Spectral preprocessing and instrumentation for the identification of camelid and goat fibers using artificial intelligence

Max Quispe, Luis Serrano-Arriezu, Edgar Quispe, Miguel Beruete & Jesús D. Trigo

Published online: 02 Sep 2025.

Submit your article to this journal ☑

View related articles ☑

View Crossmark data ☑

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS · Check for updates

Spectral preprocessing and instrumentation for the identification of camelid and goat fibers using artificial intelligence

Max Quispe[a,b] (iD), Luis Serrano-Arriezu[a,c], Edgar Quispe[d,e], Miguel Beruete[a,c] and Jesús D. Trigo[a,c]

[a]Institute of Smart Cities (ISC), Public University of Navarre (UPNA), Pamplona, Spain; [b]Research and Technological Development (RTD), MAXCORP Technologies S.A.C, Lima, Peru; [c]Navarra Institute for Health Research (IdiSNA), Pamplona, Spain; [d]Center for Research and Technological Development, Natural Fiber's Tech S.A.C, Lima, Peru; [e]Department of Animal Production, Universidad Nacional Agraria La Molina, Lima, Peru

ABSTRACT

Luxury fibers, derived from South American camelids (SAC)—alpaca and vicuña—and goats (cashmere and mohair), are highly valued in fashion and haute couture for their softness, lightness, thermal properties, and scarcity. They are obtained through a meticulous process and hold great cultural value. For this reason, they are closely linked to key factors such as price, quality, and comfort. It is therefore essential to identify their origin and characteristics in order to carry out proper evaluation and quality control. In a previous study, the authors developed and validated two machine learning models: deep neural networks (DNN) and support vector machine (SVM), to quickly, easily, and automatically identify Fourier-Transform Infrared (FTIR) spectra of SAC and caprine fibers. However, in such effort, the traditional spectral band range with no preprocessing was used. In addition, only an FTIR spectrometer in Attenuated Total Reflectance (ATR) mode was used. Thus, in this paper, the effects of, on one hand, pre-process the spectra and, on the other hand, using a FTIR microspectrometer in ATR mode are analyzed. For the former, two separate actions were tested: removing the CO_2 band and selecting a characteristic spectral band (usually referred to as "the fingerprint region"). Combining all these options (i.e. machine learning algorithm, spectral pre-processing and FTIR equipment), five additional different experiments were carried out, which were compared with the previously achieved results. Samples from SAC (alpaca: $n = 51$, llama: $n = 50$, vicuña: $n = 50$) and goats (mohair: $n = 35$, cashmere: $n = 20$) were evaluated, yielding valid FTIR-ATR spectra ($n = 1236$) and m-FTIR-ATR spectra ($n = 1220$). A 100% accuracy was achieved when evaluating m-FTIR-ATR spectra of alpacas with the DNN model, and a 98.3% accuracy with the SVM model. After removing the CO_2 band from the FTIR-ATR spectra, accuracy dropped to 95.12%, which is 1.63 percentage points lower than the 96.75% reported in a previous study by the authors, where the whole spectra was evaluated. On the other hand, evaluating the fingerprint region with the models resulted in an accuracy of up to 89.43%, indicating that important information is lost when the rest of the spectrum region is removed. Regardless of the spectrometry technique used, DNN models provided higher accuracy than SVM models, although performance improved with FTIR-ATR spectra. In conclusion, machine learning models can be used to identify and assess spectral differences between the fibers of diverse SAC and Caprine species.

Introduction

Luxury fibers, sourced from species such as South American camelids (SACs)—including alpaca, vicuña, llama, and guanaco—and goats (cashmere and mohair), among others, are integral to the production of textiles and garments. These animal fibers are closely linked to key factors such as price, quality, and comfort. Therefore, it is essential to know the origin of the fibers and their characteristics in order to carry out an adequate evaluation and quality control. The price per kg varies within a wide range: The wool with a diameter of less than 22 µm costs between US\$ 3.4 and 3.9, while dehaired vicuña fiber can range from US\$ 300 to 800 (Quispe et al., 2014). Cashmere (referred to as "soft gold") often exceeds US\$ 100 (Tan et al., 2019), alpaca Royal Baby, Baby and Fine adult fibers from Australia varied between US\$ 13.4 to 24.11 per kg (Shim, 2003), while the mohair fiber costs around US\$ 4.2.

CONTACT Max Quispe ✉ maxdavid22@gmail.com · Institute of Smart Cities (ISC), Public University of Navarre (UPNA), Arrosadia Campus, Pamplona, 31006 Spain.

As previously described by Quispe et al. (2025), as well as by other authors (Lu et al., 2019; Tan et al., 2019; Wu & He, 2008; Xing et al., 2020), the unfaithful substitution of high price fleeces (such as vicuña or cashmere) for cheaper fibers is a common issue in the market. This undermines the interests of both producers and consumers and makes the identification of cashmere and fibers from other species especially crucial to promoting an honest textile production ecosystem. Furthermore, as Quispe et al. (2025) noted, various methods for the identification and discrimination of natural fibers such as wool, cashmere, mohair, and alpaca have been explored in recent years, including optical and electron microscopes, chemical methods, physical spectrometry, deoxyribonucleic acid (DNA) analysis, and advanced statistical techniques. Some studies, including those by McGregor et al. (2018), have shown that certain spectrophotometry techniques have not yielded particularly promising results for fiber identification by species.

Fourier-Transform Infrared (FTIR)-based spectroscopy involves emitting infrared radiation onto a sample and analyzing the wavelengths it absorbs. Each compound exhibits a unique absorption pattern in the infrared spectrum, which enables the identification of the materials present in the sample. However, a limitation of this technique is the presence of CO_2 in the environment, which generates an absorption band in the infrared range. This band can interfere with measurements, particularly when analyzing samples that require high precision, as it may mask or distort the absorption signals of the compounds of interest (Hsu, 1997; Kuc, 2018; Nayak et al., 2020). Another nontrivial issue is the selection of an appropriate band to be analyzed. In the particular case of caprines and SACs fibers, this is usually in the range of 4000–600 cm^{-1} (McGregor et al., 2018, Peets et al., 2019). However, arguably, there appear to be higher differences in the band from 1000–4000 cm^{-1}, a region of the spectrum where key molecular vibrations are typically observed (McGregor et al., 2018).

In contrast to conventional FTIR spectroscopy, FTIR microspectroscopy (m-FTIR) incorporates an infrared microscope, often combined with optical imaging for sample visualization, allowing spectroscopic measurements at the microscopic level. This technique is valuable for studying small or complex samples, as it provides detailed information about the chemical distribution of compounds in specific areas of the sample, thereby enabling the analysis of heterogeneous materials, such as biological tissues or polymers (Barraza-Garza et al., 2013).

Peets et al. (2019) indicate that spectrometry is an appropriate technique for the analysis of textile samples. Furthermore, in their research, they employed both FTIR spectrometry and m-FTIR, using the Attenuated Total Reflectance (ATR) mode in both cases, and observed slight differences between the spectra obtained with the two techniques. When using the FTIR-ATR spectrometer, the spectrum quality was poor, with low absorbance intensity and a noisy baseline; however, absorbance bands characteristic of certain polyamide-based fibers were still identifiable.

Currently, the emerging technology known as Artificial Intelligence (AI) has gained significant momentum. More specifically, deep learning techniques have proven to be effective in rapid object recognition (Goel et al., 2019; Krizhevsky et al., 2017), and are already being employed for the analysis, evaluation, and recognition of medullation types in fibers from SACs and goats (Quispe et al., 2022a; 2022b; 2023).

In a previous study, the authors employed FTIR spectrometry in ATR mode, complemented by two supervised learning AI models: deep neural networks (DNN) and support vector machine (SVM) for the identification of SACs and goat fibers based on their spectra (Quispe et al., 2025). The results were highly promising, as these AI models demonstrated remarkable precision in identifying the spectra of alpaca, llama, vicuña, mohair, and cashmere fibers. The authors also concluded that the DNN model is the most suitable for evaluating both SACs and goat fibers. However, both the DNN and SVM models are effective for evaluating and identifying vicuña fibers. Conversely, the SVM model was found to be the best candidate for evaluating and identifying cashmere fibers. The spectra used in that and in the present study correspond to animal-based fibers.

