Research Article
Volume 12:3, 2023

Journal of Medical Microbiology & Diagnosis

ISSN: 2161-0703 Open Access

À la carte, Streptococcus pneumoniae capsular typing. MALDI-TOF mass spectrometry and machine-learning algorithms as complementary tools for the determination of PCV13 serotypes and the most prevalent NON PCV13 serotypes according to Argentina's epidemiology

Jonathan Zintgraff[1,3*], Florencia Rocca[2,3], Nahuel Sánchez Eluchans[1], Lucía Irazu[2], Maria Alicia Moscoloni[1], Claudia Lara[1] and Mauricio Santos[1]

[1]Servicio Bacteriología Clínica, Instituto Nacional de Enfermedades Infecciosas (INEI) – Administración Nacional de Laboratorios e Institutos de Salud (ANLIS) "Dr. Carlos G. Malbrán", Buenos Aires, Argentina
[2]Instituto Nacional de Enfermedades Infecciosas (INEI) – Administración Nacional de Laboratorios e Institutos de Salud (ANLIS) "Dr. Carlos G. Malbrán", Buenos Aires, Argentina
[3]Red Nacional de Espectrometría de Masas aplicada a la Microbiología Clínica (ReNaEM Argentina), Argentina

Abstract

Laboratory surveillance of *Streptococcus pneumoniae* serotypes is crucial for the successful implementation of vaccines to prevent invasive pneumococcal diseases. The reference method of serotyping is the Quellung reaction, which is labor-intensive and expensive. In the last few years, the introduction of MALDI-TOF MS into the microbiology laboratory has been revolutionary. In brief, this new technology compares protein profiles by generating spectra based on the m/z ratio. We evaluated the performance of MALDI-TOF MS for typing serotypes of *S. pneumoniae* isolates included in the PCV13 vaccine using a machine learning approach. We challenged our classification algorithms in "real time" with a total of new 100 isolates of *S. pneumoniae* from Argentinian nationwide surveillance. Our best approach could correctly identify the isolates with a sensitivity of 58.33% ([95%CI 40.7-71.7]); specificity of 81.48% ([95%CI 53.6-79.7]); accuracy of 63.0% ([95%CI 61.9-93.7]); PPV of 80.77% ([95%CI 64.5-90.6]) and NPV of 59.46% ([95%CI 48.9-69.2]). Furthermore, this approach allowed us to optimize the use of the antiserum used for capsular typing by 10.2% compared to the traditional "blind" typing scheme. In this work, it was possible to demonstrate that the combination of MALDI-TOF mass spectrometry and multivariate analysis allows the development of new strategies for the identification and characterization of Spn isolates of clinical importance.

Keywords: *Streptococcus pneumoniae* • MALDI TOF • Mass spectrometry • Conjugates vaccines • Capsular typing • Machine learning

Introduction

Streptococcus pneumoniae (Spn) is a human pathogenic microorganism [1], responsible for a wide spectrum of infectious diseases and invasive processes that constitute an important cause of morbidity and mortality in the world. Infectious diseases occur when the bacteria gain access to generally sterile areas of the respiratory tract, thus producing the dissemination, colonization and invasion of pneumococcus, mainly the middle ear (producing otitis media), the lungs (pneumonia), the bloodstream (bacteremia) or the central nervous system (meningitis). The bacteremia that is related to higher mortality and morbidity is generally caused by complications of pneumonia. Spread between different individuals occurs by direct contact with secretions from colonized individuals [2].

According to estimates by the World Health Organization (WHO), sepsis, neonatal and community pneumonia accounted for 15% of annual deaths in children under 5 years of age in 2015, with *S. pneumoniae* being the most common cause of pneumonia in both developed and developing countries [3-5]. The incidence of invasive pneumococcal disease (IPD) varies geographically, and is highly related to poverty [3].

The antigenic diversity of the capsular polysaccharide allows pneumococcus to be classified into different serotypes. The reference technique for serotyping was described by Neufeld in 1902 and is called the Quellung Reaction [6-8]. Briefly, this technique utilizes a microscope and specific pneumococcal antisera and is commonly used in reference and research laboratories worldwide. This method uses a chessboard system, in which the pneumococcus is sequentially tested with antisera pools until a positive reaction is observed. Each pool contains different mixtures of antisera against pneumococcal serotypes. Once a pool is established, the individual type and group antisera that are included in the reactive pool are tested individually in sequence in order to determinate the final serotype.

Contemporary studies led to the development of a system that allowed serogroups to be distinguished from serotypes: Serotype: it is defined as an isolate that produces a polysaccharide with unique chemical and immunological properties. Serogroup: it is defined as a group of serotypes that share several immunological properties or antigenic determinants (cross-reactions).

Currently, 100 immunologically different capsular serotypes have been

***Address for Correspondence:** Jonathan Zintgraff, Servicio Bacteriología Clínica, Instituto Nacional de Enfermedades Infecciosas (INEI) – Administración Nacional de Laboratorios e Institutos de Salud (ANLIS) "Dr. Carlos G. Malbrán", Buenos Aires, Argentina, Tel- 01159234601, E-mail: jczintgraff@gmail.com*

Received: 01 April, 2023, Manuscript No. jmmd-23-95504; **Editor Assigned:** 03 April, 2023, PreQC No. P-95504; **Reviewed:** 17 April, 2023, QC No. Q-95504; **Revised:** 22 May , 2023, Manuscript No. R-95504; **Published:** 29 May, 2023, DOI: 10.37421/ 2161-0703.2023.12.403

described [9]. The importance in determining the pneumococcal capsular serotype is fundamentally based on the fact that its distribution is related to a series of factors, such as age, sex, clinical picture, geographical area, antibiotic sensitivity and many other epidemiological data that allow correlate with the prevalence of certain serotypes. On the other hand, given that it is a vaccine-preventable disease and that the vaccines developed and implemented are directed at specific capsular serotypes, the determination of circulating serotypes is extremely important.

Pneumococcal vaccines provide serotype-specific protection, it is important that vaccines prevent disease caused by the most clinically relevant serotypes. Therefore, vaccines provide the greatest impetus for recognizing capsular diversity and serotype epidemiology. Pneumococcal vaccines are an important public health tool and have undergone dramatic changes in recent years. Therefore, there are many excellent overviews of pneumococcal vaccines [10-12]. In Argentina, in 2012,the pneumococcal conjugate vaccine, PCV13 (which includes purified capsular polysaccharide for serotypes 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 19A, 19F, 18C, and 23F) was included in the National Immunization Program in a 2+1 schedule (2, 4, 12 months) [13].

The impact of PCV13 introduction on serotype distribution may vary with time and location, as the epidemiology of serotypes is quite variable both geographically and temporally. In general, conjugate vaccines have been effective in preventing IPD and carriage while offering variable immunity against cross-reactive serotypes. Nevertheless, several studies have shown that the introduction of PCVs has resulted in changes in epidemiology of pneumococcal diseases and can drive an increase in the frequency of preexisting resistant variants of non-vaccine serotypes due to the removal of competition from vaccine serotypes [14-20].

MALDI-TOF MS has acquired great importance in the identification of different pathogens, such as bacteria and yeast for example [21-23]. The technique analyzes protein spectra that contain peaks with a determinable mass-charge ratio (m/z). The potential of this methodology combined with machine learning algorithms for the detection of profiles in a wide variety of samples and its use as a screening technique is expanding, due to its low-cost and high performance [24]. Although the initial investment in the purchase of a mass spectrometer is relatively high, the cost for the identification of a single isolate is lower than previously used biochemical or molecular techniques [25,26]. Among the advantages, the simplified workflow is one of the key points for the rapid acceptance of MS in laboratories and by the broad range of applications that it could be extended.

In this work, we proposed to explore the potential of MALDI-TOF MS technology as a complementary and screening tool for Spn capsular typing by developing classification models based on MALDI-TOF MS in order to discriminate Spn strains from PCV13 and NON PCV13 isolates. To achieve these two objectives, we created a spectral database with isolates of serotypes included in the PCV13 vaccine and, according to local epidemiology, the first 10 most prevalent NON PCV13 serotypes; we Implemented unsupervised models from MALDI-TOF MS spectra to establish the basis for designing predictive classification models. Then we developed supervised classification models based on MALDI-TOF MS spectra that allowed to predict Spn isolates depending on whether they belong to the PCV13 or NON PCV13 class and finally we validated the models developed in the previous point with an independent set of samples. The final goal of this work was that applying this approach as a screening tool for capsular typing and use the reference methodology in a more focused way and not blindly, which would significantly reduce the costs of this technique.

