International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:13, No:9, 2019

# Machine Learning for Aiding Meningitis Diagnosis in Pediatric Patients

Karina Zaccari, Ernesto Cordeiro Marujo

***Abstract*—**This paper presents a Machine Learning (ML) approach to support Meningitis diagnosis in patients at a children's hospital in Sao Paulo, Brazil. The aim is to use ML techniques to reduce the use of invasive procedures, such as cerebrospinal fluid (CSF) collection, as much as possible. In this study, we focus on predicting the probability of Meningitis given the results of a blood and urine laboratory tests, together with the analysis of pain or other complaints from the patient. We tested a number of different ML algorithms, including: Adaptative Boosting (AdaBoost), Decision Tree, Gradient Boosting, K-Nearest Neighbors (KNN), Logistic Regression, Random Forest and Support Vector Machines (SVM). Decision Tree algorithm performed best, with 94.56% and 96.18% accuracy for training and testing data, respectively. These results represent a significant aid to doctors in diagnosing Meningitis as early as possible and in preventing expensive and painful procedures on some children.

***Keywords*—**Machine learning, medical diagnosis, meningitis detection, gradient boosting.

## I. Introduction

IN the last years the use of ML in Medical Diagnosis has gradually increased [6], as a result of the constant development of techniques to extract statistical information or hidden patterns from medical databases. ML represents a practical and assertive way to assist doctors and healthcare professionals in diagnosing diseases more efficiently and safely [34]. A series of ML technologies have been tried in biomedical problems, mainly for the diagnosis, therapeutic planning and patients' prognosis [27].

This article presents an approach to support Meningitis diagnosis on patients from a children's hospital in São Paulo, Brazil. By applying ML techniques, we aim to help detect the disease as soon as possible and to avoid invasive procedures such as CSF collection.

Meningitis is an acute inflammation of the protective membranes lining the brain and spinal cord caused by viruses, bacteria, parasites, or fungi [35]. Even with the introduction of new polysaccharide and conjugate vaccines in the last decades [38], Meningitis remains a source of substantial morbidity and mortality in many countries [31].

In Brazil, 20% of children vaccinated against Meningitis do not receive the booster dose. In 2018 there were over 3,000 deaths from the disease, out of 15,706 cases in the year, according to the Ministry of Health. Even though the disease is considered rare in the country, the amount of exams performed on children to detect it is still high because the diagnosis can only be confirmed or excluded with the CSF collection [35].

The CSF is a sterile, limpid and colorless body fluid, found in the subarachnoid space in the brain and spinal bone marrow, brain ventricles and the bone marrow central channel [14]. Its collection and analysis are necessary to diagnose neurological pathologies, staging and vascular processes complications, infectious, inflammatory or neoplastic syndromes of organs surrounded by this liquid [7].

The CSF collection can be carried out in three ways, being the lumbar puncture the most used, followed by suboccipital and ventricular [13]. Overall, the test is recommended to diagnose meninges infections, subarachnoid hemorrhage, primary or metastatic malignancy, and demyelinating diseases [25]. The analysis of patient's data at issue showed that only 16.6% of the patients submitted to this procedure were diagnosed with Meningitis. The remaining patients received diagnosis in which brain liquor exam was not necessarily needed.

Studies on children's behavior in hospitals [37] show that examination procedures are a source of significant stress. Hospital professionals have demonstrated a significant concern about the negative effects that the CSF collection might produce on patients.

The purpose of this work is to develop a quantitative measure to assist healthcare professionals in deciding whether or not patients need to undergo the CSF exam, thus avoiding invasive and unnecessary procedures during the diagnosis phase.

## II. Methodology

### A. Database

To accomplish this project five databases were necessary. They were provided by a Sao Paulo hospital considered excellent in terms of medical services quality. The hospital also provided us with the tables produced by the laboratory specialists of reference values for each exam.

The datasets do not permit the identification of patients. No patient was interviewed.

In all datasets, a primary key represents each patient and enables us to merge datasets concerning clinical exams with data from emergency room and hospitalization registers. Careful planning of this merging data process assured that each data segment remains unique [12] and is easily accessible. The databases concern the period from March/ 2014 to September/2018 and comprise information on (1)

Karina Zaccari is with the Instituto Tecnológico de Aeronáutica, São José dos Campos, SP Brazil, and with the Itaú Unibanco, São Paulo, SP Brazil (e-mail: karina.zaccari@itau-unibanco.com.br).

Ernesto Cordeiro Marujo, is with the Instituto Tecnológico de Aeronáutica, São José dos Campos, SP Brazil (e-mail: marujo@ita.br).

patients' information registered in the emergency room, (2) patients' information registered when hospitalized, (3) clinical exams results carried out on each patient, (4) clinical exams reference values, (5) results and reference values of brain liquor exam. This collection of datasets contains information on every entrance in the hospital during the period from March/2014 to September/2018. Patients' age varied from 0 to 17 years and the vast majority of entrances refers to medical issues that do not relate to Meningitis.

There are 507,406 registers in total in the period, implying an average of approximately 9,000 entrances per month and of those, an average of 15 cases were diagnosed with Meningitis per month. Positive diagnosis for Meningitis was split not quite evenly between girls (43%) and boys (57%). This and other variables unbalance in the dataset required the use of special techniques as we discussed in Section III.

