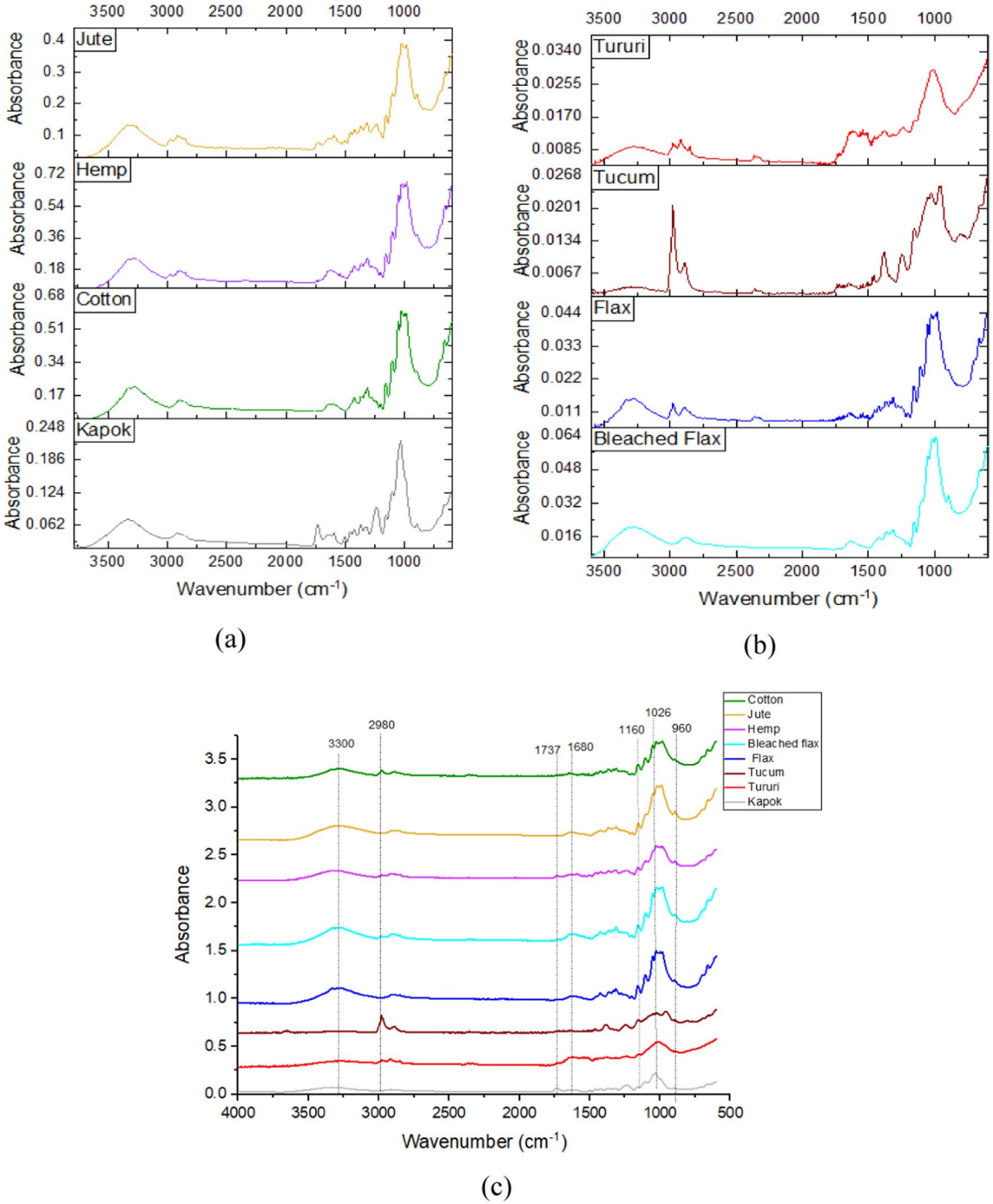


Fig. 1 FTIR spectra from 4000 to 500 cm.$^{-1}$ with absorbance values (Brucker® model ALPHA with ATR, 64 scans, Spectragryph 1.2 software) for **a** cotton, kapok, hemp, and jute; **b** non-bleached flax, flax, tucum and tururi; and **c** summarization of all spectra with some common bands: bleached flax (light blue), jute (yellow), tucum (brown), cotton (dark green), tururi (red), flax (blue), kapok (grey), and hemp (purple)

The PCA of data normalization by dividing the absolute sum of the absorbance of each fiber (Fig. 2b) was more effective for separating the fibers. Tucum fiber, after this normalization, forms a group distinct from other fibers. The dispersion in fiber data occurs mainly for morphological reasons, and there is material dispersed on the fiber, making the region of interaction of infrared radiation with the fiber contact area non-homogeneous.