Materials and Methods

Bacterial isolates

First, a total of 23 isolates were selected, of which 13 corresponded to the so-called PCV13 group (serotypes included in the PCV13 vaccine) and 10 NON PCV13 group. All of them belonging to the strain collection at the National Reference Laboratory for Meningitis and Bacterial Acute Respiratory Infections (NRL). For the 10 NON PCV13 isolates, we selected the most prevalent serotypes, according to Argentina's epidemiology, otherwise this approach could not be possible due to the many serotypes of Spn.

Prior to freezing, all isolates were identified using the optochin sensitivity and bile solubility tests. Pneumococci were serotyped using the Quellung reaction with sera provided by the Statens Serum Institute (Copenhagen, Denmark); isolates were stored at -70°C in Trypticase Soya (TS) broth with 15%–20% glycerol. In this way, the first 23 selected strains constituted the called "training set". The information on the samples used can be found in Table 1.

Culture conditions

Starting from the isolates preserved at -70°C, sheep blood agar plates (5%) were inoculated with a small aliquot of them. After 24-48 hours, replication was carried out until growth of the order of 104 CFU was obtained.

Sample preparation and MALDI-TOF MS spectra acquisition

Mass spectra acquisition was performed using a Microflex LT mass spectrometer (Bruker Daltonics, Germany) and the Flex Control software (version 3.4) with default parameter settings. Evaluation of the mass spectra was carried out using the Flex Analysis v3.4 software (BrukerDaltonics, Bremen, Germany) and MALDI Biotyper 3.1 software and library (version 9.0, Bruker Daltonics Germany). First, we generated a dataset containing 69 mass spectra from 23 biological samples with 3 technical replicates, constituted the as mentioned before the training set. On the other hand, a second dataset with 200 spectra from 100 biological samples with 2 technical replicates were obtained constituting the test set.

Only one isolate of each serotype was included in the training set due to a lack of isolates for some serotypes (i.e., serotype 5, 18C, 13, 7C, 14, 23F, 11A, and 15B). Some of these serotypes practically disappeared from circulation, meaning that, they were not recovered from any invasive pneumococcal disease in the last years, making it very difficult to obtain new fresh isolates. Regardless of the fact, that we had more isolates of other serotypes, we consider that one isolate per serotype would be representative and homogeneous than including several isolates of some serotypes and few of others, at least for this first approach. Seeding was performed by the direct method, routinely used in clinical laboratories as indicated by the manufacturer [27,28].

One colony of the sample was placed in each well of the steel plate (MSP 96; BrukerDaltonics), then allowed to dry for a few minutes at room temperature and covered with 1 ul of the HCCA matrix (α-cyano-acid-4-hydroxycinnamic acid solution diluted in 500 μL of acetonitrile, 250 μL of 10% trifluoroacetic acid and 250 μL of HPLC grade water). This matrix enables highly sensitive measures of peptides and proteins from 0.7 to 20 kDa and nucleotides. After dehydration of the sample-matrix mixture, the plate was introduced into the Micro Flex LT instrument and vacuum conditions were generated.

Table 1. Training set: Isolates used for the creation of the machine learning models.

Serotype	ID	Year	PCV13/NON PCV13
1	20380	2016	PCV13
3	20117	2016	PCV13
4	19202	2015	PCV13
5	21128	2017	PCV13
6A	20342	2016	PCV13
6B	20372	2016	PCV13
7F	23112	2019	PCV13
9V	21329	2017	PCV13
14	20437	2016	PCV13
18C	23100	2019	PCV13
19A	23049	2019	PCV13
19F	21287	2017	PCV13
23F	20018	2016	PCV13
7C	23157	2019	NON PCV13
8	23201	2019	NON PCV13
11A	22167	2018	NON PCV13
12F	23165	2019	NON PCV13
13	23210	2019	NON PCV13
15B	23148	2019	NON PCV13
16F	23115	2019	NON PCV13
22F	23143	2019	NON PCV13
23B	23189	2019	NON PCV13
24F	23207	2019	NON PCV13

Mass spectra were recorded in the spectral region from 2000 to 20,000 Da (in linear positive ionization mode). The operation of the equipment was controlled by Flex Control v3.4 software. Each spectrum was a sum of 240 laser shots collected in increments of 40. Their recording was carried out at 40% of the maximum laser energy. The platform was pre-calibrated according to the manufacturer's instructions using the Bruker Daltonics bacterial test standard (BrukerDaltonics, Bremen, Germany). The quality of the recorded spectra was evaluated using Flex Analysis software, in this step important spectrum features for the mass spectral quality were defined such as the number peaks detected, intensity of the peaks and the reproducibility of peaks between technical replicates. At the same time null spectra, were removed to perform the analysis. All recorded spectra were stored in files (mzXML) for pre-processing and further analysis.

Data analysis

All isolates were identified using the MALDI Biotyper RTC software and compared to the Bruker Biotyper database (reference library version 9.0). According to the manufacturer's recommendations, the identification was considered reliable at species level when the score value was greater than 2.0 and at the genus level when the score value was between 1.7 and 1.99; and it was considered 'No Identification' when the value of the score was equal or lower than 1.69 [29].

To perform data analysis, ClinPro Tools software (version 3.0, Bruker Daltonik GmbH, Bremen, Germany) and Flex Analysis v3.4 software were used. All spectra were exported as mzXML files using CompasXport CXP3.0.5 according to standard Bruker setup.

Data pre-processing

Regardless ClinPro Tools, the following data pre-processing steps were carried out according to the literature [30,31]:

Baseline correction: A polynomial function fitted to a subsection of the baseline was used. For this, the “top-hat” option was selected, whose length was set at 10% of the minimum width of the baseline.

Spectra recalibration: Recalibration was performed at 1,000 ppm maximum peak offset and 30% coincidence with the calibrating peaks, excluding null or out-of-range spectra.