Another issue that we had to address was the fact that we did not have the Reference Values for many laboratory exams. The database contains information on approximately 1,700 distinct laboratory exams. However, only 34 of those laboratory exams had reference values registered in the dataset. The proposed alternatives to tackle this issue would be: (1) to search the reference values from other information sources, or (2) to apply neural networks algorithms in order to train a proxy of reference values for these results. In both cases, we considered that the risk of introducing spurious results was too high. In fact, for the laboratory tests concerned, the establishment of reference values was considered very challenging by specialists working in clinical laboratories. Thereupon, we opted to continue with the research considering only the data from the 34 exams which had trustworthy reference values and were validated by the laboratory control experts.

Of course, in future works one may study the use of neural networks algorithms in order to train a proxy of reference values for these laboratory exam results. However, for the purpose of this research, only 34 were considered:

- All blood exams held in emergency room were considered: *Anomalous Lymphocytes %*, *Atypical Lymphocytes %*, *Basophils*, *Calcium*, *Chlore*, *Creatinine*, *Direct bilirubin*, *Eosinophil %*, *Erythrocytes*, *Hemoglobin Dosage*, *Indirect bilirubin*, *INR-Prothrombin*, *Leukocytes*, *Lymphocytes %*, *Magnesium*, *MCH - Mean Corpuscular Hemoglobin*, *MCHC - Mean Corpuscular Hemoglobin Concentration*, *MCV - Mean Corpuscular Volume, Metamyelocytes %*, *Monocytes*, *Myelocytes %*, *Neutrophils*, *Partial thromboplastin time*, *Partial thromboplastin time in seconds*, *Platelet count*, *Potassium*, *Prothrombin time*, *Segmented %*, *Sodium*, *Total bilirubin*, *Typical Lymphocytes %*;
- Urine exams: *Urea*, *Urine Aspect* and *Urine coloration*.

Since each exam has a particular unit of measurement, we opted to exclude the symbol of such unit and consider only the result values in float decimal numbers. For tests whose results are expressed in text format, we decided to categorize each possibility and register them using encoding techniques.

Another hindrance was detected when trying to study the brain liquor exam results. The bases which had these results were not structured; they were in JSON format (JavaScript Object Notation) and contained rich text format (RTF) columns. In this format (RTF), the columns are written as:

\{*?\\[^{}]+}|[{}]|\\\n?[A-Za-z]+\n?(?:-?\d+)?[ ]?

and generate a result such as shown in Fig. 1.

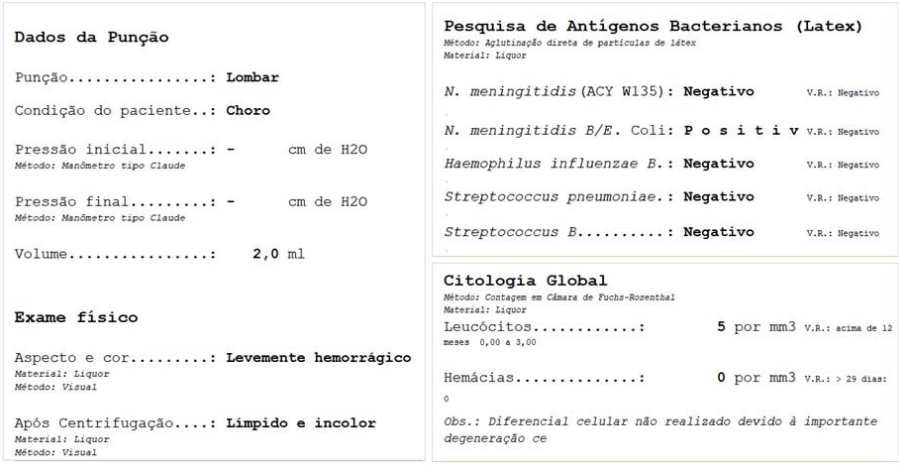


Fig. 1 Example of result field in the brain liquor exams database *(words in Portuguese)*

To enable the information extraction, specific non-structured data-solving techniques were needed. We used the Search and Validation by Regular Expression technique [11] using the open code algorithm found in [18].

### *B. Dependent Variable*

In this work, we focused our attention only on patients who had been suspected for Meningitis.

We used the International Statistics Classification of Diseases and Health Problems (ICD) [32]. In that document,

the World Health Organization (WHO) codifies and assigns a unique category to diseases depending on signs, symptoms, anomalous aspects, complaints, social circumstances and external causes of injury according to the medical diagnosis attributed to each patient.

Accordingly, we started by considering only the patients who would be asked to take the CSF exam. These included those classified in the following ICD codes [32]: *A392 – Acute meningococcemia*, *A87 – Viral Meningitis* , *A878 – Other viral Meningitis* , *A879 – Non-specific Viral Meningitis* , *G00 – Non-classified bacterial Meningitis in other part*, *G009 – Non-classified bacterial Meningitis* , *G03 – Meningitis due to other causes and non-specified causes*, *G038 – Meningitis due to specific causes*, *G039 – Non-specific Meningitis*. These patients formed the total set of patients that had been submitted to the CSF exam to confirm the diagnose Meningitis. We assign the value "1" to patients that have taken the exam and had the disease, and value "0" to patients that have taken the exam but were not diagnosed with Meningitis.

The final database had data relative to 3.265 patients in total. Recall that these are the patients that have had their brain liquor taken and examined during their stay in the emergency room, or during the hospital admission. Among this public, only 542 patients (16.6%) tested positive for Meningitis. The remaining patients were diagnosed with other pathologies and the brain liquor collection could perhaps have been avoided.

