UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENICAS QUÍMICAS

Departamento Química Analítica

TESIS DOCTORAL

Identification and discrimination of microorganisms by Laser Induced Breakdown Spectroscopy (LIBS) and mathematical algorithms

Identificación y discriminación de microorganismos mediante espectroscopía láser (LIBS) y algoritmos matemáticos

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Sadia Manzoor

Director

Jorge Cáceres Gianni

Madrid, 2018

UNIVERSIDAD
COMPLUTENSE
MADRID

Department of Analytical Chemistry

Identification and Discrimination of Microorganisms by Laser Induced Breakdown Spectroscopy (LIBS) and Mathematical Algorithms

Identificación y Discriminación de Microorganismos Mediante Espectroscopia Laser (LIBS) y Algoritmos Matemáticos

Director
Jorge Cáceres Gianni

Química Láser
Dpto Química Analítica

Tesis Doctoral

Sadia Manzoor

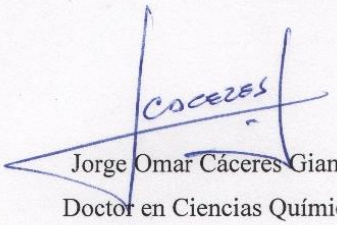
Madrid 2017

Certificado

Jorge Omar Cáceres Gianni, Director del Grupo de Investigación de Química Laser, Doctor en Ciencias Químicas, Profesor Titular del Departamento de Química Analítica de la Facultad de Ciencias Químicas de la Universidad Complutense de Madrid.

CERTIFICA:

Que el presente trabajo titulado "Identificación y Discriminación de Microorganismos Mediante Espectroscopia Laser (LIBS) y Algoritmos Matemáticos" ha sido realizado por Sadia Manzoor en el laboratorio de Química láser del Departamento de Química Analítica de la Universidad Complutense de Madrid bajo mi dirección. Las publicaciones que lo avalan no han sido utilizadas en tesis anteriores, reuniendo a mi juicio los requisitos necesarios para optar al grado de Doctor por la Universidad Complutense de Madrid, por lo que autorizo su presentación.



Jorge Omar Cáceres Gianni
Doctor en Ciencias Químicas
Director de Grupo de Química Láser
Universidad Complutense de Madrid

Madrid, 06 de Febrero de 2017

To my Parents, Siblings,
Teachers,
And a special mention to Teresa Ugidos Lozano

..... without their support I could never have reached this far

Acknowledgements

The time which brought me to Spain five years ago, has finally come. During this long journey, there have been so many people who came in my life and went away but then there are those who left very prominent marks in my research and personal life. To begin with, the most important of all is my thesis supervisor, Prof. Jorge Caceres. I would like to thank him for his immense support and guidance from the beginning of my stay in the lab. In addition, offering the possibility to keep progressing and working on my PhD thesis under several circumstances. I cannot thank him enough for guiding me, not only through my research work but helping me go through every thick and thin during the last years. I would also like to thank worthy members of our research group, Prof. Roberto Izquierdo and Prof. Daniel Rosales, along with Prof. Fernando Navarro for their contribution in my learning process. The role of Professor Juan Alfonso Ayala has been pivotal on the microbiological aspect of the work. To my colleagues, Samuel Moncayo, Alicia Marin and Lydia Ugena, with whom I shared most of my time at the lab. Carmen Rico, Jesus Sanchez Flores, Eirini Stamati and Elena Anastasiou who spent short term research stays in the group for all the discussions about the projects, conversations and fun times. I feel very fortunate to have known some wonderful people who were a continuous support and motivation to me. Here I would like to mention the person who played the most important role to get me to this point, Teresa Ugidos, who made me a member of her family and was the strong support that lent me a back whenever I needed it. Away from my family; Teresa, Gonzalo and Miriam became my Spanish family and I can never forget the fun I have had with them. It does not stop here and I would like to thank Jorge´s family, Roxana Ramondelli, Matias Caceres and Lucas Caceres for their supportive role and listening to me when I needed them. Thanks to all my wonderful friends, especially Shukriya, Kareemah, Mayte, Mariluz, Ana, Aurelis, Ioanna, Sepideh and Enas for being the fun part of my life. Back home in Pakistan; my great and lovely parents (Manzoor Hussain and Balquees Begum), sisters (Suryyia, Shamaila and Fareeha) and my brother Ahmad Mustafa who were always present to offer advice and motivate me to keep going on. They were always ready to give me moral support, and encouraged me whenever I felt down. They gave me immense support to stand back and face all the hurdles, difficulties and obstacles that came my way. It is because of them; I stood up again and again and went through every problem with my chin up in the air. There are so many people I need to thank here but time, limited space and humility make me stop here.

TABLE OF CONTENTS

RESUMEN

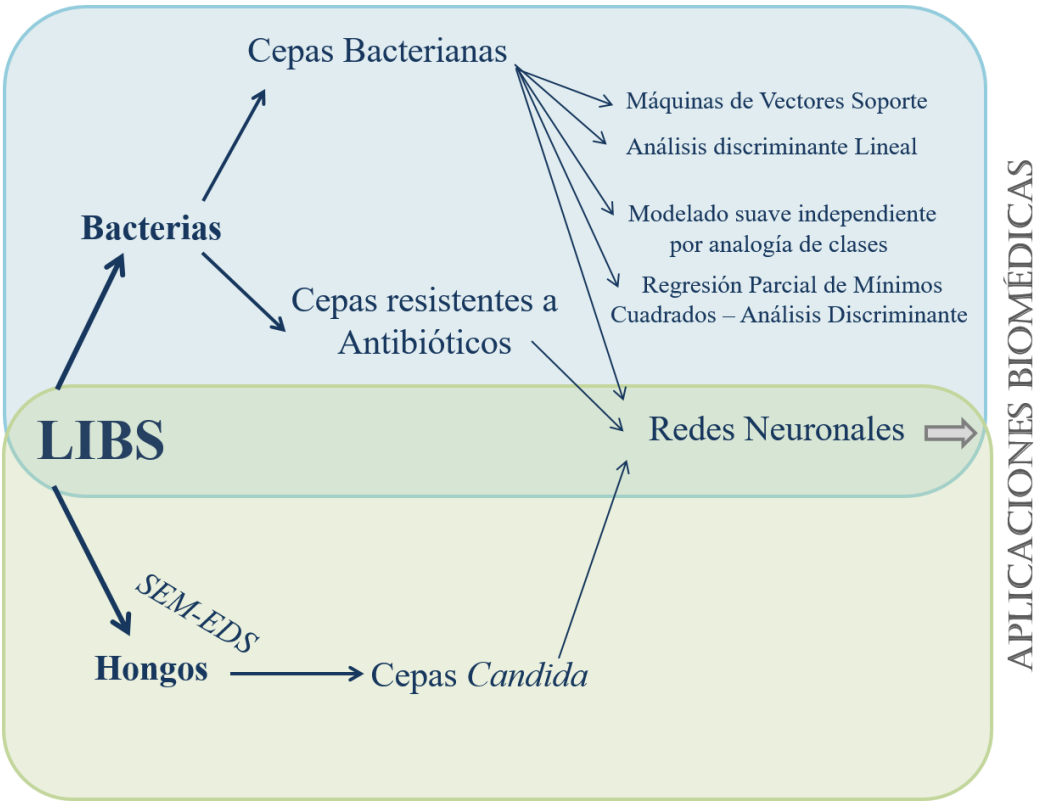
La espectroscopia de ablación láser es una técnica espectroscópica de emisión óptica que consiste en irradiar una parte de la muestra con un pulso láser de alta potencia, formando un plasma en la superficie de la muestra mediante un fenómeno denominado descomposición. El análisis espectroscópico de las emisiones del plasma permite el análisis de la composición elemental de la muestra. La adquisición instantánea de espectros ofrece la posibilidad de obtener información sobre la composición de un material mediante irradiación directa con pulsos láser, tratamiento mínimo de la muestra, simplicidad y versatilidad, análisis en cualquier estado de la materia en diferentes condiciones y distancias ambientales, destrucción mínima de la muestra, flexibilidad en la configuración experimental, portabilidad de la instrumentación y la rentabilidad de los equipos, así como los análisis, debido a que no requieren reactivos, proporcionan una excelente combinación de características de LIBS, lo que lo convierte en una gran herramienta analítica para el análisis cualitativo y cuantitativo. Durante las últimas décadas, no solo se ha incrementado el uso de esta técnica en los laboratorios, sino que también ha conducido a su transformación en una técnica portable en el campo. Debido a estas características intrínsecas de la técnica LIBS y a las ventajas que ofrece, los investigadores han estado trabajando continuamente para evaluar la capacidad de esta técnica en nuevos aspectos para su aplicación en diversos campos de la ciencia.

En la primera fase, la técnica LIBS, se limitó principalmente a los estudios relacionados con el análisis de materiales, el monitoreo ambiental y el control de procesos. El interés en este campo ha atraído a los investigadores a utilizarlo tanto en el estudio de muestras tan pequeñas como bacterias como en otras a mayor escala. Aunque

el principio básico de LIBS es el mismo, no existe una única fórmula para llevar a cabo el análisis LIBS de muestras muy diferentes, por lo tanto, dependiendo de la muestra, es necesario optimizar los parámetros experimentales que afectan a la interacción láser - materia y a los métodos de extracción de información de las mediciones espectrales obtenidas.

El plasma es el elemento básico de una medición LIBS, que se genera debido a la ablación del material, y las emisiones producidas por su enfriamiento proporcionan la información sobre los elementos constitutivos del material ablacionado. Las líneas de emisión, con sus intensidades situadas en longitudes de onda específicas, dan lugar a una característica "huella digital espectral" de la muestra en análisis, que permite realizar un análisis cualitativo o cuantitativo de la muestra. Un subconjunto de análisis cualitativo es la identificación material que incluye la creación de modelos de clasificación utilizando diversas técnicas quimiométricas para determinar la identidad o pertinencia de una muestra dentro de un cierto conjunto limitado de materiales de referencia.

En esta tesis se ha estudiado el enfoque cualitativo de la técnica LIBS en el área de la biomedicina, explorando su potencial para identificar y discriminar muestras biológicas, a partir de los cambios en su composición atómica. La motivación de estos estudios fue evaluar la capacidad de LIBS para proporcionar una identificación rápida en comparación con los métodos bioanalíticos tradicionales, aprovechando la posibilidad de combinarlo con los métodos quimiométricos para aumentar el rendimiento de la técnica y demostrar su potencial de uso como un método diagnóstico en el ámbito clínico. En la figura 1.1 se presenta un esquema que resume el trabajo presentado en esta tesis.



Esquema de la tesis

Con el fin de desarrollar los modelos de clasificación, se emplearon diferentes enfoques de los métodos quimiométricos y se compararon, para encontrar el mejor enfoque para proporcionar la solución a este problema. Una parte integral de los estudios es el desarrollo de modelos de clasificación utilizando los algoritmos de Red Neural Artificial (NN) como la herramienta quimiométrica para el análisis de datos espectrales LIBS en la identificación y discriminación de materiales biológicos moleculares complejos como bacterias y hongos. La selección de la NN fue fomentada por un estudio en el que la comparación entre varios métodos quimiométricos, incluyendo Análisis Discriminante Lineal (LDA), Máquinas de Soporte Vectorial (SVM), Modelado Suave Independiente de Analogía de Clase (SIMCA), Análisis Discrecional de mínimos cuadrados parciales (PLS-DA) y Redes Neuronales Artificiales (NN), demostró el mejor desempeño de la NN en la clasificación de cepas bacterianas.

Hay una multitud de técnicas que se emplean para identificar los patógenos bacterianos y fúngicos presentes en una muestra biológica que causa enfermedades infecciosas. Todas estas técnicas ofrecen varias ventajas y proporcionan buenos resultados, pero en muchos casos los factores como el tiempo y el coste del análisis son limitados. Por lo tanto, en esta tesis se pretendió desarrollar una metodología basada en la combinación LIBS-NN para realizar la identificación y discriminación de estos patógenos en la identificación y discriminación de bacterias y hongos en muestras biológicas; con especial referencia a la mejora de los estándares de seguridad hospitalaria, particularmente desde el punto de vista microbiológico, mediante el diagnóstico precoz de las infecciones adquiridas en el hospital (HAI) y su uso en las aplicaciones clínicas en el diagnóstico de enfermedades resultantes de infecciones bacterianas y fúngicas.

La tesis se divide principalmente en tres partes: Introducción a los fundamentos de las técnicas utilizadas, procedimiento Experimental y Resultados.

La primera parte se centra en la introducción al objetivo principal de esta tesis, es decir, el análisis de muestras de microorganismos utilizando la técnica LIBS junto con el estado del arte de la técnica LIBS en el análisis de las muestras de microorganismos. También cubre la teoria fundamental de la técnica de LIBS junto con un acercamiento teórico detallado a la quimiometría en LIBS, particularmente las técnicas.

- El capítulo 1 da una visión teórica de la técnica incluyendo los fundamentos de la técnica LIBS, instrumentación y configuración comúnmente utilizados para los análisis. Además, este capítulo abarca los antecedentes del análisis basado en LIBS de microorganismos que discute el principio de la técnica, la importancia de las muestras bacterianas

estudiadas hasta el momento y los avances realizados en esta técnica para sus aplicaciones médicas.

- El capítulo 2 incluye una introducción general a la quimiometría y las figuras de mérito que se necesitan tener en cuenta para la evaluación del desempeño de los métodos quimiométricos en las tareas de clasificación. Específicamente, las técnicas incluyen métodos lineales supervisados tales como SIMCA, PLS-DA, LDA y métodos supervisados no lineales tales como SVM y NN, discutidos desde un punto de vista teórico.
- En el capítulo 3 se describe la configuración experimental utilizada para las mediciones LIBS y los componentes del sistema experimental desarrollado en el laboratorio. También trata las muestras bacterianas incluidas en el estudio y la preparación de estas muestras. Además, se dan los detalles sobre los programas utilizados para el desarrollo de los modelos de clasificación.

La tercera parte de la tesis incluye los resultados experimentales y las conclusiones obtenidas en el estudio.

- El capítulo 4 trata sobre la metodología desarrollada de LIBS-NN para la identificación y discriminación de muestras bacterianas. Esta parte se divide además en tres secciones.

1. El primer estudio se basa en la aplicación de LIBS y NN para el análisis de cepas bacterianas, estudiando por un lado la

combinación de estas técnicas para llegar a la identificación y discriminación de estas muestras de microorganismos, mientras que por el otro se estudia el efecto del medio de crecimiento sobre el análisis y los resultados.

2. La metodología basada en LIBS-NN se extendió al análisis de cepas bacterianas resistentes a los antibióticos, que forman la segunda sub-sección.
3. También se realizó una comparación entre las prestaciones de los diferentes métodos quimiométricos, tal como se menciona en la Figura 1.1, aunque realizadas al principio, formando la última parte de este capítulo. Esta comparación constituyó la base para continuar con el uso de NN para generar los modelos de clasificación.

- El capítulo 5 presenta el análisis de muestras de hongos, específicamente cepas de *Cándida,* por LIBS con su posterior caracterización por LIBS y SEM-EDS y aplicando la metodología LIBS-NN para su discriminación. Primer análisis centrado completamente en hongos y cándida.

- Finalmente, la última parte del capítulo 6 discute las conclusiones sobre el trabajo en esta tesis y discute los desafíos que enfrenta la técnica debido a su versatilidad en su aplicación e instrumentación.

Al final de la tesis, se han añadido tres anexos que incluyen:

- Interfaz de usuario desarrollada para el análisis de datos LIBS.
- Publicaciones en revistas científicas, incluyendo todos los proyectos realizados durante el desarrollo de esta tesis.
- Finalmente, el trabajo de investigación consistente en los resultados iniciales obtenidos de la clasificación de diferentes cepas bacterianas de la misma especie publicado como en proceedings de Congreso "BioMicroWorld2013".

OBJETIVOS

El objetivo general de esta tesis fue desarrollar un conjunto de herramientas analíticas que permitiesen un análisis simple, económico, fiable y preciso para identificar y discriminar los microorganismos patógenos y la extrapolación de estos métodos a la detección de patógenos en muestras de origen biológico, y evaluar la potencial de la técnica LIBS en combinación con el análisis de datos utilizando algoritmos matemáticos de correlación para su aplicación a nivel clínico.

Específicamente, los objetivos que se persiguen se resumen a continuación:

1. Diseño de protocolo para el análisis LIBS de las muestras de microorganismos incluidos en el estudio. Para ello, se evaluaron las variables y condiciones de medición, optimización de parámetros experimentales, preparación de muestras para el análisis LIBS y evaluación de la detección directa de cultivos sin tratamiento de muestra.
2. Estudiar el efecto del medio para cultivar microorganismos en la identificación basada en LIBS, dada la necesidad de aislar bacterias de

muestras biológicas antes del análisis LIBS debido al efecto matriz y la importancia de cultivar microorganismos como método básico en microbiología. Este estudio se realizó para averiguar si el medio de cultivo influye en la composición elemental de las células y afecta a la identificación de microorganismos por LIBS.

3. Desarrollo de modelos de clasificación basados en el estudio de métodos matemáticos de análisis de datos. Esto incluye la comparación de modelos quimiométricos controlados lineales y no lineales, y procedimientos de correlación para determinar el método que proporciona mejores resultados de clasificación. También se pretendió evaluar el poder de clasificación de las redes neuronales en combinación con LIBS y el desarrollo de una interfaz para el usuario que permite el control en tiempo real de los parámetros matemáticos NN de los modelos utilizados.
4. Evaluación de la metodología LIBS-NN para discriminar e identificar diferentes bacterias y hongos patógenos sobre la base de su composición atómica con el fin de desarrollar una metodología rápida y sencilla para aplicaciones biomédicas. Por un lado, se pretendía lograr la identificación y discriminación de microorganismos, por otro, el estudio pretendía ampliar la escala de estudio aumentando el número de especies como enterobacterias, bacterias gram-positivas y gram-negativas y también diferentes cepas de estas bacterias y hongos.
5. Determinación de la resistencia a antibióticos en cepas bacterianas más comúnmente aisladas. Incluye la evaluación de la capacidad de LIBS-NN para clasificar cepas bacterianas basadas en su huella espectral característica detectando las diferencias de composición debidas a mutaciones resultantes.

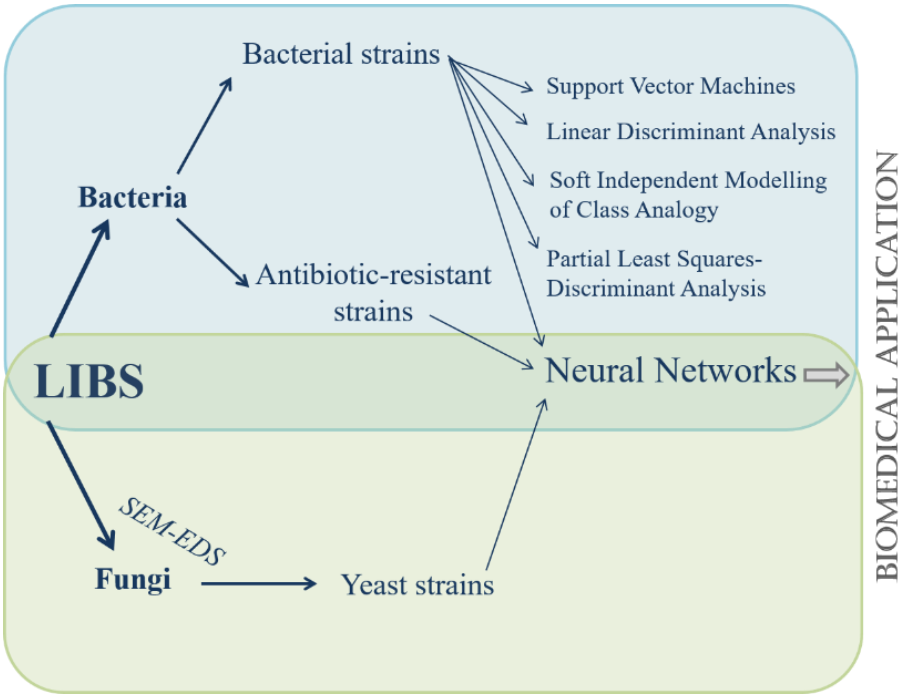
SUMMARY

Laser Induced Breakdown Spectroscopy (LIBS) is an optical emission spectroscopic technique that is based on irradiating a portion of the sample with a short and high power laser pulse, forming the plasma on the sample´s surface by a phenomenon called breakdown. The emissions from this plasma are analysed spectroscopically providing a spectrum which allows obtaining the compositional information of the sample. Instant spectra acquisition, possibility to directly irradiate the sample with laser pulses in any state of matter under different ambient conditions and variable distances, minimal sample treatment, simplicity and versatility, minimal or negligible destruction of the sample, flexibilty in experimental configuration, portability of the set-up, and cost-effectiveness of both the equipment and analyses, obviating reagents and cumbersome sample preparation steps, provide an excellent combination of features to LIBS as an analytical tool for qualitative and quantitative analysis for a wide range of samples. Last two decades have seen a rapid increase in the use of this technique in different fields, that has not been restricted only to the laboratory experiments but has led to its maturation and transformation into a field-deployable technique [1-3]. The excellent features and advantages offered by LIBS have motivated the researchers to work continuously evaluating the ability of this technique in newer aspects finding its application in various fields of science. This has resulted using this technique, only limited to material analysis, environmental monitoring and process control during the earlier phase of its use to the study of samples as small as bacteria, fungi and animal tissues.

Although the basic working principle of LIBS is same but there is not a single formula to carry out LIBS analysis of very different samples, therefore, depending on the sample, there is a need to optimize experimental parameters affecting laser - matter

interaction and the methods of extracting information from the spectral measurements obtained. Plasma is the basic element of a LIBS measurement. Emissions produced from cooling of plasma provide the information about constituent elements of the ablated material. Emissions lines with their intensities located at specific wavelengths give rise to a "spectral fingerprint" characteristic of the sample under analysis, which then allows carrying out qualitative or quantitative analysis on the sample. A subset of qualitative analysis is material identification that includes creating classification models using various chemometric techniques to determine the identity or pertinence of a sample within a certain limited set of reference materials.

In this thesis, qualitative approach of LIBS technique has been studied in the area of biomedicine, exploring its potential to identify and discriminate biological samples, based on the changes in their atomic composition. A scheme is shown in Fig. 1.1 summing up the work presented in this thesis.



Scheme of the thesis

The motivation of these studies was evaluating the ability of LIBS in order to provide fast identification of bacterial pathogens as compared to traditional bioanalytical methods, benefitting from the possibility to combine it with chemometric methods in order to increase the performance of the technique and demonstrate its potential for use as a diagnostic method in clinical set ups.

In order to develop the classification models different approaches of chemometric methods were employed and compared to find out the best approach to provide the solution to this problem. An integral part of the studies is the development of classification models using the Artificial Neural Network algorithms (NN) as the chemometric tool for the analysis of LIBS spectral data for the identification and discrimination of molecularly complex biological materials such as bacteria and fungi. The selection of NN was encouraged by a study where comparison among several chemometric methods including Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Soft Independent Modeling of Class Analogy (SIMCA), Partial least Squares – Discriminant Analysis (PLS-DA) and Artificial Neural Networks (NN) demonstrated the best performance in the classification of bacterial strains by NN classification model.

There is a multitude of techniques that are employed to identify the bacterial and fungal pathogens present in a biological sample causing infectious diseases. All of these technique offer several advantages one over the other and provide good results but in many cases the factors like time and cost of analysis have to be compromised. Therefore, in this thesis, it was aimed to develop a methodology based on LIBS-NN combination to perform the identification and discrimination of these pathogens the identification and discrimination of bacteria and fungi in the biological samples; with special reference to improve the standards of hospital security, particularly from the microbiological point of

view by an early diagnosis of Hospital Acquired Infections (HAI) and its use in the clinical applications for the diagnosis of diseases resulting from bacterial and fungal infections.

The thesis is mainly divided into three parts: Introduction to the fundamentals of the techniques used, Experimental and Results.

The first part focuses on the introduction to the main objective of this thesis i.e. analysis of the microorganisms samples using LIBS technique along with the state of art of LIBS technique in the analysis of the microorganisms samples. It also covers the fundamental of LIBS technique along with a detailed theoretical approach to the chemometrics in LIBS, particularly the techniques. Particularly,

- Chapter 1 gives a theoretical view of the technique including the fundamentals of the LIBS technique, instrumentation and configuration commonly used for the analyses. In addition, this chapter covers the background of the LIBS based analysis of microorganisms discussing the principle of the technique, importance of bacterial samples studied so far and the advancements made in this technique for its medical applications.
- Chapter 2 includes a general introduction to chemometrics and the figures of merit that are needed to be taken into account for the evaluation of the performance of chemometric methods in the classification tasks. Specifically, the methods include supervised linear methods such as SIMCA, PLS-DA, LDA and non-linear supervised methods such as SVM and NN, discussed from a theoretical point of view.

The second part composing the

- Chapter 3 discusses the experimental set-up used for the LIBS measurements and the components of the experimental system developed at the laboratory. It also includes the bacterial samples included in the study and the preparation of these samples. The details about the programs used for developing the classification models is also given.

The third part of the thesis includes the experimental results and conclusions obtained in the study.

- Chapter 4 includes the developed LIBS-NN methodology for the identification and discrimination of bacterial samples. This part is further divided into three sections.

1. The first study is based on the application of LIBS and NN for the analysis of bacterial strains, studying on one hand the combination of these techniques to reach the identification and discrimination of these microorganism samples while on the other the effect of growth medium on the analysis and results.
2. The LIBS-NN based methodology was extended to the analysis of antibiotic resistant bacterial strains, which forms the second sub-section.
3. A comparison was also performed between the performances of different chemometric methods as mentioned in Fig. 1.1, although performed in the beginning, forms the last part of this

chapter. This comparison formed the basis to continue with the use of NN to generate the classification models.

- Chapter 5 presents the analysis of fungi samples of the class yeast, specifically Candida strains by LIBS with their further characterization by LIBS and SEM-EDS and applying LIBS – NN methodology for their discrimination.

- Finally, the last part chapter 6 discusses the conclusions on the work in this thesis and discussing the challenges the technique faces because of its versatility in application and instrumentation.

In the end of the thesis, three annexes have been added, which includes:

- User Interface developed for the analysis of LIBS data.
- Publications in scientific journals, including all projects carried out during the development of this thesis.
- Finally, the research work consisting of initial results obtained for the classification of different bacterial strains of same species published as Proceedings.

OBJECTIVES

The overall objective in this thesis was developing a set of analytical tools that allow a simple, economical, reliable and accurate analysis to identify and discriminate the pathogenic microorganisms and the extrapolation of these methods to the detection of pathogens in samples of biological origin, evaluating the potential of the LIBS technique

in combination with data analysis using mathematical algorithms of correlation for its application at the clinical level.

Specifically, the objectives to be achieved are outlined below:

1. Designing protocol for LIBS analysis of the microorganism samples included in the study. This involved evaluating the variables and conditions of measurement, optimization of experimental parameters and set-up, preparing samples for LIBS analysis and evaluation of direct detection of cultures without any sample treatment.
2. Study the effect of culture medium to grow microorganisms on LIBS based identification, given the need of isolating bacteria from biological samples prior to LIBS analysis due to matrix effect and importance of cultivating microorganisms as basic method in microbiology. This study was performed to find out if culture medium influence the elemental composition of the cells and affect the identification of microorganisms by LIBS
3. Development of classification models based on the study of mathematical methods of data analysis. This included comparison of linear and non-linear supervised chemometric models and correlation procedures in order to determine the method which provides better classification results. It was also intended to evaluate the classification power of Neural Networks in combination with LIBS and development of an interface for the user that allows real time control of mathematical parameters used NN models.
4. Evaluation of LIBS-NN methodology to discriminate and identify different pathogenic bacteria and fungi on the basis of their atomic composition in order to develop a fast and simple methodology for biomedical applications. Where on one hand it was intended to achieve the identification and discrimination of

microorganisms, on the other the study was intended to expand the scale of study by increasing the number of species like enterobacteria, gram-positive and gram-negative bacteria and also including different strains of these bacteria and fungi.

5. Determination of the antibiotic resistance in most commonly isolated bacterial strains. It included evaluation of the ability of LIBS-NN to classify bacterial strains based on their characteristic spectral fingerprint by detecting the compositional differences due to mutations resulting from acquisition of antibiotic resistance.

LIST OF PUBLICATIONS, CONFERENCES AND COURSES

PUBLICATIONS

1. "Laser induced breakdown spectroscopy for the discrimination of Candida strains." Sadia Manzoor, Lydia Ugena, Jesús Tornero-Lopez, Humberto Martin, Maria Molina, Joaquin Camacho. Talanta. Volume 155 (2016) 101–106.
2. "Identification and discrimination of brands of fuels by Gas Chromatography and Neural Networks algorithm in forensic research" Lydia Ugena, Samuel Moncayo, Sadia Manzoor, Juan D. Rosales, Jorge O. Cáceres. Journal of Analytical Methods in Chemistry. Vol 2016 (2016).
3. "Evaluation of supervised chemometric methods for sample classification by Laser Induced Breakdown Spectroscopy" S. Moncayo, S. Manzoor, F. Navarro Villoslada, J. O. Caceres. Chemometric and Intelligent Laboratory Systems 146 (2015) 354-364.
4. "Discrimination of Human bodies from Bones and Teeth Remains by Laser Induced Breakdown Spectroscopy and Neural Networks". S. Moncayo, S. Manzoor, T. Ugidos, F. Navarro-Villoslada, J. O. Cáceres. Spectrochimica Acta Part B Atomic Spectroscopy 101 (2014) 21-25.
5. "Fast Bacterial Identification by Laser Induced Breakdown Spectroscopy". Industrial, Medical and Environmental Applications of Microorganisms: Current Status and Trends." S. Manzoor, S. Moncayo, F. Navarro-Villoslada, J. A. Ayala, R. Izquierdo-Hornillos, F. J. Manuel de Villena, J. O. Cáceres. Wageningen Academic Publishers (2014) 647 – 652.

6. "Rapid identification and discrimination of bacterial strains by laser induced breakdown spectroscopy and neural networks". S. Manzoor, S. Moncayo, F. Navarro-Villoslada, J. Ayala, R. Izquierdo-Hornillos, F. J. Manuel de Villena, J. O. Cáceres. Talanta 121 (2014) 65-70.
7. "Determination of the postmortem interval by Laser Induced Breakdown Spectroscopy using swine skeletal muscles". A. Marín-Roldan, S. Manzoor, S. Moncayo, F. Navarro-Villoslada, R.C. Izquierdo-Hornillos, J.O. Caceres. Spectrochimica Acta Part B Atomic Spectroscopy. 88 (2013).186-191
8. Lycopene: A Review of Chemical and Biological Activity Related to Beneficial Health Effects. 40, ELSEVIER, 01/04/2013. ISBN 978-0-444-59603-1
9. New developments in lycopene analysis by spectroscopic and chromatographic techniques, accompanied with mathematical modelling. *ISHS Acta Horticulturae 1081: XIII International Symposium on Processing Tomato,* 2015

INTERNATIONAL CONFERENCES

1. 12 – 16 September, 2016. Euro-Mediterranean Symposium on LIBS. Chamonix, France. Poster – Analysis of Indium bearing mineral by Laser Induced Breakdown Spectroscopy.
2. 13 – 18 September, 2015. Euro-Mediterranean Symposium on LIBS. Linz, Austria.

   Poster – Evaluation of supervised chemometric methods for sample classification by Laser Induced Breakdown Spectroscopy.
3. 30 September - 02 October, 2013. International Conference on Environmental, Industrial and Applied Microbiology - BioMicroworld 2013, Madrid, Spain.

Oral presentation – Fast Bacterial Identification by Laser Induced Breakdown Spectroscopy.

4. 15 – 30 September, 2013. Wetsus Congress 2013. Leeuwarden, Holland.

   Oral Presentation (Invited) – Rapid Bacterial Strain Identification by Laser Induced Breakdown Spectroscopy and Neural Networks.

5. 16 – 21 January, 2013. The RME Conference Series, 8th Conference, Food Feed Water Analysis, Innovations and Breakthroughs. Noordwijkerhout, Holland.

   Oral Presentation (Invited) - Application of Laser Induced Breakdown Spectroscopy and Neural Networks: Rapid identification and discrimination of bacterial strains

6. 26 October, 2012. Organizzazione e disagio da lavoro: nuovi strumenti per migliorare competitività e prevenzione a livello europeo. Rome, Italy.

   Oral Presentation (Invited) – Discrimination and identification of bacterial strains by Laser Ablation Spectroscopy.

NATIONAL CONFERENCES

1. 20 – 22 July, 2016. XXIV National Meeting of Spectroscopy – VIII Iberian Congress of Spectroscopy. Logroño, Spain.

   Poster Presentation – Laser induced breakdown spectroscopy for the analysis of Candida strains and their discrimination.

2. 09 – 11 July, 2014. XXIV National Meeting of Spectroscopy - VIII Iberian Congress of Spectroscopy. Logroño, Spain.

   Oral Presentation – Biomedical applications of laser induced breakdown spectroscopy in bacterial identification.

3. 09 – 11 July, 2014. XXIV National Meeting of Spectroscopy – VIII Iberian Congress of Spectroscopy. Logroño, Spain.

   Poster presentation – Identification and discrimination of bones remains by laser induced breakdown spectroscopy and neural networks.

4. 16 – 18 June, 2013. XVIII Meeting of Spanish Society of Analytical Chemistry. Ubeda, Spain.

   Poster presentation – Determination of Post-Mortem interval using skeletal tissues by Laser Ablation Spectroscopy.

5. 17 – 20 September, 2012. XXIII National Meeting of Spectroscopy – VII Iberian Congress of Spectroscopy. Cordoba, Spain.

   Poster (Best Poster Award) - Application of Laser Ablation Spectroscopy and Neural Networks for the discrimination of bacterial strains of same species.

6. 17 – 20 September, 2012. XXIII National Meeting of Spectroscopy – VII Iberian Congress of Spectroscopy. Cordoba, Spain.

   Poster – How can Chemometrics support analysis of LIBS data? Application of chemometric methods in the discrimination of individuals using the remains of human bones and of same species.

7. 17 – 20 September, 2012. XXIII National Meeting of Spectroscopy – VII Iberian Congress of Spectroscopy. Cordoba, Spain.

   Poster – Application of Laser Ablation Spectroscopy to resolve forensic problem: Determination of post-mortem interval in muscle tissues.

COURSES

1. 22 – 24 June, 2015. Laser ablation spectroscopy and chemometric methods: multidisciplinary analysis (forensic, archaeology, food industry and health care). International University of Menendez Pelayo, Santander, Spain.
2. 27 – 28 February, 2014. Applications of lasers in Archaeology, Geology, Forensic Sciences and Food industry, Winter conference, Group of Laser Chemistry, Complutense University of Madrid, Spain.
3. 15 – 19 July, 2013. Laser ablation spectroscopy and chemometric methods: multidisciplinary analysis (forensic, archaeology, food industry and health care). Complutense University of Madrid, El Escorial, Spain.

   Oral Presentation – New Laser Spectroscopy based *in vitro* diagnostic methods for the identification of infections caused by fungi methods of diagnosis.

Chapter 1

Laser Induced Breakdown Spectroscopy based identification of microorganisms

Pathogen detection methods are of peculiar interest because the spread of pathogens for instance can not only endanger whole ecosystems, but can cause serious economic damage as well. Test procedures are required in order to control the quality of drinking water, food or pharamceutical substances and prevent placing them on the market and associated health risks to the population [4]. Although all of the above mentioned sectors are important, pathogen detection and identification has high significance in the medical field and therefore the need to improve the public health standards has impulsed the focus of this thesis on the medical significance of microorganisms (bacteria and fungi) in its clinical context .

Mostly the identification of the microorganism in a specimen is aided by the taxonomic classification based on binomial nomenclature determining the genus and species. The genus represents a collection of similar species where species is referred to the strains that are highly similar among each other with respect to other strains. For example, in *Klebsiella pneumoniae, Klebsiella* represents the genus and pneumoniae specifies the species. Revealing this information about a microbe along with the type of strain helps the clinicians to deal with a diverse microbiota and establish the level of pathogenicity and risk of infections related to the microorganism in an epidemiological investigation. Furthermore, this information allows designing the strategies for dealing with Hospital Acquired Infections (HAI) or Nosocomial infections. HAI include those infections which are not present at the time of admission of a patient to a healthcare setting

and are acquired within hours after admission from the surrounding pool of the microorganisms.

A wide range of microorganisms are responsible for causing the infectious diseases in human beings but the epidemiological reports by EPINE (Study of Prevalence of Nosocomial Infections in Spain) [5], indicate a high prevalence of certain groups of bacteria and fungi in the nosocomial or Hospital Acquired Infections (HAI) as shown in Fig. 1.1.

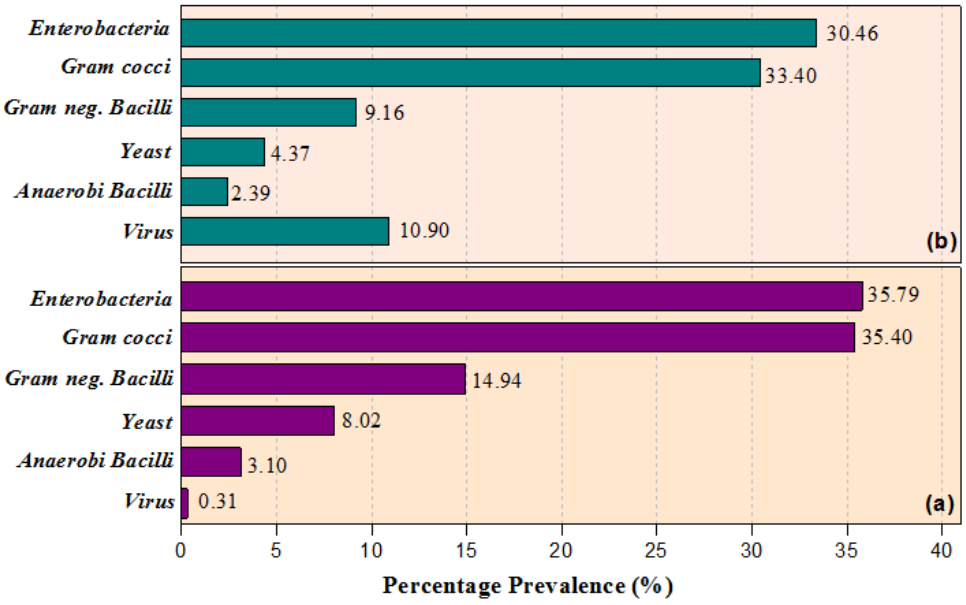


Fig. 1.1. Distribution of microorganisms isolated from patients affected by a) Hospital Acquired Infections and b) Community Infections [5]

According to this EPINE report [5,6], approximately 7% of hospitalized patients get affected by a nosocomial infection, which causes an increase of about 5 days in the hospital stay and 2% of these patients ultimately die from these infections. The direct and indirect costs for diagnosing and treatment can reach several million Euros. Bacteria cause about 90 % of HAI, 3% correspond to fungi and less than 1% to virus. Over 90% of all invasive fungal infections are caused by species of the genera *Candida* (most

common) and Aspergillus. Although fungi have a low incidence as compared to bacteria but they cause some serious infections and in most of the cases, it results in rapid death of the patient.

A serious concern posed by these pathogenic bacteria is the acquisition of antibiotic resistance and a consequence of the lack of control towards these infectious diseases is that the microbes acquiring resistance to drugs continue emerging. This results in another serious threat i.e. emerging infectious diseases (EIDs), which cause a significant burden on global economies and public health [7,8]. Studies have shown that 54.3% of EID events are caused by bacteria mainly, among them a large number belongs to drug-resistant microbes [9]. This requires developing an effective and efficient methodology capable of performing a faster microbial identification that could provide an instant diagnosis of the diseases.

1.1. MICROBIAL IDENTIFICATION METHODS

Many clinical laboratories still rely on the use of classical methods of microbial identification and characterization and will not be completely replaced, however, there is a continuous research going on in this direction that is leading to the development of new and innovative strategies. This has resulted to the addition of attractive alternatives and sophisticated instrumental techniques based on different aspects of microbial analysis providing significant advantages over traditional methods [10]. Various methods that have been brought up to deal with this matter are summarised in Fig. 1.2.

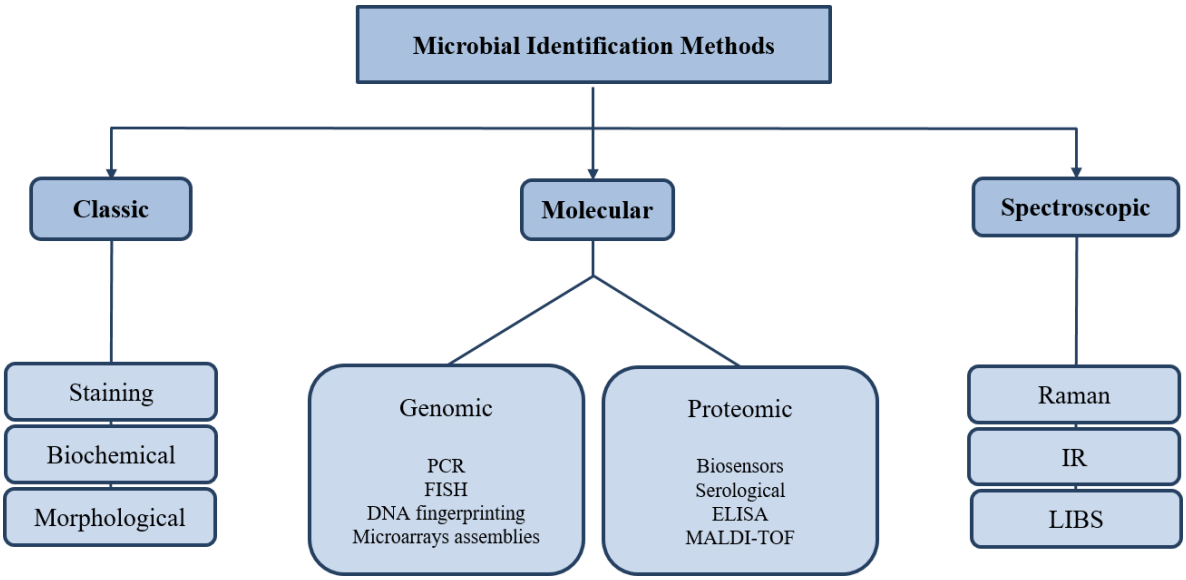


Fig. 1.2. Methods used for the identification of microorganisms

Classical methods for microbial identification are based on culturing processes coupled to the recognition of similarities and differences in morphology, growth, enzymatic activity, metabolism, gram staining, growth conditions observing the presence or absence of oxygen, growth ability on different culture media, bacterial fermentation, parasitic morphology, and viral cytopathic effects with those of type cultures. These methods employ more traditional and laborious tasks such as microscopic observation and biochemical characterization of sample microorganisms with those of type cultures and the reliability of this kind of identification is in direct proportion to the number of similar phenotypic characteristics [11,12].

Genotypic and Proteomic analysis of microorganisms have gained momentum during the last decades and have proven to be very useful in the identification and characterization of bacteria and fungi strains overcoming the shortcomings of traditional methods in failing to identify complex similar strains. The genotypic methods include nucleic acid based detection methods and rely on the amplification, sequencing,

fingerprinting and hybridization of DNA or RNA via polymerase chain reaction (PCR), where real time–Polymerase Chain Reaction and microarrays assemblies are the most commonly used techniques currently [13]. Other molecular techniques used include FISH [14] and sequencing of 16s RNA [15,16]. Proteomic methods of bacterial identification include the analysis of proteins (or specific markers) in a sample [12,17]. This includes the serological proteome analysis based on immunological techniques, which exploits the specificity, and sensitivity of the interaction between antigen and antibody is measured. In some cases, fluorescent or radioactive substances are added to the antibody which later allows the detection and quantification of antigen and hence the microorganism present [18]. Enzyme–linked immunosorbent assay (ELISA) is a serological technique used fr the detection of an antigen or antibody. It can be used to detect a bacterial antigen by an enzyme–substrate reaction by conjugating a specific antibody with an enzyme [19]. Other methods include two dimensional gel electrophoresis and obtaining the sequence of proteins generating a map, which is compared to the reference database to identify the bacteria [20].

Another proteomic technique that is extensively being used for microbial identification is Matrix Assisted Laser Desorption Ionization – Time-of-Flight – Mass Spectrometry (MALDI-TOF). It offers a gel – free spectrometry proteomics involving the absorption of UV laser at 337 nm by the matrix containing microbial sample causing soft ionization of sensitive biological molecules, generally producing charged ions of high molecular weight that are detected by the mass spectrometer providing a mass spectrum fingerprint [12,21],.

Spectroscopic methods have also been reported to be employed in microbiology over the last years providing an alternative physiochemical approach to the analysis of microorganisms. Due to their feature of displaying the molecular composition of the

sample with unique specificity, providing the fingerprint of the chemical composition of each microorganism based on the application of vibrational spectroscopic techniques i.e. infrared (IR) and Raman spectroscopy, these techniques have been the subject of various studies [22-30].

Although molecular methods are used for identification as well as characterization of microorganisms, the spectroscopic based methods mainly deal with the identification and discrimination between the microbes. These methods deal with different aspects of microbial cells such as cell wall composition, cell proteins, nucleic acids, molecular and atomic composition, offering different features and advantages and hence each method has its own potentials and limitations. When a clinical diagnostic is needed to be performed, an ideal method is the one that is able to provide an accurate and reliable identification of the infectious agent in a short period of time and at low cost per analysis. On the methodological basis, the diagnostic method should be simple and non-laborious to perform, easily available, temporally and economically feasible and present a simple experimental set-up, data analysis and results inference. When a diagnostic method is evaluated based on these features, the comparison between the efficiency and possibility of a routine use shows that none of the methods can be determined as an optimal or ideal method for the detection of pathogens in clinical samples. Thus new techniques and methodologies are continuously needed, the use of which will not replace the existing techniques, however, their development will be an ongoing process [10].

The focus for the new techniques is the development of rapid, cost effective and simple methodologies as these aspects are of utmost importance in medical diagnosis due to the tremendous impact of these infectious diseases on human health as well as economic issues. In order to fulfil these demands, it is necessary to employ adequate tools that provide reliable, timely analytical results essential for day-to-day analysis that are

temporally and economically feasible. Growing evidence suggests that early, appropriate antibiotic therapy improves outcomes, and such therapy should therefore be a goal; however, this strategy needs to be coupled with an early reassessment of both diagnosis and therapy, usually within 48 to 72 hours. Provided a quick accurate diagnosis is enabled, in the majority of cases, the medical treatment can be targeted to a more specific regimen or can be discontinued, if an alternative microorganism is identified [4,31].

The molecular methods provide accurate and reliable results and have been established with defined protocols and are currently extensively being used but the economic aspects related to them because of laborious nature, long time, high cost of analysis due to the consumables required and huge set–up cost put limitations to their use in routine clinical analysis [32]. Moreover, there are other challenges faced by most molecular methods that include the long time period for culturing and high volumes of it required in order to obtain a sufficient quantity of culture to be able to carry on with the analysis, high sensitivity to contamination, if the sample matrix contains substances inhibiting the PCR reaction, the need of expertise in both molecular biology and bioinformatics to interpret results for sequence alignments and analysis and direct contact with the hazardous organisms . The classical methods are laborious and heavily rely on culturing of microorganisms, rendering them extremely slow as some microorganisms take several days to provide a critical biomass for analysis. Furthermore, due to the increase in the diversity and complexity of the microorganisms with time highly related microbes are complicated to be differentiated based on phenotypic characteristics. The bias or lack of inexperience towards a particular isolate may result in misidentification when a biochemical test are interpreted to fit to a specific profile as reported in certain cases [11]. Serological methods are highly specific and expensive due to the high cost of the antibodies used. MALDI-TOF is currently increasing the interest in this field but its

big size and need to maintain at very low vapour pressure imply it to be installed in special set-ups and the high cost of equipment can be limiting factors. The spectroscopic methods based on Raman spectroscopy are coming up but are mostly kept back off the clinical laboratories due to complexity in maintaining the equipment as well its high cost.

Despite the fact that classical, genomic and proteomic techniques are widely used for pathogen identification and provide accurate results but mostly require expensive consumables and are time consuming, therefore, are not efficient enough to deal with situations where a fast diagnosis is required. In some cases, it has been reported that the genomic methods may also provide a considerable number of false positives, failing to provide accurate results as compared to the slow conventional methods [33]. This is why further research and development of more sophisticated, cost-effective and fast approaches is underway to bring advances in this challenging territory.

Laser Induced Breakdown Spectroscopy (LIBS) is a budding technique, and has captured the interest of researchers because of its capability to generate an instant spectral fingerprint of the sample based on the atomic composition of the microorganism that could be used to identify and discriminate microorganisms. Such approach is often referred to as a 'whole-organism' fingerprint in microbiological studies, as each different microorganism produce a unique spectrum [34,35].

LIBS, because of its working principle and being an elemental analysis, was previously used in analytical applications such as inert matter mining, nuclear energy, environmental chemical pollution, art materials, etc., and in the physical studies of plasma, however, its potential to analyze complex microbiological systems was not reported until 2003. It was in 2001 when the events of bioterrorism hit the United States where spores of lethal bacteria *Bacillus anthracis* sent by mail and caused serious infections (categorized as deliberately emerging diseases) [8] and deaths in few cases.

The need of developing a fast identification method for timely detection of such hazardous microorganisms motivated the researchers to explore it for the detection of microorganisms based on the quest that if a bacterial sample could be identified based on its elemental composition. The advantages mainly high speed of analysis, minimal sample preparation and no direct handling of pathogenic samples made it more appealing to this end. The first study in this direction was carried out by Morel *et. al.* [36] who analysed *Bacillus globigii,* a non-pathogenic surrogate for *Bacillus anthracis*. This study focused more on characterization of bacterial samples by LIBS revealing the information that plasma could provide on atomic and molecular level and therefore finding the markers to achieve discrimination between them. The occurrence of several events such as outbreaks of food-borne illness that caused deaths and serious infections led to the appearance of more research in this field focusing on bacterial strains causing infectious diseases [37-39].

Following is the description about the working principle, state of the art analysis on how this technique has been used for the study of microorganisms and a brief description on fundamentals of LIBS from theoretical point of view.