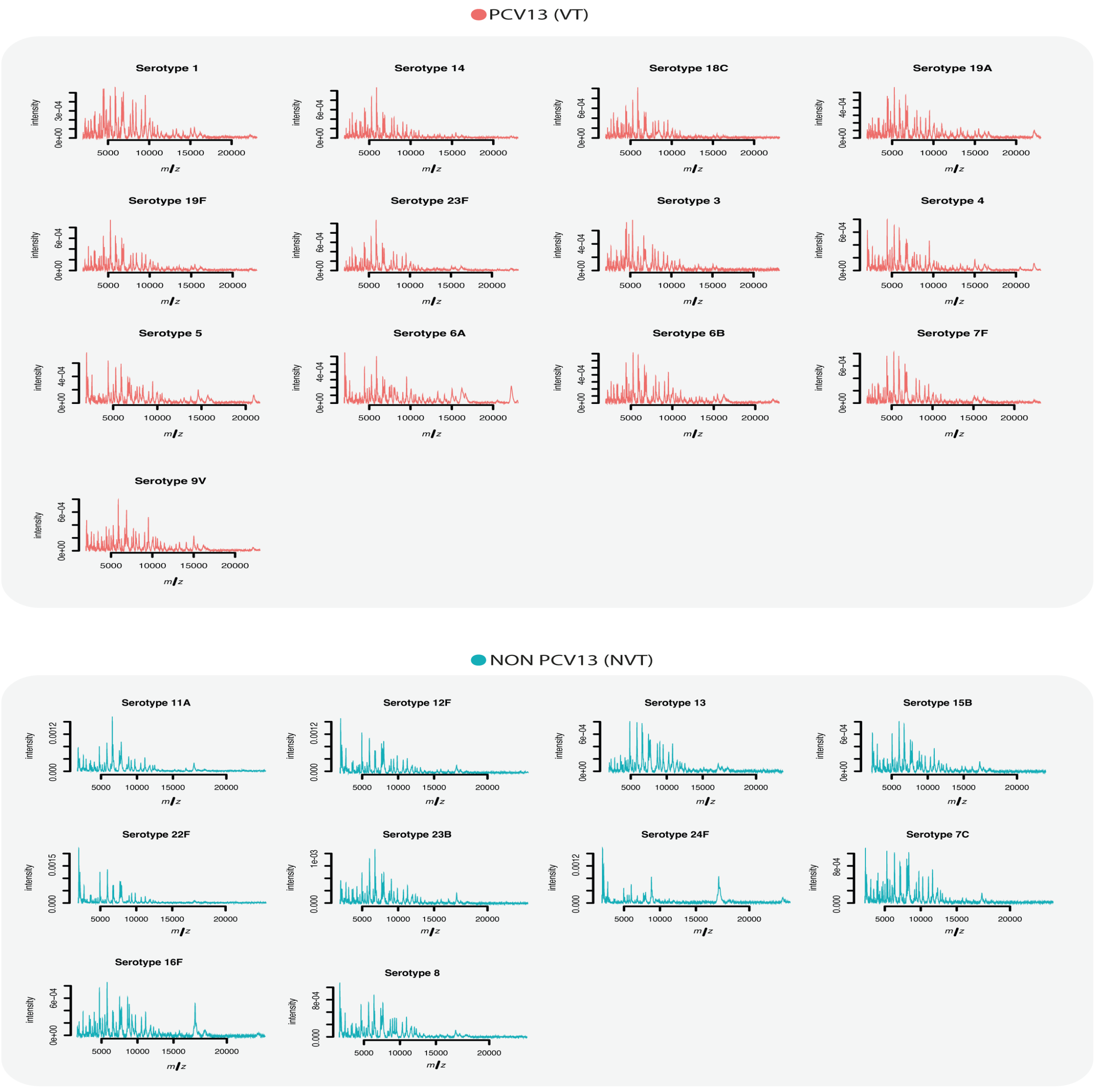


Figure 1. Individual average spectra of all isolates included in the training set.

Smoothing: With this function the noise level generated by the matrix and the components external to the sample was reduced in order to amplify the information contained in the spectra. Smoothing was performed in one cycle, on the order of 7 Da wide (Figure 1).

Unsupervised analysis models

Principal Component Analysis (PCA): In order to evaluate the possible distributions or "clusters" on the isolates of both classes (PCV13 and NON PCV13), a PCA analysis was performed on the registered spectra. For this, the ClinPro Tools software (version 3.0, BrukerDaltonikGmbH, Bremen, Germany) was used. For this analysis, 10 principal components (PC) were defined for the entire spectral range (2000-20000Da).

Hierarchical Cluster Analysis (HCA): From the generated spectra, a dendrogram was constructed based on the similarities of the peak intensity and mass signals. To carry out this point, the ClinPro Tools software (version 3.0, Bruker DaltonikGmbH, Bremen, Germany) was used.

Supervised analysis models

Peak selection

ClinPro Tools: To select the characteristic peaks of the two classes (PCV13 and NON PCV13) the following statistical tests were used: t-test/analysis of variance ANOVA (PTTA), Wilcoxon or Kruskal–Wallis test (W / KW) and Anderson test – Darling (AD). A P value of 0.05 was established as the cut-off point [27]:

-if p is <0.05 in the AD test, a characteristic peak is selected if the corresponding value of P in the W/KW test is also <0.05.

-if p is 0.05 in the AD test, then a characteristic peak is selected if the corresponding p value in ANOVA is also <0.05 [32].

Peaks were identified by class comparison using the "Peak Statistic Table" function in ClinPro Tools followed by manual confirmation, using Flex Analysis. The discriminative power for each peak was further described by receptor operating characteristic (ROC) area under the curve (AUC) analysis. The ROC curve provides a graphical description of the specificity and sensitivity of a test, and in this case an assessment of the discrimination quality of a peak. An AUC value of 0 indicates that the considered peak does not discriminate, while an AUC of 1 indicates that the considered peak discriminates.

Then, three supervised classification models were calibrated using ClinPro Tools software: Genetic Algorithm (GA), Supervised Neural Networks (SNN) and Quick Classifier (QC). In turn, the QC algorithm was calculated using the ANOVA variance test (QC-ANOVA), the Kruskal-Wallis test (QC-WKW) and the Anderson-Wallis test (QC-DaV); All provided by ClinPro Tools software. For all cases, the selection of the maximum number of best peaks was set to 100, and the maximum number of generations was set to 10. For the GA model, which selects the best peaks for classification, the k-NN (nearest neighbor) algorithm set to 3 was used for binary classification. The developed calibration models were validated by cross validation in 10 iterations, leaving out 20% of the samples in each cross.

Independent classification of supervised classification models

To assess the robustness of the developed models, an independent set of isolates was selected for classification. A total of 100 isolates corresponding to the national surveillance of the 2020-2021 periods were used (Table 2).

Capsular typing

Gold standard of pneumococcal serotyping (the Neufeld-Quellung reaction) was performed using pool, group, type and factor specific commercial antisera produced by the Statens Serum Institute (Copenhagen, Denmark). Capsular types were assigned in accordance with the Danish nomenclature system [8].

Briefly, the serotypes/groups of all isolates were identified first with a Pneumotest-Latex kit (comprising 14 latex reagents pools (A to I and P to T). Each pool contains different mixtures of antisera against pneumococcal serotypes. The kit could identify serotypes at the type/group level by all 14 pools using the “chessboard” identification system. Once a pool is stablished, the individual type and group antisera that are included in the reactive pool are tested individually in sequence in order to determinate the final serotype Usually, we used a predeterminate order to start serotyping, involving these sequential schemes: first pool P, Q, R, S, T, then A, B, C, D, E, F, G, H and finally I. This order was pre-established due to the national circulating serotypes included in each pool. However, this order is still a blind use of the technique.

“Real time classification”

The strategy applied was based on the use of classification models in “real time”. Each *Streptococcus pneumoniae* isolate sent to the National Reference Laboratory for serotyping was previously challenged by the classifying models, and according to its prediction, either PCV13 or NON PCV13, the reference technique was carried out for capsular typing.

Figure 2 Workflow used for serotyping unknown isolates. In the first place, all isolate was processed by the conventional identification methods and MALDI-TOF MS (MicroFlex platform), if the results were in agreement, then, the spectrum was loaded into the ClinPro Tools software, in which, depending on the prediction of the models, we performed the personalized Quellung reaction according to the national epidemiology of circulating serotypes.

Statistical analysis

For statistical analysis, GraphPad Prism 9.0 software (GraphPad Software, Inc., San Diego, California, USA) was used. For the comparison of the classifying models with respect to the gold standard, contingency tables were made for each model (using binary variables), calculating sensitivity, specificity, precision, positive predictive value and negative predictive value.

