Journal of the Brazilian Society of Mechanical Sciences and Engineering (2022) 44:338
https://doi.org/10.1007/s40430-022-03645-1

TECHNICAL PAPER

Online near-infrared spectroscopy for automatic polymeric material identification

Daniel Aparecido Lopes Vieira da Cunha[1] · Rafael Luis Ribessi[2] · André Carmona Hernandes[3] · Ivo Milton Raimundo Jr.[2] · Rafael Vidal Aroca[4] · Marcia Cristina Branciforti[1]

Received: 28 October 2021 / Accepted: 21 June 2022 / Published online: 14 July 2022

Abstract

Additive manufacturing (AM) offers several advantages for transforming productive chains, such as generation of complex geometries on demand. However, challenges must be overcome toward increasing manufacturing precision and quality of the parts produced, reducing production time, and standardizing the processing parameters. One example of common failure is the incorrect processing parameter selection for the filament-based AM, which can damage 3D printing machines, such as hotend clogging. In that way, this paper introduces a method that identifies AM's polymeric materials through in situ near-infrared (NIR) spectroscopy and classifies them into poly(lactic acid), acrylonitrile butadiene styrene, and poly(ethylene glycol terephthalate), also enabling a manual parameter input. A low-cost NIR spectrophotometer was used to analyze 16 filaments with color and manufacturer variability. Each filament was probed 3 times in 3 distinct places, raising 144 spectra. Chemometrics were applied to identify relevant peaks for functional groups, and a linear regression was used to filter out data that showed no such peaks. In a second stage, a second-derivative Savitzky–Golay was used to aid in class separation, and a principal component analysis was performed to reduce data dimensionality. The resulting projections were classified by an LDA algorithm, and 3 study cases conducted with data augmentation tested the classifier. The results show the proposed method is robust to bias variation and can handle blends of up to 70%–30% mix and correctly separate signals with and without peaks. Such responses have proved the feasibility of the classification system, especially when fed with a highly varied data set.

Keywords Additive manufacturing · Machine learning · NIR spectroscopy · Chemometrics · FDM · Automation

1 Introduction

The development of recent digital and electronic technologies has led to a series of production strategies in global industries that have adopted advanced manufacturing techniques based on work automation, robotics, Internet of Things (IoT), and data intelligence. Industries that apply such technologies, coordinated to confer competitiveness and optimization of the productive chain, are inserted in the era of industries 4.0 [7, 13, 21]. Among the modernities provided by this new industrial phase, additive manufacturing (AM), also called 3D printing, is defined as a group of processes that produces objects in three dimensions from a digital model [11, 16, 28]. In the scope of Industry 4.0, the need for customization of manufactured goods has made AM essential, given its great ease of adaptation and customization. Also, as the disadvantages of the process are overcome, AM is transformed, especially by the inclusion of electronic

Technical Editor: Lincoln Cardoso Brandao.

✉ André Carmona Hernandes
andre.hernandes@ufscar.br

[1] Department of Materials Engineering, São Carlos School of Engineering, University of São Paulo, Avenida Trabalhador São-Carlense, 400, São Carlos 13566-590, Brazil

[2] Institute of Chemistry, University of Campinas, Cidade Universitária Zeferino Vaz, Campinas 13083-970, Brazil

[3] Electrical Engineering Department, Federal University of São Carlos, Rodovia Washington Luiz, km 235, São Carlos 13565-905, Brazil

[4] Computer Science Department, Federal University of São Carlos, Rodovia Washington Luiz, km 235, São Carlos 13565-905, Brazil

components, such as electrical circuit boards and sensors [9, 10, 27].

Fused Deposition Modeling (FDM) is an AM technique widely used in the market and recognized for its manufacture of high-quality products in shorter processing time [9, 27]. In this process, the polymeric filament is introduced in a heated nozzle through the liquefier until the corresponding melting temperature. The molten material is deposited by the print head on the manufacturing platform according to the guidelines in the XY plane of the digital model. The movement in the Z plane is controlled by either the platform, or the print head itself. Despite this simple functionality, FDM generates complex geometries precisely and safely and has been widely used not only in production, but also in offices and educational centers [9, 32, 35].

The increased popularity of the low-cost FDM equipment, called desktops, has led to the emergence of different polymeric filaments, among which compounds made of acrylonitrile–butadiene–styrene (ABS), poly(lactic acid) (PLA), and poly(ethylene glycol terephthalate) (PETG) are the most widespread on the market [31], [35]. However, the use of FDM equipment in production environments is still limited, mainly by the difficult standardization of best processing configurations that are dependent on those materials [17, 20, 24]. Therefore, several studies have explored the relationship between the rheological properties of materials and the optimization of processing parameters [8, 9, 25], exposing the lack of norms and standards for the reproducibility of the process and consequent difficulty in adapting the system to mass production. The literature also reports the importance of a proper adjustment of processing parameters toward a better quality of the objects produced [20, 24]. In contrast, according to a review conducted by Vyavahare et al. [35], only 12% of specific literature has dealt with recent advances in the FDM technique, such as monitoring, control, and automation of the process.

Automation corresponds to the computerized use of machines and devices in the manufacturing process for the development of different tasks with minimum human effort [8, 19], therefore, robotics, hardware, and assistive devices, as well as the fundamentals of artificial intelligence and machine learning (ML), have been widely used and applied [14, 33]. Haverkort and Zimmermann [15] claimed that when learning from production data, machine learning algorithms are able to improve both efficiency and safety of a process, making it more dynamic and intelligent. Regarding AM processes, according to a review conducted by Goh et al. [14], ML concepts have been applied as object design, material adjustment, process optimization, cloud services, and cybersecurity.

In the context of this work, ML enables the creation of a classifier that previously identifies the classes of materials and automatically configures the processing conditions and automates the AM process. The system must collect the feature data of each material, differentiate their peculiarities, and adjust the respective best processing parameters, whereas the sensing device must necessarily be able to read such peculiarities and communicate directly with the classifier. As a sensing device, the alternative is the use of near-infrared (NIR) spectroscopy (NIRS), a type of spectroscopy capable of identifying compounds by employing vibrational frequencies in the wavelength range from 780 to 2500 nm [29]. When aligned with analytical chemistry procedures (specifically Chemometrics), NIR provides a particular fingerprint of the materials [23, 30]. In addition to this advantage, in the last decade, its technology has experienced an extensive transformation with the emergence of extremely portable, miniaturized, accessible, and connected devices. Some near-infrared spectrometers are already considered smart objects, since they can interconnect the physical world with online systems using cloud computing resources [2, 3, 23]. The next step in the evolution of the idealized system is its integration with other emerging technologies associated with AM through, for example, its adaptation to monitoring and control systems, such as those proposed by Goh et al. [14] and [22].

This manuscript presents a classifier based on NIR spectroscopy and chemometrics that automatically identifies the material used in AM and guides the settings of its processing parameters. The objective of the system is to reduce manufacturing failures resulting from an improper selection of processing parameters for each material, as well as to increase the accessibility of the technology to the public by decreasing the dependence of the process on the user.

2 System design

The conceptual project proposes the association of sensing devices and 3D printers toward a system that automatically identifies the raw material and, subsequently, adjusts the processing parameters. The selection of the sensing method is directly dependent on the way the materials are introduced in the process. Regarding FDM equipment, in which the material is introduced tractioning the polymeric filament in the extruder, the use of portable near-infrared spectrophotometers for capturing material information is an excellent alternative, due to their technological evolution and miniaturization of spectroscopic devices [23].

