Streamlining Bacterial Infection Diagnosis: Rapid Gram Classification Using FTIR Spectroscopy

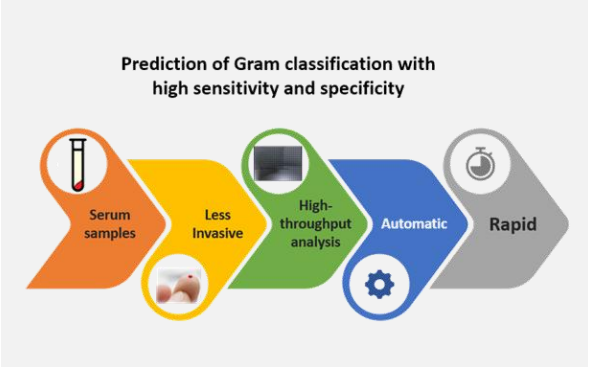
POSTER

(Topic: Infections)

R. Araújo[1,2,], L. Ramalhete[2,3], T. Fonseca[4], C. von Rekowski[4], L. Bento[1,2,5], C. Calado[4,6]*

[1] CHRC, R. Câmara Pestana 6, Lisbon, Portugal; [2] NMS, R. Campo dos Mártires da Pátria 130, Lisbon, Portugal, [3] IPST, Alameda das Linhas de Torres, Lisbon, Portugal,[4] ISEL, R. Conselheiro Emídio Navarro 1, Lisbon, Portugal, [5] CHULC, R. José António Serrano, Lisbon, Portugal,[6] CIMOSM, R. Conselheiro Emídio Navarro 1, Lisbon, Portugal

***Corresponding author:** rubenalexandredinisaraujo@gmail.com*



Short abstract: In a hospital setting, diagnosing infections typically involves a complex process that includes the collection of biological samples and growing a culture for organism isolation, followed by its characterization. However, these methods are slow, require multiple steps and are often limited by the need of specialized equipment and skilled personnel. In this preliminary study, it was analysed the serum, by FTIR spectroscopy, of 29 critically ill COVID-19 patients in an ICU. It was analysed the effect of varied pre-processing methods and spectral sub-regions on t-SNE. Through the optimization of SVM models, it was possible to achieve a very good gram predictive model with a sensitivity and specificity of 90 and 89% respectively. As an accurate classification of bacterial strains is crucial to guide effective antimicrobial therapy and prevent the spread of multidrug-resistant bacteria, FTIR spectra, acquired in a simple, economic, and rapid mode, presents therefore the potential for development of new classification methods that would greatly enhance the ability to manage bacterial infections.

Introduction

Gram classification is of the utmost importance in hospitals, in order to properly diagnose bacterial infections and determine the appropriate treatment [1]. However, correctly identifying the Gram classification of a bacterial infection can be challenging due to potential misinterpretation of staining results, overlapping cell morphology, and emerging antibiotic-resistant bacteria [2]. Therefore, accurate and timely Gram classification is crucial for guiding appropriate antibiotic therapy, preventing infection spread and to help minimize antibiotic resistance [3]. These challenges only serve to highlight the importance of rapid, economic, and reliable diagnostic methods in a clinic or hospital environment, a critical step to effectively treat and prevent serious complications and life-threatening events for the patients. Spectroscopy [4], especially FTIR spectroscopy present itself as a technique that can enable this goal, as it is can acquire the metabolic status of the biological system in a high sensitivity and specificity mode. Indeed, FTIR spectroscopy, especially in the mid-infrared region (MIR), from 400 to 4000cm^{-1}, has been widely used in biomedicine applications, from discrimination between B and T-lymphocytes [5], infection processes [6], capturing the human physiological state through serum and plasma analysis [7], as well as for medical diagnosis, prognosis [8], among many others.

Objectives

This work aims to evaluate if Fourier Transform Infrared (FTIR) spectroscopic analysis of human serum, would allow for the prediction and correct gram classification in critically ill patients in an ICU environment.

Methods

A. Biological Assay

Peripheral blood was collected in a serum tube with no anticoagulant VACUETTE®, using standard blood collection procedures. Samples were maintained at 4ºC until blood centrifugation at 3000 rpm for 10 minutes (Micro 220T, Hettich, Tuttlingen, Germany). Serum samples were kept at -20ºC until FTIR spectra acquisition. A total of 29 patients, with COVID-19 and bacteraemia, and admitted to the ICU of *Hospital São José, Centro Hospital Universitário Lisboa Central*, were considered. All participants provided a signed informed consent before enrolment in the study approved by the Hospital's Ethics Committee.