Fig. 2 shows the number of patients who tested positive and the total number of patients submitted to the CSF test in a month by month basis.

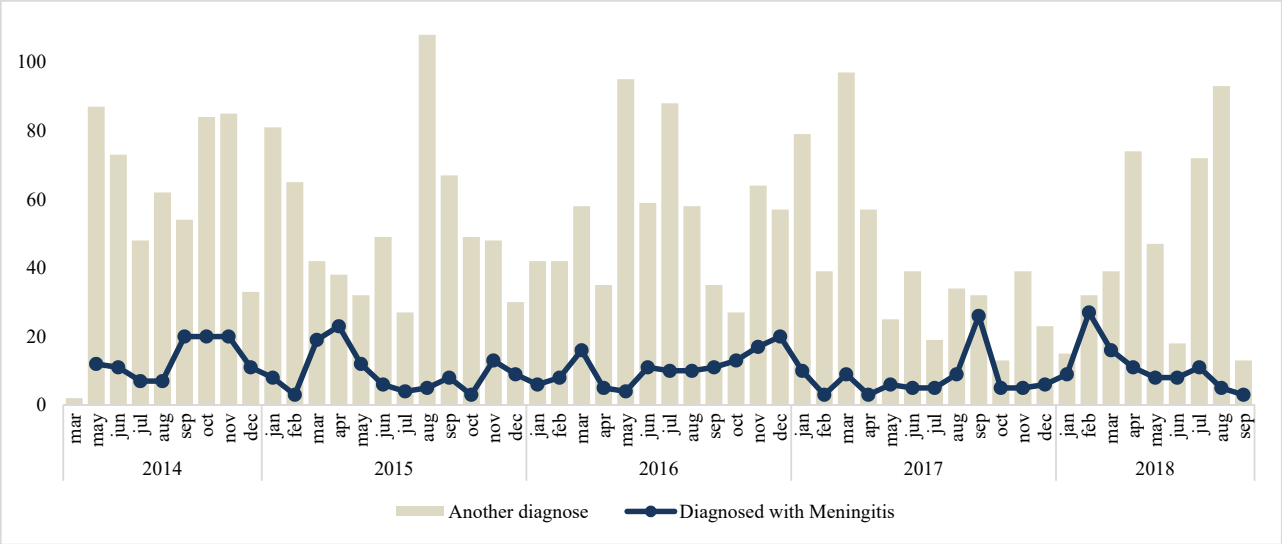


Fig. 2 Number of patients who were tested positive and the total number of patients that underwent the brain liquor exam in the period from March/2014 to September/2018

The final database is made of 3.265 individuals (rows), and 37 features (columns). One of these columns is the binary variable that indicates whether or not the individual was diagnosed with Meningitis.

It is out of the scope of this paper to report on the practical challenges we had to overcome in order to arrive at this table: natural language processing, encoding, conversion of units, and others. In the next paragraph we describe the ML techniques used to process this dataset.

## III. ML Techniques

In this work, several ML techniques were tested. We explored popular and novel techniques for classification since the labels (1, if the individual was diagnosed with Meningitis and 0 otherwise) are discrete [2]. The techniques we used are reportedly excellent ML classification tools [40]. The relative performance of each of these techniques, however, is highly dependent on the context and this work might help indicating the pros and cons of them in the context of diagnosing Meningitis.

This section contains a brief explanation about the classification methods used to accomplish this project.

All techniques were programmed using scikits.learn. This is an open code library for the programming language Python. In that library, one can find several ML algorithms for classification, regression and grouping [33].

It is important to note that our data is strongly unbalanced: the number of observations that correspond to Meningitis is much smaller than the number of observations not diagnosed with Meningitis.

If we do not apply a balancing technique, the ML algorithm will tend to bias the classification assigning new individuals in the majority class more than it should [23].

Balancing techniques include the oversampling techniques [10]. The idea of oversampling techniques is to increase the number of data points corresponding to individuals in the minority class, in our case, those diagnosed with Meningitis.

Several procedures could be used for balancing the dataset.

In this research, we used the SMOTE – Synthetic Minority Over-sampling Technique, which creates synthetic samples from the minor class instead of creating copies [10]. A pseudo-code of this algorithm is shown in Fig. 3.

**Algorithm** *SMOTE*(T, N, k)
**Input:** Number of minority class samples $T$; Amount of SMOTE $N\%$; Number of nearest neighbors $k$
**Output:** $(N/100) * T$ synthetic minority class samples
1. $(*$ *If* $N$ *is less than 100%, randomize the minority class samples as only a random percent of them will be SMOTEd.* $*)$
2. **if** $N < 100$
3. **then** Randomize the $T$ minority class samples
4. $T = (N/100) * T$
5. $N = 100$
6. **endif**
7. $N = (int)(N/100)$ $(*$ *The amount of SMOTE is assumed to be in integral multiples of 100.* $*)$
8. $k =$ Number of nearest neighbors
9. $numattrs =$ Number of attributes
10. $Sample[\,][\,]$: array for original minority class samples
11. $newindex$: keeps a count of number of synthetic samples generated, initialized to 0
12. $Synthetic[\,][\,]$: array for synthetic samples
$(*$ *Compute* $k$ *nearest neighbors for each minority class sample only.* $*)$
13. **for** $i \leftarrow 1$ **to** $T$
14. Compute $k$ nearest neighbors for $i$, and save the indices in the $nnarray$
15. Populate($N$, $i$, $nnarray$)
16. **endfor**

$Populate(N,\ i,\ nnarray)$ $(*$ *Function to generate the synthetic samples.* $*)$
17. **while** $N \neq 0$
18. Choose a random number between 1 and $k$, call it $nn$. This step chooses one of the $k$ nearest neighbors of $i$.
19. **for** $attr \leftarrow 1$ **to** $numattrs$
20. Compute: $dif = Sample[nnarray[nn]][attr] - Sample[i][attr]$
21. Compute: $gap =$ random number between 0 and 1
22. $Synthetic[newindex][attr] = Sample[i][attr] + gap * dif$
23. **endfor**
24. $newindex{+}{+}$
25. $N = N - 1$
26. **endwhile**
27. **return** $(*$ *End of Populate.* $*)$
End of Pseudo-Code.