The working principle of LIBS technique relies on the use of a highly energetic, short pulse of laser light focused on to the target which causes the vaporization of a fraction of target material forming on its surface a plume "plasma" within the focal region of the laser [1,3,37],. As a result of the vaporization or ablation or laser induced breakdown of the irradiated material is converted into the constituent components and are retained in the plasma in the form of energetic atoms, ions and electrons. The light emitted from this plasma is collected and recorded as a pattern referred to as "spectral fingerprint" consisting of emissions lines from the elements forming the target with their intensities located at specific wavelengths. This spectral fingerprint is a characteristic feature of the

sample under analysis and hence provides instant information about the sample composition [1,40-42].

The ability of pulsed focused laser radiation to initiate plasma directly irradiating a sample in-situ and at a distance, perform fast spectral signature acquisition in seconds time taking into account the time from generation of plasma to collection of its light enabling simultaneous scanning of constituting elements of sample, provides a quick and versatile method for analysing different types of samples that can be inaccessible or tedious using conventional analytical techniques. The possibility of (a) elimination of cumbersome sample preparation steps; (b) performing analysis in any state of matter; (c) requiring minute amount of sample in micro- or nanograms; (d) setting up a simple experimental system; (e) low cost and compactness of equipment; (f) remote measurements for the analysis of hazardous, highly contagious pathogenic targets, and (g) automation of analysis make LIBS a simple and potent technique for its application in microbiological systems [40,41].

1.2. LASER INDUCED BREAKDOWN SPECTROSCOPY FOR MICROBIAL ANALYSIS

Microorganisms although unicellular organisms, possess a highly complicated cellular structure. Although they are basically made up of carbon based organic molecules and water, and composed of genome, cell organelles, cytoplasm, and different membranes, that can be found common in all but at the structural and molecular level there exist variations between different microorganisms. Therefore, they can be differentiated and/or classified on the basis of different cellular structures such as presence or absence of cell wall, nuclear membranes, types of cell wall, serotype, metabolism etc.

Bacteria is a biological structure with mainly carbon, hydrogen, nitrogen, phosphorus and oxygen forming the base along with other trace inorganic elements such as calcium, sodium, magnesium, potassium that are involved in the metabolic activities, enzymatic activities of the cells and maintaining the cellular structures [37]. Depending on the environment and genetic make-up, microbial cells can intake other elements from the medium. Furthermore, on structural basis bacterial cells can be seen sharing some common cellular elements and some variable structures such as pili, flagella or capsule as shown in Fig. 1.3, which also contribute to differences between different types [36].

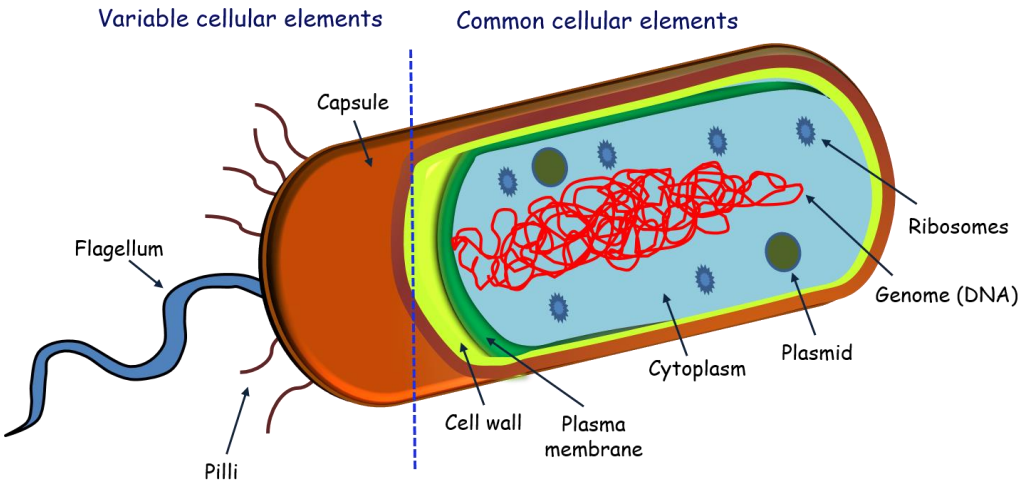


Fig. 1.3 Basic structural division of bacteria [36]

In LIBS analysis, all these variations can be of great help as depending on the characteristics of a microbe a highly specific spectral fingerprint can be generated which can allow achieving identification and discrimination between wide ranges of microorganisms. The laser pulse vaporizes several bacteria from the surface of bacterial target breaking it into its constituents and distributing this elemental and molecular information over the entire spectral range. This spectral fingerprint acts as a marker for the particular microorganism in order to provide information about its identity and related

pathogenicity [36]. Taking the advantage of possibility of coupling LIBS with other techniques such as Laser Induced Fluorescence Spectroscopy (LIFS), the same year, Hybl *et al.* [38] investigated the potential of LIBS–LIFS for detecting and discriminating aerosolized bacterial spores from more common background fungal/mold spores and pollens [41]. These studies highlighted the meaningfulness of the LIBS spectral fingerprint demonstrating the possibility of elemental based identification of microorganisms and marked the onset of a completely new direction of research integrating LIBS and microbiology with each other.

The possibility of using LIBS to identify and discriminate pathogens in clinical diagnosis and public health has prompted investigations into the use of more realistic samples from a clinical analysis perspective. Almost all of the studies so far have emphasized studying bacterial samples and not much attention has been paid on the fungi samples except the inclusion of few strains in some studies. Most commonly studied bacteria include *Escherichia coli, Pseudomonas aeruginosa*, along with different species and strains belonging to the genus *Staphylococcus, Streptococcus, Salmonella, Mycobacterium, Enterobacter and Bacillus.* Among fungi samples a strain of *Candida albicans* was included whereas just a single study reported the analysis of virus i.e. *Hantavirusstrains* [43].

All of the studies have employed different methods for the mounting of samples for LIBS analysis, and therefore there is not present any standard method of manipulation of the microbial samples and different protocols have been implemented for this purpose. A review of the bacterial based LIBS studies show that in some cases a previous centrifugation [44] of bacterial culture was done and the semi-liquid pellet obtained was deposited onto a nutrient free agar bed, others performed analysis by immobilizing the

bacterial slurry directly on a nutrient-free agar bed [45], microscope glass slides or substrate coupons made of polycarbonate, stainless steel and aluminium foil [43,46].

Like many other techniques LIBS analysis requires pure culture and therefore isolation of microbial culture, however, Rehs*e et al.* [47] studied the presence of a second bacteria in a mixed culture and its effect on the identification capability. Furthermore, the dependence of LIBS signal intensity with the bacterial cell number was studied making different dilutions of the pure culture and it was demonstrated that a linear relationship existed between them. It was a very interesting analysis demonstrating the ability of LIBS in the detection of mixed infection. However, studies report that about 90% of the bacteremia patients were found to contain only one pathogenic bacteria at a time and infection resulting from multiple pathogens was found to be rare [33,48,49].

A review of the instrumentation employed for LIBS measurements of bacterial samples reveals that almost all studies used Nd:YAG laser. This laser is usually preferred by LIBS community over other lasers because of its size, compactness, easy maintenance, repetition rate, high power throughput, short pulse duration, transmission of laser radiation via fiber optics, efficient frequency conversion, high amplification, better thermal and mechanical properties and easy-to-use source [50]. Another important component of the LIBS system is the spectrometer and detection system as this is the part where plasma emission is resolved and the spectral fingerprint is generated. Different systems have been used based on Czerny Turner combined with Charged Coupled Device detectors (CCD) but the introduction of high resolution Echelle spectrometers coupled with ICCD detectors have given some highly resolved spectra and signal enhancement [1,2,51,52]. The emission from the plasma is greatly affected by the ambient gases and enhancement in the signals was observed in the spectrum with helium and argon as ambient gases and hence an improvement in the discrimination capability [53]. However,

measurements have been usually made in air at room conditions that also allowed an efficient discrimination, furthermore kept the analysis simple and experimental set-up easy to handle [54-57]. The spectrum recorded may be the result of an irradiation with a single or multiple laser shots. Some LIBS experiments relied on single shot spectra while others recorded spectra by accumulating and averaging individual measurements including one point averaging or multiple points averaging or a combination of both in order to increase the S/N ratio, LIBS analytical performance, precision and reproducibility in the spectral set.

A very important parameter that affects this is the chemometric method selected for the analysis of data and its tolerance to the shot-to-shot variations in data. Most of the times the method selected is the one available at hand. Different chemometric methods have been evaluated by other research groups, such as principal components analysis (PCA), soft independent modeling of class analogy (SIMCA), linear discriminant analysis (LDA) and partial least-squares discriminant analysis (PLS-DA). Some widely used chemometric methods sometimes do not offer satisfactory solutions to many practical problems due to the complexity of data or problem to be addressed. Research exists demonstrating that the use of NN can provide better results [58,59]. Therefore, a comparison of some of these methods has been performed in this thesis.

1.3. Fundamentals of Laser Induced Breakdown Spectroscopy

The process of LIBS involves several stages mainly dealing with the interaction of laser with the sample then with the ablated material and finally coupling of the remaining pulse energy with the plasma [60] (Fig. 1.4). The process is initiated with the irradiation of a sample surface by a high power energy laser pulse typically at 1 – 10 $GW.cm^{-2}$ [61]. This marks the onset of a sequence of events. The laser shot which is

usually a nanosecond pulse causes the breakdown of a portion of sample in the focal volume of the laser pulse by local heating, melting and vaporization leading to its atomization.

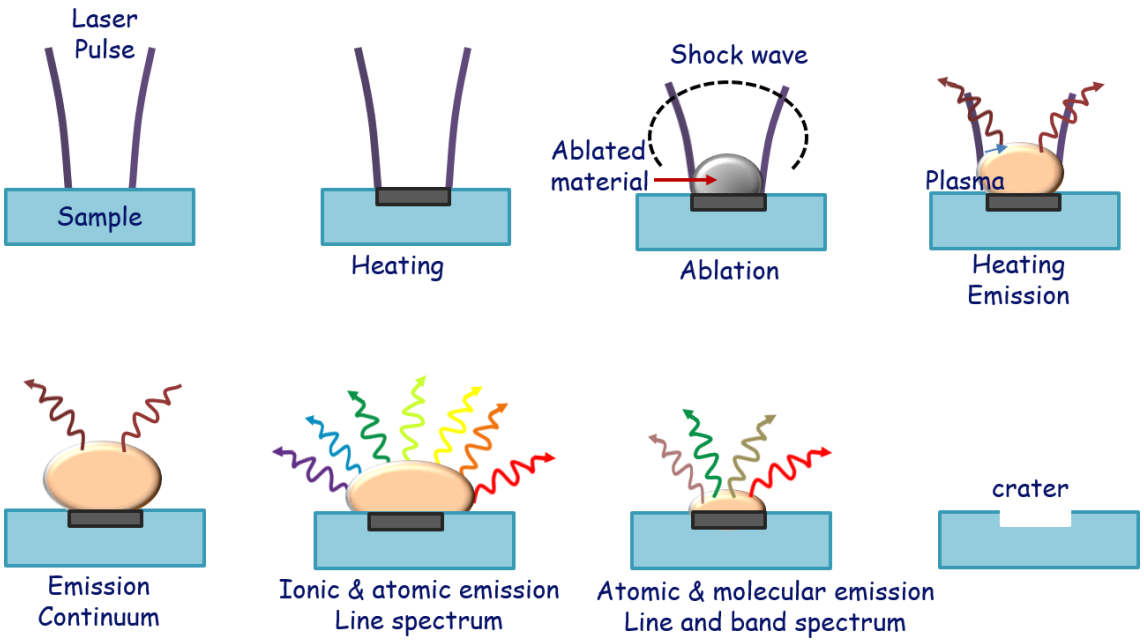


Fig. 1.4. Main stages during the formation and emission of LIBS plasma

The effects of laser-matter interaction and breakdown threshold to generate plasma are usually specified as the minimum irradiance defined as power per unit area [60]. The breakdown caused by optical excitation begins with the presence or generation of free electrons (seeding electrons) that absorb the incident laser energy through three-body collisions with photons and neutrals followed by avalanche ionization in the focal region of the laser pulse. During the time period between collisions the electric fields of the optical pulse tend to accelerate the free electrons leading to thermalizing their energy distribution [3]. The increase in the energy of electrons and their collisions result in the ionization of species in plasma, generation of more electrons, hence more of the energy absorbed producing an avalanche. This leads to formation of high pressure plasma plume

with a temperature as high as tens of thousands of degrees and starts expanding ultrasonically outwards in all directions at a high speed creating a shock wave and also producing a loud sound. The expansion of plasma is accompanied by changes in temperature and density of electrons and depends on ablated mass, spot size, energy coupled to the sample, and environmental conditions (state of the sample, pressure, etc.). The expansion of shockwave and plasma causes pressure changes in the surrounding gases creating a negative pressure gradient collapsing the plasma plume [3,60],. The plasma finally cools and decay emitting energy in different ways leaving crater behind.

Pulses of femtosecond duration have also been used to perform LIBS and the mechanism of plasma formation, expansion and the profile of spectral lines and the interaction with ambient atmosphere differs from the one using ns pulses. In case of fs laser pulses non-thermal processes dominate the ionization and the highly energetic laser pulses induce excitation of sample by effects such as multiphoton absorption and ionization, tunneling, and avalanche ionization giving rise to highly charged ions by Coulomb explosion [51]. Femtosecond pulses being short prevent thermal effects producing a much-defined crater without melted edges and deposited material around it. On the contrary in comparatively longer ns laser pulses the heating time of electron-lattice is approximately 10^{-12} s that is much shorter than the complete pulse duration causing thermal effects to dominate ionization and the increase in cause the ionization. However, if the irradiance is high non-thermal processes also accompany to ionize the sample. During the irradiation of a sample by laser pulse a part of the pulse energy is consumed in the ablation of the material, forming a plasma plume which becomes opaque between 10^{-9} s and 10^{-8} s, acting as a shield and preventing more sample to be ablated therefore a significant remaining fraction of energy is absorbed by the expanding plasma leading to the excitation of species (atoms and ions) constituting the plasma plume. This

phenomenon, called plasma shielding, produces a crater with melted edges and deposited material around it. Because of these phenomenon the lifetime and size of ns LIP is greater than fs LIP [51]. The evolution of plasma has been given by three models i.e. laser-supported combustion (LSC), laser-supported detonation (LSD), and laser-supported radiation (LSR) waves, differing in the opacity of the plasma and energy transfer to the ambient [3].

Optical spectroscopic studies of plasma reveal its temporal evolution and the emissions from it containing continuum and line radiation as a result of bremsstrahlung, recombination and de-excitation of atoms [52]. As the laser induced plasma (LIP) is formed by a pulsed laser source the lifetime of the plasma is very short approximately 300 ns to more than 40 μs range depending on the ambient gas, its density among other factors [62]. The timeline of events occurring and radiations emitting from different species constituting the plasma have been shown in Fig. 1.5. Initially the plasma is rich with ions but later the ion – electron recombination forms the neutral atoms and finally some of those combine with each other or ambient atmosphere forming molecules. As in the beginning, the plasma is rich in ions and electrons, early stages of plasma emission is dominated by a white light continuum that arises from the bremsstrahlung effect (free – free) and recombination (free – bound).

$$X^{+} + e^{-} \rightarrow X + h\nu \qquad \text{(recombination)}$$

$$X^{+*} \rightarrow X^{+} + h\nu \qquad \text{(de-excitation of ions)}$$

$$X^{*} \rightarrow X + h\nu \qquad \text{(de-excitation of atoms)}$$

This continuum ranges from 2 – 600 nm approximately and dominates in the first hundreds of nanoseconds of plasma life, interfering with the characteristic emissions from ionic, neutral and molecular species. During the subsequent stages, the continuum decays

with time and the spectral lines and molecular band spectra arising from de-excitation of ions, atoms and molecules start appearing sequentially.

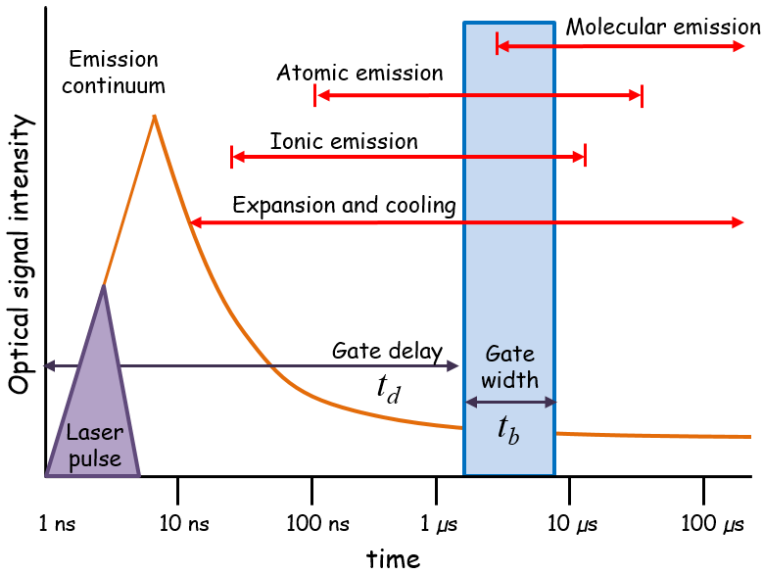


Fig. 1.5. Timeline of events showing the evolution of plasma and observed emission

Although the continuum minimizes with time, it does exist along the entire length of plasma emission therefore in order to record a spectrum that could allow for the LIBS compositional analysis of the sample, a compromise between high line intensity and minimum background is required. This is achieved by time resolution of plasma and finding the delay "t_d" i.e. time passed since the laser shot is fired to the opening of window "t_b" defined as time interval during which spectral acquisition takes place. The values of "t_b" and "t_d" are important parameters and optimized depending on the application and instrumental set-up [62]. Of all stages of LIBS process, the last stage that is recording of optical emission is crucial and basic component of any LIBS measurement as it provides the LIBS spectral fingerprint revealing the information about the sample [1,40]. The analytical performance of LIBS technique is strongly dependent on the experimental parameters such as laser wavelength, the pulse energy, time interval of spectra acquisition

and the geometrical set-up of the collecting optics. Furthermore, achieving plasma conditions where Local Thermodynamic Equilibrium (LTE) and optically thin plasma is the underlying premise for the use of LIBS as an analytical technique provided that the sample undergoes stoichiometric ablation i.e. the composition of the luminous plasma reflects that of the target [62]. LTE is defined by the condition when material particles balance has attained equilibrium but the light-matter balance does not. This assumption is considered in case the McWhirter and the Boltzmann distribution criteria are fulfilled, which state that the critical value for electron density ($N_e(cm^{-3})$) in the plasma should be greater than $1.6 \times 10^{12} T^2(K)(\Delta E_{nm})^3(eV)$. Here, T represents the plasma temperature and ΔE_{nm} represents the energy gap between two adjacent transition levels n and m [63]. In order for LTE condition to be present in plasma, electron density, N_e, should have a value of 4.7 X 10^{15} cm^{-3} [64]. According to Griem's analysis, for a temperature of 1 eV (~11 000 K), irradiance of >10^8 W/cm^2 and at atmospheric pressure, LTE exists at an electron density of 10^{17}/cm^3 [65].

The spectrum obtained has discrete lines and bands belonging to the constituent elements and molecules formed as a result of recombination in plasma that can be characterized based on their wavelength, intensity and shape, and depend on the parameters such as the structure of the emitting atoms and their environment [1,40,51]. Simultaneous multi–elemental scanning of a sample to provide a characteristic spectrum enables LIBS to provide qualitative and quantitative information about the sample, which requires a previous identification of emission lines to detect the constituent elements. Several factors are considered when identifying a particular line such as *a priori* information about the expected composition of samples, significant and/or intense emission lines, ionization stage of emitting species (neutral atom, ion or molecule), relative intensities of the lines of interest and experimental conditions (including ambient

gas and focused power densities). Plasma temperature is another very important factor to take into account in order to identify the emission lines unequivocally as several lines corresponding to different ionization states can be observed for an element. However, generally singly ionized species are found in LIBS plasma, nonetheless, other oxidation states unlikely to be present, can be observed under special conditions [1,66,67]. In addition to emissions from atoms and ions, emissions from simple molecules formed such as CaOH, AlO, CN, etc., because of recombination between atoms in the plasma or interaction with ambient air are also observed [1]. The emission lines in the LIBS spectra can be identified by comparing them manually with those provided in the spectral data bases such as National Institute of Standards and Technology (NIST) [68], Vienna Atomic Line Database (VALD) and Kurucz atomic and molecular database (Harvard) [69]. These databases can be assessed using wavelength of the emission line as prime identification parameter however additional data including ionization stage, relative intensity and energy level designations with energies are also provided [1]. Characteristic patterns observed in the spectra such as doblets, triplets or profile facilitate the assignation of certain emission lines and bands in the spectrum. Another approach that can be used in order to assign an emission line of interest in a sample is by generating a database using inorganic salts containing the element of interest. Superposition of lines from different elements, dependence of line intensity on experimental conditions, complexity of LIBS spectra and subjectivity in assignation can sometimes make manual identification tedious and lead to incorrect results especially in case of weak emission lines [70]. Therefore, in view of the need to perform a reliable, fast and automatic identification of lines, different programs and software have been developed for particular applications [71-73] as well as generic experimental data such as SALIPS (Software for the Analysis of Laser-Induced Plasma Spectra) [70]. Simulations of a set of spectra corresponding to different plasma

temperatures and electron intensity is also carried out to develop algorithm for an automatic identification of emission lines in LIBS spectra. The lines can be assigned by finding the best correlated pair for model and experimental spectra. Plasma temperature, nature of emission lines (atomic or ionic), different excitation conditions, local thermodynamic equilibrium, line broadening and spectrographic features are essential paramets to consider [73].

The qualitative analysis encompasses finding out the elemental composition of a sample, monitoring any specific element/s compositional changes and the chemical pattern recognition. In case of pattern recognition, the analysis can be based on selected wavelength ranges or entire spectral fingerprint recognition. LIBS spectrum is complex emission pattern and as it reflects the samples composition, differs for different kind of samples. These differences are sometimes more evident while in some cases very subtle differences are present. Therefore, pattern recognition methods are used to develop classification models in order to reveal the identity of a sample within a certain limited set of materials [1].

The quantitative analysis by LIBS can be basically divided into two main groups; empirical and theoretical [74]. The empirical approach includes the conventional univariate analysis and multivariate approaches, both based on obtaining calibration models. Multivariate analysis make use of chemometric tools and require calibration of LIBS spectral data. These methods require an internal standard for calibration, however, this makes these methods highly sensitive to matrix effects and the physical state of the samples. In order to deal with these problems, matrix matched standards are required which are mostly not easily available. In the area of quantitative LIBS, this issue was dealt by developing the theoretical approach, “Calibration Free – Laser Induced Breakdown Spectroscopy (CF-LIBS)” by Ciucci et. al. [75], which is based on calculation

of some plasma parameters such as measuring the ablated mass, the plasma electron temperature and the electron number density from the LIBS spectra. Few other approaches have also been put forward, employing Neural Networks for data analysis in CF-LIBS as well as some hybrids or modifications of LIBS such as Optimized Calibration (OC-LIBS) [74]. All this progress in quantitative analysis has been possible due to introduction of advanced instrumentation involving broadband high resolution spectrometers that have allowed increasing the sensitivity of LIBS.

In this thesis, qualitative approach of LIBS has been applied, particularly, pattern recognition and classification. Therefore, more stress has been given on qualitative aspect and chemometric methods used for the data analysis. These pattern recognition methods form part of chemometric techniques and are discussed in Chapter 2.

Chapter 2

Chemometric methods for the analysis of LIBS Data

Chemometrics is an interdisciplinary field combining statistics, mathematics and computer science in order to extract useful information from the multivariate data. The origin of chemometrics lies in chemistry and it emerged as a response to the problem of dealing with large amount of data generated by the computerized instrumental techniques. In 1960s, some groups working in analytical chemistry started using mathematical and statistical techniques for data analysis. Early studies were based on Principal Component Analysis (PCA), Factor analysis, sequential simplex optimization, canonical correlation, k-nearest-neighbours among others involving the use of mainframe computers. The availability of personal computers in 1980s and advancements made in the field of computer technology allowed a sudden progress and expansion of this approach to acquisition, processing and interpretation of chemical data [76]. In the present time chemometrics has become a very mature and extensively worked field being widely used in analytical chemistry, medicine, biochemistry and engineering among others [76-93].

Typically, in spectroscopic studies, chemometric techniques are applied to signal processing, experimental design, extracting relevant information, exploring the patterns associated within the data, cluster analysis, pattern recognition and quantification. The use of chemometrics offers the advantages of noise reduction, neglecting the effects of interfering signals, and the exposure of outlier samples [1,94-98],.

Chemometrics for Laser Induced Breakdown Spectroscopy (LIBS) data analysis was first used in 1994 and since then the number of papers, using LIBS–chemometrics

combination is constantly increasing. Although in the beginning chemometrics was employed for quantitative analyses, recent studies are using it for qualitative purposes focusing on pattern recognition and sample classification. LIBS is being proposed more and more as a high-throughput technology for the sample analysis. On one hand this can be attributed to the high speed of spectral data acquisition while on the other the application of chemometric techniques for data analysis has allowed it to move ahead even more quickly, and is one of the reason of the success of LIBS as a potential tool in various applications. A wide range of chemometric methods have been employed in the LIBS studies and the literature is expanding with the continuous introduction of new approaches and software tools for LIBS data analysis. Even there have been introduced some commercial LIBS systems that have chemometric analysis software embedded in them [1].

Mostly LIBS experiments include recording a set of data with a large number of measurements and each laser shot produces a spectrum containing tens of thousands of intensity values (variables) generating extremely large data sets that are needed to be statistically modelled. The use of chemometrics enables multivariate analysis of complex spectral data involving computations with only a few clicks and seconds time [99]. Data analyses based on the selection of one or two more variables from the dataset keeping out a range of variables provides the disadvantage of missing out the information that can be obtained by establishing the relationship between the variables and their usefulness to the sample under study. Moreover, in LIBS spectra an element has usually more than one emission line, and there are many examples where several emission lines used in combination give dramatically better results than any of the individual variables used alone. With a multivariate approach, models can be constructed taking into account all

the emission lines related to a specific element, along with the peak profiles; therefore more robust models can be developed [100].

In case of LIBS – chemometrics combination, a detailed analysis of the spectra is not required and an instantaneous identification of the sample is done using a correlation procedure to recognize the spectra from different samples [41]. Fig. 2.1 shows the scheme of such qualitative analysis.

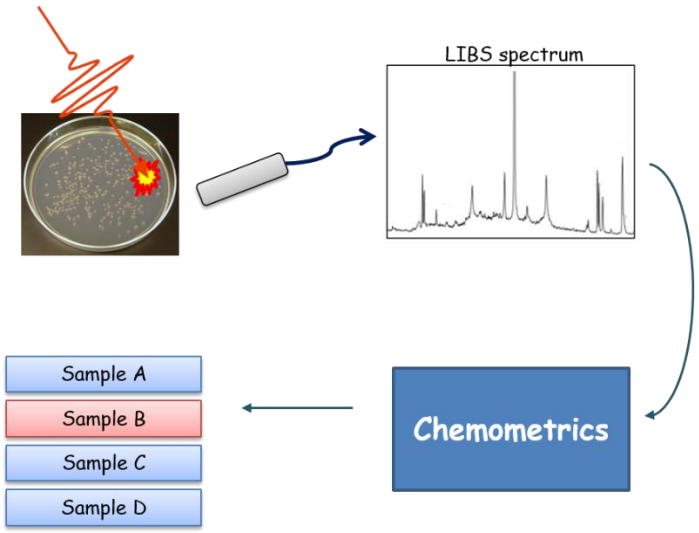


Fig. 2.1. LIBS and chemometrics combined for the classification tasks

In this thesis, chemometric techniques have been employed for the purpose of pattern recognition, to obtain classification models able to recognize the membership of each object to its class based on a set of measurements. These methods are based on finding mathematical relationships between a set of descriptive variables (measurements) and a qualitative variable (membership to a defined category) [101]. Depending on the different characteristics and properties, these methods have been put into several categories. Regarding classification methods, two procedures are mainly used; unsupervised and supervised.

Unsupervised methods do not require *a priori* model on the set of samples used for classification purpose. These methods involve only one dataset "matrix **X**" or "matrix of objects" defined by *I* rows (objects) and *J* columns (explanatory variables), composed by G categories (classes). These methods generate the model, determining the number of classes present in the dataset by revealing the patterns of association and groupings and characterize the variables that define each class, also reducing the complex data to a more comprehensible form.

Supervised methods are defined by two datasets; matrix **X** composed by objects (input variables) and classes **Y** (target) and attempt to discover the relationship between input and target, representing in a structure referred to as a model, characterized by a set of features that define the classification process. This approach consists of two steps, calibration or training and second being the validation of the generated model where these models can be used to predict the membership of new objects (unknowns for the models) on the basis of their similarity to a certain class in the model [102].

Another criterion to classify these methods is based on the division of hyperspace such as pure classification and class modelling. *Pure classification* also called discriminant classifiers bases on separating the hyperspace into the regions equal to the number of classes, whereas *Class modelling* involves modelling each class individually by a defined boundary based on similarities between the objects of the class. Fig. 2.2 shows the creation of a model by pure classification and class modelling on a set of objects belonging to three classes separating the hyperspace into three groups and posterior projection of three new objects T1, T2 and T3.

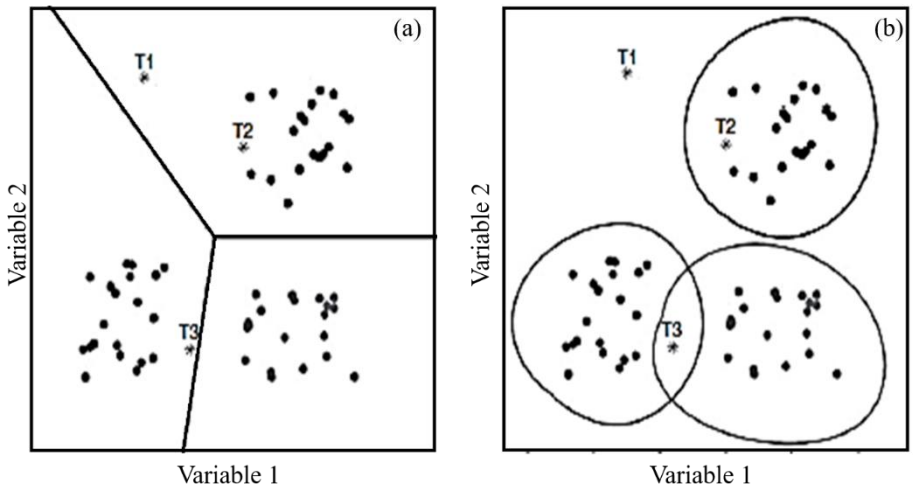


Fig. 2.2. Projection of new objects T1, T2 and T3 on the models created by (a) pure classification and (b) class modelling on a data set defined by 60 objects and two variables grouped into three classes.

The Fig. 2.2 demonstrates the drawback of pure classification methods as each object is always assigned to one of the one of the regions or predefined classes depending on the modelled space and if an object does not belong to any predefined class; it is also assigned to one of the modelled classes. On the other hand, class modelling enables recognizing the objects belonging to none of the class spaces or the object can can be assigned to more than one class and therefore considered as a confused object [103]. From the mathematical point of view, algorithms employed by these methods can be classified into two types based on the mathematical form of the decision boundary, i.e. linear and non-linear. This refers to the ability of the method to detect linear or non-linear boundaries between classes. In case of linear classification method, the best linear boundary separating the classes is calculated by using linear combination of variables to estimate the classification model. The non-linear methods base on finding the best curve (non-linear boundary) and employ more complex algorithms that increase the sophistication improving the classification capacity of the method [58,101,103-108].

A combination of these strategies give rise to a variety of statistical and computational algorithms, therefore, the selection of a chemometric method depends on the analytical application, the information available and desired outcome. Although chemometric methods offer a rapid and simple approach to data analysis, it requires best quality spectral data, and knowledge to interpret the results and significant understanding of the mathematical algorithms as well as optimizing the working parameters accordingly in each case [99].

Once representative data is collected, termed as objects that were spectra in our case, it is combined in a data matrix X when a supervised chemometric method is selected whereas in supervised methods an additional Y matrix is also created which contains the class information. The data matrix X is sometimes passed through some pre-treatment steps that involves data processing in order to standardize the data using approaches such as normalization, mean-centring, auto-scaling, baseline correction, signal filtering etc. In case of high dimensional data, reduction in dimensionality of data is performed prior to model development by projecting it onto a lower-dimensional space, retaining important information on the original variables that offers some sort of optimality properties. This can be achieved by selection of variables, also called *feature selection* or creating a reduced set of linear or nonlinear transformations of the input variables, also referred to as *feature extraction*. For this purpose, Principal Component Analysis (PCA) is one of the commonly used techniques to achieve the purpose. As PCA is an exploratory data analysis, it allows obtaining an overview of the data improving the performance of the method and sometimes used for variables reduction as well. When a supervised method is selected, training or calibration is done considering the Y matrix, and the model is optimized mostly by cross-validation and analysing the parameters such as root mean square error of calibration (RMSEC), Receiver-Operating characteristics (ROC), Area

under Curve (AUC) etc. Final step includes validation of the optimized model, which may be internal or external, and is performed using the samples not included in training.

2.1. DESCRIPTION OF CHEMOMETRIC METHODS USED

In the development of this thesis, the chemometric methods commonly employed in LIBS studies have been taken into account for the data analysis. Therefore, this chapter includes a brief description of these methods such as Partial Least Squares – Discriminant Analysis (PLS-DA), Linear Discriminant Analysis (LDA), Soft Independent Modelling of Class Analogy (SIMCA), Support Vector Machines (SVM) and Neural Networks (NN). Principal Component Analysis (PCA) has also been included because of its importance in relation to SIMCA. The figures of merits for qualitative classification to evaluate the performance of the models have also been discussed briefly in the end of chapter. The scientific literature consulted for this chapter includes the references [1,76,94,97,101,104,109-127].

2.1.1. Principal Components Analysis

Principal Component Analysis (PCA) is the most commonly employed exploratory data analysis and is generally used to display the structure in the multivariate data studying the relationship between variables and observations [104]. PCA itself is not a classification method, however, the models generated by PCA can be used to perform the classification by making other predictive models. It is a non-supervised technique that projects the data in a reduced hyperspace defined by Principal Components (PCs), retaining the maximum amount of variability. PCs are new variables obtained as linear combinations of the original variables, orthogonal to each other. The number of PCs obtained corresponds to the number of variables (dimensions) in the original dataset and

the largest variance is shown by the first principal component. Each of the successive PCs shows a continuous decrease in the variance. The variance explained by each principal component allows retaining a lower number of PCs as compared to the variables present in original matrix by selecting the ones that are more important to the model. This leads to data reduction by keeping maximum information of the original data set in least possible dimensions. Plotting these PCs allows visualizing the data providing the relationship between the variables and observations, revealing the hidden patterns and groupings in the dataset, distribution of samples, observing anomalous objects (outliers) and aliens. Usually 2D plots are generated that mostly take into account PC1 and PC2 as these are the components representing largest variance; however, for the purpose of discrimination other PCs may also serve useful.

The key concept of PCA is the transformation of the original matrix (Fig. 2.3) into a product of two matrices – score and loading and can be represented as:

$$X = TP + E \qquad \text{Eq. 1}$$

where *X* is the data matrix with experimental data with dimensions *I* X *J, I* being the rows containing the objects (spectra) and *J* represents the columns containing variables (intensity values at each wavelength); *T* is the score matrix with *I* rows and *A* columns (number of PCs); *P* represents the matrix containing loadings with *A* rows and *J* columns; and E represents the error matrix *(I* X *J),* consisting of residuals, representing the deviation of model with respect to original matrix X [128].

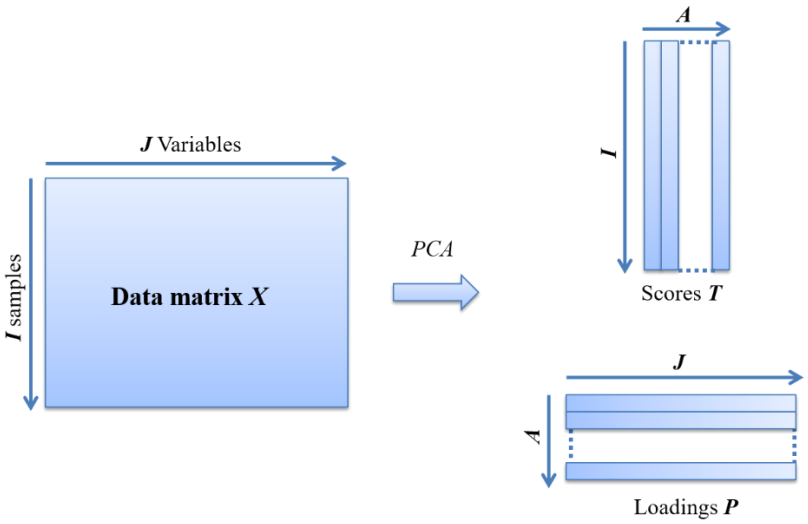


Fig. 2.3. Transformation of original matrix by PCA into scores and loading matrix

Each column of T, t_a, and each row of P, p_a, can be represented as a column vector that generate the PC, a, and can be given as:

$$x_{ij} = \sum_{a=1}^{A} t_{ia} p_{aj} + e_{ij} \quad \text{Eq. 2}$$

where i is the individual observations and j is the individual variables. The computation of PCs is done by using algortihms such as Non-linear iterative partial least squares (NIPALS) or more powerful methods that are based on matrix diagonalization such as Singular Value Decomposition (SVD), or bidiagonalization such as partial least squares [76].

2.1.2. Soft Independent Modelling of Class Analogy

Soft Independent Modelling of Class Analogy (SIMCA) is a supervised classification method that uses the modelling properties of PCA to generate a PC model (disjoint class models) for each class independently and calculating a critical distance with probabilistic meaning [117]. Fig. 2.4 shows the scheme for the generation of SIMCA model.

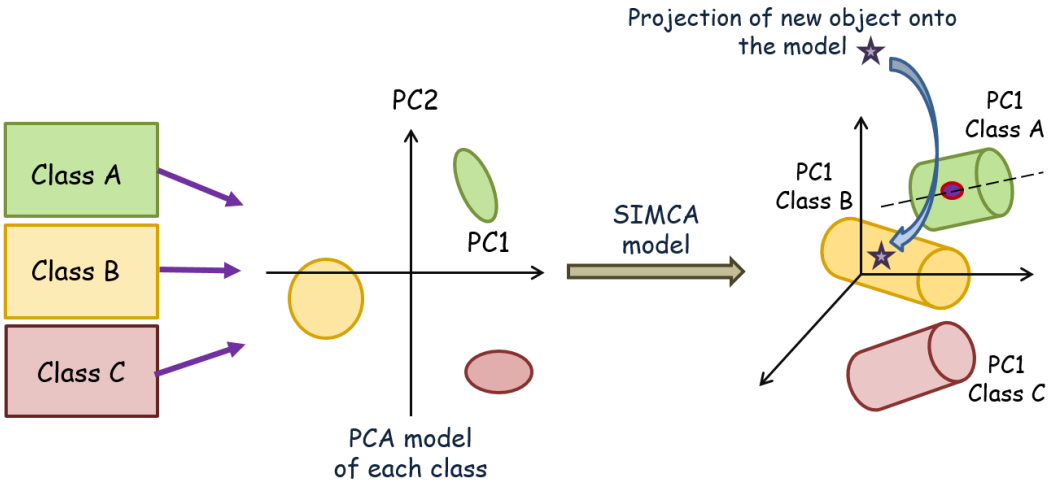


Fig. 2.4. Developing SIMCA model using PCA models and prediction of class membership of new object by the model

For each model the optimal number of PCs is selected independently. Since the shape and structure of each class is different this number can therefore be different for each class [116]. In order to optimize this parameter cross validation is performed and once every data point has been left out the Principal Component model with minimum error is selected. SIMCA generates a model taking into account the PCA models belonging to G classes, generating a space where each class is defined by a distinct confidence region at a given probability. The distance at which these regions or planes are situated depends on the spread of the data around the plane and can be given by the total residual variance of each class,

$$s^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{e_{ij}^2}{(p-d)(n-d-1)} \qquad \text{Eq. 3}$$

where e_{ij}^2 stands for the residual of each element, p is the number of variables, n is the number of objects and d is the number of principal components of each class. In order to find the membership of a new object, u, the object is projected in the class space of each of the disjoint models of G classes and the distance to each class is calculated.

Mathematically, this is done by calculating the scores and residuals with respect to each PCA model are calculated. This allows obtaining the standard deviation (s_u^2) for the object *u* as:

$$s_u^2 = \sum_{j=1}^{p} \frac{e_{uj}^2}{(p-d)} \quad \text{Eq. 4}$$

providing the information on object-to-model distance and hence the ability of each model to describe the newly projected data [116].

The quantity s_u^2 obtainedcan be compared to the class variance (s^2) using the *F*-ratio.

$$F = \frac{s_u^2}{s^2} \quad \text{Eq. 5}$$

If *F*-ratio > *F*-ratio$_{(critical)}$ at a given level of significance, the distance of object u is larger than that of the class as a whole. It can be given that,

$$s_{lim}{}^2 = s^2 . F^{0.05} \quad \text{Eq. 6}$$

If $(s_u)^2 \leq (s_{lim})^2$, the object *u* is assigned as belonging to a class [76,94,109,111,112].

2.1.3. Partial Least Square Discriminant Analysis

Partial Least Square Discriminant Analysis (PLS-DA) is a supervised linear method which comprises of partial least squares regression approach in combination with discriminant analysis [104,109,110]. PLS is a well-known dimension reduction method and has been extensively used in the past for quantitative analysis however it has been adapted to perform the high dimensional classification tasks by the application of Discriminant Analysis [109]. PLS algortihm is an evolution to PCA, the difference being that PLS is applied to a dataset divided into two matrices, X and Y [115]. The matrix X contains variables forming the experimental data and Y matrix contains information on the membership of samples to the G classes with rows (samples) and G columns (class information, response, calibration data and/or observed outcome) [100]. Each value y_{ig} of

matrix Y, expressed with categorical (dummy) variable, assigns class membership *"g-th"* to the *"i-th"* sample of X matrix.

The PLS regression algorithm tends to reduce the data dimensions using PLS1 when only one dependent Y variable is involved and PLS2 when several dependent Y variables are being dealt. The algorithm searches for latent variables *LV* (orthogonal linear combinations of the original variables) with a maximum covariance with the Y variables. The weight vectors that are used for the computation of LV are called loadings. The PLS2 algorithm transforms X and Y matrices as given in equations 7 and 8, and constructs a relationship between both matrices.

$$X = TP + E \quad \text{Eq. 7}$$

$$Y = UC + F \quad \text{Eq. 8}$$

where T and U are scores, P and C are the loadings, and E and F are the corresponding residuals. The number of LV taken into account to generate the model is selected by cross-validation. The developed PLS2 regression model is then calibrated on the Y matrix, calculating the value y_{ig}^{calc} for each *i-th* sample for each *g-th* class, which allows estimating the probability of an object to each of the G classes. The value of y_{ig}^{calc} lies between 0 and 1, where values close to 0 means no match and close to 1 means belonging to a class. In order to assign the membership of an object, a threshold is defined for each class based on the Bayes theorem, and the sample is assigned to its corresponding class [97,113,114].

2.1.4. Linear discriminant analysis

Linear discriminant analysis (LDA) is a supervised linear classification method, which is based on finding the plane of projection or directions (linear discriminants) that allows maximizing the variance between classes by maximizing the distance between the

mean of the classes and minimizing the variance within the class [117]. This hyperplane is useful in the classification tasks. LDA bases on the calculation of orthogonal linear discriminant functions in a number equal to the number of classes minus one. The variables significant for the construction of a classification model are evaluated using Wilk's lambda selection criterion, and an *F*-statistic to determine the significance of the changes in lambda when a new variable is tested [117]. It is a parametric method and therefore same spread and normal distribution is assumed for all classes. Moreover, the covariance matrix for all classes (S_b) is considered homoscedastic, and S_w and S_b are intra-class and inter-class variance [94,112,113] given as:

$$S_p = \frac{\sum_{i=1}^{n}(n_{k-1})S_g}{k-n} \quad \text{Eq. 9}$$

$$S_b = \sum_{i=1}^{n} n_{k(\mu_i-\mu)(\mu_i-\mu)^T} \quad \text{Eq. 10}$$

$$S_w = \sum_{i=1}^{n} \sum_{x \in c_i} (x-\mu_i)\left(x_j-\mu_i\right)^T \quad \text{Eq. 11}$$

where *n* stands for number of classes; n_k, number of objects in each class; *k*, total number of objects; *x*, object in a class; μ_i, mean of a class; and μ is the global mean of all classes. The weights of the linear discriminant functions (λ, eigenvalues) are found as the eigenvectors (v) of the following matrix:

$$S_w^{-1} S_b \, v = \lambda \, v \quad \text{Eq. 12}$$

The eigenvector, v_1, with the highest value of λ, provides the first linear discrimination function. With the residual x data, the second largest discriminant function is calculated and in the same way all the DFs needed to solve a problem are found. In order to find the membership of a new object, it is inserted into DFs in order to transform its coordinates in the same way as for original data set and then the object is assigned to the class for which the Euclidian distance is the smallest [58,129],.

The limitations possessed by LDA include singularity problem i.e. it fails when dealing with singular scatter matrices. This implies the need of having the number of variables equal to the number of objects in matrix X, and therefore a previous reduction in data dimension is required. Mostly PCA is used for this purpose, however increases the costs of analysis in terms of time and space due to the Eigen decomposition involving the scatter matrices [108]. Secondly, being a pure classification technique and parametric method always assigns a class membership to the object even if it does not belong to any of the classes in target matrix Y, therefore giving false positives [109].

2.1.5. Support Vector Machine

Support Vector Machines (SVM) is a machine learning method, enters in a new generation of classification methods, and follows the supervised learning approach. The key feature of SVM is the formulating the decision boundaries between the classes which can be linear or can become non-linear when dealing with a situation where boundaries are complex. The decision boundaries are calculated by minimizing the empirical error from the training dataset.

In case when the samples *x* (classes) have membership to two classes given by y = ±1, the simplest approach can be a separation by a linear decision boundary (hyperplane) given by:

$$y(x) = sgn(w_c^T x + b_c) \quad \text{Eq. 13}$$

where ($w_c b_c$) are the corresponding weight and bias parameters determined from the training set. A linear boundary can work for simpler cases; however, in a complex situation where a linear decision boundary cannot produce separation between groups, SVM uses kernel function defined as:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad \text{Eq. 14}$$

where K is the Kernel function, x_i and x_j are the training sets. Different kernel functions are used for the transformation, such as Radial Basis Function (RBF), Polynomial and Sigmoidal Functions, and RBF is most commonly used. Radial basic function (RBF) can be given as:

$$K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad \text{Eq. 15}$$

where σ defines implicitly the non-linear mapping from input space to feature space [130,131]. It works by mapping the coordinates of the objects (original input space) into a new transformed high dimensional space (feature space) where the samples are projected by means of feature function Φ(x) that maximizes the margins [130]. When the hyperplane from the new feature space is projected back to its original input space, it results in complex non-linear decision boundaries producing separation between the classes (Fig. 2.5). The location of the decision boundaries is directly affected by some data points (support vectors), which lie closest to the hyperplanes and therefore determine the separation between the training sets [119-121].

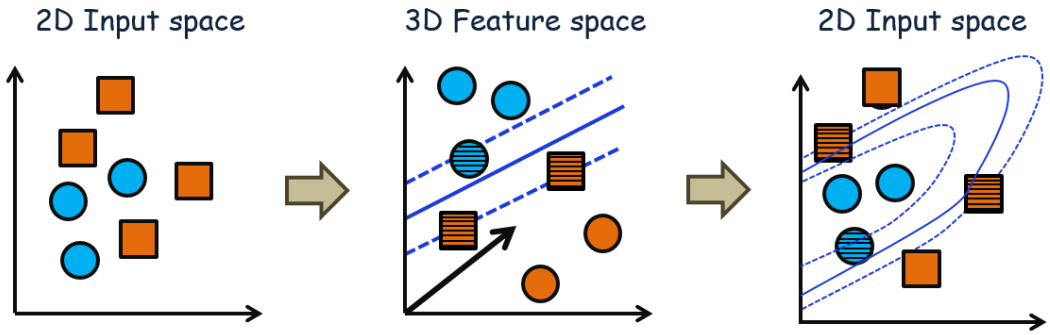


Fig. 2.5. Creation of the boundary for a non-separable case. Projection of data into a high dimensional space by SVMs searching the optimal separating hyperplane. Squares represent the support vectors.

2.1.6. Neural Networks

Neural Networks (NN) are computational models that are motivated by the structure of biological nervous system, defined by a set of parameters that enable them to perform certain tasks such as pattern recognition and generating classification models [76,122,123 ,124,126,132,133]. NN are made up of simple elements "neurons" highly connected to each other and arranged in layers where the input layer receives the signals, the output layer represents the results of computation and between these layers might be present additional layers of neurons called hidden layers. The connections between the neurons are controlled by numerical factors that are called weights *w*. A simplified version of NN computation is presented in Fig. 2.6 showing the operation of a single neuron *j*.

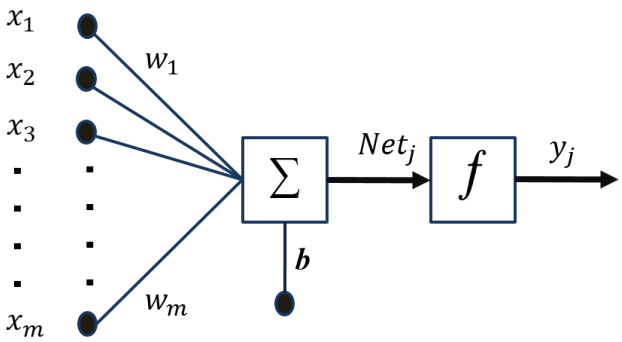


Fig. 2.6. Operation of a single neuron

The input signals are received by the neuron that are scalar inputs x_i (i = 1,…, m) which are multiplied by the scalar weights w_i (i = 1,…, m) to form $x_i w_i$ one of the terms that is sent to the sum function. Each neuron has a bias b that is also passed to the sum function. It is also a numerical value just like weights but has a constant input of 1. The sum funtion gives the summation of these inputs which is often referred to as the net input Net_j.

$$Net_j = \sum_{i=1}^{m} x_i w_i + b \quad \text{Eq. 16}$$

A transformation is performed on Net_j by a transfer function (activation function), producing a scalar neuron output y_j.

$$y_j = \int Net_j \quad \text{Eq. 17}$$

The output of a neuron (or neural network) depends on the kind of transfer function chosen. Different types of transfer functions are used, mainly classified as linear, threshold and sigmoid functions (Fig. 2.7). Apart from linear transfer functions all other transformations lead to a non-linear output signal.