However, in such previous research, the authors used the traditional spectral band range with no preprocessing. In addition, only the FTIR-ATR mode was used. Therefore, two underlying hypotheses are uppermost in this approach. First, pre-processing the signals acquired from the device with different techniques could lead to better identification results. These techniques would be based on removing parts of the signal that *a priori* could be considered not necessary to identify the species properly or that may be source of undesirable noise. Second, using a different device for acquiring the data could provide less noisy signals that would result in a better performance of the AI models, thereby achieving a higher percentage of correct classification guesses.

Thus, in this paper, the effects of, on one hand, pre-processing the spectra and, on the other hand, using a FTIR microspectrometer in ATR mode are analyzed. For the former, two separate actions were tested: removing the CO_2 band and selecting a characteristic spectral band (usually referred to as "the fingerprint region"). Combining all these options (i.e. machine learning algorithm, spectral pre-processing and FTIR equipment), five additional different experiments were carried out, which were compared with the previously achieved results.

Materials and methods

Research sites and duration

The AI models were developed and evaluated at the Research and Technological Development Laboratory of Maxcorp Technologies SAC (Lima, Peru). Fiber sample preparation was conducted at the Textile Fiber Laboratory of Natural Fiber's Tech SAC (Lima, Peru), while the analysis of fiber samples using the FTIR spectrometer Bruker Vertex 80v and Bruker Hyperion 3000 microscope was carried out at the Multispectral Biosensing Laboratory of Navarrabiomed (Pamplona, Spain). The field and laboratory work lasted approximately 11 months.

Characteristics of the fibers

The alpaca, llama, vicuña, cashmere, and mohair fiber samples were obtained from the regions of Huancavelica and Puno (Peru) and Bariloche (Argentina). The fibers samples of two genera (goats and SAC) and five species (mohair and cashmere, alpaca, llama, vicuña) were used. The fiber samples per species were as follows: alpaca = 51, llama = 50, vicuña = 50, mohair = 35, and cashmere = 20.

All samples were collected from the mid-side area, located over the third-to-last rib, halfway between the midline of the belly and the midline of the back at the level of the tenth dorsal rib, following the guidelines of McGregor et al. (2012).

Moreover, as a reference, the average fiber diameter (AFD) was calculated using de Fiber EC equipment. The values, expressed as "average ± standard deviation", are as follows: alpaca $(n = 51)$: $21.79 \pm 5.67\,\mu m$; llama $(n = 50)$: $24.16 \pm 4.62\,\mu m$; vicuña $(n = 50)$: $13.13 \pm 3.07\,\mu m$; mohair $(n = 35)$: $30.94 \pm 6.87\,\mu m$; and cashmere $(n = 20)$: $15.26 \pm 3.65\,\mu m$.

Finally, alpaca, llama and mohair samples were of white color; while cashmere ones were grey and those from vicuñas were beige.

Sample preparation

Similar to the method described by Quispe et al. (2025), the fiber samples were washed in a solution consisting of 7 parts 96% ethyl alcohol and 3 parts benzene. They were then dried using a towel pressed with a roller (Quispe et al., 2022a). Fiber snippets, ranging from 1.5 to 2 mm in length, were obtained using a fiber cutter. These fragments were properly labeled and kept in Eppendorf tubes. Temperature and relative humidity preconditioning was set at $20 \pm 2\,°C$ and $65 \pm 3\%$, respectively.

FTIR-ATR microspectroscopic analysis

The Bruker Hyperion 3000 FTIR microscope (m-FTIR) was used in ATR mode (m-FTIR-ATR). Measurements were performed with the following specifications: LN-MCT-D316-025/TE-InGaAs detector, cooled with liquid nitrogen, KBr beam splitter, and a Globar mid infrared MIR source. The following parameters were used: spectral range = $400–4000\,cm^{-1}$; resolution = $4\,cm^{-1}$; aperture = 1 mm; scan rate = 40 kHz; and number of scans = 8. The ATR objective 20x with a conical germanium crystal (100 μm) was used, with a pressure force of 40% (second level). First, the samples were placed on an aluminum plate. Second, a single fiber was located and focused from the set of fibers using the microscope in visible mode (see Figure 1A). Third, pressure was applied to the fiber using the germanium crystal (see Figure 1B). Finally, the samples were scanned and analyzed. Data were collected using version 7.5 of Bruker's OPUS software.

To expand our dataset and minimize potential noise artifacts caused by pollution, varying fiber diameters, or differences in medullation (factors that could affect the effective training of our artificial intelligence model), each sample was divided into 3 subsamples. Each subsample was then scanned twice at different points along the fiber. As a result, 6 FTIR-ATR spectra were collected per identified sample. After filtering and removing spectra with artifacts, a total of 1220 m-FTIR-ATR spectra were available for analysis and evaluation.

FTIR-ATR spectroscopic analysis

In accordance with the methodology used by Quispe et al. (2025), the analysis was performed using the FTIR Bruker Vertex 80v spectrometer, which was equipped with the Platinum ATR accessory. This accessory features a diamond prism, upon which the samples were placed for analysis. The spectrometer
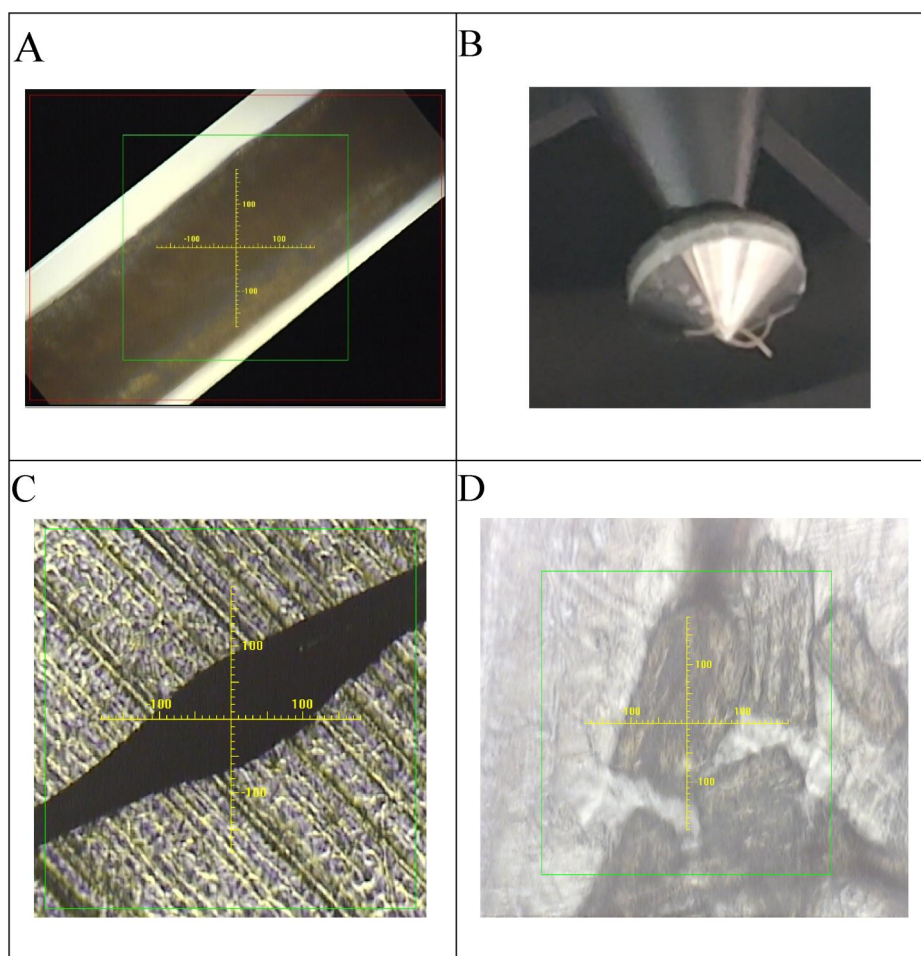
Figure 1. A) Focused alpaca fiber sample and aperture closure to obtain spectra with the FTIR microscope from the focused area, B) conical germanium crystal used for pressing and obtaining spectra with the FTIR microscope, C) fiber after being pressed with the conical germanium crystal, D) fiber damaged after being pressed with the conical germanium crystal.