Fig. 3 SMOTE Pseudo-Code [10]

Besides SMOTE, we also tried a K-fold Cross Validation. By analyzing both scenarios, the one which had better performance to support our unbalanced data with the classificators was the combination of SMOTE technique with the K-fold Cross Validation with 10 folds. Other balancing techniques can be studied in future work.

### *A. Decision Tree*

A Decision Tree is a technique that partitions the data at each step. Each partition is the result of a condition applied to an attribute. The operation of the algorithm resembles the branching of a tree where each branch represents the result of applying a condition on the value of an attribute of all data points. At the end, the result is a set of "leafs" representing a single class, considering all attributes of the tree [24].

### *B. KNN*

The KNN method is not a parametric model. It considers the data of each individual as a point in a Euclidean space [26] and tries to find groups of k neighbors where, in each group, the elements are close to each other and distant from elements in other groups.

The KNN is currently widely used for classification problems and ML regression [5]. The main idea of this algorithm is to determine the classification label of a sample based on neighbor samples derived from a training set. It has just one free parameter (k, the number the neighbors in each group) which is controlled by the user with the aim to obtain a good classification.

### *C. Logıstic Regression*

Logistic Regression is a statistical modelling technique that enables us to forecast the value of a binary variable of an individual, considering the knowledge of the value of other variables associated to that individual [1].

### *D. Random Forest*

Random Forest method is an ensemble of decision trees built in the training data where each of the trees has randomly and independently sampled values [22]. After the random

decision tree construction, each tree will generate a classification, in accordance to the problem concerned. The algorithm final result will assign an individual to the most voted class among all decision trees [8]. One major focus when using this algorithm is that it has commonly shown overfitting in the training data [16], besides low performance in the test data.

### *E. SVM*

SVM are algorithms that could be used for classification, regression and other supervised ML tasks. SVMs have demonstrated equivalent performance to other ML algorithms, such as Artificial Neural Network (ANN) [20].

The main advantages of SVM are: (1) good generalization capacity; (2) robustness in big dimension data, (3) presence of just one global minimum, once it implies in the optimization of a cubic function and (4) strong statistics theoretical basis [26]. The main idea of this algorithm is to map the set of its original space for a new, bigger dimension one [21], aiming at creating a decision surface from a great hyperplane with a good separation margin among the data from different families [20].

### *F. Gradient Boosting*

The Gradient Boosting is an algorithm which generates classifications from an ensemble of predicted models. Usually, these predictive models are decision tree models. Each classification model goal is to minimize a cost function (loss function) and gradient descending methods are used [17].

### *G. Adaptive Boosting*

The Adaptive Boosting (AdaBoost) method is a classification algorithm that has become very popular. It involves a combination of classifying models. The idea is to, step by step, modify the set of weights for each point in the training data. At first, the weights are all equal and the classification models are applied. In the next step, the classification models are used again, but considering a new distribution of weights where the weights of those points that have been incorrectly classified are increased. Adaboost promotes the intense execution of the best classification model in the training data [15]. One of the advantages of AdaBoost is that it does not require the previous knowledge of good classifying models, because it adapts to incorrect predictions [15].

### *H. Classificators Performance Evaluation*

Subsequent to the application of algorithms, the resulting classifications must be evaluated and compared. We used some common metrics and procedures, as the confusion matrix shown in Table I.

A **Confusion matrix** is a table showing the results of a classification exercise in four cells: True Positive (TP), False Positive (FP) in first line and, number of False Negative (FN) and True Negative (TN) in second line. This table is the basis for evaluating the performance of the different algorithms using classical metrics as Accuracy, Sensitivity and Specificity.

TABLE I
EXAMPLE OF A CONFUSION MATRIX

| | | *Real Condition* | | |
|---|---|---|---|---|
| | | *Positive* | *Negative* | |
| *Predicted Condition* | *Positive* | *True Positive (TP)* | *False Positive (FP)* | *TP + FP* |
| | *Negative* | *False Negative (FN)* | *True negative (TN)* | *FN + TN* |
| | | *TP + FN* | *FP + TN* | |

**Accuracy** measures the fraction of correctly classified samples and is calculated by:

$$Accuracy = \frac{TP+TN}{Total\ number\ of\ samples}$$

**Sensitivity** measures the algorithm capacity to find positive cases, also called True Positive Rate, calculated by:

$$Sensitivity = \frac{TP}{TP+FN}$$

**Specificity** measures the algorithm capacity to identify those individuals that are negative, also called True Negative Rate, calculated by

$$Specificity = \frac{TN}{TN+FP}$$

## IV. RESULTS

### *A. Exploratory Data Analysis*

First, we produced a correlation matrix in order to study what are the relationships among the attributes [29]. We used Pearson correlation coefficient, which assess the correlation degree between two variables [3]. We should note that Pearson coefficient adequately measures the correlation for a pair of variables only under the assumption that the relationship between them is linear or quasi linear. Also, it should be noted that the Pearson correlation coefficient is not appropriate if the distribution of the points does not follow a Gaussian distribution [30].