Results

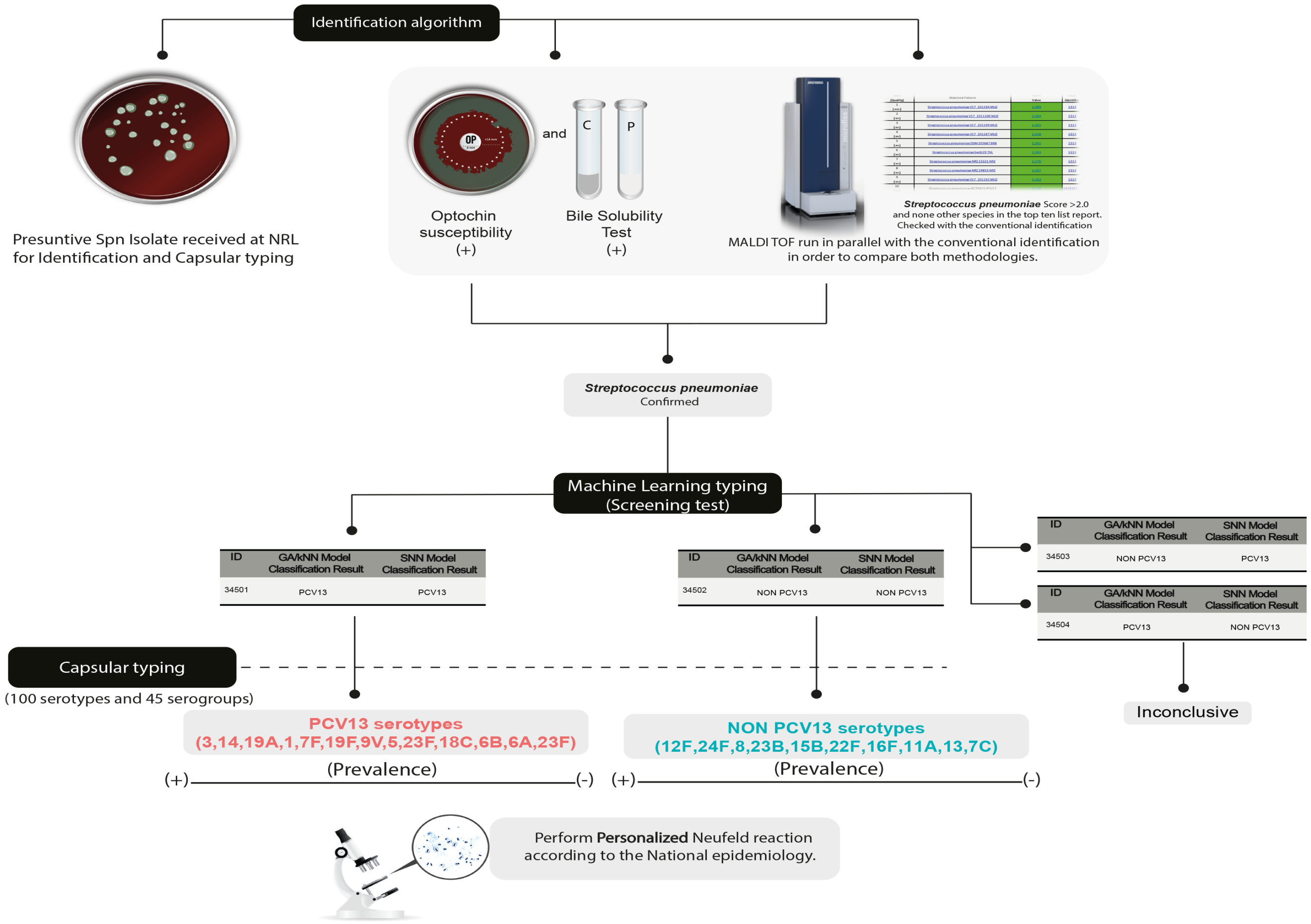
Unsupervised analysis

These analyses were performed only for the training set. Principal Component Analysis (PCA) was performed for the 23 isolates used as training. (Figure 3) shows the graph of the accumulated variance for the new components. In this case, 10 principal components (PCs) were defined. In this way it was possible to reduce all the information contained in the MALDI-TOF MS spectra in a few new variables. This means that if the spectrum is thought of as a multivariate system, each peak or signal as a function of the charge mass (m/z) represents a

Table 2. The best 10 peaks calculated by ClinPro tools software.

Index	Mass	Dave	PTTA	PWKW	PAD
78	4703.69	0.57	0.00299	0.00302	0.807
170	10785.95	0.15	0.00299	0.00358	0.423
81	4794.85	0.57	0.00693	0.00302	0.194
180	13561.74	0.2	0.00739	0.00358	0.0895
116	6253.35	0.3	0.0121	0.0188	0.635
117	6270.43	2.56	0.0127	0.0211	0.121
43	3173.19	1.02	0.0168	0.0188	0.132
118	6436.73	0.65	0.0176	0.0188	0.875
171	10991.5	0.55	0.0176	0.0188	0.817
39	2983.78	1.15	0.0176	0.0188	0.192

DAve=Difference between the maximum and minimum intensity of the average peak of all classes; PTTA=p-value obtained through the t-test, range 0-1; where 0: good and 1: bad; VPWKW=p value obtained using the Wilcoxon / Kruskal-Wallis test; range 0-1; where 0: good and 1: bad; PAD=p-value obtained by the Anderson-Darling test: range 0-1; 0: non-normal distribution, 1: normal distribution


Identification algorithm
Presuntive Spn Isolate received at NRL for Identification and Capsular typing
OP
Optochin susceptibility (+)
and
C
P
Bile Solubility Test (+)
Streptococcus pneumoniae Score >2.0 and none other species in the top ten list report. Checked with the conventional identification
MALDI TOF run in parallel with the conventional identification in order to compare both methodologies.
Streptococcus pneumoniae Confirmed
Machine Learning typing (Screening test)
Capsular typing
(100 serotypes and 45 serogroups)
PCV13 serotypes (3,14,19A,1,7F,19F,9V,5,23F,18C,6B,6A,23F)
(+) (Prevalence) (-)
NON PCV13 serotypes (12F,24F,8,23B,15B,22F,16F,11A,13,7C)
(+) (Prevalence) (-)
Inconclusive
Perform **Personalized** Neufeld reaction according to the National epidemiology.

Rank (Quality)	Matched Pattern	Score Value	NCBI Identifier
1 (+++)	Streptococcus pneumoniae V17_201194 MUZ	2,389	1313
2 (++)	Streptococcus pneumoniae V17_2011100 MUZ	2,284	1313
3 (++)	Streptococcus pneumoniae V17_201199 MUZ	2,265	1313
4 (++)	Streptococcus pneumoniae V17_201197 MUZ	2,218	1313
5 (++)	Streptococcus pneumoniae DSM 20566T BRB	2,202	1313
6 (++)	Streptococcus pneumoniae bes5t29 THL	2,183	1313
7 (++)	Streptococcus pneumoniae NRZ 23321 NRZ	2,176	1313
8 (++)	Streptococcus pneumoniae NRZ 29853 NRZ	2,167	1313
9 (++)	Streptococcus pneumoniae V17_201192 MUZ	2,163	1313
10	Streptococcus pneumoniae BC20415-PCV13	2,139	1419391

ID	GA/kNN Model Classification Result	SNN Model Classification Result
34501	PCV13	PCV13

ID	GA/kNN Model Classification Result	SNN Model Classification Result
34502	NON PCV13	NON PCV13

ID	GA/kNN Model Classification Result	SNN Model Classification Result
34503	NON PCV13	PCV13

ID	GA/kNN Model Classification Result	SNN Model Classification Result
34504	PCV13	NON PCV13

Figure 2. Workflow used for serotyping unknown isolates. In the first place, all isolate was processed by the conventional identification methods and MALDI-TOF MS (MicroFlex platform), if the results were in agreement, then, the spectrum was loaded into the ClinPro Tools software, in which, depending on the prediction of the models, we performed the personalized Quellung reaction according to the national epidemiology of circulating serotypes.

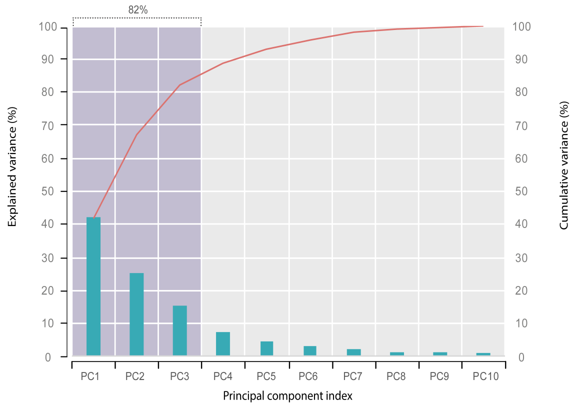


Figure 3. Variance attributable to each principal component generated.

variable, then the PC's represent new variables that contain all the information of the multivariate system (spectrum). This allows us to graphically represent all the spectra together in three and two dimensions on the Score plot.

Figures 4 shows the score plots for the PC's. Clusters appeared as two slightly overlapping groups. Specifically, cluster 1 achieved 100% (12/12) of homogeneity for spectra corresponding to PCV13 serotypes and 92.3% (12/13) of PCV13 spectra fell in this cluster, while for cluster 2, a 90% (10/11) homogeneity was achieved for spectra corresponding to NON PCV13 serotypes group and 100% (10/10) of this group spectra fell in this cluster. This result shows that both groups may have distinctive protein signatures that allow their discrimination.