was set up with the following specifications: Room Temperature-Deuterated L-alanine doped Triglycine Sulphate (RT-DLaTGS) detector, KBr beam splitter, and a Globar MIR source. The measurement settings included: spectral range $= 400$–$4000\,cm^{-1}$; resolution $= 4\,cm^{-1}$; aperture $= 6\,mm$; scan rate $= 10\,kHz$; sample temperature (room temperature) $= 20\,°C$; number of scans $= 8$. Spectra were obtained and processed using Bruker's OPUS software (Version 7.5). A gripper was utilized to place a small portion of the sample onto the ATR crystal, and pressure was applied before obtaining the FTIR-ATR spectrum. The interferometer was vacuum purged to minimize atmospheric gas absorption artifacts, but normal atmospheric pressure was preserved at the sample compartment.

Using a procedure similar to the one described for m-FTIR-ATR, 6 FTIR-ATR spectra were obtained for each identified sample. In this case, the total number of FTIR-ATR spectra available for analysis and evaluation was 1,236.

Application of machine learning

The data analysis method used was supervised learning. Initially, several machine learning models were trained, with the best results obtained using two specific models (DNN and SVM). The models were iteratively trained with various hyperparameters, which were adjusted in a heuristic yet supervised manner, until the established accuracy criteria were met. Python was employed as the programming language, together with TensorFlow, an open-source framework for machine learning applications.

The valid dataset obtained with FTIR-ATR ($n = 1,236$) was divided into three subsets as follows: the training dataset ($n = 792$) and the validation dataset ($n = 198$), both randomly selected during each training phase, and the test dataset ($n = 246$), which was set aside for the subsequent evaluation of the chosen models. The valid dataset obtained with m-FTIR-ATR ($n = 1,220$) was also divided into three subsets in the same manner and for the same purpose

Table 1. Summary of machine learning experiments conducted in this study, including key configurations and one prior experiment by Quispe et al. (2025).

Spectral region	Experiment	Technique	ML Algorithms	Hyperparameters
Whole region	0*	FTIR-ATR	DNN	layers = 4, model = sequential, epochs = 1000, batch = 64, LR = 0.005, optimizer = Adam, loss = categorical_crossentropy, activation = ReLU/Softmax
			SVM	classifier = SVC, kernel = RBF, $C = 80$
	1	m-FTIR-ATR	DNN	LR = 0.01; others as in experiment 0
			SVM	same as experiment 0
CO_2 band removed	2	FTIR-ATR	DNN	same as experiment 0
	3	m-FTIR-ATR	DNN	batch = 32, LR = 0.01; the rest as in experiment 0
Fingerprint region	4	FTIR-ATR	DNN	same as experiment 0
	5	m-FTIR-ATR	DNN	LR = 0.01; the rest as in experiment 0

*Experiment 0 was conducted by Quispe et al. (2025) using FTIR-ATR spectra and is included here for comparison purposes.
Abbreviations: m-FTIR-ATR = Fourier transformed infrared microspectroscopy with attenuated total reflectance. FTIR-ATR = Fourier transformed infrared spectroscopy with attenuated total reflectance. DNN = deep neural network. SVM = support vector machine. LR = learning rate.

described in the previous paragraph, resulting in the training dataset ($n = 779$), validation dataset ($n = 195$), and test dataset ($n = 246$).

Each model was trained using the training dataset and initially validated with the validation dataset. Subsequently, the accuracy and performance of the two chosen models were validated using the test dataset. The evaluation of the chosen models for both DNN and SVM was performed with FTIR-ATR and m-FTIR-ATR spectra, divided into the following quantities: alpaca (60), llama (60), vicuña (60), cashmere (42), and mohair (24) for both cases.

The hyperparameters used for training the DNN model with the m-FTIR-ATR, in the whole region ($400–4000 \, cm^{-1}$), were: layers = four, model = sequential, epochs = 1000, batch size = 64, learning rate = 0.01, optimizer = Adam, loss = categorical_crossentropy, and activation = relu and softmax. The hyperparameters used for training the SVM model, in the whole region, were: classifier = support vector classifier, $C = 80$, and kernel = radial basis function.

The hyperparameters used for training the DNN models with the m-FTIR-ATR and FTIR-ATR spectra, excluding the CO_2 region, were: layers = four, model = sequential, epochs = 1000, batch size = 32 and 64 (respectively), learning rate = 0.01 and 0.005 (respectively), optimizer = Adam, loss = categorical_crossentropy, and activation = ReLU and softmax.

The hyperparameters used for training the DNN models with the m-FTIR-ATR and FTIR-ATR spectra, within the range of 750 to $1750 \, cm^{-1}$, were: layers = four, model = sequential, epochs = 1000, batch size = 64, learning rate = 0.01 and 0.005 (respectively), optimizer = Adam, loss = categorical_crossentropy, and activation = ReLU and softmax.

To improve clarity and facilitate comparison across the experiments, the main configurations of each machine learning implementation—comprising model type, spectroscopic technique, and key hyperparameters—are summarized in Table 1. This table also includes Experiment 0, previously reported by Quispe et al. (2025), as a reference point.

Experiments

In order to analyze the spectra, there are different choices available, namely: the measured spectra, the instrumentation, and the machine-learning algorithm. Regarding the measured spectra, it can be chosen the whole region, the fingerprint region, or the whole region without the CO_2 band. As regards the equipment, either the spectroscope or the microscope combined with FTIR can be used. Finally, two machine learning models are considered (DNN or SVM). Thus, in this research, 5 different experiments were performed (see Table 1).

For the fiber samples spectra obtained with the m-FTIR-ATR, whose band range is between 400 and $4000 \, cm^{-1}$ (see Figure 2), and in order to ensure that no potentially relevant information is lost, two machine learning models (DNN and SVM) were developed and validated. These experiments will hereafter be referred to as m-FTIR-ATR+DNN and m-FTIR-ATR+SVM (Experiment 1).

For the spectra obtained with both m-FTIR-ATR and FTIR-ATR, and with the goal of eliminating potential biases, reducing the obtained spectra, and identifying specific bands that could contain information differentiating the various animal species previously evaluated, the CO_2 region was removed from the fiber spectra (see Figure 3A and 3B). Two new DNN models were then developed (Experiments 2 and 3). Subsequently, the spectra were reduced to the range between 750 and $1750 \, cm^{-1}$ (see Figure 3C
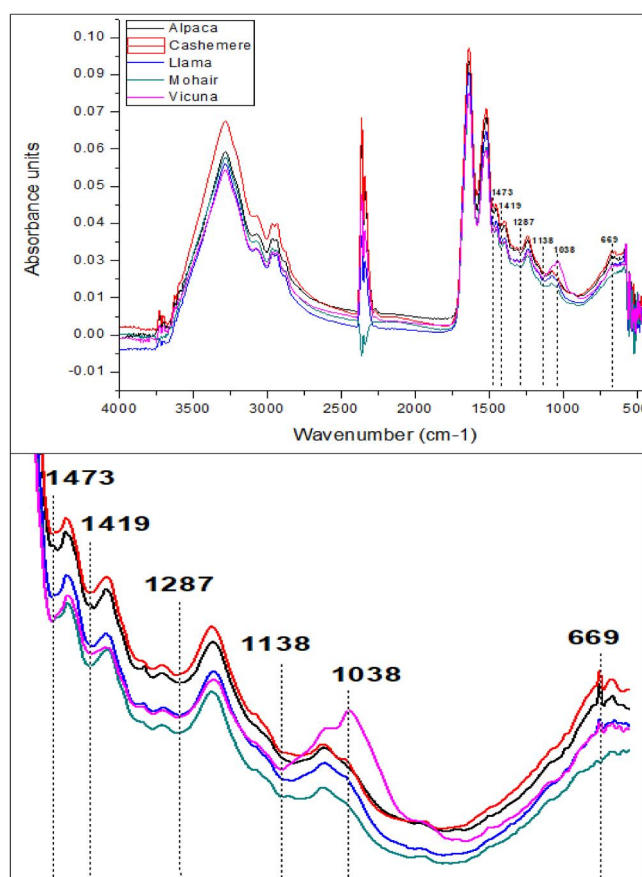
Figure 2. Spectra of alpaca, llama, vicuña, mohair and cashmere fibers obtained with fourier Transform infrared microspectrometer with Attenuated total Reflectance (m-FTIR-ATR) at wavelengths between 4000 and 400 cm^{-1} (top). Zoomed spectra showing the differences in peaks among the different fibers with the respective wavenumbers annotated (bottom).

