Vibrational Spectroscopy 120 (2022) 103378

Contents lists available at ScienceDirect

Vibrational Spectroscopy

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

ATR-FTIR spectroscopy combined with machine learning for classification of PVA/PVP blends in low concentration

Thiago Franca [a], Daniel Goncalves [b], Cicero Cena [a,*]

[a] UFMS – Federal University of Mato Grosso do Sul, Optics and Photonic Lab (SISFOTON-UFMS), Campo Grande, MS, Brazil
[b] UFGD – Federal University of Grande Dourados, Dourados, MS, Brazil

ARTICLE INFO

Keywords:
Machine learning
FTIR
Polymeric blends
PVP/PVA
Multivariate analysis

ABSTRACT

The use of Fourier transformed infrared spectroscopy (FTIR) and multivariate analysis for sample classification has been extensively explored in the literature. The overall accuracy obtained in most of the studies raises questions about their reliability due to technical limitations and misuse of the analysis driving to overfitting/underfitting of data. There are established procedures to avoid overfitting/underfitting, but there is a lack of studies to understand the relationship between the limitation of the data provided by the FTIR and the associated method of analysis for sample classification. In this study, FTIR spectra were obtained from thick films of poly (vinyl alcohol) (PVA)/poly (vinyl pyrrolidone) (PVP) polymeric blends, with PVP concentrations below 3% (wt./wt.). We analyzed the FTIR spectra by using simple algorithms for sample classification to evaluate the predicting model accuracy at low concentrations of PVP into the PVA matrix. The raw data were submitted to pre-treatment by standard normal variate (SNV), and different spectral ranges were explored by principal component analysis (PCA). Then, PCA and FTIR-SNV data were used for supervised analysis by using linear discriminant analysis (LDA); k-nearest neighbors (KNN); and support vector machines (SVM) algorithms. The method was able to classify samples with 0.1 wt% of PVP with an overall accuracy of 100% with quadratic SVM by the proper choice of the spectral range and number of PCs. Finally, we show that the FTIR associated with multivariate analysis can be used for sample classification at low concentration changes.

1. Introduction

Poly (vinyl alcohol) (PVA) and Poly (vinyl pyrrolidone) (PVP) are water-soluble polymers able to form miscible polymeric blends due to interactions through H-bonds. These H-bonds also promote multiple intra- and intermolecular bonds between PVA hydroxyls (O-H) at crystalline regions; single H-bonds between PVA hydroxyls at amorphous regions; and cross-links between the PVP carbonyls groups (C═O) and PVA hydroxyls groups [1,2].

In general, the final properties of polymeric blends will be determined by the concentration of the mixture and the nature of their components. The precise control of such concentration has motivated some studies in the literature to quantify and/or classify these samples according to the concentration of their components [3–5]. A promising alternative to access such information is the use of molecular vibrational spectroscopy with machine learning to provide predicting models with high overall accuracy, based on an easy data obtention technique, with low cost and fast response [6–8]. However, the accuracy value achieved in multivariate analysis models is often difficult to be understood and/or accepted, unlike univariate analysis approaches, which offer a clear data variation and bias identification [8].

Studies on polymeric blends are very promising to test and prove the power of optical spectroscopy and machine learning as reliable methods for sample classification, since, the control of experimental parameters allows to easily know the sample characteristics, such as concentration and homogeneity, besides the facile reproducibility, test, and validation. Near-infrared (NIR) and Partial Least Squares (PLS) were shown to be more effective than thermal analysis to determine the fraction of polypropylene below 1 wt% into polypropylene/high-density polyethylene, and polypropylene/ acrylonitrile-butadiene-styrene blends [9]. Similar results were obtained to evaluate polycarbonate blends with different amounts of butadiene, styrene, and acrylonitrile [10].

NIR, Fourier transform infrared spectroscopy (FTIR) and Raman spectroscopy was tested to analyze the concentration of polylactic acid, propylene polycarbonate, and poly (butylene adipate co-terephthalate) into biodegradable blends. The data obtained for different

* Corresponding author.
E-mail address: cicero.cena@ufms.br (C. Cena).

https://doi.org/10.1016/j.vibspec.2022.103378
Received 10 February 2022; Received in revised form 21 March 2022; Accepted 24 April 2022
Available online 4 May 2022
0924-2031/© 2022 Elsevier B.V. All rights reserved.

concentrations were analyzed by PCA (Principal Component Analysis) and PLS, revealing the best performance for Raman spectroscopy [3]. FTIR applied for sample classification showed great accuracy for ethylene-propylene-diene composites by using KNN (K-nearest neighbor) algorithm [11]; and identifying different types of monomers (octane - C8, hexane - C6, and butane - C4) in coextruded polyethylene films by using PCA [12].