Table 2 FTIR (mid-infrared) bands for plant material constituents (STUART 2004; PINHEIRO et al. 2022)

Element	Wavenumber (cm^{-1})
α -D-glucose	915, 840
β-D-glucose	915, 900
β-D-fructose	873, 869
β-D-cellulose	916, 908
Cellulose	1170–1150, 1050, 1030
Lignin	1590, 1510
Hemicellulose	1732, 1240
Pectin	1680–1600, 1260, 955

Second Approach—PCA Preceded by Standardization (Gaussian)

After applying the standard scaling, it was noticed that there is a tendency for the values referring to the tururi and kapok fibers to form distinct groups from the other fibers (Fig. 3); the spread of the kapok fiber data is due especially regarding morphology issues since the difficulty to fiber compaction in the device does not allow an exact reproduction of the measurement, and there is no uniformity of the interaction of the IR beam with the fiber surface.

According to Fig. 3, the dispersion in the data, mainly in kapok fiber, is because the measured signal is a deterministic contribution junction, that is, the true signal with relevant information. Another part of the measured signal comes from the stochastic contribution as a source of noise that does not add to the true signal. The kapok fiber disperses the radiation and makes it difficult to reproduce the signal in the same region. Therefore, even though it forms a group distinct from other fibers, there is dispersion in the group of kapok fibers.

The average spectrum of the fibers shows that the spectrum of the fibers is highly similar due to the similar chemical composition (Fig. 4a). The accumulated variance values (Fig. 4b) show that with just two main components, it is possible to obtain approximately 85% of the explained variance, and with only four main components responding with 95% of the data variability, only the four main components will be adopted.

Third Approach—PCA Preceded by Mean and Logarithm Normalization

It was observed that the mean normalization of the data separated a greater number of fibers into subgroups (Fig. 2b) than standardization (Fig. 3). So, the logarithm in base 10

Table 3 Example database absorbance for FTIR wavenumbers (described in rows) versus samples (described in lines)

	1799.705887	1797.644368	1795.582849	1793.521330	1791.459812	1789.398293	1787.336774	1785.275255
dados\ELVIS 16_11_22\ATR_algodao_100%f_med10_64_0.dx	0.239914	0.239997	0.237747	0.228752	0.216106	0.219364	0.231063	0.236964
dados\ELVIS 16_11_22\ATR_algodao_100%f_med9_64_0.dx	0.336450	0.322537	0.305475	0.296521	0.301307	0.318864	0.331600	0.329633
dados\ELVIS 16_11_22\ATR_algodao_100%_med5_64_0.dx	-0.006818	-0.012591	-0.015163	-0.021785	-0.028723	-0.018985	-0.007766	-0.006824
dados\ELVIS 16_11_22\ATR_algodao_100%_med6_64_0.dx	0.248586	0.249738	0.239999	0.221599	0.215353	0.229079	0.238348	0.237342
dados\ELVIS 16_11_22\ATR_algodao_100%_med11_64_0.dx	0.258607	0.257706	0.259623	0.265807	0.262323	0.246814	0.235467	0.228923
...	...	...	...	...	...	...	...	...
dados\ELVIS 16_11_22\ATR_tururi_med19a_64_0.dx	-0.831153	-0.823646	-0.819586	-0.828186	-0.830313	-0.815811	-0.807711	-0.806726
dados\ELVIS 16_11_22\ATR_tururi_med20_64_0.dx	-0.666270	-0.668797	-0.667976	-0.665368	-0.654010	-0.640927	-0.640450	-0.638944
dados\ELVIS 16_11_22\ATR_tururi_med7_64_0.dx	-1.022497	-1.031750	-1.035382	-1.035508	-1.030751	-1.021948	-1.013803	-1.004018
dados\ELVIS 16_11_22\ATR_tururi_med8_64_0.dx	-0.998196	-1.000867	-1.002458	-1.003864	-1.000733	-0.988231	-0.973326	-0.964809
dados\ELVIS 16_11_22\ATR_tururi_med8_64_1.dx	-0.890706	-0.890422	-0.889619	-0.889811	-0.885727	-0.879846	-0.875490	-0.870083

Table 4 Example of FTIR database transposed matrix

	dados\ELVIS 16_11_22\ATR_algodao_100%f_med10_64_0.dx	dados\ELVIS 16_11_22\ATR_algodao_100%f_med9_64_0.dx	dados\ELVIS 16_11_22\ATR_algodao_100%_ med5_64_0.dx	dados\ELVIS 16_11_22\ATR_algodao_100%_ med6_64_0.dx
1799.705887	0.239914	0.336450	-0.006818	0.248586
1797.644368	0.239997	0.322537	-0.012591	0.249738
1795.582849	0.237747	0.305475	-0.015163	0.239999
1793.521330	0.228752	0.296521	-0.021785	0.221599
1791.459812	0.216106	0.301307	-0.028723	0.215353
...	...	...	...	...
608.148035	0.139677	0.324328	-0.098557	0.227733
606.086516	0.129919	0.319645	-0.096429	0.228995
604.024998	0.115195	0.313349	-0.092986	0.228233
601.963479	0.110438	0.307972	-0.090898	0.223386
599.901960	0.123932	0.311182	-0.094250	0.221571