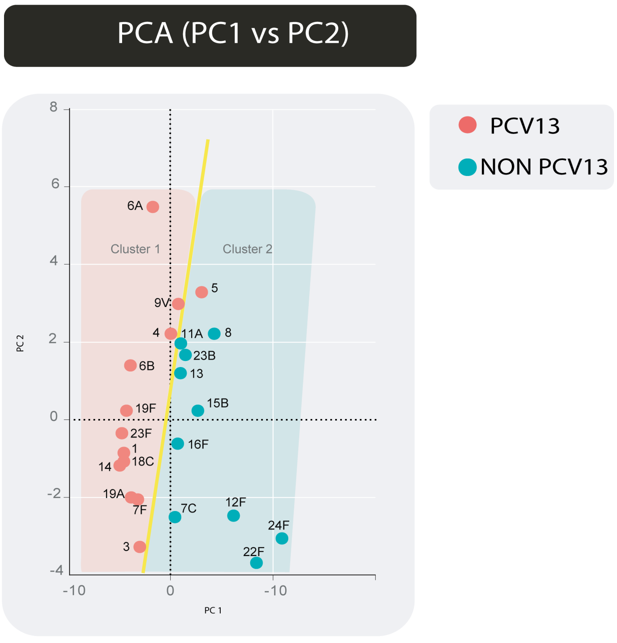


Figure 4. Score plot. Dimensional image of PCA showing the distinction between the 23 isolates; 13 PCV13 isolates (pink) and 10 NON PCV13 isolates (green).

At the same time, an unsupervised hierarchical clustering analysis (HCA) was carried out, from which the dendrogram that can be seen in Figure 5. The horizontal axis represents the distance calculated in the clustering, which is shown in relative units, corresponding to the similarity of the MALDI-TOF MS spectra.

Figure 5, shows the visualization of the respective relationship between the isolates realized with ClinProo Tools. Based on a distance approximate above of 1.1, two clusters were present. Cluster 1 achieved 90% (10/11) of homogeneity for spectra corresponding to PCV13 serotypes and 77% (10/13) of PCV13 spectra fell in this cluster. Only one isolate of NON PCV13, serotype 7C, in miss included in this cluster. While for cluster 2 a 75% (9/12) homogeneity was achieved for spectra corresponding to NON PCV13 serotypes group and 90% (9/10) of this group spectra fell in this cluster. As can be observed, three PCV 13 isolates were miss included in this cluster.

Supervised analysis models

Subsequently, with the additional information of each isolate to define each class (PCV13 / NON PCV13), the supervised multivariate analysis was performed. To recognize the spectral patterns of each class, all the recorded spectra were imported into the Clin ProTools software. Data were entered into two groups (PCV13 and NON PCV13) according to the results of previously obtained Quellung serotyping.

Figure 6 shows the two-dimensional distribution graph of all the spectra of each class based on the two best peaks obtained for their classification; which were 170; 10786 Da and 78;4703 Da. The peak number and m/z values of these are shown on the x and y axes, while the ellipses represent the 95% confidence interval. On the other hand, Figure 6 shows the ROC curves of the two selected peaks. The area under the curve (AUC) represents the discriminatory potential of each biomarker peak. The best 10 peaks calculated by the software, with their statistical values, are summarized in Table 2.

DAve=Difference between the maximum and minimum intensity of the average peak of all classes;

PTTA= p-value obtained through the t-test, range 0-1; where 0: good and 1: bad

VPWKW=p value obtained using the Wilcoxon / Kruskal-Wallis test; range 0-1; where 0: good and 1: bad

PAD= p-value obtained by the Anderson-Darling test: range 0-1; 0: non-normal distribution, 1: normal distribution

Classification models

A total of 5 algorithms were calculated: GA/kNN, SNN, QC-DAv, QC-Anova and QC-WKW, the results of the different parameters of each algorithm are summarized in Table 3.

Based on the results obtained, it was decided to use only GA/kNN and SNN since they presented the highest values of theoretical recognition.

For the real time classification of the selected models, 100 isolates were used of which the serotype was unknown. The results of the performance of the predictive models with respect to the gold standard are summarized in Tables 4 and 5 and (Figure 7).

Table 5 shows the statistical parameters of the GA/kNN plus SNN algorithms

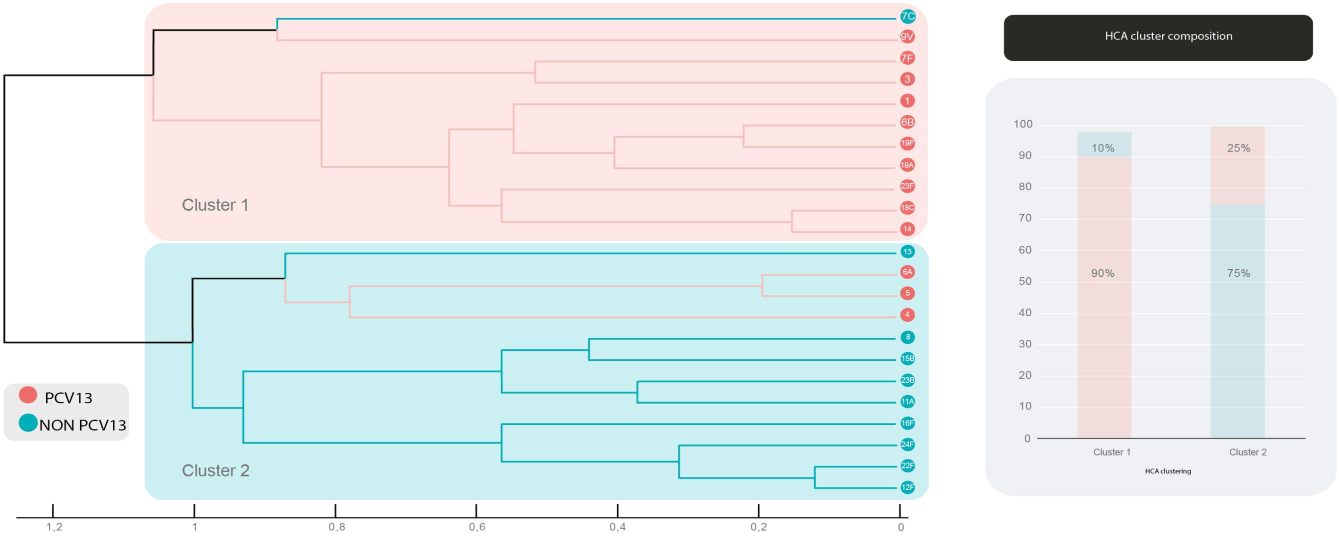


Figure 5. Dendrogram corresponding to unsupervised hierarchical clustering.

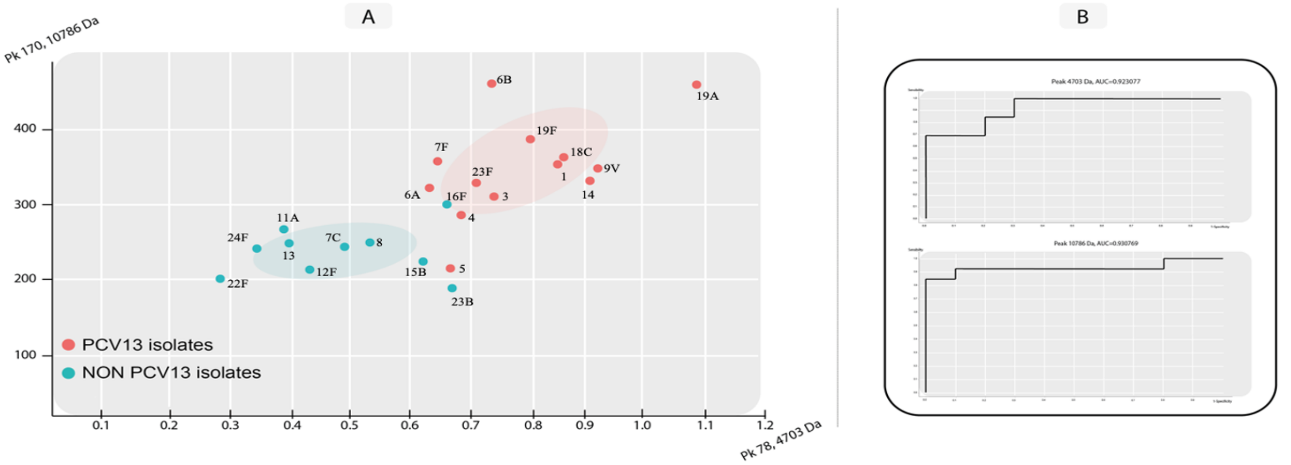


Figure 6. (A) Two-dimensional distribution of the characteristic peaks of each generated class and **(B)** ROC curves with their corresponding AUC values of those peaks.