Figure 1 illustrates the system design, which is the polymeric filament (1) organized by a spool (2) and inserted into an extrusion head (3) that manufactures the part (4). Attached to the manufacturing environment is a portable near-infrared spectrophotometer (5) that, aided by a reference bulkhead (6), receives and sends spectral information to an acquisition board (7). An effective functioning of the

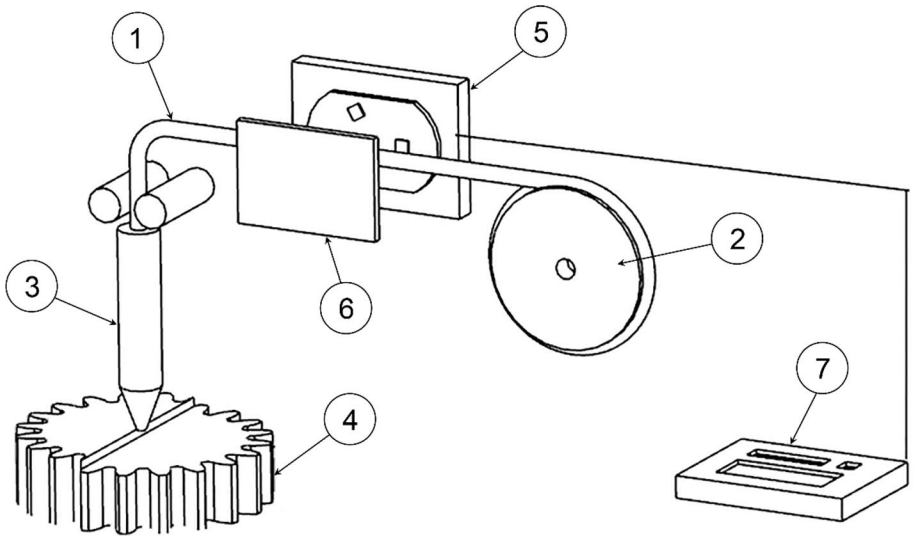

Fig. 1 Material identification system applied to the FDM process. (1) Polymeric filament; (2) Spool; (3) Extrusion head; (4) Part; (5) Near-infrared spectrophotometer; (6) Reference bulkhead; and (7) Acquisition board. Adapted from [1]

system requires the filament to have direct contact with the capture windows while a spectrum is obtained. Information of the raw material is sent to the acquisition board, which operates in conjunction with the classifier developed in this study. The system checks the identity of the material according to the spectral characteristics specific to each class. In the case of identification of a material, the inputs of the processing parameters specific to the material are given. The configuration information can be stored in a cloud-connected database. In this way, the user can enter and share the unknown settings of the equipment. The manufacturing process is then started, and an informative report of the process is sent to the cloud to be available to the user. On the other hand, if the material is not identified, the system guides the operator to manually configuring the processing parameters. The entered settings, as well as an informative report of the selected process conditions, are uploaded and stored in the cloud database, and then, manufacturing begins.

At this point, it is important to note that the methodology and the results presented in this manuscript describe the development of the first version of the classifier, which was able to identify the materials studied by aligning machine learning and chemometrics techniques.

3 Development of the algorithm

The research roadmap is described below and illustrated in Fig. 2.

This study examines in situ polymeric filaments intended for the FDM technique and the hypothesis that portable NIR spectroscopy sensors can provide information for the discrimination of material features, enabling the machine learning classification process. Following the guidelines of [26], the supervised machine learning employed in this study is divided into four stages, namely (1) data collection; (2) data preparation (exploratory data analysis and feature engineering); (3) training of the classification model; and (4) testing. In the first, spectral data of PLA, ABS, and PETG polymer filaments are collected and analyzed. Next, the exploratory analysis procedures start with the pre-treatment of data by second-derivative Savitzky–Golay (SG) algorithm and end with Principal Component Analysis (PCA), reaching a decision to use the first 3 components as features to a linear classifier. The third step was separated as just LDA classifier, to serve as baseline for the classification, then enhancing it by performing a pre-classification step to remove the confusing

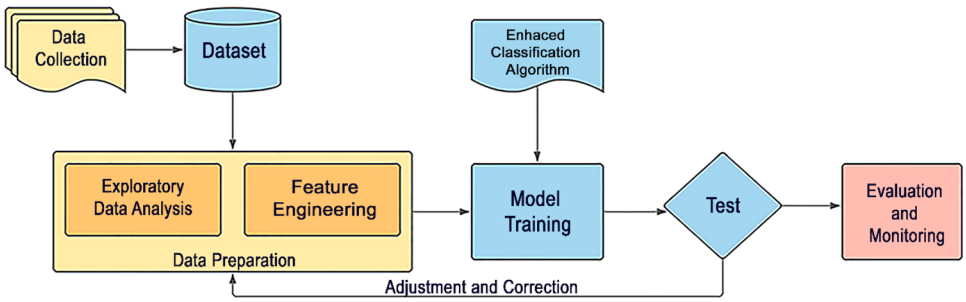

Fig. 2 Methodology roadmap

group in an earlier stage. Finally, the last stage was done by performing 3 study cases with data augmentation.

3.1 Data collection

Initially, 19 commercial polymeric filaments of PLA, ABS, and PETG of 1.75 ± 0.05 mm diameter and different colorations were selected from four different manufacturers. More information about these filament samples is viewed in Table 1 of the supplementary material. A model DPL NIRscan Nano Texas Instrument portable spectrophotometer acquired the spectral data; it performs absorbance measurements in at most seven seconds and operates in the 900 to 1700 nm range. Measurements are taken in triplicate for three different positions of each polymeric filament, totaling $3 \times 3 \times 19 = 171$ sample spectra with 228 absorbance values. Polytetrafluoroethylene (PTFE) was used as a support and calibration material for the measurement spectroscopy data of the raw materials, since it does not interact significantly with NIR radiation, reflecting most of the radiation back to the detection system; therefore, it is ideal for such an application as a reference material. Figure 3 illustrates the collection spectral data from polymeric filaments by a portable spectrophotometer.

Figure 4 shows two examples of PLA spectra obtained from the second reading of the blue and black colored filaments from Manufacturer 4, using the DPL NIRscan Nano Texas spectrophotometer.

The first analysis evaluated the appearance of the spectral data by promoting a band assignment study toward understanding of the main particularities of the samples. Band assignment allows not only understanding relevant information about the identity of the analyzed material but also evaluating the bond behavior of the chemical groups belonging to it. Note that not all spectral data generated provide band information relevant to classification, such as the spectrum of the black PLA filament shown in Fig. 4. According to Becker, Sachsenheimer and Klemenz [4], the lack of band information in spectra of darkly colored polymers is explained by the fact that these materials are composed of pigments/colorants (e.g., soot and carbon

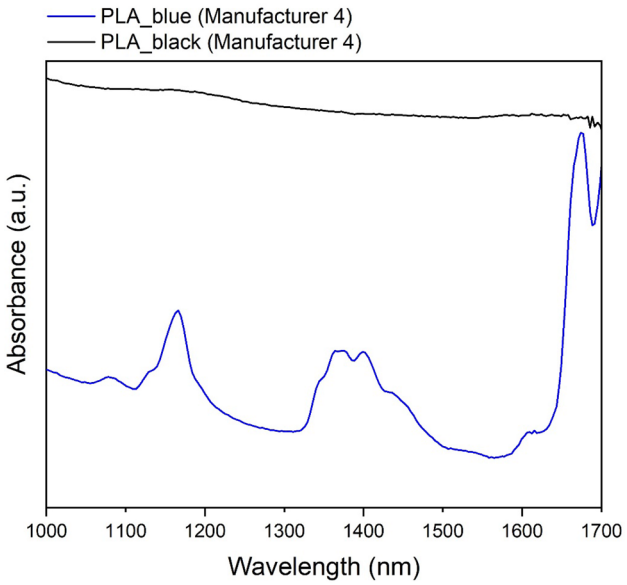

Fig. 4 Spectra of samples that show (blue-colored PLA filament) and do not show (black colored PLA filament) absorption of near-infrared radiation