FTIR with multivariate analysis studies for blends classification usually is performed in samples with a minimum concentration above 1 wt% [9–12]. Sample classification with concentration below 1 wt% has been little explored, probably because it reaches the limits of univariate distinguishable FTIR spectra. But multivariate analysis can access small deviations in the FTIR spectra for correct sample classification, and also the most interesting questions lie on low concentration range, such as changes in biofluids concentrations for disease diagnosis, or even for DNA screaming [13,14]. In this study, we investigate the use of FTIR with machine learning algorithms for sample classification, by using Poly (vinyl alcohol)/ Poly (vinyl pyrrolidone) polymeric blends with concentrations from 3 wt% to 0.1 wt%.

2. Materials and methods

2.1. Sample preparation and data acquisition

Poly (vinyl alcohol) (PVA), $(C_2H_4O)_x$, 86.5–89.5% hydrolyzed, was supplied from Dinamica Chemical Co., and Poly (vinyl pyrrolidone) (PVP), $(C_6H_9NO)_n$, with an average molecular weight of 360.00 was supplied in powder form from Sigma-Aldrich Chemical Co. The PVA/PVP solution was prepared by slow dissolution of precursor powders in 25 ML of hot deionized water (50 °C) under constant stirring for 2 h. The blend thick films were obtained by casting the precursor solution into a Poly (tetrafluoroethylene) (PTFE) mold and drying at 50 °C per 240 min. The samples were produced in five different concentrations, 3.0 wt%, 1.0 wt%, 0.5 wt% and 0.1 wt% of PVP into PVA matrix. A total of 30 samples (in duplicate) were produced for each concentration and kept under a vacuum before the measurements to avoid water adsorption.

The samples were measured in a Fourier transform infrared (FTIR) spectrophotometer (Spectrum 100, Perkin Elmer) using an attenuated total reflectance (ATR) accessory. The thick films were directly introduced in the ATR accessory, and the FTIR spectra were taken in the range from 4000 to 600 cm^{-1}, with a 4 cm^{-1} resolution and 10 scans. The average spectra were collected by measuring each sample in 3 different spots of each duplicate.

2.2. Data analysis and sample classification

The data analysis was performed with the aid of Matlab R2018a software. First, FTIR raw spectra were subjected to a standard normal variate (SNV) preprocessing step to remove offset [15,16], then average FTIR-SNV spectra were calculated for each sample. An exploratory analysis was performed by Principal Component Analysis (PCA) [17,18] by using the average FTIR-SNV spectra in three different ranges: (i) 4000–600 cm^{-1}; (ii) 3800–2800 cm^{-1}; and (iii) 1800–800 cm^{-1}. The PCA reduces the data dimensionality by a linear orthogonal transformation of the initial data into a new set of uncorrelated variables, called principal components (PCs), which carry information about the data variance. The PCs plot (score plot) allows to observe the clustering trend of each sample group in a cartesian graph according to the data variance for each PC generated (or PC axis). Related to each PC we also obtain the weight (loading plot) from linear transformation, which is related to the original variables, and indicate the molecular vibrational modes with the highest variance for those PCs.

The PCA data originated from the averaged FTIR-SNV spectra were used for the supervised analysis step, by using machine learning algorithms. We chose to use the following machine learning (ML) algorithms to build the predicting models: linear discriminant analysis (LDA) [19,20]; k-nearest neighbors (KNN) [21], and support vector machines (SVM) [22] with different functions. The discriminant analysis uses a linear (LDA) combination of features to separate the samples groups, defining a border between the groups in a spatial distribution; SVM performs the group separation by the construction of hyperplanes (linear, quadratic, cubic); and KNN classifies each sample based on the number of closest neighbors from a determined group, basically using the distances between then. The algorithms test different functions for each classifier to build the predicting model and use a different number of PCs to perform the sample classification with higher accuracy. To avoid underfitting and overfitting [22,23] the ideal number of PCs was determined by analyzing the overall accuracy value as a function of the number of PCs, we adopt the number of PCs corresponding to the first maximum value of the overall accuracy.

Finally, the overall accuracy of the predicting model is tested by leave-one-out cross-validation (LOOCV) [24]. In the LOOCV, a sample from the data set is withdrawn from the training set, then the predicting model is built and tested in this sample. This process is repeated until all the samples are tested. The accuracy in each interaction is registered, and in the end, a confusion matrix is generated exhibiting the average value of the overall accuracy.