Linear transfer function: The output of the neuron is directly proportional (or equal) to the net input Net_j.

Hard limit Threshold function: The output is set at two levels, 0 or 1, and depends on whether the argument function comes out to be greater or smaller than the fixed threshold value. The output is 0 if $Net_j < 0$, or 1 if $Net_j \geq 0$.

Log Sigmoid function: The output of the neuron keeps varying with the input in a non-linear way. The input which may lie between + and – infinity is squashed to produce an output in the range of 0 to 1, according to $\int Net = \frac{1}{1+e^{-c.Net}}$. This function is most commonly used particularly in multilayer network topologies that are trained using backpropagation algorithm and also because it is easily differentiable. Some applications also employ a *hyperbolic tangent sigmoid transfer function*, yielding output values in the range [-1, +1]. Once a transfer function is chosen the parameters of the neural network are adjusted by employing some learning rule in order to establish an optimum input – output relationship in the neuron to achieve a specific goal. In addition to transfer function, the architecture and learning rule for the training process are the parameters that define a network.

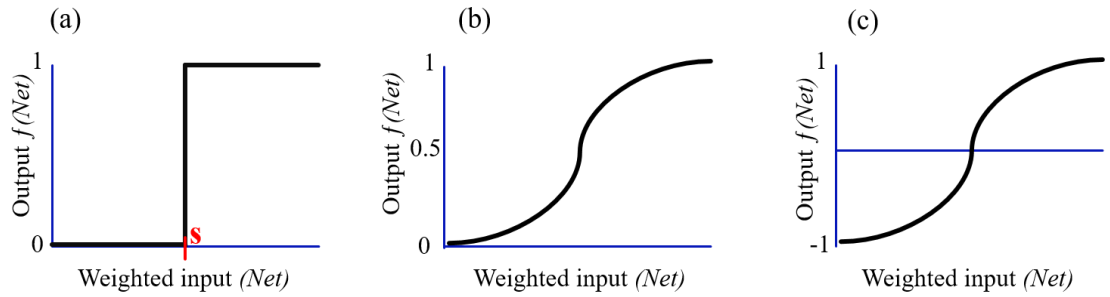


Fig. 2.7. Transfer functions for a) Hard limit Threshold function,"s-threshold", b) Log Sigmoid Function and c) Hyperbolic tangent sigmoid function

2.1.6.1. Architecture of Neural Networks

The two layered network model, perceptron, made up of only input and output layer, was the earliest model studied. The perceptron model presents limitations in lacking flexibility therefore multilayered networks are mostly employed where the interconnections between layers increase the flexibility to transform the inputs for each layer of neurons. In a multilayered model (as an example Fig. 2.8), all neurons of a layer are connected to all neurons of the previous and next layer, and each neuron j in the layers receives the input data X ($x_1, x_2, \ldots\ldots, x_m$), transformed by a matrix of weights W with elements w_{ji} and transfer function generating an output value y_{ji} specific to this neuron. The output $\boldsymbol{Out}^{\boldsymbol{l}}$ from each layer ***l,*** is fed to the next layer as an input, that arrives with a different value as each connection between neurons of successive layers is associated with a different weight $\boldsymbol{w_{ji}}$ factor, the magnitudes of which are optimized by the learning process.

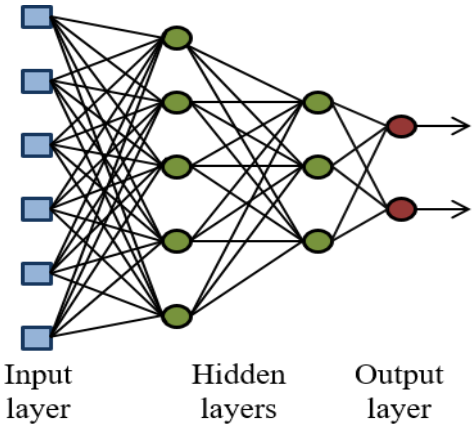


Fig. 2.8 Neural Network with input and three layers of active neurons

The architecture of the network and connections between the neuron of consecutive layers have a great influence on the performance of an NN model. There is a wide range of architectures available that are mainly classified as:

Feedforward Neural Networks that allow the information to flow in only forward direction, i.e. from input to output layer. Pattern recognition tasks most commonly rely on this type of architecture. These networks offer multilayered structures with one or more hidden layers; input layer consisting of neurons equal to the number of variables in data matrix; and output layer containing the neurons same as the number of classes to be assigned. The number of hidden layers are selected depending on the system´s complexity and influence the transformation on the data matrix to find the desired classification results between the objects based on the similarity between them.

Recurrent Neural Networks that have feedback mechanism where the outputs of the neurons are connected to its inputs. In this topology each neuron in addition to being connected to the neurons of previous and next layers is also connected to itself as well as the neurons of the same layer. The connections are similarly controlled by weights that are determined by the learning process. These networks are potentially more powerful

than feedforward networks but at the same time exhibit complicated dynamics that make the training process very difficult.

Symmetrically connected Neural Networks are similar to the recurrent networks except that the connections between neurons are always symmetrical, i.e. the magnitude of weights for each connection between the layers is same in both directions. These networks are less powerful than recurrent networks but offer simplicity in operation and learning algorithms. Symmetrically connected networks can be distinguished into Hopfield nets (with no hidden layers) and Boltzmann machines (those having hidden layers) where later is more powerful due to the transformations by the hidden neurons.

2.1.6.2. Learning paradigms

The neural networks have the ability to store a set of patterns and memorise information, which can be used for relating the information between patterns and responding to the new one presented to the network by revealing whichever pattern lies more closely. This ability is called associative memory and stores the patterns through the weights associated with the neurons. Its function is to recognize previously learned input vectors. The association of patterns can be distinguished into two types, *autoassociation* (recognizing two objects belong to the same class even when one of the patterns is distorted) and *heteroassociation* (recognizing that one pattern is related to another object).

Neural networks capture hidden information in the data, through a process called "learning" or "training", which mathematically refers to adjusting the weights between the neurons so that some conditions are fulfilled and desired classification results are achieved. NN can be used in reinforcement, unsupervised or supervised learning approach.

Reinforcement Learning: It involves the processing and transformation of input matrix into output resul providing a feedback based on how well the system performed the computation. The feedback given reflects the correctness of transformation in terms of an an extra reiforcement signal, which rewards or penalizes the network depending on its performance. Using trial and error method the network keeps computing in order to minimize the penalties and maximize the rewards until it achieves correct transformations [134].

Unsupervised learning: In this type of learning, the desired outcome is not known and the network is trained with only input data matrix. Using the information in the input data the system performs autocorrelation finding patterns within the data and settles down after some number of iterations grouping the objects based on their features. This approach is useful in data exploration and data reconstructions. Kohonen self organizing maps (SOM) is a more typical example using this type of learning, working similar to PCA; however SOM offer the advantages in providing advanced visualization, establishing non-linear relations and can be applied to a large dataset.

Supervised learning is controlled by the user where in addition to the input data matrix X consisting of the patterns, the network receives the information about the clases in form of a target Y. The network then categorizes the patterns on the basis of this predefined set of classes that helps to adjust the weights, teaching the network to reach to an output which can associate input and output target data. Once the network is trained it is should be able to generalize, i.e. classify the patterns that are new and unknown but similar

enough to the examples used used in the training. Multilayer feedforward networks are the common examples of this type of learning.

2.1.6.3. Supervised Learning Algorithms

In supervised learning each input value, x_i, is coded by the target value, y_t, that corresponds with the input vector forming training pairs (x_i, y_t). The learning process consists of using a set of randomly guessed synaptic weights and calculating an output y_j, which is then compared with the target output Y. As learning is an iterative process involving epochs, using the learning algorithm, it keeps updating and adjusting the parameters and outcome of the network with each epoch to minimize the error by applying a correction value (δ) until an optimum set of synaptic weights that fit the neural network outputs to an input-target data set has been achieved in order to perform the required task such as classification.

The algorithms that use this type of learning are called supervised learning algorithms and are given as follows:

Perceptron Learning Algorithm: The perceptron learning rule classifies the input vectors that are linearly separable by a hyperplane whereas many classification tasks include the problems that are not separable linearly and there the perceptron model exhibits limitations to solve such problems. Although the basic perceptron learning rule presents limitations it provides the basis for understanding the lraning rules in more powerful networks.

Delta rule or Least Mean Square Error method: used by ADALINE or MADALINE networks. Delta rule is similar to perceptron learning rule with the difference that it can be derived for any differentiable output activation function whereas perceptron learning rule only uses threshold output function. It works by choosing an error function that allows the system to estimate the the difference between the network output and target

output and minimze the error. This difference can be quantified by a function given by eq. 17,

$$E(W) = \frac{1}{2}\sum_{p}\sum_{j}\left(y_t^p - y_j^p\right)^2 \qquad \text{Eq. 18}$$

where *W* represnts the weights in the neural net adapted for any set p of known objects. Eq. 17 can be given in a normalized form as:

$$MSE = \frac{1}{2PN}\sum_{p}\sum_{j}\left(y_t^p - y_j^p\right)^2 \qquad \text{Eq. 19}$$

where MSE is mean square error, P is total number of training patterns and N is the number of neurons in the output layer.

The Generalized Delta rule: used by multilayered feedforward networks in the backpropagation procedure for learning internal representations of the network.

2.1.6.4. Gradient descent method

The learning algorithms use the principle of gradient descent for the minimization of mean square error. As the objective of learning is to achieve an optimum set of weights and minimize the mean square error, the gradient descent algorithm works by finding a "global minimum" of the error function within solution´s vector space that is a hyperparabolic surface. The global minimum represents the minimum error value encountered, and at this point the outputs agree with the target output values and learning is said to be converged to a solution.

In this method, as shown in Fig. 2.9, a gradient is defined in terms of an error function E(W) that keeps indicating the error variations as a function of set of synaptic weights (W), tweaking the weights in every epoch. The gradient descent algorithm takes small calculated steps in the direction negative to the gradient of the function along the steepest gradient of the error surface.

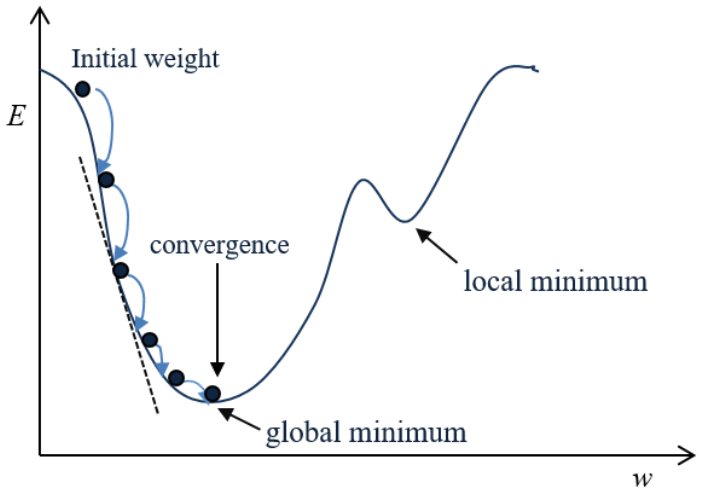


Fig. 2.9. Decrement of error along the steepest curve in the hyperbolic error space according to the Gradient Descent method

The algorithm continues to obtaining a decrement in the error until a global minimum is achieved, according to the Eq. 18:

$$W(t+1) = W(t) - \alpha E(W) \qquad \text{Eq. 20}$$

where α represents the learning rate i.e. the size of each step in each epoch. This is sometimes called delta rule since it bases on the difference between target and network output and adjusts the weights so as to minimize the mean square error.

The error surface is hyperbolic and rough in nature therefore an infinitely small value of α will lead to entrapping in a local mínimum whereas a greater value results in jumping around the global mínimum without reaching to the best solution. α is a critical parameter as these training parameters are not not a priori therefore several iterations take place before the system reaches the optimum state.

2.1.6.5. Backpropagation

Supervised neural networks make use of a wide range of learning algorithms however back propagation is the one most commonly used for the generation of

classification models which relies on the delta rule since backpropagation requires the activation function by the network to be differentiable. As the original perceptron learning procedure does not apply to more than one layer, therefore, backpropagation rule that is infact a generalization of the delta rule was developed to find a way to train multi-layered networks. Backpropagation is a special learning process using chain rule based on the fact that adjustments to the neural net's weights can be calculated on the basis of well-defined equations. In a multilayered network, the input data passes through the hidden layers in such a way that each of the hidden layers produces an output Out^{l} which serves as the input of the neurons in the next layer eventually providing an output Out^{last} by the final layer. A comparison between Out^{last} and target Y matrix determines the error ε between the neurons outputs which is then used to correct the weights in the output layer. Derivation of the error gives rise to the correction value δ required for the adjustments of the weights Δw_{ji}^{last} in the final layer. The chain rule leads to the following equations.

$$E = \sum_{j=1}^{m} \left(y_t - out_j^{last}\right)^2 \qquad \text{Eq. 21}$$

$$\delta_j^{last} = -\partial E / \partial Net_j^{last} \qquad \text{Eq. 22}$$

$$\Delta w_{ji}^{last} = \eta \delta_j^{last} out_j^{last-1} \qquad \text{Eq. 23}$$

Once the weights for the final layer are corrected, the weights in the penultimate layer are corrected with regard to the error from the final layer. Similarly the error is determined and fed back to the preceding layers one by one correcting the errors and adjusting the weights, flowing in the direction counter to the input data, hence called back propagation [135] (Fig. 2.10).

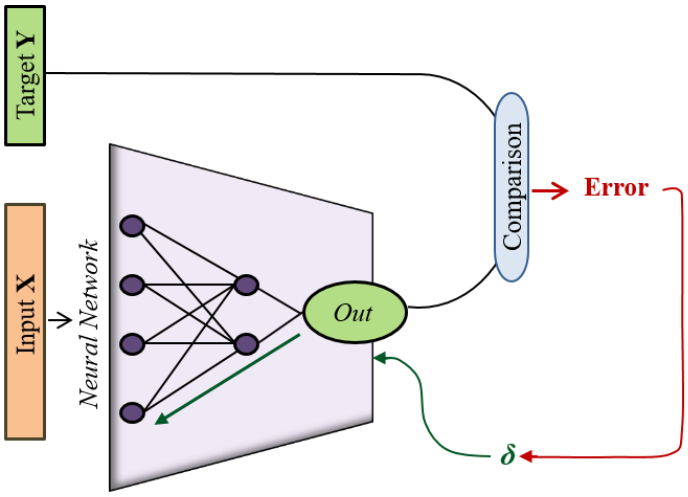


Fig. 2.10. Learning process of Neural Networks by Feedforward computation and back-propagation of errors correcting the weights

Eq. 23 can be generalized for weight adjustments of any single layer of the network as:

$$\Delta w_{ji}^{l} = \eta \sum_{k=1}^{r} \delta^{l+1} w_{jk}^{l+1} f'\left(Net_j^l\right) out_i^{l+1} \qquad \text{Eq. 24}$$

where η is the learning rate. In the process of errors and weights correction, the adjustments from the previous cycle are also considered, which is accomplished by extending the Eq. 22 by the contribution of $\mu \Delta w_{ji}^{l(previous)}$ where μ reprsents the momentum of learning. The parameter η determines the speed of convergence between the networks´s current solution and the global minmum whereas μ determines the extent to which the weight adjustments from the previous cycle are considered. The momentum factor renders a certain amount of inertia to the learning process helping the network to avoid entrappment in a local minima and provide the solution at or close to global miniumum.

2.1.6.6. Generalization

Generalization is the ability of a classification technique to recognize or classify correctly new patterns presented to it once it has been trained sufficiently and at this point it can be employed as an analytical tool. During the training or learning process the neural networks establish associations between input and output data extracting essential features from the training set which can later enable to predict the pertinence of new data not presented to the network before. The input-output data relationships are established by weight adjustments made in the network based on the training error as given by equations previously however in order for the network to be able to generalize an error of generalization called validation error is determined. To do this the input data set is divided into two sets: a training subset used in the learning process to compute gradient and adjust weights, and a validation subset to test the generalizability. Training is stopped when the network provides good results on both the training data and the validation data obtaining a minimum error for the validation (Early stopping criteria). However, it is also possible that the network instead of extracting the essential and representative features of the input data, memorizes it so specifically that it has lost the the ability of generalization. This is called overfitting and is characterized by an increase in the validation error during the training process. Such network is an unwanted situation in classification tasks in real word problems.

2.2. Evaluation of classification performance of Chemometric Methods

Classification or pattern recognition involve fundamental chemometric techniques to establish mathematical models able to assign an object defined by a set of features to one of the classes previously learned by the model based on a set of measurements. The developed classification model therefore serves as an analytical tool for the analysis of

test samples, where each object can be defined as belonging or not to a class in question based on its characteristic features. The outcome of the classification model for an object in an identification process can be associated to a qualitative binary response, yes (Positive) or no (Negative). This binary classifier therefore can provide four possible results as given by the confusion matrix in Table 2.1. TP is the case when the outcome of a prediction and the actual value are both positive and FP occurs when prediction is positive but the actual vaue is negative. TN occurs when the outcome of prediction is and actual value are both negative whereas FN occurs when the predicted value is negative but the actual value is positive [136, 137].

Table 2.1. Confusion Matrix for a classifier

Predicted	**Actual**		
	Positive	*Negative*	
Positive	True Positive (TP)	False Positive (FP)	R_p
Negative	False Negative (FN)	True Negative (TN)	R_n
	C_p	C_n	N

The C_n and C_p represent total number of truly negative and positive observations while R_n and R_p are total number of predicted negative and positive observations and N is the the total number of observations ($N = C_n + C_p = R_n + R_p$). The confusion matrix allows extracting meaningful information obtaining several parameters that can used for the evaluation of the performance of classification models such as correct classification rates measured in terms of True Positive rate (TPR) and True Negative Rate (TNR), incorrect classification rate defined by false positives (FP) or false negatives (FN), sensitivity, specificity, accuracy, selectivity, generalization ability and robustness.

Sensitivity (*S*) of a classification model also called the true positive rate (TPR) can be given by:

$$S = \frac{TP}{TP + FN} \quad \text{Eq. 25}$$

When $S = 1$ it indicates that all samples are correctly identified.

Specificity (Sp) determines the ratio of fault-free recognitions and can be calculated as:

$$Sp = \frac{TN}{FP + TN} \quad \text{Eq. 26}$$

Moreover, False Positive rate (FPR) gives the ratio of objects erroneously classified, where FPR = 1 − SP. The sensitivity and specificity terms in chemometrics are used in a context different from the analytical chemistry.

A model´s classification ability cannot be represented in terms of sensitivity or specificity alone hence the combination of both metrics is needed. Therefore, accuracy (A), which involves the relative frequency of correct and incorrect classification is the main parameter used for the estimation of the efficiency of a model, calculated as:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{Eq. 27}$$

and can be therefore defined as the ratio between number of correct predictions by the model and total number of predictions on a set of data. Although most commonly used, accuracy is not considered as a parameter sufficient to evaluate a model alone [138] and other metrics such as Receiver Operating Characteristic (ROC) and Area Under Curve (AUC) are suggested by statisticians.

Receiver Operating Characteristic (ROC): ROC plot analysis was developed in the context of electronic signal detection in the early 1950s [139] and since then this metric is extensively used in machine learning as a primary indicator to evaluate the

performance of a model and its visualization in order to select a suitable operating point, or decisión threshold [140]. For its construction sensitivity (S) and specificity (SP) are calculated for consecutive cut-off points with respect to the predicted probabilities from the models [141] and ROC graph is plotted with TPR (S also called recall) on y-axis and FPR (1−SP) on x-axis. ROC plot serves as a tool to plot all possible combinations between sensitivity and specificity for classification problem. An ideal ROC plot would be obtained at the operating point with TPR = 1 and FPR = 0. ROC curves for four classifiers are shown as an example in Fig. 2.11 where classifier I and II seem to dominate III; however, in many cases it becomes difficult to mark a clear domination between different curves. In such cases, area under the ROC curve (AUC) provides a good "summary" to evaluate a model or compare between different classifiers.

Area Under Curve (AUC): It is widely suggested that the area under curve (AUC) provides a better measure than accuracy for evaluating the predictive ability of the classification [142, 143]. The use of the area under ROC curve as a performance metric exhibits several desirable properties over the conventional overall accuracy in presenting increased sensitivity in Analysis of Variance (ANOVA) tests, Independence of the decision threshold, invariant to a priori class probability distributions [138,140].

AUC can be calculated [138,144] according to the eq.:

$$AUC = \frac{S_0 - n_0\,(n_0 + 1)/2}{n_0 n_1} \qquad \text{Eq. 28}$$

where n_0 and n_1 are the numbers of points in the test set belonging to class 0 and 1 respectively, and $S_0 = \sum r_i$. The probability (G) is calculated at a threshold *t* that is

randomly chosen according to the distribution F is calculated for class 0 and 1 points. According to the definition of ROC curve, AUC is given as:

$$\int G(u)dF(u) = \int G(u)\, f\ (u)du \qquad \text{Eq. 29}$$

can be estimated by replacing the G and f theoretical functions by the observed distributions $g_i and\ f_i$ of sample values and arranging them gives the Rank r_i. AUC with an area of 1.0 represnts an ideal case where all the predictions were made correctly. Fig. 2.11 shows a plot of three ROC curves, each representing one of the three classifiers.

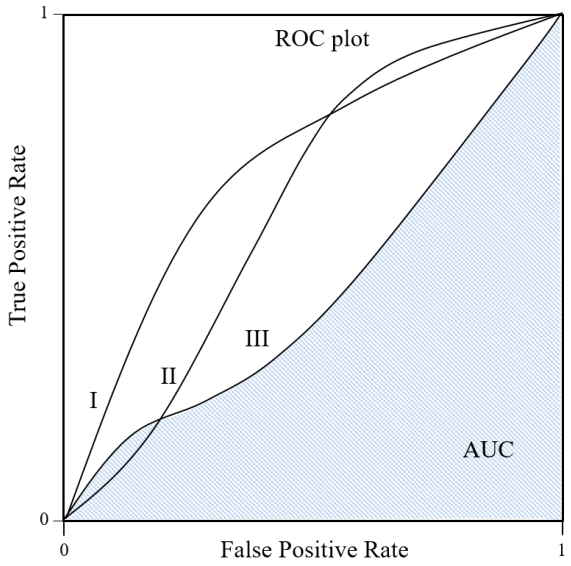


Fig. 2.11 shows a plot of four ROC curves, each representing a classifier [138]

In this example, I and II dominate III, however, often there is no clear dominating relation between two ROC curves I and II as they are not dominating each other in the whole range. In such cases, the area under the ROC curve, or simply AUC, provides a good summary for comparing the two ROC curves [138].

Validation tests

In addition to the metrics mentioned above to evaluate a classification model, different types of validation tests can be performed in order to use it for real applications. These include:

Sensitivity: ability to classify correctly the samples that were included in training to construct the model.

Generalization ability: classification of samples of the same class not included in the training set.

Robustness: ability of system to estimate the correct result when a sample from a class not included in the training set is tested by the model. Such a sample that does not belong to any of the classes included in training set must not be classified as false positive to any class and should be predicted as "unknown" by the model. The test of robustness is performed using Leave One Out (LOO) procedure where one dataset (class) is removed from the from the training set and the model´s ability to "see" it as "unknown" is evaluated. This is a positive identification for the statistical analysis as a classification model must be able to recognize foreign objects and not misclassify them [41,42]. The success of a classification model lies in its ability to present high sensitivity, high generalization ability and high robusteness [58].

Chapter 3

Experimental – LIBS Instrumentation

LIBS has been mainly a subject of academic laboratories research for long and the continuous efforts to increase the efficiency of LIBS as an analytical technique have resulted in a great advancement in the components of LIBS system in order to perform highly sophisticated experiments. However, the past years have seen a successful transference to the industrial applications with the availability of compact commercial LIBS instruments on the market, offering direct measurements [62]. The set-ups used in LIBS experiments whether on the academic or commercial level share the same basic components but depending on the analyte of interest and its concentration, type and state of sample and its physical and chemical characteristics, the kind of analysis (qualitative or quantitative), environment, distance from the sample, the need to maintain some specific ambient conditions among others, the components and their specifications are tailored to the particular application [3]. A typical LIBS instrumental set-up consists of a *laser system* generating pulsed beam that is focused on a sample placed on a *moving stage* that acts as a sample holder. The pulsed laser beam is directed and focused onto the sample using *optical system*. The emission from the plasma is collected and taken by another *optical circuit* to the *spectrograph* that is coupled with a *detector* and finally a *computer* that controls the whole system and stores spectral data.

3.1. LIBS SET-UP

The assembly of LIBS components and most of the experimental parameters were kept the same for the measurements of all microorganism samples, however, depending on the samples, sample manipulation and emissions observed, different spectrometers were used for bacterial and fungi samples. The scheme of experimental set-up for LIBS measurements is shown in Fig. 3.1 and the description of the components forming the LIBS set-up is mentioned in the coming text.

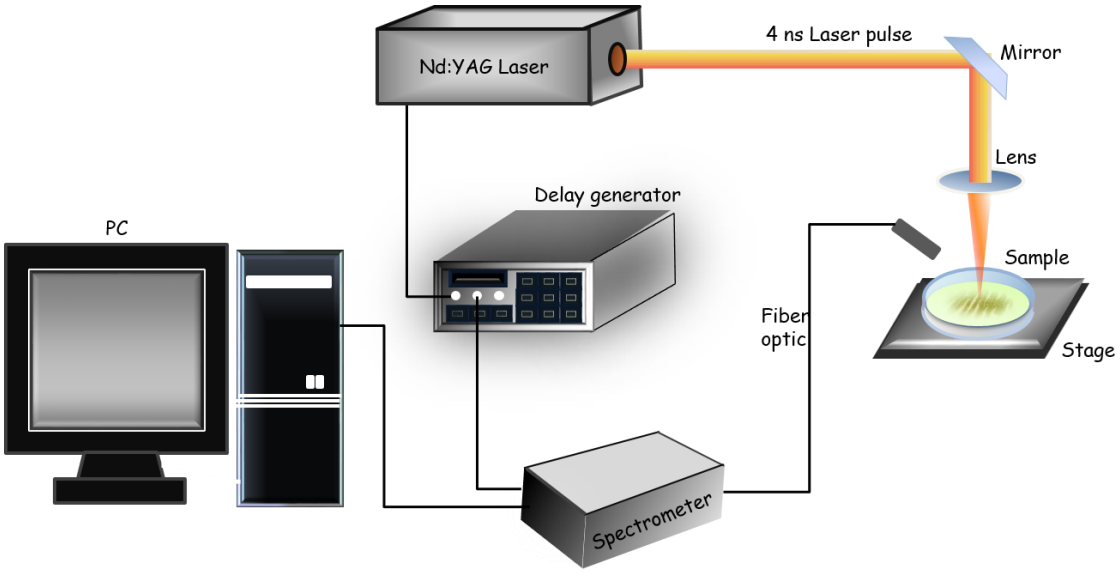


Fig. 3.1. Scheme of LIBS set-up for the study of samples in this thesis

3.1.1. LASER SYSTEM

The laser system used as excitation source for all measurements was Q-switched Nd:YAG laser, Brio model commercialized by Quantel, operating at its fundamental wavelength of 1064 nm. The temporal profile of the laser beam or the pulse duration at Full Width Half Maximum (FWHM) was 4 ns with a beam diameter of 4 mm and a

divergence of 0.6 mrad. Although the energy of the laser pulses could be varied, the laser energy was set to 20 mJ for the experiments.

3.1.2. Construction and working of Nd:YAG laser

Nd:YAG laser is a solid state laser where the rod Yttrium aluminium garnet ($Y_3Al_5O_{12}$) crystal acts as a host and is doped with 1% of Nd^{+3} ions that act as gain medium responsible for lasing. The energy levels of Nd^{+3} ion make Nd:YAG a four level laser. The optical pumping is done by a lamp of noble gas mostly Krypton or Xenon to get the population inversion, the latter was present in our laser system. The pumping of Nd:YAG laser occurs at wavelengths of 800 – 820 nm, causing the transition from the ground state into the absorption bands. This transition is followed by a fast non-radiative decay in the order of nanoseconds to the upper laser level "$^4F_{3/2}$" which has a longer lifetime, the metastable state, resulting in the population inversion. Transition from this level to the lower laser level "$^4I_{11/2}$" leads to the production of laser radiation at 1064 nm, the fundamental line of Nd:YAG laser, followed by another fast non-radiative decay to the ground state "$^4I_{9/2}$". In addition to this fundamental wavelength in the IR region, Nd:YAG can be operated in visible region at second, third harmonics at wavelengths 532 nm, 355 nm and UV region at fourth harmonic 266 nm respectively by using internal frequency doublers, depending on the sample and application. Nd:YAG laser can be operated in Continuous wave or pulsed mode, where pulsed mode is used in LIBS experiments because of the high power generated required to cause sample breakdown. The pulsed mode is typically obtained by Q-switching, generating high power laser pulses required for the ablation of sample in the range of 4 – 20 ns and power in the order of MW. Nd:YAG lasers are the most commonly used ones in LIBS because of the possibility of using at different harmonics, compactness, reliability and easy maintenance. The

working principle of Nd:YAG laser and the laser system used in the study have been shown in Fig. 3.2.

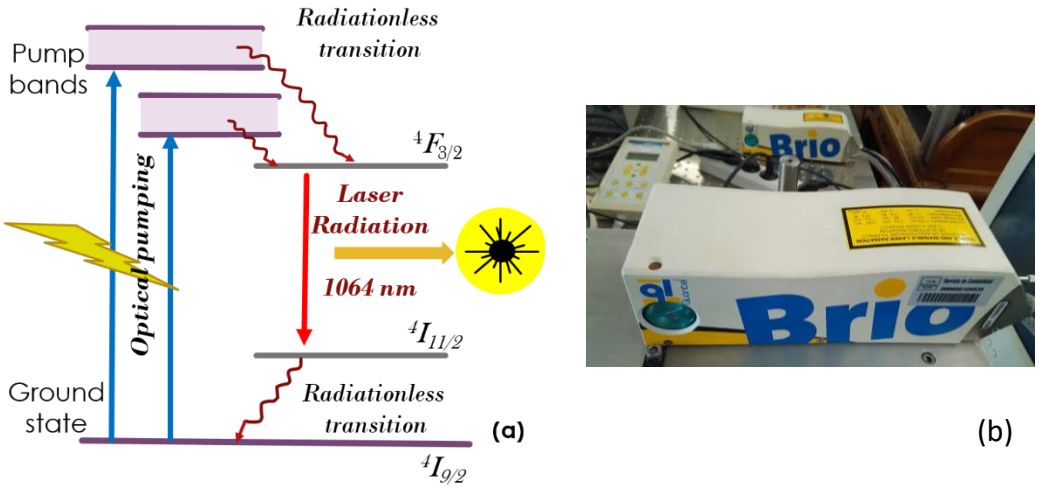


Fig. 3.2. (a) Scheme of the Nd:YAG laser (b) Laser system used in the study

3.1.3. Positioner/Stage

Samples were placed on a translational stage perpendicularly with respect to the laser beam and hence the beam had to be directed to the sample using optical components. The sample stage or micro-metric positioner allowed movement in X – Y – Z axes with a displacement of 0.5 μm at every coordinate to ensure that each laser pulse impinged on a fresh sample. The dimensions of the stage was sufficient to allow an easy manipulation of the sample in petri dishes.

3.1.4. Laser Focusing and plasma emission collecting optics

As the sample was placed perpendicular to the laser beam a specially coated mirror for infrared radiation was used to deflect the laser beam onto the sample. The laser beam was then focused on the sample by using a Plano-convex lens of 100 mm focal length to a fine spot of micrometric range on the target.

The second optical circuit served for the collection of the light emitting from the plasma. The LIBS set-ups make use of lens or mirrors or alternatively a fibre optic cable (FOC), placed at an angle to the plasma. The use of FOC provides advantage in simplifying the light collection and taking it to the spectrometer, especially when samples or detection system are at long distances and therefore FOC coupled with a collimator was used in all of our experiments.

The collimator was 7 mm, made of fused silica and with an aperture of 4 mm that collected the emission from plasma. It was placed at an angle from normal and a few centimetres distance from the sample. The collected light was then focused into the FOC that was 1 m length with a core diameter of 1000 μm and a numerical aperture of 0.22, which was coupled to the entrance of the spectrometer.

3.1.5. Spectrometer

The spectrometer consists of a resolution device (spectrograph) that resolves the incoming light into its component wavelengths which is read out by a detector (camera) presenting this information as a function of wavelength in the form of a spectrum on a computer. Hence the important features of a spectrograph that are considered while performing a LIBS analysis are the spectral range and the resolution i.e. the minimum wavelength separation that can be achieved between two emission lines, as these are essential factors in determining the quality of an experiment and accurate identification of constituent elements in a spectrum.

There exist various configurations of spectrographs such as Littrow, Ebert–Fastie, Czerny Turner and Echelle, depending on the number and mounting of the mirrors, prisms and grating, however, the last two are most commonly used in LIBS studies. Czerny–Turner spectrograph coupled to a photodiode array (PDA) or Charged Coupled Device

(CCD) array has been mostly used in LIBS analysis. On the other hand, Echelle spectrograph combined with Intensified Charged Coupled Device (ICCD) has recently caught great interest due to high resolving power and wide wavelength range covering emission lines from most elements on one hand while highly sensitive detection of charge on the other. However, Czerny-Turner configuration takes advantage over Echelle and other spectrometers when it comes to price, compactness, simple design, miniaturization and therefore suitability for designing of the portable instruments in field and industrial applications. In our experiments, compact spectrometers were used employing Czerny Turner configuration couple with CCD detectors. This design consists of an entrance slit, two mirrors, and diffraction grating as shown in Fig. 3.3.

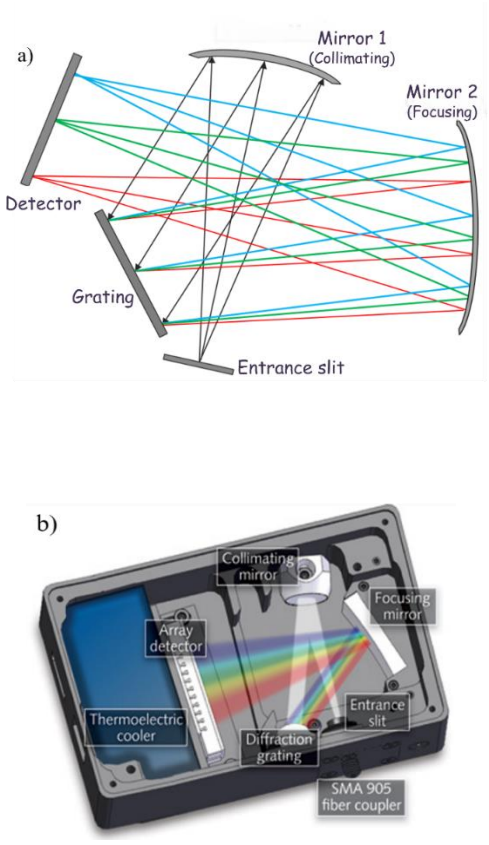


Fig. 3.3. a) Crossed Czerny – Turner design b) inside view of a typical compact spectrometer

The incident light enters through the entrance slit and falls on the first mirror which collimates it onto the diffraction grating where it is spatially dispersed at different angles into spectral components depending on the component wavelengths. This dispersed light is then focussed by the second mirror on the focal plane of the detector where these photons are imaged by the detector. Because of this crossing of the beams this configuration is also known as crossed Czerny Turner. Variation in the focal length of the two mirrors and the number of lines on the diffraction grating allow obtaining different resolving power and spectral range of the spectrometer, hence the spectrometer can be optimized depending on the use and purpose of experiments. When the light falls on the CCD detector, that is made up of pixels arranged linearly or in two-dimensional pattern, it is imaged and converted into electrons which are digitalized. The electric signals are interpolated depending on the pixels of the camera and the spatial dispersion by the diffraction grating to create a calibration and converting this data into a plot of the intensity of each resolved spectral component as a function of wavelength, the spectrum

3.1.6. Spectrometers used in the study

As mentioned previously the spectral range and the resolution of the Czerny-Turner spectrometers can be tuned varying the parameters of the internal components depending on the application therefore in order to carry out the analysis of bacterial and fungal samples two different spectrometers were used based on the emission signals.

Analysis of Bacterial samples was performed using compact spectrometer EPP2000, StellarNet, Tampa, FL, USA (Fig. 3.4a). It consisted of a diffraction grating of 300 l/mm and a slit of 7 µm. The CCD detector was linear, composed by 2048 pixels. This spectrometer offered a spectral resolution of 0.5 nm over a spectral range of 200 to 1000 nm.

Analysis of Candida samples was performed using compact spectrometer single-fibre system USB4000, Ocean Optics, Dunedin, FL, U.S.A (Fig. 3.4b). It consisted of a grating of 600 l/mm and an entrance slit of 10 µm, with a gated CCD detector, 10 µm entrance slit and 3648 pixels. The spectral range covered was from 200 to 900 nm with a resolution of 1 nm.

During the final year of the thesis, a Dual channel spectrometer was also employed in the experiments in order to achieve spectra with higher resolution. The spectrometer was compact, AvaSpec StarLine family, Dual-Channel Fibre optic, commercialized by Avantes (Fig. 3.4c). Based on the need of the experiment, analyte of interest and the spectral range required to be scanned; these spectrometers can be configured independently in different channels with different gratings and resolving power. The spectrometer covered spectral ranges between 190 – 415 nm and 390 – 900 nm, with resolution of 0.08 nm (600 l/mm) and 0.17 nm (1200 l/mm) respectively, and was equipped with two CCD detectors of 3648 pixels each and an entrance slit of 10 µm.

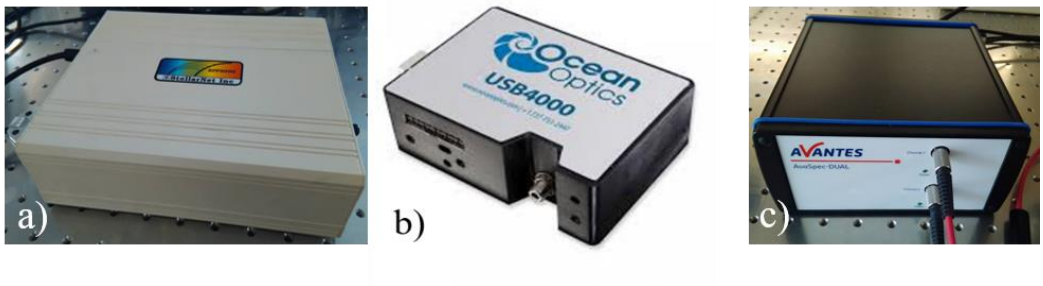


Fig. 3.4. Spectrophotometers used in the experiments

The spectral range and resolution of the spectrometers used enabled to measure the emission from atoms, ions and molecules in the samples. Although traditionally classified as an elemental analysis technology, thanks to these of broadband high-

resolution spectrometers that have extended LIBS applications to molecular species identification. The ability to detect molecular and elemental signatures with a single laser pulse offers unprecedented performance for emerging biomedical and biological applications [3].

3.1.7. Time consideration

In order to optimize the signal to noise ratio the delay time for each study was optimized. The LIBS system was controlled using a delay generator. For the study of bacterial samples, DG535, Stanford Research Systems and in the study of fungal samples, Stanford model DG645 was used.

3.1.8. Experimental set-up

The experimental set-up used for the study of bacterial samples at the laboratory of Laser Chemistry Research Group has been shown in Fig. 3.5 and the experimental assembly for the study of Candida samples set at Los Madroños Hospital has been shown in fig 3.6.

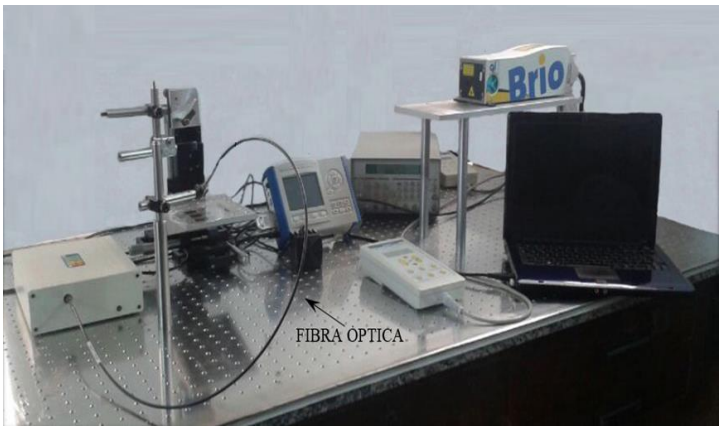


Fig. 3.5. Experimental Set-up at the laboratory of Laser Chemistry Research Group, Complutense University

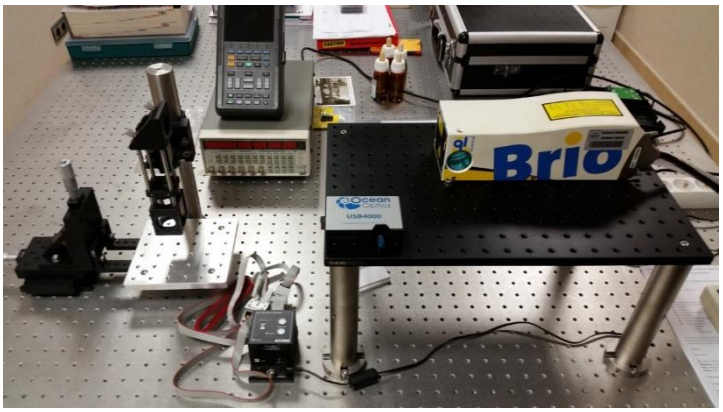


Fig. 3.6. Experimental Set-up for the analysis of Candida samples

The experimental configurations were simple construction and mounting of LIBS set-up, without extensive optics. The experimental parameters such as laser fluence, delay time, integration time, angle and distance of fibre optic cables and the spot size have been mentioned in the respective studies in the results section.

3.2. LIBS Measurements

Single shot spectra were recorded for all microbial samples. Many studies dealing with bacterial identification have used averaged spectrum irradiating multiple laser shots at the same point and/or different points on the bacterial culture whereas others used single shot measurements. Although averaged spectra may be suitable for many applications in relation to reproducibility, data processing and the ability to obtain quantitative measurements, the use of single shot measurements can be preferable in many studies [145] depending on the type of sample and the purpose of application. In the work included in this thesis, single-shot LIBS measurement was selected because of technical and methodological reasons. Firstly, the microbial samples are fragile, the layer of the microbial culture grown on the medium is thin, and the measurements were made directly on the agar medium. Although the agar medium was observed to be almost transparent to Laser beam but still there could be chances that if multiple laser shots were

irradiated on the thin layered fragile samples, it might penetrate and there might appear some signals from the nutrients in the medium interfering with the signals from microorganisms ablation. Moreover, more quantity of microbial culture would be required for such measurements. As the aim of LIBS based methodology is a future medical application, the speed of analysis is a very important factor. On one hand using single shot measurements reduces the time of analysis while on the other small quantity of microbial culture can be sufficient for each test sample thus reducing the time of culture. Therefore, keeping in mind the future application of this technique and its integration into clinical set-ups single shot measurements were selected to perform the spectral analysis.

3.3. Assignation of spectra

The emission lines in the spectra were assigned using the NIST atomic spectral database and the observed emission molecular bands were identified using the spectroscopic information available in References [146,147]. Furthermore, previous studies in the same field of application were also consulted [45,47,148,149].

3.4. SEM-EDS Analysis

In the study of Candida samples, analyses by Scanning Electron Microscopy (SEM) and Energy – Dispersive Spectroscopy (EDS) were carried out. SEM is extensively used to obtain information in order to create high-resolution surface image of the microstructures and enables characterization of the samples. It makes use of electron beam, which due to having shorter wavelength provide greater resolution, on the order of about 1 – 5 nm. The schematic diagram of SEM is shown in Fig. 3.7a. It uses a beam of electrons, produced by an electron gun present at the top of the column, which is then

passed through a series of electromagnetic lenses (at different voltages) and apertures, and finally interacts with the sample placed in a specimen chamber at the bottom. A high vacuum is established inside the column (10^{-10} – 10^{-11} Torr) and specimen chamber (10^{-5} – 10^{-6} Torr) area to maintain the beam coherency by preventing its interaction with gas molecules. The lenses (condenser and objective lens) and micrometric apertures control the properties of the electron beam and its interaction with the sample surface. A pair of deflection coils is present above the objective lens that enables the rastering or scanning across the sample surface on such a local area by the electron beam. The electron beam impinges the sample surface to a few micrometres depending on the accelerating voltage and the characteristic of samples, generating various signals such as secondary electrons, backscattering electrons, Auger electrons, X-rays and photons of various energies. In SEM, the signals of greatest interest are secondary and backscattering electrons, since they reflect the differences in the topography of the sample. These electrons generated have energies less than 50 eV. These electrons fall on the detector and processed to produce a two-dimensional grey-scale image on the monitor following a one to one correspondence rastering pattern.

The X-rays possess characteristic energy depending on the parent atom, and the analysis of this characteristic energy and its intensity by an Energy Dispersive Spectrometer, provides the information about the qualitative and quantitative distribution of elements in the sample at micrometric scale [150].

In the present study, *Candida* samples were analysed using Hitachi S-3000N microscope, to study the external morphology of the *Candida* culture. Energy – Dispersive Spectroscopy (EDS) analyses were performed for the analysis of elemental composition of *Candida* samples with an attached EDS analyser (Oxford Instruments

INCAx-sight) as shown in Fig. 3.7b. The SEM-EDS analyses were performed at high vacuum with an accelerating voltage of 3.5 kV and at a working distance of 4 mm.

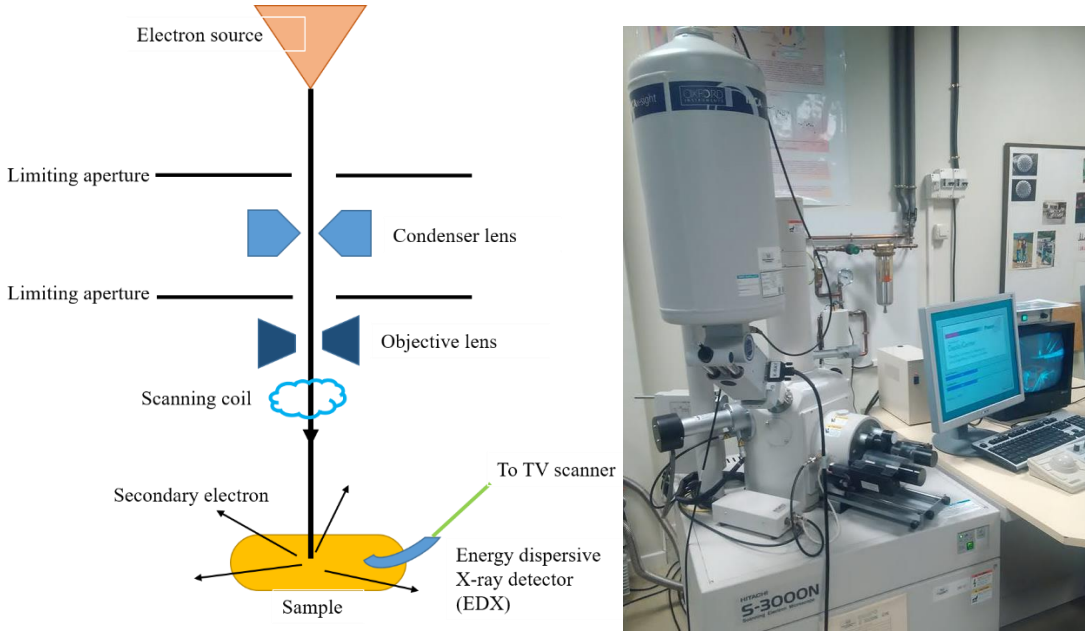


Fig. 3.7. a) Schematic diagram of Scanning Electron Microscope/ Energy Dispersive Spectrometer b) Scanning Electron Microscope/ Energy Dispersive Spectrometer facility at Autonoma University of Madrid

Chapter 4

Laser Induced Breakdown Spectroscopy to Bacterial Analysis

This chapter deals with the analysis of bacteria by Laser Induced Breakdown Spectroscopy in order to evaluate its potential for the development of a methodology to perform the diagnosis of infectious diseases. Such diagnoses require the identification of the causative agent in order to understand the pathology of the disease and start the required antibiotic treatment. Different phenotypic, molecular and proteomics based methods are employed in routine practice to perform this task, however, these methods can be very expensive, restricted to some centralized laboratories, requiring specialized staff to perform lengthy and complicated molecular procedures that may lead to delays of several hours to days even in providing the results [151]. This directly poses risks for the patients' health, as an early and appropriate antibiotic treatment can reduce the costs of treatment and mortality rate in case of serious infections. According to a study, each hour of delay in antimicrobial treatment was associated with an increment in the mortality of 7.6 % [152,153]. Given the situation, the patients are mostly put to an inappropriate antibiotic regimen that leads to the emergence of antibiotic resistant strains, bringing up further challenges to deal with these microbes.

Regarding public health, it is expected that these diseases be treated as early as possible to prevent unwanted outbreaks and series of epidemic events. Where on one hand strict measures should be taken to prevent such risks, on the other hand, highly

selective analytical methods are desirable that can allow a fast and cost-effective detection of infectious agents for a timely antimicrobial treatment.

This thesis has therefore the main goal to take economic, temporal and methodological advantages of LIBS technique in microbiological field to propose a methodology based on LIBS and chemometrics to perform a reliable and robust analysis of bacteria of clinical origin and their antimicrobial susceptibility, improving and/or complementing current detection systems. Particularly, this chapter encompasses the study of effect of medium used to culture bacteria and its significance on LIBS measurements, detection of bacteria with a range of mutations and acquired antibiotic resistance, and finally a comparative study of chemometric methods to develop an appropriate LIBS-chemometrics based bacterial detection strategy.

I - Identification and Discrimination of Bacterial Strains by Laser Induced Breakdown Spectroscopy and Neural Networks.