We normalized the original data set and computed all possible Pearson correlation coefficients. To visualize these correlations we use a heat map as a graphic form to represent the individual values contained in a matrix. The dark colors indicate positive correlation, and the light colors indicate negative correlations [39], as we can see in Fig. 4.

From the correlation matrix, it is possible to visually extract important information. For instance:

1) The colors among the variables direct bilirubin, indirect bilirubin and total bilirubin indicate these are highly correlated. Such fact can be expected since indirect bilirubin coming from hemoproteins catabolism is converted into direct bilirubin through the connection with glucuronic acid molecules [36], and total bilirubin means the total sum of both;
2) The correlation between the variables hemoglobin and erythrocytes were also highly positive. This occurs because erythrocytes (red blood cells) are in charge of the

transportation of hemoglobin in the blood [9], showing there is a strong dependence relation between them.

Moreover, it is possible to verify that the target (answer variable) *Class* has stronger positive correlation with some variables (such as *platelet count*, *erythrocytes*, *hemoglobin dosage* and *segmented percentage*), and negative correlation with other variables (such as *lymphocyte percentage* and *average corpuscular hemoglobin*).

Another variable, Main complain, registered in the emergency room is possibly very valuable but has not been included in the algorithmic analysis of Meningitis prediction. Since it is a categorical variable with many possible values, we have used a simple procedure to map its importance and defer a deep study on the use of this variable to aid in diagnosing Meningitis for later work.