B. FTIR spectra acquisition

Triplicates of 25 µL of serum diluted at 1/10 in water were transferred to a 96-wells Si plate and then dehydrated for about 2.5 h, in a desiccator under vacuum (ME 2 pump, Vacuubrand, Wertheim, Germany). Spectral data was collected using a FTIR spectrometer (Vertex 70, Bruker, Germany) equipped with an HTS-XT (Bruker, Germany,) accessory. Each spectrum represented 64 coadded scans, with a 2cm^{-1} resolution, and was collected in transmission mode, between 400 and 4000 cm^{-1}. The first well of the 96-wells plate did not contain a sample and the corresponding spectra was acquired and used as background. Medians of triplicate spectra were used.

C. Spectra pre-processing and processing

All spectra were submitted to atmospheric correction, using OPUS® software, version 6.5 (Bruker, Germany, Billerica, USA). A baseline correction was also performed. Second derivative spectra (based on a Savitzky-Golay filter, and a 2nd order polynomial over a 15-point window) and unit vector normalization and spectra processing were conducted by Orange 3 Data Mining Toolbox (Faculty of Computer and Information Science, University of Ljubljana, Slovenia. Spectra processing included t-distributed stochastic neighbour embedding (t-SNE) and Support Vector Machine (SVM) models optimization.

Results

A total of 29 COVID-19 patients, all hospitalized at the ICU, were considered, all presenting bacteremia, as based on microbiological analysis. Patients, between the two groups, i.e., gram positive and gram negative, did not present significant differences concerning gender, age, and body mass index ($p>0.1$).

Among the many pre-processing tested (Table 1), from non-derivative to derivative spectra (data not shown), it was the atmospheric correction spectra with a normalized (unit vector normalization) second derivative (Table 1D, Figure 1), which

proved to be the most effective at separating the two groups of patients. This was possible due to the minimization effect that normalization has on the spectra, regarding the impact of sample quantity under analysis. Also, the second derivative contributes to a superior differentiation of groups, as it resolves superimposed bands, therefore increasing the information retrieved from the spectra. As derivatives also increase noise, care must be taken however, as any region that might decrease the signal-to-noise ratio should, therefore, be removed.

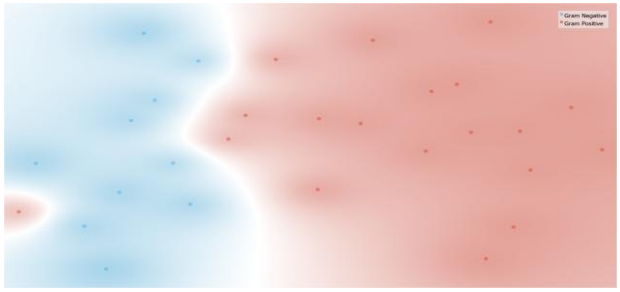


Figure 1. t-SNE of patients with gram negative (blue) and gram positive (red) bacteria, based on serum spectra after a normalized second derivative between 1040-1045 cm^{-1}, 1345-1350 cm^{-1}, 2190 to 2195 cm^{-1} and 2965-2970 cm^{-1}.

Having identified the best pre-processing through means of the spectra t-SNE with the best scores separation between the two groups of patients, SVM models were developed to predict Gram classification, based on non-derivative spectra and second derivative spectra, based on the whole spectra, or select sub-regions. For that, 80% of patient's data were used for model training and the remaining 20% used for an independent dataset for model validation. A total of 100 models were built for each spectra pre-processing model and regions evaluated, based on random selection of data for the model's training and validation. The various model's performance, for each pre-processing, can be observed in Table 1.

Table 1. SVM models' performance to predict Gram classification on spectra from serum of 29 patients with atmospheric and baseline correction (A), complete spectra normalized second derivative (B), and select sub-regions (C), normalized second derivative between 1040-1045 cm^{-1}, 1345-1350 cm^{-1}, 2190 to 2195 cm^{-1} and 2965-2970 cm^{-1} (D), and normalized second derivative between 600-1800 and 2800-3100 cm^{-1} (E) as well as select sub-regions (F).