4.1.1. ABSTRACT

A method based on Laser Induced Breakdown Spectroscopy (LIBS) and neural networks (NN) has been developed and applied to the identification and discrimination of some medically relevant bacteria (*Pseudomonas aeruginosa*, *Escherichia coli* and *Salmonella typhimurium)*. In order to develop a strategy for its posterior application in the detection of bacteria, the dependence of type of medium to culture these microbes on LIBS analysis has been investigated. For this purpose, three different media i.e. *LB agar*, *MacConkey* and *Brucella* were selected. Furthermore, the samples used in this study were divided into three groups, each prepared on a different day. A dependence of medium was found out on LIBS signal which affected the identification. However, when medium was kept the same for all microorganisms, a successful identification and discrimination of the samples was achieved using the spectral libraries, showing that only a difference between the bacteria could cause identification. Due to the nature of samples, LIBS measurements were made using a single laser pulse that resulted sufficient for clear identification of the bacterial strains by NN models, obtaining the identification of the bacteria with a certainty of over 95%. The results showed the potential of LIBS-NN system for fast, reliable and robust measurements. In order to prove its practicality in clinical laboratories, it is needed to analyse a diverse range of bacteria.

4.1.2. INTRODUCTION

Laser Induced Breakdown Spectroscopy (LIBS) analysis by direct measurement of the optical emission from laser-induced plasma has been the subject of research for many years [40,154]. This technique provides a quick and versatile method for analysing different types of samples that can be inaccessible or tedious using conventional analytical techniques, and is particularly useful for analysing samples with a complex matrices [155,156]. LIBS as an instrumental platform has replaced some analytical techniques successfully in some industries especially in material analysis [157,158], and is finding extensive use for the analysis of biological samples such as muscle tissues, cancerous tissues, plants and microorganisms among others [159-165]. Of which analysis of microorganisms for the identification of pathogens to detect food contamination, infectious diseases or deal with bioterrorism have been important directions of LIBS microbiological approach [36,37,39,44,166].

In many situations, a considerable number of alternative techniques with a higher sensitivity are available, however, LIBS provides several advantages over these conventional methods for elemental analysis. These include, elimination or minimization of laborious sample preparation steps, possibility of analysis in any state of matter (solid, liquid, gas), high speed of analysis, micrograms or nanograms amount of sample ablated from its surface, and simultaneous detection of a range of elements and molecular patterns representing sample´s composition [45,167]. Specifically, remote measurements allowing the analysis of hazardous, highly contagious and pathogenic targets at a safe distance and ability to computerize the spectral as well as data analysis, without being mandatory to have the expertise of a trained microbiologist make it an interesting pathogen detection tool.

Detection and identification of biological samples, in particular, bacteria using the LIBS technique has been studied by several research groups. Nevertheless, it was not until 2003, when the possibility of using LIBS to microbial analysis was tested [36,47,168,169]. The first of these studies focused on the use of LIBS as a system for early real-time detection of biological weapons. The motivation for most of these studies was LIBS' ability to provide extremely fast identification compared to traditional methods of identifying bacteria.

In preliminary experiments performed in 2003, Morel et al. investigated the detection of six strains of bacteria and two pollens [36]. They placed particular emphasis on *Bacillus globigii*, which acts as a non-pathogenic surrogate for *Bacilus anthracis* (anthrax), demonstrating the ability of LIBS to detect bacteria. That same year, Hybl et al. investigated the technique's potential for detecting and discriminating aerosolized bacterial spores from more common background fungal/mold spores and pollens [38]. More recently, the possibility of using LIBS to detect and identify bacteria in clinical diagnosis and public health has prompted investigations into the use of more realistic samples from a clinical analysis perspective. As from that date, the increase of publications and reports of the use of LIBS for bacterial analysis has been exponential.

The intrinsic features of LIBS technique on one hand while combination of this technique with chemometrics became a reason to catalyze research in this direction. Chemometrics is a multivariate data analysis tool coupled to the data rich instrumental methods for qualitative or quantitative analysis [170]. LIBS has been combined to a variety of chemometric techniques for bacterial identification; however, the correct identification rate in some cases fell below 90%. Thus, there is a clear need for more thorough and systematic studies that include new approaches. Many chemometric methods have been evaluated in several research works, such as Principal Components

Analysis (PCA), Soft Independent Modelling of Class Analogy (SIMCA), and Partial Least-Squares Discriminant Analysis (PLS-DA). Those methods are not able to give satisfactory solutions to many practical problems that can be attributed mainly to uncertainty in identification. Research exists demonstrating that the use of Neural Networks (NN) can provide better results. An interesting comparison of some of these methods with the use of NN has been performed [59,171]. Therefore, NN was selected for data analysis as it is reported to significantly improve the identification capability without considerably increasing the difficulty of implementation.

The aim of this paper is to evaluate and develop a simple and direct methodology based on LIBS and Neural Networks to identify and discriminate bacterial samples. Specifically, in this work, it was intended to improve the recognition capacity by developing a method capable of identifying extremely similar samples that have few physical and spectral differences between them. Moreover, the ultimate goal of evaluation of this methodology for bacterial identification is to develop a reliable pathogen detection system, research on the factors affecting analysis were also investigated. Previous studies in the field of microbiology have reported that the composition and behavior of microbes is strongly dependent on their growth environment [172], in particular on the nature of the nutrient. This holds true not only for all cellular constituents and metabolic pools but also for the elemental composition of the microbial biomass [172,173]. As LIBS technique is sensitive to changes in the elemental composition of the samples, the effect of type of medium used to cultivate bacteria on LIBS analysis was also studied. This test was performed to find out if the chemical composition of media could introduce changes to the bacterial cells that consequently affect LIBS based identification.

4.1.3. MATERIAL AND METHODS

4.1.3.1. Experimental set-up

The experimental configuration has already been shown in Chapter 3. Experiments were performed by using a Q-switched Nd:YAG laser (Quantel, Brio model) operating at 1064 nm, with a pulse duration of 4 ns full width at half maximum (FWHM), 4 mm beam diameter and 0.6 mrad divergence. The samples were placed directly over an X-Y-Z manual micrometric positioner with a 0.5 µm stage of travel at every coordinate to ensure that each laser pulse impinged on a fresh sample. The laser beam was focused onto the sample surface with a 100 mm focal-distance lens, producing a spot of about 100 µm in diameter. The experimental configuration with this focal distance, distance of the collimator and resulting spot diameter made it possible to maintain spatial resolution on the sample surface without irradiating the target on the previously shot area, as well as recording the plasma emission without any interference. The pulse energy was 20 mJ, and the repetition rate was 1 Hz. The emission from the plasma created was collected with a 4 mm-aperture, with a 7 mm focus fused silica collimator placed at a distance of 3 cm from the sample, and then focused into an optical fibre (with a 1000 µm core diameter and 0.22 numerical aperture), which was coupled to the entrance of the spectrometer. The spectrometer system was a user-configured miniature single-fibre system (EPP2000, StellarNet, Tampa, FL, U.S.A.) with a gated CCD detector. A grating of 300 l/mm was selected; a spectral resolution of 0.5 nm was achieved with a 7 µm entrance slit. The spectral range from 200 to 1000 nm was used. The detector integration time was set to 1 ms. To prevent the detection of bremsstrahlung, the detector was triggered with a 5 µs delay time between the laser pulse and the acquired plasma radiation using a digital delay generator (Stanford model DG535). The spectrometer was computer-

controlled using an interface developed with Matlab, which allowed for data processing and real-time NN analysis.

4.1.3.2. Methodology

The LIBS-NN methodology of bacterial identification is mainly based on isolation of bacteria from the biological samples, obtaining LIBS spectra and finally comparison of test spectra from bacterial sample against the generated spectral database providing the identity. In this case, no detailed chemical analysis of bacteria is sought; rather an instant identification of the sample using a unique "characteristic LIBS spectral fingerprint" or "marker" of that sample is achieved. This characteristic LIBS spectral fingerprint is due to the nature of the emission spectra, representative of the main compounds and the bacterial matrix that constituted the sample. The matrix structure and composition strongly affect the intensity of the emission lines, and have often inhibited a possible direct relationship between the elemental concentration of a sample and the intensities of its spectral lines. Thus, LIBS provides a unique spectrum, corresponding only to the sample under analysis. Using a correlation procedure, the LIBS-NN system developed can be trained to recognize spectra from different samples, which means evaluating the similarity of unknown spectra against a spectral library of classified samples.

4.1.3.3. Bacterial samples

Keeping in view the medical importance of microorganisms, medically relevant species and strains of bacteria were analysed included in this study. It has already been mentioned in chapter 1 that the focus was particularly to the pathogens that are common in the nosocomial infections or HAI. Specifically, the bacteria studied included strains belonging to the species *Escherichia coli*, *Pseudomonas aeruginosa* and *Salmonella typhimurium* were studied. As the matrix effect causes significant interference in LIBS

signals, the analysis was always performed on isolated bacterial samples obtained from biological specimens.

Bacterial samples were grown on three different culture media, having a different chemical composition and nutrient value. The media selected were Luria-Bertani (LB) agar (from Difco Microbiology, Lawrence, KS, U.S.A), MacConkey agar (from Difco Microbiology, Lawrence, KS, U.S.A.) and Brucella anaerobic agar (from bioMerieux SA, Marcy l'Etoile, France). LB agar is a nutritionally rich medium, which is highly referenced for growing a wide range of bacteria and traditionally has been used as an optimum medium for Escherichia coli. MacConkey Agar is a selective and differential medium used to isolate gram-negative bacteria. It contains bile salts as a main component, which inhibits the growth of gram-positive bacteria. Brucella Agar is a nutrient-rich medium used for *Brucella* particularly; however, it is also recommended for the isolation and cultivation of other fastidious and non-fastidious bacteria from clinical and non-clinical specimens of public health important [174]. Due to different composition and functions, the effect of these media on the growth of bacteria and LIBS analysis was studied.

For preparing each culture medium, 20 g of nutrient agar powder was suspended in one litre of distilled water and heated to dissolve it completely. The pH was adjusted to seven and the flask containing liquid medium was autoclaved and sterilized. After cooling the agar medium was transferred to sterile petri dishes (8.9 cm in diameter), and allowed to solidify. The bacterial strains were taken from frozen culture and then streaked onto the surface and incubated at 37 ºC for 18 hours, to get the required biomass for LIBS analysis. I would like to thank Prof. Juan Alfonso Ayala from Centro de Biología Molecular "Severo Ochoa", Autonoma University of Madrid, Spain, for providing

bacterial strains used in this study and sharing his expertise in the microbiology field for the development of this thesis.

The bacterial samples were of wild-type strains, *Pseudomonas aeruginosa* M841, *Escherichia coli* OV2 and *Salmonella typhimurium* LB5010 were designated as B1, B2 and B3. They were prepared on three different days, with a 10-day gap in between. The media i.e. LB agar, MacConkey agar and Brucella anaerobic agar were designated as culture media M1, M2 and M3, respectively. Because the NN handles numbers, numerical and alphanumerical nomenclatures were used to label each sample. Table 4.1.1 shows the name structure used for the samples. The alphanumerical names indicate the bacterial strain and culture medium. The numbers in brackets correspond to the identification number of the bacterial sample. For example, sample 23 corresponds to B2M3, which means, E.coli (B2) grown on Brucella anaerobic agar (M3).

Table 4.1.1. Nomenclature of bacterial samples

Bacteria	**LB Agar**	**MacConkey Agar**	**Brucella Anaerobic Agar**
Pseudomonas aeruginosa	B1M1 (11)	B1M2 (12)	B1M3 (13)
Escherichia coli	B2M1 (21)	B2M2 (22)	B2M3 (23)
Salmonella typhimurium	B3M1 (31)	B3M2 (32)	B3M3 (33)

4.1.3.4. LIBS measurements

All samples were used with no further preparation than that described herein. LIBS spectra were recorded directly on the biomass deposited on the culture media in each petri dish. Due to the delicate and thin-layered structure of bacterial biomass, single

laser shots were irradiated. A minimum of 3 replicate tests for each sample were performed. The samples analysed were 3 bacteria × 3 culture media × 3 days × 3 replicate analyses which made a total of 81 samples analysed. The analysis was made under room experimental conditions. Other studies have shown that LIBS measurements in inert atmosphere such as Argon increase the signal to noise ratio, which could allow easing in data analysis; however, it may increase the complexity of experimental set-up as well as cost of analysis. To keep the apparatus simple for a posterior application room conditions were chosen. Each sample was irradiated with 100 laser pulses. For each pulse, the generated plasma spectrum was acquired and stored as a column on a dataset. Thus, the dataset contains the intensity at different wavelengths in rows and the spectra in columns. Thus, our dataset has 2048 rows (one for each wavelength) and 100 columns or spectra for each sample.

4.1.3.5. Creation of Spectral Libraries

Each of the individual worksheets containing the spectra for a specific sample constitutes a spectral library. From the spectral library, 80 spectra were used to create the training libraries to calibrate the NN model and 20 spectra were used to test the classification model. The greater the number of spectra used in the fingerprint of a sample, the better the recognition capacity of the method. A more thorough study of how recognition affected model identification is shown in the results section. The acquisition of these 100 spectra is very fast (<2 minutes, taking into account the integration time and 1 Hz laser pulse repetition). Although the data matrix can be considerably large, the computation time for training the NN was always below 10 s.

4.1.4. Data Analysis by Neural Networks

The NN models were based on a multilayer perceptron, feed-forward networks and supervised learning algorithm. They consisted of several neurons (information processing units) arranged in two or more layers. Each layer receives information from all of the neurons in the previous layer. The connections are controlled by a weight that modulates the output from the neuron before inputting its numerical content into a neuron in the next layer. The process that optimizes the weights is called the learning or training process [175,176], which was based on a back-propagation (BP) algorithm [176]. The inputs from each neuron are added by an activation function, and the result is transformed by a transfer function that limits the amplitude of the neuron output. In this work, the hyperbolic tangent sigmoid function was used as the NN transfer function. When the NN parameters are adjusted by slightly refreshing the weights, the NN is able to learn from its environment. The theoretical fundamentals behind these concepts have already been mention in Chapter 2. The NN models were designed using Matlab software (Matworks, 2010a).

4.1.4.1. Description of the learning and verification set

As the NN models were based on a supervised algorithm, in order to optimize the weight matrix, it was necessary to use input and output data that adequately characterized the process to be modelled. The input data was a linear combination of the dataset libraries. Data from the spectral training libraries (mentioned in section 4.1.3.5) were randomly distributed into the learning (80% – 64 spectra) and verification (20% – 16 spectra), internally by the NN algorithm. Once the learning and verification processes were carried out, the network parameters were transferred to a program developed in Matlab that provided real-time identification during data acquisition.

4.1.4.2. NN model calibration and verification process

The NN model consisted of three layers (input, hidden and output), a topology widely used to model systems with a similar level of complexity [177]. In particular, the input layer consisted of 2048 nodes (intensity values in the 200 - 1000 nm wavelength range). The output layer was comprised of J neurons (where J = number of reference samples used) for estimating the similarity between the reference sample spectrum and the testing sample spectrum. The identification process was based on the ability of the NN to detect the degree of similarity between the new spectrum and each of the reference spectra used in the learning process [177].

During the training process, each sample used as a reference was associated with an identification number (usually the same number assigned to the sample) in the output layer. Thus, a perfect identification was obtained if the output from the NN model for the test sample matched the identification number assigned to the reference. It is possible to use more than two identification numbers simultaneously, e.g., when analysing a large number of samples. If the test sample is not assigned to any class, it is taken as unknown.

NN training was achieved by applying the BP algorithm, based on the conjugate gradient method [178], one of the general-purpose, second-order techniques that help minimize the goal functions of several variables. Second order indicates that such methods use the second derivatives of the error function, whereas a first-order technique, such as standard back-propagation, uses only the first derivatives.

To determine when the training should be stopped, an early stopping criteria based on the validation set was used [126]. The number of epochs was not relevant in this case. To avoid an overfitting of the NN model, the learning process was repeated while the verification mean square error (MSE), defined in Eq. 30, was decreased.

$$MSE = \frac{1}{N}\sum_{k}^{N}(r_k - y_k)^2 \qquad \text{(Eq. 30)}$$

where N, y_k, and r_k are the number of input data, the response from each output neuron, and the real output response, respectively. A detailed description of the calculation process is provided in the literature [126,175].

4.1.5. Process for testing the calibrated NN Models

4.1.5.1. Performance parameters for classification models

Once NN models were trained, they were validated by test libraries. The confidence of the prediction can be expressed by a conditional probability, i.e. the rate of correct classification within the classified spectra. The confidence was estimated by a parameter termed as Spectral correlation (SC) given by Eq. 31. The higher the SC, the better the efficiency of the network for identifying a bacterial strain.

$$SC = \frac{100}{N}\sum_{i}\delta_i \qquad \text{Eq. 31}$$

where δ_i is the number of spectra classified correctly and N is the total number of spectra. The performance of the models generated was evaluated based on the metrics explained previously in Chapter 2. In each case, confusion matrix was obtained and the results were obtained in terms of s*ensitivity* (S), *specificity* (SP) and *Accuracy* (A) [126,143,179,180] based on the number of TP, FP, TN and FN in the validation set. Whereas accuracy was the main parameter of the recognition procedure for decision-making, and the reason the metrics for assessing detection processes are so important. These involve the relative frequency of correct and incorrect identification obtained from the results. The receiver operating characteristic (ROC) curve which is a standard tool for plotting all possible combinations of sensitivity and specificity in a screening process was obtained. Area

Under Curve (AUC) was calculated as it provides a better measure than accuracy for evaluating the predictive ability of the classification.

The robustness of the model was tested by assessing the ability of LIBS-NN to estimate the correct result when an unknown sample was input into the network model. In other words, the higher the robustness, the better the efficiency of the network model for identifying a sample not included in the training step as an "unknown," and not identifying it as another bacteria [181]. In order to test this, one dataset (input data spectra) from the training set was removed. The results obtained with the remaining ones were then checked, in terms of the probability of correct identification. This was alternately repeated for each bacterial strain.

4.1.5.2. LIBS-NN model validation process

Four validation processes were carried out, each with its respective set of input data. In all cases, the reference datasets (spectral training libraries) were formed by the spectra recorded on day 1.

First validation was carried out to evaluate the model's capacity to recognize and identify different bacterial species in the same culture media. The test samples were comprised of the day 1 bacterial samples. Each bacterial sample was individually analysed and contrasted with another from the same day in a binary manner.

Second validation was performed to evaluate the validity of NN models and spectral libraries forming reference datasets in identifying a bacterial strain cultivated on different days. Bacterial samples from the other two days were then individually compared against the reference datasets, both in the same and different culture media.

Third validation included the evaluation of the NN models to identify bacterial strains grown of different culture media, in order to find out if media composition affected bacterial cells in way that could interfere with LIBS based identification.

Fourth validation was an *independent external validation*, carried out to test the model's ability to identify unknown samples. To test the robustness of the model, the spectra of samples from day 1 were used as the library reference, and one bacterial dataset was removed and alternately repeated for each bacterial strain. The model was considered robust if the sample removed was identified as unknown and the NN output for these bacteria was zero.

4.1.6. RESULTS AND DISCUSSION

Fig. 4.1.1 shows an example of normalized single-shot spectra for three bacterial species i.e. *Pseudomonas aeruginosa*, *Escherichia coli* and *Salmonella typhimurium* on LB agar medium. The compositions of these samples seems to be very similar, as evident from their spectra, hence making visual identification difficult.

Due to very fast testing conditions and bearing in mind the high spectral similarity between the samples (for which a single shot might be the only sampling event), the spectra were not averaged. The emission lines can be seen to be coming from the sample as a result of ablation as well as contribution from the air is also present, due to the experimental conditions.

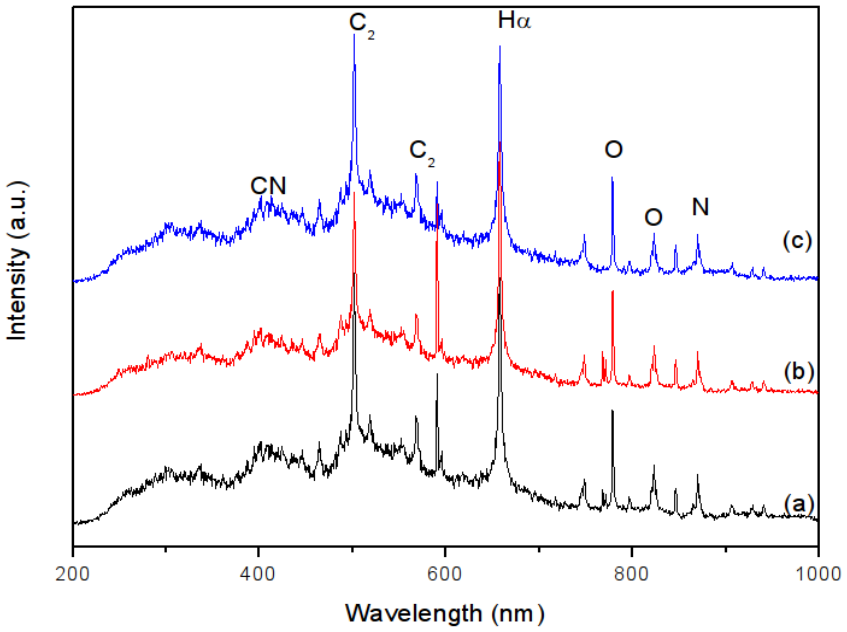


Fig. 4.1.1. Normalized LIBS single-shot spectra for three bacterial strains from day 1: (a) P. aeruginosa (B1M1), (b) E. coli (B2M1), (c) S. typhimurium (B3M1).

Taking into account that the plasma is generated from evaporated and ionized sample materials mixed with ambient gas [61], some spectral lines from air were present in the spectra, which might make elemental identification less effective. On the other hand, it has been demonstrated [182], that spectra from air can be correlated with the use of a simple multiplicative scaling factor, thus demonstrating that changes in relative spectral contributions from oxygen and nitrogen are not occurring. Furthermore, the contribution of different atomic ionization levels in the spectra does not change significantly. It is also necessary to consider the kinetic effect of the native CN molecular band usually observed in LIBS spectra for organic and biological samples [183,184], which corresponds to either native CN molecules vaporized from the sample, or CN bond formations due to recombination with the air.

Spectral emission lines and a continuum background emission are typical components of the spectra created in LIBS. In addition, the intensity of the spectra can change from pulse to pulse and from day to day; however, to ensure that it does not affect the system's ability to identify the sample; two aspects were taken into consideration. First, the system is trained with a set of spectra that were recorded using all possible variations of the ablation parameters, such as changes in the lens-sample distance and the laser pulse energy. Second, both the intensity ratio and the bandwidth were analysed by the NN, which made discrimination and identification possible. Furthermore, in order to avoid data variations due to changes in the laser pulse energy, each spectrum was normalized by the intensity of one specific spectral line (i.e., with the most intense assigned to hydrogen Hα, Fig. 4.1.1).

4.1.6.1. Variable selection

Optimization of the appropriate number of variables to be used as inputs for the training of a chemometric model is a very important aspect to get a successful classification results. Selection of the “optimum” subset of full spectrum for multivariate calibrations is widely investigated in literature [185]. On the other hand, selection of emission lines of a specific element or using intensity/area of multiple emission lines or ratios between them has also been among the approaches used for variable selection. It has been shown that the selection of intensities at some wavelengths for elements such as P, C, Mg, Ca, and Na provides sufficient information to achieve the identification of bacteria [47]. However, in the present study two approaches were considered i.e. selecting shorter spectral ranges with few peaks, applying Principal Component Analysis (PCA) and secondly using the entire spectral range as spectral fingerprint for the sample. Tests performed using shorter spectral ranges with few peaks, selected by PCA analysis, showed that the model’s recognition ability decreased and the NN model could not

discriminate between different bacterial species. This can be attributed to highly similar spectra and composition of bacterial samples. This imposed extreme requirements on the identification model to take into account the broadest spectral range in order to cover the greatest number of spectral characteristics of the sample. In many other chemometric methods, this can limit their analytical performance due to massive data volume. However, the size of the data for the NN (denoted hereafter as input-data) can be quite large; therefore, a mathematical algorithm efficient and effective enough to handle a large amount of data was optimized. In the light of these results, it can be said that the whole spectrum serves as the biomarker for the bacterial samples. This has the disadvantage of decreasing the spectral resolution, which makes it even more difficult to unequivocally identify the elements responsible for those lines, but has the advantage of reducing the analysis time. Employing a spectrometer with greater resolution might allow a reduction in the variables.

4.1.6.2. Number of spectra used in the training set

The size of training set has great influence on each algorithm and its performance, therefore, number of spectra needed to provide a robust classification was investigated. To improve the correct identification rate, the number of spectra for each sample must increase to improve the spectral correlation. It is therefore necessary to select a high number of spectra, and because each spectrum comes from a single laser pulse, the time taken to collect the data (1 second per spectrum) is not an issue. The broad spectral range used also plays a key role in correct identification.

Finally, to evaluate the optimum number of spectra used in the training process, the variations in robustness as a function of the number of spectra used in the training matrix was studied. Fig. 4.1.2. shows a plot of these results. As we can observe, robustness increases rapidly with the number of spectra. Even for a very low number (8

spectra), the robustness is acceptable but it could present problems in discrimination if a large number of diverse microorganisms are tested.

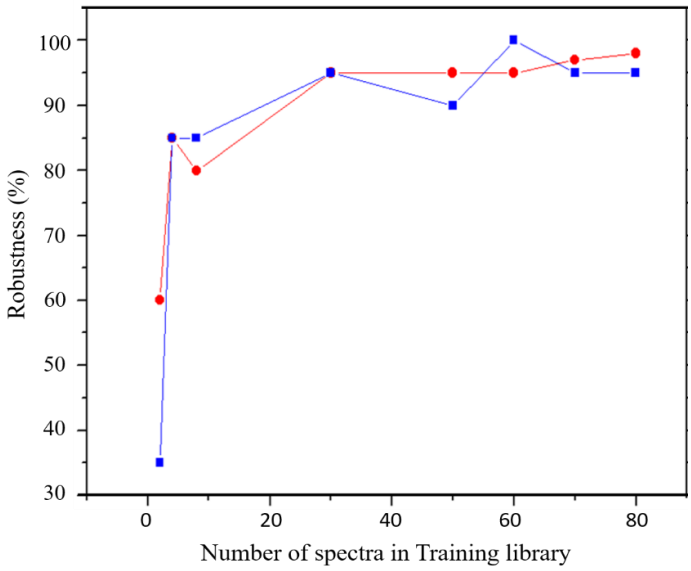


Fig. 4.1.2. Robustness obtained for the samples B1M1 (■) and B2M2 (●), as a function of the number of spectra used in the process of training the NN.

Therefore, to develop a robust classification model for microbial identification a higher number should be considered, as the Fig. 4.1.2 shows that when spectra number in training library reaches 80, the robustness reaches closer to 100%. Although a bigger number can be used but it would increase the time of LIBS analysis as well as for data manipulation. Keeping in view time factor and robustness achieved, the number of spectra in training libraries was optimized to 80.

The NN was trained for spectral characteristics using the input data of known samples, with the optimized number of spectra. The subsequent network model parameters were then handled by an identification program that performs real-time identification during data acquisition.

4.1.6.3. LIBS-NN model validation

First validation: The LIBS/NN correlation method was first applied to the identification of bacteria in the same day. A collection of single-shot spectra for each sample were analysed and contrasted with the spectra for reference samples in the same day, in a binary manner. In each binary analysis, part of the library for the sample was used as reference. As an example, Fig. 4.1.3 shows the analysis of sample B1M1 and B2M1 (*Pseudomonas aeruginosa* and *Escherichia coli* on the same cultured media LB agar). 80 spectra for each sample were used as a reference sample set, where 11 and 21 were assigned as identification numbers, respectively.

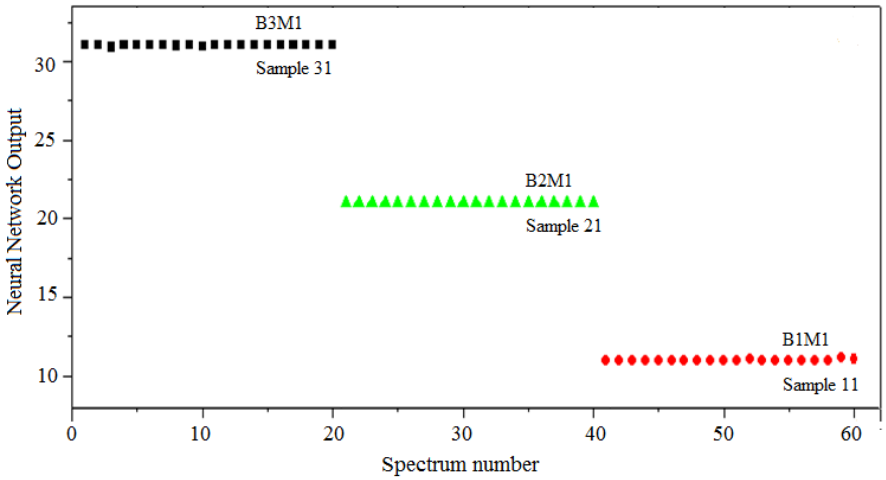


Fig. 4.1.3. LIBS–NN correlation method applied to the identification of bacterial strains from day 1. Squares correspond to B3M1 (sample 31), triangles correspond to B2M1 (sample 21), Circles correspond to B1M1 (sample 11)

Twenty spectra for samples not included in the reference were tested. The references were assigned the same bacterial identification numbers for the NN output, taking into account the bacteria and matrix. Thus, if the NN has an output equal to 11, this corresponds to the identification of bacterial strains B1M1 shown in Table 4.1.1. If

the NN output contains an output equal to 21, this corresponds to the identification of bacteria B2M1, and so on. With the first laser shot, the network recognizes the spectrum as belonging to a reference dataset, and the NN output is 21. Therefore, the NN model cannot "see" the difference between the analyzed sample and the reference sample with an identification number equal to 21. The same result was obtained up to spectrum 20. Then, when test sample B2M1 is replaced by sample B1M1, the network output is 11. At this point, the NN correctly assign the spectra to the identification number equal to 11 used for this sample. Most of the spectra for this sample were assigned correctly, and only two deviated from the expected behaviour, returning values in between the two values assigned.

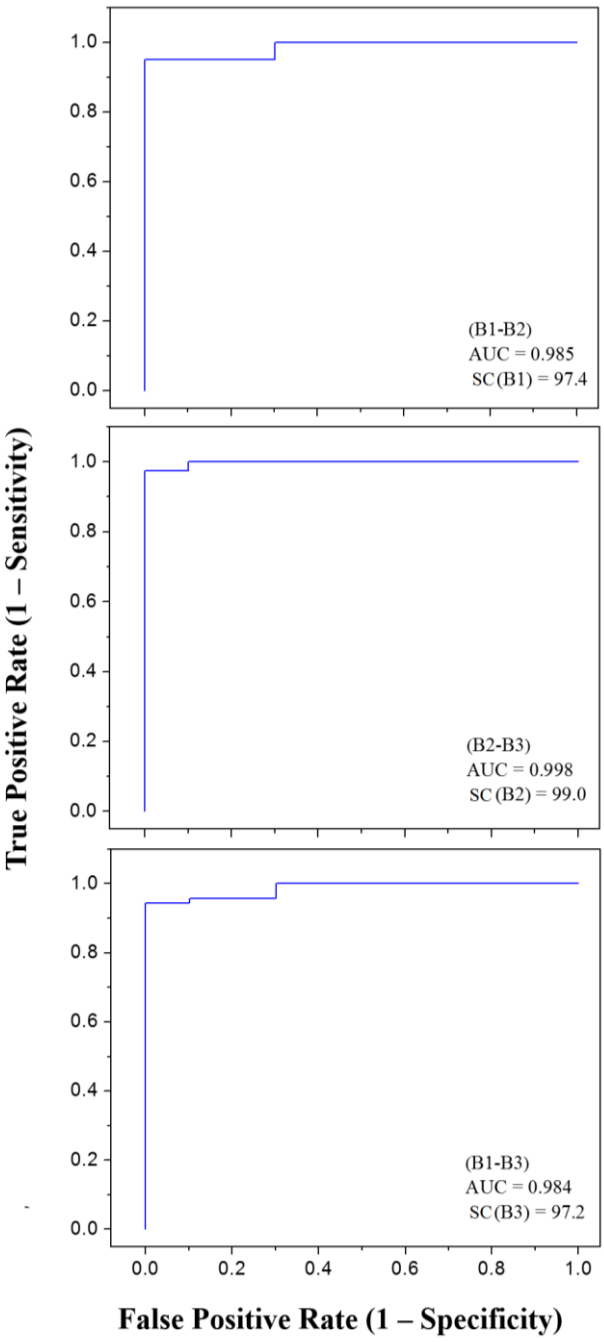


Fig. 4.1.4. ROC plot obtained for the NN models, presenting AUC and Spectral correlation

This behaviour strongly affects both the spectral correlation and the AUC, which were 97.4% and 0.985, respectively. Fig. 4.1.4 shows the ROC plot and spectral correlation obtained for this case, and the capacity of the LIBS/NN to identify those bacterial strains. Given that the spectra

analysed came from a single laser shot, the disturbance spectra observed in only 2 out of 20 is not only more than acceptable, but is essential for taking into account the spectral correlation. Exactly the same results were obtained when using the data set for all other bacterial strains from day 1 as a reference, and updating the NN parameters, as shown in Table 4.1.2.

Table 4.1.2. Classification results from Neural Network analysis

Classification results							
		Identification test				**Robustness test**	
Bacteria	**Medium**	**MI (%)**	**Unidentified**	**Misidentified**	**Correctly identified**	**Misidentification**	**Correct as unknown**
Pseudomonas aeruginosa	*LB*	97.4	0	0	100	0	100
	Mac Conkey	98.2	0	0	100	0	100
	Brucella	98	0	0	100	0	100
Escherichia coli	*LB*	99	0	0	100	0	100
	Mac Conkey	100	0	0	100	0	100
	Brucella	100	0	0	100	0	100
Salmonella typhimurium	*LB*	97.2	0	0	100	0	100
	Mac Conkey	98	0	0	100	0	100
	Brucella	100	0	0	100	0	100

It means that the differences between the two colonies cannot be attributed to matrix variations. Only differences between bacteria can cause identification within the

same matrix. Another important result is obtained from this test. The high spectral correlation shows that, despite not taking into account the contributions of both air and the culture medium, a correct identification can be achieved, and helps to decrease the analysis time without significantly affecting the model discrimination capacity.

Second validation: As part of a tough test for evaluating the validity of the reference matrix with regard to time and the method's ability to identify bacteria cultivated on different days, the spectra for the bacterial strains in day 1 were introduced into the NN model as references. The mathematical procedure followed was similar to the training and validation process described above. Measured on different days, but still each strain was classified to its respective species. Fig. 4.1.5 shows the results obtained for the samples from day 2 (B1M1, B2M2 and B3M3). For all other bacterial samples, the classification results were averaged and presented in Table 4.1.2. It can be seen that the samples have been correctly identified from the first laser shot. Results also show the network's capacity to work simultaneously with more than one fingerprint, without significantly increasing the computing time. The training process for all samples from day 1 used as references required 7.2 seconds of execution time on a standard computer. Fig 4.1.5 shows the results obtained for *E. Coli* (B2) from day 3 (last day) in three different media. The bacteria were correctly identified, even in this case where the identification was made using the first library (the oldest) as a reference, and comparing to libraries obtained subsequently. This confirmed the temporal validity of the libraries, at least during this study.

Third validation: In principle, it might be assumed that the identification of the bacteria should not be dependent on the culture medium, as LIBS measures the cellular content.

Nevertheless, the Fig. 4.1.5 shows that same bacteria was classified differently when cultivated in different media. This difference was interpreted as a real and not unexpected elemental alteration of the membrane of the bacteria cultured in that medium, and did not represent an inherent limitation of the LIBS technology.

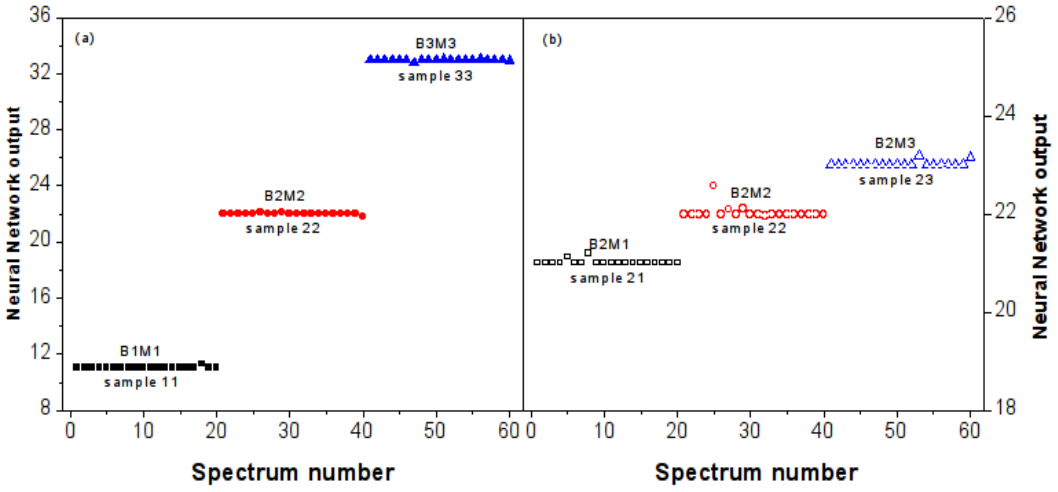


Fig 4.1.5. a) NN output for 20 spectra for sample B1M1, B2M2 and B3M3 from day 2 (not included in the training), and b) The NN output for 20 spectra for samples B2M1, B2M2 and B2M3 from day 3. This confirms that the spectral library used in the training process was useful for all samples analysed.

The alteration of fundamental bacteria chemistry was attributed to the presence of bile salts in the MacConkey medium, which is known from biochemistry to disrupt membrane integrity. Because standard serological (antibody-based) diagnostics are also membrane-based, rather than genetically based, these competing microbiological techniques might misidentify such bacteria if their outer membrane or surface has been significantly altered.

Although growth on different media affected the elemental composition, and consequently LIBS analysis, no significant difference was observed in the quantitative

aspect of bacterial growth. Virtually every microbiological experiment starts with the cultivation of microbes that makes it a fundamental methodology of microbiology [186]. This suggests that for LIBS based bacterial identification, a standard medium should be selected in order to avoid any alterations in the cellular structure and thus in the spectra.

Fourth validation: In order to complete the validation of the prediction capability of the optimized NN, a fourth independent validation set was carried out in order to test the model's ability to identify unknown samples as unknown. To test the robustness of the model, the spectra of samples from day 1 were used as the library reference, and one bacterial dataset was removed and alternately repeated for each bacterial strain. The NN compares the analysed sample spectra with those stored as a reference (like two fingerprints). Therefore, if they match, the output from the network is satisfactory, and the value assigned to the reference and NN output match. If the fingerprints (spectra) differ slightly, the NN output must be zero. This means that the bacteria are not present in the training set, and are therefore unknown. As an example, Fig. 4.1.6 shows a possible stronger test. In this case, all samples from day 1 were used as a reference, and the spectra for Salmonella (B3) were removed. The NN output for 20 spectra from each sample in day 2 (B3M1, B2M1 and B1M1) were tested. The first 20 spectra correspond to sample B2M1, the next 20 to sample B1M1, and the last 20 to sample B3M1. All samples were correctly identified. Even sample B3M3 was correctly identified as unknown, and not as another sample present during training, which demonstrates the robustness of the method used.

Looking at the results, it can be said that the time required to obtain the spectra is very fast. Once stored, they are selected for further analysis in real time. Analyses carried out with different types of bacteria (generated at different times) showed that the libraries

were adequate for correct identification of the bacterial strains, even with small variations in the experimental conditions, such as changes in laser energy, room temperature, or sample distance.

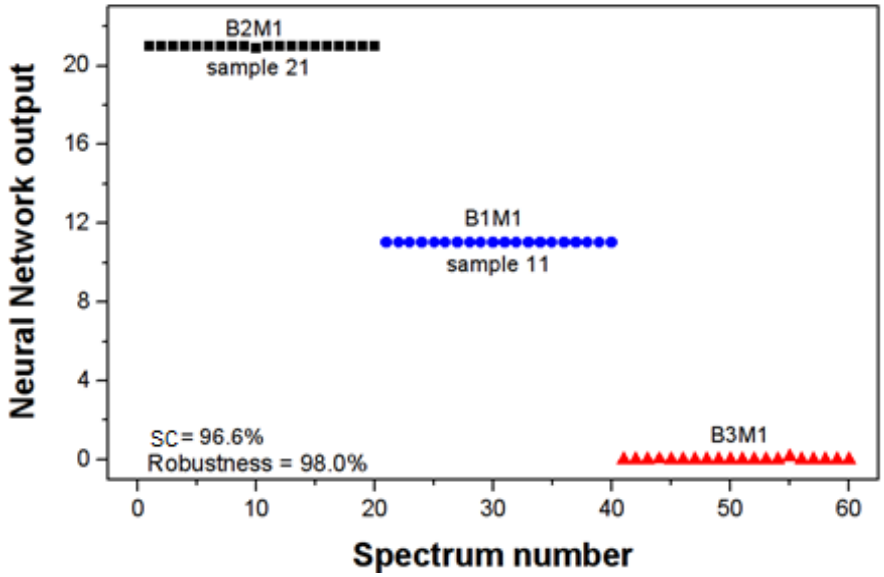


Fig 4.1.6. NN outputs for B1M1, B2M1 and B3M1, where the latter corresponds to a sample set not included in the training set. Thus, this B3M1 sample was unknown for the NN model. Sample B3M3 was correctly assigned a zero, the identification number assigned to unknown samples. The spectral correlation was 96.6%, and the robustness was 98.0%.

Single-shot measurements were sufficient for clear identification of the bacterial strains studied. In light of these results, the optimized NN model provides reliable results (sample identification) for all samples analysed. This result is the best indicator of the capacity of the methodology presented.

4.1.7. CONCLUSIONS

It has been shown that accurate sample analyses can be obtained using LIBS-NN. Tests performed on bacteria samples demonstrated 100% reliable identification of known and unknown samples with very similar spectral characteristics. In addition, in studies

where the only variation was the type of bacteria, the identification was correct, and therefore did not depend on the culture medium. Only differences between bacteria resulted in identification. Despite not taking into account the contributions of the air and culture medium, a correct identification can be achieved, which helps decrease the analysis time without significantly affecting the model's discrimination capacity.

The identification analysis was stable over a long period, and minor changes in experimental conditions, such as the intensity of the LIBS single-shot regimen and continuum background, were not relevant for sample identification. The system was able to perform a correct identification even with a single laser shot. The most important conclusion is that in the 200-1000 nm range, each spectrum is a true fingerprint of the sample, allowing correct differentiation of each bacterial strain using the NN model.

Multivariate techniques are known to be efficient methods for sorting and classifying data. However, the results of this study show that better reproducibility data and the introduction of advanced statistical models are needed to produce robust classification models. Clearly, the sample size was small, representing the lower limit of practical application. However, the verification test emphasized that the methodology used in this work can provide a measure of confidence classification that may have practical significance. It was also demonstrated that the medium to cultivate bacteria must be same because it changes LIBS spectral fingerprint and affects classification capacity. The study will be extended to characterize different bacteria and, more importantly, to differentiate pathogenic bacterial strains, thus demonstrating medical diagnosis potential. This work is currently underway in our laboratories. The equipment and methods used in this work can be developed for quick, automatic, reliable and robust measurements in real time.

II – Rapid identification and discrimination of antibiotic resistant bacterial strains by laser induced breakdown spectroscopy and neural networks

Analysis of bacteria with major and minor mutations

4.2.1. ABSTRACT

Identification and discrimination of bacterial strains of same species exhibiting resistance to antibiotics using Laser Induced Breakdown Spectroscopy (LIBS) and Neural Networks (NN) algorithm is reported. The method has been applied to identify 40 bacterial strains causing Hospital Acquired Infections (HAI), i.e. *Pseudomonas aeruginosa*, *Escherichia coli*, *Klebsiella pneumoniae*, *Salmonella typhimurium*, *Salmonella pullurum* and *Salmonella salamae*. The strains analysed included both isolated from clinical samples and constructed in laboratory that differ in mutations because of their resistance to one or more antibiotics. Small changes in the atomic composition of the bacterial strains, as a result of their mutations and genetic variations, were detected by the LIBS-NN methodology and led to their identification and classification. This is of utmost importance because solely identification of bacterial species is not sufficient for disease diagnosis and identification of the actual strain is also required. The proposed method was successfully able to discriminate strains of the same bacterial species. The optimized NN models provided reliable bacterial strain identification with an index of spectral correlation higher than 95% for the samples analyzed, showing the potential and effectiveness of the method to address the safety and social-cost HAI-related issue.

4.2.2. INTRODUCTION

Hospital Acquired Infections (HAI) have been widely studied in the past 30 years and have been raised to top-priority issue due to the associated economic and social costs [187]. Therefore, many preventive campaigns as well as new protocols have been implemented [188,189]. Infectious diseases have been a major challenge to public health and remains as a leading cause of mortality and ill health worldwide. However, the rate of morbidity and mortality is found high in infants, children and people from developing countries [8]. On average 5-7% of hospitalized patients are affected by HAI, and 1% of such unwanted events result in the patient's death [190]. Bacteria are responsible for 95% of HAI, *Escherichia coli* (18.2%), *Staphylococcus* (18.1%), *Pseudomonas* (6.0%), *Enterococcus* (15.4%), *Klebsiella* (3.7%), *Acinetobacter* (0.8%), and *Salmonella* (2.8%) being the most relevant ones.

An important issue highlighted in recent years has been the increasing emergence of bacteria that are resistant to many antimicrobial therapies, sometimes resulting in multidrug-resistant strains or “super bugs”. Studies have reported that each hour of delay in the use of antibiotic lead to an increase in the mortality rate of 7.6 % [153]. This implies clinicians to start a treatment as early as possible; however, in many cases due to the difficulty of obtaining a quick diagnosis of the infection, an inappropriate and early use of antibiotics in an attempt to treat the infection is practiced, producing resistant bacteria [191]. This antibiotic resistance is evolved under the treatment regimens of single or multidrug combinations as a result of the mutations [192]. Problems associated with BSI include infection with multidrug resistant pathogens, which are difficult to treat and are associated with increased mortality [193]. Emerging infectious diseases (EIDs) are a significant burden on global economies and public health. 54.3% of EID events are caused by a large number of drug-resistant bacteria [9]. EPINE has reported the prevalence of

most common antibiotic drugs taken by the patients (Fig. 4.2.1) and consequently the same type of resistance was found in the bacteria isolated from biological samples. Therefore, the bacterial samples included in the present study were selected based on the resistance to a majority of these antibiotics.

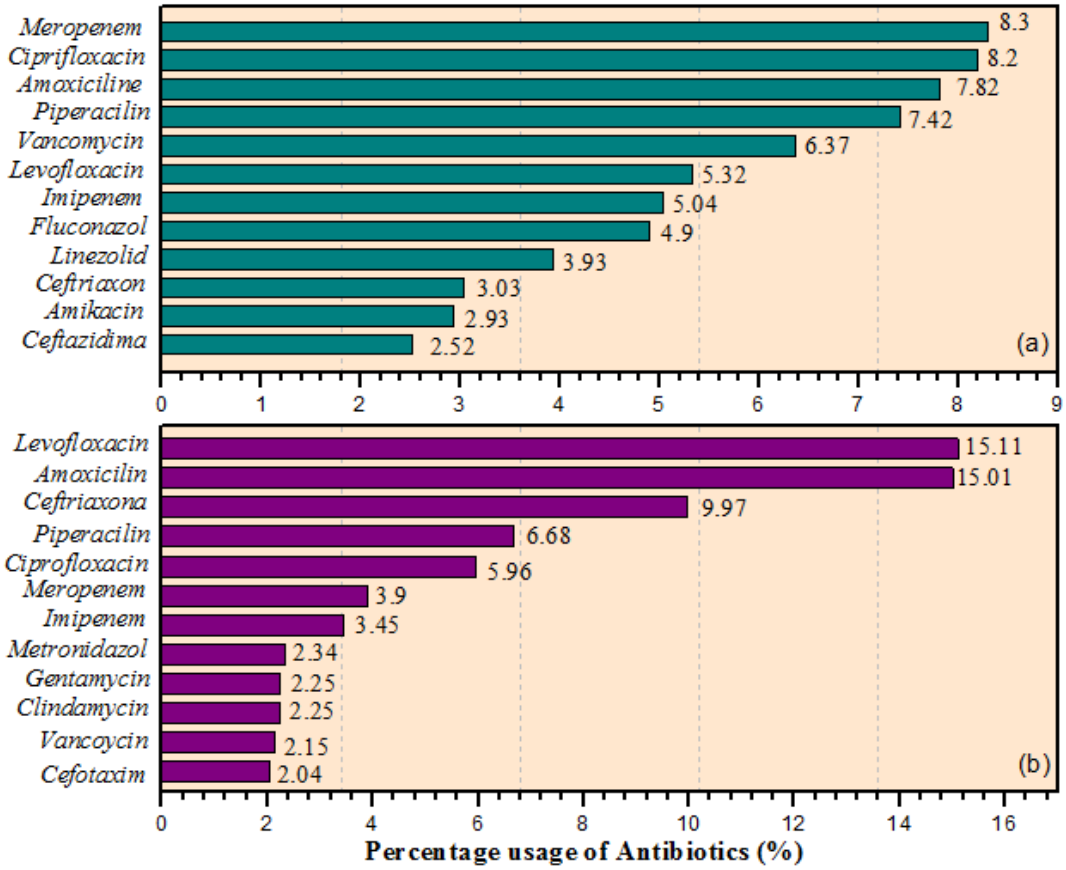


Fig. 4.2.1. Percentage of Antibiotic drugs used by patients of a) Hospital Acquired Infections and b) Community Infections

Most of these antibiotics belonged to β-lactams, Quinolones and Tetracycline. β-lactam antibiotics are responsible for causing changes in the bacterial cell wall [194]. Quinolones affect DNA and Tetracycline acts on ribosomes [195-197]. These modifications in the cellular structures might be helpful for their analysis by LIBS technique.

During the past decades several methods have been proposed to optimize the identification of bacterial strain, which are based on molecular techniques such as fluorescent probes [198], microarray assemblies [199,200] and polymerase chain reaction [201,202]. However, these methodologies present some difficulties and drawbacks such as use of consumables, primer, probes or fluorescently labeled RNA antibodies [47]. Moreover, sometimes the sequences in the database are not accurate or up-to-date and micro-heterogeneity is also found common in 16S rRNA gene sequence within a species [203,204]. The phenotypic similarities between the strains of the same bacterial species restrict their identification using routine diagnostic methods [205].