583 rows × 91 columns

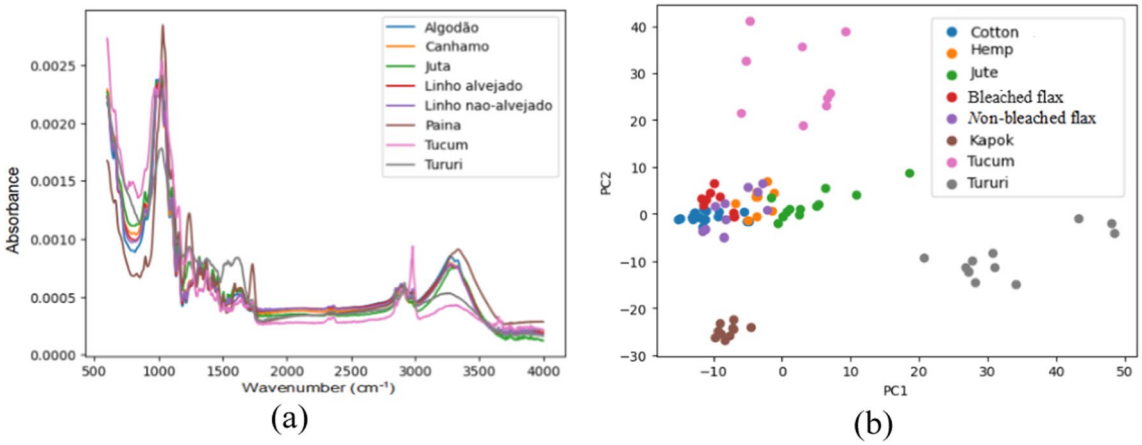


Fig. 2 **a** Average spectrum of fibers after mean normalization; **b** PC1 X PC2 in mean normalization

normalization was adopted for this database to enhance this differentiation. The results show that the two normalizations (mean and logarithm) together achieve a more effective separation in PCA. Five of the eight selected fibers formed distinct groups; cotton and flax fibers have very similar composition characteristics and make separation by PCA difficult (Fig. 5a and b).

From Fig. 5c, it is possible to notice that when removing the cotton and flax fibers, all the other fibers form distinct clusters. In Fig. 5d, the behavior of the flax and cotton fibers without the other fibers is clear.

The greatest difficulty in separating and distinguishing between fibers by PCA occurs between cotton and flax fibers (Fig. 5d); both fibers have a very similar chemical composition

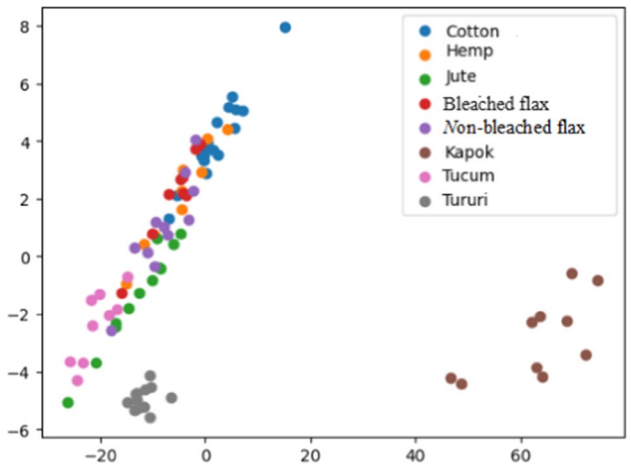


Fig. 3 PCA preceded by standardization

and proportion of cellulose, making distinguishability difficult, and some authors described the impossibility of distinguishing these fibers using only the unsupervised methodology (Peets et al. 2017, 2019; Riba et al. 2020; Saito et al. 2021).

LDA Preceded by Mean and Logarithm Normalization

To calculate the LDA model, first, the size of the datasets was reduced (i.e., training, testing, and real samples) (Table 5). In this LDA run, only the fingerprint range was employed (400–1700 nm reflectance spectra), and the number of PCs was selected based on the minimum number needed to obtain the minimum classification error in the LDA model to avoid overfitting. The best normalization that provided the most separation between the fibers was the normalization by mean and logarithmic carried out in the original database. In this way, the LDA was performed by taking this transform. With this supervised methodology, it was possible to separate all the fibers more efficiently (Fig. 6), although the values for flax fibers (both bleached and non-bleached) still coalesced into the same group, but the flax and cotton fiber values, which were the most problematic in the unsupervised methodology, were properly separated by LDA.

Comparison between the present study and literature

Noticeably, the three normalization metrics adopted individually for the database and previously applied to the PCA resulted in different results. As the third approach (mean and logarithm $\log_{10}$ normalization) was the most effectively associated with PCA, then this last one was the only one adopted for LDA. This procedure sequence is illustrated in the flowchart presented in Fig. 7.

