2023 International Conference on Optical MEMS and Nanophotonics (OMN) and SBFoton International Optics and Photonics Conference (SBFoton IOPC) | 979-8-3503-0402-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/OMN/SBFOTONIOPC58971.2023.10230933

A 3D discriminant analysis for Hyperspectral FTIR images

Sajid Farooq
Instituto de Pesquisas Energéticas e Nucleares (IPEN)
São Paulo, Brazil
sajid.f@ipen.br

Gleice Germano
Instituto de Pesquisas Energéticas e Nucleares (IPEN)
São Paulo, Brazil
gleice.germano@ipen.br

Kleber Adriani Stancari
Instituto de Pesquisas Energéticas e Nucleares (IPEN)
São Paulo, Brazil
kleberstancari@yahoo.com.br

Rocío Raffaeli
Universidad Nacional de La Plata
La Plata, Argentina
rocioraffaeli@gmail.com

Maria Virginia Croce
Universidad Nacional de La Plata
La Plata, Argentina
mariavirginiacroce@gmail.com

Adela E. Croce
Universidad Nacional de La Plata
La Plata, Argentina
acroce@inifta.unlp.edu.ar

Denise Maria Zezell
Instituto de Pesquisas Energéticas e Nucleares (IPEN)
São Paulo, Brazil
zezell@usp.br

***Abstract*—Here, we apply a 3D discriminant analysis approach to analyze FTIR hyperspectral images of normal vs malignant Melanoma (MM) samples for skin cancer diagnosis. For this porpose we used 2 samples, for Normal (49k) and for MM(90k) . Our results evidence the outstanding performance with accuracy up to 81% for big data ($> 100k$).**

***Index Terms*—FTIR, Machine learning, Hyperspectral images, Skin cancer, Cancer diagnosis**

I. Introduction

According to World Health organization (WHO) data, more than 500k will be estimated to diagnose melanoma cancer skin per year from 2020-2040, predicting deaths of around 100.000 per year[1]. As known early diagnosis of this disease means a better prognosis and more well-being for patients[2]. Currently, pathologists make the final cancer diagnosis by the morphological characteristics and immunohistochemical markers. However, normally this diagnosis occurs in late stage and the development of new technologies has been encouraged[2].FTIR spectroscopy with machine learning has become an screening and diagnosis tool due to its ability to distinguish for qualitatively and quantitative manner analysis. Moreover, FTIR spectroscopy is a non-destructive and powerful tool that enables both spatial and chemical signatures of an available sample[3-5].

In this study, we utilize a computer modeling framework based on 3D discriminant analysis approach, which was composed by quadratic discriminant analysis (3D-PCA-QDA) and linear discriminant analysis (3D-PCA-LDA) to evaluate FTIR hyperspectral images in normal vs melanoma samples[6].

II. Methods And Materials

For the purpose of acquiring spectrum images, we combined an Agilent Technologies (USA) Cary 660 FTIR spectrometer with a Cary 620 FTIR microscope of 5.5μm of spatial resolution and a focal plane array detector (FPA) of 32 by 32 components, device can acquire 1024 spectrum per acquisition. A fixed spectral resolution of $4cm^{-1}$ was used to acquire the spectra, which ranged from 3950 to 900 cm with scans of the background and sample adjusted to 128 and 64 scans respectively. All spectra was pre-processed and analysed using python with a lab-made routine.The pre-processing, stage where is possible eliminate unwanted artifacts like paraffin and water vapor, was performed following the steps smoothing using Savitzky Golay (SG) filter (window of points 15 and 2^{nd} polynomial), biofingerprint truncation, region which has the most biochemical information ($2000-900cm^{-1}$). Normalization the spectra and de-waxing by cut the region between ($1500-1400cm^{-1}$). after that, the quality test was performed using a T-squared test with the assistance of Partial Linear regression [5].

III. Computational modeling analysis

In this study we investigated the 3D discriminant analysis approaches i.e. 3D-PCA-LDA, and 3D-PCA-QDA with the intention to reduce the dimensions and to acquire the high efficiency of applied algorithms .

In order to evaluate our model perforamcne, we applied two important classifiers QDA and LDA algorithms combined with PCA to the mean scores (L_{st}) and (Q_{st}), as given below[6]:

$$L_{st} = (x_s - \bar{x_t})^T C_{pooled}^{-1}(x_s - \bar{x_t}) - 2log_e\pi_t \quad (1)$$

$$Q_{st} = (x_s - \bar{x_t})^T C_t^{-1}(x_s - \bar{x_t}) + log_e|C_t| - 2log_e\pi_t \quad (2)$$

where x_s represent $1 \times N$ is the mean scores of T for samples s and x_t is a row-vector $1 \times N$ presenting the mean scores of class t for the respective PCs. In addition, the variables like C_{pooled}^{-1} as follow.

$$C_{pooled} = \frac{1}{n}\sum_{t=1}^{T} n_t C_t \quad (3)$$

979-8-3503-0402-2/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on November 13,2023 at 18:35:43 UTC from IEEE Xplore. Restrictions apply.