and 3D), to focus on the region known as the fingerprint, and two additional DNN models were developed (Experiments 4 and 5).

Statistical analysis

Bar graphs were used to represent the hit and miss percentages for the different spectra obtained through m-FTIR-ATR and processed with the DNN and SVM models. A proportion comparison test with Yates' correction was then performed to compare the number of hits with the total number of m-FTIR-ATR spectra evaluated for each species type, for each configuration under analysis. This analysis was performed using the free software RStudio.

Results

Spectral comparison of m-FTIR-ATR and FTIR-ATR

The averaged spectra of mohair, cashmere, llama, alpaca, and vicuña fibers obtained with m-FTIR-ATR are shown in Figure 2. Clear differences between the fibers can be observed in the regions of 1473, 1419,

1287, 1138, 1038, and 669 cm^{-1}. At the wavelengths of 1473 and 1419 cm^{-1}, alpaca and llama fibers show slightly higher peaks compared to cashmere and mohair fibers, while vicuña fibers lack this characteristic. Furthermore, at 1287 cm^{-1}, alpaca fibers exhibit a slight peak, while the spectra of the other species do not show this feature. Between the bands at 1138 and 1038 cm^{-1}, the greatest differences are observed between vicuña fibers and the other types of fibers. At the wavelength of 669 cm^{-1}, mohair fibers do not display a peak compared to the other fibers.

The averaged spectra obtained using FTIR-ATR within the spectral range of 400–4000 cm^{-1} were similar to those obtained by Quispe et al. (2025). Meanwhile, the spectra (m-FTIR-ATR and FTIR-ATR) with the CO_2 region removed and subsequently reduced within the 750–1750 cm^{-1} region (fingerprint region) are shown in Figure 3. It can be observed that the FTIR-ATR spectra (Figure 3B) for all species shows bands around 2000 cm^{-1}, whereas this characteristic is not present in the m-FTIR-ATR spectra (Figure 3A). Additionally, a smaller slope is observed between the peaks of the FTIR-ATR spectra compared
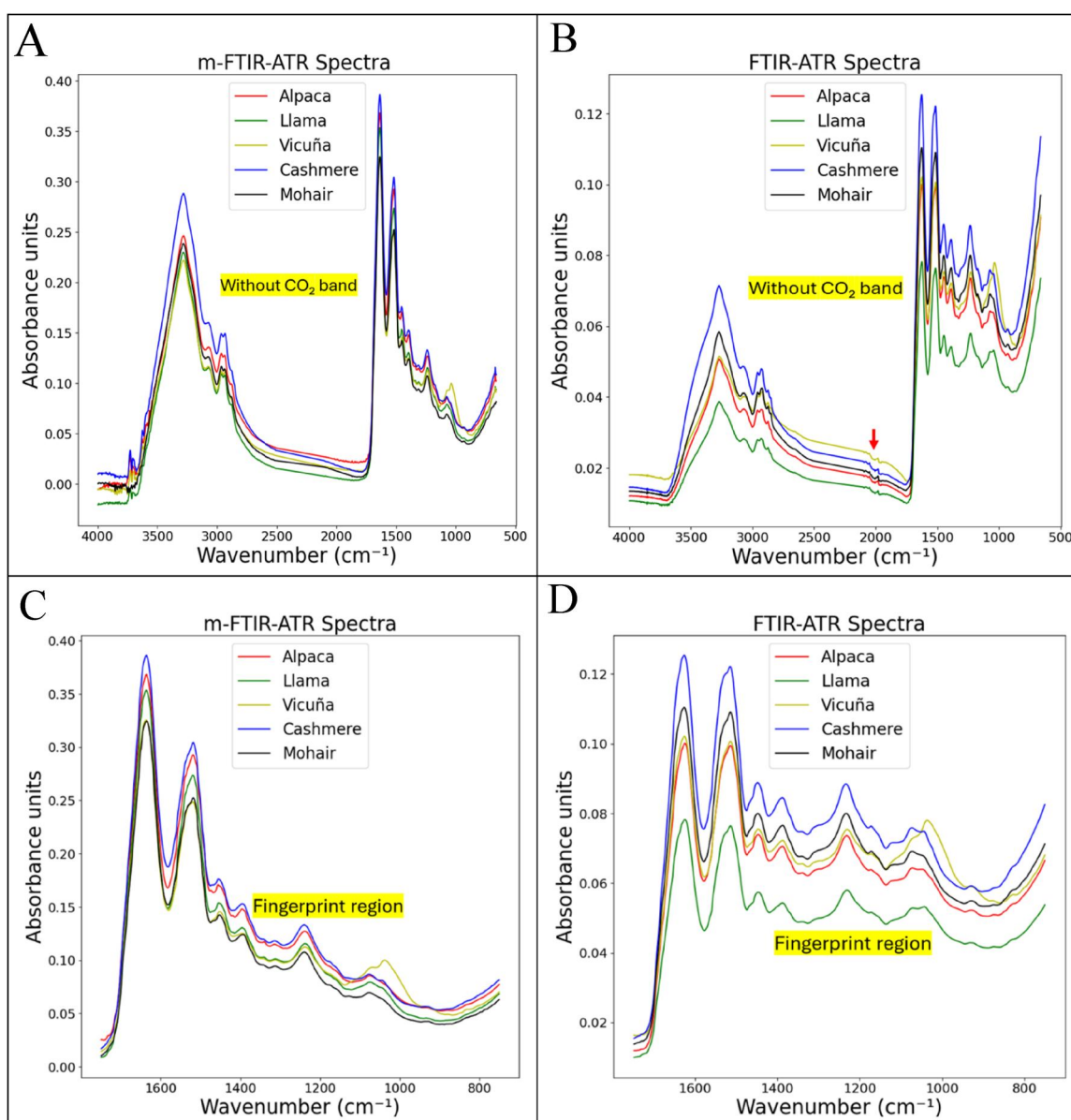
Figure 3. A) Spectra obtained with an FTIR micro-spectrometer in ATR mode (m-FTIR-ATR) with the CO_2 band removed (regions 661–2100 and 2500–4000 cm^{-1}), B) spectra obtained with an FTIR spectrometer in ATR mode (FTIR-ATR) with the CO_2 band removed, C) spectra obtained with an FTIR micro-spectrometer in ATR mode (m-FTIR-ATR) from the fingerprint region (750–1750 cm^{-1}), D) spectra obtained with an FTIR spectrometer in ATR mode (FTIR-ATR) from the fingerprint region.

to those obtained from m-FTIR-ATR. Furthermore, the absorbance levels for the FTIR-ATR spectra were lower in comparison with the m-FTIR-ATR spectra obtained.

Accuracy of AI models per species and genera using m-FTIR-ATR

Table 2 shows the accuracy when the training, validation, and test datasets are evaluated using the m-FTIR-ATR spectra. Results show a higher percentage

Table 2. Accuracy of training, validation and test of the two artificial intelligence models applied to the m-FTIR-ATR spectra of SAC and goat fibers. Left: Deep neural network (DNN). Right: Support vector machine (SVM).