The approach of mean and logarithm normalization was more successful in PCA performance since these treatments

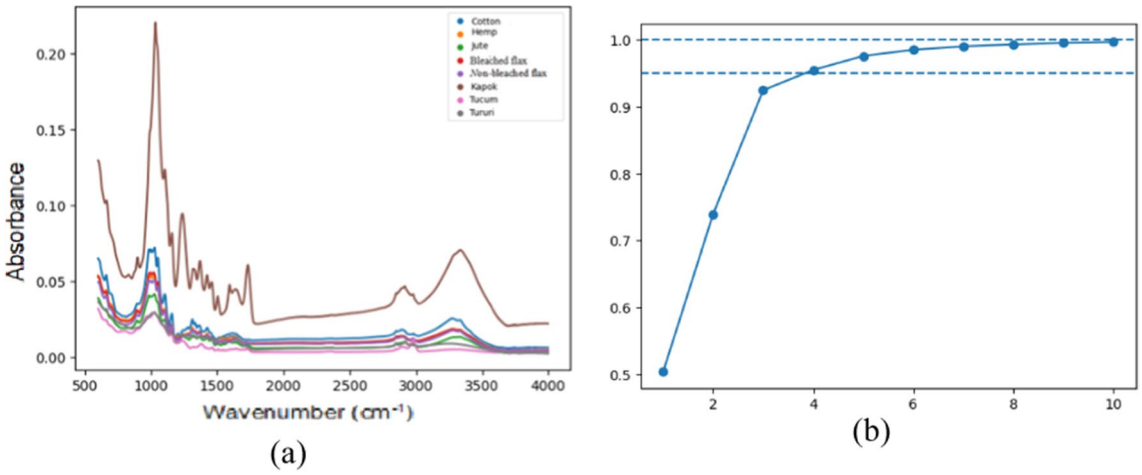


Fig. 4 **a** Average spectrum of fibers after standardization; **b** accumulated variance values

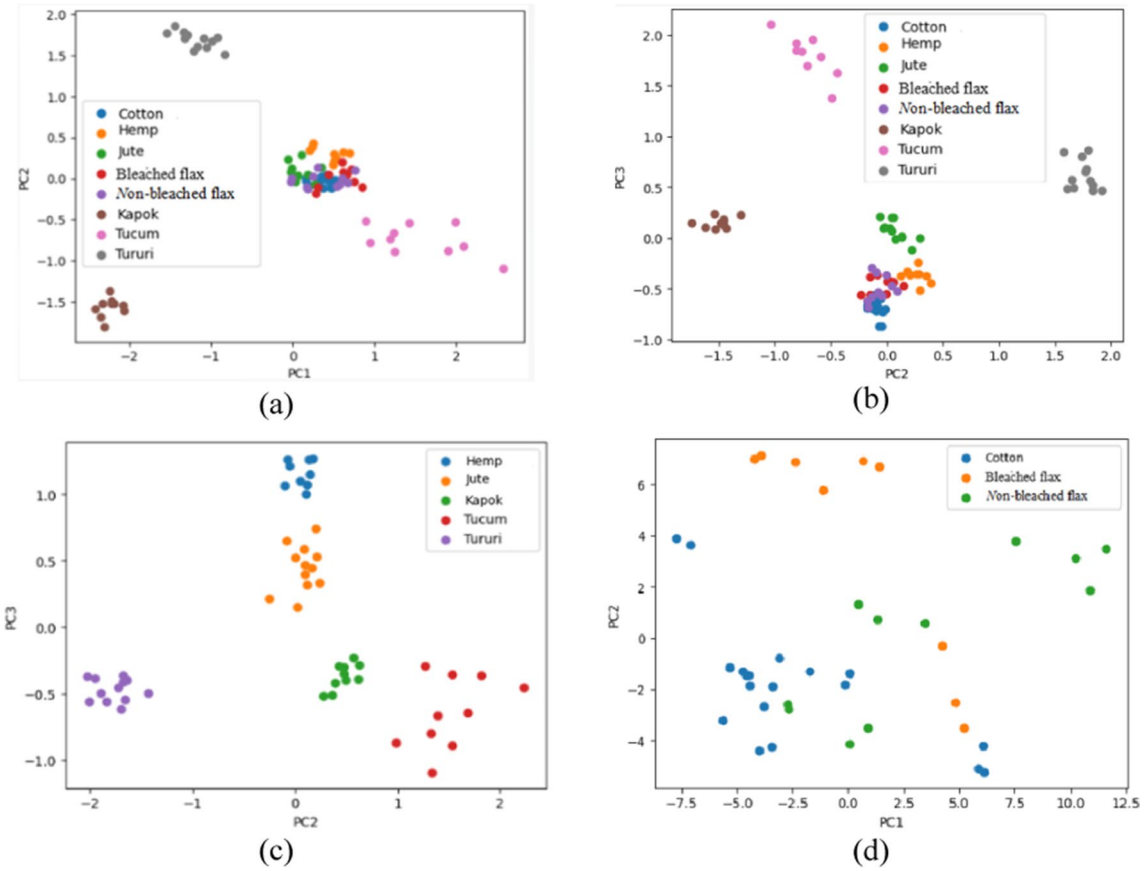
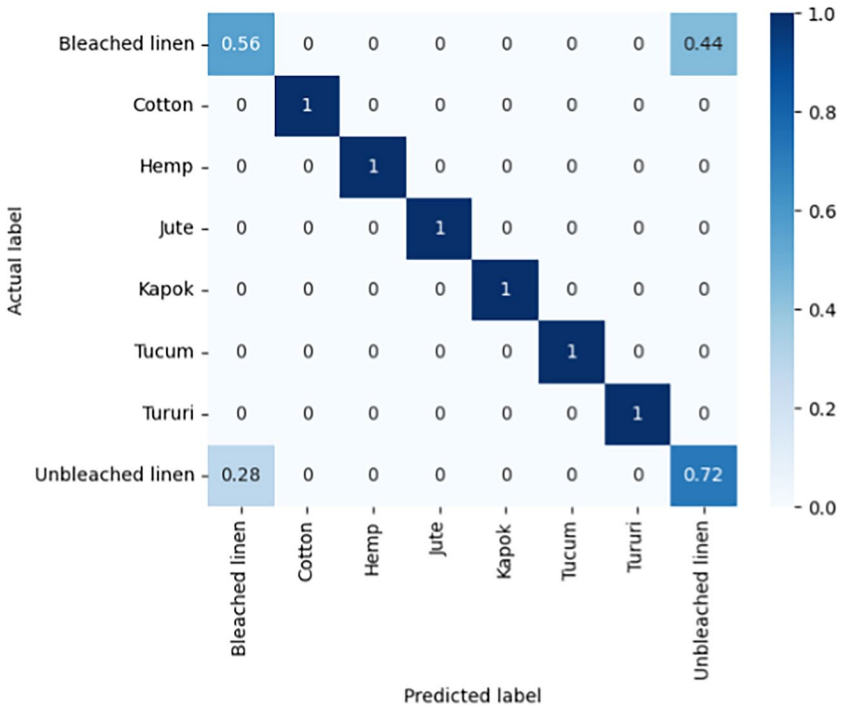