Although these methods provide a reliable and accurate bacterial identification, special sample treatment methods, the high costs and low speed to perform such analysis limit their use as rapid diagnostic methods in common laboratories in order to provide quick results which leads to an increase in the rates of infectious diseases in clinical settings [151]. Further, the direct handling of these potentially pathogenic bacterial samples poses health-associated security risks. At present, clinical safety, procedures and cost-related considerations do not allow an easy routine analysis of highly dangerous pathogenic bacterial specimens causing Hospital Acquired Infections. Nevertheless, bacterial identification within the first 24 hours of infection allows the use of a more effective and less risky targeted-therapy decreasing unnecessary hospitalization days and costs. It is also necessary to point out the importance of sample preparation, which is an important step to achieve a significant result within a reasonable amount of time, thereby, avoiding or reducing the need for time-consuming culture enrichment steps. Thus, the increasing need for high speed and precision illustrates the importance of sophisticated methods for sampling and sample preparation within the overall process. The proper development and adaptation of sample preparation towards the detection method, is

essential for exploiting the whole potential of the complete workflow of any diagnostic method.

Detection and identification of biological samples and, in particular bacteria, using Laser Induced Breakdown Spectroscopy (LIBS) analysis has been studied by several research groups [36,44,167,169,206,207]. The motivation of these studies was to evaluate the ability of LIBS to provide fast identification compared to traditional bioanalytical methods, benefitting from the possibility to combine it with chemometric methods in order to increase the performance of the technique. Morel et al. [36] investigated the detection of six bacteria by time resolved LIBS. They placed particular emphasis on *Bacillus globigii*, which acts as a non-pathogenic surrogate for *Bacillus anthracis* (anthrax), demonstrating the ability of LIBS to detect bacteria. Baudelet et al. [148] showed an unambiguous discrimination of different bacteria based on the concentration profile of trace elements. Recently, Rehse et al. [44,47], studied the effect of different experimental conditions (e.g., bacteria dilution and nutrient deprivation) on bacteria identification by discriminant functional analysis showing successful bacterial classification. Multari et al. [169] applied partial least squares regression analysis to differentiate *E. coli* from *S. aureus* strains. Although these studies present good results, in some cases correct identification rate of bacterial strains or correlation falls below 85%. Thus, there is a clear need for more thorough and systematic studies that include new approaches making it possible to take this methodology in clinical setting for diagnosis of diseases and public health. Therefore, the motivation behind this study is to use a classification model for bacterial identification using artificial intelligence algorithms like Neural Networks (NN) to improve the accuracy and precision of the classification process. In a previous study by our group [41] artificial intelligence algorithms like Neural Networks (NN) have been used that have shown to be a promising chemometric

methodology to classify and predict bacterial samples at genus level with a high degree of precision and accuracy. The full sets of variables (intensities at each wavelength) that constitute the sample spectrum are important in the process of comparison performed by the NN, which constitutes the basis of their ability to carry out discrimination. The NN is able to compute internal parameters (weights and bias) in the learning process for classifying a given set of input variables as belonging to particular sample with a high tolerance for noise and the presence of outliers [42].

In this study, we have intended to extend the previous study to investigate the application of LIBS-NN to discriminate different antibiotic resistant strains of same bacterial species and address its use as a rapid potential diagnostic methodology. The aim is to determine if genetic variations between bacterial strains of the same bacterial species, even when there is a difference in only one gene, generate sufficient or significant changes in their atomic composition, can be detected by LIBS-NN method in order to achieve their discrimination and identification.

4.2.3. MATERIAL AND METHODS

4.2.3.1. LIBS set-up

The LIBS technique and the methodology used in the present work together with the most significant experimental conditions have been described in the previous section [41]. Same experimental setup and parameters were established. A study of integration and delay time resulted in the selection 1 ms and 5 µs. The spectrometer was computer-controlled using an interface developed with Matlab, which allowed for data processing and real-time analysis.

4.2.3.2 Bacterial samples

All the bacterial strains were cultivated in LB agar (Difco Microbiology, Lawrence, KS, U.S.A) at 37 ºC for 12 hours in three Petri dishes (8.9 cm in diameter). A total of 40 strains of different bacterial species i.e. *Escherichia coli* (Ec) [208], *Pseudomonas aeruginosa* (Pa) [209], *Klebsiella pneumoniae* (Kp) [210], *Salmonella typhimurium* (St) [211], *Salmonella pullorum* (Sp) [211] and *Salmonella salamae* (Ss) [211] were included in the study. Kp, Ec and Pa strains showed multidrug antibiotic resistance and multiple genes mutations (Table 4.2.1), whereas St, Sp and Ss were resistant to kanamycin and differed in only one gene (Table 4.2.2). Two bacterial strains of the *Salmonella* species (strains 2 and 3 of Table 4.2.2) were constructed by inactivation by directed mutagenesis of specific gene in the wild type background (strains 1 of Table 4.2.2). Sample ID has been represented in XYZ format, where X is the genus, Y the bacterial species and Z the type of strain. Thus, for example, Kp1 refers to the first strain of *Klebsiella pneumoniae* (K21P).

Table 4.2.1. Description of bacterial strains with multidrug antibiotic resistance.

Bacteria	Strain	Antibiotic Resistance Phenotypes Diffusion Disk[(b)(c)]	Sample ID	Origin
Klebsiella pneumoniae	K21P[a]	Amp, Amo, Amc, Tet, Nal, Fox, Cip, Tsu, Ctx, Caz, Azt	Kp1	Peritoneal fluid
	K18P[a]	Amp, Amo, Amc, Clo, Tet, Nal, Cip, Stm, Tsu, Caz, Azt	Kp2	Urine
	K17P[a]	Amp, Amo, Amc, Tet, Nal, Cip, Tsu, Ctx, Caz, Azt	Kp3	Urine
	K16R[a]	Amp, Amo, Amc, Clo, Stm, Tsu, Ctx, Caz	Kp4	Urine
	K11CM[a]	Amp, Amo, Amc, Clo, Tet, Nal, Fox, Cip, Tsu, Ctx, Caz, Azt	Kp5	Urine
	K11P[a]	Amp, Amo, Amc, Clo, Tet, Nal, Cip, Stm, Tsu, Ctx, Caz, Azt	Kp6	Urine
	K7P[a]	Amp, Amo, Amc, Clo, Tet, Nal, Fox, Cip, Tsu, Ctx, Caz, Azt	Kp7	Urine
	K6P[a]	Amp, Amo, Amc, Clo, Tsu, Ctx, Caz, Azt	Kp8	Blood
	K3C[a]	Amp, Amo, Tet, Ctx (Int.)	Kp9	Urine
	K2P[a]	Amp, Amo, Tet, Nal, Tsu, Caz, Azt	Kp10	Urine
Escherichia coli	MC6-RP11	MC6RP1 *leu+ ftsA3*(Ts)	Ec1	
	QCB1	MC6RP1 *ponB*::*Spc*^{r}	Ec2	
Pseudomonas aeruginosa	PA1	Serotype 010,	Pa1	Secretion
	PA2	Serotype 014, Gen, Amk, Tob, Azt, Tic, Pip, Clo, Car, Eri.	Pa2	Secretion
	PA3	Serotype 011,	Pa3	Secretion
	PA4	Serotype N/A,	Pa4	Secretion
	PA5	Serotype 04, Azt, Tic, Pip, Cip, Clo, Car, Eri.	Pa5	Secretion
	PA6	Serotype 011,	Pa6	Secretion
	PA7	Serotype 014,	Pa7	Blood
	PA8	Serotype 01, Azt, Tic, Pip, Cip, Clo, Car, Eri.	Pa8	Secretion
	PA9	Serotype 09,	Pa9	Sputum
	PA10	Serotype 06, Azt, Tic, Pip, Cip, Clo, Car,Eri.	Pa10	Secretion
	PA11	Serotype 010,	Pa11	Sputum
	PA12	Serotype 015,	Pa12	Secretion
	PA13	Serotype 09,	Pa13	Blood
	PA14	Serotype 011,	Pa14	Secretion
	PA15	Serotype 011,	Pa15	Sputum
	PA16	Serotype 011, Gen, Amk, Tob, Azt, Tic, Pip, Imp, Cip	Pa16	Secretion
	PA17	Serotype 010, Gen, Amk, Tob, Azt, Tic, Pip, Cip, Imp, Clo, Car, Eri	Pa17	Blood
	PA18	Serotype 015,	Pa18	Secretion
	PA19	Serotype 011,	Pa19	Secretion

P, Peritoneal fluid; U, Urine; B, Blood; S, Secretion; Sp, Sputum

[(a)] *K. pneumoniae* isolates resistant or intermediate to cephalosporin by the disk diffusion method.

[(b)] *Amp*, ampicillin; *Amo*, amoxicillin; *Amc*, amoxicillin-clavulanate; *Clo*, chloramphenicol; *Tet*, tetracycline; *Nal*, nalidixic acid; *Fox*, cefoxitin; *Cip*, ciplofloxacin; *Stm*, streptomycin; *Tsu*, trimethoprim/sulfamethoxazole; *Ctx*, cefotaxime; *Caz*, ceftazimide; *Azt*, aztreonam; (Int.), Intermediate

[(c)] *Gen*, gentamicin; Amk, amikacin; Tob, tobramicin; Azt, aztreonam; Tic, ticarcilin; Pip, piperacillin; Cip, ciplofloxacin; Imp, imipenem; Clo, chloramphenicol; Car, carbenicillin; Caz, ceftazimide; Eri, erithromicin

Table 4.2.2. *Salmonella* strains that differ in a single gene.

Bacterial species	**Strain**			**Sample ID**
	1	2	3	
S. pollorum	1JVC	1/1Km	2/1Km	Sp
S. salamae	2JVC	1/2Km	2/2Km	Ss
S. typhimurium	SL-1344	1/22Km	2/22Km	St

JVC: Wild type codification

Km: Mutant in a gene by insertion of a Kanamycin resistance cassette

4.2.3.3. LIBS measurements and spectral libraries

Bacterial samples were measured directly in the Petri dish at room experimental conditions. It was observed that the water content of the bacterial samples affected the LIBS signals therefore; it was reduced by flowing air over the bacterial surface before LIBS spectra acquisition. For each bacterial strain, four replicate Petri dishes were measured. Eighty single-laser-shot spectra from one Petri dish and twenty single-laser-shot spectra each from the remaining three Petri dishes were acquired. The acquisition time of these 140 spectra was approximately 2 min. Because the emission intensity signal may vary with laser pulse, spectra were normalized by the most intense emission line, i.e. Na (I), to avoid instrumental variations (Fig. 4.2.2). The 80 spectra from the first Petri dish were used to train the NN model for each bacterial strain (spectral training library), whereas the 20 spectra from the remaining three Petri dishes each were used to test the model (spectral test library). Although the matrix dataset was considerably large, the computation time for training each NN model was below 10 s.

4.2.3.4. Neural Network model

The details about the topology, learning and verification process have been given in the previous sections. Therefore, only the essential details have been mentioned here. The NN topology consists of three layers (input, hidden and output), which is widely used to model systems with a similar level of complexity [212]. In particular, the input layer consisted of 2048 nodes (spectral response in the 200–1000 nm wavelength range). The number of neurons in the hidden layer was 10. The output layer (classification result) was comprised of J neurons (where J = number of reference samples used) for estimating the similarity between the reference sample spectra and the testing sample spectrum. Because the NN is a supervised method, in order to optimize the weight matrix it is necessary to use input and output data that adequately characterized the system to be modelled. The spectral data of the training library was randomly divided as a part of the training process into two subsets: 80% (64 spectra) for training and 20% (16 spectra) for self-validation of the model. Once the training and self-validation process was carried out, the models were validated by testing the 60 spectra from the remaining three replicas of the bacterial strains.

4.2.3.5. Neural Network model validation

Three validation tests were carried out for each NN model. *First validation* test performed was inter-bacterial species discrimination where the ability of NN models was evaluated to classify the bacterial strains based on their species. The *second* and *third* validation tests were performed to evaluate the capacity of the LIBS-NN methodology to discriminate between strains of the same bacterial species with small changes in the atomic composition because of their resistance, mutations and genetic variations. Second validation test included strains which differed in resistance to a set of different antibiotics, whereas, the third validation test included strains which differed in only one gene.

The model performance was evaluated by its accuracy, i.e. the rate of correct bacterial classification or correlation within the classified spectra. The model accuracy was estimated by the parameter "spectral correlation", which is expressed as the percentage of test-set spectra classified correctly. The higher the index of spectral correlation of a test set, the better the capacity of the NN model for discriminating a sample. A spectral correlation value higher than 90% was considered as correct identification of the sample.

4.2.4. RESULTS AND DISCUSSION

LIBS experiments were performed for some typical HAI-causing bacteria and therefore important pathogens from medical point of view. Fig. 4.2.2 shows a normalized LIBS spectrum of an Ec bacterial sample, identifying the elements with the most intense emission lines.

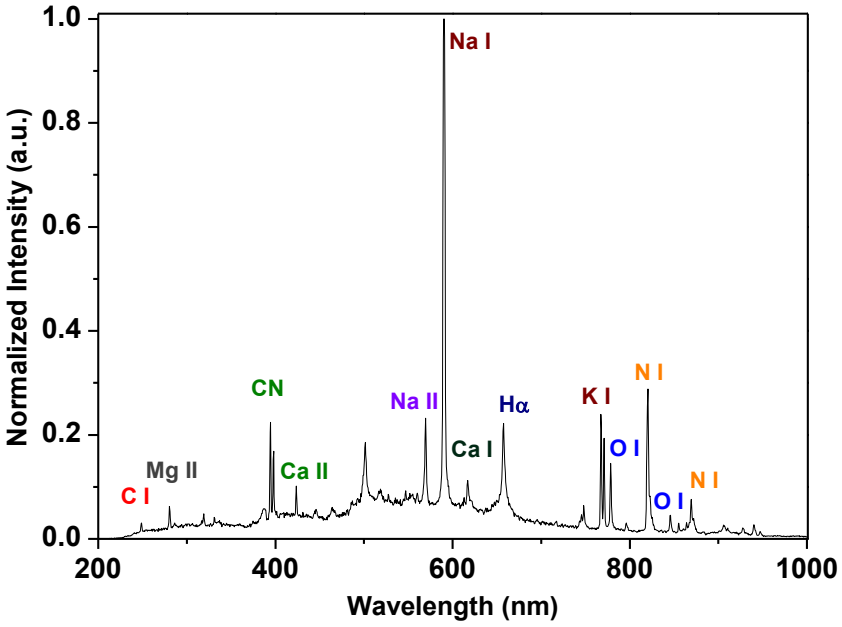


Fig. 4.2.2. Normalized LIBS spectrum of Escherichia coli

It has been shown that elements such as P, C, Mg, Ca and Na provide sufficient information to achieve bacterial identification [45,47,148]. However, due to the high similarity between the spectra of the bacterial strain samples (especially in the case of the differences is only in one gene), a broad spectral range (200-1000 nm) was used in order to cover the greater number of spectral characteristics and improve the performance of the NN model. Several papers have shown [213] that reducing the number of variables in the training of NN using spectral zones with few peaks, selected by PCA, decreases the performance of the NN model. Therefore, the selection of certain variables does not imply that the other variables are redundant with respect to selected. Thus, the larger the information (representative data) is used in the training of the NN model, the better is the predictive capability of the model. On the other hand, neural networks have a much better noise performance.

Despite that there are not significant variations in the spectra to easily discriminate the bacterial strains, from the mathematical point of view each bacterial strain can be discriminated based on its complete spectral fingerprint [42]. Therefore, the whole set of variables that constitute the sample spectrum is important in the classification process performed by the NN model by computing internal parameters (weights and bias) in the training step. This fact constitutes the basis of their ability to carry out the discrimination between the bacterial strains with high tolerance for noise and the presence of outliers.

Because samples were measured at room conditions, emission signals from air were also observed in the LIBS spectra. However, as it has been demonstrated, these signals do not significantly interfere the relative spectral contribution of the trace elements of the samples [182]. In order to improve the signal to noise ratio of the emission lines, the water content of the samples was reduced by flowing air before LIBS spectra

acquisition. On the other hand, a continuum background emission was also observed as a typical component of LIBS spectra.

4.2.4.1. First validation: Inter-bacterial Species Discrimination

Inter-bacterial species discrimination test involved the classification of bacterial strains based on their species. The capacity of LIBS-NN method to discriminate bacterial species, independent of the strain of the bacterial species used to train the NN model, was evaluated. For carrying out this validation test, the NN was trained with two libraries of different bacterial species, and tested with the libraries of the bacterial strains of all species. Even though it is possible to include the spectral datasets of all the bacterial strains in the training phase of the Neural Network, we have used only two bacterial libraries to train the neural network as a first approach to test whether there are sufficient significant spectral differences between different bacterial strains that allow for discrimination. *The results have been shown in* Fig. 4.2.3.

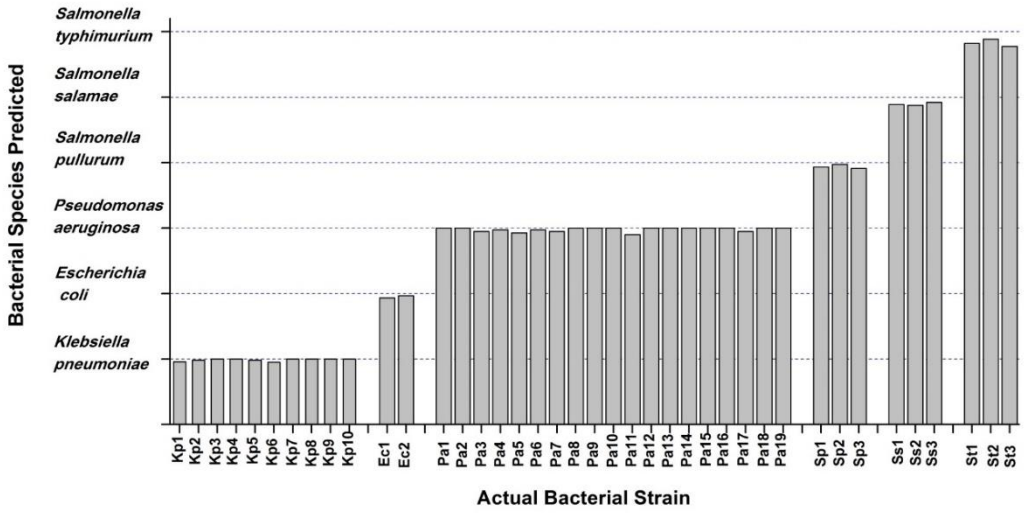


Fig. 4.2.3. Inter-bacterial species classification.

It can be seen that the LIBS-NN methodology developed was able to correctly classify all the bacterial strains to their corresponding bacterial species, achieving accurate and reliable results. All bacterial strains were assigned to their species (95% spectral correlation) independent of the bacterial strains used to train the NN model, demonstrating the robustness of the methodology. Hence, classification is based on the major differences in the spectral fingerprints of the bacteria at species level.

4.2.4.2. Second validation: Intra-bacterial strain discrimination of multidrug resistant bacteria

Strains of Kp, Ec and Pa which acquired resistance to multiple antibiotics were selected for LIBS analysis. These strains exhibited different antibiotic resistance patterns and, hence, different mutations. In this validation, NN models were estimated training the NN with two libraries of different strains of the same bacterial species and validated with all the test libraries of the strains of that species. An average spectral correlation for each bacterial strain and estimated model was calculated. Fig. 4.2.4 shows the classification results (spectral correlation) obtained for all the models and bacterial species.

A spectral correlation higher than 92%, 98.3% and 95% for *Klebsiella pneumoniae*, *Escherichia Coli* and *Pseudomonas aeruginosa* was achieved, respectively, demonstrating the good performance of the LIBS-NN methodology to classify the strain samples for each of the bacterial species. The fact that 100% spectral correlation was not achieved could be attributed to that each spectrum was recorded with a single laser shot. Nevertheless, this lack of complete accuracy was considered no significant in the overall performance of the LIBS-NN method developed to discriminate strains of different bacterial species. Therefore, these results showed that NN trained with datasets of different bacterial strains allowed to discriminate strains of the same bacterial species

with multidrug resistance and genetic variations that imparted changes in their elemental composition providing a characteristic fingerprint and hence, enabling LIBS-based bacterial identification.

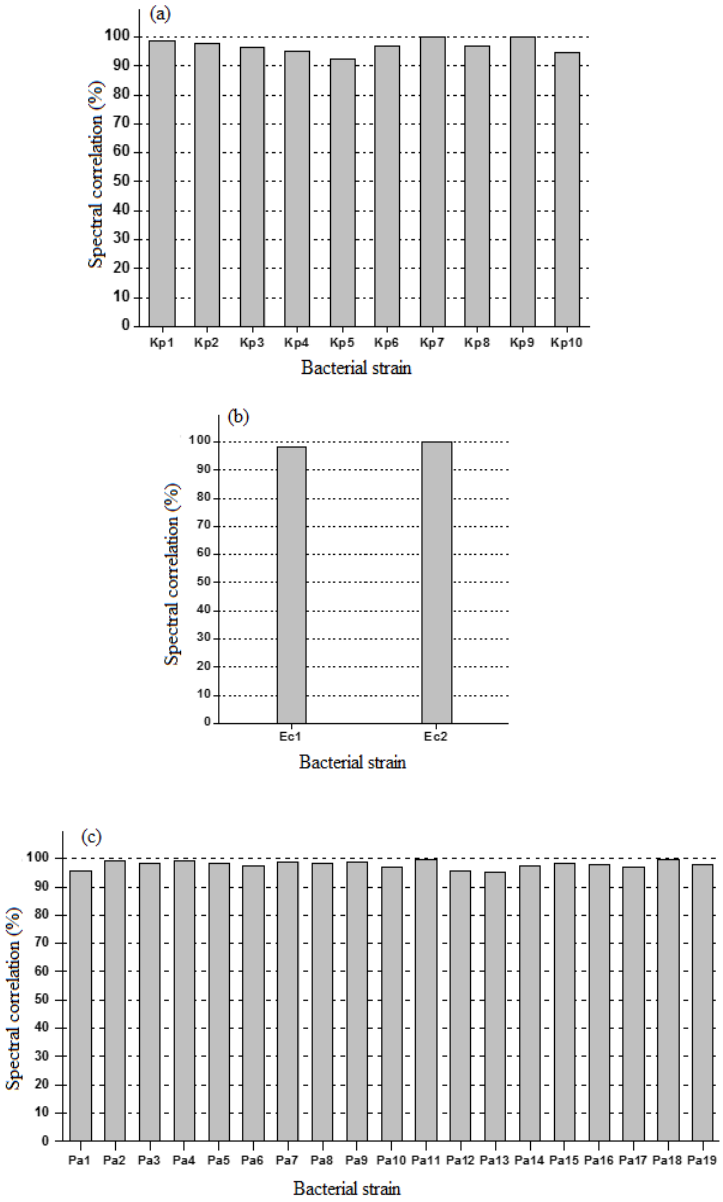


Fig. 4.2.4. Intra-bacterial strain classification of multidrug resistant strains of (a) Klebsiella pneumoniae (b) Escherichia coli (c) Pseudomonas aeruginosa.

4.2.4.3. Third validation: Single-gene intra-bacterial strain discrimination

A single-gene intra-bacterial strain discrimination of *Salmonella* species was carried out. Three different strains for each species with genetic variation of only one gene were selected. This minor variation in the genetic code of the bacterial strains made the training of the NN to take more time (ca. 3 second) in order to reach the minimum MSE. In this validation, NN models were estimated training the NN with two libraries of different strains of the same bacterial species and validated with all the strains' libraries of the same species. An average spectral correlation for the three estimated models for each bacterial strain and bacterial species was calculated. Table 4.2.3 shows that a spectral correlation higher than 95%, 95% and 96.6% for *S. pullorum*, *S. salamae* and *S. typhimurium* was achieved, respectively, demonstrating the capacity of the LIBS-NN methodology to classify bacterial strain samples that differed in a single gene.

Table 4.2.3. Single-gene Intra-bacterial strain classification results

Bacterial species	**Spectral Correlation %**		
	Strain		
	1	2	3
S. pullorum	97.3	95.0	97.0
S. salamae	95.0	96.0	97.2
S. typhimurium	96.6	97.5	97.5

These results demonstrate that when the NN models were trained with libraries of different species, they were able to correctly identify strains based on their species. Moreover, the discrimination capability of the NN models is not only based on their ability to find similarities between spectra but also due to their capacity to tolerate small

spectral variations among them. This feature led to classify the strains to their corresponding species and the intra-bacterial strain discrimination.

4.2.5. CONCLUSIONS

A method based on Laser Induced Breakdown Spectroscopy (LIBS) and Neural Networks (NN) algorithms was developed and applied to achieve rapid identification and discrimination of different bacterial species and strains causing Hospital Acquired Infections, including multidrug resistance and single gene variation bacterial strains based on their characteristic spectral fingerprint. Single shot LIBS measurements combined with a supervised Neural Network method were sufficient for a clear identification and classification of bacterial strains of different species due to the reliability and robustness of the estimated non-linear classification models.

The results show that the LIBS-NN methodology proposed is able to discriminate bacterial species and strains with high accuracy. Therefore, it may be considered a quick, simple and cost-effective alternative for the slower and more expensive biological methods to discriminate strains of the same bacterial species. The three validation tests carried out show the ability of the NN models for rapid identification of bacterial samples from species to strain level with minor genetic variations. Moreover, single gene variations are enough to discriminate strain of the same bacterial specie. From the medical point of view, these capabilities would allow an early diagnosis of the bacterial infections and their treatment, which may reduce the recurrence of HAI.

III – A Comparison of Linear and Non-linear Supervised Chemometric methods for the classification of microorganisms using LIBS spectra

4.3.1. ABSTRACT

Considering the case in routine laboratories where diverse microbiota is isolated from biological fluids and different bacteria are needed to be classified principally on the basis of their species and discriminated from other strains, an analysis by Laser Induced Breakdown Spectroscopy (LIBS) and chemometrics using different strains from five bacterial species was performed. Five most commonly used linear and nonlinear supervised chemometric methods such as Linear Discriminant Analysis (LDA), Soft Independent Modelling of Class Analogy (SIMCA), Partial least Squares–Discriminant Analysis (PLS-DA), Support Vector Machines (SVM) and Artificial Neural Networks (NN) were studied to perform the discrimination of bacterial samples based on LIBS measurements. Sensitivity, generalization ability and robustness were the parameters taken into account to select the chemometric method for the development of classification models in LIBS based diagnostic methodology. These strains included *Pseudomonas aeruginosa, Escherichia coli, Klebsiella pneumoniae, Salmonella typhimurium* and *Staphylococcus aureus*. The results showed that NN outperformed SIMCA, PLS-DA, LDA and SVM with respect to all the above-mentioned parameters to discriminate highly similar bacterial samples.

4.3.2. INTRODUCTION

Development of a methodology for the identification and discrimination of a microorganism in biological samples with a high accuracy is very important for the clinical laboratories in the diagnosis of infectious diseases. Furthermore, the technical, economical and temporal aspects of the microbial analysis involving acquisition of the data as well as the processing of data to provide results are of high significance for a methodology to be useful for a real time analysis. Laser Induced breakdown Spectroscopy (LIBS) due to its intrinsic properties and aided by the possibility to combine with chemometric methods may offer such a diagnostic tool in future to analyze the clinical samples. However, advancements on both levels are required to transform its microbiological application from LIBS laboratories to clinical set-ups.

LIBS performs an instant and simultaneous multi-elemental scanning of elements providing a "spectral fingerprint", which is composed by a huge amount of data [154]. Here, chemometrics serves as a useful tool to convert huge amount of measurements into significant and useful information for multivariate calibration and classification, thus contributing to a significant improvement in the analytical performances of LIBS technique [42,56,58,95-98,100,131,181,214-216]. In order to obtain maximum information from the raw data, researchers have developed and applied a variety of statistical and computational algorithms. These methods attempt to discover a relationship between objects (input) and classes (target), referred to as a model, which represents a set of features that define the classification process. New objects that are unknown to the model are predicted based on their correlation to a certain class (target) in the model. These algorithms used to develop a model are based on two different approaches i.e. linear (linear combination of variables) or nonlinear (employ complex algorithms). As nonlinear methods apply more complex algorithms, they tend to increase

the sophistication in finding the boundaries between classes, hence enhancing the classification capacity of the method [58]. The selection of a chemometric method for data analysis is one of the very challenging tasks because it depends on the addressed analytical application, required information and the structure of the data. Generalizing an algorithm or operational parameters to all tasks and sets of data is therefore not straightforward, hence, the raw data must be carefully studied and methods properly evaluated for a specific task in order to achieve a satisfactory classification result [217].

A wide range of chemometric approaches are available to carry out classification and discrimination processes [101,113,164,218,219] and various studies have focused on the comparison between the performance of chemometric methods on a wide range of samples and data sets [58,212,219-223]. Several studies have demonstrated the viability of LIBS spectrum to identify and discriminate bacterial samples employing a range of chemometric methods for the data analysis [37,43,45,47,148,169,206,207,224-226]. However, mostly the chemometric methods used are the ones available at hand, which might lead to a bias towards that method. Among various chemometric methods employed for bacterial identification such as Principal Component Analysis (PCA), Linear Discriminant analysis (LDA), Partial Least Squares - Discriminant Analysis (PLS-DA), Neural Networks (NN) and Support Vector Machines (SVM), this has not still been established that which method could provide a superior classification performance for a given set of data and choice of independent variables. In case of bacterial analysis, Putnam et al. [207] carried out a comparison between PLS-DA and Discriminant Function Analysis (DFA) performing a sensitive and specific genus level classification of LIBS spectra from live bacterial specimens using independent variable models and obtained a better performance by PLS-DA [227]. If LIBS-chemometrics is to be used as a routine

diagnostic method, standardization of the methodology on every level is required so that reproducible results can be obtained between laboratories.

Certain evaluation parameters such as sensitivity (testing the samples included in the training), generalization ability (samples of the same class not included in the training) and robustness (identifying a sample not belonging to any class as unknown or unassigned) are required to be taken into account when using a chemometric method to develop a classification model. In order to present a solution to real world problem, only performing the sensitivity test does not make much difference, therefore, an appropriate chemometric method must be able to perform a successful test on generalization ability and robustness [185].

In this work, the performance of five most commonly used linear and nonlinear supervised chemometric methods, i.e., Soft Independent Modelling of Class Analogy (SIMCA), Partial Least Squares–Discriminant Analysis (PLS-DA), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and Artificial Neural Networks (NN) have been evaluated for their performance to deal with the bacterial identification and discrimination.

4.3.3. MATERIALS AND METHODS

4.3.3.1. Experimental Setup

The experimental configuration and parameters are same as explained in section 4-I. The delay time for the acquisition of spectra was studied and optimized to 5 µs.

4.3.3.2. Samples and LIBS measurements

Nine bacterial strains belonging to five different species, i.e. *Klebsiella pneumoniae, Escherichia coli, Pseudomonas aeruginosa, Staphylococcus aureus* and *Salmonella typhimurium* were analysed to generate LIBS spectral data. Bacterial culture

preparation and LIBS spectra acquisition were performed following the same procedure as mentioned in the section 4-I. The bacterial strains along with the nomenclature used are given in table 4.3.1. For each bacterial sample, 100 single shot spectra were recorded. 80 spectra forming the spectral training libraries were used to calibrate the models, whereas 20 spectra formed spectral test library to test the models.

Table 4.3.1. Nomenclature of the bacterial samples selected in this study

Bacteria	Strain	Sample ID
Klebsiella pneumoniae	KBC	Kp1
Escherichia coli	MC6-RP11	Ec1
	MG1655	Ec2
	QCB1	Ec3
Pseudomonas aeruginosa	M8A4	Pa1
	M9A1	Pa2
	M8A1	Pa3
Staphylococcus aureus	CP	Sa1
Salmonella typhimurium	LB5010	St1

4.3.4. DATA ANALYSIS METHODS

All data analysis was performed using Matlab software (Mathworks, 2012a). PCA and NN models were developed using home-made routines. SIMCA, LDA, PLS-DA was performed using the freeware toolbox developed by Ballabio et al [103]. For all the chemometric methods used, entire spectral range was considered. The spectral libraries were constructed with intensities (2048 variables) at each wavelength versus number of spectra (observations). The spectra were normalized by the most intense emission line, i.e. Na I, in order to avoid instrumental variations. The samples used for the calibration

of classification models were Kp1, Ec1, Pa1, Sa1, St1. In sensitivity test, the number of classes (N) was five whereas in robustness test N = 4. Once the models were generated, they were validated using the spectral test libraries and prediction results were obtained. The essential details related to the development of classification models by each chemometric method is given below:

4.3.4.1. Soft Independent Modelling of Class Analogy (SIMCA)

In order to calibrate SIMCA classification model, a prior PCA is run for each class. Leave-one-out cross-validation procedure was used to obtain the number of Principal Components (PCs). In the validation step, each spectrum is assigned the class membership based on its analogy to a reference class. The distance between the spectrum and centre of each class is calculated and finally the spectrum is allocated to the class with highest probability of membership. A spectrum was classified as unknown if its residual variance exceeded the upper limit for every class. This limit was fixed following the method described by Sirven et al., [228], assuming that the distances of training sample to centre of the corresponding class followed a normal distribution. A spectrum was considered as unclassified when the membership probability for every class was less than 5%.

4.3.4.2. Partial least Squares – Discriminant Analysis (PLS-DA)

The classification model by PLS-DA was generated using both PLS1 and PLS2 algorithms. The calibration was performed using two matrices, **X (**reference) composed by spectral training libraries, and **Y** (target) composed by the class information. For Sensitivity test, **Y** matrix had a dimension of 400 X 25, obtained by associating 80 spectra of each training library with a vector, such as (1,0,0,0,0) for class 1 “Kp1”, (0,1,0,0,0) for

class 2 “Ec1, and so on for remaining classes. The prediction results for the test spectra in validation of the model are given in the form of a vector where each element represents the probability of the spectrum to each class.

4.3.4.3. Linear Discriminant Analysis (LDA)

LDA model was estimated using 5 classes as mentioned above, taken simultaneously, creating N-1 discriminant function. Afterwards the tests libraries are introduced into the model to obtain the prediction membership results. Being a method for the discrimination of objects, it maximizes inter-class variance to the intra-class variance ratio [229]. In this case, test spectra are always classified to a class and no unknown assignation can be made.

4.3.4.4. Support Vector Machines (SVM)

SVM classification model was created by optimizing two parameters i.e. regularization parameter (γ) and RBF kernel function parameter. Parameter “γ” allows establishing a trade-off between maximizing the models´ performance and minimizing model complexity. RBF kernel function is the bandwidth that defines the nonlinear mapping from input space to some high-dimensional feature space [130,131],. A fine search of the γ and C model parameters was performed in order to select the values that better fit the data and provide the minimum error in the cross-validation.

4.3.4.5. Artificial Neural Networks (NN).

NN model was developed in the same way as explained in section 4-I. The calibration was carried out with the same training spectral libraries as in all methods. Each individual was assigned with an identification number in the output layer. Once a suitable

training is achieved, all spectral test libraries for all individual were introduced into the NN model for validation.

4.3.5. RESULTS AND DISCUSSION

Three performance evaluation criteria i.e. sensitivity, generalization ability and robustness of the models were assessed in a stepwise manner to determine the most suitable classification method [41]. The successful classification rate for each model, determined by calculating the percentage of bacterial samples correctly classified over the total number of samples was used as a measure of the classification sensitivity, generalization ability and robustness.

To consider bacteria correctly classified, the prediction of the model must match with the actual class by an arbitrary threshold of Spectral Correlation (SC) higher than 90% and less than 20% to the other classes, otherwise it was considered incorrectly classified. A bacterial sample was classified as unknown when the SC was less than 90% for all classes. For a single spectra to be correctly classified a probability of 5% as a decision threshold was used. Thus, any spectra surpassing this limit was allocated as unknown or unassigned spectrum.

4.3.5.1. Sensitivity test

The sensitivity of the models was evaluated by their capacity to classify the spectral test libraries of the samples (Kp1, Ec1, Pa1, Sa1 and St1) used in the calibration process. The results of classification by all models are given in Table 4.3.2, in terms of the spectral correlation obtained for each set of 20 test spectra. For SIMCA model, the spectral correlation for all test sets was very low and less than 90%. Only in case of St1,

SC obtained was higher but it was still less than the threshold established to obtain a correct classification.

Table 4.3.2. Classification results in Sensitivity test by five classification models

		Predicted (Spectral correlation %)						
	Actual	***Ec***	***Pa***	***Kp***	***Sa***	***St***	***Unknown***	***Correctly classified***
SIMCA	Ec1	10	40	0	0	0	50	×
	Pa1	0	55	0	0	0	45	×
	Kp1	0	0	0	0	0	100	×
	Sa1	0	45	0	10	0	45	×
	St1	0	0	0	0	75	25	×
PLS-DA	Ec1	75	0	0	5	0	20	×
	Pa1	5	70	0	15	0	10	×
	Kp1	0	0	100	0	0	0	✓
	Sa1	0	0	0	100	0	0	✓
	St1	0	0	0	0	95	5	✓
LDA	Ec1	90	0	0	10	0	-	✓
	Pa1	60	35	0	5	0	-	×
	Kp1	0	0	100	0	0	-	✓
	Sa1	5	0	0	95	0	-	✓
	St1	0	0	0	0	100	-	✓
SVM	Ec1	90	0	0	0	10	0	✓
	Pa1	0	100	0	0	0	0	✓
	Kp1	0	0	100	0	0	0	✓
	Sa1	0	0	0	100	0	0	✓
	St1	0	0	0	0	100	0	✓
NN	Ec1	90	0	0	0	0	10	✓
	Pa1	0	100	0	0	0	0	✓
	Kp1	0	0	100	0	0	0	✓
	Sa1	0	0	0	100	0	0	✓
	St1	0	0	0	0	100	0	✓

The SIMCA model gave a classification success rate of 0% as none of the bacteria was correctly classified. However, PLS-DA and LDA classification models showed a better sensitivity as can be seen from the spectral correlation obtained for the test-sets. A

classification success rate of 60% and 80% was achieved, respectively, along with a decrease in the percentage of unknown spectra PLS-DA. The highest classification success rate in sensitivity test was obtained by SVM and NN models. For all of the spectral test libraries, a spectral correlation higher or equal to 90% was provided. Although it is noteworthy that NN did not provide any false positives whereas SVM classified two spectra of Ec1 as St1. Both classification models, NN and SVM, provided a classification success rate in sensitivity of 100%.

4.3.5.2. Generalizability

In the generalizability test (Internal independent validation), the models created previously for the sensitivity test were tested by other strains not used in the training of the model. For this validation, two remaining strains of *E. coli*, Ec2 and Ec3, and of *P. aeruginosa*, Pa2 and Pa3 were used as spectral test libraries. As the comparison was performed to find the most suitable classification method, taking into account the prediction results by all models in the sensitivity test, the models with high classification success rate were selected for generalizability test. Therefore, LDA, SVM and NN models selected. The classification results of generalization test are given in Table 4.3.3. The classification success rate by LDA model was 50%, moreover, more false positives were obtained as reflected by the spectral correlation for Ec2 and Pa3. Furthermore, LDA is incapable of producing “unknown” assignations and produced higher false positives, it could lead to incorrect results. SVM demonstrated to be a sensitive method on the calibration samples but provided false positives for Ec2 and Ec3, classifying them incorrectly as *S. typhimurium*, thus incorrectly identifying the species of the bacterial strains. Thus, a low classification success rate of 50% was achieved for SVM in the generalization test. Whereas NN model still produced a high spectral correlation >90%

for each set and therefore a high classification success rate in generalizability. Comparing the results of these three methods, it is clear that in case of a real world application where different strains can be isolated from biological samples and genus- and species- specific identification is needed, NN models can be a better option.

Table 4.3.3. Classification results in Generalizability test by selected models

		Predicted (Spectral correlation %)						
	Actual	***Ec***	***Pa***	***Kp***	***Sa***	***St***	**Unclassified**	***Correctly classified***
LDA	Ec2	55	40	0	5	0	-	×
	Ec3	95	5	0	0	0	-	✓
	Pa2	0	100	0	0	0	-	✓
	Pa3	0	25	0	75	0	-	×
SVM	Ec2	5	0	0	10	85	0	×
	Ec3	0	0	0	40	60	0	×
	Pa2	0	100	0	0	0	0	✓
	Pa3	0	100	0	0	0	0	✓
NN	Ec2	95	0	0	0	0	5	✓
	Ec3	100	0	0	0	0	0	✓
	Pa2	0	100	0	0	0	0	✓
	Pa3	0	95	0	0	0	5	✓

4.3.5.3. Robustness Test

Moreover, it is also important that the model shows robustness, that is, the ability to identify bacterial samples not included in the training step as an unknown and not classifying as another bacteria [230]. This can be a normal situation in a clinical laboratory when a new or bacterial samples different from the species or strains used in calibration is isolated and it is required to be classified correctly as not belonging to any class. In order to perform this test, new models were generated by calibrating with four

classes, removing the fifth class and introducing its spectral library to validate the classification model. As NN model showed high sensitivity as well as generalizability, its capacity to provide robustness was tested. The results obtained were then analysed in terms of the probability of correct classification and are given in the table 4.3.4. In order for achieving a correct classification as "unknown" the spectral correlation criteria was kept as < 20% for each class. In all of the tests this criteria was fulfilled and a high robustness was produced.

Table 4.3.4. Results of External Independent validation (Robustness)

	Robustness test						
		Predicted (Spectral correlation %)					**Correctly unknown**
	Actual	***Ec1***	***Pa1***	***Kp1***	***Sa1***	***St1***	
NN	Ec1	*	10	0	5	0	✓
	Pa1	5	*	0	0	0	✓
	Kp1	0	0	*	10	5	✓
	Sa1	5	0	5	*	15	✓
	St1	0	5	15	0	*	✓

A comparison of the overall performance of the models has been shown in Fig. 4.3.1. SIMCA was unable to classify spectra not used in the calibration, showing both low sensitivity. PLS-DA and LDA provided better sensitivity than SIMCA but still in case of PLS-DA classification the classification success rate was not considered acceptable to deal with bacterial identification. Due to 100% sensitivity by SVM and NN and higher sensitivity of LDA model as compared to SIMCA and PLS-DA, they were tested for generalization ability. Among these models, NN provided the best results but high misclassification rate was obtained in LDA and SVM. NN model was tested for its

robustness where it also provided high classification success rate. The results indicate NN outperformed all the models giving the best results in sensitivity, generalizability, robustness and time-dependent capacity. Thus, despite highly similar composition of bacterial samples a correct classification was achieved by NN.

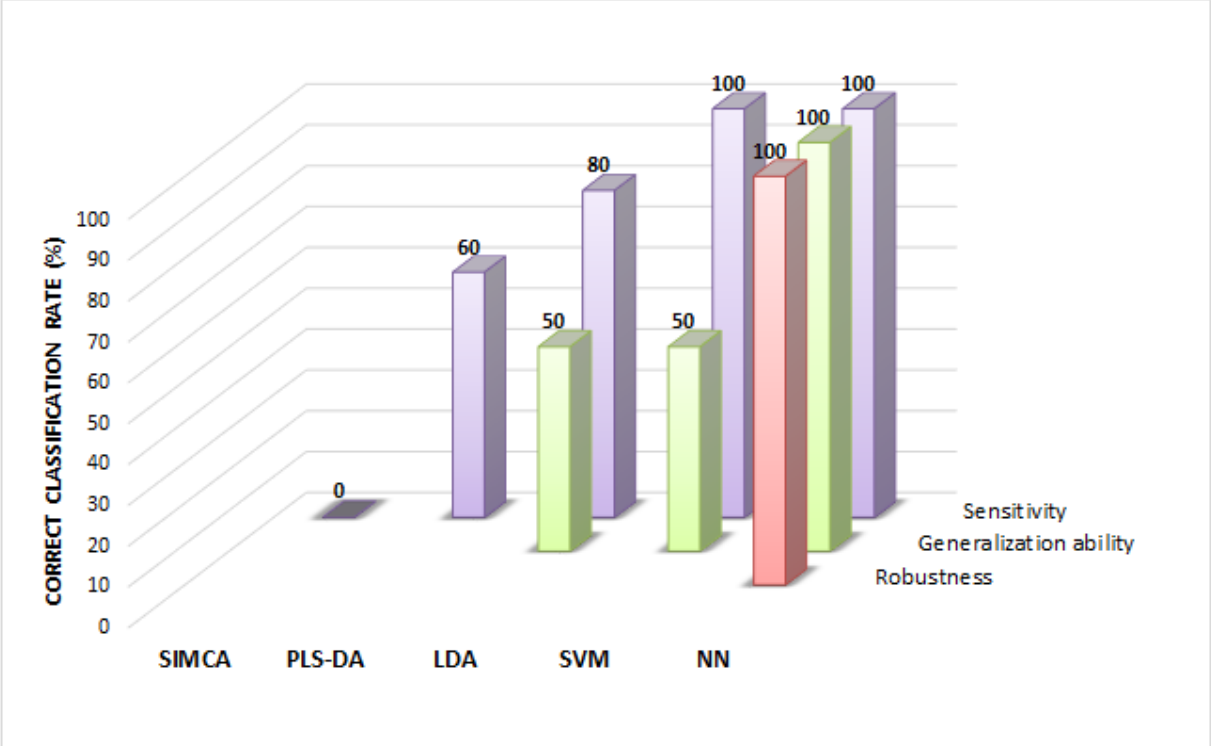


Fig.4.3.1. Stepwise selection of chemometric method for bacterial classification based on the performance evaluation parameters

4.3.6. CONCLUSIONS

Laser Induced Breakdown Spectroscopy (LIBS) technique and five supervised chemometric methods (SIMCA, PLS-DA, LDA, SVM and NN) have been evaluated to develop a methodology for the identification and discrimination of bacteria. SVM and NN performed best in sensitivity, however, the generalizability was found highest for NN. This can be attributed to the non-linear nature of algorithm used for the model generation and its ability to model complex nonlinear input-target relationships. As NN

models provided the best results in all evaluation parameters, LIBS combined with NN seem to offer a robust methodology for bacterial classification, taking into account the entire spectral fingerprint. Working with NN reveals that these algorithms may be more difficult to implement than other standard chemometric methods, due to the diversity of functions and architectures, the results obtained in this study demonstrate that LIBS combined with NN offers significant advantages as a screening tool, given its speed, high throughput, minimal destructive, micro-analysis and ease of use.

Chapter 5

Laser Induced Breakdown Spectroscopy and Neural Networks for the discrimination of Candida strains

5.1. ABSTRACT

The present study reports the evaluation of Laser Induced Breakdown Spectroscopy (LIBS) and Neural Networks (NN) for the discrimination of different strains of various species of *Candida*. This genus of yeast was selected due to its medical relevance as it is commonly found in cases of fungal infection in humans. Twenty-one strains belonging to seven species of *Candida* were included in the study. Scanning Electron Microscopy with Energy-Dispersive X-ray Spectroscopy (SEM-EDS) was employed as a complementary technique to provide information about elemental composition of *Candida* cells. The use of LIBS spectra in combination with optimized NN models provided reliable discrimination among the distinct *Candida* strains with a high spectral correlation index for the samples analysed, without any false positive or false negative. Therefore, this study indicates that LIBS-NN based methodology has the potential to be used as fast fungal identification or even diagnostic method.

5.2. INTRODUCTION

Biomedical applications of Laser Induced Breakdown Spectroscopy (LIBS) have been extensively explored since the last decade. Among these studies the analysis of microorganisms especially bacteria with particular relevance to infectious diseases in humans have made a greater contribution [41,47,148,224,231,232]. The increased interest in this area has been due to the advantages offered by LIBS in providing a speedy and cost effective analysis as well as to the use of modern chemometric techniques to deal with the large amount of data produced by LIBS. Whereas most of the studies have focused on the identification and discrimination of bacterial samples [47, 57,149,184,232,233] , not much has been done towards the analysis of fungi samples by LIBS. Hybl *et. al* [38] and Samuels *et. al.* [39] in 2003 analysed few fungal specimens such as *Penicillium, Corn Smut, Oat Smut* and *Alternaria tenuis, Cladosporum herbarum* respectively. Their work was based on studying the elemental composition and elemental ratios obtained by the observed emission lines in order to discriminate them from other bacterial samples. These studies were the pilot studies in LIBS based microbial analysis but posteriorly all studies dealt with identification and discrimination of pathogenic bacteria.

Among the fungal pathogens that cause the infections, species of *Candida*, *Aspergillus* and non-*Aspergillus* molds appear to be on the rise in the patients. Infections from these pathogens range from localized infections (endogenous microbiota of digestive and urogenital tract) [234,235] to widespread haematogenous (blood stream infections) and mucocutaneous dissemination [236], include patients undergoing blood and organ transplantations, surgeries, AIDS, cancer and premature infants [237-239]. Epidemiological studies reveal that among all these fungal pathogens *Candida* species

are ubiquitous and found to be the most common causative agents of these infections [240-243].

Studies report an increase in these fungal infections [244,245] being *Candida* to be a leading cause of nosocomial infections posing a major concern associated with a great deal of attributable morbidity and mortality accompanied by an increased hospitalization period and treatment costs [237,246,247]. In fact, *Candida* is the fourth most commonly recovered organism from blood cultures of hospitalized patients, with an estimated mortality rate of 38 to 75% [248]. Candidiasis poses considerable *diagnostic and therapeutic challenges and t*his implies the need of improvement on one hand in the defining preventive measure and standards of care while very importantly develop fast diagnostic techniques to start a timely and targeted antifungal therapy [249,250].

Out of approximately 200 species of the genus *Candida*, about 20 species are implicated in clinical infections. Among them, *C. albicans*, *C. glabrata*, *C. tropicalis*, *C. parapsilosis* and *C. krusei* are found more frequently than others. Some species such as *C. guillermondii*, *C. kefyr* and *C. dubliniensis*, although are less frequent, also cause serious infections [234,244,245,251]. Therefore, keeping in view the medical importance of fungi, the *Candida* genus and these strains were selected for this study.

The taxonomic identification of a fungal sample is conventionally done by methods based on the analysis of macro and microscopic morphology and physiology. A common method relies on the analysis of the ability to develop germ tubes, which is a characteristic feature of *Candida albicans* that allows its differentiation from the non-*albicans* species. However, 5% of *C. albicans* isolates do not produce germ tubes, while some *C. tropicalis* isolates exhibit germ tube formation. Other identification procedures include biochemical-based assays, like carbohydrate fermentation and assimilation,

serological and molecular analyses such as ELISA and PCR tests, chromogenic studies and fluorogenic assays [240].

Moreover, sophisticated spectroscopic techniques like Fourier transform infrared spectroscopy (FTIR), Raman and matrix-assisted laser desorption/ionization time of flight/Mass spectrometry (MALDI-TOF/MS) have also been employed to identify *Candida* isolates [236,248,252,253]. Conventional methods are time consuming and may take up to 96 hours for a complete identification. Main drawbacks of most of the molecular methods are the high cost of the analysis and the requirement of complex equipment [240,254]. Thus, there is a clear need for rapid and cost-effective novel approaches to be used in clinical setting for the diagnosis of fungal diseases.