	DNN	SVM
Total Spectra	1220	1220
Sample	974	974
Train	728/779	778/779
Train	93.45%	99.87%
Validation	182/195	192/195
Validation	93.33%	98.46%
Test	232/246	222/246
Test	94.31%	90.24%

Abbreviations: m-FTIR-ATR = Fourier transformed infrared microspectroscopy with attenuated total reflectance, SAC = South American Camelids.
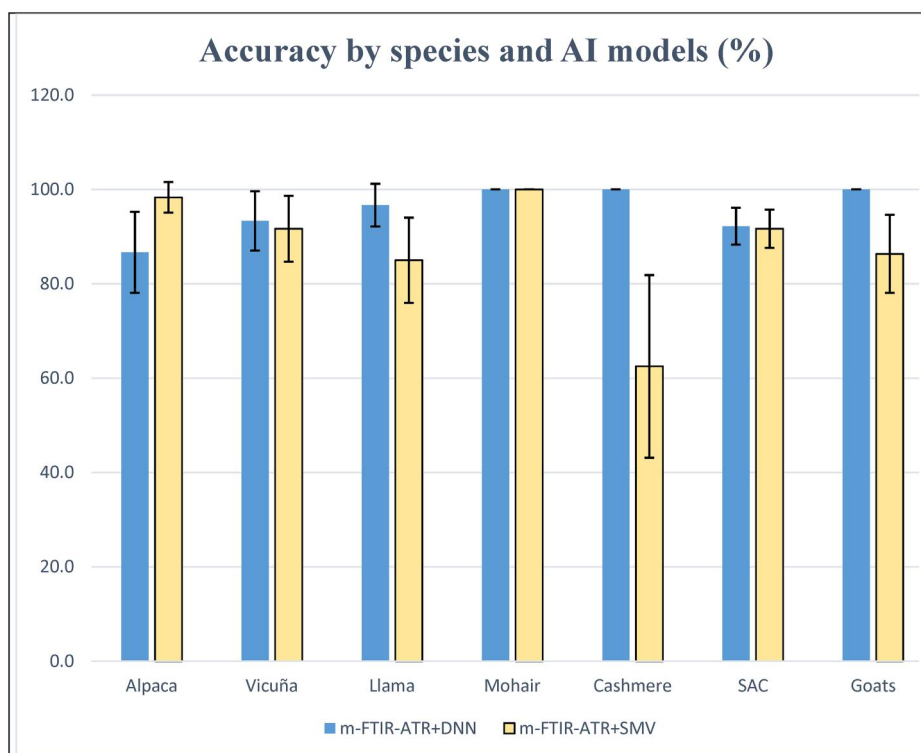
Figure 4. Accuracy (%) with confidence intervals obtained for the two artificial intelligence (AI) models: ([■]) fourier transformed infrared microspectroscopy (m-FTIR) with attenuated total reflectance (ATR), and deep neural Network (DNN); ([□]) m-FTIR-ATR+support vector machine (SVM); SAC: South American camelids.

Table 3. Percentages of hit ± error in the identification of different types of fibers using m-FTIR-ATR complemented with deep neural network (DNN) and support vector machine (SVM).

Species	DNN	SVM
Alpaca	86.7 ± 8.6	98.3 ± 3.2
Vicuña	93.3 ± 6.3	91.7 ± 7.0
Llama	96.7 ± 4.5	85.0 ± 9.0
Mohair	100.0 ± 0.0	100.0 ± 0.0
Cashmere	100.0 ± 0.0	62.5 ± 19.4
SAC	92.2 ± 3.9	91.7 ± 4.0
Goats	100.0 ± 0.0	86.4 ± 8.3
Average	**95.6 ± 3.3**	**87.9 ± 7.3**

Abbreviations: FTIR-ATR = Fourier transformed infrared microspectroscopy with attenuated total reflectance.

of hits for DNN (94.31%), closely followed by SVM (90.24%), with a difference of 4.07 percentage points.

The accuracy percentages and their confidence intervals for the identification of the 5 fiber types under study, using m-FTIR-ATR complemented with the DNN and SVM artificial intelligence models, are shown in Figure 4. Additionally, Table 3 presents the same data numerically and includes a row and a column with the averages by AI model and by species, respectively. When looking at the hits per species, the average percentage of hits (± error) indicates that the DNN model generally outperforms the SVM model,

with the only exception being alpaca fibers, where SVM achieves 98.3 ± 3.2% compared to DNN's 86.7 ± 8.6%. When results are grouped by genus, DNN shows superior performance across both evaluated genera: 92.2 ± 3.9% for SACs and 100.0 ± 0.0% for goats, compared to 91.7 ± 4.0 and 86.4 ± 8.3%, respectively, for SVM. Overall, when comparing average performance across all species, DNN achieves on average better performance (95.6 ± 3.3%), while SVM produces a close but lower percentage of hits (87.9 ± 7.3%).

Other possible analysis is by type of fiber independently. For cashmere, fibers throw excellent results when using the DNN model (100.0%). In the case of mohair, the two models (100.0%) show no error for the samples analyzed. For alpaca, the SVM model shows a percentage higher than 98%. As regards vicuña, both DNN and SVM models show a percentage higher than 91%. For llama, the DNN model shows a percentage higher than 96%.

Misclassification analysis of fibers using m-FTIR-ATR and AI models

In addition to the accuracy percentages previously presented in Table 3, Figure 5 provides a complementary and detailed visualization of the total spectra, hits
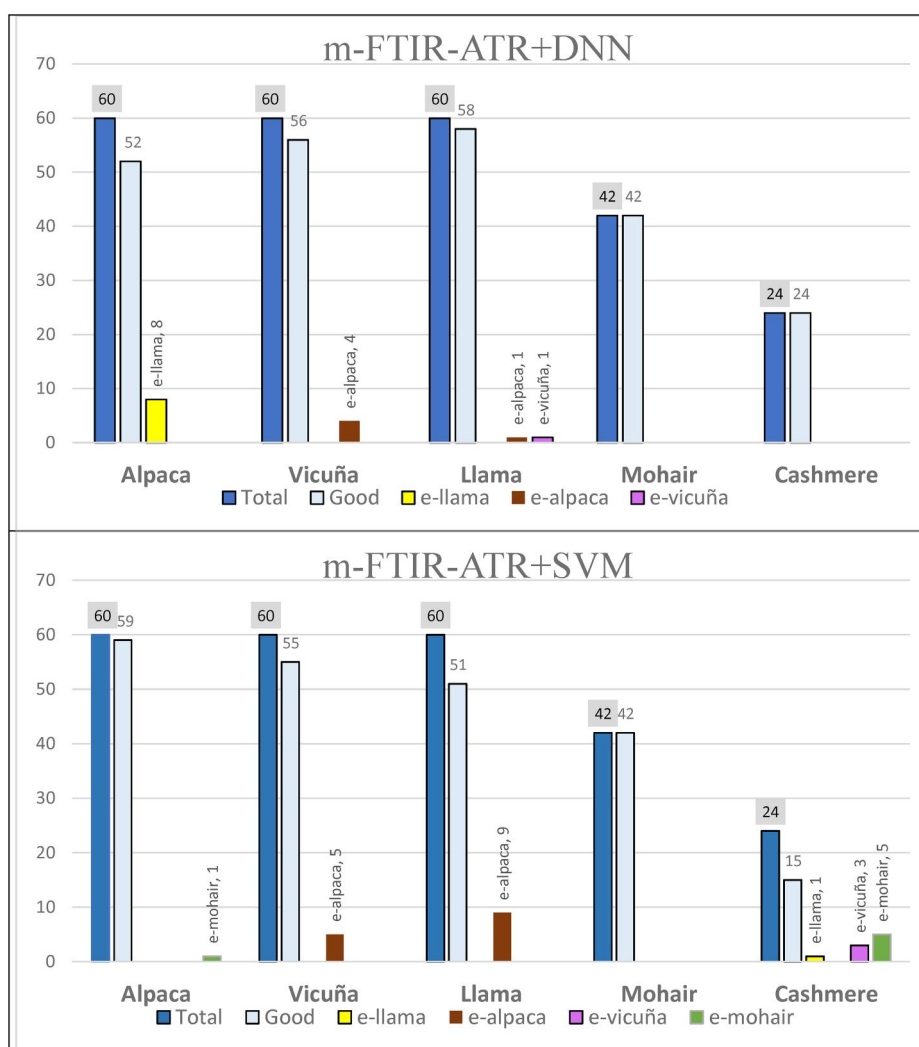
Figure 5. Total, hits and misses numbers when using deep neural Network (DNN) and support vector machine (SVM) by different fiber types with fourier Transform infrared microspectroscopy with Attenuated total Reflectance (m-FTIR-ATR). Top: m-FTIR-ATR + DNN. Bottom: m-FTIR-ATR + SVM. Prefix e-: chosen fiber by the system when misidentifying a fiber as a different species.

