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/SBFOTONIOPC58971.2023.10230957

Monitoring changes in urine from diabetic rats using ATR-FTIR and Machine Learning

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

Daniella Lúmara Peres
Instituto de Pesquisas Energéticas e Nucleares (IPEN)
São Paulo, Brasil
daniellalumara@usp.br

Douglas Carvalho Caixeta
Universidade Federaql de Ubelndia (UFU)
Uberlandia - Brasil
caixetadoug@gmail.com

Cassio Lima
University of Liverpool
Liverpool - United Kingdom
Cassio.Lima@liverpool.ac.uk

Robinson Sabino da Silva
Universidade Federal de Uberlândia (UFU)
Uberlandia - Brasil
robinsonsabino@ufu.br

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

***Abstract*—Here, we aim to better characterize diabetes mellitus (DM) by analyzing 149 urine spectral samples, comprising of diabetes versus healthy control groups employing ATR-FTIR spectroscopy, combined with a 3D discriminant analysis machine learning approach. Our results depict that the model is highly precise with accuracy close to 100%.**

***Index Terms*—FTIR, machine learning, diabetes mellitus**

I. Introduction

Diabetes mellitus (DM) is, a chronic metabolic disorder, characterized by hyperglycemia as a result of impairment to the pancreatic β-cells' ability to secrete insulin. This disease affects the endocrine system and is expected to affect approximately 693 million adults by 2045, making it a major public health concern [1].

The chronic metabolic disorder because of DM is attributed to microvascular and macrovascular complications, including diabetic kidney disease, cardiovascular disease, and diabetic retinopathy, which can pose the risks of mortality, kidney failure, blindness, as well as a diminished quality of life for individuals living with diabetes [2]. Therefore, early diagnosis and control of diabetes are vital to prevent complications and maintain good health. Recently, Fourier Transform Infrared Spectroscopy (FTIR) is considered a highly successful approach for urine examination, and Attenuated Total Reflectance (ATR)-FTIR is especially beneficial for measuring several components of urine, including creatinine, urea, phosphate and sulfate, uric acid, pH, and cystinuria [3].

ATR-FTIR quantifies the infrared spectrum of the constituents absorbed at certain wavelengths, and this method has benefits such automated analysis, cost-effective, non-invasive nature, and no sample preparation [4]. The use of FTIR spectroscopy might result in a rapid and more accurate identification of diseases, regardless the fact that the procedures for measuring samples using the technique can be fatigue and computationally exhaustive. The precision of FTIR spectroscopy in disease detection is expected to upgrade with the continuous advancement of machine learning algorithms [5].

This study employed ATR-FTIR spectroscopy to measure urine samples and identify changes brought on by diabetes and non-diabetes in rats. We used a 3D discriminant analysis technique based on three-dimensional-Principal Component Analysis-Linear Discriminant Analysis (3D-PCA-LDA) to extract chemical signatures from biological fingerprints.

II. Methods and Materials

A. Sample Preparation

The study was managed on isogenic male Wistar rats without any other disease to induce diabetes. The rats were kept under standard conditions, and diabetes was induced by injecting them with streptozotocin. The diabetic rats were divided into placebo-treated, while non-diabetic rats received a placebo treatment. The experiment was conducted according to ethical guidelines and approved by an Ethics Committee.

B. ATR-FTIR analysis

The spectroscopic data was collected using an FTIR equipment in ATR mode, with 100 scans per spectrum and a spectral resolution of 4 cm^{-1}, covering a range of 4000-400 cm^{-1}. The spectra were pre-processed by smoothing, baseline correction, and vector normalization before data processing.

C. 3D discriminant modeling analysis

To evaluate the 3D-discriminant analysis approaches, namely 3D-PCA-LDA, we employed the linear discriminant analysis (LDA), on the mean-scores of 3D-PCA. The calculation of scores for 3D-PCA-LDA (L_{ij}) is as follows according to [6]:

$$L_{ij} = (x_i - \overline{x_j})^T C^{-1}_{pooled}(x_i - \overline{x_j}) - 2log_e\pi_j \qquad (1)$$

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

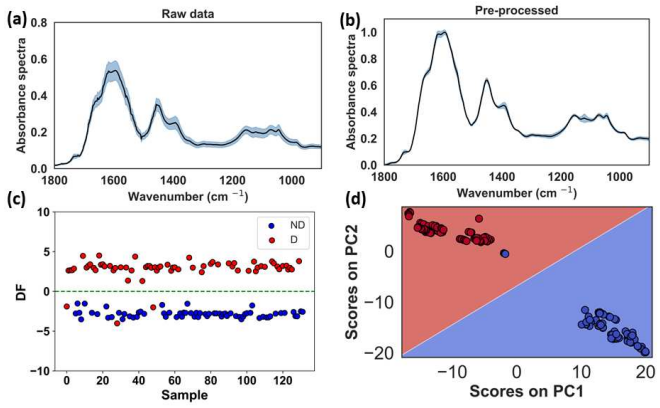

Fig. 1. (a) Mean spectra of raw data with standard deviation (b) Pre-processed data with standard deviation (c) Discriminant function (DF) plot (d) boundary decision plot

$$C_{pooled} = \frac{1}{n}\Sigma_{j=1}^{J} n_j C_j \tag{2}$$

$$C_j = \frac{1}{n_j - 1}\Sigma_{i=1}^{n_j}(x_i - \overline{x_j})^T \tag{3}$$

$$\pi_j = \frac{n_j}{n} \tag{4}$$

In order to explore the performance of our computational model, we use the accuracy parameter by determining the area under the curve (AUC) of receiver operating characteristics (ROC) curve, given as:

$$ACC(\%) = \left[\frac{TP + TN}{TP + FP + TN + FN}\right] \times 100 \tag{5}$$

where TN stands for true negative, TP for true positive, FN for false negative and FP for false positive.