Fig. 5 **a** PC1 X PC2 logarithm base 10 normalization; **b** PC2 X PC3 logarithm base 10 normalization; **c** PC2 X PC3 logarithm normalization subgroups; **d** PC1 X PC2 logarithm normalization subgroups

Table 5 Train test and real sample prediction matrix



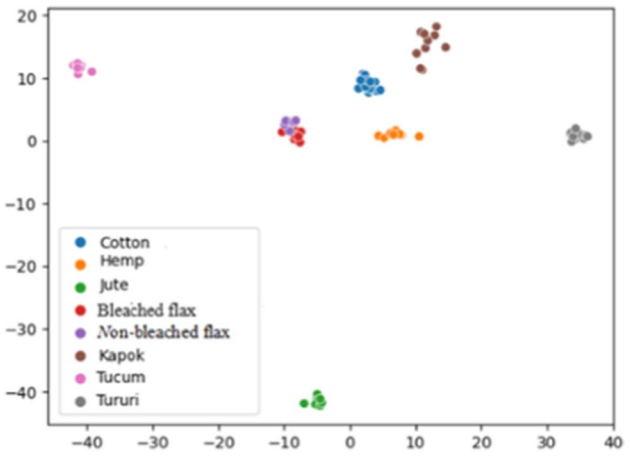


Fig. 6 LDA Classification for all studied fibers (cotton, kapok, hemp, non-bleached flax, bleached flax, jute, tucum, and tururi)

directly reduce noise associated with the fiber morphology. This noise is a result of non-linear and unequal absorption of the fiber regions irradiated by infrared. The normalization, by logarithm transform in particular, implies a linear relationship between reflectance and molar concentration, reducing signal multiplicative effects generated by the optical path when insufficient fiber quantity or poor filling is present, which absorbs the radiation in an inhomogeneous way.

The applied multivariate analysis procedures in this study and the main results expressed by the success in the separation of fibers are shown in Fig. 7, which are compared with the literature (Table 6).

Some authors have attempted to characterize and differentiate lignocellulosic fibers (Table 6), using spectroscopy in the mid-infrared region and with PCA algorithms (Peets et al. 2017, 2019), without success in differentiating very similar textile fibers such as cotton and flax. FTIR spectroscopy with PCA alone appears not to be sufficient to differentiate mixtures, but it easily separates and groups other textile species (Zhou et al. 2018, 2019) which corroborates this article regarding the difficulty in differentiating flax and cotton by PCA.