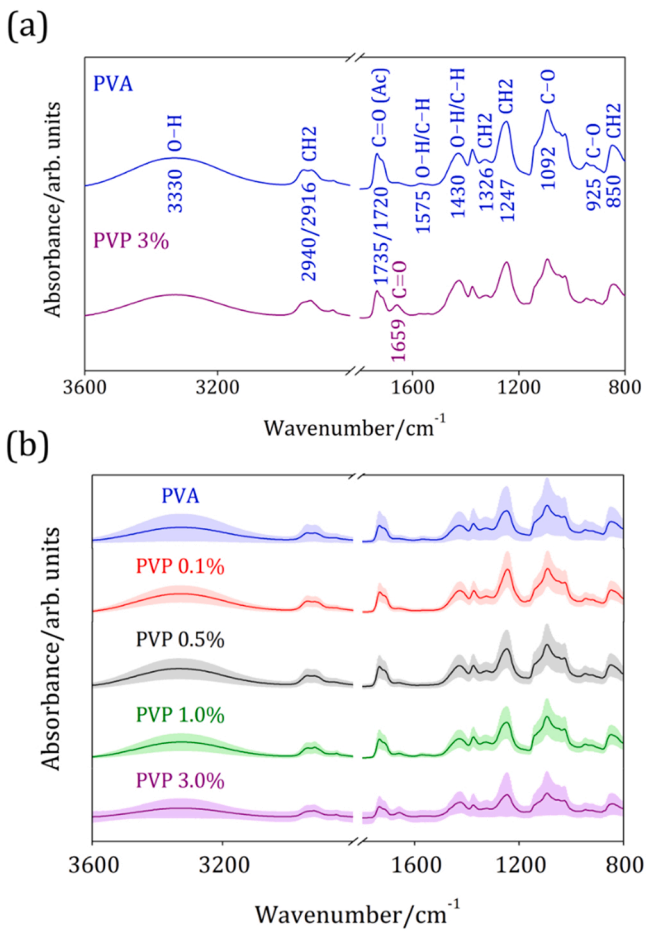

Fig. 1. (a) Average FTIR-SNV spectra main vibrational bands from pure PVA and PVA/PVP (3 wt%); (b) average FTIR-SNV spectra with standard deviation (shadow) for PVA/PVP blends with different concentrations of PVP (0, 0.1, 0.5, 1.0, and 3 wt%).

3. Results and discussion

The average FTIR-SNV spectra for pure PVA, and PVA/PVP (3 wt%) blend thick films, Fig. 1(a), exhibit some dissimilar bands around 2940 cm^{-1} and 2916 cm^{-1} assigned to CH_2 asymmetric stretching; and 1430