(correctly identified), and misclassified by fiber type when using DNN and SVM models with m-FTIR-ATR. The top panel displays results for m-FTIR-ATR + DNN, and the bottom panel for m-FTIR-ATR + SVM. Each bar represents the total number of spectra evaluated per species (blue), the correctly classified spectra (labeled "Good" in light blue), and the misclassified ones (in other colors, labeled with the prefix "e-" to indicate the incorrectly assigned species).

For instance, in the case of alpaca fibers, the DNN model correctly identified 52 out of 60 spectra, while misclassifying 8 as llama (e-llama, 8), resulting in an accuracy of 86.7%, as shown in Table 3. Vicuña spectra were mainly confused with alpaca, although the DNN still achieved 93.3% hits. Meanwhile, the SVM model performed better with alpaca (98.3%), misclassifying only one spectrum as mohair (e-mohair, 1). For llama, the DNN reached 96.7% hits, but the SVM

only reached 85.0%, showing greater confusion with vicuña and alpaca.

It is noteworthy that the DNN model achieved perfect accuracy (100%) for goat fibers, correctly identifying all mohair and cashmere samples. Additionally, despite some misclassifications among SAC species—such as alpaca fibers being misclassified as llama, vicuña fibers as alpaca, and llama fibers as alpaca and vicuña—all predictions remained within the SAC group. This suggests the model was able to discriminate at the genus level, even if individual species distinctions presented challenges.

In contrast, the SVM model struggled more with some fibers, particularly cashmere (62.5%), although it maintained strong performance for mohair (100.0%) and alpaca (98.3%). Overall, the integration of Figure 5 with the numerical data in Table 3 strengthens the analysis by providing both detailed

Table 4. Accuracy of training, validation, and testing for two additional experiments evaluated with deep neural Network (DNN) models. Left: Spectra in the 750–1750 cm^{-1} band (fingerprint region) obtained with FTIR-ATR and m-FTIR-ATR. Right: Spectra in the $661–2100 + 2500–400$ cm^{-1} region (CO_2 band removed) obtained with FTIR-ATR and m-FTIR-ATR.

	Region 750–1750 cm^{-1}		Region $661–2100 + 2500–4000$ cm^{-1}	
	FTIR-ATR	m-FTIR-ATR	FTIR-ATR	m-FTIR-ATR
Total spectra	1236	1220	1236	1220
Sample	990	974	990	974
Train	779/792	520/779	781/792	759/779
Train	98.36%	66.75%	98.61%	97.43%
Validation	185/198	124/195	189/198	181/195
Validation	93.43%	63.59%	95.45%	92.82%
Test	220/246	191/246	234/246	224/246
Test	89.43%	77.64%	95.12%	91.06%

Abbreviations: m-FTIR-ATR = Fourier transformed infrared microspectroscopy with attenuated total reflectance. FTIR-ATR = Fourier transformed infrared spectroscopy with attenuated total reflectance.

performance metrics and a clear visualization of error distribution, together offering a comprehensive view of how each model classifies the different fiber types.

However, when comparing the number of hits found using m-FTIR-ATR spectra complemented with DNN against the total number of alpaca, llama, vicuña, mohair and cashmere fibers, significant differences were found (p-value = 0.024). On the other hand, with SVM highly significant differences were found at the 99% confidence level (p-value < 0.001). These results indicate that DNN would be a better complementing AI model than SVM.

Accuracy of AI models with band reduction

Table 4 shows the accuracy of the DNN models when the training, validation, and test datasets are evaluated using spectra with the CO_2 bands removed, and then with the region including the fingerprint region, from 750 to 1750 cm^{-1}.

For the spectra where the CO_2 band was removed, the results show a higher percentage of accuracy, in descending order, for FTIR-ATR spectra (>95%), closely followed by m-FTIR-ATR spectra (>91%). In contrast, the results show a lower percentage of accuracy, in descending order, for spectra processed within the 750–1750 cm^{-1} region, with FTIR-ATR spectra (>89%) followed by m-FTIR-ATR spectra (>77%). Therefore, the best results were obtained when only the CO_2 band was removed, regardless of the spectrometry technique used. Additionally, a better accuracy was achieved, regardless of the region removed, with FTIR-ATR spectra compared to m-FTIR-ATR spectra.

Discussion

About the spectra obtained

It has been shown that there are slight differences in the structure and chemical composition of SAC and goat fibers (McGregor, 2018; Rane & Barve, 2010), which explains the variations observed in certain bands of the spectra obtained through FTIR-ATR and m-FTIR-ATR. In the fiber samples spectra obtained with FTIR-ATR, the quality was relatively lower, with low absorbance intensity and a noisy baseline. However, the absorbance bands were still detectable and represented the distinctive features of the fibers, indicating that, although the spectral resolution is lower, the technique is still able to identify the main components of the samples. In comparison, the spectra obtained with m-FTIR-ATR exhibited greater definition and clarity, which aligns with previous results reported in the literature (Peets et al., 2019).

On the other hand, m-FTIR-ATR revealed a greater number of distinctive bands between specie fibers compared to the study by Quispe et al. (2025) using the FTIR-ATR technique. However, when processing the spectra obtained with m-FTIR-ATR through the DNN model, the precision did not exceed that reported by Quispe et al. (2025) for spectra obtained with FTIR-ATR, with a difference of 2.44 percentage points. This discrepancy could be attributed to factors related to the procedure itself, such as noise caused by saturation of the IR detector, a suboptimal aperture, or excessive pressure applied to more fragile fibers, which may have caused deformation or even breakage (see Figure 1C and 1D). These issues could have resulted in scans with noise and anomalous peaks, potentially influenced by variations in feeding patterns of the species (McGregor et al., 2018), small remnants of dirt on the fibers, the occurrence of medullation (Quispe et al., 2022b), or other factors that might affect classification accuracy.

Several studies have linked FTIR wavelengths with chemical bonds in fibers (Gallagher, 2005; McGregor et al., 2018; Quispe et al., 2025). In vicuña fibers, the peaks at 1038, 1138, and 1287 cm^{-1}, which were obtained using the m-FTIR-ATR technique,

correspond to S–O, C–O, and C–H bonds, while the bands at 1419 and 1473 cm^{-1} are associated with C–N (Carbonaro & Nucara, 2010; McGregor et al., 2018). The spectral differences observed in vicuña fibers were already explained by Quispe et al. (2025), who attribute these variations to their diet and high-altitude habitat, which influence their chemical composition and amino acid levels, which are lower compared to other animal fibers such as those of goats or domestic camels (McGregor & Tucker, 2010).