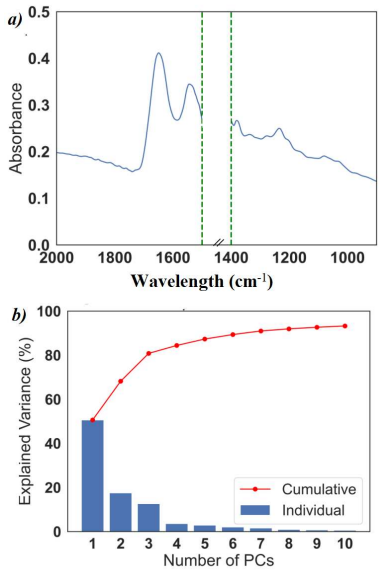

Fig. 1. (a)Spectrum after preprocessing that follow (SG, fingerprint truncation, removing paraffin band by cutting the spectrum ((1500 − 1400 cm $^{-1}$), (b) the explained variance ratio indicating a number of principal component analyses.

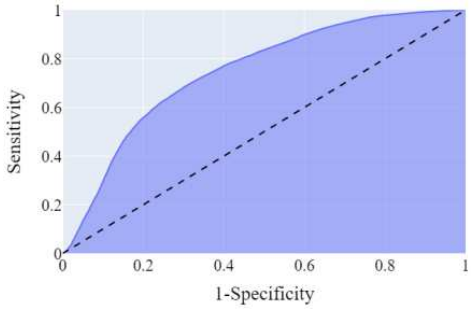

Fig. 2. A typical ROC of the independent test of model PCA-LDA classification.

TABLE I
THE MODELING PARAMETERS USING UNFOLDED (PCA-LDA, PCA-QDA) AND THREE-DIMENSIONAL ARRAYS (3D-PCA-LDA, 3D-PCA-QDA)

Data	Model	Accuracy
Unfolded	PCA-LDA	74
	PCA-QDA	75
3D	3D-PCA-LDA	79
	3D-PCA-QDA	81

The performance was determined by accuracy (ACC), as given following:

$$ACC(\%) = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \qquad (4)$$

IV. RESULTS AND DISCUSSION

In Fig. 1a is possible see the data after all step in the pre-processed procedure that are crucial to achieve better results. Fig.1b presents the unsupervised analysis using PCA, indicating the number of PCs vs. explained variance scores. One can see that first three scores of PC show the potential and significance for the computational analysis, having up to 80% of variance with three PCs.

As shown in Fig.2, it is depicted by a typical receiver operating characteristic curve (ROC) for the model classification of PCA-LDA where the area under the curve (AUC) represents the accuracy; that is (AUC = 74%). Table 1 depicts the classification performance achieved through the implementation of unfolded and 3D-discriminant analysis approaches. In the training set, the accuracy rates were observed to be 81% when employing 3D modeling algorithms (i.e. 3D-PCA-QDA), which outperforms the accuracy achieved by both unfolded procedures (PCA-LDA and PCA-QDA), which were recorded at 74% and 75% respectively. Our results show compelling evidence that the utilization of the 3D discriminant analysis approach significantly outperforms other methods i.e. conventional approaches, resulting in superior accuracy.

V. CONCLUSION

We showed a computational simulation approach based on 3D algorithms (3D-PCA-LDA and 3D-PCA-QDA) to evaluate the big data of cancer cells i.e. normal vs MM total number of spectra for MM is 93k and for Normal 49k. These 3D discriminant models are highly effective to explore hyperspectral images and discriminate the classes, comparing with unfolded approaches (PCA-LDA and PCA-QDA). The accuracy obtained is up to 80%, even for big data much more than 100k. These 3D discriminant analysis algorithms show fast class differentiation and present high performance compared to unfolded approaches.

ACKNOWLEDGMENT

This work was supported by CNPq (INCT-INTERAS 406761/2022-1), INCT-INFO (465763/2014-6); Sisfoton (440228/2021-2); PQ (314517/2021-9); CAPES Finance code 001 and FAPESP (17/50332-0) and FAPESP (21/00633-0)

VI. CONFLIT OF INTEREST

The authors declare no conflict of interest

REFERENCES

[1] WHO. World Health Organization.2022.URL:https://www.iarc.who.int/news-events/melanoma-awareness-month-2022.(accessed in 04/25/2023)
[2] Su KY, Lee WL. Fourier Transform Infrared Spectroscopy as a Cancer Screening and Diagnostic Tool: A Review and Prospects. Cancers 2020, 12, 115.
[3] Giuffrida R, et al. Use of noninvasive imaging in the management of skin cancer. Curr Opin Oncol. 2020;32,98-105.
[4] Johansen,TH et al. Recent advances in hyperspectral imaging for melanoma detection, wiley interdisciplinary Reviews: computational statistics 2020, 12 e1465
[5] Morais C. L, et al, A three-dimensional discriminant analysis approach for hyperspectral images, Analyst, 2020, 145, 5915
[6] Baker, M et al. Using Fourier transform IR spectroscopy to analyze biological materials. Nat Protoc 2014, 9, 1771.

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on November 13,2023 at 18:35:43 UTC from IEEE Xplore. Restrictions apply.