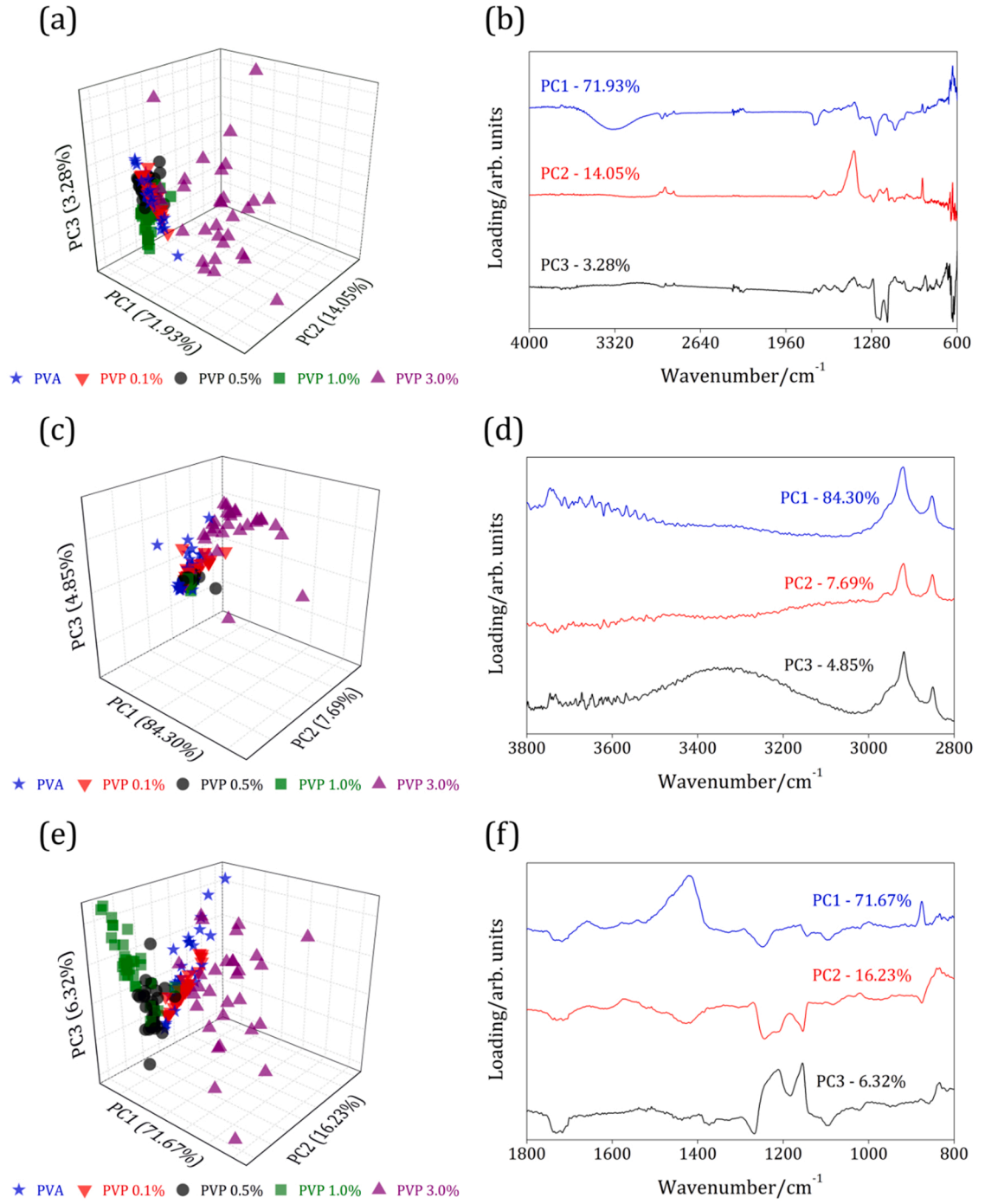

Fig. 2. PCA results for FTIR-SNV data by using three different spectral ranges: (a-b) 4000–600 cm^{-1};(c-d) 3800–2800 cm^{-1}; (e-f) 1800–800 cm^{-1}. In the left column, we observe the score plot, and in the right column the loading plot.

cm^{-1} assigned to stretching modes from phenols and alcohols groups [25–27]. But the main difference is due to the 1659 cm^{-1} vibrational band, assigned to C═O stretching from the carbonyl group, which is more prominent at PVA/PVP blend. We know that this band can be used to differ pure PVA from PVA/PVP blend by univariate analysis, so the idea is to work in concentrations below 3 wt% and observe how the FTIR spectra of PVA/PVP blend became similar with pure PVP, and also if multivariate analysis by using machine learning can correctly classify the sample in this situation.

Fig. 1(a) also shows the wideband around 3330 cm^{-1} assigned to O-H symmetric stretching from hydroxyl groups; CH_2 asymmetric stretching bands from the methyl group, 1326 cm^{-1} (out of plane), 1247 cm^{-1} (bending), and 850 cm^{-1} (wagging); C-O symmetric stretching wide band around 1096 cm^{-1}, and 925 cm^{-1}; and C-H bending at 962 cm^{-1} [25–30]. The average FTIR-SNV spectra (strong line) with respective standard deviation (shadow) for PVA/PVP blends with different concentrations of PVP (0, 0.1, 0.5, 1.0, and 3 wt%), Fig. 1 (b), exhibit high similarities between the samples, except by the 1659 cm^{-1} vibrational band assigned to C═O, and C-N stretching. The 1659 cm^{-1} band intensity decreases with the concentration of PVP, while the other vibrational bands seem to remain unchanged [28–30].