Having LIBS focused more towards bacterial identification, the utility of this technique in the biomedical field for the analysis of fungal strains remains to be established. Therefore, the motivation of this study was to evaluate the potential of LIBS technique in achieving discrimination between different *Candida* strains based on the information given about the LIBS elemental composition. Furthermore, Neural Network algorithms were applied to the analysis of LIBS data in order to design the classification models to identify and discriminate the *Candida* strains. In order to reveal the cellular structure, three-dimensional visualization [255] and chemical elemental composition of *Candida*, analysis by scanning electron microscopy (SEM) and Electro Dispersion X-Ray Spectroscopy (EDS) was also performed.

5.2. MATERIAL AND METHODS

5.2.1. Samples

Keeping in view the medical importance of microorganisms, medically relevant species and strains of bacteria and fungi were analysed included in this thesis. It has

already been mentioned in Chapter 1 that the focus was particularly to the pathogens that are common in the nosocomial infections or HAI. Among the fungi samples different species and strains of genus Candida composing *Candida parasilopsis, Candida dubliniensis, Candida krusei, Candida glabrata, Candida guilliermondii, Candida tropicalis* and *Candida albicans* were analysed. Type of strains and the details about these strains are given in respective sections.

5.2.2. Sample preparation

In case of Candida samples selective media was used for their growth. YPD (10 g/L of Yeast extract, 20 g/L of Peptone, 20 g/L of Dextrose and 20 g/L of agar) plates were prepared and Candida cells were grown at 30 °C for 48 h. Then, cells were resuspended in sterile water and 70 µL of this suspension spread on the surface of a small petri dish (5 cm in diameter) containing selective Sabouraud Dextrose Agar medium (5 g/L of pancreatic digest of casein, 5 g/L of peptic digest of animal tissue, 40 g/L of dextrose, 15 g/L of agar and 0.05 g/L of chloramphenicol for 48 hours at 30 ºC. In order to ensure the presence of only the Candida cells, antibiotic "chloramphenicol" was also added to the medium once it was cooled. This inhibits the growth of bacteria in the fungal culture.

5.2.3. Sample preparation

All samples were used with no further preparation than that described herein. *Candida* strains were grown on YPD plates (10 g/L of Yeast extract, 20 g/L of Peptone, 20 g/L of Dextrose and 20 g/L of agar) for 48 hours at 30 ºC. Then, cells were resuspended in sterile water and 70 µL of this suspension spread on the surface of a small petri dish (5 cm in diameter) containing selective Sabouraud Dextrose Agar medium (5 g/L of pancreatic digest of casein, 5 g/L of peptic digest of animal tissue, 40 g/L of dextrose, 15

g/L of agar and 0.05 g/L of chloramphenicol (to inhibit the growth of bacteria) for 48 hours at 30 ºC. Table 1 shows the *Candida* strains analyzed. Strains were kindly provided by Ana Alastruey-Izquierdo from the Instituto de Salud Carlos III, Spain (CNM-strains) and by Elvira Marín from Dept. Microbiología II, Universidad Complutense, Spain.

Table 5.1 shows the *Candida* strains analyzed. Strains were provided by Ana Alastruey-Izquierdo from the Instituto de Salud Carlos III, Spain (CNM-strains) and by Elvira Marín from Dept. Microbiología II, Universidad Complutense, Spain.

Table 5.1. *Candida strains used in this study.*

Candida spp.	**Strain**	**Sample ID**
C. parapsilosis	*ATCC CP22019*	*CpA*
	CNM-CL9608	*CpB*
	CNM-CL9607	*CpC*
C. dubliniensis	*CBS 7987*	*CdA*
	CNM-CL9531	*CdB*
	CNM-CL9472	*CdC*
C. krusei	*ATCC 6258*	*CkA*
	CNM-CL9526	*CkB*
	CNM-CL9524	*CkC*
C. glabrata	*CBS138*	*CglA*
	CNM-CL9600	*CglB*
	CNM-CL9555	*CglC*
C. guilliermondii	*CNM-CL9533*	*CgA*
	CNM-CL9603	*CgB*
	CNM-CL9602	*CgC*
C. tropicalis	*CNM-CL9537*	*CtA*
	CNM-CL9690	*CtB*
	CNM-CL9470	*CtC*
C. albicans	*CNM-CL9534*	*CaA*
	CNM-CL9535	*CaB*
	CNM-CL9536	*CaC*

5.2.4. Mounting protocols for LIBS analysis

For the study of *Candida* samples, during the initial attempts to optimize the LIBS signal and deal with the sample manipulation in case of microbiological samples, different sample preparation methods were used such as drying the biomass, preparing a suspension followed by filtration, freezing the culture and direct irradiation. The mounting protocols are described as under:

- *Drying the biomass*

The *Candida* cells were collected from the petri dish using a spatula and deposited onto a filter paper. It was then left to dry for two hours by flowing the air over the surface.

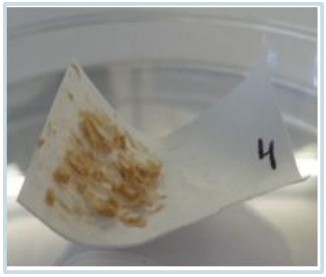


Fig. 5.1. Candida biomass harvested from agar medium and dried on filter paper

- *Preparing a suspension and filtration*

2 g of *Candida* cells were mixed with distilled water to get a final volume of 10 mL. The suspension was made and filtered through a filter paper using vacuum suction filter device, confining to a circle of diameter of 1.5 cm in order to get a uniform layer of *Candida* cells.

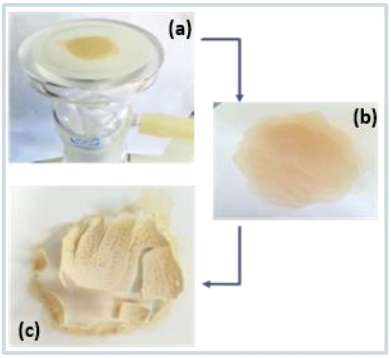


Fig. 5.2. Sample manipulation protocol showing a) filtration of Candida cells suspension b) formation of smooth layer c) appearance of cracks and folding of the layer

- *Freezing the culture*

A part of the layer of *Candida* cells along with the bottom agar medium was cut and placed into a second petri dish. This piece was frozen to –18 ºC into a solid sample.

Fig. 5.3. Piece of frozen culture along with agar media

- *Direct irradiation*

As evident, it involved irradiating the laser beam directly on the *Candida* in the petri dish above the agar medium.

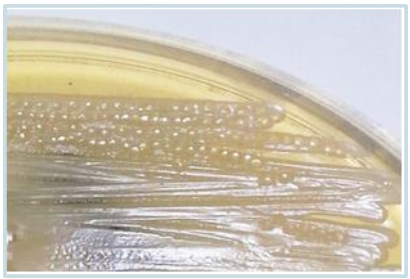


Fig. 5.4. Direct analysis in the petri dish

5.2.5. Experimental Set-up

The LIBS technique and methodology used have been previously described [42]. The more relevant parameters are discussed here. The experimental configuration used for the LIBS measurements included Q-switched Nd:YAG laser (Quantel, Brio model) at 1064 nm, with a pulse duration of 4 ns full width at half maximum (FWHM), 4 mm beam diameter and 0.6 mrad divergence. The laser beam was focused onto the sample surface with a 125 mm focal-distance lens. The diameter of the spot on sample surface was ≈ 150 μm. The laser fluence was set at 20 J/cm^2 per pulse and the repetition rate was set to 1 Hz. Collection of light was done by a 4 mm-aperture fiber optic, (with a 1000 μm core diameter and 0.22 numerical aperture), coupled with a 7 mm focus fused silica collimator placed at 45 degrees with respect to the surface normal, and at a distance of 5 cm from the sample. The optical fiber was coupled to the entrance of the spectrometer.

The spectrometer system was a user-configured miniature single-fiber system (USB4000, Ocean Optics, Dunedin, FL, U.S.A., grating 600 l/mm, spectral resolution 1 nm, 10 μm entrance slit) with a gated CCD detector. The spectral range covered was from 200 to 900 nm. The detector integration time was fixed at 100 ms, obtaining entire spectral information for ions, atoms and molecules. It is necessary to point out that the optical transmission properties of the optical fibre, the wavelength dependence of the spectrometer, and the wavelength efficiency of the detector elements all contribute to an overall wavelength dependence in the sample signal [41]. A delay of 2 μs was set to trigger the detector using a digital delay generator (Stanford model DG645) in order to prevent the detection of bremsstrahlung. The spectrometer was computer-controlled using an interface developed with Matlab (Mathworks, 2014b).

Scanning Electron Microscopy (SEM) analyses was carried out using a microscope (Hitachi S-3000N). Energy – Dispersive X-Ray Spectroscopy (EDS) analyses

were performed with an attached EDS analyzer (Oxford Instruments INCAx-sight). In the present study, *Candida* samples were analysed using Hitachi S-3000N microscope, to study the external morphology of the *Candida* culture. Energy – Dispersive Spectroscopy (EDS) analyses were performed for the analysis of elemental composition of *Candida* samples with an attached EDS analyser (Oxford Instruments INCAx-sight). The SEM-EDS analyses were performed at high vacuum with an accelerating voltage of 3.5 kV and at a working distance of 4 mm.

5.2.6. LIBS Measurements

A single spectrum was obtained for each laser shot, recording 150 spectra for each petri dish. From the recorded set of spectra, 100 were used to create the training library while the second set of 50 spectra composed the test library. Each library contains the intensity at different wavelengths in rows and the spectra in columns. Thus, our training library has 3648 rows (one for each wavelength) and 100 columns or spectra for each sample. The remaining 50 spectra were used to construct the test library in similar way. The pre-treatment of data was done by normalizing the spectra using the intensity emission line H_α (i.e. 656 nm) in order to avoid data variations.

5.3. NN Analysis

Statistical analysis of LIBS spectral data was performed by artificial intelligence algorithms i.e. NN in order to develop the classification models. The models were generated in a similar way as discussed previously in chapter 5 and only the more relevant aspects are given here. In order to estimate the NN model, the spectral data of the training library was randomly divided into two subsets: 80% for training and 20% for self-validation of the model. The number of neurons in the hidden layer was optimized to 10

and the number of outputs corresponded to the number of strains included in the training of NN model. Once the training and self-validation process was carried out, the models were validated by test libraries. The “spectral correlation” (SC) parameter [58], defined as the percentage of test spectra correctly classified, was used to evaluate the models performance.

5.4. RESULTS AND DISCUSSION

LIBS analysis on all these targets revealed the advantages and disadvantages offered by each method, based on what a grading was done as shown in Fig. 5.5., showing that direct analysis of *Candida* samples offered the best mounting method for the spectra acquisition.

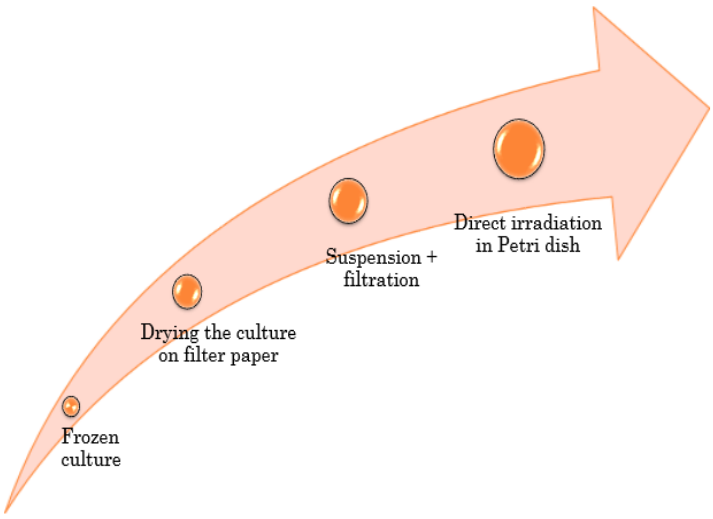


Fig. 5.5. Comparison of sample preparation methods

The frozen culture was firstly difficult to manage during the measurements due to rapid melting because of temperature changes and secondly the signal intensity was low. The samples deposited and dried on filter paper provided good signal to noise ratio but

the difference in lens to sample distance with each shot varied significantly because of non-uniform deposition and surface, giving greater variations in the signal intensity. This directed towards the preparation of sample by making suspension and then filtration to obtain a smooth and uniform surface for an even ablation. In the beginning, the deposited layer was uniform but along the course of experiment, the bacterial layer started drying and detaching from the filter paper resulting in folding and formation of cracks. This again led to the variations in lens to sample distance and therefore more scatter in the data. The same sample mounting protocol has been discussed in a recent study by Malenfant et. al. [46] where similar problems were encountered with respect to sample surface and signals scattering. Finding and standardizing a particular sample mounting protocol for microorganisms is still an area that needs much attention. The last method i.e. direct laser irradiation on the samples on one hand provided good signals intensity while on the other hand was the fastest method of analysis, which did not include any sample treatment steps and hence was selected as an optimal method.

The potential of LIBS for the analysis and discrimination of distinct *Candida* strains has been addressed. LIBS measurements were taken by directly irradiating the sample grown in the Petri dish at room conditions. Fig. 5.4. shows an example of an irradiated *C. albicans* culture and the craters formed by the laser shots. The typical normalized LIBS spectra of *Candida guilliermondii* is shown in Fig. 5.6 where mainly ionic and molecular bands along with few neutral atomic emission lines can be seen. For the assignment of the atomic/ionic lines the information tabulated in NIST Atomic Spectral Database was used [256]. The observed molecular emission bands are identified using the spectroscopic information available in References [146,147]. The assignation of the emissions showed that the spectrum mainly consisted of emissions from C, N, O and H reflecting the organic nature of fungal cells. Interestingly emission lines from

inorganic elements were not observed. Emissions in all the spectra from analyzed strains corresponded to these same elements (Fig. 5.5b). A visual comparison of the emission lines showed high similarity among the distinct LIBS spectral fingerprints. This can be attributed to a very similar chemical composition of these fungal cells.

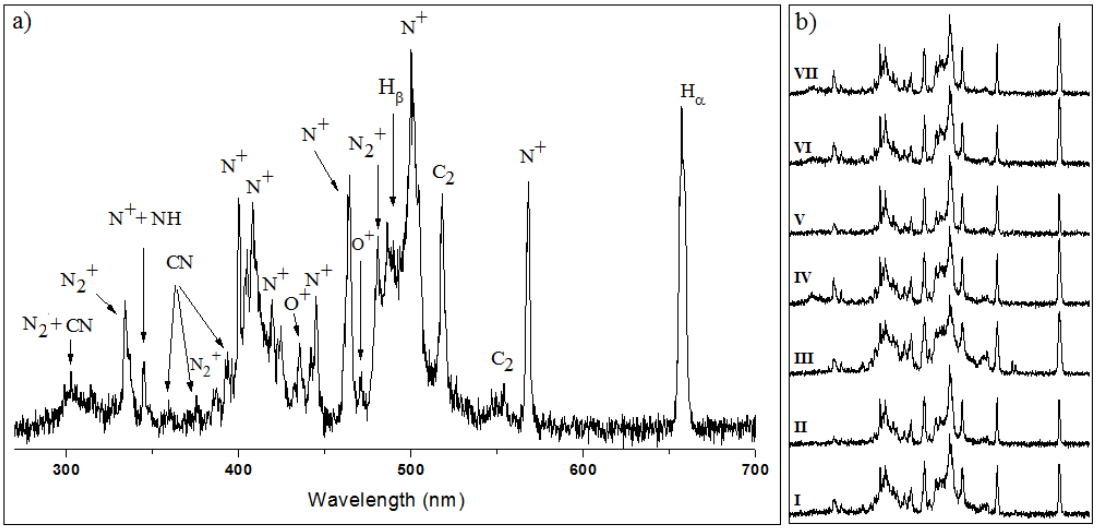


Fig. 5.6. a) Normalized LIBS spectrum of *Candida guilliermondii* with the assignation of emission lines. b) Normalized spectra of (I) *C. parasilopsis (ATCC CP22019,)* (II) *C. dubliniensis (CBS 7987),* (III) *C. krusei (ATCC 6258),* (IV) *C. guilliermondii (CNM-CL9533),* (V) *C. glabrata (CBS138),* (VI) *C. tropicalis (CNM-CL9537),* (VII) *C. albicans (CNM-CL9534)* within the spectral range of 300 – 700 nm

In order to understand the absence of signals from inorganic elements in LIBS spectra and to obtain more information about the elemental composition of Candida cultures SEM-EDS analyses were also performed. Fig. 5.7 illustrates the morphological aspect of a Candida culture as observed by SEM. EDS analyses of these strains provided the elemental composition in terms of weight percent of each element. As observed in Fig. 5.8., the major contribution was from elements such as C, N and O whereas Na, P, K and S were present in very small concentration.

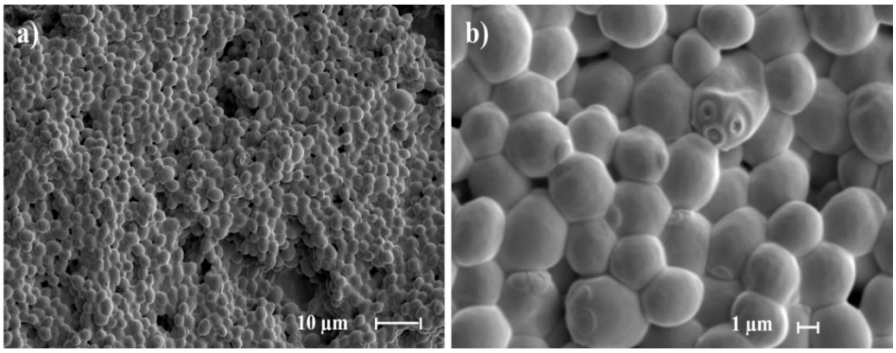


Fig. 5.7. a) Wide and b) magnified view of Candida albicans (CNM-CL9535) by SEM

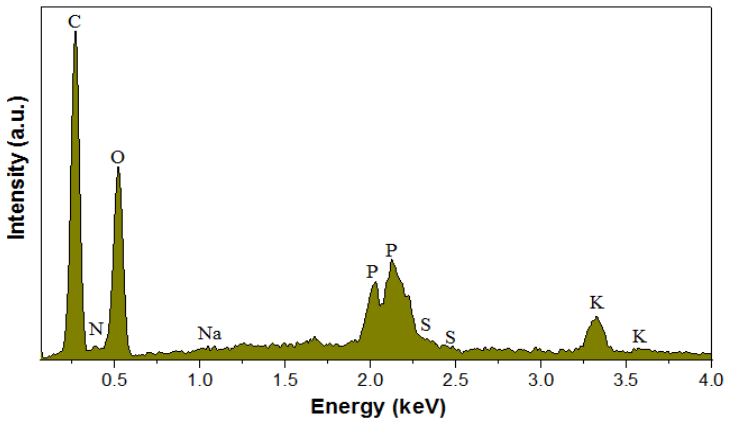


Fig. 5.8. Energy – Dispersive X-Ray Spectrum of Candida albicans (CNM-CL9535)

Table 5.2 illustrates the highly similar elemental profile of these *Candida* strains as exemplified with *C. dubliniensis, C. tropicalis* and *C. albicans.* These results were consistent with the LIBS spectra in which emissions from the inorganic elements were not observed. As expected, hydrogen was impossible to detect in the EDS spectrum. However, emission lines from H (H_α, H_β) were observed in LIBS spectra (Fig. 5.6) which are characteristic lines in LIBS analysis of organic samples.

Table 5.2. Element weight % obtained by Energy–Dispersive Spectroscopy of *Candida* strains

Element	Weight%		
	C. dubliniensis (CBS-7987)	***C. tropicalis*** (CNM-CL9537)	***C. albicans*** (CNM-CL9535)
C	46 ± 1	48 ± 1	48 ± 1
N	12 ± 1	11 ± 1	9 ± 2
O	40 ± 1	38 ± 1	39 ± 1
Na	0.13 ± 0.05	0.21 ± 0.07	*
P	0.92 ± 0.08	0.96 ± 0.08	1.49 ± 0.13
S	0.21 ± 0.06	0.34 ± 0.06	0.47 ± 0.09
K	1.02 ± 0.05	1.02 ± 0.06	1.29 ± 0.09

**The weight % for Na was below the threshold.*

LIBS studies focusing on the organic compounds suggest that the diatomic carbon emissions (around 516 nm) called Swan bands (C_2:$d^3\Pi_g - a^3\Pi_u$) are intrinsic to the sample whereas other emissions such as CN ($B^2\sum^+ - X^2\sum^+$), N, and O arise both from the sample as well as its combination with ambient air. The information extracted from C, N, H, O, CN and C/H, C/O and C/N intensity ratios proved highly useful for the characterization of compounds or samples of organic origin [55,257-260]. Thus, the cellular chemical structure of *Candida*, being predominantly organic, presented the spectral fingerprint with emissions from C, N, H, O and CN [54].

A recent study of our group demonstrated that NN is the best model to analyze spectral data from LIBS experiments [58]. Therefore, the NN model has been evaluated to analyze the LIBS data from the studied *Candida* samples. The NN was selected because it can significantly improve the identification capability without considerably increasing the difficulty of implementation [41]. This methodology improves the recognition capacity because the NN model is capable of identifying extremely similar samples that

have few physical and spectral differences. The selection of a full set of variables (intensity values of emission lines at each wavelength) constituting the spectrum was considered an important parameter in the development of the NN model for the classification process.

In order to achieve the classification of the *Candida* strains, a NN model was estimated using training libraries of all of the 21 *Candida* strains and then it was validated by test libraries. Each spectrum from the test library (unknown for the model) was evaluated using NN model, providing the prediction about the sample identity. The results from the NN model for the discrimination between strains are given in table 5.3.

Table 5.3. NN classification results for *Candida* samples

Candida strain	**Classification results**		
	Identification test		**Robustness test**
	Spectral	**Misidentified**	**Correct as**
CpA	100	0	✓
CpB	100	0	✓
CpC	100	0	✓
CdA	100	0	✓
CdB	100	0	✓
CdC	100	0	✓
CkA	100	0	✓
CkB	100	0	✓
CkC	100	0	✓
CglA	100	0	✓
CglB	100	0	✓
CglC	100	0	✓
CgA	100	0	✓
CgB	100	0	✓
CgC	100	0	✓
CtA	100	0	✓
CtB	100	0	✓
CtC	100	0	✓
CaA	100	0	✓
CaB	100	0	✓
CaC	100	0	✓

All of the spectra for *Candida* strains were assigned correctly to their species, achieving a spectral correlation of 100 % in all cases. The training process for all samples used as references required ≈ 1 min of execution time on computer whereas the test was carried out in approx. 2 seconds. Thus, the results also showed the network's capacity to deal simultaneously with more than one fingerprint (samples), without significantly increasing the computing time. Moreover, none of the test spectra of *Candida* strains were misidentified i.e. assigned to any other class. All of the strains included in the study were successfully classified according to their species and strain without producing any false positive or false negative.

On the other hand, the robustness of NN model was determined by assessing the ability of NN to estimate the correct result when an unknown sample was introduced into the network model [41,58]. This was evaluated as the efficiency of the NN model in recognizing a sample not included in the training step as "unknown" and not assigning it as one of the *Candida* strains used in the training. For this purpose, one *Candida* strain was removed from the training set and the NN model's prediction was checked. This process was continued until every strain was once removed and tested. The results of this analysis show that NN models were capable of perfectly discriminating all strains. None of the strains was incorrectly assigned to another strain. Therefore, these results, without any false positive or false negative, indicate that LIBS-NN is a highly robust method for strain discrimination.

In spite of the noise observed in the spectra, mainly due to direct analysis on the *Candida* culture with high humidity content, the mathematical treatment based on the complete spectral fingerprint constitutes the basis to achieve the discrimination between the samples. The NN is able to compute internal parameters (weights and bias) in the

training process for classifying a given set of input variables as belonging to a particular strain, with a high tolerance for noise and the presence of outliers [42].

5.5. CONCLUSION

LIBS spectra obtained for the *Candida* strains mainly showed emissions from elements that constitute organic compounds highlighting the organic nature of yeast cells. This evidence was further supported by the SEM–EDS analysis. The CN, C_2 band system, H, N, O and other molecular bands in LIBS spectra offered sufficient information for the discrimination of twenty-one *Candida* strains. In spite of high similarity between LIBS spectra, the mathematical analysis by NN algorithms that have significant identification capability and a relatively simple implementation, allowed achieving discrimination between *Candida* strains based on the single shot LIBS spectral fingerprint. This is the first LIBS-based study focused on the analysis of fungal samples and particularly of *Candida*, which present an outstanding clinical significance. The fact that NN models did not misidentify any strain makes LIBS-NN methodology very promising for classification purposes. Without the need of complex sample manipulation, LIBS-NN allowed fast spectral acquisition providing valuable compositional information, without the need of complex sample manipulation, and instant mathematical analysis of the data to predict the identity of the test sample. The methodology shows a great potential for rapid identification and discrimination between highly similar *Candida* strains.

Acknowledgements

I pay special thanks to my co-authors for their great help throughout this project. Prof. Maria Molina and Prof. Humberto Martin for sharing their expertise on the microbiological side and Prof. Joaquin Camacho for his help in SEM-EDS analysis. Great

thanks to Jesus Tornero Lopez from Hospital Los Madroños (Brunete), Complutense University of Madrid (UCM), DGICYT (Project MICINN:CTQ2013-43086-P) and Madrid, Spain for this research project. I also thank Ana Alastruey from Instituto de Salud Carlos III and Elvira Marín from Departamento de Microbiología II (UCM) for providing the samples of *Candida* strains.

Chapter 6

Conclusions and General Discussion

The main goal of this thesis was to explore the potential of Laser Induced Breakdown Spectroscopy (LIBS) for the identification and discrimination of microorganisms and present a possible application in biomedical field. Another significant aspect focused was the study of appropriate chemometric method for data analysis and development of classification models. Essential and related background knowledge about the techniques used along with the experimental details to achieve this goal have been explained in the thesis. Based on the objectives formulated and results obtained in each study, following conclusions can be drawn:

- A methodology based on the combination of Laser Induced Breakdown Spectroscopy with advanced chemometric methods i.e. Neural Networks (NN) has been developed for the identification and discrimination of clinically relevant bacterial and fungal strains. The methodology has been successfully validated in the identification of bacterial samples in a hierarchical manner, based on their species first and discriminating from other strains of the same species. Bacterial strains isolated from different biological fluids were correctly identified to their species. Furthermore, a study using LIBS technique solely focused on fungal samples, in particular, *Candida* species and their identification was for the first time performed in this thesis. Test performed on

Candida strains with very similar spectral characteristics provided successful classification results.

- Undoubtedly, the incorporation of computer-based signal processing techniques (chemometric algorithms) for the classification and identification of pathogenic microorganisms has added the ultra-rapid character required for the evaluation of biological threats and clinical diagnosis. On the other hand, it has improved its capacity of analysis and identification without false positives or false negatives turning it into a very robust technique.
- Given the accuracy of the results obtained, viability of LIBS spectral fingerprint, high speed of analysis on part of LIBS as well as NN algorithm for data analysis, this methodology demonstrated its potential for performing a reliable, automated, simple and robust analysis in clinical laboratories. From the medical viewpoint, this would allow an early diagnosis of the infectious diseases and a timely their treatment, which may reduce the recurrence of Hospital Acquired Infections.
- Spectral differences are difficult to detect by visual analysis of highly similar LIBS spectra. Herein, multivariate techniques are known to be efficient methods for sorting and classifying data, however, selection of an appropriate statistical method for data analysis is not a straightforward task as it depends on the application and type of dataset. Therefore, a comparison was carried out between five supervised chemometric methods (SIMCA, PLS-DA, LDA, SVM and NN) based on parameters such as sensitivity, generalization ability and robustness, demonstrated that NN outperformed all methods in all evaluation parameters. This can be attributed to the non-linear nature of algorithm used for the model generation and its ability to model complex nonlinear input-target

relationships. Therefore, neural networks models were taken forward to achieve the objectives formulated in this thesis, specifically designing algorithms for the particular application.

- Although NN may be more difficult to implement than other standard chemometric methods due to the diversity of functions and architectures, the results obtained in this study demonstrate that LIBS combined with NN offers significant advantages as a screening tool, given its speed, high throughput, minimal destructive, micro-analysis and ease of use. This can be attributed to the supervised, non-linear approach of NN models, their inherent fault tolerance, noise immunity and statistical sensitivity. These characteristics enabled obtaining a clear identification and discrimination of microbes using single shot LIBS measurements and tolerating minor changes in experimental conditions, such as the intensity of the LIBS single-shot regimen and continuum background, when samples were measured on different days. However, the importance of quality of recorded data cannot be ignored. Data analysis of different spectral sets recorded along the course of the experimental work showed that only well-measured, structured and reproducible data could lead to a successful application of the chemometric method.
- A key point while using a chemometric method is the selection of the input variables for data treatment. In this thesis, the complete set of variables (intensities at each wavelength) that make up the spectrum of the sample along entire spectral range presented the features to the NN models that were important in the comparison process to achieve a correct classification. Thus, whole LIBS spectral fingerprint was considered important for the identification of each strain analyzed.

- LIBS analysis requires minimal or no sample preparation, nevertheless, isolation of microorganisms and cultivating on a nutrient medium to obtain sufficient biomass for LIBS spectra acquisition is a required step, as in any other techniques used. It is necessary to emphasize that the time spent to generate the required biomass is much lesser comparatively as no culture enrichment steps are required, thus, this LIBS based methodology could provide the results in a reasonable period. However, studies on optimizing the minimum amount and time of culture are needed to be carried out in future.
- The study using three different culture media to grow bacteria demonstrated that the type of medium used affects LIBS based microbial analysis. This might happen due to possible elemental alteration of the membranes and cellular content induced by different nutrient composition of the culture media. However, these results do not represent any inherent limitations of LIBS technique and are expected when samples´ elemental composition is altered. In experiments where measurements were made using same culture media and the only variation was the type of bacteria, a correct identification and discrimination between the samples was made. This implies the use of a standard medium to grow microorganisms in order to avoid any changes in the characteristic LIBS spectral fingerprint that might affect the classification capacity.
- The experimental configuration including different spectrometers and detectors along with other components of the system and taking measurements at room experimental conditions, as described in Chapter 3 demonstrate the simplicity of the set-up and analysis. Despite the contribution of the air in LIBS spectra, a

correct identification was achieved, which allowed keeping the analysis simple, fast and cost-effective.

- LIBS-NN methodology allowed classifying and predicting bacterial samples at the level of genus, strain and even at the level of differences of a single gene between bacterial strains of the same species with a high degree of precision and accuracy. It has been shown that variations acquired by bacteria in their cell wall or genetic make-up due to antibiotic resistance generated sufficient or significant changes in their atomic composition that could be detected by the LIBS technique allowing their discrimination. This was a very fine test and successful result as such minute differences are hard to detect with other conventional techniques.
- A first fully focused study on the analysis of Candida species of fungi was also performed. The initial results for twenty-one different strains of Candida were satisfactory as successful discrimination between them was achieved. The spectral profile of bacterial and fungal samples was found to be very distinct from each other. The emissions in bacterial spectra presented molecular bands as well as atomic and ionic lines from inorganic elements whereas LIBS spectra obtained for the *Candida* strains mainly showed band emissions such as CN, C_2 system and lines from H, N, O along with other elements that constitute organic compounds. The inorganic elements were not observed because of lower concentration. SEM-EDS analysis also supported the compositional analysis as reflected by LIBS spectra. Nevertheless, these spectral emissions offered sufficient information for the discrimination of highly similar *Candida* strains applying the mathematical analysis by NN algorithms, providing a robust analysis. An important component of this study was evaluation of mounting

protocols of the sample for LIBS analysis. All methods gave satisfactory results but direct analysis of fungal culture deposited on the petri dish was selected due to ease of analysis and possibility of a quick analysis. The thesis includes initial results obtained showing the potential of LIBS-NN for the analysis of *Candida* strains, obtaining and spectrally characterizing the fingerprints of these samples.

- It is important to note that the fraction of total spectral intensity observed in the LIBS spectrum is not a measure of the total elemental concentration in the bacterial cell. The relative emission intensity of a given elemental emission line depends not only on the concentration of that element/analyte in the cell, but also on the relative atomic or ionic transition probabilities. Therefore, the emission lines observed in the LIBS spectrum give qualitative information of the elements.

Advanced classification algorithms "NN" have been used to analyze the LIBS spectrum, based on the spectral characteristics of the specimen, which allowed assigning an unknown sample to a species or strain according to a predefined reference library. Due to the possibility of automation and converting into a portable system, the results of the diagnosis could be provided in real time, not requiring much intervention by the user, although additional diagnostic information may be included to aid in the interpretation of the spectrum and classification. Therefore, it can be considered a quick, simple and cost-effective alternative to the slower and more expensive biological methods.

The most important objective in this thesis was to develop and program LIBS-NN based methodology, efficient enough to produce results without false positives or false negatives, which has been sought fortunately. However, some challenges are still to be dealt such as introducing improvements in the experimental system, a better detection system and sensitive light collection system, the number of variables to be treated, the

ability to analyze diverse range of microorganisms, sample contamination, low cell count and mixed samples.

In spite of being fast, rapid and accurate, LIBS is still a young technique in biomedical field particularly and still in the phase of research. This is evident by use of different experimental configurations and chemometric methods to get the results. Different types of lasers have been used, although the fundamental frequency (1064 nm) of nanosecond Nd:YAG lasers are the most common. As for the experimental conditions, the energy may vary according to the experiment and the research group. Different geometries are used for collecting plasma emission, and different atmospheres (Air, He, Ar). There is no standard protocol of analysis nor for the preparation and mounting of microbial samples recommended by the scientific community, which makes it difficult to make a meaningful comparison between the analyses made in different laboratories. As such, there is still no agreement on how a standard "typical" LIBS spectrum should look like for a given organism and therefore there is no "reference LIBS spectral fingerprint" to be used as a standard by laboratories around the world.

To date, no effort has been made to standardize the method for the identification of microorganisms. This situation is understandable since the existing results are relatively recent and that LIBS laboratories working around the world with microbiological samples are not coordinating their efforts. This will be impulsed by investing the funds and scientific support in this area of research. If standard measurement protocols were adapted, several research groups could come together to build a large and good quality database but even then the nature of the chemometric algorithms used could greatly affect the sensitivity and specificity of the technique.

Although currently commercial instruments specially designed for real-time LIBS analyzes of biomedical samples are not yet a reality, it is evident from the experiments

summarized here that hardware and software are present to make such an instrument possible. During the development of this thesis, efforts were made in the use of LIBS technology and designing a prototype for specific clinical applications by working closely with a private hospital in the Community of Madrid. In addition, the microbiological experiments have been carried out to demonstrate that a LIBS-based analysis is not only sensitive and specific but also robust against many sources of error and uncertainty common to other methods. Therefore, it is required now is to increase the investment of both state and private resources to consolidate protocols for the standardized preparation of microbiological samples and the procedures for analysis and validation using a large number of clinical samples as well as chemometric classification algorithms.

REFERENCES

[1] D.A. Cremers, L.J. Radziemski, Handbook of laser-induced breakdown spectroscopy, 2nd Ed, Wiley-Blackwell, Oxford - UK, 2013.

[2] D.A. Cremers, R.C. Chinni, Laser-Induced Breakdown Spectroscopy—Capabilities and Limitations, Applied Spectroscopy Reviews. 44 (2009) 457–506.

[3] A.W. Miziolek, V. Palleschi, I. Schechter, Laser Induced Breakdown Spectroscopy (LIBS), 1st ed., Cambridge University Press, New York - USA, 2008.

[4] S. Pahlow, S. Meisel, D. Cialla-May, K. Weber, P. Rösch, J. Popp, Isolation and identification of bacteria by means of Raman spectroscopy, Advanced Drug Delivery Reviews. 89 (2015) 105–120.

[5] EPINE: Estudio de Prevalencia de las Infecciones Nosocomiales en España., (n.d.). http://hws.vhebron.net/epine/ (accessed September 4, 2016).

[6] J. Vaqué, J. Rosselló, J.L. Arribas, Prevalence of nosocomial infections in Spain: EPINE study 1990–1997, Journal of Hospital Infection. 43 (1999) S105–S111.

[7] S. Binder, A.M. Levitt, J.J. Sacks, J.M. Hughes, Emerging Infectious Diseases: Public Health Issues for the 21st Century, Science. 284 (1999) 1311–1313.

[8] D.M. Morens, G.K. Folkers, A.S. Fauci, The challenge of emerging and re-emerging infectious diseases, Nature. 430 (2004) 242–249.

[9] K.E. Jones, N.G. Patel, M.A. Levy, A. Storeygard, D. Balk, J.L. Gittleman, P. Daszak, Global trends in emerging infectious diseases, Nature. 451 (2008) 990–993.

[10] M.J. Desai, D.W. Armstrong, Separation, Identification, and Characterization of Microorganisms by Capillary Electrophoresis, Microbiol Mol Biol Rev. 67 (2003) 38–51.

[11] C.A. Petti, C.R. Polage, P. Schreckenberger, The Role of 16S rRNA Gene Sequencing in Identification of Microorganisms Misidentified by Conventional Methods, J. Clin. Microbiol. 43 (2005) 6123–6125.

[12] K. Böhme, I. C., J. Barros-Velzquez, J. M., B. Caas, P. Calo-Mat, Species Identification of Food Spoilage and Pathogenic Bacteria by MALDI-TOF Mass Fingerprinting, in: K. Kapiris (Ed.), Food Quality, InTech, 2012. http://www.intechopen.com/books/food-quality/species-identification-of-food-spoilage-and-pathogenic-bacteria-by-maldi-tof-mass-fingerprinting (accessed March 23, 2015).

[13] D. Emerson, L. Agulto, H. Liu, L. Liu, Identifying and Characterizing Bacteria in an Era of Genomics and Proteomics, BioScience. 58 (2008) 925–936.

[14] B.G. Kornreich, M. Craven, S.P. McDonough, D.V. Nydam, V. Scorza, S. Assarasakorn, M. Lappin, K.W. Simpson, Fluorescence in-situ hybridization for the identification of

bacterial species in archival heart valve sections of canine bacterial endocarditis, J. Comp. Pathol. 146 (2012) 298–307.

[15] S. Mignard, J.P. Flandrois, 16S rRNA sequencing in routine bacterial identification: A 30-month experiment, Journal of Microbiological Methods. 67 (2006) 574–581.

[16] P.C.Y. Woo, S.K.P. Lau, J.L.L. Teng, H. Tse, K.-Y. Yuen, Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories, Clinical Microbiology and Infection. 14 (2008) 908–934.

[17] M. Natesan, R.G. Ulrich, Protein Microarrays and Biomarkers of Infectious Disease, Int J Mol Sci. 11 (2010) 5165–5183.

[18] A. Álvarez-Barrientos, J. Arroyo, R. Cantón, C. Nombela, M. Sánchez-Pérez, Applications of Flow Cytometry to Clinical Microbiology, Clin Microbiol Rev. 13 (2000) 167–195.

[19] J. Verma, S. Saxena, S. Babu, ELISA-Based Identification and Detection of Microbes, in: D.K. Arora, S. Das, M. Sukumar (Eds.), Analyzing Microbes, Springer Berlin Heidelberg, 2013: pp. 169–186.

[20] N.T. (Kasetsart Univ, B. (Thailand) D. of P. Pathology), Using of gel electrophoresis for bacterial identification, 1987.

[21] K. Böhme, I.C. Fernández-No, J. Barros-Velázquez, J.M. Gallardo, B. Cañas, P. Calo-Mata, Comparative analysis of protein extraction methods for the identification of seafood-borne pathogenic and spoilage bacteria by MALDI-TOF mass spectrometry, Analytical Methods. 2 (2010) 1941.

[22] P. Rösch, M. Harz, M. Schmitt, K.-D. Peschke, O. Ronneberger, H. Burkhardt, H.-W. Motzkus, M. Lankers, S. Hofer, H. Thiele, J. Popp, Chemotaxonomic Identification of Single Bacteria by Micro-Raman Spectroscopy: Application to Clean-Room-Relevant Biological Contaminations, Appl. Environ. Microbiol. 71 (2005) 1626–1637.

[23] D. Helm, H. Labischinski, G. Schallehn, D. Naumann, Classification and identification of bacteria by Fourier-transform infrared spectroscopy, J Gen Microbiol. 137 (1991) 69–79.

[24] D. Naumann, Infrared Spectroscopy in Microbiology, in: Encyclopedia of Analytical Chemistry, John Wiley & Sons, Ltd, 2006. http://onlinelibrary.wiley.com/doi/10.1002/9780470027318.a0117/abstract (accessed January 20, 2017).

[25] C. Xie, J. Mace, M.A. Dinno, Y.Q. Li, W. Tang, R.J. Newton, P.J. Gemperline, Identification of Single Bacterial Cells in Aqueous Solution Using Confocal Laser Tweezers Raman Spectroscopy, Anal. Chem. 77 (2005) 4390–4397.

[26] R.M. Jarvis, A. Brooker, R. Goodacre, Surface-Enhanced Raman Spectroscopy for Bacterial Discrimination Utilizing a Scanning Electron Microscope with a Raman Spectroscopy Interface, Anal. Chem. 76 (2004) 5198–5202.

[27] S. Stöckel, S. Meisel, M. Elschner, P. Rösch, J. Popp, Identification of Bacillus anthracis via Raman Spectroscopy and Chemometric Approaches, Anal. Chem. 84 (2012) 9873–9880.

[28] A. Walter, A. März, W. Schumacher, P. Rösch, J. Popp, Towards a fast, high specific and reliable discrimination of bacteria on strain level by means of SERS in a microfluidic device, Lab Chip. 11 (2011) 1013–1021.

[29] X. Lu, H.M. Al-Qadiri, M. Lin, B.A. Rasco, Application of Mid-infrared and Raman Spectroscopy to the Study of Bacteria, Food Bioprocess Technol. 4 (2011) 919–935.

[30] K. Maquelin, C. Kirschner, L.-P. Choo-Smith, N. van den Braak, H.P. Endtz, D. Naumann, G.J. Puppels, Identification of medically relevant microorganisms by vibrational spectroscopy, Journal of Microbiological Methods. 51 (2002) 255–271.

[31] A.Y. Peleg, D.C. Hooper, Hospital-Acquired Infections Due to Gram-Negative Bacteria, N Engl J Med. 362 (2010) 1804–1813.

[32] M. Fakruddin, K.S. Bin Mannan, R.M. Mazumdar, A. Chowdhury, N. Hossain, Identification and characterization of microorganisms: DNA-fingerprinting methods, Songklanakarin Journal of Science & Technology. 35 (2013) 397–404.

[33] A. Pechorsky, Y. Nitzan, T. Lazarovitch, Identification of pathogenic bacteria in blood cultures: Comparison between conventional and PCR methods, Journal of Microbiological Methods. 78 (2009) 325–330.

[34] R. Goodacre, éadaoin M. Timmins, R. Burton, N. Kaderbhai, A.M. Woodward, D.B. Kell, P.J. Rooney, Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks, Microbiology. 144 (1998) 1157–1170.

[35] T. Schmid, A. Sebesta, J. Stadler, L. Opilik, R.M. Balabin, R. Zenobi, Tip-enhanced Raman spectroscopy and related techniques in studies of biological materials, in: 2010: pp. 758603-758603–13.

[36] S. Morel, N. Leone, P. Adam, J. Amouroux, Detection of Bacteria by Time-Resolved Laser-Induced Breakdown Spectroscopy, Appl. Opt. 42 (2003) 6184–6191.

[37] S.J. Rehse, Laser-Based Identification of Pathogenic Bacteria, Phys. Teach. 47 (2009) 152–156.

[38] J.D. Hybl, G.A. Lithgow, S.G. Buckley, Laser-Induced Breakdown Spectroscopy Detection and Classification of Biological Aerosols, Appl. Spectrosc. 57 (2003) 1207–1215.

[39] A.C. Samuels, F.C. DeLucia, K.L. McNesby, A.W. Miziolek, Laser-induced breakdown spectroscopy of bacterial spores, molds, pollens, and protein: initial studies of discrimination potential, Applied Optics. 42 (2003) 6205.

[40] A.W. Miziolek, V. Palleschi, I. Schechter, Laser-induced breakdown spectroscopy (LIBS), Cambridge, UK, Cambridge University Press, 2006.

[41] D. Marcos-Martinez, J.A. Ayala, R.C. Izquierdo-Hornillos, F.J.M. de Villena, J.O. Caceres, Identification and discrimination of bacterial strains by laser induced breakdown spectroscopy and neural networks, Talanta. 84 (2011) 730–737.

[42] J.O. Caceres, S. Moncayo, J.D. Rosales, F.J.M. de Villena, F.C. Alvira, G.M. Bilmes, Application of Laser-Induced Breakdown Spectroscopy (LIBS) and Neural Networks to Olive Oils Analysis, Applied Spectroscopy. 67 (2013) 1064–1072.

[43] R.A. Multari, D.A. Cremers, M.L. Bostian, Use of laser-induced breakdown spectroscopy for the differentiation of pathogens and viruses on substrates, Appl. Opt. 51 (2012) B57–B64.

[44] Q. Mohaidat, S. Palchaudhuri, S.J. Rehse, The Effect of Bacterial Environmental and Metabolic Stresses on a Laser-Induced Breakdown Spectroscopy (LIBS) Based Identification of Escherichia coli and Streptococcus viridans, Appl. Spectrosc. 65 (2011) 386–392.

[45] S.J. Rehse, J. Diedrich, S. Palchaudhuri, Identification and discrimination of Pseudomonas aeruginosa bacteria grown in blood and bile by laser-induced breakdown spectroscopy, Spectrochim. Acta B. 62 (2007) 1169–1176.

[46] D.J. Malenfant, D.J. Gillies, S.J. Rehse, Bacterial Suspensions Deposited on Microbiological Filter Material for Rapid Laser-Induced Breakdown Spectroscopy Identification, Appl. Spectrosc., AS. 70 (2016) 485–493.

[47] S.J. Rehse, Q.I. Mohaidat, S. Palchaudhuri, Towards the clinical application of laser-induced breakdown spectroscopy for rapid pathogen diagnosis: the effect of mixed cultures and sample dilution on bacterial identification, Appl. Opt. 49 (2010) C27–C35.

[48] N. Wellinghausen, A.-J. Kochem, C. Disqué, H. Mühl, S. Gebert, J. Winter, J. Matten, S.G. Sakka, Diagnosis of Bacteremia in Whole-Blood Samples by Use of a Commercial Universal 16S rRNA Gene-Based PCR and Sequence Analysis, J. Clin. Microbiol. 47 (2009) 2759–2765.

[49] I.-F. Cheng, H.-C. Chang, T.-Y. Chen, C. Hu, F.-L. Yang, Rapid (<5 min) Identification of Pathogen in Human Blood by Electrokinetic Concentration and Surface-Enhanced Raman Spectroscopy, Scientific Reports. 3 (2013) 2365.

[50] R. Noll, Laser-Induced Breakdown Spectroscopy, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. http://link.springer.com/10.1007/978-3-642-20668-9 (accessed July 13, 2015).

[51] F. Anabitarte, A. Cobo, J.M. Lopez-Higuera, F. Anabitarte, A. Cobo, J.M. Lopez-Higuera, Laser-Induced Breakdown Spectroscopy: Fundamentals, Applications, and Challenges, Laser-Induced Breakdown Spectroscopy: Fundamentals, Applications, and Challenges, International Scholarly Research Notices, International Scholarly Research Notices. 2012, 2012 (2012) e285240.

[52] J.P. Singh, S.N. (Ed) Thakur, Laser-Induced Breakdown Spectroscopy, Elsevier, Amsterdam, Netherland, 2007.

[53] S.J. Rehse, Q.I. Mohaidat, The effect of sequential dual-gas testing on laser-induced breakdown spectroscopy-based discrimination: Application to brass samples and bacterial strains, Spectrochimica Acta Part B: Atomic Spectroscopy. 64 (2009) 1020–1027.

[54] C. López-Moreno, S. Palanco, J.J. Laserna, F.D. Jr, A.W. Miziolek, J. Rose, R.A. Walters, A.I. Whitehouse, Test of a stand-off laser-induced breakdown spectroscopy sensor for the detection of explosive residues on solid surfaces, J. Anal. At. Spectrom. 21 (2006) 55–60.

[55] J. Anzano, R.-J. Lasheras, B. Bonilla, J. Casas, Classification of polymers by determining of C1:C2:CN:H:N:O ratios by laser-induced plasma spectroscopy (LIPS), Polym. Test. 27 (2008) 705–710.

[56] S. Moncayo, S. Manzoor, T. Ugidos, F. Navarro-Villoslada, J.O. Caceres, Discrimination of human bodies from bones and teeth remains by Laser Induced Breakdown Spectroscopy and Neural Networks, Spectrochimica Acta Part B: Atomic Spectroscopy. 101 (2014) 21–25.

[57] S. Manzoor, S. Moncayo, F. Navarro-Villoslada, J.A. Ayala, R. Izquierdo-Hornillos, F.J.M. de Villena, J.O. Caceres, Rapid identification and discrimination of bacterial strains by laser induced breakdown spectroscopy and neural networks, Talanta. 121 (2014) 65–70.

[58] S. Moncayo, S. Manzoor, F. Navarro-Villoslada, J.O. Caceres, Evaluation of supervised chemometric methods for sample classification by Laser Induced Breakdown Spectroscopy, Chemometrics and Intelligent Laboratory Systems. 146 (2015) 354–364.

[59] P. Inakollu, T. Philip, A.K. Rai, F.-Y. Yueh, J.P. Singh, A comparative study of laser induced breakdown spectroscopy analysis for element concentrations in aluminum alloy using artificial neural networks and calibration methods, Spectrochim. Acta B. 64 (2009) 99–104.

[60] C. Pasquini, J. Cortez, L.M.C. Silva, F.B. Gonzaga, Laser Induced Breakdown Spectroscopy, Journal of the Brazilian Chemical Society. 18 (2007) 463–512.

[61] A. Sarzyski, W. Skrzeczanowski, J. Marczak, Physical phenomena disturbing LIBS analysis, in: Proceedings of SPIE, Munich, Germany, 2007: p. 66180V–66180V–10.

[62] E. Tognoni, V. Palleschi, M. Corsi, G. Cristoforetti, Quantitative micro-analysis by laser-induced breakdown spectroscopy: a review of the experimental approaches, Spectrochimica Acta Part B: Atomic Spectroscopy. 57 (2002) 1115–1130.