Separation of flax and cotton fibers with a 100% accuracy rate was achieved only using supervised along unsupervised methods (Riba et al. 2020). In this present work, it was only possible to effectively separate flax fibers using the supervised LDA method, which shows that for these cellulosic fibers with very similar chemical composition, separation using supervised classification methods is more effective.

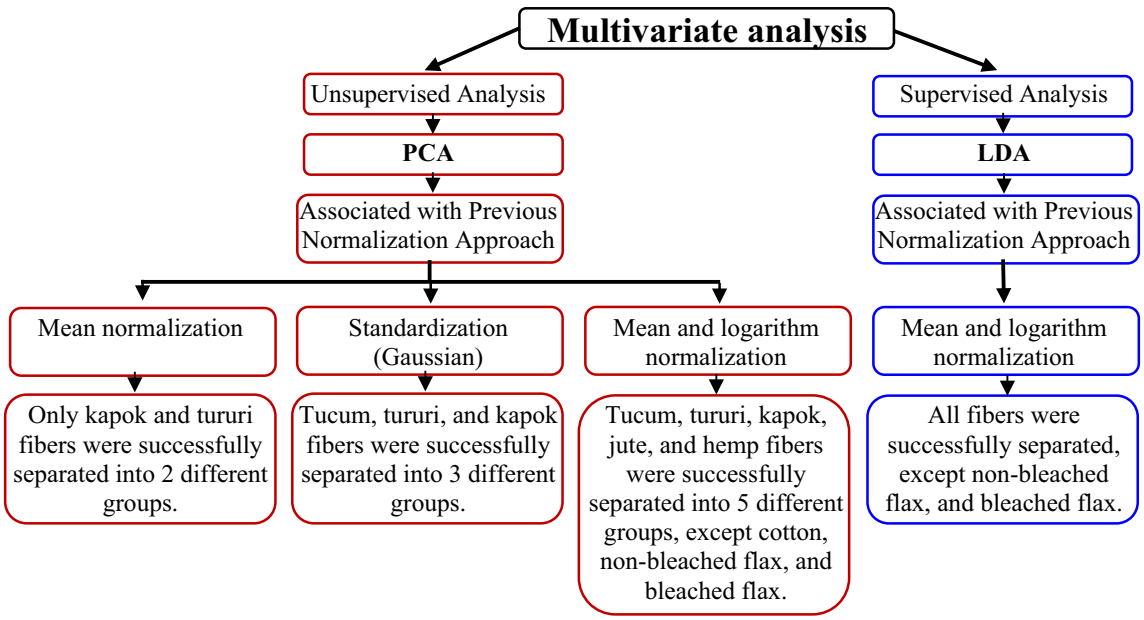


Fig. 7 Multivariate analysis procedures and main results

Table 6 Comparison of literature and the present study taking into account fiber characterization

Aim	Type of spectrum	Types of textile fibers	Mathematic algorithms	Recognition rate (%)	Reference
Differentiation	ATR-FTIR	Cotton, ramie, and linen (natural cellulosic); rayon/viscose, cupro, and lyocell (regenerated cellulosic)	PCA, FDOD	100% employing FDOD for cotton and bast fibers (linen and ramie) and for all other combinations (natural fiber − regenerated fiber, rayon − (cupro + lyocell), cupro − lyocell) 67% employing only PCA for cotton and bast fibers (linen and ramie)	Saito et al. (2021)
Differentiation	NIR	Cotton (natural cellulosic), wool, and silk (natural animal protein), and their mixtures	PCA, LDA, SIMCA	99.8% for single fibers and 97.5% for mixtures employing SIMCA	Quintero Balbas et al. (2022)
Textile recycling	ATR-FTIR	Polyamide and polyester (synthetic), viscose (regenerated cellulosic), cotton and flax (natural cellulosic), and wool and silk (natural animal protein)	PCA, CVA + kNN[a]	100%	Riba et al. (2020)
Textile recycling	NIR	Cotton (natural cellulosic); polyester, polyamide and acrylic (synthetic fibers); and wool and silk (natural animal protein)	MSC, SNV, SIMCA[b]	100% for polyester, polyamide, acrylic, silk and wool 90% for cotton and polyester fabric	Zhou et al. (2018)
Textile recycling	NIR	Cotton (natural cellulosic); lyocell/tencel (regenerated cellulosic); wool and cashmere (natural animal proteinic), polyethylene terephthalate/polyester (PET) and polypropylene (PP) (synthetic); polylactic acid (PLA) (corn starch bioplastic)	PCA, SIMCA, LDA	100% for cotton, tencel, PP, PLA, and PET Wool and cashmere fibers yielded confusing results	Zhou et al. (2019)
Quality control	ATR-FTIR	Cotton and flax (natural cellulosic); cellulose acetate, lyocell/tencel and viscose (regenerated cellulosic); wool and silk (natural animal protein); polyester, polyamide, polyacrylic and elastane (synthetic) and 15 two-component textiles	PCA	The main problems with ATR-FTIR-PCA classification are (1) difficulties in getting high quality spectra from some textiles (e.g., polyacrylic), (2) inhomogeneity of the textile fibers in the case of two-component fibers, and (3) intrinsic similarity between the spectra of some fibers (e.g., cotton and flax)	Peets et al. (2017)
Quality control	ATR-FTIR	Cotton, linen, jute, sisal (natural cellulosic); viscose, cellulose acetate, tencel/lyocell (regenerated cellulosic); wool and silk (natural animal protein); fiberglass (inorganic); polyester, polyamide, polyacrylic, elastane, polyethylene and polypropylene (synthetic)	PCA, random forest classification	Cellulose-based fibers (cotton, linen, jute, sisal, viscose), which are very similar in their chemical composition, are difficult to differentiate	Peets et al. (2019)
Textile recycling	NIR and imaging	Cotton (natural cellulosic), viscose, lyocell, modal (regenerated cellulosic), and fiber blends	PCA	Difficulties were related to the separation employing only PCA on NIR spectra data. However, the differentiation was highly enhanced by associating imaging spectroscopy analysis	Mäkelä et al. (2021)