A sample classification can be more easily performed if a clustering

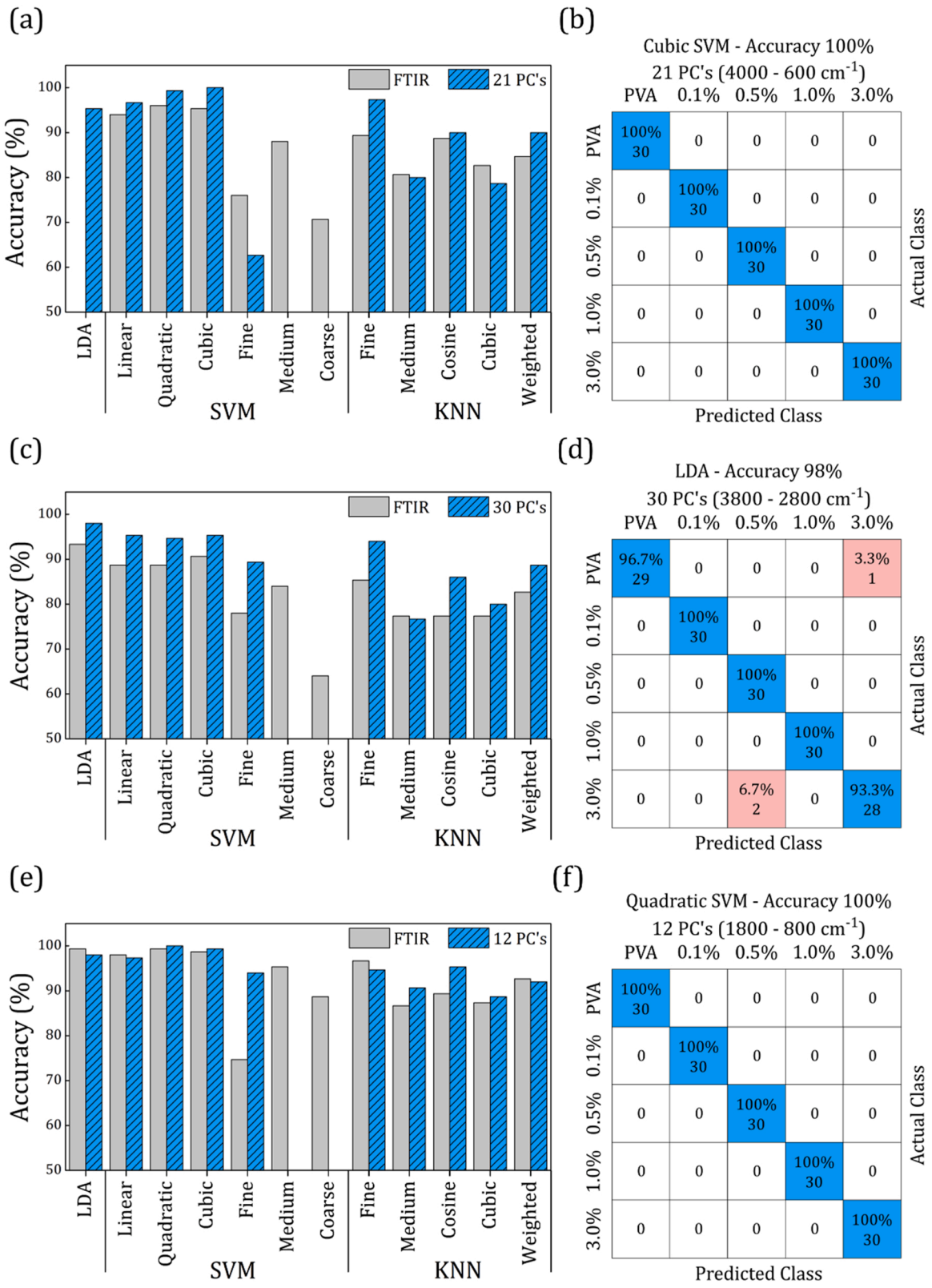

Fig. 3. Overall accuracy for different ML algorithms (left side) for FTIR-SNV and PCA dataset; and confusion matrix (right side) for the ML algorithm with highest overall accuracy. Different spectral ranges were analyzed: (a-b) 4000–600 cm^{-1};(c-d) 3800–2800 cm^{-1}; (e-f) 1800–800 cm^{-1}.

tendency is observed in the score plot. If the PCA data carries a huge number of highly correlated variables the clustering formation is hampered, and to improve the clustering tendency we may select a proper spectral range by removing high correlated variables [31–33].

Fig. 2 shows the PCA results for FTIR-SNV data by using three different spectral ranges, the left column we observe the score plot, and the right column the loading plot.

The 4000–600 cm^{-1} score plot, Fig. 2(a), shows a dense clustering

formation with no clear boundaries between the groups with concentration below 3 wt%, which showed a large dispersion along with the first 3 PCs. These first three PCs are responsible for 89.26% of data variance. The respective loading plot, Fig. 2(b), exhibits great data variance in the 1800–800 cm^{-1}, followed by the 3800–2800 cm^{-1} spectral range.