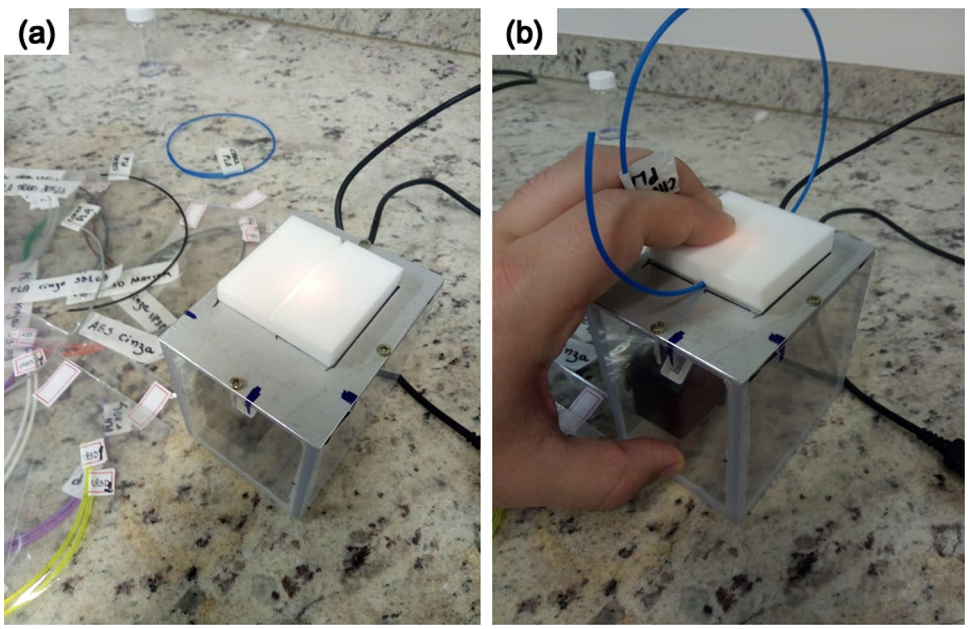

Fig. 3 Data Collection: **a** reference measurement; **b** polymer filament measurement

black), which absorb all light in the NIR spectral region. Therefore, spectral data without relevant band information, such as gray and black filaments, were removed from the dataset in the first step.

Figure 5 shows the main absorption bands in the 1000 to 1700 nm infrared region of NIR spectra of PLA, ABS, and PETG filament samples, respectively. According to Burns and Ciurczak [6], the absorption bands located in the NIR region for polymers (800 to 2500 nm) contain overtones and combinations related to stretches of C–H, N–H, and O–H bonds. For PLA (Fig. 5a), bands associated with vibrations are observed in the region of the second C–H stretching overtone (1100 nm to 1200 nm), specifically at wavelengths between 1132 and 1166 nm. In the sequence, there are bands assigned to asymmetric and symmetric stretching vibrations, especially of methyl groups (CH_3) present in the first overtone of the C–H combination (1300 to 1500 nm and 1600 to 1700 nm). The ABS spectra (Fig. 5b) contain two representative bands in the 1139 to 1196 nm range, inserted in the second stretching overtone of the aromatic and methyl C–H group, respectively. A prominent absorption peak is observed in the 1406 nm wavelength region and related to the first overtone of the CH_2 and C–H combination band, and two peaks are in the range of the first C–H stretching overtone (1600 to 1700 nm). Burns and Ciurczak (2008) claimed this sequence of absorption bands is typical of styrene-derived polymers. Similarly, the vibrational spectra of PETG (Fig. 5c) show two absorption bands in the region of the second C–H stretching overtone (1100 and 1200 nm), with peaks mainly at wavelengths near 1121 and 1189 nm and related to the aromatic C–H group. PETG is a modified version of PET, which has a high hydroxyl content in the near-infrared region [5]. The results show the distinguishing factor between the data to be classified is based on the difference in the functional groups present in the chemical structure of the polymeric filaments. The peaks identified by assigning absorption bands are analogous and stable for most of the samples, except for the data from the black-colored PLA filaments and the gray-colored PLA and ABS.

Another peculiarity observed in the spectral data shown in Fig. 5 is the clear variation in the baselines of samples of the same material, but different colorations/manufacturers. The literature reports that these variations are caused by the equipment itself, for example, due to temperature variations in the radiation source, and non-uniformity of the samples, differences in granulometry and changes of pigments and dyes in the material identified this specificity by performing a study related to NIR spectroscopy applied to samples with PLA and PET aggregation [5, 12, 18]. They have shown that as PLA concentrations increase, there is a change in the absorption of infrared radiation causing the change in the baseline. The variation in the baseline is a determining factor for the predictive ability of the classifier. Thus, it is implicit the importance of applying derivative methods in the spectral data treatment stage to correct both baseline shift and slope. Also, such methods can increase peak detection sensibility, which, as described, are relevant to band assignment.

3.2 Data preparation

Data preparation comprises two correlated processes, namely exploratory analysis and feature engineering. The former applies statistics and visualization methods to better understand the data identifying features, trends, qualities, and possible classifications factors. After the best features are identified and selected, the classification model can be created by feature engineering. In this study, the exploratory analysis performed used Unscrambler X version 10.4.

Initially, the spectral data are imported into the software in a matrix format with 171 objects (samples) and 228 variables (wavelength), and the treatment begins with a second derivative by Savitzky–Golay (SG) algorithm. Figure 6 shows the result of SG application to the spectral data, calculating the second derivative in a second-order polynomial

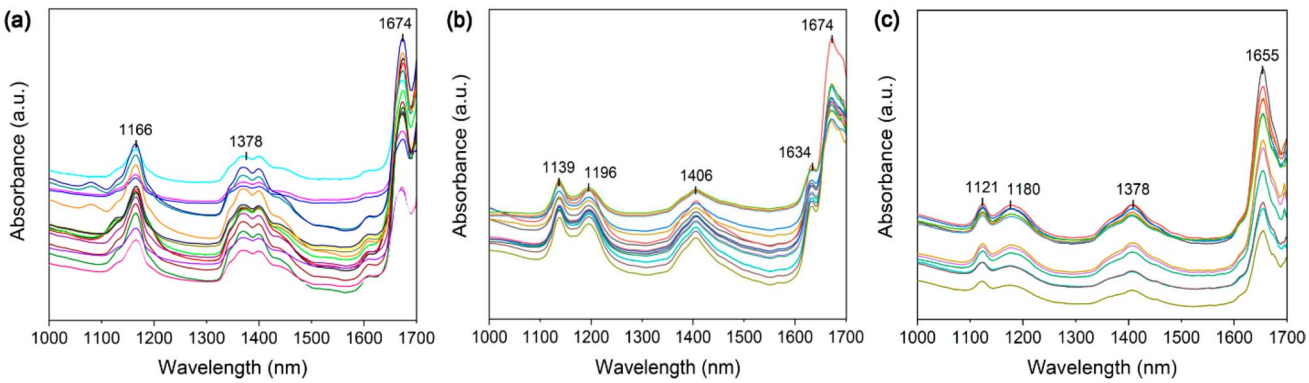

Fig. 5 Example of identifiable near-infrared (NIR) absorption spectra of polymeric filaments for FDM processes: **a** PLA; **b** ABS; and **c**; PETG