Table 6 (continued)

Aim	Type of spectrum	Types of textile fibers	Mathematic algorithms	Recognition rate (%)	Reference
Differentiation	ATR-FTIR	Cotton, kapok, hemp, non-bleached flax, bleached flax, jute, tucum, and tururi (natural cellulosic)	PCA, LDA	Distinction between fibers employing PCA, except between cotton and flax (bleached and non-bleached) 100% employing LDA, except for bleached and non-bleached flax	Present study

Next, the nearest neighbor (kNN) classifier is applied, and this algorithm provides as many outputs normalized variables within the range 0–1 as types of textile fibers or classes defined in the problem, thus assigning an incoming textile sample to the class having the highest output value

FDA Fisher's discriminant analysis, *FDOD* Fisher's discriminant analysis orthogonal decomposition (enhanced FDA calculation method), *NIR* near-infrared spectroscopy (from 12,000 to 4000 cm^{-1}), *SIMCA* soft independent modeling of class analogy

[a]Principal component analysis (PCA) algorithm is applied followed by the canonical variate analysis (CVA) algorithm

[b]Multiplicative scattering correction (MSC) and standard normal variate (SNV) on the noise reduction of the spectrum were also studied. Meanwhile, the soft independent modeling of class analogy (SIMCA)

Conclusions

This study highlighted the advantages and disadvantages of two exploration or classification techniques (PCA and LDA) for the non-invasive classification of textile fibers (cotton, kapok, hemp, non-bleached flax, bleached flax, jute, tucum, and tururi) employing FTIR "fingerprint" database (400–1800 cm^{-1}). Despite the known applicability of these techniques in industry as well as statistical modeling, a limited number of publications report this methodology for the identification of textile fibers from a recycling perspective. The present PCA and LDA approach showed good performance for both techniques when analyzing individual fibers. However, flax (both non-bleached and bleached) and cotton fiber values without other types of pre-processing such as application of filters or derivatives are ineffective via PCA, whereas LDA showed the best performance.

The results suggest that PCA is a powerful tool in studying textiles with a relatively simple structure, while objects with a more complex or very similar composition, for example, LDA statistics, are more advantageous. Despite this good performance in the samples of this model, the study must continue and investigate the separation in textiles found in recycling scenarios, that is, textiles of different compositions, different industrial processing processes, and degree of degradability, and assess whether these conditions change significantly in the results of supervised multivariate statistics.

Acknowledgements The authors gratefully acknowledge CAPES (Coordination for the Improvement of Higher Education Personnel of Brazilian Education Ministry).

Author Contribution The authors, ECS, MAR, PHSR, and JBR, carried out laboratory experiments, data validation process, and design of the study. ECS, AABS, RRAF, MAR, PHSR, and JBR carried out analyses of experimental results associated with technical and statistical processes. The manuscript was written by ECS, while AABS, RRAF, MAR, PHSR, and JBR reviewed critically all versions of the manuscript. All authors, ECS, AABS, RRAF, MAR, PHSR, and JBR, have given final approval for the version to be published.

Data Availability All data generated or analyzed during this study are included in this published article, more information is available with the corresponding author on reasonable request.

Declarations

Competing Interests The authors declare no competing interests.

References

Abit (2023) Valor da produção de vestuário teve aumento de 0,5% em 2022. In: https://www.abit.org.br/noticias/valor-da-producao-de-vestuario-teve-aumento-de-05-em-2022#:~:text=Varejo%20%2D%20Em%202022%2C%20o%20consumo,8%25%20na%20compara%C3%A7%C3%A3o%20com%202021.