Hits and misses using m-FTIR-ATR spectra with artificial intelligence models

Similar to the findings of Quispe et al. (2025), this study has shown that the combination of m-FTIR-ATR spectra complemented with artificial intelligence models (DNN and SVM) provides a more accurate classification than that achieved through visual comparison alone, as illustrated in the figures and tables. Although the highest accuracy obtained with m-FTIR (94.31%) was slightly lower than the 96.75% reported by Quispe et al. (2025) using FTIR-ATR, the results still highlight the potential of the m-FTIR technique. This technique offers advantages such as higher spatial resolution and the ability to capture detailed spectral information, which enhances fiber discrimination. However, despite the visual differences observed in the spectra of SAC and goat fibers, the identification of these fibers based solely on the spectra would be less precise and more prone to errors. Thus, the combination of m-FTIR and AI models demonstrates great potential for optimizing classification and improving accuracy in fiber analysis.

Although the DNN model proved to be more efficient and accurate compared to the SVM model, as also reported by Quispe et al. (2025) using FTIR-ATR spectra, the SVM model in this study also stood out for its faster training and lower computational demand. However, when evaluating accuracy, several studies (such as Samah et al., 2020) indicate that DNN models tend to be more robust and accurate, often outperforming SVM models, especially when large volumes of labeled data are available or when the data present classification challenges, similar to those in our study.

The results obtained with m-FTIR-ATR in this study show high accuracy in the methods used. The identification of mohair fibers using DNN and SVM, as well as cashmere fibers using SVM, demonstrated perfect accuracy (100%), given the number of samples

evaluated. In this regard, when compared to the work of Lu et al. (2019), which included wool and cashmere fibers and achieved an average accuracy of 94%, our method exhibits high recognition accuracy. Similarly, these results show superior accuracy compared to those reported by Xing et al. (2020), who attempted to identify cashmere fibers (93%). Regarding the recognition of mohair fibers, the results obtained in this study are more accurate than those reported by Quispe et al. (2025) when evaluating fiber spectra with DNN (95.2%), although the same result was obtained for alpaca fiber identification using the SVM model (98.3%).

About accuracy with band reduction

The best results obtained by evaluating the spectra with DNN, both for spectra without the CO_2 band (95.12%) and for spectra within the fingerprint region (89.43%), were achieved using spectra obtained with the FTIR-ATR technique. In contrast, lower accuracy was obtained (91.06 and 77.64%, respectively) when using spectra obtained with the m-FTIR-ATR technique. A higher accuracy (95.12%) was also achieved when evaluating FTIR-ATR spectra without the CO_2 band, compared to the spectra evaluated with m-FTIR-ATR (94.31%) in the full mid-infrared range (400–4000 cm^{-1}). These results indicate that the FTIR-ATR technique is more effective for evaluating and identifying fibers from SAC and goats compared to the m-FTIR-ATR technique. The higher accuracy observed with FTIR-ATR suggests that this methodology offers better spectral discrimination, making it a more reliable option for the analysis of these fibers.

In comparison, when evaluating the fiber samples spectra using the m-FTIR technique, processing the full spectrum (94.31%) resulted in higher accuracy than the results obtained when analyzing the spectra without the CO_2 band (91.06%) and within the fingerprint region (77.64%). This suggests that analyzing the entire spectrum allowed for the capture of additional information, contributing to better identification and differentiation of the fibers. The results indicate that using the m-FTIR technique to process the full spectrum improves accuracy in classifying SAC and goat fibers compared to an analysis limited to the CO_2 band exclusion and the fingerprint region.

On the other hand, regardless of the spectroscopic technique used, the accuracy of the results obtained when evaluating spectra without the CO_2 band (95.12 and 91.06%) was found to be better compared to the results obtained with spectra within the fingerprint

Table 5. Summary of the experiments performed using FTIR-ATR and m-FTIR-ATR techniques, with artificial intelligence models (DNN and SVM), showing the classification accuracy for SAC and goat fibers.

Spectral region	Experiment	Technique	ML algorithm	Accuracy (%)	Performed
Whole region	0	FTIR-ATR	DNN	96.75	Quispe et al. (2025)
			SVM	95.12	Quispe et al. (2025)
	1	m-FTIR-ATR	DNN	94.31	This research
			SVM	90.24	This research
CO[N/A] band removed	2	FTIR-ATR	DNN	95.12	This research
	3	m-FTIR-ATR	DNN	91.06	This research
Fingerprint region	4	FTIR-ATR	DNN	89.43	This research
	5	m-FTIR-ATR	DNN	77.64	This research

Abbreviations: m-FTIR-ATR = Fourier transformed infrared microspectroscopy with attenuated total reflectance. FTIR-ATR = Fourier transformed infrared spectroscopy with attenuated total reflectance. DNN = deep neural network. SVM = support vector machine.

region (89.43% and 77.64%). This suggests that relevant information may have been lost within the 1751 to 4000 cm^{-1} range, where subtle differences between the studied animal species could exist. The exclusion of this spectral band may have reduced the accuracy of the DNN artificial intelligence models by losing crucial data necessary for the proper identification and differentiation of SAC and goat fibers.

Finally, the best accuracy (95.12%) achieved when evaluating the DNN with FTIR-ATR spectra without the CO_2 band is equal to the results obtained by Quispe et al. (2025) when evaluating FTIR-ATR spectra with their SVM model. It is also very close, though not better, to the result obtained when evaluating FTIR-ATR spectra with the DNN model (96.75%), with only a difference of 1.63 percentage points. This relatively small difference may be attributed to variations in preprocessing parameters or differences in the datasets used, such as the number or quality of the fiber samples. These results highlight the ability of DNN to handle and extract relevant features from the spectra, even when specific components such as the CO_2 band are removed. The accuracy achieved with DNN in this study underscores the effectiveness of this methodology for making precise and reliable classifications in fiber analysis, particularly when noise elements that could interfere with the process are eliminated. These findings reinforce the potential of DNN as a powerful tool in spectral analysis applications, suggesting that, despite the slight difference with previous studies, the use of DNN remains a highly promising option for improving accuracy in fiber identification.

To provide a clearer overview of the classification accuracies obtained in each experiment, Table 5 summarizes the results. This table includes the spectral region, the type of FTIR technique used (FTIR-ATR or m-FTIR-ATR), the artificial intelligence model applied (DNN or SVM), and the corresponding accuracy percentage achieved in each case. The purpose of this summary is to facilitate comparison between experiments and better visualize how spectral processing choices and AI algorithms impact classification performance.

Conclusions

In this work, we explored different options for analyzing the spectra, including the choice of spectral region (whole spectrum, fingerprint region, or whole spectrum without the CO_2 band), the type of equipment (FTIR spectrometer or FTIR microscope), and the machine learning model (DNN or SVM). Based on these variables, five distinct experimental configurations were tested, building upon previous research to evaluate the most effective combination for identifying SAC and goat fibers.

The fiber samples spectra obtained through FTIR-ATR, from which the CO_2 band was subsequently removed, exhibit better accuracy when evaluated with DNN models, compared to the m-FTIR-ATR spectra evaluated with both DNN and SVM models, regardless of whether part of the mid-infrared band was removed. This difference could provide an advantage in terms of operational efficiency, simplification of the data acquisition process, and cost reduction, as FTIR-ATR spectra are easier and faster to obtain than spectra acquired with m-FTIR-ATR. Furthermore, its implementation could facilitate future evaluations by being quicker to process, enabling greater automation in analyses and enhancing the feasibility of its use in high-volume applications.

The DNN model is the most suitable for evaluating both SAC and goat fibers. However, the combination of the m-FTIR-ATR technique with the DNN model proves to be the most effective option for goat fiber identification. For alpaca fibers, the combination of m-FTIR-ATR with the SVM model emerges as the most efficient alternative. These results conclude that FTIR-ATR spectroscopy techniques, combined with machine learning models, provide a reliable method for identifying SAC and goats through the spectral

analysis of their fibers. Implementing these models and techniques could optimize processes in the textile industry (such as classification, quality control, and traceability), enabling precise identification of fibers based on their origin. This, in turn, could positively impact the sustainability and value of the studied species.