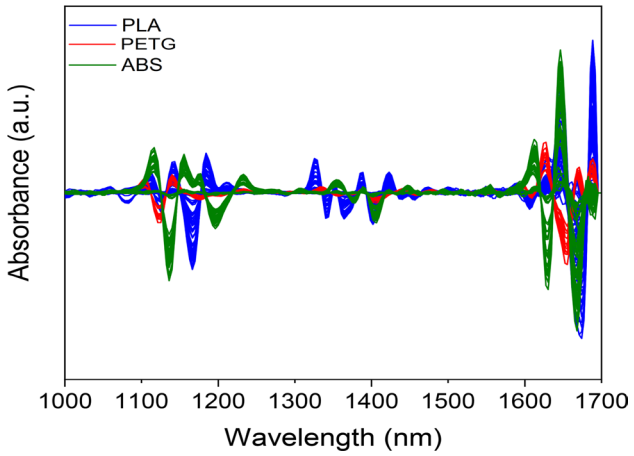

Fig. 6 Savitzky-Golay digital filtering treatment of NIR spectra of PLA, ABS, and PETG applying the second derivative in a 2nd order polynomial and seven-point window

with a seven-point window. The criterion for the choice of parameters was based on the need for corrections in the baseline of the spectra, reduction in the light scattering effects, and amplification of the characteristic peaks of each material (Fig. 5) toward increasing the classification capability of the algorithm. According to Fig. 6, the application of the second derivative maximized the information on the curve's concavity, facilitating the location of the data absorbance peaks. Therefore, the treatment identified the wavelength regions of location of each material's absorbance bands, reducing the number of overlapping peaks and increasing the ability to distinguish between classes.

The importance of SG application for spectral data classification performance is proved by the PCA projection of the untreated data (Fig. 7). In this project, PCA was designed by the Singular Value Decomposition (SVD) method, which is an algorithm that distinguishes dimensions with highest variance data [34]. PCA's objective is to reduce the dimensionality of the original dataset space without affecting the properties of the samples. Therefore, correlated samples are grouped in a specific region of the score plot to be more evident for analyses by a classification model. Figure 7 shows the projection of the raw data generated a nonlinear distribution of the sample with no homogeneous clustering of the spectra of the same material.

The effect of SG application to the spectral data is best seen in Fig. 8, which illustrates the dispersions of pretreated PLA, ABS, and PETG spectra on the PCA projection score plot.

Along PC-1, a distributed group of PLA samples is formed in the positive quadrants of the axis (I and IV), and ABS and PETG groups are formed in the region of negative values (II and III). The distinction between the scatters of the latter samples is clear along PC-2, where groups of PETG and ABS are located in the positive (II) and negative (III) quadrants of the axis, respectively. The results show the parameters selected for the SG treatment as well as the projection of the data by PCA were sufficient for grouping similar samples and distancing dissimilar ones, in such a way they are linearly separable via PC1 and PC2 features.

Nevertheless, a group of samples from different classes (PLA and ABS) is observed in quadrant II—they are spectra with no visible absorption bands in the 1000 to 1700 nm range. As will be seen below, this result will be critical for the predictive ability of the classification model, since the selection of the correct frontier would be a difficult task for a linear decision algorithm.

For sake of completeness, Fig. 9 shows a normalized comparison between the curves of the inverse PC-2 loading plot and the characteristic spectra of the samples that varied the most along this axis, i.e., those related to ABS.

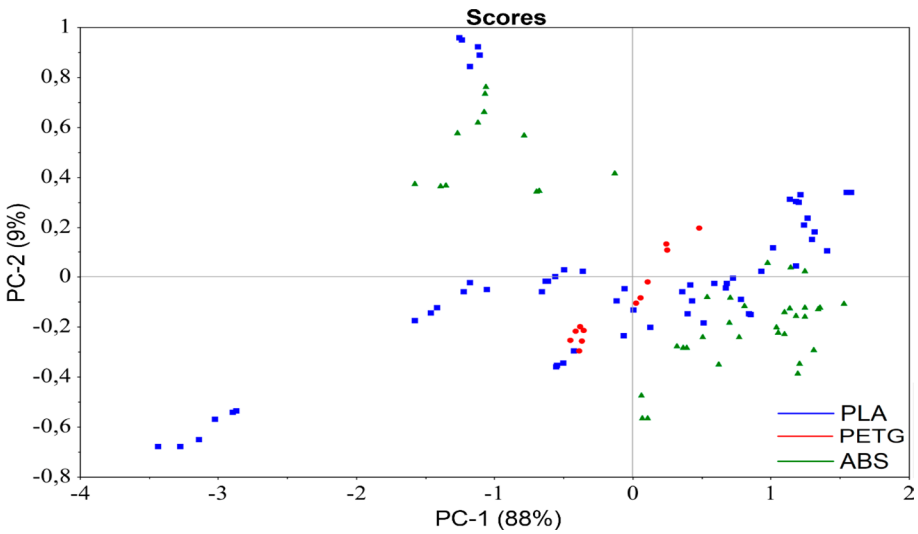

Fig. 7 PCA score plot projected to the data set with no pretreatment. The projection shows an inhomogeneous scatter (PC-1 explains 88% of total variance while PC-2 explains 9%)

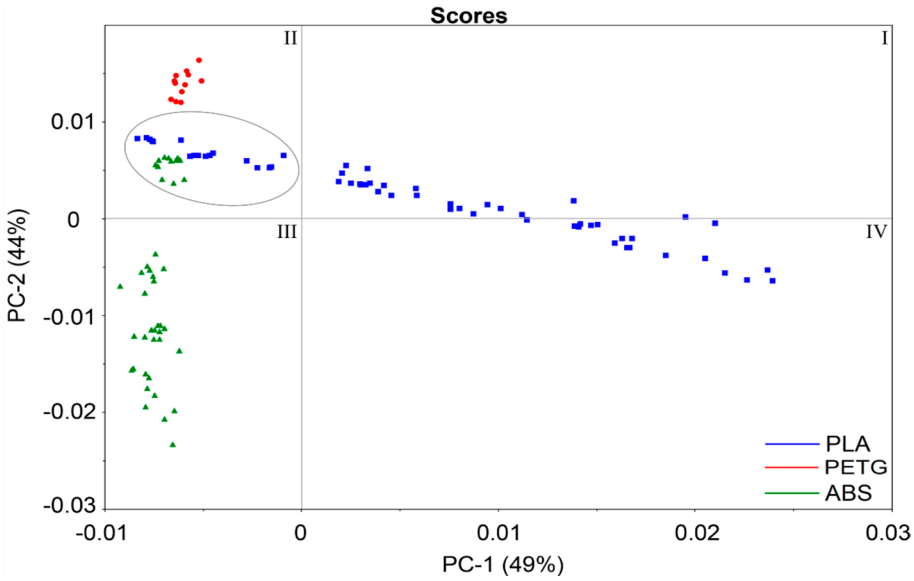

Fig. 8 PCA score plot projected for the data set with pretreatment. The data treatment defined the clustering of the samples. As indicated, distinct class groupings will be adjusted in the algorithm steps

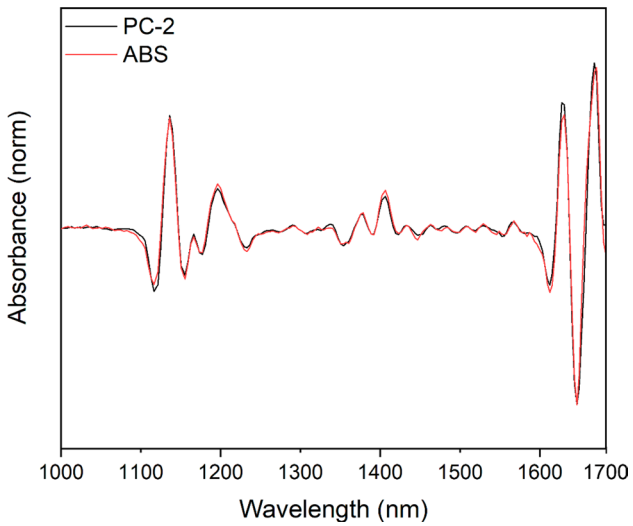

Fig. 9 Comparison between the loading plot that contributed to PC-2 formation and the typical spectrum of the ABS sample. The loading identified characteristic ABS curvatures over the entire wavelength range, facilitating the well-defined dispersions of the samples in PCA

The wavelength where the loading peak occurs indicates the most relevant regions for the formation of the rescaled axis of the PCA. As can be seen, the characteristic wavelength intervals of PC-2 formation are analogous to the absorbance bands of ABS throughout the analysis range. This corroborates the fact that the SG has enhanced the curvatures of the peaks to facilitate the projection of the PCA according to the distribution trend of the absorbance bands relative to each material, hence acting as a peak template, without the need to artificially create such structure.