Table 3. Parameters of each calculated model.[d]

Classification Algorithm	Cross Validation[a] (%)			Recongnition Capability[b] (%)			Peaks (m/z) used in the model					
	Overall[c]	PCV13	NON PCV13	Overall	PCV13	NON PCV13						
Genetic Algorithm (GA/kNN)	82.00%	88.00%	76.00%	95.00%	100.00%	90.00%	4703.69	10785.95	5003.4	7667.2	16025.6	
Supervised Neural Network (SNN)	81.90%	84.60%	79.20%	100.00%	100.00%	100.00%	4703.69	7054.35	7527.1	4794.9	9978.26	10785.95
Quick Classifier (DaV)	67.30%	84.60%	50.00%	74.60%	69.20%	80.00%	2035.26					
Quick Classifier (WKW)	77.90%	80.80%	75.00%	96.10%	92.30%	100%	4703.69	4794.85				
Quick Classifier (ANOVA)	76.30%	69.20%	83.30%	96.10%	92.30%	100%	4703.69	4794.85	10786			

[a]Cross Validation is a statistical measure of reliability for the calculated model and the normalized value of the relative ability of prediction;
[b]Recognition Capability is a measure to explain the capability of calculated model and is calculated as the relative number of data correctly classified by the models. It is equal to sensitivity;
[c]Overall is the value obtained by averaging Cross Validation and Recognition Capability;
[d]The sort model is "p value tta"

Table 4. Results of the parameters evaluated for the selected algorithms.

Algorithms	Sens	Spec	Acc	PPV	NPV
	(CI95%)	(CI95%)	(CI95%)	(CI95%)	(CI95%)
GA / kNN	57.45	67,92	63	61.36	64.29
	(42.1-71.7)	(53.6-79.7)	(52.7-72.4)	(50.0-71.1)	(55.17-72.7)
SNN	55.32	64,15	60	57.78	62.82
	(40.1-69.8)	(49.-76.8)	(49.7-69.6)	(46.7-68.0)	(52.3-70.2)

Sens=sensitivity; Spec=Specificity; Acc=Accuracy; PPV=Positive Predictive Value; NPV=Negative Predictive Value

Table 5. Results of the parameters evaluated considering only the concordances between both models. (n=63).

Algorithm	Sens (CI95%)	Spec (CI95%)	Acc (CI95%)	PPV (CI95%)	NPV (CI95%)
GA / kNN + SNN	58.33 (40.7-71.7)	81.48 (53.6-79.7)	63.00 (61.9-93.7)	80.77 (64.5-90.6)	59.46 (48.9-69.2)

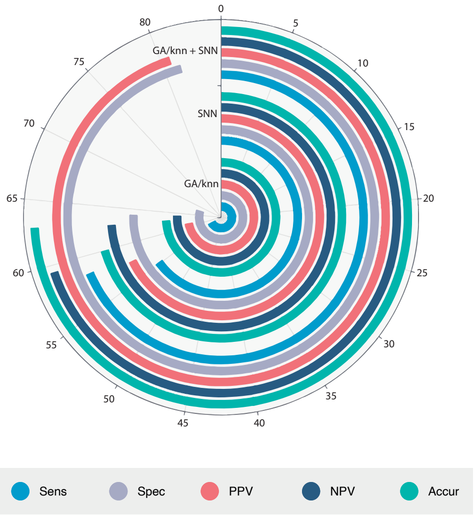


Figure 7. Radial representation of the different parameters evaluated.

with respect to the gold standard, considering only the isolates that showed concordance between both classifying models (63%). It could be seen that 37% of the isolates could not be categorized, resulting inconclusive by this approach.

Discussion

Invasive disease caused by *Streptococcus pneumoniae* (pneumococcus) is a serious infection. It can produce a wide spectrum of clinical manifestations including sepsis, meningitis, bacteremic pneumonia, arthritis, osteomyelitis, cellulitis, and endocarditis.

The people with the highest risk of suffering from this pathology are those under 2 years of age, the elderly and people with immune disorders or certain respiratory, cardiac, renal pathologies, among others.

Pneumococcal vaccination is intended to reduce the incidence, complications, sequelae, and mortality from pneumonia and invasive pneumococcal disease. Two types of vaccines are available: a polysaccharide vaccine against 23 serotypes (PPSV23) and a conjugate vaccine against 13 serotypes (PCV13) [33].

Currently, serotype identification is performed by the Quellung technique. This technique requires expensive reagents, is very laborious, requires specialized staff and involves long times until the result is obtained. In our country, the determination of the capsular type is performed only in the NRL and, due to this; it is difficult to obtain the results in real time. For all the above reasons, the search for reliable, fast and cheap alternative methods for Spn serotyping is of great interest in the field of public health.

In this sense, in recent years MALDI-TOF mass spectrometry has revolutionized the field of clinical microbiology [34]. Although its application is fundamentally related to microbiological identification [35] applications combined with other bioinformatic analysis platforms have recently emerged [36,37]. In line with the above, the use of artificial intelligence has disruptively entered the field of health, thus becoming a new tool to be considered for the diagnosis of different pathologies [38-40]. The first attempts to discriminate Streptococcus from the viridans group, using mass spectrometry, were encouraging according to the published results which were based on a small number of isolates or creating a specific database for this group of microorganisms [41,42]. However, it was later found that Bruker's Biotyper 3.0 system could not resolve the low specificity in the identification of this genus [43-46]. That is why several authors have ventured into various approaches to overcome this limitation. Werno AM, et al. [47] proposed the use of specific peak analysis to confirm identification, which implied an improvement in the typing of different species of *Streptococcus mitis*, including *Streptococcus pneumoniae*.

Subsequently, Ikryannikova LN, et al. [48] published an article referring to the difficulty of discriminating pneumococcus from *Streptococcus mitis* by mass spectrometry; for which the authors proposed to use, in addition to biomarker peaks, artificial intelligence classifier algorithms. However, other authors [49-51] tried to replicate this methodology but were unable to find the peaks described. It is important to note that this limitation occasionally arose during the creation of the spectral database and during the "real-time" classification of the unknown isolates performed in this manuscript.

In this context, an inclusion criterion was established for the isolates to be incorporated in this work (either as part of the training set as well as in the validation set). In this way, only those isolates that did not present discrepancies in the top ten of the results obtained by MS and that also showed a score > 2.0 were included, by doing this, we could affirm that our spectra accomplished the quality we needed to perform this approach.

The objective of this work was to evaluate the application of MALDI-TOF MS in combination with artificial intelligence algorithms, as screening methods in the serotyping of *Streptococcus pneumoniae.*

To this end, two-class models were proposed to differentiate PCV13 vaccine serotypes from NON PCV13 serotypes, which would allow to performed the Quellung serotyping in a more targeted manner combined with the most prevalent circulating serotypes in our country, substantially saving both reagents and man-hours.

First, a calibration set comprised of isolates of all the most frequent vaccine and non-vaccine serotypes within the local epidemiology was used, and a principal component analysis (PCA) was performed from the spectra obtained by MS. This unsupervised analysis was carried out with the aim of exploring the behavior of these objects (Spn isolates) in function of the new variables defined by this study. It was possible to see that the first three components explained the highest percentage of the variance of the data, which gave an accumulated variance of 82%.

In the unsupervised hierarchical clustering analysis, in the dendrogram obtained, a high percentage of discrimination between both classes can be seen, with only 1 isolate (PCV13) that was grouped outside of those expected.

Once the behavior of the study objects was explored in an unsupervised manner, supervised training was implemented, which aims to give a predictive approach to analysis, considered a subdomain of artificial intelligence, in which the computer uses algorithms to learn from a set of past data to make predictions about new data. In this way, it is possible to classify according to a previously established criterion in the training phase.