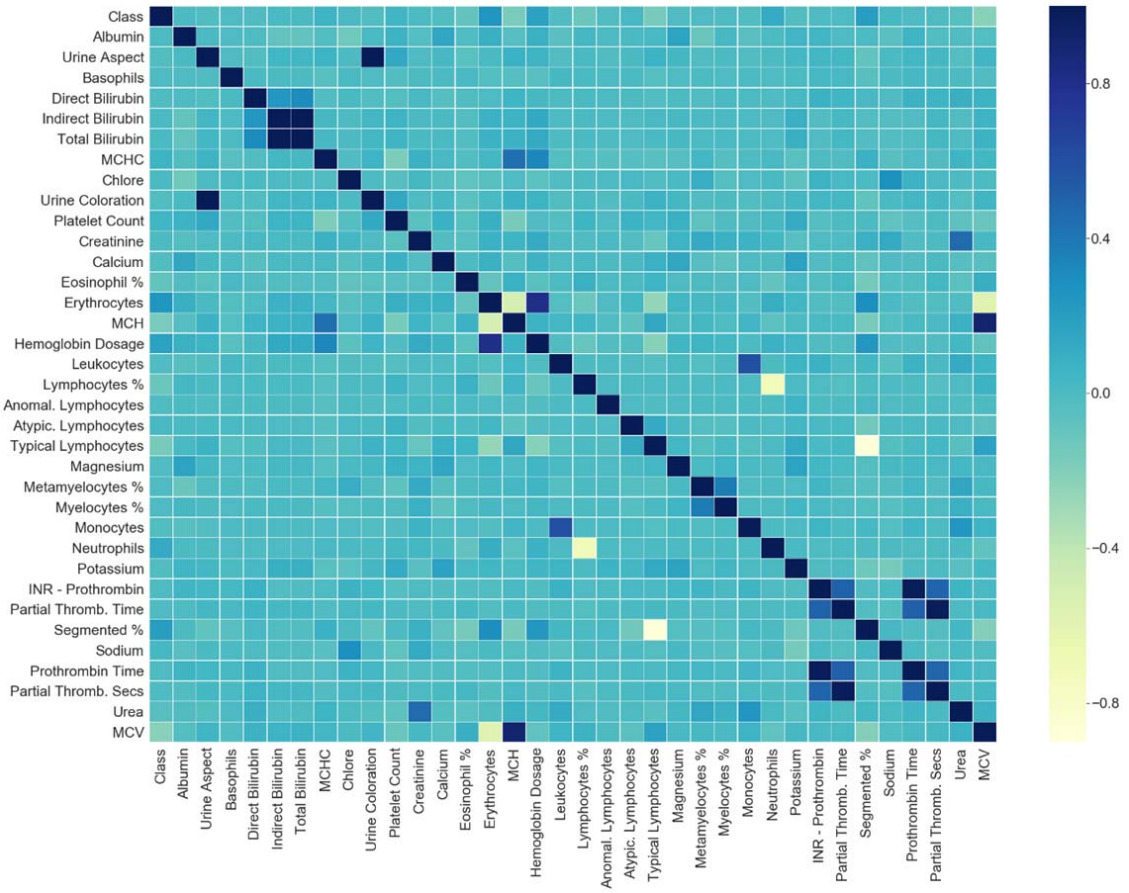


Fig. 4 Correlation Matrix from the final base

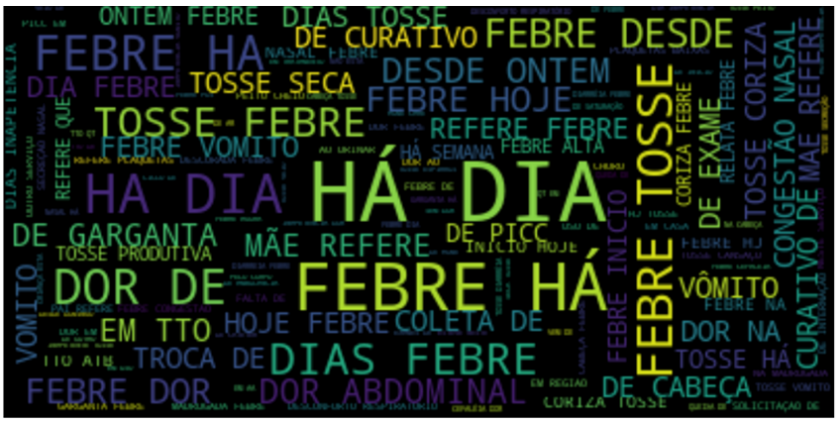


Fig. 5 Word Cloud (in Portuguese) of all patients who undertook the exam

Such variable is a text variable written by a nurse during patient screening. It contains information about complaints verbalized by the patients as soon as they arrive at the hospital.

As we argued that, due to the nature of such text variable and its open possibilities, we chose to produce a word cloud diagram. This is a visual text data presentation for texts in free format, where the importance of each word is shown according to the size or font color [19].

In the word cloud diagram we identify that the main complains were: pain, high temperature (fever) for days, vomit, and others, as illustrated in Fig. 5.

If we considered the subset of population that includes only those who had Meningitis, and use the same word cloud procedure, it is noticeable that besides these complains, we would find complains about otalgia (ear-ache), inappetence (lack of appetite), cephalea (headache), productive cough (when there is mucus or catarrh), nausea (dizziness or sickness), diarrhea, tiredness, abdominal pain and nasal bleeding, as illustrated in Fig. 6.


Fig. 6 Word Cloud (in Portuguese) of all patients who took the exam and were diagnosed with Meningitis

The difference in these two word clouds indicate that it might be worthy to study how to better incorporate this variable together with others in a ML algorithm to aid the Meningitis diagnosis.

*A. Results from ML Algorithms Application*

The ML methods cited in the previous section were applied to two parts of the dataset. First to a subset of the data that we call "Training data" used to estimate the parameters of the model [28]. The second subset segregated from the first one is called "Test data" and is a sample for testing the performance of the techniques used.

The size of these sets should be determined depending on the number of parameters that our algorithm uses. If there are too many parameters to tune, we will need a large set in the Training data, so that we can get enough data to yield statistically meaningful results.

The separation of the data into Training data and Test data should be such that the Test set presents the same characteristics of the Training set but should not be "contaminated" by data in the Training set. "Contamination" generally occurs in time-series data where one observation depends on the previous one. Such phenomenon might be present in our case. Our dataset is time-based, and the number and characteristics of Meningitis cases in one day might interfere in the prognosis of Meningitis in the next day. However, we assumed time-independency and left this issue for future work.

In this work we simply split the data according to the period of collection: For the Training sample we considered the data from the first period: from March/2014 to January/2017. For the Test set, the data came from February/2017 to September/ 2018. The proportion of patients that were diagnosed with Meningitis in both sets was approximately equal after we used the balancing technique SMOTE.

The results of all techniques applied are shown in Tables II and III.

TABLE II
RESULTS FROM THE APPLIED TECHNIQUES IN TRAINING SET

| | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| Adaptive Boosting | 80.34% | 83.57% | 77.12% |
| Decision Tree | 94.56% | 100% | 89.13% |
| Gradient Boosting | 85.75% | 90.84% | 80.67% |
| KNN | 80.64% | 74.27% | 87.02% |
| Logistic Regression | 71.12% | 68.75% | 73.49% |
| Random Forest | 94.29% | 98.11% | 90.47% |
| SVM | 93.01% | 96.91% | 89.09% |

TABLE III
RESULTS FROM THE APPLIED TECHNIQUES IN TEST SET

| | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| Adaptive Boosting | 87.27% | 87.27% | 87.27% |
| Decision Tree | 96.18% | 100.00% | 92.36% |
| Gradient Boosting | 91.90% | 94.54% | 89.27% |
| KNN | 80.81% | 74.54% | 87.09% |
| Logistic Regression | 66.01% | 62.54% | 69.45% |
| Random Forest | 95.90% | 99.81% | 92.00% |
| SVM | 95.36% | 99.09% | 91.63% |

If the primary objective is to avoid unnecessary CSF exams,

than sensitivity is an important feature of the algorithm. It measures the proportion of true positives in all cases that were forecasted as positive by the algorithm [4].

Analysing the results, we see that the Decision Tree is the preferable method to detect the True Positive in both data sets, Train and Test data.

Other algorithms also demonstrated high capacity for aiding Meningitis diagnosis: Gradient Boosting, Random Forest and SVM with good Sensitivity as well as Specificity in both Training and Testing data.

The classification algorithms we used estimate the probability that a certain individual belong to Class 1 (those with Meningitis). If the value of this probability is high, then the algorithm assigns this individual to Class 1 and, conversely, if the probability is low to Class 0. The threshold between low and high probability was automatically determined by the algorithm. The default assumption used was that the goal was to have the minimum number of misclassifications, with the same penalty for misclassifying a positive or a negative case. But it is not hard to investigate how these algorithms would perform if we simply change the classification threshold.