3.3 Classification step

The whole dataset was used in an LDA classifier for the creation of a baseline and assessment of the extension to which the group in Fig. 8 might cause a misclassification. Initially, the spectral data set of 171 samples was divided into two subsets, namely calibration (70%) and test (30%). The randomness of the data set partition was guaranteed by Kennard–Stone algorithm, which selects samples that accurately represent the variability in the evaluated spectra. The calibration subset was separated for training the LDA classification model, while the test one validated the classifier. As a result, the created model performed the training with 95% accuracy, verified by only four inaccurate specifications. In the validation process, the classifier prediction accuracy was 79.63%, verified by eight incorrect classifications. All incorrect classifications refer to ABS samples with reduced absorbance signals being categorized as PLA, exactly the ones identified in Fig. 8. The precision, recall, and macro-f1 metrics of the classifier validation resulted in 59.62%, 76.73%, and 0.67, respectively.

Despite LDA's high classification performance for spectral data, the model was unable to classify PLA and ABS samples grouped in the same region of the PCA score plot (Fig. 8, quadrant II). Specifically, these are samples where the levels of pigmentation, dye, or other additives directly influenced the absorption in the near-infrared. Figure 10 depicts the spectral signal difference between samples that show (orange-colored ABS filament) and do not show (gray-colored ABS filament) absorption of near-infrared radiation. An in-depth analysis of the interaction of the additive mask with the intrinsic spectral analysis of the polymer should be further investigated.

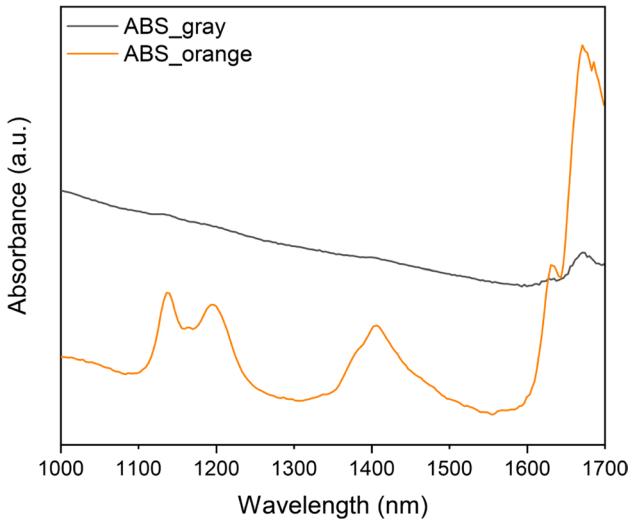

Fig. 10 Spectra of samples that show (orange-colored ABS filament) and do not show (grey colored ABS filament) absorption of near-infrared radiation

3.4 Enhanced classification algorithm

The formulation of the enhanced classification algorithm was based on chemometric activities applied to spectral data and previous exploratory analysis results. The exploratory analysis procedures were replicated in MATLAB® version 8.5, developed by MathWorks Inc. Initially, the spectral data were imported into the software in a 171-sample matrix format for 228 wavelengths (nm) values. Numerical response vectors created identified the classes (− 1, 0, and 1 were the indicators for PLA, ABS, and PETG, respectively). The possible outputs of the program were classification of the PLA, ABS, and PETG polymer filament spectra or guidance for a manual operation of the system.

The first decision node inserted at the beginning of the proposed classification algorithm aims to solve the problem of data with no near-infrared absorbance, discriminating the input spectra with no absorbance information. This decision node consists of calculating the angular coefficient of spectra in the 901.32 to 1701.23 nm range for discriminating spectral data with low absorbance information. According to Eq. 1, function b returns the slope line of the linear regression of the spectra formed by x wavelength values and y absorbance measurements. Here, the possible outputs are the continuation of the screening process or a manual selection of the processing parameters of the 3D printing equipment, depending on threshold value b.

$$b = \frac{\sum \left(x - \underline{x}\right)\left(y - \underline{y}\right)}{\sum \left(x - \underline{x}\right)^2} \quad (1)$$

The function in charge of the decision node defines the following two outputs: (1) if the angular coefficient times one thousand is higher than − 0.092, the spectrum will undergo classification, and (2) if the angular coefficient is lower than − 0.092, the system directs the user to manual operation. Value − 0.092 was acquired by an outlier detection algorithm, called Median Absolute Deviation. A 0.5% cutoff was chosen for each side, resulting in a value of 4.6, and the average between the smallest and the largest value of the distribution was calculated, generating a − 0.092 threshold value.

After the first decision node, only 3 × 3x16 = 144 samples were selected; the spectra selected for classification were subjected to the pre-treatment discussed in the Data Preparation section. Figure 11 shows the score plot of the PCA projection that illustrates the ability of the enhanced algorithm to treat and separate the spectral data toward facilitating the construction of the decision planes of the LDA model.

3.5 Model training

In the training stage, the treated spectra are resized by the projection of the PCA, specifically by three principal components. The result of the PCA projection aided the training performed by the Classification Learner application of MATLAB®, in which the LDA method was selected as the classification model. For training, the generalization ability of the classification model was evaluated by a 10-segment cross-validation.

The classification performance of the model can be better understood by analyzing the LDA mechanism when aided by the PCA projection. Figure 12 shows the decision planes (PLA-ABS, PLA-PETG, and ABS-PETG) created by LDA, which are responsible for the separations between the sample scatters in the score plot. PLA, ABS, and PETG samples are clustered in limited and distinct regions in the three-dimensional space of the principal components (PC-1, PC-2, and PC-3), which explains the maximum percentage of the training accuracy.

4 Algorithm tests

Test cycles check the classification accuracy of the samples and generate a final evaluation of the algorithm. In this step, the model is subjected to new spectral data examples for the quantification of the input data, as well as the understanding and generalization of the classifier prediction. In this context, the data augmentation technique has proved extremely useful in providing modified copies of the spectra and meeting the shortage of new data with specific traits. The data augmentation applied in this study generated three distinct subsets of data namely, A, B, and C. Subset A is formed by