First, the best discriminatory peaks for each class were identified, according to the previously established parameters, which yielded two peaks for each class (4703.69 Da and 10785.95 Da). The performance values observed in the detection of these peaks were acceptable, since two well-defined groupings were observed in the two-dimensional plot of classes. It's well known that one of the limitations of MALDI TOF profiling approaches is the lack of information about the identity of the information obtained. Even the most accurate MALDI-TOF mass spectrometers have a limited power in terms of resolution and mass accuracy even in a reflector "on" mode. Acquiring in linear mode is known to provide higher sensitivity, but the return is the loss in resolution and mass accuracy. The idea of comparing the observed mass obtained by MALDI TOF operating in linear mode for a given peak, with the theoretical mass of a protein indexed in a database for protein identification is profound conceptual mistake. That's the reason why cannot refer those peaks to any specific proteins.

Nakano SY, et al. [52] used ClinPro Tools software to create prediction models for the ten most prevalent serotypes in Japan, but were only able to validate the assay for three serotypes (3, 15A, and 19A). Subsequently Pinto TCA, et al. [53], through the use of biomarker peaks and using the BioNumerics 7.6 software, found encouraging results but only with the serotypes 6A, 6B, 6C, 9N, 9V and 14. However, both authors conclude that it is necessary to carry out an external validation to corroborate the true reproducibility of the evaluated approaches.

In this work, five classifier models were calibrated using the ClinPro Tools software, of which the recognition capacity and cross-validation values showed greater efficiency for the GA/k-NN and SNN algorithms. An important result to highlight is the following, and it is that in accordance with what was found in the unsupervised analysis, more precisely in the two-dimensional figure, it is that the peaks selected to be able to make this figure, are in agreement with the peaks used by some of the classifying models, providing more robustness to the results obtained.

When implementing these models independently, low sensitivity and specificity values were observed, which is an inappropriate option to use them in this way as a screening technique for the serotyping of unknown isolates. However, when applying both algorithms in parallel and combined, a notable improvement in specificity (82%) and therefore in the positive predictive value (81%) was achieved. However, negative sensitivity and predictive value values continued to be around 60%, yielding inconclusive results in 37/100 isolates.

As a perspective to improve the predictive parameters, we can mention the increase in the number of isolates both to create the training set as well as the number of isolates to challenge.

Although the results obtained for this work were not as expected, by implementing the models developed in the form of screening, the use of antisera was reduced by 10.2% compared to the blindly Quellung technique that we used to do, as it was mentioned before.

Undoubtedly, the development and application of EM has contributed to meeting the demands in the field of microbiological diagnosis due to the important advantages it presents compared to more traditional methodologies. Among these advantages, its versatility can be mentioned (since it can be applied to bacterial, mycological, virus, parasite cultures, even from the clinical sample itself). Additionally, minimal sample preparation is required and it is a fast, easy-to-use analysis technique that allows parameters to be evaluated in real time [54].

In this work it was possible to demonstrate that the combination of MALDI-TOF MS and multivariate analysis allows the development of new strategies for the identification and characterization of Spn isolates of clinical importance. MALDI-TOF mass spectra generate a large amount of data, which requires appropriate analysis methods to make the most of the information contained in them. In this sense, multivariate analysis models (both supervised and unsupervised) allow extracting information from the spectra (multivariate data set) and correlating it with different properties of the samples.

The results of this work represent the bases to continue exploring the combination of MALDI-TOF MS with multivariate analysis, with prospects of improving the predictive parameters so that the developed models can be more robust and reliable.

On the other hand, the development of this work provided a solid background in multivariate analysis as a tool to extract useful information and produce inferences from a large amount of data.

Acknowledgments

Thanks to John McQuiston, Team Lead, Special Bacteriology Reference Laboratory at CDC, your opinions and suggestions were fundamental to conclude with the writing of this manuscript.

References

1. Denapaite, Dalia and Regine Hakenbeck. "A new variant of the capsule 3 cluster occurs in *Streptococcus pneumoniae* from deceased wild chimpanzees." PLoS One 6 (2011): e25119.

2. Kadioglu, Aras, Jeffrey N. Weiser, James C. Paton and Peter W. Andrew, et al. "The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease." *Nat Rev Microbiol* 6(2008): 288-301.

3. UNICEF, y WHO (2006). Pneumonia: the forgotten killer of children.

4. UNICEF (2015). UNICEF global databases 2015.

5. OMS (2016). Neumonía.

6. Cooper, Georgia, Carolyn Rosenstein, Annabel Walter and Lenore Peizer, et al. "The further separation of types among the pneumococci hitherto included in group IV and the development of therapeutic antisera for these types." *J Exp Med* 55 (1932): 531.

7. Beckler, Edith and Patricia MacLeod. "The Neufeld method of pneumococcus type determination as carried out in a public health laboratory: A study of 760 typings." *J Clin Investig* 13 (1934): 901-907.

8. Lund, Erna. "Laboratory diagnosis of pneumococcus infections." *Bull World Health Organ* 23 (1960): 5.

9. Ganaie, Feroze, Jamil S. Saad, Lesley McGee and Andries J. van Tonder, et al. "A new pneumococcal capsule type, 10D, is the 100th serotype and has a large cps fragment from an oral *Streptococcus*." MBio 11 (2020): e00937-20.

10. Grabenstein, J. D. and K. P. Klugman. "A century of pneumococcal vaccination research in humans." *Clin Microbiol Infect* 18 (2012): 15-24.

11. Oliveira, Giuliana S., Maria Leonor S. Oliveira, Eliane N. Miyaji and Tasson C. Rodrigues, et al. "Pneumococcal vaccines: Past findings, present work, and future strategies." Vaccines 9 (2021): 1338.

12. Lineamientos tecnicos Vacunacion contra Neumococo. DiCEI. Ministerio de Salud de la Nacion. Acceso.

13. Bell, Brian G., Francois Schellevis, Ellen Stobberingh and Herman Goossens, et al. "A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance*BMC Infect Dis* 14 (2014): 1-25.

14. Flasche, Stefan, Albert Jan Van Hoek, Elizabeth Sheasby and Pauline Waight, et al. "Effect of pneumococcal conjugate vaccination on serotype-specific carriage and invasive disease in england: A cross-sectional study." *PLoS Med* 8 (2011): e1001017.

15. Obaro, Steven and Richard Adegbola. "The pneumococcus: Carriage, disease and conjugate vaccines." *J Med Microbiol* 51 (2002): 98-104.

16. Obolski, Uri, José Lourenço, Craig Thompson and Robin Thompson, et al. "Vaccination can drive an increase in frequencies of antibiotic resistance among nonvaccine serotypes of *Streptococcus pneumoniae*." Proc Natl Acad Sci 115 (2018): 3102-3107.

17. Lehtinen, Sonja, François Blanquart, Nicholas J. Croucher and Paul Turner, et al. "Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage." Proc Natl Acad Sci 114 (2017): 1075-1080.

18. Moore, Matthew R., Ruth Link-Gelles, William Schaffner and Ruth Lynfield, et al. "Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: Analysis of multisite, population-based surveillance." *Lancet Infect Dis* 15 (2015): 301-309.

19. Pilishvili, Tamara and Nancy M. Bennett. "Pneumococcal disease prevention among adults: Strategies for the use of pneumococcal vaccines." Vaccine 33 (2015): D60-D65.

20. Kallow, Wibke, Marcel Erhard, Haroun N. Shah and Emmanuel Raptakis, et al. "MALDI-TOF MS for microbial identification: Years of experimental development to an established protocol." Mass spectrometry for microbial proteomics (2010): 255-276.

21. He, Ying, Haijing Li, Xuedong Lu and Charles W. Stratton, et al. "Mass spectrometry biotyper system identifies enteric bacterial pathogens directly from colonies grown on selective stool culture media." *J Clin Microbiol* 48 (2010): 3888-3892.

22. Cherkaoui, Abdessalam, Jonathan Hibbs, Stéphane Emonet and Manuela Tangomo, et al. "Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level." *J Clin Microbiol* 48 (2010): 1169-1175.