A similar pattern was observed in the 3800–2800 cm^{-1} spectral range, the sample group with 3 wt% forms a separated cluster with large data dispersion in the first three PCs, Fig. 2(c), while a dense cluster with no clear boundaries between the group samples was formed. These first three PCs are responsible for 96.84% of data variance, and the loading plot shows that the main variance occurs for the 2940 cm^{-1} and 2916 cm^{-1} vibrational bands, assigned to CH_2 asymmetric stretching.

A promising result was obtained at 1800–800 cm^{-1} spectral range, Fig. 2(e), with 94.22% of data variance for the first three PCs. The clustering formation for 3 wt% concentration remains with the same character as before, well-dispersed with a clear boundary separation from the others. Followed by the 1 wt% group, in which some samples share the same region with the 0.5 wt% group, and both are separated from the others. Finally, the 0.1 wt% and pure PVA group are in the center of the score plot, but also exhibit some level of separation between them. The PC1 loading, Fig. 2(f), shows the main variance around the 1430 cm^{-1} vibrational band, then PC2 and PC3 loading exhibit their main variance in the 1300–1100 cm^{-1} spectral range. A careful analysis of Fig. 2(e) evidenced the main role played by PC2 and PC3 for the sample group separation.

Machine learning algorithms were applied to sample group classification by using the FTIR raw data and PCA data, Fig. 3. In general, the overall accuracy was slightly improved by using PCA data [34,35], only for 1800–800 cm^{-1} spectral range a small difference between in the overall accuracy was observed for both datasets. We focused on the use of PCA data since it improves the computational processing because contains a reduced dataset. The overall accuracy of 100% for Cubic SVM in the 4000–600 cm^{-1} range, Fig. 3(a-b), was obtained by using 21 PCs with 99.34% of data variance.

The overall accuracy of 98% at 3800–2800 cm^{-1} range, Fig. 3(c), was obtained by the LDA method by using the 30 first PCs, caring 99.94% of data variance. The confusion matrix, Fig. 3(d), showed a 100% accuracy for low concentrations groups (1 wt%, 0.5 wt%, and 0.1 wt%), while 1 pure PVA sample and 2 PVA/PVP blend with 3 wt% was misclassified, it is probably related to sample outliers caused by the 3330 cm^{-1} (O-H) vibrational band assigned to water molecules. The outlier's removal must be performed before the ML classification [36,37], here we choose to keep every sample from the dataset to observe this situation. Finally, an overall accuracy of 100% at the 1800–800 cm^{-1} range, Fig. 3(e), was obtained for Quadratic SVM by using 12 PCs, caring 99.44% of data variance.

Besides the overall accuracy of 100% in two different spectral ranges, the method which provides accuracy using the smallest possible number of PCs from the score matrix (1800–800 cm^{-1}) is very interesting for applications purposes, because it uses less computational performance with faster response [38]. Then, Quadratic SVM can be considered the most promising algorithm, since it reached 100% accuracy with the smallest number of PCs, by using a smaller data matrix than the full spectrum range.

4. Conclusions

The main band around 1659 cm^{-1} assigned to the PVP presence in the PVA/PVP blends seems to disappear below 1 wt%, reducing the possibility of identifying PVP in low concentrations by using univariate analysis. But the multivariate classification results allowed the distinction between all sample groups analyzed with an overall accuracy of 100% in the 4000–600 cm^{-1}, and 1800–800 cm- 1 spectral range with Cubic SVM and Quadratic SVM, respectively. The Quadratic SVM in the 1800–800 cm^{-1} range, used only the first 12 PCs to correctly classify each sample, while the Cubic SVM in the 4000–600 cm^{-1} range, used the first 21 PCs, and both carries similar data variance around 99.4%. The outlier's remotion is a necessary step in this kind of study, even when the samples can be produced with high control of concentration and homogeneity. Finally, Fourier Transform Infrared Spectroscopy Technique associated with the multivariate analysis (machine learning) demonstrates great potential to classify samples with low concentration deviation.

CRediT authorship contribution statement

Cicero Cena: Conceptualization, Methodology, Supervision, Writing – review & editing. **Thiago Franca.**: Data curation, Software, Investigation, Writing – original draft. **Daniel Goncalves**: Visualization, Validation.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Thiago Franca reports financial support was provided by Coordination of Higher Education Personnel Improvement.

Acknowledgments

This study was financed by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grant number: 408127/2018-0; 440214/2021-1. The authors also thank to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.vibspec.2022.103378.