III. Results and Discussion

A quality check was performed on a raw spectral data to identify outliers and anomalous spectra, furthermore we smoothed raw spectra (Fig. 1a), with Savitzky-Golay (SG) filter (window = 11, poly =2, deriv = 1), in the fingerprint region ranging from 900 to 1800 cm^{-1}. Baseline correction and normalization were performed in order to achieve the pre-processed data for ML simulations, as shown in Fig.1b. The parameters employed to evaluate the performance were selected optimal during this study. One can observe that the pre-processing procedure reduced the outliers, eliminating background noise before employing for computational simulations. Moreover, we applied SMOTE-TOMEK technique to transform imbalanced to balanced data to confirm the model accurate precision. For 3D-PCA, we selected three optimal parameters of PCs to achieve higher accuracy using AUC applying ROC curve.

Furthermore, the performance of the proposed model was evaluated in order to monitor diabetes and non-diabetes samples. Our study was used to analyse glucose, creatinine and urea in urine [8]. The results are depicted in Fig.1c, which displays the discriminant function (DF) based on the score plot of 3D-PCA-LDA. Our analysis shows that the proposed method was able to correctly identify diabetes and non-diabetes samples up to 99% accuracy.

Moreover, Fig.1d shows the efficiency of the proposed method through the use of a decision boundary plot. In particular, the first two principal components (PCs), which together account for more than 90% of the relevant information in the variance of the data, were chosen. The plot highlights the potential of 3D-PCA-LDA as an outstanding analytical tool for diabetes diagnosis and shows how well the suggested technique performs in properly categorizing data. One can see the strong separation of diabetes and ND samples, evaluating accuracy using 3D discriminant analysis. These results demonstrate the potential of the suggested strategy to increase the precision and effectiveness of diabetes diagnosis.

IV. Conclusion

In conclusion, we used a computational approach based on 3D-discriminant analysis (3D-PCA-LDA) and ATR-FTIR to monitor diabetes and healthy urine groups. Our results demonstrated that the computational model achieved accuracy of up to 99% and depicted interest to use in diagnosing diabetes.

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).

References

[1] Cole, J B and Florez, J C, "Genetics of diabetes mellitus and diabetes complications". Nature reviews nephrology, vol. 16, pp.377–390, 2020.
[2] Ogurtsova K, Guariguata L, Barengo, N. C., Ruiz,P. L. D, Sacre,J. W, Karuranga, S, Sun, H, Boyko, E. J. and Magliano D. J, "Idf diabetes atlas: Global estimates of undiagnosed diabetes in adults for 2021". Diabetes research clinical practice, vol 183, 2022.
[3] Sarigul N, Kurultak İ, Uslu Gökceoğlu A, and Korkmaz F, "Urine analysis using ftir spectroscopy: A study on healthy adults and children," J. Biophotonics, vol 14, 2021.
[4] Ergün E, Kart S, Zeybek D. K, and Zeybek B, "Simultaneous electrochemical determination of ascorbic acid and uric acid using poly (glyoxal-bis (2-hydroxyanil)) modified glassy carbon electrode", Sensors Actuators B: Chem., vol. 224, pp. 55–64, 2016.
[5] Bernardes-Oliveira E, Freitas, D. L. D. de, Morais,C. d. L. M. de , Cornetta, M. d. C. d. M. , Camargo J. D. d. A. S, Lima, K. M. G. de, and Crispim, J. C. d. O., "Spectrochemical differentiation in gestational diabetes mellitus based on attenuated total reflection fourier-transform infrared (atr-ftir) spectroscopy and multivariate analysis," Sci. Reports321, vol.10, 2020.
[6] Morais, Camilo LM, et al. "A three-dimensional discriminant analysis approach for hyperspectral images". Analyst, vol. 145.17, pp. 5915-5924, 2020.
[7] Farooq, Sajid and Del-Valle, Matheus and dos Santos, Moises Oliveira and dos Santos, Sofia Nascimento and Bernardes, Emerson Soares and Zezell, Denise Maria, "Rapid identification of breast cancer subtypes using micro-FTIR and machine learning methods", Applied Optics, vol. 62, pp. C80–C87, 2023.
[8] Caixeta, Douglas Carvalho and Lima, Cassio and Xu, Yun and Guevara-Vega, Marco and Espindola, Foued Salmen and Goodacre, Royston and Zezell, Denise Maria and Sabino-Silva, Robinson, "Monitoring glucose levels in urine using FTIR spectroscopy combined with univariate and multivariate statistical methods",Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 290, 2023.