23. Rocca, María Florencia, Jonathan Cristian Zintgraff, María Elena Dattero and Leonardo Silva Santos, et al. "A combined approach of MALDI-TOF mass spectrometry and multivariate analysis as a potential tool for the detection of SARS-CoV-2 virus in nasopharyngeal swabs." *J Virol Methods* 286 (2020): 113991.

24. Tran, Anthony, Kevin Alby, Alan Kerr and Melissa Jones, et al. "Cost savings realized by implementation of routine microbiological identification by matrix-assisted laser desorption ionization–time of flight mass spectrometry." *J Clin Microbiol* 53 (2015): 2473-2479.

25. Dhiman, Neelam, Leslie Hall, Sherri L. Wohlfiel and Seanne P. Buckwalter, et al. "Performance and cost analysis of matrix-assisted laser desorption ionization–time of flight mass spectrometry for routine identification of yeast." *J Clin Microbiol* 49 (2011): 1614-1616.

26. Wang, Hsin-Yao, Frank Lien, Tsui-Ping Liu and Chun-Hsien Chen, et al. "Application of a MALDI-TOF analysis platform (ClinProTools) for rapid and preliminary report of MRSA sequence types in Taiwan." PeerJ 6 (2018): e5784.

27. MALDI Biotyper 3.1 User Manual.

28. Bruker Daltonik GmbH, 2011.ClinPro Tools User Manual Version 3.0.BrukerDaltonik GmbH, Bremen.

29. Camoez, M., J. M. Sierra, M. A. Dominguez and M. Ferrer-Navarro, et al. "Automated categorization of methicillin-resistant *Streptococcus aureus* clinical isolates into different clonal complexes by MALDI-TOF mass spectrometry." *Clin Microbiol Infect* 22 (2016): 161-e1.

30. Zhang, Hongjuan, Jing Cao, Lin Li and Yanbin Liu, et al. "Identification of urine protein biomarkers with the potential for early detection of lung cancer." *Sci Rep* 5 (2015): 11805.

31. Stephens, Michael A. "EDF statistics for goodness of fit and some comparisons." *J Am Stat Assoc* 69 (1974): 730-737.

32. Khot, Prasanna D. and Mark A. Fisher. "Novel approach for differentiating Shigella species and *E. coli* by matrix-assisted laser desorption ionization–time of flight mass spectrometry." *J Clin Microbiol* 51 (2013): 3711-3716.

33. Satzke, Catherine, Paul Turner, Anni Virolainen-Julkunen and Peter V. Adrian, et al. "Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*: updated recommendations from the world health organization pneumococcal carriage working group." Vaccine 32 (2013): 165-179.

34. Lavigne, Jean-Philippe, Paula Espinal, Catherine Dunyach-Remy and Nourredine Messad, et al. "Mass spectrometry: A revolution in clinical microbiology?." *Clin Chem Lab Med* 51 (2013): 257-270.

35. Rocca, Maria Florencia, Marisa Almuzara, Claudia Barberis and Carlos Vay, et al. "presentation of the national network for microbiological identification by mass spectrometry website. Guide for the interpretation of maldi-tof ms results." *Rev Argent Microbiol* 52 (2020): 83-84.

36. Kubo, Yumi, Osamu Ueda, Sawa Nagamitsu and Hachiro Yamanishi, et al. "Novel strategy of rapid typing of shiga toxin-producing *E. coli* using maldi biotyper and clinprotools analysis." *J Infect Chemother* 27 (2021): 1137-1142.

37. Fiamanya, Selali, Lucía Cipolla, Mónica Prieto and John Stelling. "Exploring the value of MALDI-TOF MS for the detection of clonal outbreaks of burkholderia contaminans." *J Microbiol Methods* 181 (2021): 106130.

38. Gubbi, Sriram, Pavel Hamet, Johanne Tremblay and Christian A. Koch, et al. "Artificial intelligence and machine learning in endocrinology and metabolism: The dawn of a new era." *Fron Endocrinol* 10 (2019): 185.

39. Landberg, Göran, Paul Fitzpatrick, Pauline Isakson and Emma Jonasson, et al. "Patient-derived scaffolds uncover breast cancer promoting properties of the microenvironment." *Biomater* 235 (2020): 119705,

40. Lau, Anna F. "Matrix-assisted laser desorption ionization time-of-flight for fungal identification." *Clin Lab Med* 41 (2021): 267-283.

41. Rupf, S., K. Breitung, W. Schellenberger and K. Merte, et al. "Differentiation of mutans streptococci by intact cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." *Oral Microbiol Immunol* 20 (2005): 267-273.

42. Friedrichs, C., A. C. Rodloff, G. S. Chhatwal and W. Schellenberger, et al. "Rapid identification of viridans streptococci by mass spectrometric discrimination." *J Clin Microbiol* 45 (2007): 2392-2397.

43. La Scola, Bernard and Didier Raoult. "Direct identification of bacteria in positive blood culture bottles by matrix-assisted laser desorption ionisation time-of-flight mass spectrometry." PloS One 4 (2009): e8041.

44. Stevenson, Lindsay G., Steven K. Drake and Patrick R. Murray. "Rapid identification of bacteria in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry." *J Clin Microbiol* 48 (2010): 444-447.

45. Risch, Martin, Darko Radjenovic, Jong Nam Han and Monica Wydler, et al. "Comparison of MALDI TOF with conventional identification of clinically relevant bacteria." *Swiss Med Wkly* 140 (2010): w13095-w13095.

46. Welker, Martin and Edward RB Moore. "Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology." *Syst Appl Microbiol* 34 (2011): 2-11.

47. Werno, Anja M., Martin Christner, Trevor P. Anderson and David R. Murdoch. "Differentiation of *Streptococcus pneumoniae* from nonpneumococcal streptococci of the *Streptococcus mitis*group by matrix-assisted laser desorption ionization–time of flight mass spectrometry." *J Clin Microbiol* 50 (2012): 2863-2867.

48. Ikryannikova, L. N., A. V. Filimonova, M. V. Malakhova and T. Savinova, et al. "Discrimination between *Streptococcus pneumoniae*and *Streptococcus mitis*based on sorting of their maldi mass spectra." *Clin Microbiol Infect* 19 (2013): 1066-1071.

49. Chen, Jonathan HK, Kevin KK She, Oi-Ying Wong and Jade LL Teng, et al. "Use of MALDI biotyper plus clinprotools mass spectra analysis for correct identification of *Streptococcus pneumoniae* and *Streptococcus mitis/oralis*." *J Clin Pathol* 68 (2015): 652-656.

50. Marín, Mercedes, Emilia Cercenado, Carlos Sánchez-Carrillo and Adrián Ruiz, et al. "Accurate differentiation of *Streptococcus pneumoniae* from other species within the *Streptococcus mitis* group by peak analysis using MALDI-TOF MS." *Front Microbiol* 8 (2017): 698.

51. Ercibengoa, María, Marta Alonso, Diego Vicente and Maria Morales, et al. "Utility of MALDI-TOF MS as a new tool for *Streptococcus pneumoniae* serotyping." PLoS One 14 (2019): e0212022.

52. Nakano, Satoshi, Y. Matsumura, Y. Ito and T. Fujisawa, et al. "Development and evaluation of MALDI-TOF MS-based serotyping for *Streptococcus pneumoniae*." *Eur J Clin Microbiol Infect Dis* 34 (2015): 2191-2198.

53. Pinto, Tatiana CA, Natalia S. Costa, Luciana FS Castro and Rachel L. Ribeiro, et al. "Potential of MALDI-TOF MS as an alternative approach for capsular typing *Streptococcus pneumoniae* isolates." *Sci Rep* 7 (2017): 45572.

54. Wang, Yueling, Yan Jin, Yuanyuan Bai and Zhen Song, et al. "Rapid method for direct identification of positive blood cultures by MALDI−TOF MS." *Exp Ther Med* 20 (2020): 1-1.

How to cite this article: Zintgraff, Jonathan., Florencia Rocca, Nahuel Sánchez Eluchans and Lucía Irazu, et al. “A *la carte, streptococcus pneumoniae* capsular Typing. MALDI-TOF mass spectrometry and machine-learning algorithms as complementary tools for the determination of PCV13 serotypes and the most prevalent NON PCV13 serotypes according to Argentina’s epidemiology.” *J Med Microb Diagn* 12 (2023): 403.