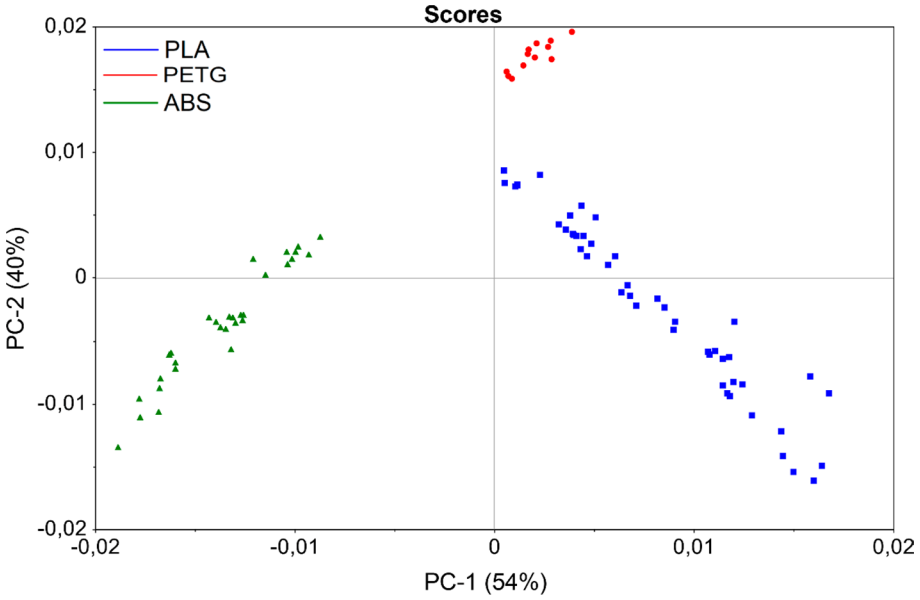

Fig. 11 PCA resulting from the processes defined for the algorithm. PLA, ABS, and PETG samples are grouped in limited and distinct regions in the score plot, justifying the high accuracy of the classifier

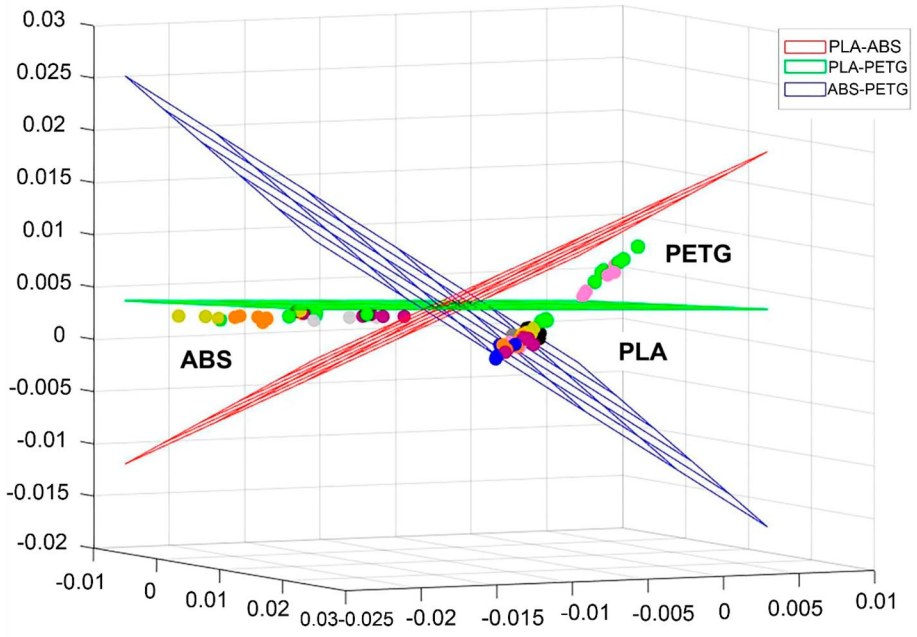

Fig. 12 Distribution of NIR spectra samples in three-dimensional space (PC-1, PC-2, and PC-3) and separated by the PLA-ABS, PLA-PETG, and ABS-PETG decision planes of the Linear Discriminant Analysis (LDA)

samples created from the averaging of the absorbance of spectra of a same material, but different colors. Subset B was generated by the percentage addition of the absorbance of spectra from different materials, but of the same coloration. Finally, subset C was obtained by percent addition of the absorbance of spectra of the same material (ABS), but of distinct signal strengths. Case studies developed from such subsets can evaluate the classification capability of the algorithm.

Figure 13 shows the spectral data resulting from the data augmentation procedure—the spectra in Fig. 13a, b, and c refer to subsets A, B, and C, respectively—note the evident similarity in the configurations of the spectra in Fig. 13a and the original PLA, ABS, and PETG spectra (Fig. 5). In contrast, Fig. 13b shows the increase in the absorption signal percentage from one material to the other led to a loss of pattern information in the original spectral data, as indicated by the arrows. Thus, the ability of the classifier was measured using similar (subset A) and distinct (subset B) spectra to those already observed during training. Figure 13c, on the other hand, shows the mixing of absorbance signals from the spectra of yellow ABS (good signal) and gray ABS (bad signal) filaments reduced absorbance amplitude, as seen in the samples with a 10%/90% ratio.

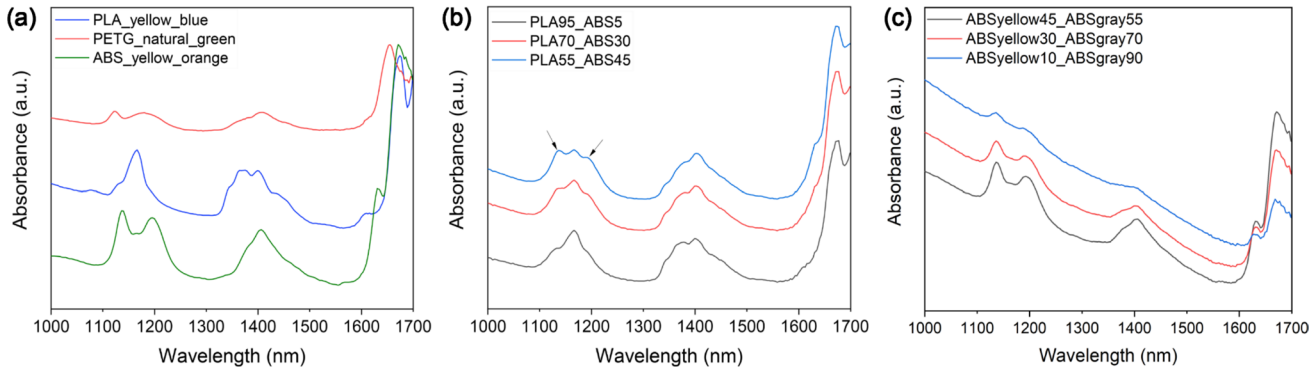

Fig. 13 Data augmentation results. **a** Subset A: average absorbance of spectra of same material and different colorations. **b** Subset B: percentage difference of the absorbance of spectra of different materials and same coloration. **c** Subset C: percentage difference of spectra with high and low sign of absorbance

The case study that used spectral data from subset A confirmed the stability of the classifier, showing a maximum result of the evaluation metrics, as expected, since the characteristic absorbance bands are maintained after the data production. Another important finding is that the entire data processing applied to the test subsets ensured the baseline variation in the spectra would not impair the classification accuracy, thus showing the effectiveness of the algorithm in classifying spectra typical of the analyzed materials, regardless of the baseline variation.

The case study that used spectral data of the subset B consisted of a stress test of the classifier, since the algorithm was subjected to spectra with mixtures of absorbance signals from two different materials, obtaining spectra previously unseen in training. The goal was an analysis of the system´s behavior and obtaining its classification limit. Considering the correct classifications given to the material with the highest percentage of absorbance signal, the results of the evaluation metrics showed maximum classification accuracy up to the mixing ratio of 70%/30%. In case, 12 incorrect identifications out of the 54 samples tested were detected and referred to the 55%/45% ratio (77.78% of accuracy). In conclusion, it is observed that the algorithm performs the classification according to the comparison of similarities between the spectra learned in training and approximation of the scatter groupings delimited by the decision plans described in Fig. 11.