For a balanced data set, the Gradient Boosting algorithm had chosen a threshold of 0.14. The resulting Confusion matrix in Table IV shows that 32 (=9+23) misclassifications were made.

TABLE IV
GRADIENT BOOSTING - CONFUSION MATRIX

| | | *Real* | |
|---|---|---|---|
| | | *0* | *1* |
| *Predicted* | *0* | 379 | 9 |
| | *1* | 23 | 66 |

If we change the threshold, obviously, the Confusion matrix changes and the performance measures change. Since we depart from the default threshold, the number of misclassifications should increase but this is not necessarily bad.

Table V shows the results obtained using the Gradient Boosting algorithm with different thresholds.

TABLE V
GRADIENT BOOSTING WITH DIFFERENT THRESHOLDS

| | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|
| *Precision* | 70% | 76% | 82% | 88% | 97% | 100% | 100% | 100% |
| *Recall* | 90% | 87% | 79% | 74% | 66% | 56% | 31% | 13% |
| *F1-score* | 78% | 81% | 80% | 80% | 79% | 72% | 48% | 24% |

We noticed that, if the goal is to maximize the Precision, a threshold of 0.7 or higher should be used.

The following Confusion matrix (Table VI) was produced using the threshold of 0.7. This matrix indicates that there are zero cases of False Negatives. On the other hand, the number of False Positives increased to 39.

TABLE VI
GRADIENT BOOSTING - CONFUSION MATRIX WITH THRESHOLD = 0.7

| | | *Real* | |
|---|---|---|---|
| | | *0* | *1* |
| *Predicted* | *0* | 388 | 0 |
| | *1* | 39 | 50 |

For this particular case, 388 individuals would not have been required to take the CSF exam because the algorithm would have correctly inferred that they were free of Meningitis. Only 39 individuals without Meningitis would have taken the CSF exam because the algorithm "suspected" that they could be suffering with Meningitis. Therefore, the choice of the threshold is an important instrument to gauge the algorithm to produce a more, or less, conservative procedure.

Based on our case study, it seems that it would be possible to avoid unnecessary CSF tests to diagnose Meningitis in many cases without compromising the risk of not testing an ill individual.

## V. CONCLUSIONS

We have tested various ML techniques that could be used to help diagnosing Meningitis prior to the CSF exam.

We considered that in the data, all patients who had Meningitis were tested with CSF exam and that this exam does not produce false positives nor false negatives. Therefore, the register of a positive CSF result could be understood as a certainty on the occurrence of Meningitis.

When applying the ML techniques we use, as data, the results of blood and urine exams, and complains reports. The results of these exams and reports are not hard to obtain for any new patient and areis considerably less invasive and disturbing than the CSF exam.

We used data from the period of March/2014 to January/ 2017 to train the model. Data from February/2017 to September/2018 was used to test the model.

The Decision Tree model presented the best performance with 96.18% accuracy; 100% sensitivity and 92.36% sensibility. Therefore, we could not claim that the ML model is anywhere close to substitute the CSF exam. The prospect is that the ML model could be another aid to the doctor in deciding whether or not to submit a patient to the CSF exam to confirm or risk out the diagnosis of Meningitis.

It is important to mention that ML can also be used by healthcare professionals to systematically incorporate patients complains and other symptoms as qualitative inputs to help diagnosing Meningitis.

We have demonstrated that the use of balancing methods such as SMOTE and Cross Validation is important to improve the efficiency of ML classification algorithms. For example, when using the Decision Tree algorithm with unbalanced data,

the Accuracy of the Test data was 94.75%. After balancing the data, the Accuracy reached 96.18%.

ML classification algorithms produce estimates for the probability of a certain individual belonging to a certain class. If this probability is higher than a certain threshold, than the individual is assigned to that class. Different contexts call for different thresholds. For instance, if the cost of a false positive is very small, we should choose a small threshold. This is not the case for the Meningitis diagnosis since a false positive means that the patient would be required to take the CSF exam. For this reason we have studied the effects of changing the threshold. The results we obtained suggests that the ML algorithms could help prevent the majority of CSF exams without any apparent damage to the risk of false negatives.

## VI. Future works

One area of interest for future work would be to include the non-structured data obtained by registering patients' complaints in the ML model.

In this work, the accuracy of the CSF test was considered to be 100%. We also considered that the diagnosis is binary: the patient either has the disease or not. In practice, these two assumptions might be considered to be too strong and should be reexamined in future research.

The time-dependency of Meningitis occurrences was neglected and future work might show the existance of such dependency in predicting Meningitis.