Bro R, Smilde AK (2014) Principal component analysis. Anal Methods 6:2812–2831. https://doi.org/10.1039/C3AY41907J

Chowdhury NR, Paul SK, Sarker T, Shi Y (2023) Implementing smart waste management system for a sustainable circular economy in the textile industry. Int J Prod Econ 262:108876. https://doi.org/10.1016/j.ijpe.2023.108876

de Oliveira Neto GC, Correia JMF, Tucci HNP et al (2022) Sustainable resilience degree assessment of the textile industrial by size: incremental change in cleaner production practices considering circular economy. J Clean Prod 380:134633. https://doi.org/10.1016/j.jclepro.2022.134633

Ferreira MMC (2015) Quimiometria: conceitos, métodos e aplicações. Editora da Unicamp

Gaustad G, Krystofik M, Bustamante M, Badami K (2018) Circular economy strategies for mitigating critical material supply issues. Resour Conserv Recycl 135:24–33. https://doi.org/10.1016/j.resconrec.2017.08.002

Haque F, Fan C, Lee Y-Y (2023) From waste to value: addressing the relevance of waste recovery to agricultural sector in line with circular economy. J Clean Prod 415:137873. https://doi.org/10.1016/j.jclepro.2023.137873

Hospodarova V, Singovszka E, Stevulova N (2018) Characterization of cellulosic fibers by FTIR spectroscopy for their further implementation to building materials. Am J Analyt Chem 09:303–310. https://doi.org/10.4236/ajac.2018.96023

Krystofik M, Luccitti A, Parnell K, Thurston M (2018) Adaptive remanufacturing for multiple lifecycles: a case study in office furniture. Resour Conserv Recycl 135:14–23. https://doi.org/10.1016/j.resconrec.2017.07.028

Mäkelä M, Rissanen M, Sixta H (2021) Identification of cellulose textile fibers. Analyst 146:7503–7509. https://doi.org/10.1039/D1AN01794B

Navarro A, Puig R, Martí E et al (2018) Tackling the relevance of packaging in life cycle assessment of virgin olive oil and the environmental consequences of regulation. Environ Manage 62:277–294. https://doi.org/10.1007/s00267-018-1021-x

Nayak RK, Padhye R, Fergusson S (2012) Identification of natural textile fibres. Handb Nat Fibres 314–344. https://doi.org/10.1533/9780857095503.1.314

Palacios-Mateo C, van der Meer Y, Seide G (2021) Analysis of the polyester clothing value chain to identify key intervention points for sustainability. Environ Sci Eur 33:2. https://doi.org/10.1186/s12302-020-00447-x

Peets P, Leito I, Pelt J, Vahur S (2017) Identification and classification of textile fibres using ATR-FT-IR spectroscopy with chemometric methods. Spectrochim Acta A Mol Biomol Spectrosc 173:175–181. https://doi.org/10.1016/J.SAA.2016.09.007

Peets P, Kaupmees K, Vahur S, Leito I (2019) Reflectance FT-IR spectroscopy as a viable option for textile fiber identification. Herit Sci 7:93. https://doi.org/10.1186/s40494-019-0337-z

Quintero Balbas D, Lanterna G, Cirrincione C et al (2022) Non-invasive identification of textile fibres using near-infrared fibre optics reflectance spectroscopy and multivariate classification techniques. Eur Phys J Plus 137. https://doi.org/10.1140/epjp/s13360-021-02267-1

Riba J-R, Cantero R, Canals T, Puig R (2020) Circular economy of post-consumer textile waste: classification through infrared spectroscopy. J Clean Prod 272:123011. https://doi.org/10.1016/j.jclepro.2020.123011

Saccani N, Bressanelli G, Visintin F (2023) Circular supply chain orchestration to overcome circular economy challenges: an empirical investigation in the textile and fashion industries. Sustain Prod Consum 35:469–482. https://doi.org/10.1016/j.spc.2022.11.020

Saito K, Yamagata T, Kanno M et al (2021) Discrimination of cellulose fabrics using infrared spectroscopy and newly developed discriminant analysis. Spectrochim Acta A Mol Biomol Spectrosc 257:119772. https://doi.org/10.1016/J.SAA.2021.119772

Sandin G, Peters GM (2018) Environmental impact of textile reuse and recycling – a review. J Clean Prod 184:353–365. https://doi.org/10.1016/j.jclepro.2018.02.266

Sousa-Zomer TT, Magalhães L, Zancul E, Cauchick-Miguel PA (2018) Exploring the challenges for circular business implementation in manufacturing companies: an empirical investigation of a pay-per-use service provider. Resour Conserv Recycl 135:3–13. https://doi.org/10.1016/j.resconrec.2017.10.033

Stuart BH (2004) Infrared spectroscopy: fundamentals and applications. Wiley

Zhou C, Han G, Via BK et al (2018) Rapid identification of fibers from different waste fabrics using the near-infrared spectroscopy technique. Text Res J 89:3610–3616. https://doi.org/10.1177/0040517518817043

Zhou J, Yu L, Ding Q, Wang R (2019) Textile fiber identification using near-infrared spectroscopy and pattern recognition. Autex Res J 19:201–209. https://doi.org/10.1515/aut-2018-0055