References

[1] M. Teodorescu, M. Bercea, S. Morariu, Biomaterials of PVA and PVP in medical and pharmaceutical applications: Perspectives and challenges, Biotechnol. Adv. 37 (1) (2019) 109–131.
[2] I.S. Elashmawi, H.A. Baieth, Spectroscopic studies of hydroxyapatite in PVP/PVA polymeric matrix as biomaterial, Curr. Appl. Phys. 12 (1) (2012) 141–146.
[3] L. Bokobza, Some applications of vibrational spectroscopy for the analysis of polymers and polymer composites, Polymers 11 (7) (2019) 1159.
[4] J. Moros, S. Garrigues, M. Guardia, Vibrational spectroscopy provides a green tool for multi-component analysis, TrAC Trends Anal. Chem. 29 (7) (2010) 578–591.
[5] M. Rani, C. Marchesi, S. Federici, S. Rovelli, I. Alessandri, I. Vassalini, S. Ducoli, L. Borgese, A. Zacco, F. Bilo, E. Bontempo, L. Depero, Miniaturized near-infrared (MicroNIR) spectrometer in plastic waste sorting, Materials (12) (2019). Article 2740.
[6] R. Chen, X. Huang, M. Wang, G. Jin, Evaluation and comparison of mid-infrared, Raman, and near-infrared spectroscopies for characterization and determination of the compositions in fully biodegradable poly (lactic acid)/poly (propylene carbonate)/poly (butylene adipate-co-terephthalate) blends combined with chemometrics, J. Macromol. Sci. - Part A 53 (6) (2016) 354–361.
[7] D.A. Silva, Daniel José, H.élio wiebeck, Using PLS, iPLS, and siPLS linear regressions to determine the composition of LDPE/HDPE blends: a comparison between confocal Raman and ATR-FTIR spectroscopies, Vib. Spectrosc. 92 (2017) 259–266.
[8] Benita Percival, Miles Gibson, Justine Leenders, Philippe B. Wilson, Martin Grootveld, Univariate and Multivariate Statistical Approaches to the Analysis and Interpretation of NMR-based Metabolomics Datasets of Increasing Complexity, in: Computational Techniques for Analytical Chemistry and Bioanalysis, Royal Society Of Chemistry, 2020, pp. 1–40.
[9] W. Camacho, S. Karlsson, NIR, DSC, and FTIR as quantitative methods for compositional analysis of blends of polymers obtained from recycled mixed plastic waste, Polym. Eng. Sci. 41 (9) (2001) 1626–1635.
[10] Y. Sulub, J. Derudder, Determination of polymer blends composed of polycarbonate and rubber entities using near-infrared (NIR) spectroscopy and multivariate calibration, Polym. Test. 32 (4) (2013) 802–809.
[11] J.R.R. Ruiz, T. Canals, R. Cantero, Supervision of ethylene propylene diene M-class (EPDM) rubber vulcanization and recovery processes using attenuated total