The last case study evaluated the effectiveness of the first decision node in discriminating spectra with low absorbance signals. The data augmentation procedure mixed the information from low and high absorbance spectra, thus generating subset C, and a linear regression calculated for the spectral data in this subset obtained slope values (b), which were compared with the -0.092 threshold value. As addressed elsewhere, b must be greater than − 0.092 in the first decision node for the algorithm to guide the collected spectrum to the data treatment procedures and subsequent classification. The results showed the only samples in subset C that completely met this condition were composed of 45% information from spectra with good absorbance signal and 55% from spectra with low absorbance signal (dark gray curve in Fig. 13c). The other samples, i.e., those composed of 30%/70% and 10%/90% (red and blue curves in Fig. 13c, respectively), were discriminated by the algorithm, proving the first decision node corresponds to an excellent filter of spectra with a low absorbance signal.

In addition to contributing to the understanding of the prediction limit of the classifier, the testing step guided the development of the roadmap for the best prediction condition of the algorithm. Figure 14 shows a flowchart of the classification algorithm developed in this study.

5 Conclusions

This manuscript introduced a novel classifier that identifies the raw polymeric material used in AM (or in other applications), enabling the construction of a system that automatically adjusts the equipment's processing parameters (mainly plate temperature, hotend temperature, and extrusion speed). The design of the roadmap of the classification algorithm was based on the alignment of Chemometrics procedures and machine learning fundamentals. The implementation steps were essential for obtaining training with maximum classification metrics when applying a 10-segment cross-validation, and tests with new data accurately recognized spectra similar to those presented in training and classified spectral data of varying aspects. For these latter, the classifier's limit of correct predictions was reached for samples with percent absorbance signal mixtures of 70%/30%.

The exploratory data analysis proved the potential of NIR spectroscopy as a sensing method that accurately identified the characteristics of each class of material studied. The features of the samples were more prominent when maximized

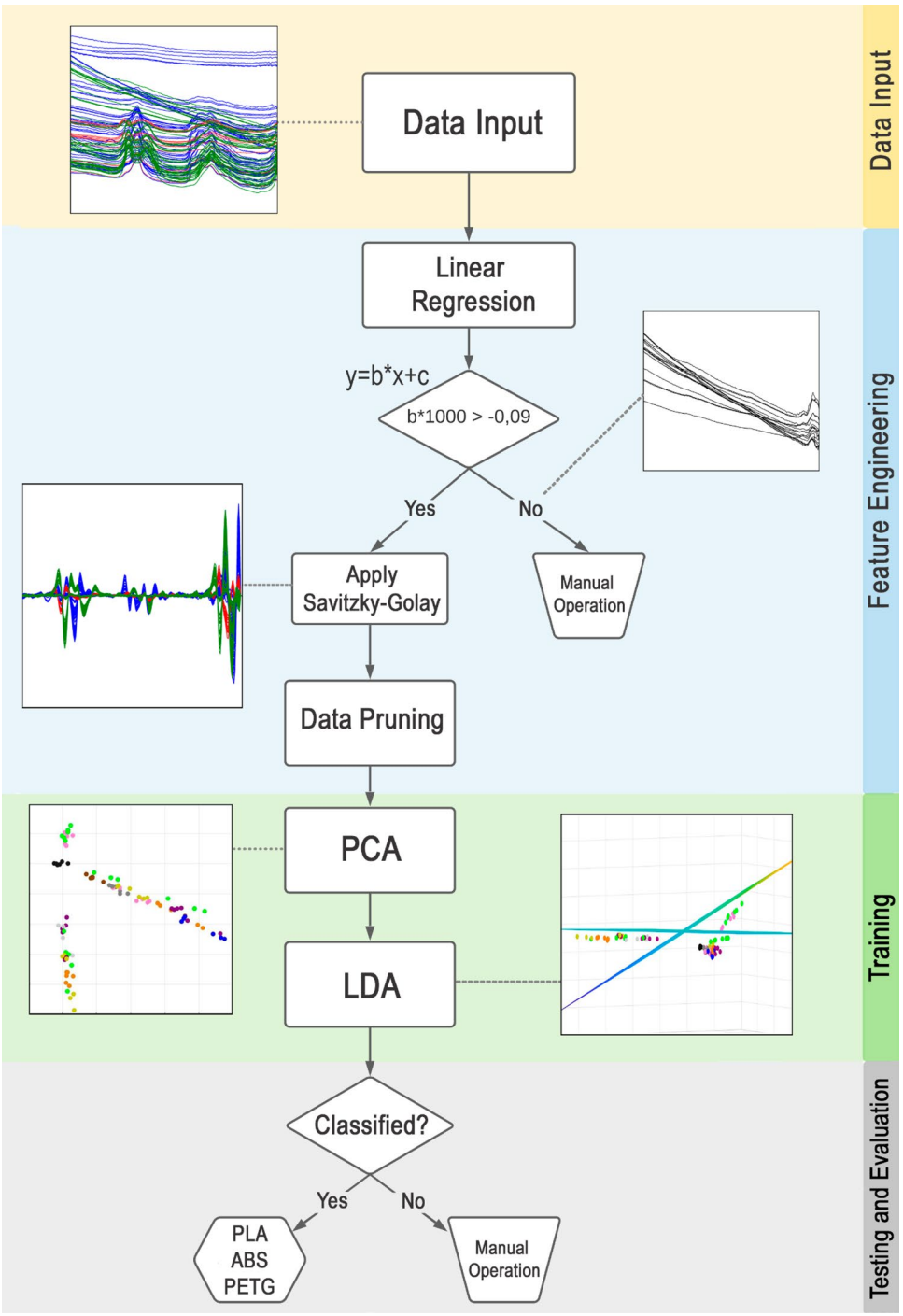

Fig. 14 Flowchart of the development of the material identification algorithm divided into data input, feature engineering, training, testing, and evaluation

by digital filtering and noise smoothing processes, such as the Savitzky–Golay method. The exploratory analysis also showed the relevance of data dimensionality reduction performed by PCA concerning the classification performance of linear models such as LDA which created decision hyperplanes that discriminated scatters from samples in the PCA score plot. Therefore, the capability of the classifier depends directly on feature engineering, as well as on the quality of the pre-treatment of the spectral data, effective PCA projection, and choice of the best classification model. Furthermore, the algorithm proved its high capacity for discriminating the input of spectra with little spectral information from the analyzed material.

According to the results, the proposed classifier amounts to a promising resource for optimizing a program that automates the usual MA processes. The field of application of 3D printing has been significantly expanded, and less costly modern miniaturized sensing devices have emerged. Studies

of additive processes have currently focused on optimizing the relationship between processing parameters and production quality. Therefore, our classification algorithm grants the initial procedures of a system that, aligned with the aforementioned studies, increase both the accessibility and productivity of AM in several manufacturing domains.

Finally, the integration of the method with the open-source work of Parsekian et al. [22] would enable online and real-time automatic parameters adjustments on low-cost 3D printing machines using G-CODEs based on the RepRap project. Even a less experienced user would be able to load a 3D printer with PLA, ABS or PETG filament, and the printer automatically configures the printing parameters during the process, even if the filament is changed during the 3D printing process.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40430-022-03645-1.

Acknowledgements The authors acknowledge Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Brazil—finance code 001 (fellowship 88887.200686/2018-00) for the financial support, 3D Fila®, 3D Lab®, Cliver®, and UP3D® companies for the donation of the polymer filaments, and Prof. J.J.R. Rohwedder for the use of NIR spectroscopy. INCTAA (CNPq 465768/2014-8 and FAPESP 2014/50951-4) is also acknowledged for financial support.

Authors' contribution All authors contributed to the conception and design of the study. Material preparation and data collection were carried out by Daniel Aparecido Lopes Vieira da Cunha, Marcia Cristina Branciforti, Rafael Vidal Aroca, Rafael Luis Ribessi, and Ivo Milton Raimundo Jr. Data analysis was performed by Daniel Aparecido Lopes Vieira da Cunha, André Carmona Hernandes, Marcia Cristina Branciforti, and Rafael Vidal Aroca. The first draft of the manuscript was written by Daniel Aparecido Lopes Vieira da Cunha and André Carmona Hernandes, and all authors commented on the previous versions of the manuscript. All authors have read and approved the final manuscript.