## References

[1] A. Agresti, *Categorical data analysis.* John Wiley & Sons, 2003, vol.482.
[2] E. Alpaydin, *Introduction to machine learning.* MIT press, 2009.
[3] D. G. Altman, *Practical statistics for medical research.* CRC press, 1990.
[4] D. G. Altman and J. M. Bland, "Diagnostic tests. 1:(Sensitivity and specificity." *BMJ: British Medical Journal*, vol. 308, no. 6943, p. 1552, 1994.
[5] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp.175–185, 1992.
[6] P. Baldi, S. Brunak, and F. Bach, *Bioinformatics: the machine learning approach*. MIT press, 2001.
[7] A. G. Bonavigo, V. Gelinski, G. d. M. Costa, J. Plewka, and M. A. Costa, "Comparação entre a contagem manual e automatizada de células no líquido cefalorraquidiano," *Rev. bras. anal. clin*, vol. 41, no. 1, pp. 47–50, 2009.
[8] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[9] G. Caruso, L. Genovese, G. Maricchiolo, and A. Modica, "Haematological, biochemical and immunological parameters as stress indicatorsin dicentrarchus labrax and sparus aurata farmed in off-shore cages," *Aquaculture International*, vol. 13, no. 1-2, pp. 67–73, 2005.
[10] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote:synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
[11] D. Chopra, N. Joshi, and I. Mathur, *Mastering Natural Language Processing with Python*. Packt Publishing Ltd, 2016.
[12] E. F. Codd, "A relational model of data for large shared data banks," *Communications of the ACM*, vol. 13, no. 6, pp. 377–387, 1970.
[13] S. R. Comar, N. de Araújo Machado, T. G. Dozza, and P. Haas, "Análise citológica do líquido cefalorraquidiano," *Estudos de Biologia*, vol. 31, no. 73/75, 2009.
[14] R. J. Ferro and R. L. Makinistian, "El líquido cefalorraquídeo," *Publicación digital de la 1ra Cátedra de Clínica Médica y Terapéutica y laCarrera de Posgrado de especialización en Clínica Médica. Facultadde Ciencias Médicas-Universidad Nacional de Rosario*, 2011.
[15] Y. Freund, R. E. Schapireet al., "Experiments with a new boostingalgorithm," in *Icml*, vol. 96. Citeseer, 1996, pp. 148–156.
[16] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning.* Springer series in statistics New York, NY, USA:, 2001, vol. 1, no. 10.
[17] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.
[18] "Function to extract text in rtf files," Gilson Filho. (Online). Available:https://gist.github.com/gilsondev/7c1d2d753ddb522e7bc22511cfb08676
[19] M. J. Halvey and M. T. Keane, "An assessment of tag presentation techniques," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 1313–1314.
[20] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
[21] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.
[22] T. K. Ho, "Random decision forests," in *Document analysis and recogni-tion, 1995., proceedings of the third international conference* on, vol. 1.IEEE, 1995, pp. 278–282.
[23] N. Japkowicz, "The class imbalance problem: Significance and strategies," in *Proc. of the Int'l Conf. on Artificial Intelligence*, 2000.
[24] B. Kamiński, M. Jakubczyk, and P. Szufel, "A framework for sensitivity analysis of decision trees," *Central European journal of operations research*, vol. 26, no. 1, pp. 135–159, 2018.
[25] D. Karcher and R. McPherson, "Cerebrospinal, synovial, serous body fluids and alternative specimens," *Henry's clinical diagnosis and management by laboratory methods, 22th edition. Richard A. McPherson, Matthew R. Pincus eds. Elsevier Saunders, Philadelphia (PA)*, pp. 480–506, 2011.
[26] A. C. Lorena and A. Carvalho, "Introdução as máquinas de vetores suporte," *Relatório Técnico do Instituto de Ciências Matemáticas e de Computaçao (USP/Sao Carlos)*, vol. 192, 2003.
[27] R. A. Miller, K. F. Schaffner, and A. Meisel, "Ethical and legal issues related to the use of computer programs in clinical medicine," *Annals of Internal Medicine*, vol. 102, no. 4, pp. 529–536, 1985.
[28] M. Mohri, "*Foundations of machine learning lecture* 11."
[29] P. A. Morettin and W. O. BUSSAB, *Estatística básica.* Editora Saraiva, 2017.
[30] M. M. Mukaka, "A guide to appropriate use of correlation coefficient inmedical research," *Malawi Medical Journal*, vol. 24, no. 3, pp. 69–71, 2012.
[31] H. B. Neuman and E. R. Wald, "Bacteral meningitis in childhood at the children's hospital of pittsburgh: 1988-1998," *Clinical pediatrics*, vol. 40, no. 11, pp. 595–600, 2001.
[32] W. H. Organizationet al., "International classification of diseases (icd)," *http://www. who. int/classifications/icd/en/*, 2006.
[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion,O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourget al.,"Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
[34] M. Sabbatini, "Uso do computador no apoio ao diagnóstico médico," *Revista Informédica*, vol. 1, no. 1, pp. 5–11, 1993.
[35] X. Sáez-Llorens and G. H. McCracken Jr, "Bacterial meningitis inchildren," *The lancet*, vol. 361, no. 9375, pp. 2139–2148, 2003.
[36] M. I. Schinoni, "Fisiologia hepática," *Gazeta Médica da Bahia*, vol. 76, no. 2, 2008.
[37] V. V. Soares and L. J. E. de Souza Vieira, "Percepção de crianças hospitalizadas sobre realização de exames," *Rev Esc Enferm USP*, vol. 38, no. 3, pp. 298–306, 2004.
[38] M. N. Theodoridou, V. A. Vasilopoulou, E. E. Atsali, A. M. Pan-galis, G. J. Mostrou, V. P. Syriopoulou, and C. S. Hadjichristodoulou,"Meningitis registry of hospitalized cases in children: epidemiological patterns of acute bacterial meningitis throughout a 32-year period," *BMC Infectious Diseases*, vol. 7, no. 1, p. 101, 2007.
[39] L. Wilkinson and M. Friendly, "The history of the cluster heat map," *The American Statistician*, vol. 63, no. 2, pp. 179–184, 2009.
[40] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philipet al., "Top 10 algorithms in data mining," *Knowledge and information systems*, vol. 14, no. 1, pp.1–37, 2008.