	AUC	CA	Precision	Sensitivity	Specificity
A	0.487	0.643	0.460	0.643	0.324
B	0.460	0.662	0.443	0.662	0.331
C	0.950	0.892	0.892	0.892	0.868
D	**0.969**	**0.897**	**0.898**	**0.897**	**0.883**
E	0.906	0.797	0.797	0.797	0.658
F	0.904	0.880	0.879	0.880	0.815

The best predictive model was achieved based on the spectra with atmospheric correct and normalized second derivative, on the spectral sub-regions 1040-1045 cm^{-1}, 1345-1350 cm^{-1}, 2190 to 2195 cm^{-1} and 2965-2970 cm^{-1} (Table 1D), highlighting the relevance of the regions ≈1045 cm^{-1}, to the stretching vibrations of the C-O bond in cyclic ethers such as tetrahydrofuran and oxirane, ≈1350 cm^{-1} associated with the bending vibrations of the C-H bond in the CH3 group, ≈2195 cm^{-1}, associated with the stretching vibrations of the C≡N bond in nitriles. This absorption is typically very strong and narrow and can be used to identify nitrile functional groups in various molecules. The ≈2970 cm^{-1} region associated with the stretching vibrations of the C-H bond in methyl (CH3) groups. This absorption can be used to identify the presence of methyl groups in various molecules, including lipids and other organic compounds. The relevance of these regions, cannot be overstated, for example, changes in the position or intensity of the absorption bands at 1030-1070 cm^{-1} can be indicative of changes in the molecular composition of lipids and carbohydrates, which are common in many infectious agents [9]. These regions relevance is highlighted by the very good prediction model developed (Table 1D), which resulted in a sensitivity and specificity of 90% and 89%, respectively.

Conclusion

The present work points to an alternative technique to classify, using FTIR spectroscopy, between positive and negative Gram bacteria in the blood stream, through a simple, rapid, and economic mode, in ICU patients. This allows adequate treatment to critically ill patients, leading to fast infection detection rates, and individually concocted course of antibiotics, at the same time avoid its unnecessary use, which ultimately would ease the metabolic burden on the patients.

Acknowledgements

This work was supported by the project grant DSAIPA/DS/0117/2020 supported by *Fundação para a Ciência e a Tecnologia*, Portugal and by the project grant NeproMD/ISEL/2020 financed by *Instituto Politécnico de Lisboa*.

References

[1] N. Tripathi and A. Sapra, *Gram Staining*. Treasure Island (FL), 2022. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/14925025

[2] L. P. Samuel, J.-M. Balada-Llasat, A. Harrington, and R. Cavagnolo, "Multicenter Assessment of Gram Stain Error Rates," *J. Clin. Microbiol.*, vol. 54, no. 6, pp. 1442–1447, Jun. 2016, doi: 10.1128/JCM.03066-15.

[3] S. Vasoo, J. N. Barreto, and P. K. Tosh, "Emerging Issues in Gram-Negative Bacterial Resistance," *Mayo Clin. Proc.*, vol. 90, no. 3, pp. 395–403, Mar. 2015, doi: 10.1016/j.mayocp.2014.12.002.

[4] S. Berezin, Y. Aviv, H. Aviv, E. Goldberg, and Y. R. Tischler, "Replacing a Century Old Technique – Modern Spectroscopy Can Supplant Gram Staining," *Sci. Rep.*, vol. 7, no. 1, p. 3810, Jun. 2017, doi: 10.1038/s41598-017-02212-2.

[5] L. Ramalhete, R. Araújo, and C. R. C. Calado, "Discriminating B and T-lymphocyte from its molecular profile acquired in a label-free and high-throughput method," *Vib. Spectrosc.*, vol. 111, p. 103177, Nov. 2020, doi: 10.1016/j.vibspec.2020.103177.

[6] R. Araújo, L. F. N. Bento, T. A. H. Fonseca, C. P. Von Rekowski, B. R. da Cunha, and C. R. C. Calado, "Infection Biomarkers Based on Metabolomics," *Metabolites*, vol. 12, no. 2, p. 92, Jan. 2022, doi: 10.3390/metabo12020092.

[7] R. Araújo, L. Ramalhete, E. Ribeiro, and C. Calado, "Plasma versus Serum Analysis by FTIR Spectroscopy to Capture the Human Physiological State," *BioTech*, vol. 11, no. 4, p. 56, Dec. 2022, doi: 10.3390/biotech11040056.

[8] L. M. Ramalhete, R. Araújo, A. Ferreira, and C. R. C. Calado, "Proteomics for Biomarker Discovery for Diagnosis and Prognosis of Kidney Transplantation Rejection," *Proteomes*, vol. 10, no. 3, p. 24, Jul. 2022, doi: 10.3390/proteomes10030024.

[9] G. Larrouy-Maumus, "Lipids as Biomarkers of Cancer and Bacterial Infections," *Curr. Med. Chem.*, vol. 26, no. 11, pp. 1924–1932, Jun. 2019, doi: 10.2174/0929867325666180904120029.