Funding Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES): 88887.200686/2018–00 INCTAA (CNPq 465768/2014–8 and FAPESP 2014/50951–4).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Availability of data and materials As supplementary material.

Code availability Upon request.

Consent to participate All authors consent to participate.

Consent for publication All authors consent for publication.

References

1. Aroca RV, Branciforti MC, Cunha DALV, Cardinal GC, Endo MT (2021) Sistema para ajuste automático de máquinas de processo de manufatura de materiais plásticos por identificação de materiais. INPI Patent No. BR 10 2021 007758 1. 04/23/2021
2. Beć KB, Grabska J, Huck CW (2020) Principles and applications of miniaturized near-infrared (NIR) spectrometers. Chem Eur J 27(5):1514–1532. https://doi.org/10.1002/chem.202002838
3. Beć KB, Grabska J, Siesler HW, Huck CW (2020) Handheld near-infrared spectrometers: Where are we heading? NIR News 31(3–4):28–35. https://doi.org/10.1177/0960336020916815
4. Becker W, Sachsenheimer K, Klemenz M (2017) Detection of black plastics in the middle infrared spectrum (MIR) using photon up-conversion technique for polymer recycling purposes. Polymers 9:435. https://doi.org/10.3390/polym9090435
5. Burns D (2008) Handbook of near-infrared analysis. CRC Press, Boca Raton
6. Burns DA, Ciurczak EW (2008) Handbook of near infrared analysis. CRC Press, Boca Raton, EUA
7. Cui Y, Kara S, Chan KC (2020) Manufacturing big data ecosystem: A systematic literature review. Robot Computer-Integr Manuf 62:101861. https://doi.org/10.1016/j.rcim.2019.101861
8. Cummings MM (2014) Man versus machine or man machine? IEEE Intell Syst 29(5):62–69. https://doi.org/10.1109/mis.2014.87
9. Dhinakaran V, Kumar KM, Ram PB, Ravichandran M, Vinayagamoorthy M (2020) A review on recent advancements in fused deposition modeling. Mater Today Proc 27:752–756. https://doi.org/10.1016/j.matpr.2019.12.036
10. Dilberoglu UM, Gharehpapagh B, Yaman U, Dolen M (2017) The role of additive manufacturing in the era of industry 4.0. Procedia Manufactur 11:545–554. https://doi.org/10.1016/j.promfg.2017.07.148
11. Franco D, Ganga GMD, de Santa-Eulalia LA, Filho MG (2020) Consolidated and inconclusive effects of additive manufacturing adoption: a systematic literature review. Comput Ind Eng 148:106713. https://doi.org/10.1016/j.cie.2020.106713
12. Gemperline P (2006) Practical guide to chemometrics. CRC/Taylor & Francis, Boca Raton
13. Ghobakhloo M (2020) Industry 4.0 digitization, and opportunities for sustainability. J Clean Prod 252:119869. https://doi.org/10.1016/j.jclepro.2019.119869
14. Goh GD, Sing SL, Yeong WY (2020) A review on machine learning in 3D printing: applications, potential, and challenges. Artif Intell Rev 54(1):63–94. https://doi.org/10.1007/s10462-020-09876-9
15. Haverkort BR, Zimmermann A (2017) Smart industry: How ICT will change the game! IEEE Internet Comput 21(1):8–10. https://doi.org/10.1109/mic.2017.22
16. Korner MEH, Lambán MP, Albajez JA, Santolaria J, del Carmen Ng Corrales L, Royo J (2020) Systematic literature review: Integration of additive manufacturing and industry 4.0. Metals, 10(8), 1061. https://doi.org/10.3390/met10081061
17. Kumar MB, Sathiya P (2021) Methods and materials for additive manufacturing: a critical review on advancements and challenges. Thin-Wall Struct 159:107228. https://doi.org/10.1016/j.tws.2020.107228
18. McLauchlin AR, Ghita O, Gahkani A (2014) Quantification of PLA contamination in PET during injection moulding by in-line NIR spectroscopy. Polym Test 38:46–52. https://doi.org/10.1016/j.polymertesting.2014.06.007
19. Nardo M, Forino D, Murino T (2020) The evolution of man–machine interaction: the role of human in industry 4.0 paradigm.

Product Manuf Res 8(1):20–34. https://doi.org/10.1080/21693277.2020.1737592
20. Ngo TD, Kashani A, Imbalzano G, Nguyen KT, Hui D (2018) Additive manufacturing (3D printing): a review of materials, methods, applications and challenges. Compos B Eng 143:172–196. https://doi.org/10.1016/j.compositesb.2018.02.012
21. Oztemel E, Gursev S (2018) Literature review of industry 4.0 and related technologies. J Intell Manuf 31(1):127–182. https://doi.org/10.1007/s10845-018-1433-8
22. Parsekian PHL, Cunha DALV, Watanabe FY, Branciforti MC, Aroca RV (2020) Failure monitoring and recovery system during manufacturing process. IEEE Lat Am Trans 18(02):407–413. https://doi.org/10.1109/tla.2020.9085297
23. Pasquini C (2018) Near-infrared spectroscopy: a mature analytical technique with new perspectives—a review. Anal Chim Acta 1026:8–36. https://doi.org/10.1016/j.aca.2018.04.004
24. Pereira T, Kennedy JV, Potgieter J (2019) A comparison of traditional manufacturing vs additive manufacturing, the best method for the job. Procedia Manuf 30:11–18. https://doi.org/10.1016/j.promfg.2019.02.003
25. Sanchez LC, Beatrice CAG, Lotti C, Marini J, Bettini SHP, Costa LC (2019) Rheological approach for an additive manufacturing printer based on material extrusion. Int J Adv Manuf Technol 105(5–6):2403–2414. https://doi.org/10.1007/s00170-019-04376-9
26. Sarkar D (2018) Practical machine learning with python: a problem-solver's guide to building real-world intelligent systems. Apress, Springer
27. Savolainen J, Collan M (2020) How additive manufacturing technology changes business models?—review of literature. Addit Manuf 32:101070. https://doi.org/10.1016/j.addma.2020.101070
28. Schwab K (2016) The fourth industrial revolution. Crown Business, New York
29. Siesler HW (2002) Near-infrared spectroscopy: principles, instruments, applications. Wiley-VCH
30. Siesler HW et al (2008) Near infrared spectroscopy: principles, instruments, applications. Wiley-VCH, Weinheim, DE
31. Singh S, Ramakrishna S, Singh R (2017) Material issues in aditive manufacturing: a review. J Manuf Process 25:185–200. https://doi.org/10.1016/j.jmapro.2016.11.006
32. Stern A, Rosenthal Y, Dresler N, Ashkenazi D (2019) Additive manufacturing: an education strategy for engineering students. Addit Manuf 27:503–514. https://doi.org/10.1016/j.addma.2019.04.001
33. Sun B, Jämsä-Jounela S-L, Todorov Y, Olivier LE, Craig IK (2017) Perspective for equipment automation in process industries. IFAC-Papers OnLine 50(2):65–70. https://doi.org/10.1016/j.ifacol.2017.12.012
34. Tanwar S, Ramani T, Tyagi S (2018) Dimensionality reduction using PCA and SVD in big data: a comparative case study. Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering (pp. 116–125). Springer International Publishing. https://doi.org/10.1007/978-3-319-73712-6_12
35. Vyavahare S, Teraiya S, Panghal D, Kumar S (2020) Fused deposition modelling: a review. Rapid Prototyp J 26(1):176–201. https://doi.org/10.1108/rpj-04-2019-0106