reflection Fourier transform infrared (ATR FT-IR) spectroscopy and multivariate analysis, Appl. Spectrosc. 71 (1) (2017) 141–151.
[12] E. Caro, E. Comas, Polyethylene comonomer characterization by using FTIR and a multivariate classification technique, Talanta 163 (2017) 48–53.
[13] G. Larios, M. Ribeiro, C. Arruda, S.L. Oliveira, T. Canassa, M.J. Baker, B. Marangoni, C. Ramos, C. Cena, A new strategy for canine visceral leishmaniasis diagnosis based on FTIR spectroscopy and machine learning, J. Biophotonics 14 (11) (2021). Article e202100141.
[14] T.G. Rios, G. Larios, B. Marangoni, S.L. Oliveira, C. Cena, C.A.N. Ramos, FTIR spectroscopy with machine learning: a new approach to animal DNA polymorphism screening, Spectrochim. Acta – Part A (261 () (2021). Article 120036.
[15] J. Engel, J. Gerretsen, E. Szymanska, J. Jansen, G. Downey, L. Blanchet, L.M. C. Buydens, Breaking with trends in pre-processing? TrAC Trends Anal. Chem. 50 (2013) 96–106.
[16] A. Rinnan, F. Van Den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends Anal. Chem. 28 (10) (2009) 1201–1222.
[17] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philoso. Trans. R. Soc. A: Math. Phys. Eng. Sci. 374 (2065) (2016) 20150202.
[18] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (6) (1933) 417.
[19] R.G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pmerantsev, O. Rodionova, J. M. Roger, B. Walczak, R. Tauler, Chemometrics in analytical chemistry—part II: modeling, validation, and applications, Anal. Bioanal. Chem. 410 (26) (2018) 6691–6704.
[20] A. Tharwat, G. Tarek, I. Abdelhameed, H.A. Ella, Linear discriminant analysis: a detailed tutorial, AI Commun. 30 (2) (2017) 169–190.
[21] A. Mucherino, P.J. Papajorgji, P.M. Pardalos, K-Nearest neighbor classification, Data Min. Agric. 34 (2009) 83–106.
[22] J. Luts, F. Ojeda, R.V. Plas, B.D. Moor, S.V. Huffel, J.A.K. Suykens, A tutorial on support vector machine-based methods for classification problems in chemometrics, Anal. Chim. Acta 665 (2) (2010) 129–145.
[23] C. Maione, B.L. Batista, A.D. Campiglia, F. Barbosa, R.M. Barbosa, Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry, Comput. Electron. Agric. 121 (2016) 101–107.
[24] S. Wold, Cross-validatory estimation of the number of components in factor and principal components models, Technometrics 20 (4) (1978) 397–405.
[25] E.M. Abdelrazek, I.S. Elashmawi, S. Labeeb, Chitosan filler effects on the experimental characterization, spectroscopic investigation, and thermal studies of PVA/PVP blend films, Phys. B: Condens. Matter 405 (8) (2010) 2021–2027.
[26] C.R. Cena, M.J. Silva, L.F. Malmonge, J.A. Malmonge, Poly(vinyl pyrrolidone) sub-microfibers produced by solution blow spinning, J. Polym. Res. (25 () (2018). Article 238.
[27] J.L. Viana, T. França, C. Cena, Solution blow spinning poly(vinyl alcohol) sub-microfibers produced from different solvents, Orbital: Electron. J. Chem. 12 (1) (2020) 1–6.
[28] W.H. Eisa, Y.K.A. M 1 am, A.A. Shabaka, A.E.M. Hosam, In situ approach induced growth of highly monodispersed Ag nanoparticles within free-standing PVA/PVP films, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 95 (2012) 341–346.
[29] M.H.A. Taleb, Thermal and spectroscopic studies of poly (N-vinyl pyrrolidone)/ poly (vinyl alcohol) blend films, J. Appl. Polym. Sci. 114 (2) (2009) 1202–1207.
[30] L. Li, C.M. Chan, L.T. Weng, The effects of specific interactions on the surface structure and composition of miscible blends of poly (vinyl alcohol) and poly (N-vinyl-2-pyrrolidone), Polymer 39 (11) (1998) 2355–2360.
[31] A.E. Casaril, A.G. Santos, B.S. Marangoni, S.M. Lima, L.H.C. Andrade, W. S. Fernandes, J.O.M. Infran, N.O. Alves, M.D.G.L. Borges, C. Cena, A.G. Oliveira, Intraspecific differentiation of sandflies specimens by optical spectroscopy and multivariate analysis, J. Biophotonics 14 (1) (2021). Article e202000412.
[32] E. Scaglia, G.D. Sockalingum, J. Schmitt, C. Gobinet, N. Schneider, M. Manfait, G. Thiefin, Noninvasive assessment of hepatic fibrosis in patients with chronic hepatitis C using serum Fourier transforms infrared spectroscopy, Anal. Bioanal. Chem. 401 (2011) 2919–2925.
[33] I.C. Oliveira, T. Franca, G. Nicolodelli, C.P. Morais, B.M. Marangoni, G. Bacchetta, D.M.B.P. Milori, C.Z. Alves, C. Cena, Fast and accurate discrimination of Brachiaria brizantha (A. Rich.) Stapf seeds by molecular spectroscopy and machine learning, ACS Agric. Sci. Technol. 1 (5) (2021) 443–448.
[34] T. Howley, M.G. Madden, M.L. O`Connell, A.G. Ryder, 2005. The effect of principal component analysis on machine learning accuracy with high dimensional spectral data. Proceedings of AI-2005, (2005).
[35] A.T. Ikram, A.K. Cherukuri, Improving accuracy of intrusion detection model using PCA and optimized SVM, J. Comput. Inform. Technol. 24 (2) (2016) 133–148.
[36] D.R. Jensen, D.E. Ramirez, Use of Hotellings T2: outlier diagnostics in mixtures, Int. J. Stat. Probab. 6 (6) (2017) 1–11.
[37] B. Brownfiel, J.H. Kalivas, Consensus outlier detection using sum of ranking differences of common and new outlier measures without tuning parameter selections, Anal. Chem. 89 (9) (2017) 5087–5094.
[38] Konrad Mulrennan, et al., A soft sensor for prediction of mechanical properties of extruded PLA sheet using an instrumented slit die and machine learning algorithms, Polym. Test. 69 (2018) 462–469.