[63] R. Gaudiuso, M. Dell'Aglio, O.D. Pascale, G.S. Senesi, A.D. Giacomo, Laser Induced Breakdown Spectroscopy for Elemental Analysis in Environmental, Cultural Heritage and Space Applications: A Review of Methods and Results, Sensors. 10 (2010) 7434–7468.

[64] O. Samek, D.C.S. Beddows, H.H. Telle, J. Kaiser, M. Liska, J.O. Cáceres, A. Gonzáles Ureña, Quantitative laser-induced breakdown spectroscopy analysis of calcified tissue samples, Spectrochim. Acta, Part B. 56 (2001) 865–875.

[65] H.R. Griem, Principles of Plasma Spectroscopy, Cambridge University Press, 2005.

[66] J.J. Camacho, J.M.L. Poyato, L. Díaz, M. Santos, Laser-induced breakdown spectroscopy of trisilane using infrared CO2 laser pulses, Journal of Applied Physics. 102 (2007) 103302.

[67] J.J. Camacho, L. Díaz, J.P. Cid, J.M.L. Poyato, Time-resolved spectroscopic diagnostic of the CO2 plasma induced by a high-power CO2 pulsed laser, Spectrochimica Acta Part B: Atomic Spectroscopy. 66 (2011) 698–705.

[68] NIST USA, Atomic Spectroscopy Information, (n.d.). http://www.nist.gov/atomic-spectroscopy.cfm (accessed October 30, 2012).

[69] VALD, Vienna Atomic Line Database (VALD), (n.d.). http://vald.astro.uu.se/ (accessed September 25, 2016).

[70] M.P. Mateo, G. Nicolás, V. Piñón, J.C. Alvarez, A. Ramil, A. Yáñez, Versatile software for semiautomatic analysis and processing of laser-induced plasma spectra, Spectrochimica Acta Part B: Atomic Spectroscopy. 60 (2005) 1202–1210.

[71] R. Barbini, F. Colao, R. Fantoni, A. Palucci, S. Ribezzo, H.J.L. van der Steen, M. Angelone, Semi-quantitative time resolved LIBS measurements, Applied Physics B: Lasers and Optics. 65 (1997) 101–107.

[72] S. Palanco, A. Alises, J. Cuñat, J. Baena, J.J. Laserna, Development of a portable laser-induced plasma spectrometer with fully-automated operation and quantitative analysis capabilities, J. Anal. At. Spectrom. 18 (2003) 933–938.

[73] T.A. Labutin, S.M. Zaytsev, A.M. Popov, Automatic Identification of Emission Lines in Laser-Induced Plasma by Correlation of Model and Experimental Spectra, Anal. Chem. 85 (2013) 1985–1990.

[74] J. Frydenvang, K.M. Kinch, S. Husted, M.B. Madsen, An Optimized Calibration Procedure for Determining Elemental Ratios Using Laser-Induced Breakdown Spectroscopy, Anal. Chem. 85 (2013) 1492–1500.

[75] A. Ciucci, M. Corsi, V. Palleschi, S. Rastelli, A. Salvetti, E. Tognoni, New Procedure for Quantitative Elemental Analysis by Laser-Induced Plasma Spectroscopy, Appl. Spectrosc., AS. 53 (1999) 960–964.

[76] M. Otto, Chemometrics: Statistics and Computer Application in Analytical Chemistry, Wiley, 1999.

[77] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis, Anal. Chem. 70 (1998) 2796–2804.

[78] D.-S. Lee, B.-S. Noh, S.-Y. Bae, K. Kim, Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics, Analytica Chimica Acta. 358 (1998) 163–175.

[79] R. Kizil, J. Irudayaraj, K. Seetharaman, Characterization of Irradiated Starches by Using FT-Raman and FTIR Spectroscopy, J. Agric. Food Chem. 50 (2002) 3912–3918.

[80] E.C. López-Díez, G. Bianchi, R. Goodacre, Rapid Quantitative Assessment of the Adulteration of Virgin Olive Oils with Hazelnut Oils Using Raman Spectroscopy and Chemometrics, J. Agric. Food Chem. 51 (2003) 6145–6150.

[81] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, Pearson Education, 2005.

[82] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, Journal of Pharmaceutical and Biomedical Analysis. 44 (2007) 683–700.

[83] T. Salomonsen, H.M. Jensen, D. Stenbæk, S.B. Engelsen, Chemometric prediction of alginate monomer composition: A comparative spectroscopic study using IR, Raman, NIR and NMR, Carbohydrate Polymers. 72 (2008) 730–739.

[84] J.S. Torrecilla, M. Cámara, V. Fernández-Ruiz, G. Piera, J.O. Caceres, Solving the Spectroscopy Interference Effects of β-Carotene and Lycopene by Neural Networks, J. Agric. Food Chem. 56 (2008) 6261–6266.

[85] H. Stenlund, R. Madsen, A. Vivi, M. Calderisi, T. Lundstedt, M. Tassini, M. Carmellini, J. Trygg, Monitoring kidney-transplant patients using metabolomics and dynamic modeling, Chemometrics and Intelligent Laboratory Systems. 98 (2009) 45–50.

[86] S. Nakajima, H. Hoshina, M. Yamashita, C. Otani, N. Miyoshi, Terahertz imaging diagnostics of cancer tissues with a chemometrics technique, Applied Physics Letters. 90 (2007) 041102.

[87] J.T. Brindle, H. Antti, E. Holmes, G. Tranter, J.K. Nicholson, H.W.L. Bethell, S. Clarke, P.M. Schofield, E. McKilligin, D.E. Mosedale, D.J. Grainger, Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics, Nat Med. 8 (2002) 1439–1445.

[88] J. van der Greef, A.K. Smilde, Symbiosis of chemometrics and metabolomics: past, present, and future, J. Chemometrics. 19 (2005) 376–386.

[89] A. Sussulini, A. Prando, D.A. Maretto, R.J. Poppi, L. Tasic, C.E.M. Banzato, M.A.Z. Arruda, Metabolic Profiling of Human Blood Serum from Treated Patients with Bipolar Disorder Employing 1H NMR Spectroscopy and Chemometrics, Anal. Chem. 81 (2009) 9755–9763.

[90] R. Madsen, T. Lundstedt, J. Trygg, Chemometrics in metabolomics—A review in human disease diagnosis, Analytica Chimica Acta. 659 (2010) 23–33.

[91] J.A. Lopes, P.F. Costa, T.P. Alves, J.C. Menezes, Chemometrics in bioprocess engineering: process analytical technology (PAT) applications, Chemometrics and Intelligent Laboratory Systems. 74 (2004) 269–275.

[92] J.A. Cramer, Approaching the Chemometric Modeling of Realistically Diverse Biochemical Data, in: 40 Years of Chemometrics – From Bruce Kowalski to the Future, American Chemical Society, 2015: pp. 283–298. (accessed June 20, 2016).

[93] A.L. Pomerantsev, O.Y. Rodionova, Process analytical technology: a critical view of the chemometricians, J. Chemometrics. 26 (2012) 299–310.

[94] Massart, Handbook of Chemometrics and Qualimetrics, Volume Part A, 1st ed., Elsevier Science, Amsterdam, The Netherlands, 1997.

[95] S.D. Brown, Chemometrics, in: Encyclopedia of Analytical Chemistry, John Wiley & Sons, Ltd, Chichester, UK, 2006.

[96] P. Gemperline, Practical Guide To Chemometrics, 2nd ed., CRC Press, Boca Raton- FL, USA, 2006.

[97] R.G. Brereton, Applied Chemometrics for Scientists, John Wiley & Sons Ltd., Chichester, UK, 2007.

[98] S.D. Brown, R. Tauler, B. Walczak (Ed.), Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Elsevier, Amsterdam, Netherlands, 2009.

[99] B.K. Lavine, J. Workman, CHEMOMETRICS, Anal. Chem. (2012)..

[100] N. Labbé, I.M. Swamidoss, N. André, M.Z. Martin, T.M. Young, T.G. Rials, Extraction of information from laser-induced breakdown spectroscopy spectral data by multivariate analysis, Applied Optics. 47 (2008) G158.

[101] D. Ballabio, R. Todeschini, Multivariate Classification for Qualitative Analysis, in: D.-W. Sun (Ed.), Infrared Spectroscopy for Food Quality Analysis and Control, Academic Press, San Diego - USA, 2009: pp. 83–104.

[102] O. Maimon, L. Rokach, Data Mining and Knowledge Discovery Handbook, 2nd ed., Springer US, Boston, MA, 2010. http://link.springer.com/10.1007/978-0-387-09823-4 (accessed May 16, 2016).

[103] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, Anal. Methods. 5 (2013) 3790–3798.

[104] D. Ballabio, Chemometric characterisation of physical-chemical fingerprints of food products., PhD thesis, Universit`a degli Studi di Milano Facolt`a di Agraria Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche., 2005.

[105] C. Vanesa Di Anibal, Determination of banned sud an dyes in culinary spices through spectroscopic techniques and multivariate analysis., 2011.

[106] D. Chung, S. Keles, Sparse Partial Least Squares Classification for High Dimensional Data, Statistical Applications in Genetics and Molecular Biology. 9 (2010).

[107] P. Xanthopoulos, P.M. Pardalos, T.B. Trafalis, Linear Discriminant Analysis, in: Robust Data Mining, Springer New York, 2013: pp. 27–33.

[108] J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, Advances in Neural Information Processing Systems. 17 (2004) 1569–1576.

[109] S. Moncayo, Desarrollo y Aplicación de Métodos Quimiométricos Para el Estudio de Muestras Mediante Espectroscopia de Ablación Laser (LIBS)., Doctoral thesis, Complutense University of Madrid, 2016.

[110] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemometrics. 17 (2003) 166–173.

[111] R. De Maesschalck, A. Candolfi, D.L. Massart, S. Heuerding, Decision criteria for soft independent modelling of class analogy applied to near infrared data, Chemometrics and Intelligent Laboratory Systems. 47 (1999) 65–77.

[112] Massart, Handbook of Chemometrics and Qualimetrics, Volume Part B, Volume B, Elsevier Science, 1998.

[113] R.G. Brereton, Chemometrics for Pattern Recognition, John Wiley & Sons Ltd., Chichester, UK, 2009.

[114] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, J. Chemometrics. 28 (2014) 213–225.

[115] P. Porízka, Using Laser-Induced Breakdown Spectroscopy (LIBS) for material analysis., Doctoral thesis, Brno University of Technology, 2014.

[116] E. Marengo, M. Aceto, V. Maurino, Classification of Nebbiolo-based wines from Piedmont (Italy) by means of solid-phase microextraction–gas chromatography–mass spectrometry of volatile compounds, Journal of Chromatography A. 943 (2002) 123–137.

[117] S. Frías, J.E. Conde, J.J. Rodríguez-Bencomo, F. García-Montelongo, J.P. Pérez-Trujillo, Classification of commercial wines from the Canary Islands (Spain) by chemometric techniques using metallic contents, Talanta. 59 (2003) 335–344.

[118] P.D.S. Abe, Introduction, in: Support Vector Machines for Pattern Classification, Springer London, 2010: pp. 1–19..

[119] K.B. Lipkowitz, T.R. Cundari, D.B. Boyd, Reviews in Computational Chemistry, John Wiley & Sons, 2007.

[120] Y. Xu, S. Zomer, R.G. Brereton, Support Vector Machines: A Recent Method for Classification in Chemometrics, Critical Reviews in Analytical Chemistry. 36 (2006) 177–188.

[121] J. Luts, F. Ojeda, R. Van de Plas, B. De Moor, S. Van Huffel, J.A.K. Suykens, A tutorial on support vector machine-based methods for classification problems in chemometrics, Analytica Chimica Acta. 665 (2010) 129–145.

[122] D. Svozil, V. Kvasnicka, J. Pospichal, Introduction to multi-layer feed-forward neural networks, Chemometrics and Intelligent Laboratory Systems. 39 (1997) 43–62.

[123] J. Gasteiger, J. Zupan, Neural Networks in Chemistry, Angew. Chem. Int. Ed. Engl. 32 (1993) 503–527.

[124] B. Kröse, P. van der Smagt, An introduction to neural networks, 1996. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.493 (accessed May 24, 2016).

[125] H.B. Demuth, M.H. Beale, O. De Jess, M.T. Hagan, Neural Network Design, 2nd ed., Martin Hagan, USA, 2014.

[126] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, USA, 1996.

[127] C.M. Bishop, Pattern Recognition and Machine Learning, 1st ed., Springer, 2007.

[128] R. Bro, A. K. Smilde, Principal component analysis, Analytical Methods. 6 (2014) 2812–2831.

[129] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, The Mahalanobis distance, Chemometrics and Intelligent Laboratory Systems. 50 (2000) 1–18.

[130] L. Sheng, T. Zhang, G. Niu, K. Wang, H. Tang, Y. Duan, H. Li, Classification of iron ores by laser-induced breakdown spectroscopy (LIBS) combined with random forest (RF), J. Anal. At. Spectrom. 30 (2015) 453–458.

[131] K.-Q. Yu, Y.-R. Zhao, F. Liu, Y. He, Laser-Induced Breakdown Spectroscopy Coupled with Multivariate Chemometrics for Variety Discrimination of Soil, Scientific Reports. 6 (2016) 27574.

[132] D. Kriesel, A Brief Introduction to Neural Networks, 2007. http://www.dkriesel.com/en/science/neural_networks (accessed June 5, 2016).

[133] H. Cartwright, Using Artificial Intelligence in Chemistry and Biology: A Practical Guide, 1st ed., CRC, 2008.

[134] F.A. Camargo, Learning Algorithms in Neural Networks, 1990.

[135] R. Rojas, Neural Networks: A Systematic Introduction, Springer Science & Business Media, 2013.

[136] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Information Processing & Management. 45 (2009) 427–437.

[137] M. Bylesjö, M. Rantalainen, O. Cloarec, J.K. Nicholson, E. Holmes, J. Trygg, OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, J. Chemometrics. 20 (2006) 341–351.

[138] C. Ling, J. Huang, H. Zhang, AUC: A Better Measure than Accuracy in Comparing Learning Algorithms, in: Advances in Artificial Intelligence, Y. Xiang and B. Chaib-draa (Eds.), Springer - Verlag, Berlin, Heidelberg, 2003: pp. 329–341. http://dx.doi.org/10.1007/3-540-44886-1_25 (accessed April 18, 2010).

[139] K. Fukunaga, Introduction to Statistical Pattern Recognition (2Nd Ed.), Academic Press Professional, Inc., San Diego, CA, USA, 1990.

[140] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition. 30 (1997) 1145–1159.

[141] E.W. Steyerberg, S.E. Bleeker, H.A. Moll, D.E. Grobbee, K.G.M. Moons, Internal and external validation of predictive models: A simulation study of bias and precision in small samples, Journal of Clinical Epidemiology. 56 (2003) 441–447.

[142] F. Provost, T. Fawcett, Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions, in: In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, AAAI Press, 1997: pp. 43–48.

[143] Y. Jiang, B. Cukic, Y. Ma, Techniques for evaluating fault prediction models, Empir. Softw. Eng. 13 (2008) 561–595.

[144] D.J. Hand, R.J. Till, A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems, Mach. Learn. 45 (2001) 171–186.

[145] A.P.M. Michel, A.D. Chave, Analysis of laser-induced breakdown spectroscopy spectra: The case for extreme value statistics, Spectrochim. Acta B. 62 (2007) 1370–1378.

[146] Huber G. Herzberg, Molecular Spectra And Molecular Structure, IV. Constants Of Diatomic Molecules, K. P. Huber G. Herzberg, n.d. http://archive.org/details/MolecularSpectraAndMolecularStructureIV.ConstantsOfDiatomicMoleculesK.P.HuberG.Herzberg (accessed February 3, 2016).

[147] L.D. J. J. Camacho, Time-and space-resolved spectroscopic characterization of laser-induced swine muscle tissue plasma, Spectrochimica Acta - Part B Atomic Spectroscopy. 111 (2015) 92–101.

[148] M. Baudelet, J. Yu, M. Bossu, J. Jovelet, J.-P. Wolf, T. Amodeo, E. Frejafon, P. Laloi, Discrimination of microbiological samples using femtosecond laser-induced breakdown spectroscopy, Appl. Phys. Lett. 89 (2006) 163903–3.

[149] S.J. Rehse, N. Jeyasingham, J. Diedrich, S. Palchaudhuri, A membrane basis for bacterial identification and discrimination using laser-induced breakdown spectroscopy, J. Appl. Phys. 105 (2009) 102034–13.

[150] J. Goldstein, D.E. Newbury, P. Echlin, D.C. Joy, A.D.R. Jr, C.E. Lyman, C. Fiori, E. Lifshin, Scanning Electron Microscopy and X-Ray Microanalysis: A Text for Biologists, Materials Scientists, and Geologists, Springer Science & Business Media, 2012.

[151] G. Bou, A. Fernández-Olmos, C. García, J.A. Sáez-Nieto, S. Valdezate, Métodos de identificación bacteriana en el laboratorio de microbiología, Enfermedades Infecciosas Y Microbiología Clínica. 29 (2011) 601–608.

[152] S. Harbarth, J. Garbino, J. Pugin, J.A. Romand, D. Lew, D. Pittet, Inappropriate initial antimicrobial therapy and its effect on survival in a clinical trial of immunomodulating therapy for severe sepsis, The American Journal of Medicine. 115 (2003) 529–535.

[153] A. Kumar, D. Roberts, K.E. Wood, B. Light, J.E. Parillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar, Cheang, Duration of hypotension before initiation of effective antim... : Critical Care Medicine, Critical Care Medicine. 34 (2006) 1589–1596.

[154] D.A. Cremers, L.J. Radziemski, Handbook of Laser-Induced Breakdown Spectroscopy, Chichester-UK, John Wiley & Sons, 2006.

[155] J.O. Caceres, J. Tornero López, H.H. Telle, A. González Ureña, Quantitative analysis of trace metal ions in ice using laser-induced breakdown spectroscopy, Spectrochim. Acta, Part B. 56 (2001) 831–838.

[156] W.-B. Lee, J. Wu, Y.-I. Lee, J. Sneddon, Recent Applications of Laser-Induced Breakdown Spectrometry: A Review of Material Approaches, Appl. Spectrosc. Rev. 39 (2004) 27–97.

[157] R. Noll, H. Bette, A. Brysch, M. Kraushaar, I. Mönch, L. Peter, V. Sturm, Laser-induced breakdown spectrometry — applications for production control and quality assurance in the steel industry, Spectrochimica Acta Part B: Atomic Spectroscopy. 56 (2001) 637–649.

[158] S. Merk, C. Scholz, S. Florek, D. Mory, Increased identification rate of scrap metal using Laser Induced Breakdown Spectroscopy Echelle spectra, Spectrochimica Acta Part B: Atomic Spectroscopy. 112 (2015) 10–15.

[159] A. Marín-Roldan, S. Manzoor, S. Moncayo, F. Navarro-Villoslada, R.C. Izquierdo-Hornillos, J.O. Caceres, Determination of the postmortem interval by Laser Induced Breakdown Spectroscopy using swine skeletal muscles, Spectrochimica Acta Part B: Atomic Spectroscopy. 88 (2013) 186–191.

[160] F. Mehari, M. Rohde, R. Kanawade, C. Knipfer, W. Adler, F. Klämpfl, F. Stelzle, M. Schmidt, Investigation of the differentiation of ex vivo nerve and fat tissues using laser-induced breakdown spectroscopy (LIBS): Prospects for tissue-specific laser surgery, J. Biophoton. 9 (2016) 1021–1032.

[161] V. Motto-Ros, L. Sancey, X.C. Wang, Q.L. Ma, F. Lux, X.S. Bai, G. Panczer, O. Tillement, J. Yu, Mapping nanoparticles injected into a biological tissue using laser-induced breakdown spectroscopy, Spectrochimica Acta Part B: Atomic Spectroscopy. 87 (2013) 168–174.

[162] D. Santos Jr., L.C. Nunes, G.G.A. de Carvalho, M. da S. Gomes, P.F. de Souza, F. de O. Leme, L.G.C. dos Santos, F.J. Krug, Laser-induced breakdown spectroscopy for analysis of plant materials: A review, Spectrochimica Acta Part B: Atomic Spectroscopy. 71–72 (2012) 3–13.

[163] L.C. Trevizan, D. Santos Jr., R.E. Samad, N.D. Vieira Jr., L.C. Nunes, I.A. Rufini, F.J. Krug, Evaluation of laser induced breakdown spectroscopy for the determination of micronutrients in plant materials, Spectrochimica Acta Part B: Atomic Spectroscopy. 64 (2009) 369–377.

[164] F.-Y. Yueh, H. Zheng, J.P. Singh, S. Burgess, Preliminary evaluation of laser-induced breakdown spectroscopy for tissue classification, Spectrochim. Acta B. 64 (2009) 1059–1067.

[165] S.J. Rehse, H. Salimnia, A.W. Miziolek, Laser-induced breakdown spectroscopy (LIBS): an overview of recent progress and future potential for biomedical applications, Journal of Medical Engineering & Technology. 36 (2012) 77–89.

[166] J.L. Gottfried, Discrimination of biological and chemical threat simulants in residue mixtures on multiple substrates, Anal Bioanal Chem. 400 (2011) 3289–3301.

[167] S.J. Rehse, N. Jeyasingham, J. Diedrich, S. Palchaudhuri, A membrane basis for bacterial identification and discrimination using laser-induced breakdown spectroscopy, J. Appl. Phys. 105 (2009) 102034–13.

[168] M. Mordmueller, C. Bohling, A. John, W. Schade, Rapid test for the detection of hazardous microbiological material, (2009) 74840F–74840F.

[169] R.A. Multari, D.A. Cremers, J.M. Dupre, J.E. Gustafson, The Use of Laser-Induced Breakdown Spectroscopy for Distinguishing Between Bacterial Pathogen Species and Strains, Appl. Spectrosc. 64 (2010) 750–759.

[170] J.C. Moore, J. Spink, M. Lipp, Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010, Journal of Food Science. 77 (2012) R118–R126.

[171] R.E. Shaffer, S.L. Rose-Pehrsson, R.A. McGill, A comparison study of chemical sensor array pattern recognition algorithms, Anal. Chim. Acta. 384 (1999) 305–317.

[172] T. Egli, Microbial growth and physiology: a call for better craftsmanship, Front Microbiol. 6 (2015).

[173] D. Herbert, The chemical composition of micro-organisms as a function of their environment., Symp. Soc. Gen. Microbiol. 11 (1961) 391–416.

[174] R.M. Atlas, J.W. Snyder, Handbook of Media for Clinical and Public Health Microbiology, CRC Press, 2013.

[175] H.B. Demuth, M.H. Beale, M.T. Hagan, Neural network toolbox for use with MATLAB: User's guide 9th for version 6.0 (Release 2008a), Math Works, 2007.

[176] A.J. Maren, C.T. Harston, Handbook of Neural Computing Applications, Academic Press, San diego, USA, 1990.

[177] J.-B. Sirven, B. Bousquet, L. Canioni, L. Sarger, S. Tellier, M. Potin-Gautier, I.L. Hecho, Qualitative and quantitative investigation of chromium-polluted soils by laser-induced breakdown spectroscopy combined with neural networks analysis, Anal. Bioanal. Chem. 385 (2006) 256–262.

[178] M.F. Møller, A scaled conjugate gradient algorithm for fast supervised learning, Neural Networks. 6 (1993) 525–533.

[179] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874.

[180] G.I. Webb, K.M. Ting, On the Application of ROC Analysis to Predict Classification Performance Under Varying Class Distributions, Mach. Learn. 58 (2005) 25–32.

[181] J.-B. Sirven, B. Salle, P. Mauchien, J.-L. Lacour, S. Maurice, G. Manhes, Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods, J. Anal. At. Spectrom. 22 (2007) 1471–1480.

[182] R.J. Nordstrom, Study of Laser-Induced Plasma Emission Spectra of N2, O2, and Ambient Air in the Region 350 nm to 950 nm, Appl. Spectrosc. 49 (1995) 1490–1499.

[183] A. Portnov, S. Rosenwaks, I. Bar, Emission Following Laser-Induced Breakdown Spectroscopy of Organic Compounds in Ambient Air, Appl. Opt. 42 (2003) 2835–2842.

[184] M. Baudelet, M. Boueri, J. Yu, S.S. Mao, V. Piscitelli, X. Mao, R.E. Russo, Time-resolved ultraviolet laser-induced breakdown spectroscopy for organic material analysis, Spectrochim. Acta B. 62 (2007) 1329–1334.

[185] J. Huang, D. Brennan, L. Sattler, J. Alderman, B. Lane, C. O'Mathuna, A comparison of calibration methods based on calibration data size and robustness, Chemometrics and Intelligent Laboratory Systems. 62 (2002) 25–35.

[186] J. Monod, The Growth of Bacterial Cultures, Annu. Rev. Microbiol. 3 (1949) 371–394.

[187] K. Kaier, Economic Implications of the Dynamic Relationship between Antibiotic Use and Hospital-Acquired Infections, Value in Health. 15 (2012) 87–93.

[188] J. Vaqué, J. Rosselló, J.L. Arribas, Prevalence of nosocomial infections in Spain: EPINE study 1990–1997, J. Hosp. Infect. 43, Supplement 1 (1999) S105–S111.

[189] K.W. Lobdell, S. Stamou, J.A. Sanchez, Hospital-Acquired Infections, Surgical Clinics of North America. 92 (2012) 65–77.

[190] HAI-Net Annual Reports, (n.d.). http://www.ecdc.europa.eu/en/healthtopics/healthcare-associated_infections/epidemiological_data/pages/hai-net_annual_reports.aspx (accessed July 30, 2012).

[191] M. Zourob, S. Elwary, A. Turner, eds., Principles of Bacterial Detection: Biosensors, Recognition Receptors and Microsystems, Springer New York, New York, NY, 2008.

[192] A.C. Palmer, R. Kishony, Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance, Nat Rev Genet. 14 (2013) 243–248.

[193] S.H. Ahmed, E.A. Daef, M.S. Badary, M.A. Mahmoud, A.A. Abd-Elsayed, Nosocomial blood stream infection in intensive care units at Assiut University Hospitals (Upper Egypt) with special reference to extended spectrum β-lactamase producing organisms, BMC Res Notes. 2 (2009) 76.

[194] H. Cho, T. Uehara, T.G. Bernhardt, Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery, Cell. 159 (2014) 1300–1311.

[195] H.C. Neu, T.D. Gootz, Antimicrobial Chemotherapy, in: S. Baron (Ed.), Medical Microbiology, 4th ed., University of Texas Medical Branch at Galveston, Galveston (TX), 1996. http://www.ncbi.nlm.nih.gov/books/NBK7986/ (accessed January 31, 2017).

[196] M.A. Kohanski, D.J. Dwyer, J.J. Collins, How antibiotics kill bacteria: from targets to networks, Nat Rev Microbiol. 8 (2010) 423–435.

[197] I. Chopra, M. Roberts, Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance, Microbiol Mol Biol Rev. 65 (2001) 232–260.

[198] K. Zwirglmaier, Fluorescence in situ hybridisation (FISH) – the next generation, FEMS Microbiology Letters. 246 (2005) 151–158.

[199] D. Wang, L. Coscoy, M. Zylberberg, P.C. Avila, H.A. Boushey, D. Ganem, J.L. DeRisi, Microarray-based detection and genotyping of viral pathogens, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 15687–15692.

[200] T. Kostic, P. Butaye, J. Schrenzel, Detection of Highly Dangerous Pathogens: Microarray Methods for the Detection of BSL 3 and BSL 4 Agents, John Wiley & Sons, 2009.

[201] L.S. Garcia, Clinical Microbiology Procedures Handbook, 3 Lslf, ASM Press, 2010.

[202] American Society for Microbiology (Washington), Manual of clinical microbiology, 8th ed, ASM Press, Washington, D.C, 2003.

[203] J.E. Clarridge, Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases, Clin. Microbiol. Rev. 17 (2004) 840–862.

[204] J.L. Cloud, D. Harmsen, P.C. Iwen, J.J. Dunn, G. Hall, P.R. LaSala, K. Hoggan, D. Wilson, G.L. Woods, A. Mellmann, Comparison of Traditional Phenotypic Identification Methods with Partial 5' 16S rRNA Gene Sequencing for Species-Level Identification of Nonfermenting Gram-Negative Bacilli, Journal of Clinical Microbiology. 48 (2010) 1442–1444.

[205] A. Álvarez-Buylla, E. Culebras, J.J. Picazo, Identification of Acinetobacter species: Is Bruker biotyper MALDI-TOF mass spectrometry a good alternative to molecular techniques?, Infection, Genetics and Evolution. 12 (2012) 345–349.

[206] M. Baudelet, L. Guyon, J. Yu, J.-P. Wolf, T. Amodeo, E. Frejafon, P. Laloi, Femtosecond time-resolved laser-induced breakdown spectroscopy for detection and identification of bacteria: A comparison to the nanosecond regime, J. Appl. Phys. 99 (2006) 084701-9.

[207] R.A. Putnam, Q.I. Mohaidat, A. Daabous, S.J. Rehse, A comparison of multivariate analysis techniques and variable selection strategies in a laser-induced breakdown spectroscopy bacterial classification, Spectrochimica Acta Part B: Atomic Spectroscopy. 87 (2013) 161–167.

[208] F. Garcia del Portillo, M.A. de Pedro, Differential effect of mutational impairment of penicillin-binding proteins 1A and 1B on Escherichia coli strains harboring thermosensitive mutations in the cell division genes ftsA, ftsQ, ftsZ, and pbpB., J Bacteriol. 172 (1990) 5863–5870.

[209] C.P. Guillermo, Caracterización fenotípica y análisis de componentes macromoleculares de pseudomonas aeruginosa procedentes de aislados clínicos cubanos, Tesis - Universidad Autónoma de Madrid - Spain, 2003. http://dialnet.unirioja.es/servlet/tesis?codigo=32557 (accessed October 17, 2013).

[210] A.C.S. Lopes, D.L. Veras, A.M.S. Lima, R. de C.A. Melo, J.A. Ayala, bla(CTX-M-2) and bla(CTX-M-28) extended-spectrum beta-lactamase genes and class 1 integrons in clinical isolates of Klebsiella pneumoniae from Brazil, Mem. Inst. Oswaldo Cruz. 105 (2010) 163–167.

[211] J. Velasco, M. Araque, J.A. Ayala, Contruction of Salmonella enterica Mutants by inactivation of invG/invE y ssaJ/ssaK genes of patogenicity islans, Colombiana de Microbiología. 12 (2010) 55–66.

[212] J.-B. Sirven, B. Bousquet, L. Canioni, L. Sarger, Laser-Induced Breakdown Spectroscopy of Composite Samples: Comparison of Advanced Chemometrics Methods, Analytical Chemistry. 78 (2006) 1462–1469.

[213] A. Ramil, A.J. López, A. Yáñez, Application of artificial neural networks for the rapid classification of archaeological ceramics by means of laser induced breakdown spectroscopy (LIBS), Appl. Phys. A. 92 (2008) 197–202..

[214] F.C.D.L. Chase A. Munson, Investigation of statistics strategies for improving the discriminating power of laser-induced breakdown spectroscopy for chemical and biological warfare agent simulants, Spectrochimica Acta Part B: Atomic Spectroscopy. (2005) 1217–1224.

[215] M.Z. Martin, N. Labbé, T.G. Rials, S.D. Wullschleger, Analysis of preservative-treated wood by multivariate analysis of laser-induced breakdown spectroscopy spectra, Spectrochimica Acta Part B: Atomic Spectroscopy. 60 (2005) 1179–1185.

[216] E. D'Andrea, S. Pagnotta, E. Grifoni, S. Legnaioli, G. Lorenzetti, V. Palleschi, B. Lazzerini, A hybrid calibration-free/artificial neural networks approach to the quantitative analysis of LIBS spectra, Appl. Phys. B. 118 (2015) 353–360.

[217] C. Tan, X. Qin, M. Li, Comparison of chemometric methods for brand classification of cigarettes by near-infrared spectroscopy, Vibrational Spectroscopy. 51 (2009) 276–282.

[218] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, 1 edition, Wiley-Interscience, Hoboken, NJ - USA, 2004.

[219] A.K. Myakalwar, S. Sreedhar, I. Barman, N.C. Dingari, S. Venugopal Rao, P. Prem Kiran, S.P. Tewari, G. Manoj Kumar, Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis, Talanta. 87 (2011) 53–59.

[220] Q. Godoi, F.O. Leme, L.C. Trevizan, E.R. Pereira Filho, I.A. Rufini, D. Santos Jr., F.J. Krug, Laser-induced breakdown spectroscopy and chemometrics for classification of toys relying on toxic elements, Spectrochimica Acta Part B-Atomic Spectroscopy. 66 (2011) 138–143.

[221] F. Colao, R. Fantoni, P. Ortiz, M.A. Vazquez, J.M. Martin, R. Ortiz, N. Idris, Quarry identification of historical building materials by means of laser induced breakdown spectroscopy, X-ray fluorescence and chemometric analysis, Spectrochimica Acta Part B: Atomic Spectroscopy. 65 (2010) 688–694.

[222] N.C. Dingari, I. Barman, A.K. Myakalwar, S.P. Tewari, M. Kumar Gundawar, Incorporation of Support Vector Machines in the LIBS Toolbox for Sensitive and Robust Classification Amidst Unexpected Sample and System Variability, Anal. Chem. 84 (2012) 2686–2694.

[223] J. Moros, J. Serrano, C. Sánchez, J. Macías, J.J. Laserna, New chemometrics in laser-induced breakdown spectroscopy for recognizing explosive residues, J. Anal. At. Spectrom. 27 (2012) 2111–2122.

[224] J. Diedrich, S.J. Rehse, S. Palchaudhuri, Pathogenic Escherichia coli strain discrimination using laser-induced breakdown spectroscopy, J. Appl. Phys. 102 (2007) 014702-8.

[225] R.A. Multari, D.A. Cremers, M.L. Bostian, J.M. Dupre, J.E. Gustafson, Proof of Principle for a Real-Time Pathogen Isolation Media Diagnostic: The Use of Laser-Induced Breakdown Spectroscopy to Discriminate Bacterial Pathogens and Antimicrobial-Resistant *Staphylococcus aureus* Strains Grown on Blood Agar, Journal of Pathogens. 2013 (2013).

[226] J. Cisewski, E. Snyder, J. Hannig, L. Oudejans, Support vector machine classification of suspect powders using laser-induced breakdown spectroscopy (LIBS) spectral data, J. Chemometrics. 26 (2012) 143–149.

[227] T.-L. Zhang, S. Wu, H.-S. Tang, K. Wang, Y.-X. Duan, H. Li, Progress of Chemometrics in Laser-induced Breakdown Spectroscopy Analysis, Chinese Journal of Analytical Chemistry. 43 (2015) 939–948.

[228] J.B. Sirven, B. Salle, P. Mauchien, J.-L. Lacour, S. Maurice, G. Manhes, Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods, J. Anal. At. Spectrom. 22 (2007) 1471–1480.

[229] G. Vítková, K. Novotný, L. Prokeš, A. Hrdlička, J. Kaiser, J. Novotný, R. Malina, D. Prochazka, Fast identification of biominerals by means of stand-off laser-induced breakdown spectroscopy using linear discriminant analysis and artificial neural networks, Spectrochimica Acta Part B: Atomic Spectroscopy. 73 (2012) 1–6.

[230] J.-B. Sirven, B. Sallé, P. Mauchien, J.-L. Lacour, S. Maurice, G. Manhès, Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods, J. Anal. At. Spectrom. 22 (2007) 1471–1480.

[231] M. Baudelet, J. Yu, M. Bossu, J. Jovelet, J.-P. Wolf, T. Amodeo, E. Frejafon, P. Laloi, Discrimination of microbiological samples using femtosecond laser-induced breakdown spectroscopy, Appl. Phys. Lett. 89 (2006) 163903–3.

[232] J. Diedrich, S.J. Rehse, S. Palchaudhuri, Escherichia coli identification and strain discrimination using nanosecond laser-induced breakdown spectroscopy, Appl. Phys. Lett. 90 (2007) 163901–3.

[233] M. Baudelet, L. Guyon, J. Yu, J.-P. Wolf, T. Amodeo, E. Frejafon, P. Laloi, Spectral signature of native CN bonds for bacterium detection and identification using femtosecond laser-induced breakdown spectroscopy, Appl. Phys. Lett. 88 (2006) 063901-3.

[234] C. de Bedout, B.L. Gómez, Candida y candidiasis invasora: un reto continuo para su diagnóstico temprano, Infectio. 14, Supplement 2 (2010) 159–171.

[235] J. de C.O. Sardi, N. de S. Pitangui, F.P.G. and A.M.F.A. e M.J.S.M. Giannini, A Mini Review of Candida Species in Hospital Infection: Epidemiology,Virulence Factor and Drugs Resistance and Prophylaxis, Trop Med Surg. 1 (2013) 1–7.

[236] G. Pulcrano, D.V. Iula, A. Vollaro, A. Tucci, M. Cerullo, M. Esposito, F. Rossano, M.R. Catania, Rapid and reliable MALDI-TOF mass spectrometry identification of Candida non-albicans isolates from bloodstream infections, Journal of Microbiological Methods. 94 (2013) 262–266.

[237] M.A. Pfaller, P.G. Pappas, J.R. Wingard, Invasive Fungal Pathogens: Current Epidemiological Trends, Clin Infect Dis. 43 (2006) S3–S14.

[238] T.J. Walsh, A. Groll, J. Hiemenz, R. Fleming, E. Roilides, E. Anaissie, Infections due to emerging and uncommon medically important fungal pathogens, Clin. Microbiol. Infect. 10 Suppl 1 (2004) 48–66.

[239] A.W. Maksymiuk, S. Thongprasert, R. Hopfer, M. Luna, V. Fainstein, G.P. Bodey, Systemic candidiasis in cancer patients, Am. J. Med. 77 (1984) 20–27.

[240] M.Z. Alam, Q. Alam, A. Jiman-Fatani, M.A. Kamal, A.M. Abuzenadah, A.G. Chaudhary, M. Akram, A. Haque, Candida identification: a journey from conventional to molecular methods in medical mycology, World J. Microbiol. Biotechnol. 30 (2014) 1437–1451.

[241] M.A. Pfaller, D.J. Diekema, Rare and Emerging Opportunistic Fungal Pathogens: Concern for Resistance beyond Candida albicans and Aspergillus fumigatus, J. Clin. Microbiol. 42 (2004) 4419–4431.

[242] P.G. Pappas, J.H. Rex, J. Lee, R.J. Hamill, R.A. Larsen, W. Powderly, C.A. Kauffman, N. Hyslop, J.E. Mangino, S. Chapman, H.W. Horowitz, J.E. Edwards, W.E. Dismukes, N.M.S. Group, A Prospective Observational Study of Candidemia: Epidemiology, Therapy, and Influences on Mortality in Hospitalized Adult and Pediatric Patients, Clin Infect Dis. 37 (2003) 634–643.

[243] N. Yapar, Epidemiology and risk factors for invasive candidiasis, Ther Clin Risk Manag. 10 (2014) 95–105.

[244] R. Di Santo, Natural products as antifungal agents against clinically relevant pathogens, Nat Prod Rep. 27 (2010) 1084–1098.

[245] J. Hsu, S. Ruoss, N. Bower, M. Lin, M. Holodniy, D. Stevens, Diagnosing invasive fungal disease in critically ill patients., 37 (n.d.) 277–312.

[246] O. Gudlaugsson, S. Gillespie, K. Lee, J.V. Berg, J. Hu, S. Messer, L. Herwaldt, M. Pfaller, D. Diekema, Attributable Mortality of Nosocomial Candidemia, Revisited, Clin Infect Dis. 37 (2003) 1172–1177.

[247] J. Morgan, M.I. Meltzer, B.D. Plikaytis, A.N. Sofair, S. Huie-White, S. Wilcox, L.H. Harrison, E.C. Seaberg, R.A. Hajjeh, S.M. Teutsch, Excess mortality, hospital stay, and cost due to candidemia: a case-control study using data from population-based candidemia surveillance, Infect Control Hosp Epidemiol. 26 (2005) 540–547.

[248] M. Essendoubi, D. Toubas, M. Bouzaggou, J.-M. Pinon, M. Manfait, G.D. Sockalingum, Rapid identification of Candida species by FT-IR microspectroscopy, Biochimica et Biophysica Acta (BBA) - General Subjects. 1724 (2005) 239–247.

[249] D. Liu, Molecular Detection of Human Fungal Pathogens, CRC Press, 2011.

[250] T.E. Zaoutis, J. Argon, J. Chu, J.A. Berlin, T.J. Walsh, C. Feudtner, The Epidemiology and Attributable Outcomes of Candidemia in Adults and Children Hospitalized in the United States: A Propensity Analysis, Clin Infect Dis. 41 (2005) 1232–1239.

[251] G.D. Brown, D.W. Denning, S.M. Levitz, Tackling Human Fungal Infections, Science. 336 (2012) 647–647.

[252] K. Maquelin, L.P. Choo-Smith, H.P. Endtz, H.A. Bruining, G.J. Puppels, Rapid identification of Candida species by confocal Raman microspectroscopy, J. Clin. Microbiol. 40 (2002) 594–600.

[253] J. Nobrega de Almeida Júnior, L.B. de Souza, A.L. Motta, F. Rossi, T.S. Romano Di Gioia, G. Benard, G.M.B. Del Negro, Evaluation of the MALDI-TOF VITEK MS^{TM} system for the identification of Candida parapsilosis, C. orthopsilosis and C. metapsilosis from bloodstream infections, Journal of Microbiological Methods. 105 (2014) 105–108.

[254] N. Papon, V. Courdavault, M. Clastre, R.J. Bennett, Emerging and Emerged Pathogenic Candida Species: Beyond the Candida albicans Paradigm, PLoS Pathog. 9 (2013) e1003550. doi:10.1371/journal.ppat.1003550.

[255] S.G.W. Kaminskyj, T.E.S. Dahms, High spatial resolution surface imaging and analysis of fungal cells using SEM and AFM, Micron. 39 (2008) 349–361.

[256] N. US Department of Commerce, NIST Atomic Spectra Database, (2015). http://physics.nist.gov/PhysRefData/ASD/lines_form.html (accessed April 24, 2014).

[257] J. Anzano, M.-E. Casanova, M.-S. Bermúdez, R.-J. Lasheras, Rapid characterization of plastics using laser-induced plasma spectroscopy (LIPS), Polymer Testing. 25 (2006) 623–627.

[258] J.M. Anzano, M.A. Villoria, A. Ruíz-Medina, R.J. Lasheras, Laser-induced breakdown spectroscopy for quantitative spectrochemical analysis of geological materials: Effects of the matrix and simultaneous determination, Anal. Chim. Acta. 575 (2006) 230–235.

[259] P. Lucena, A. Doña, L.M. Tobaria, J.J. Laserna, New challenges and insights in the detection and spectral identification of organic explosives by laser induced breakdown spectroscopy, Spectrochim. Acta B At. Spec. 66 (2011) 12–20.

[260] P. Lucena, I. Gaona, J. Moros, J.J. Laserna, Location and detection of explosive-contaminated human fingerprints on distant targets using standoff laser-induced breakdown spectroscopy, Spectrochimica Acta Part B: Atomic Spectroscopy. 85 (2013) 71–77.

ANNEXES

7.1. USER INTERFACE

For the development of this thesis, a user interface was developed in order to perform the data analysis, from the generation of data matrix consisting of input variables to the creation and validation of Neural Network classification models,. Once the operational parameters are optimized, the user interface allow an easy manipulation of the recorded data and therefore can facilitate a non-specialist person to process the data. The interface and data analysis involves three main components, i.e. generation of data matrices from LIBS spectral data, training of the model and finally testing procedures. The interface offers the possibility to modify essential parameters depending on the results of operation and possesses necessary tools that allow the rapid data incorporation and integration to the developed algorithms.

The interface developed for the creation of data matrix is presented in Fig. 7.1, which allowed the combination of individual spectra and preprocessing of data. This includes normalization, adjusting the baseline and selection of variables as optimized by the user. The Fig. 7.1.1 shows that the interface has two panels. The left panel deals with loading spectral data of a sample using the root name of spectra files, introducing the first and last number of spectrum to mention the range of spectra required to create the spectral libraries or data matrix and finally a label to the library. The right panel shows the processed spectra in the form of graphs. Therefore, any error during the process can be detected and the procedure repeated.

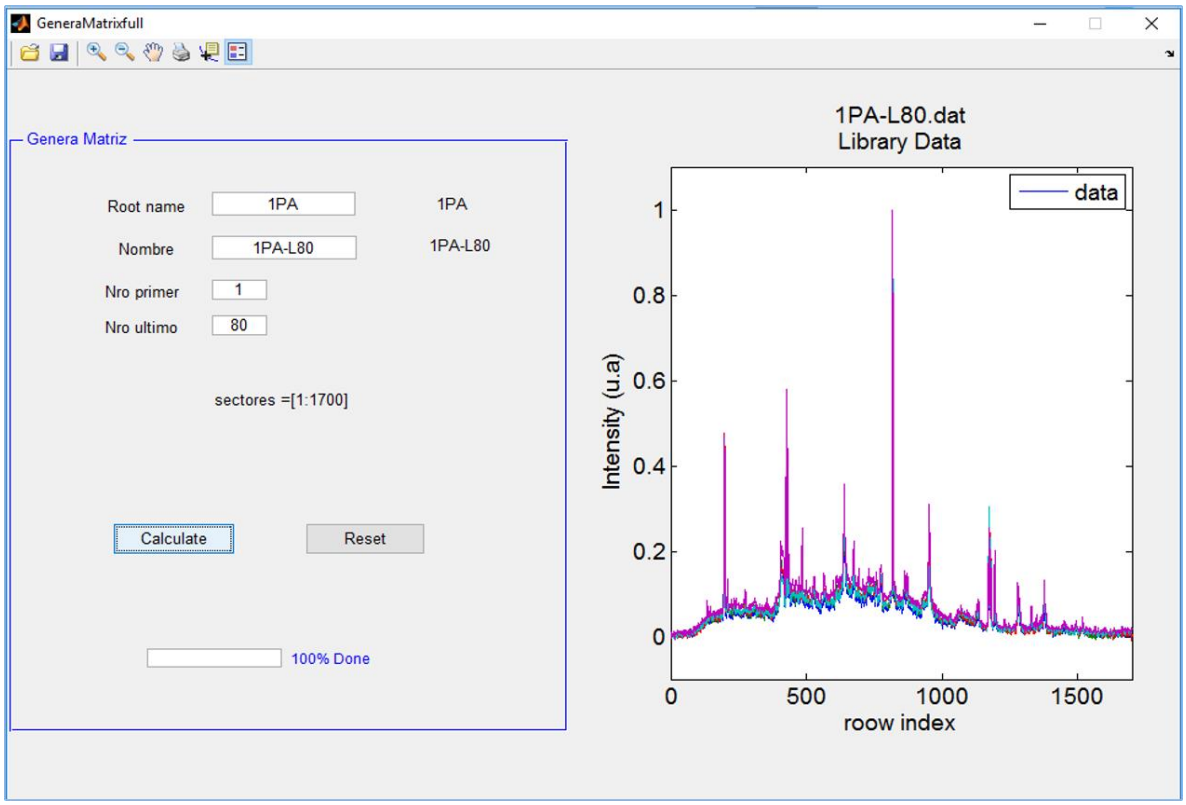


Fig. 7.1.1. Interface for creating Spectral libraries

For the analysis of data by Neural Networks, two interfaces were created in Matlab, one for the training of NN models and the other for testing to make the prediction of test samples. Fig. 7.2 shows the training interface, where the left panel contains the sections to load the spectral libraries required to train the NN model. In the given example, spectral training libraries of strains 1PA and 2PA were loaded in "*Select Reference 1*" and "*Select Reference 2*", respectively. The number of spectra forming each spectral library and number of neurons in hidden layer were mentioned and Training button is clicked. When the training is completed the performance parameters of the network appear on the left panel. The results of training are shown on the right panel. It can be seen that Reference 1 has been assigned an output of 1 and Reference 2 is 2. If the results are not satisfactory, the process is repeated until an accepted model is obtained.

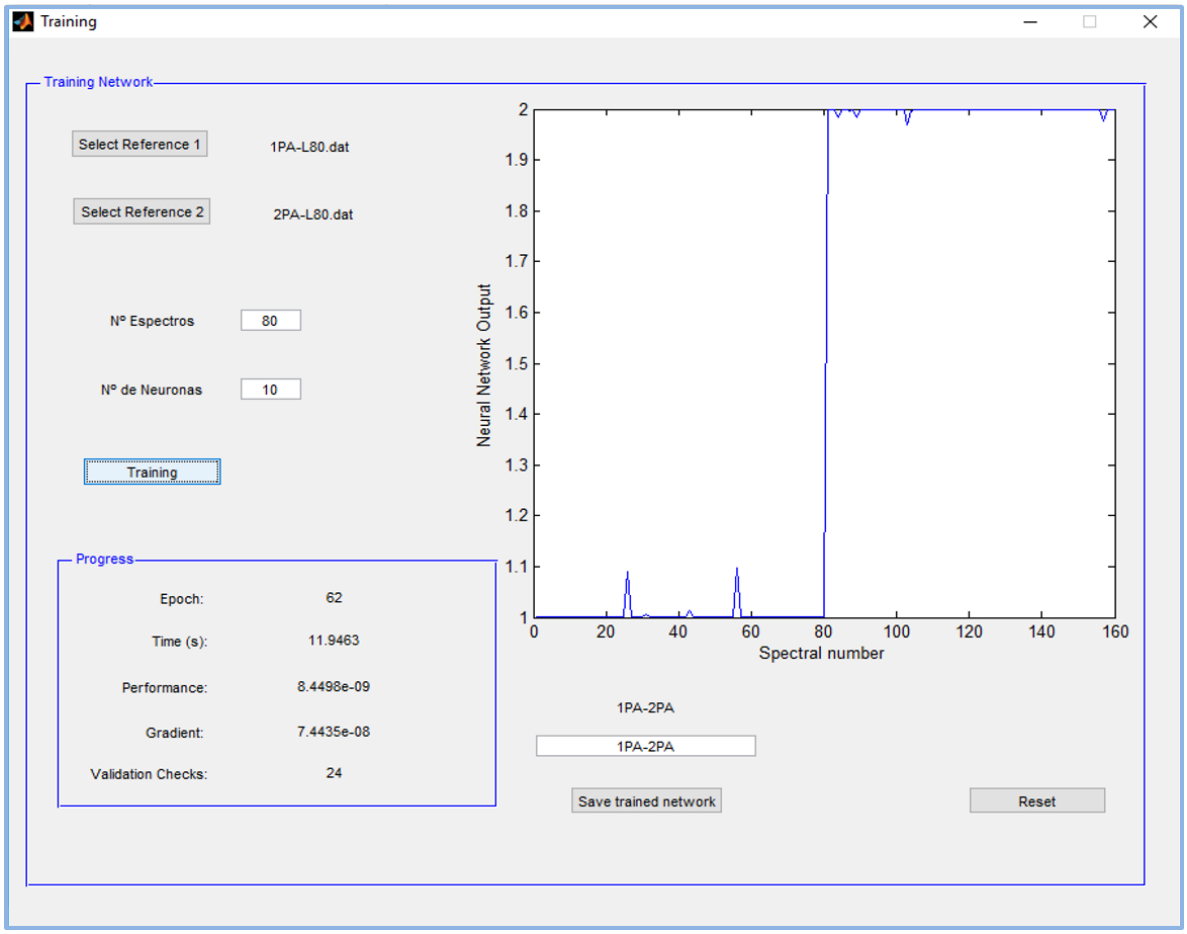


Fig. 7.1.2. Interface for the training of NN model

The trained model is saved and followed by testing process. The interface developed for this purpose is shown in Fig. 7.3, which includes loading the test sample and trained model. Clicking on the Test file button, the NN model provides the prediction results for the test sample and results are shown on the bottom left side, giving the spectral correlation for each Reference library as well as any unassigned spectrum. Spectral Correlation later replaced the term spectral correlation. In the given example, test sample used was strain 1PA, which was predicted with 95% to Reference 1 (1PA) and 5% was classified as unknown. Similar to other interfaces, the right panel showed the visual results and it can be seen that except one all spectra are present at a NN output value of 1. The visual results add value to the interface by showing how much variation or

deviation is obtained in the results. As the results were 95% correct identification and one spectrum seems to be unknown, the visual inspection shows that the deviation was out of the defined threshold value; however, it was still closer to Reference 1.