Regardless of the spectrometry technique used, spectra without the CO_2 band, when combined with DNN models, show higher accuracy than spectra evaluated within the fingerprint region (750–1750 cm^{-1}). This suggests that the spectral range from 1751 to 4000 cm^{-1}, which was excluded to focus solely on the fingerprint region, contains crucial information for differentiating between the SAC and caprine species studied. Its omission therefore reduces the effectiveness of the artificial intelligence models.

As a general conclusion, the two underlying hypotheses formulated in the Introduction section have been tested, showing mixed results. As depicted in the manuscript, in some cases, pre-processing the signals and/or using different instrumentation has produced better outcomes compared with previous research, depending on the species and/or the AI model. Nevertheless, the identification percentages have remained at comparable levels in any case, which upholds the soundness of the proposed approach.

Future work should focus on expanding the spectral database with a larger and more diverse set of fiber samples, including dyed, blended, or treated fibers, to assess the robustness of the models. Additionally, exploring other machine learning and deep learning architectures, such as convolutional neural networks (CNNs) or transformers, could further improve classification performance.

Acknowledgements

Disclosure Statement

Funding

ORCID

Max Quispe http://orcid.org/0000-0003-0884-8789

References

Barraza-Garza, G., Martinez-Martinez, A., De la Rosa, L. A., & Castillo-Michel, H. (2013). Microespectroscopía de infrarrojo con transformada de Fourier (FTIRM) en el estudio de sistemas biológicos. *Revista. Latinoamerericana de Química*, *41/3*(2013), 125–148. ISSN 0370–5943

Carbonaro, M., & Nucara, A. (2010). Secondary structure of food proteins by Fourier transform spectroscopy in the mid-infrared region. *Amino Acids*, *38*(3), 679–690. https://doi.org/10.1007/s00726-009-0274-3

Gallagher, W. (2005). FTIR analysis of protein structure. Chem455, Biochemistry Lab II—Synthesis and Characterisation of Amyloid Fibril-Forming Peptides Manuals. Retrieved from www.chem.uwec.edu

Goel, R., Sharma, A., & Kapoor, R. (2019). Object recognition using deep learning. *Journal of Computational and Theoretical Nanoscience*, *16*(9), 4044–4052. https://doi.org/10.1166/jctn.2019.8291

Hsu, S. (1997). Infrared spectroscopy. In *Handbook of Instrumental Techniques for Analytical Chemistry*. (pp. 248–283). Prentice hall PTR.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kuc, B. (2018). The concept of a spectrophotometer for soil sample testing. In *2018 International Conference BIOMDLORE* (pp. 1–4). https://doi.org/10.1109/BIOMDLORE.2018.8467212

Lu, K., Luo, J., Zhong, Y., & Chai, X. (2019). Identification of wool and cashmere SEM images based on SURF features. *Journal of Engineered Fibers and Fabrics*, *14*, 1–9. https://doi.org/10.1177/1558925019866121

McGregor, B. A. (2018). Physical, chemical, and tensile properties of cashmere, mohair, alpaca, and other rare animal fibers. In *Handbook of Properties of Textile and Technical Fibres*. (pp. 105–136). Elsevier. https://doi.org/10.1016/B978-0-08-101272-7.00004-3

McGregor, B. A., Liu, X., & Wang, X. G. (2018). Comparisons of the Fourier Transform Infrared Spectra of cashmere, guard hair, wool and other animal fibres. *The Journal of The Textile Institute*, *109*(6), 813–822. https://doi.org/10.1080/00405000.2017.1372057

McGregor, B. A., Ramos, H., & Quispe, E. (2012). Variation of fibre characteristics among 562 samples sites for Huacaya alpaca fleeces from the High Andes. *Small Ruminant Research*, *102*(2-3), 191–196. https://doi.org/10.1016/j.smallrumres.2011.07.016

McGregor, B. A., & Tucker, D. J. (2010). Effects of nutrition and origin on the amino acid, grease, and suint composition and color of cashmere and guard hairs. *Journal of Applied Polymer Science*, *117*(1), 409–420. https://doi.org/10.1002/app.31651

Nayak, R., Houshyar, S., Khandual, A., Padhye, R., & Fergusson, S. (2020). Identification of natural textile fibres. In *Handbook of Natural Fibres*. (pp. 503–534). Elsevier. https://doi.org/10.1016/B978-0-12-818398-4.00016-5

Peets, P., Kaupmees, K., Vahur, S., & Leito, I. (2019). Reflectance FT-IR spectroscopy as a viable option for

textile fiber identification. *Heritage Science*, *7*(1), 93. https://doi.org/10.1186/s40494-019-0337-z

Quispe, E., Quispe, M., Quispe, C., Poma, A., Paucar-Chanca, R., Cruz, A., & McGregor, B. A. (2022b). Relationships between the incidence and degree of medullation with the diameter of alpaca fibers evaluated using a novel device based on artificial intelligence. *The Journal of The Textile Institute*, *114*(7), 1016–1031. https://doi.org/10.1080/00405000.2022.2105110

Quispe, M. D., Quispe, C. C., Serrano-Arriezu, L., Trigo, J. D., Bengoechea, J. J., & Quispe, E. C. (2023). Development and validation of a smart system for medullation and diameter assessment of alpaca, llama and mohair fibers. *Animal: An International Journal of Animal Bioscience*, *17*(5), 100800. https://doi.org/10.1016/j.animal.2023.100800

Quispe, E. C., Sánchez, F., Filella, J. B., & Alfonso, L. (2014). Variation of commercially important characteristics among sampling sites for Vicuña (Vicugna Vicugna Mensalis) fleeces. *Journal of Camelid Science 2014*, 7, 1–14

Quispe, M. D., Serrano-Arriezu, L., Trigo, J., Quispe, C., Poma, A., & Quispe, E. (2022a). Application of artificial intelligence and digital images analysis to automatically determine the percentage of fiber medullation in alpaca fleece samples. *Small Ruminant Research*, *213*, 106724. https://doi.org/10.1016/j.smallrumres.2022.106724

Quispe, M. D., Trigo, J., Serrano-Arriezu, L., Huere, J., Quispe, E., & Beruete, M. (2025). Classification of South American Camelid and goat fiber samples based on fourier transform infrared spectroscopy and machine learning. *The Journal of The Textile Institute*, *116*(2), 198–207. https://doi.org/10.1080/00405000.2024.2324209

Rane, S., & Barve, S. (2010). Evaluating protein patterns of speciality fibres for identification to combat false labeling. *International Journal of Zoology Research*, *6*(4), 286–292. https://doi.org/10.3923/ijzr.2010.286.292

Samah, A., Nasien, D., Hashmin, H., Sahar, J., Majid, H. A., Yusoff, Y., & Shah, Z. A. (2020). Application of deep learning method in facilitating the detection of breast cancer. In *IOP Conference Series: Materials Science and Engineering*. https://doi.org/10.1088/1757-899X/864/1/012079

Shim, S. (2003). *Analytical techniques for differentiating Huacaya and Suri alpaca fibers* [The Ohio State Univerity]. https://etd.ohiolink.edu/acprod/odb_etd/ws/send_file/send?accession=osu1064202091&disposition=inline (Accessed on April 7, 2025)

Tan, C., Chen, H., Lin, Z., & Wu, T. (2019). Category identification of textile fibers based on near-infrared spectroscopy combined with data description algorithms. *Vibrational Spectroscopy*, *100*, 71–78. https://doi.org/10.1016/j.vibspec.2018.11.004

Wu, G., & He, Y. (2008). Identification of fine wool and cashmere by Vis/NIR spectroscopy technology. *Spectroscopy and Spectral Analysis*, *28*, 1260–1263.

Xing, W., Deng, N., & Yu, C. (2019). Identification of wool and cashmere based on multi-feature fusion image analysis technology. *Journal of Textile Research*, *40*(3), 146–152.

Xing, W., Liu, Y., Deng, N., Xin, B., Wang, W., & Chen, Y. (2020). Automatic identification of cashmere and wool fibers based on the morphological features analysis. *Micron (Oxford, England: 1993)*, *128*, 102768. https://doi.org/10.1016/j.micron.2019.102768