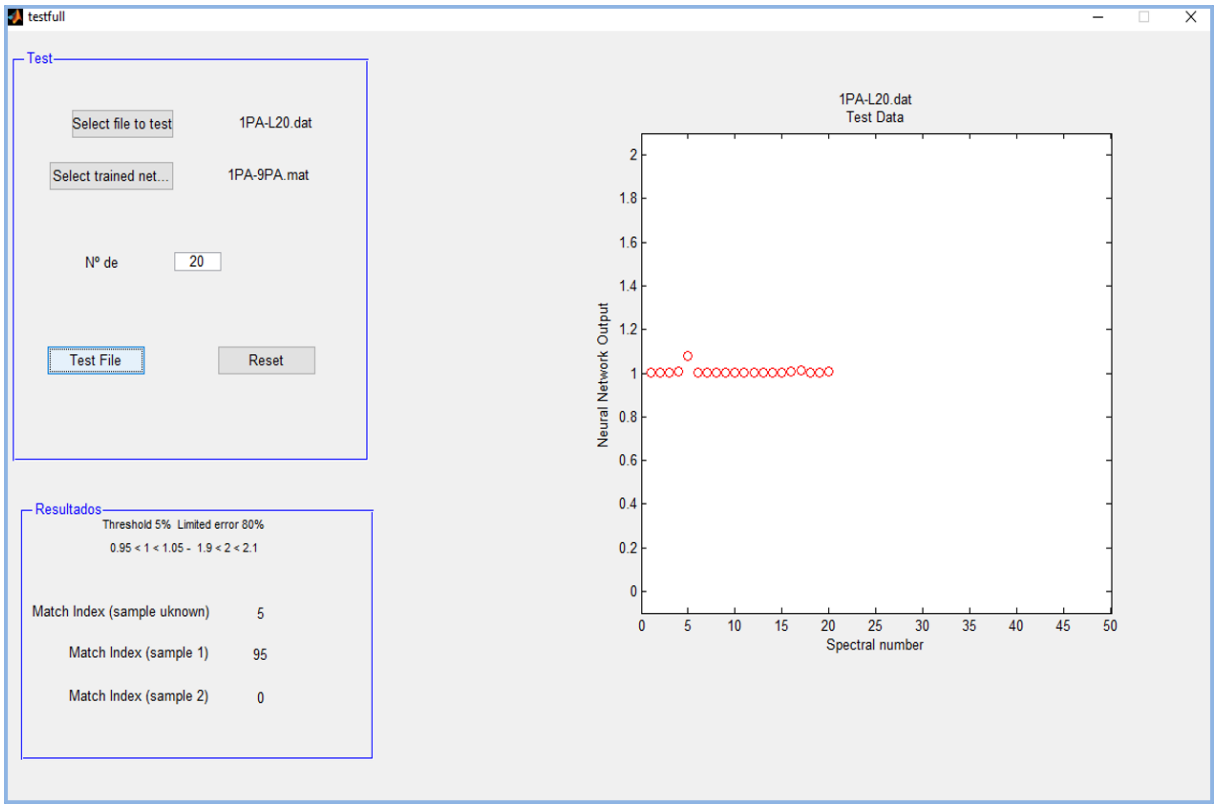
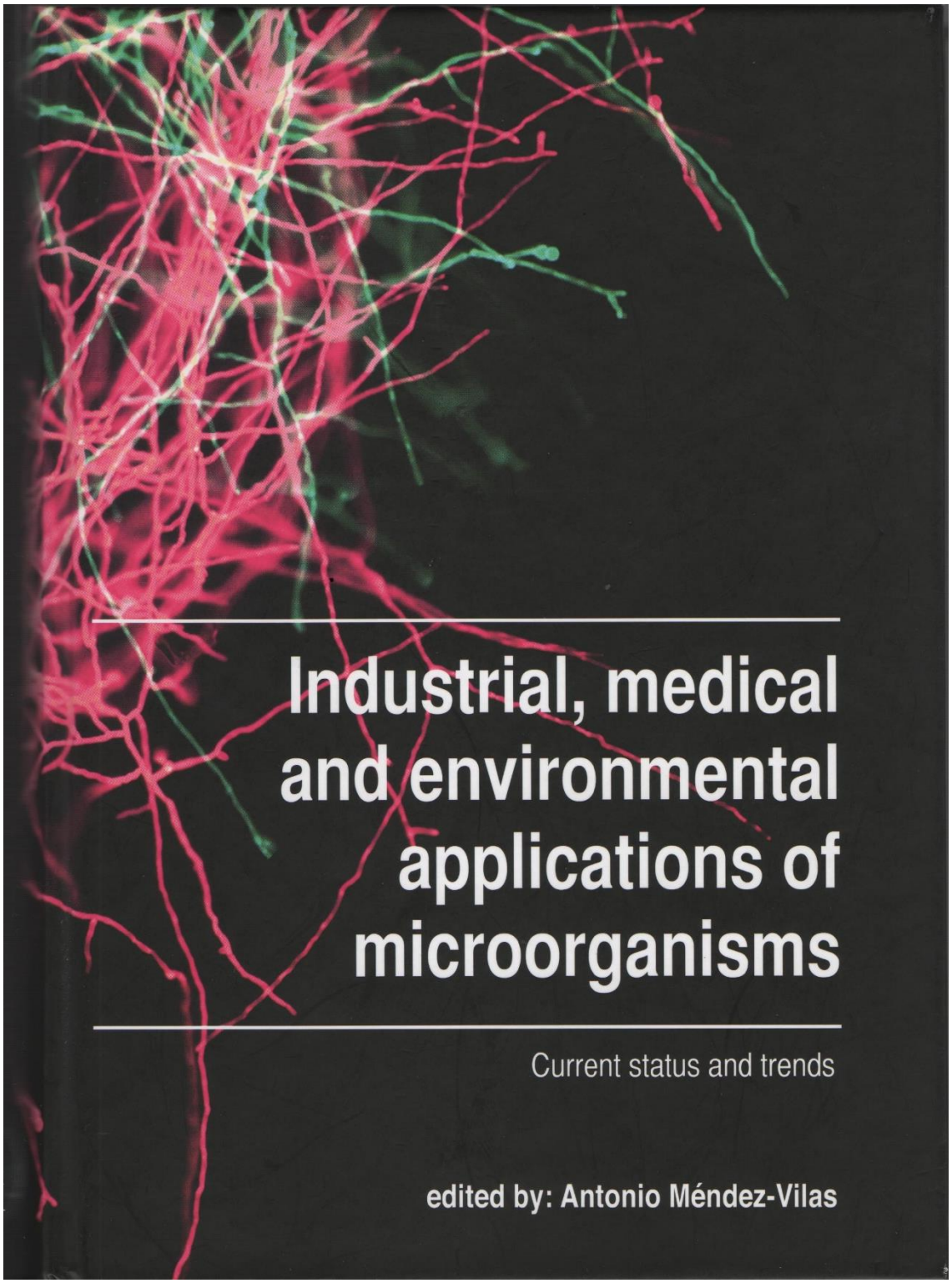


Fig. 7.1.3. Interface developed for testing process

7.2. CONFERENCE PROCEEDINGS



Fast bacterial strain identification by laser induced breakdown spectroscopy and neural networks

S. Manzoor[1], S. Moncayo[1], F. Navarro-Villoslada[1], J.A. Ayala[2], R. Izquierdo-Hornillos[1], F.J. Manuel de Villena[1] and J.O. Caceres[1]

[1]Departamento de Química Analítica, Facultad de Ciencias Químicas, Universidad Complutense, 28040 Madrid, Spain; jcaceres@ucm.es

[2]Centro de Biología Molecular 'Severo Ochoa', CSIC, C/Nicolás Cabrera 1, Cantoblanco 28049 Madrid, Spain

Abstract

A method for rapid bacterial strain identification based on Laser Induced Breakdown Spectroscopy (LIBS) and Neural Networks (NN) is reported. The study includes bacterial strains of the most relevant bacteria causing Hospital Acquired Infections (HAI), i.e. *Pseudomonas aeruginosa*, *Escherichia coli, Klebsiella pneumoniae, Salmonella typhimurium* and *Staphylococcus aureus.* LIBS/NN methodology was evaluated for its capacity to discriminate different bacterial strains using their characteristic LIBS spectra from changes in their elemental composition as a result of genetic variations. The samples were measured for two different days to evaluate the time-dependent classification capacity of the methodology. A successful classification of bacterial strains by the proposed LIBS/NN method, with accuracy above 95%, shows its potential to address the safety and social-cost HAI-related issue.

Keywords: bacterial strain discrimination; laser induced breakdown spectroscopy (LIBS); neural networks (NN); bacteria; hospital acquired infections (HAI)

1. Introduction

A rapid bacterial identification is of utmost importance in medical diagnosis because of the tremendous impact in health and economic issues. The economic and health aspects related to HAI have raised them to top-priority issue and a subject of various studies [1]. Reports show that bacteria are responsible for 95% of hospital acquired infections (HAI). The more common and relevant bacteria among them are *Escherichia coli* (18.2%), *Staphylococcus* (18.1%), *Pseudomonas* (6.0%), *Enterococcus* (15.4%), *Klebsiella* (3.7%), *Acinetobacter* (0.8%) and *Salmonella* (2.8%). There is a general consensus suggesting that every year about 5-7% of hospitalized patients are affected by HAI that increases hospitalization time and costs; whereas, 1% of cases end in death [2]. Therefore, to avoid and deal with all these unwanted situations a method that is fast enough to identify bacteria in the clinical samples is highly desirable.

There are many methods for bacterial identification which involve phenotypic methods [3], biochemical testing [4] and spectrometric techniques [5,6]. Among the molecular-based methods, Polymerase Chain Reaction [7,8], microarray assemblies [9,10] and fluorescent probes [11] have been mostly employed [12]. Although these methods provide accurate results, they involve special sample preparation steps, long incubation time, and laborious work to achieve bacterial identification. These features make these methods more time-consuming and expensive, limiting their use as routine diagnostic methods, which may increase the rate of HAI.

In the present study a fast and simple method based on a non-biological approach combining Laser Induced Breakdown Spectroscopy (LIBS) and Neural Networks (NN) is being proposed to identify bacterial samples for some very important bacterial pathogens causing HAI.

LIBS is a spectroscopic technique involving the irradiation with a laser pulse on a target and the generation of a plasma. The plasma is composed by excited atoms present in the sample that emits radiation, providing a characteristic spectral fingerprint of the sample [13]. During the last decade many studies have been carried out to demonstrate the use of LIBS combined with chemometric techniques to identify and discriminate different bacterial species based on their elemental composition [14-17]. The motivation behind these studies has been the advantages of the technique such as little or no sample preparation, simultaneous monitoring of both major and trace elements and high speed of analysis [15].

In a previous study, an NN method in combination with LIBS have showed as a promising methodology to classify and predict bacterial samples at genus level with a high degree of precision and accuracy [18].

The aim of this paper is to determine if genetic variations between different bacterial strains of the same species generate changes in their spectral fingerprints to be detected by LIBS/NN methodology to achieve their discrimination and identification.

2. Materials and methods

2.1 LIBS set-up

The LIBS technique and the methodology used in the present work together with the most significant experimental conditions have been described elsewhere [18]. Briefly, LIBS measurements were obtained using a Q-switched Nd:YAG laser (Quantel, Brio model) operating at 1064 nm, with a pulse duration of 4 ns full width at half maximum (FWHM). Samples were placed over an X–Y–Z manual micro-metric to ensure that each laser pulse impinged on a fresh sample. The laser beam was focused onto the sample surface producing a spot of 100 μm in diameter. The laser fluence was fixed to 20 J/cm^2 and the repetition rate was 1 Hz. Emission from the plasma was collected by an optical fiber), coupled to a spectrometer. The spectrometer system was EPP2000, StellarNet (Tampa, FL, U.S.A.) with a CCD detector. A spectral resolution of 0.5 nm was achieved. The wavelength range used was from 200 to 1000 nm. In order to prevent the detection of bremsstrahlung, the detector was triggered with a 5 μs delay time between the laser pulse and the acquired plasma radiation using a digital delay generator (Stanford model DG535). The spectrometer was computer-controlled using an interface developed with Matlab, which allowed for data processing and real-time analysis.

2.2 Bacterial Samples

Each bacterial strain was incubated at 37 °C for 12 h in three Petri dishes (8.9 cm in diameter). The bacterial samples were wild-type strains of *E. coli* (Ec), *Pseudomonas aeruginosa* (Pa), *Klebsiella pneumoniae* (Kp), *Salmonella typhimurium* (St) and *Staphylococcus aureus* (Sa). The bacterial strains were cultivated in LB agar Difco Microbiology (Lawrence, KS, U.S.A). Table 1 shows the ID nomenclature used for the bacterial strains, where sample ID is represented in a XYZ pattern, where X is the genus, Y the bacterial specie and Z the type of strain. Thus, Ec1 refers to the first strain of *E. coli* (MC6-RP11).

2.3 LIBS measurements and spectral libraries

Bacterial samples were measured directly in the Petri dish at room experimental conditions with 100 laser pulses/day on two different days one week apart by. For each pulse, the spectrum was acquired and stored; being the acquisition time of these 100 spectra was less than 2 min. Because the emission intensity signals vary with the laser pulse, each spectrum was normalized by the most intense spectral line (Na (I)) to avoid instrumental variations. 80 random spectra were used to train the NN model for each bacterial strain (spectral training library) and the other 20 spectra to validate the model (spectral test library). The computation time for training the NN models was below 10 s.

Table 1. Nomenclature of bacterial samples.

Bacteria	Strain	Sample ID
Klebsiella pneumoniae	KBC	Kp1
Escherichia coli	MC6-RP11	Ec1
	MG1655	Ec20
	QCB1	Ec3
Pseudomonas aeruginosa	M8A4	Pa1
	M9A1	Pa2
	M8A1	Pa3
Staphylococcus aureus	CP	Sa1
Salmonella typhimurium	LB5010	St1

Part 7

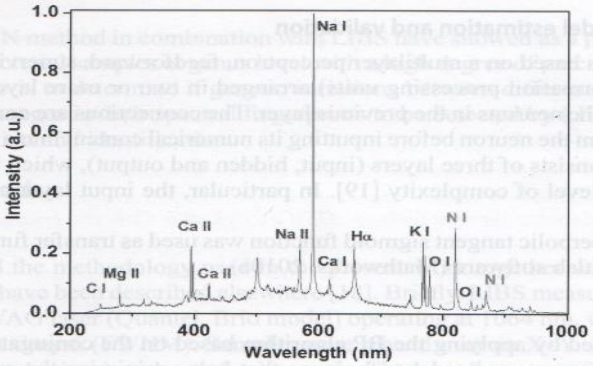


Fig. 1. Typical LIBS spectrum of *Escherichia coli*.

3.1 Intra-bacterial strain discrimination

The LIBS-NN methodology was evaluated to discriminate strains of same bacteria species (i.e. Ec and Pa). NN models were estimated with two spectral training libraries of the same bacteria species and validated with spectral test libraries of both bacterial species. All binary combinations of bacterial strains were considered in order to estimate different NN models. Table 2 shows the average match index for the analysis of the three Ec and Pa strains. In all cases the spectral correlation was higher than 97%, indicating a good performance of the models to classify the bacterial strain samples for each species. Therefore, when the NN model is trained with spectral datasets of different bacterial strains, it is able to discriminate between bacterial strains of the same bacteria species that differ in tiny genetic variations which impart a characteristic fingerprint in the LIBS spectrum.

3.2 Inter-bacterial strain discrimination

The second validation of the LIBS-NN method was the evaluation of its capacity to discriminate bacterial species independent of the strain of the bacterial specie used to train the NN model. For this purpose, the NN models were trained with two strains of different bacterial species and validated with spectral libraries of all strains of both bacterial species. All binary combinations between different bacterial strains were studied. Table 3 shows the spectral correlation for each bacterial strain. Kp1, Ec, Pa, Sa and St strains were classified to the corresponding species with a spectral correlation above 95%, 98% 95%, 100% and 98%, respectively. Therefore, the developed LIBS-NN methodology was able to produce accurate and reliable results in the identification of bacterial strains independently of the bacterial strain employed for training the NN model. On the other hand, spectral data not correctly classified by the NN model was assigned as unknown, that is, not belonging to another bacterial species either. These results demonstrate the ability and robustness of the proposed LIBS-NN methodology to discriminate different bacterial species and bacterial strains even in the case of small genetic mutations.

Table 2. Average match index for the bacterial strain discrimination of *Escherichia coli* and *Pseudomonas aeruginosa* species.

		Predicted							
		Ec1	Ec2	Ec3	Unknown		Pa1	Pa2	Pa3	Unknown
Actual	Ec1	98.3	0	0	1.7	Pa1	98.3	0	0	1.7
	Ec2	0	100	0	0	Pa2	0	97.5	0	2.5
	Ec3	0	0	100	0	Pa3	0	0	100	0

Table 3. Spectral correlation of the inter-bacterial strain discrimination.

		Predicted					
		Klebsiella pneumoniae	*Escherichia coli*	*Pseudomonas aeruginosa*	*Staphylococcus aureus*	*Salmonella typhimurium*	Unknown
Actual	Kp1	95.0	0	0	0	0	5.0
	Ec1	0	96.6	0	0	0	3.4
	Ec2	0	98.3	0	0	0	1.7
	Ec3	0	100	0	0	0	0
	Pa1	0	0	96.6	0	0	3.4
	Pa2	0	0	100	0	0	0
	Pa3	0	0	95.0	0	0	5.0
	Sa1	0	0	0	100	0	0
	St1	0	0	0	0	98.3	1.7

3.3 Time-dependent prediction capacity

To evaluate the time-dependence prediction capacity of the LIBS-NN method for the classification of the bacterial samples, each strain was measured under the same experimental conditions on two different days one week apart. During this time range bacteria were deprived of nutrition. NN models were estimated from the spectral library measured on the first day and the validation of the model was done by the spectral library of both days. Fig. 2 shows the spectral correlation and predicted classification on both days for the five bacterial species. Bacterial strains were correctly classified with an average spectral correlation higher than 93%. Therefore, no significant differences in the elemental composition (LIBS spectral fingerprint) of the bacterial specimens were observed due to nutrition deprivation.

4. Conclusions

A method based on LIBS and an NN algorithm was developed and applied to achieve rapid identification and discrimination of bacterial strains. Significant supervised Neural Network models were estimated for identification of different bacterial species and strains causing HAI. The proposed LIBS-NN methodology provided a quick identification with a high level of accuracy to successfully discriminate a wide variety of bacteria and bacterial strains. Single shot measurements were sufficient for a definite identification of the bacterial strains.

The proposed methodology may be considered as a quick, simple and cost-effective alternative for the slower and more expensive biological methods to discriminate strains of the same bacterial specie.

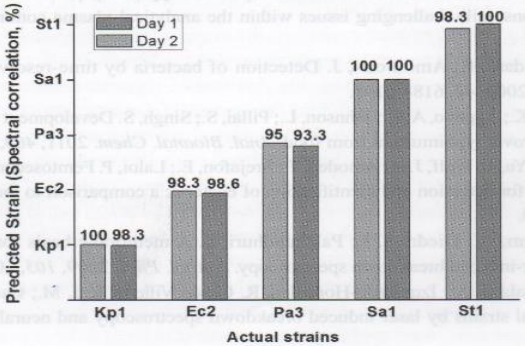


Fig. 2. Time-dependent classification capacity of the LIBS-NN method.

On the other hand, the method could be automatized for real time and reliable measurements as well as be packed into portable systems for in-situ measurements. From the medical point of view, these capabilities would allow an early diagnosis of the bacterial infections and their treatment which may reduce the recurrence of HAI.

Acknowledgements

This work was supported by the Complutense University of Madrid (grant CCG10-UCM/PPQ-4713) granted to JO Caceres and grant BFU2009-09200 from the Ministry of Science and Innovation of Spain, granted to JA Ayala.

References

[1] Kaier, K. Economic implications of the dynamic relationship between antibiotic use and hospital-acquired infections. *Value Health*. 2012, *15*, 87-93.

[2] *The European Healthcare-Associated Infections. HAI-Net Annual Reports.* http://www.ecdc.europa.eu/en/healthtopics/healthcare-associated_infections/epidemiological_data/pages/hai-net_annual_reports.aspx. julio 2012,.

[3] Reva, O. N.; Sorokulova, I. B.; Smirnov, V. V. Simplified technique for identification of the aerobic spore-forming bacteria by phenotype. *Int. J. Syst. Evol. Microbiol.* 2001, *51*, 1361-1371.

[4] Lennox, V. A.; Ackerman, V. P. Biochemical identification of bacteria by replicator methods on agar plates. *Pathology [Phila.]*. 1984, *16*, 434-440.

[5] Kuda, T.; Izawa, Y.; Yoshida, S.; Koyanagi, T.; Takahashi, H.; Kimura, B. Rapid identification of tetragenococcus halophilus and tetragenococcus muriaticus, important species in the production of salted and fermented foods, by matrix-assisted laser desorption ionization-time of flight mass spectrometry [MALDI-TOF MS]. *Food Control*. 2014, *35*, 419-425.

[6] Álvarez-Buylla, A.; Culebras, E.; Picazo, J. J. Identification of acinetobacter species: Is bruker biotyper maldi-tof mass spectrometry a good alternative to molecular techniques? *Infect. Genet. Evol.* 2012, *12*, 345-349.

[7] Garcia, L. S. Clinical Microbiology Procedures Handbook; 3 Lslf.; ASM Press, 2010.

[8] American Society for Microbiology [Washington] Manual of clinical microbiology; 8th ed.; ASM Press: Washington, D.C, 2003.

[9] Wang, D.; Coscoy, L.; Zylberberg, M.; Avila, P. C.; Boushey, H. A.; Ganem, D.; DeRisi, J. L. Microarray-based detection and genotyping of viral pathogens. *Proc. Natl. Acad. Sci. U. S. A.* 2002, *99*, 15687-15692.

[10] Kostic, T.; Butaye, P.; Schrenzel, J. Detection of Highly Dangerous Pathogens: Microarray Methods for the Detection of BSL 3 and BSL 4 Agents; John Wiley & Sons, 2009.

[11] Zwirglmaier, K. Fluorescence *in situ* hybridisation [fish] – the next generation. *FEMS Microbiol. Lett.* 2005, *246*, 151-158.

[12] Rehse, S. J.; Mohaidat, Q. I.; Palchaudhuri, S. Towards the clinical application of laser-induced breakdown spectroscopy for rapid pathogen diagnosis: the effect of mixed cultures and sample dilution on bacterial identification. *Appl. Opt.* 2010, *49*, C27-C35.

[13] Hahn, D. W.; Omenetto, N. Laser-induced breakdown spectroscopy [libs], part i: review of basic diagnostics and plasma-particle interactions: still-challenging issues within the analytical plasma community. *Appl. Spectrosc.* 2010, *64*, 335A-366A.

[14] Morel, S.; Leone, N.; Adam, P.; Amouroux, J. Detection of bacteria by time-resolved laser-induced breakdown spectroscopy. *Appl. Opt.* 2003, *42*, 6184-6191.

[15] Barnett, C.; Bell, C.; Vig, K.; Akpovo, A. C.; Johnson, L.; Pillai, S.; Singh, S. Development of a libs assay for the detection of salmonella enterica serovar typhimurium from food. *Anal. Bioanal. Chem.* 2011, *400*, 3323-3330.

[16] Baudelet, M.; Guyon, L.; Yu, J.; Wolf, J.-P.; Amodeo, T.; Frejafon, E.; Laloi, P. Femtosecond time-resolved laser-induced breakdown spectroscopy for detection and identification of bacteria: a comparison to the nanosecond regime. *J. Appl. Phys.* 2006, *99*, 084701-9.

[17] Rehse, S. J.; Jeyasingham, N.; Diedrich, J.; Palchaudhuri, S. A membrane basis for bacterial identification and discrimination using laser-induced breakdown spectroscopy. *J. Appl. Phys.* 2009, *105*, 102034-13.

[18] Marcos-Martinez, D.; Ayala, J. A.; Izquierdo-Hornillos, R. C.; de Villena, F. J. M.; Caceres, J. O. Identification and discrimination of bacterial strains by laser induced breakdown spectroscopy and neural networks. *Talanta*. 2011, *84*, 730-737.

7.3. PUBLICATION IN SCIENTIFIC JOURNALS

Talanta 121 (2014) 65–70

Contents lists available at ScienceDirect

Talanta

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

Rapid identification and discrimination of bacterial strains by laser induced breakdown spectroscopy and neural networks

S. Manzoor [a], S. Moncayo [a], F. Navarro-Villoslada [a], J.A. Ayala [b], R. Izquierdo-Hornillos [a], F.J. Manuel de Villena [a], J.O. Caceres [a,*]

[a] Departamento de Química Analítica, Facultad de Ciencias Químicas, Universidad Complutense, 28040 Madrid, Spain
[b] Centro de Biología Molecular "Severo Ochoa", CSIC. C/Nicolás Cabrera, 1, Cantoblanco, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 18 October 2013
Received in revised form
16 December 2013
Accepted 24 December 2013
Available online 4 January 2014

Keywords:
Bacterial strain discrimination
Laser induced breakdown spectroscopy (LIBS)
Neural networks (NN)
Bacteria
Hospital acquired infections (HAI)

ABSTRACT

Identification and discrimination of bacterial strains of same species exhibiting resistance to antibiotics using laser induced breakdown spectroscopy (LIBS) and neural networks (NN) algorithm is reported. The method has been applied to identify 40 bacterial strains causing hospital acquired infections (HAI), i.e. *Pseudomonas aeruginosa*, *Escherichia coli*, *Klebsiella pneumoniae*, *Salmonella typhimurium*, *Salmonella pullurum* and *Salmonella salamae*. The strains analyzed included both isolated from clinical samples and constructed in laboratory that differ in mutations as a result of their resistance to one or more antibiotics. Small changes in the atomic composition of the bacterial strains, as a result of their mutations and genetic variations, were detected by the LIBS–NN methodology and led to their identification and classification. This is of utmost importance because solely identification of bacterial species is not sufficient for disease diagnosis and identification of the actual strain is also required. The proposed method was successfully able to discriminate strains of the same bacterial species. The optimized NN models provided reliable bacterial strain identification with an index of spectral correlation higher than 95% for the samples analyzed, showing the potential and effectiveness of the method to address the safety and social-cost HAI-related issue.

1. Introduction

Hospital acquired infections (HAI) have been widely studied in the past 30 years and have been raised to top-priority issue due to the associated economic and social costs [1]. Therefore, many preventive campaigns as well as new protocols have been implemented [2,3]. On average 5–7% of hospitalized patients are affected by HAI, and 1% of such unwanted events result in the patient's death [4]. Bacteria are responsible for 95% of HAI, *Escherichia coli* (18.2%), *Staphylococcus* (18.1%), *Pseudomonas* (6.0%), *Enterococcus* (15.4%), *Klebsiella* (3.7%), *Acinetobacter* (0.8%), and *Salmonella* (2.8%) being the most relevant ones.

An important issue highlighted in recent years has been the increasing emergence of bacteria that are resistant to many antimicrobial therapies, sometimes resulting in multidrug-resistant strains or "super bugs." One of the overriding reasons for this is the widespread indiscriminate use of antibiotics to treat infections [5]. This antibiotic resistance is evolved under the treatment regimens of single or multidrug combinations as a result of the mutations [6].

During the past decades several methods have been proposed to optimize the identification of bacterial strain, which are based on molecular techniques such as fluorescent probes [7], microarray assemblies [8,9] and polymerase chain reaction [10,11]. However, these methodologies present some difficulties and drawbacks such as use of consumables, primer, probes or fluorescently labeled RNA antibodies [12]. Moreover, sometimes the sequences in the database are not accurate or up-to-date and micro-heterogeneity is also found common in 16S rRNA gene sequence within a species [13,14]. The phenotypic similarities between the strains of the same bacterial species restrict their identification using routine diagnostic methods [15]. Although these methods provide a reliable and accurate bacterial identification, special sample treatment methods, the high costs and low speed to perform such analysis limit their use as rapid diagnostic methods in common laboratories in order to provide quick results which leads to an increase in the rates of infectious diseases in clinical settings. Further the direct handling of these potentially pathogenic bacterial samples poses health-associated security risks. At present, clinical safety procedures and cost-related considerations do not allow an easy routine analysis of highly dangerous pathogenic bacterial specimens causing hospital acquired infections. Nevertheless, bacterial identification within the first 24 h of infection allows the use of a more effective and less risky targeted-therapy decreasing unnecessary hospitalization days and costs.

* Corresponding author. Tel.: +34 91 3944322; fax: +34 91 3944329.
E-mail address: jcaceres@ucm.es (J.O. Caceres).

0039-9140/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2013.12.057

Talanta 155 (2016) 101–106

Contents lists available at ScienceDirect

Talanta

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

Laser induced breakdown spectroscopy for the discrimination of *Candida* strains

S. Manzoor [a,1], L. Ugena [a,1], J. Tornero-Lopéz [c], H. Martín [b], M. Molina [b], J.J. Camacho [d], J.O. Cáceres [a,*]

[a] Departament of Analytical Chemistry, Faculty of Chemical Sciences, Complutense University, 28040 Madrid, Spain
[b] Department of Microbiology II. Faculty of Pharmacy, Complutense University, 28040 Madrid, Spain
[c] Hospital "Los Madroños", 28690 Brunete, Madrid, Spain
[d] Departament of Applied Physical Chemistry, Autonoma University of Madrid, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form
11 April 2016
Accepted 14 April 2016
Available online 16 April 2016

Keywords:
Laser Induced Breakdown Spectroscopy
Neural Networks
Candida
Candidiasis

ABSTRACT

The present study reports the evaluation of Laser Induced Breakdown Spectroscopy (LIBS) and Neural Networks (NN) for the discrimination of different strains of various species of *Candida*. This genus of yeast was selected due to its medical relevance as it is commonly found in cases of fungal infection in humans. Twenty one strains belonging to seven species of *Candida* were included in the study. Scanning Electron Microscopy with Energy-Dispersive X-ray Spectroscopy (SEM-EDS) was employed as a complementary technique to provide information about elemental composition of *Candida* cells. The use of LIBS spectra in combination with optimized NN models provided reliable discrimination among the distinct *Candida* strains with a high spectral correlation index for the samples analyzed, without any false positive or false negative. Therefore, this study indicates that LIBS-NN based methodology has the potential to be used as fast fungal identification or even diagnostic method.

1. Introduction

Biomedical applications of Laser Induced Breakdown Spectroscopy (LIBS) have been extensively explored since the last decade. Among these studies the analysis of microorganisms especially bacteria with particular relevance to infectious diseases in humans have made a greater contribution [1–5]. The increased interest in this area has been due to the advantages offered by LIBS in providing a speedy and cost effective analysis as well as to the use of modern chemometric techniques to deal with the large amount of data produced by LIBS. Whereas most of the studies have focused on the identification and discrimination of bacterial samples [2,3,6,7–9], not much has been done towards the analysis of fungi samples by LIBS. Keeping in view the medical importance of fungi, the *Candida genus* has been selected for this study. *Candida* species are ubiquitous fungi and most common human fungal pathogens [10]. *Candida* is an opportunist pathogen but can cause mucocutaneous and disseminated infections [11]. Species of *Candida* can be part of the endogenous microbiota of digestive and urogenital tract [12,13] but have also been found frequently in the patients with cancer [14]. However, they are also found to be a leading cause of nosocomial infections, which are associated with a great deal of morbidity and mortality resulting in increased hospitalization period and treatment costs. Recent studies have reported an increase in these fungal infections [15,16]. In fact, *Candida* is the fourth most commonly recovered organism from blood cultures of hospitalized patients, with an estimated mortality rate of 38–75% [17]. Out of approximately 200 species of the genus *Candida*, about 20 species are implicated in clinical infections. Among them, *Candida albicans*, *Candida glabrata*, *Candida tropicalis*, *Candida parapsilosis and Candida krusei* are found more frequently than others. Some species such as *Candida guillermondii*, *Candida kefyr and Candida dubliniensis*, although are less frequent, also cause serious infections [12,15,16,18].

The taxonomic identification of a fungal sample is conventionally done by methods based on the analysis of macro and microscopic morphology and physiology. A common method relies on the analysis of the ability to develop germ tubes, which is a characteristic feature of *Candida albicans* that allows its differentiation from the non-*albicans* species. However, 5% of *C. albicans* isolates do not produce germ tubes, while some C. *tropicalis* isolates exhibit germ tube formation. Other identification procedures include biochemical-based assays, like carbohydrate fermentation and assimilation, serological and molecular analyses such as ELISA and PCR tests, chromogenic studies and fluorogenic assays [10].

* Corresponding author.
E-mail address: jcaceres@ucm.es (J.O. Cáceres).
[1] These authors have contributed in equal form.

http://dx.doi.org/10.1016/j.talanta.2016.04.030
0039-9140/

Spectrochimica Acta Part B 101 (2014) 21–25

Contents lists available at ScienceDirect

Spectrochimica Acta Part B

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

Technical Note

Discrimination of human bodies from bones and teeth remains by Laser Induced Breakdown Spectroscopy and Neural Networks

S. Moncayo, S. Manzoor, T. Ugidos, F. Navarro-Villoslada, J.O. Caceres *

Departamento de Química Analítica, Facultad de Ciencias Químicas, Universidad Complutense, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 7 February 2014
Accepted 15 July 2014
Available online 23 July 2014

Keywords:
Laser Induced Breakdown Spectroscopy
Neural Networks
Human remains
Bone
Teeth

ABSTRACT

A fast and minimally destructive method based on Laser Induced Breakdown Spectroscopy (LIBS) and Neural Networks (NN) has been developed and applied to the classification and discrimination of human bones and teeth fragments. The methodology can be useful in Disaster Victim Identification (DVI) tasks. The elemental compositions of bone and teeth samples provided enough information to achieve a correct discrimination and reassembling of different human remains. Individuals were classified with spectral correlation higher than 95%, regardless of the type of bone or tooth sample analyzed. No false positive or false negative was observed, demonstrating the high robustness and accuracy of the proposed methodology.

1. Introduction

The proper identification of human bodies is important from the legal and administrative point of view. However, this identification issue is sometimes extremely difficult, particularly in mass disaster situations or genocide crimes [1,2]. Adequate human body identification management goes through strict methodology and protocols. This requires adopting and applying standard operating protocols such as the Interpol Disaster Victim Identification (DVI) Standing Committee guidelines [3]. When the number of remains is high, such as in natural disasters, accidents and mass graves, finding the identity of an individual is required. The necessity to achieve an accurate identification is not only important from the humanitarian approach but is also important in lawsuits. The identification process involves a multidisciplinary group of experts and techniques, i.e. fingerprint analysis, forensic pathology, forensic odontology and DNA analysis. When the bodies are non-decomposed the methods used in DVI include physiognomic data analysis such as scars or marks, personal effects, matching of fingerprints and dentition pattern (provided premortem records are available) [2,4,5]. In many situations, these methods cannot be used due to either extensive putrefaction and destruction of the remains or unavailability of appropriate medical or dental records. Common methods for human identification are not sufficient in approximately 42% of cases, because of decomposition of the body and unavailability of pre-mortem data and therefore DNA identification was requested [6]. However, the unavailability of DNA from relatives, the complications in evidence collection and the degradation and contamination of DNA extracted from bone and teeth samples can hinder the identification process [7]. Moreover, in case of huge number of remains, even identification by DNA analysis becomes difficult, requiring more time and economical resources.

Therefore, a new methodology that provides simple, direct and cost-effective analysis is needed in forensic science. In recent years, Laser Induced Breakdown Spectroscopy (LIBS) has become a powerful analytical tool because of its ability to carry out a rapid qualitative and quantitative analysis of different samples, able to provide real time spectral fingerprint of the elemental composition of the sample [8–10]. The possibility to combine LIBS with chemometric methods allows extracting significant information to classify different samples, such as providing an acceptable classification of commercial pharmaceutical tablets using PCA and SIMCA [11].

Human bones and teeth are the typical remains that can be found in forensic scenes due to their resistance to degradation and, hence, are useful as evidences in anthropology, archeology and forensic science [12,13]. These remains are useful to classify and discriminate individuals in crime scenes, accidents or mass burial sites. Although differences in the elemental composition of these samples between individuals are negligible, some trace elements can be related to diet and environment [7], providing variations in the LIBS spectra. These differences allow the achievement of the discrimination when a chemometric analysis is done.

* Corresponding author.
E-mail address: jcaceres@ucm.es (J.O. Caceres).

http://dx.doi.org/10.1016/j.sab.2014.07.008
0584-8547/

Chemometrics and Intelligent Laboratory Systems 146 (2015) 354–364

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Evaluation of supervised chemometric methods for sample classification by Laser Induced Breakdown Spectroscopy

S. Moncayo, S. Manzoor, F. Navarro-Villoslada, J.O. Caceres *

Department of Analytical Chemistry, Faculty of Chemistry, Complutense University, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 9 February 2015
Received in revised form 14 May 2015
Accepted 10 June 2015
Available online 17 June 2015

Keywords:
Laser Induced Breakdown Spectroscopy
Chemometrics
Supervised methods
Bone remains

ABSTRACT

In this work seven supervised chemometric methods have been evaluated in a real world application for the classification of human bone remains with similar elemental composition based on Laser Induced Breakdown Spectroscopy (LIBS) measurements. Bone samples belonging to five individuals were obtained from a local cemetery, exposed to uncontrolled conditions. LIBS data were processed with different linear and non-linear supervised chemometric approaches. The performance of each chemometric model was assessed by three validation procedures taking into account their sensitivity (internal validation), generalization ability and robustness (independent external validation). The accuracy of each method increased in the following order: 42% for Linear Discriminant Analysis (LDA), 48% for Classification and Regression Tree (CART), 56% for Support Vector Machines (SVM), 58% for Soft Independent Modeling of Class Analogy (SIMCA), 58% for Partial least Squares–Discriminant Analysis (PLS-DA), 66% for Binary Logistic Regression (BLR) and 100% for Artificial Neural Networks (NN). The results showed that NN outperforms in terms of sensitivity, generalization ability and robustness; whereas SIMCA, PLS-DA, LDA, CART, Logistic Regression and SVM did not show significant accuracy to discriminate the bone samples with a high degree of similarity.

1. Introduction

Chemometrics has been established as a mature methodology for converting huge amounts of measurements into significant and useful information in a multitude of disciplines and applications, for instance, signal processing, experimental design, optimization, data mining, multivariate calibration and classification [1–4]. In order to obtain maximum information from the raw data, a variety of statistical and computational algorithms have been applied. The selection of a chemometric method depends on the analytical application and the required information. Regarding classification methods, two possible procedures can be applied. Firstly, unsupervised methods, where no prior classification model is assumed over data collected in a matrix (objects). Secondly, supervised methods, which are defined by two data sets, objects (input) and classes (target). These methods attempt to discover a relationship between objects and classes, referred to as a model, which represents a set of features that define the classification process. The membership of new objects (unknowns for the models) is predicted on the basis of their similarity to a certain class in the model.

The mathematical algorithms used by these methods are based on two different approaches, linear and non-linear. Linear chemometric methods use linear combination of variables to estimate the classification model; whereas non-linear methods apply more complex algorithms increasing the sophistication and the classification capacity of the method.

In Laser Induced Breakdown Spectroscopy (LIBS) the interaction of a high energy focused laser pulse with the sample surface is capable of producing the rupture of chemical bonds and the excitation of atoms and ions, in a process called ablation [5–7]. The optical emission of the plasma contains information about the sample and its spectroscopic study provides useful data for qualitative and quantitative applications [8,9]. LIBS technique has a broad scope due to the possibility of analyzing solids, liquids or gases, being able to acquire quickly a large amount of data. LIBS provides a spectral fingerprint of the elemental composition of the sample, which allows to obtain classification results in real time [8,10].

A wide range of chemometric approaches are available to carry out classification and discrimination processes [11–13]. Nevertheless, the selection of the appropriate method in each case is not straightforward and the raw data must be carefully studied in order to achieve a satisfactory classification result.

An appropriate chemometric method must be able to perform a successful classification when all samples included in the training are tested and correctly classified (high sensitivity), as well as unknown samples of the same class not included in the training step are also correctly classified (high generalization ability). Furthermore, the model must be able to classify a sample that does not belong to any

* Corresponding author. Tel.: +34 91 3944322; fax: +34 91 3944329.
E-mail address: jcaceres@ucm.es (J.O. Caceres).

http://dx.doi.org/10.1016/j.chemolab.2015.06.004
0169-7439/

Spectrochimica Acta Part B 88 (2013) 186–191

Contents lists available at ScienceDirect

Spectrochimica Acta Part B

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

Technical note

Determination of the postmortem interval by Laser Induced Breakdown Spectroscopy using swine skeletal muscles

A. Marín-Roldan, S. Manzoor, S. Moncayo, F. Navarro-Villoslada, R.C. Izquierdo-Hornillos, J.O. Caceres *

Departamento de Química Analítica, Facultad de Ciencias Químicas, Universidad Complutense, Ciudad Universitaria, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 26 November 2012
Accepted 13 July 2013
Available online 20 July 2013

Keywords:
Postmortem interval
Swine
Laser Induced Breakdown Spectroscopy

ABSTRACT

Skin and muscle samples are useful to discriminate individuals as well as their postmortem interval (PMI) in crime scenes and natural or caused disasters. In this study, a simple and fast method based on Laser Induced Breakdown Spectroscopy (LIBS) has been developed to estimate PMI using swine skeletal muscle samples. Environmental conditions (moisture, temperature, fauna, etc.) having strong influence on the PMI determination were considered. Time-dependent changes in the emission intensity ratio for Mg, Na, Hα and K were observed, as a result of the variations in their concentration due to chemical reactions in tissues and were correlated with PMI. This relationship, which has not been reported previously in the forensic literature, offers a simple and potentially valuable means of estimating the PMI.

1. Introduction

The postmortem interval (PMI) refers to the time since death and its estimation is extremely important in criminal, civil and forensic investigations [1]. Postmortem decomposition is reported to begin immediately after death involving the breakdown of soft tissues due to the autolysis and putrefaction processes [2], and often with superimposed insect activity [3]. A death event is followed by a number of complicated chemical reactions which continue to occur within the cells, and cause specific time-dependent changes in metabolism and sub-cellular structures, enabling to determine PMI [1]. Many methods have been employed to accurately determine the PMI, also known as time of death (TOD), but the accuracy of these methods still leaves a high time uncertainty and needs improvements [4]. From an entomological point of view, since insect growth and development are heavily influenced by geographical, topological and environmental conditions, the difficulty in accounting for the conditions at the scene prior to discovery introduces some uncertainty in application. This variability in the environmental conditions made a well-characterized foundation of workable information difficult to attain [5]. Actually, methods of PMI estimation include body temperature measurements and postmortem chemical changes in the body [4]. In addition, some methods have also focused on the degradation of nucleic acids including DNA and RNA [1,6–8] and proteins within different tissues [9,10]. However, these methods require specialized staff and laboratories, and in some cases complicated and costly analysis procedures.

A significant relationship has been established between the ratio of postmortem sodium and potassium concentration variation with PMI in wistar rat serum [11] and human blood [12] samples. However, authors showed that internal and external environmental factors as well as the cause of death significantly affect PMI prediction due to the fast changes of this ratio in the serum and plasma samples (0–12 h) [13]. On the other hand, individual biological variability also limits the usefulness of PMI prediction based on these ratio measurements. Ideally, the concentration of the constituents at any given postmortem interval should display minimal individual biological variation and should be independent of environmental conditions.

Laser Induced Breakdown Spectroscopy (LIBS) has been a subject of research for the past few decades because of its unique features and wide variety of applications in various fields [14,15]. In recent years, LIBS has become a powerful analytical tool because of its ability to carry out a rapid qualitative and quantitative analysis of different samples [16,17]. LIBS analyzes a sample by direct measurement of the atomic emission of the elements from laser-induced plasma generated by the ablation of the sample, providing an immediate spectral fingerprint which is representative of its elemental composition [18]. Moreover, the technique requires little or no sample preparation as opposed to biochemical methods.

The spatial distribution of volumetric energy density generated by laser irradiation drives all pulsed laser ablation processes [19]. In case of laser–tissue interaction, this distribution is controlled by the incident radiant exposure and the optical absorption and scattering properties of the tissue. A detailed description of laser–tissue interaction has been the subject of many reviews discussing general aspects of this process [20–22]. Specific laser ablation processes resulting from nonlinear absorption have also been considered [23,24].

This paper focuses on the development of a simple, fast and direct analytical method for the determination of PMI based on LIBS using muscle tissue samples. Swine muscle tissues were selected due to similar

* Corresponding author.
E-mail address: jcaceres@quim.ucm.es (J.O. Caceres).

0584-8547/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sab.2013.07.008

Hindawi Publishing Corporation
Journal of Analytical Methods in Chemistry
Volume 2016, Article ID 6758281, 7 pages
http://dx.doi.org/10.1155/2016/6758281

Hindawi

Research Article

Identification and Discrimination of Brands of Fuels by Gas Chromatography and Neural Networks Algorithm in Forensic Research

L. Ugena, S. Moncayo, S. Manzoor, D. Rosales, and J. O. Cáceres

Department of Analytical Chemistry, Faculty of Chemical Sciences, Complutense University, 28040 Madrid, Spain

Correspondence should be addressed to J. O. Cáceres; jcaceres@ucm.es

Received 23 March 2016; Accepted 27 April 2016

Academic Editor: Miguel de la Guardia

The detection of adulteration of fuels and its use in criminal scenes like arson has a high interest in forensic investigations. In this work, a method based on gas chromatography (GC) and neural networks (NN) has been developed and applied to the identification and discrimination of brands of fuels such as gasoline and diesel without the necessity to determine the composition of the samples. The study included five main brands of fuels from Spain, collected from fifteen different local petrol stations. The methodology allowed the identification of the gasoline and diesel brands with a high accuracy close to 100%, without any false positives or false negatives. A success rate of three blind samples was obtained as 73.3%, 80%, and 100%, respectively. The results obtained demonstrate the potential of this methodology to help in resolving criminal situations.

1. Introduction

Classification and identification of petroleum fuels are a crucial challenge in the scientific investigation of arson and have high importance in cases of commercial, industrial, and forensic criminal acts. The crime of arson may be defined as a willful and malicious act to setting fire to a property and is considered one of the easiest crimes to commit and yet also one of the hardest to investigate [1, 2]. The liquids commonly used as accelerants in the criminal acts are gasoline, kerosene oil, or diesel due to their easy availability and low price [2, 3]. Regulations established for the analysis of ignitable liquid residues provide a common language to the field of forensic science for describing the characteristics of the different inflammable liquids as well as to classify them [4–6].

The quality control of the fuels is guaranteed by means of establishing some technical specifications that vary in different parts of the world (i.e., EN 228 in Europe, ASTM D48 14 in the USA, JIS K2202 in Japan, and IS 2796 in India) [7]. In particular, commercial or industrial adulteration with lower-price fuels is one of the most difficult situations to detect and is a serious problem in the petrochemical industry because of the similar composition of the adulterant fuels used.

The contents of oxygen and hydrocarbons such as aromatics and olefins vary in different samples that produce various kinds of petroleum derivatives with specific characteristics belonging to a particular refinery [1, 8]. The chemical composition of the crude oil depends on various factors that must be taken into account each time when performing analysis of these samples and collecting reference materials. These factors include the origin (oilfields and country), refining procedures, and the location of the distributors. Apart from these elements, another important factor is the contamination introduced in the fuels due to the residual level of the storage tank [3].

In many situations where fuels are used in criminal acts, their composition may vary significantly subjecting to the fire. Factors like air flow and high temperature affect the composition because the more volatile compounds could evaporate. Moreover, the fuels may get contaminated by biological processes such as bacterial degradation [1]. For these reasons, the analysis of the samples could become more difficult. However, there are sites on the crime scene such as the porous materials where the accelerant liquids stay unaffected and retain their original composition that can be recovered for the analysis [1, 8, 9].

Chapter 11

Lycopene: A Review of Chemical and Biological Activity Related to Beneficial Health Effects

Montaña Cámara*, María de Cortes Sánchez-Mata*, Virginia Fernández-Ruiz*, Rosa María Cámara†, Sadia Manzoor‡ and Jorge O. Caceres‡,1

*Department of Food Science and Nutrition, Faculty of Pharmacy, Complutense University, Madrid, Spain

†Department of Forestry Engineering, ETSI Montes, Polytechnic University of Madrid, Madrid, Spain

‡Department of Analytical Chemistry, Faculty of Chemical Sciences, Complutense University, Madrid, Spain

1Corresponding author: Tel.: +34 913944322, jcaceres@quim.ucm.es

Chapter Outline

All authors have contributed equally in this chapter.

Studies in Natural Products Chemistry, Vol. 40. http://dx.doi.org/10.1016/B978-0-444-59603-1.00011-4