

Modelling Temporal Relationships in *Pseudomonas Aeruginosa* Antimicrobial Resistance Prediction in Intensive Care Unit

Àlvar Hernández-Carnerero¹ and Miquel Sànchez-Marrè² and Inmaculada Mora-Jiménez³
and Cristina Soguero-Ruiz⁴ and Sergio Martínez-Agüero⁵ and Joaquín Álvarez-Rodríguez⁶

Abstract. In this paper, the prediction of antimicrobial resistance of *Pseudomonas aeruginosa* bacteria caused by nosocomial infections in the Intensive Care Unit (ICU) was considered. It was trained a *Logistic Regression* model using health records information from patients together with the history of past sensitivity tests (antibiograms). To predict the antimicrobial resistance for a certain patient, this study proposes to model the temporal relationships using bacterial information from the rest of the patients who are at the same time in the ICU. Furthermore, a training window with incremental size is used so that training set is always temporarily as near as possible to test instances to be predicted. Using these contributions, experiments show promising results to predict antimicrobial resistance even when few training data is available. From these results it is further inferred that resistant bacteria may be spreading among patients in the ICU and their populations rapidly mutate, changing the underlying data distribution, along time.

1 INTRODUCTION

Antimicrobial resistance has been increasing for decades, and the rate at which new antibiotics are synthesized is not as fast as it would be required to prevent this trend [7, 17]. A large proportion of infections caused by resistant bacteria occur during hospital stays, specially in the Intensive Care Unit (ICU) [16]. There, infection rates are much higher than in other hospital divisions [9]. This is due to its severely vulnerable population and to the high risk of becoming infected through multiple procedures and to the use of invasive devices

distorting the anatomical integrity-protective barriers of patients (intubation, mechanical ventilation, vascular access, etc.) [3].

It is frequent to find in the ICU some kinds of bacteria which can become multidrug-resistant. Among them, *Acinetobacter spp.*, *Enterococcus faecalis* and *Enterococcus faecium*, *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Staphylococcus aureus* are usually found. In this paper, we focus on *Pseudomonas aeruginosa* due to their prevalence and virulence. It is naturally resistant to many antibiotics and has a remarkable capacity for acquiring new resistance mechanisms, creating therapeutic problems [6]. *Pseudomonas aeruginosa* is considered to be multi-drug resistant (MDR) when it is observed a reduced in vitro susceptibility to three or more antimicrobial families [21].

It is known that infections due to MDR microorganisms are a major problem. This has a significant impact in the ICU, where they can cause additional morbidity, mortality, and health care costs [3, 11]. Inappropriate initial antimicrobial treatment of *P. aeruginosa* is statistically linked to a higher mortality compared to initial treatment with appropriate antimicrobial. The growing MDR rate of *P. aeruginosa* also increases the chance of inappropriate initial antimicrobial treatment [13].

To tackle with MDR in the hospital, a culture or microbiological analysis is usually performed to test whether the bacterium is resistant or susceptible to a set of antibiotics. For this purpose, first the germ (bacterium) is isolated and an antibiogram is carried out. The antibiogram represents the in vitro bacterium's resistance to a series of antibiotics. The set of antibiotics used in the antibiogram can be selected for the specific type of bacteria being tested. The result of the antibiogram is a vector of pairs antibiotic/sensitivity [4]. Antibiograms are often used by clinicians to assess bacteria susceptibility rates, as an aid in selecting empiric antibiotic therapy [10]. Hence, the antibiogram result could vary between bacterium species, depending on the resistance of a particular bacterium to different antibiotics. However, quite often groups of antibiotics still have similar sensitivity when tested on a given bacterium species, despite its strains [20].

In addition, the result of the antibiogram can help to reduce the bacteria spread by taking special measurements such as isolation of the patient. One of the most relevant factors of the spread of bacterial resistance is the so called cross-transmission [18], which may facilitate the spread of resistant bacteria from one patient to another. Also, some measures such as hand hygiene, skin cleansing, and contact precautions can help to prevent cross-transmission [18]. To know how and where to extreme caution, information about the kind of

¹ Campus del Poblenou, Universitat Pompeu Fabra, 08018 Barcelona, Spain, email: alvar.hernandez01@estudiant.upf.edu

² Knowledge Engineering and Machine Learning Group (KEMLG-UPC), Intelligent Data Science and Artificial Intelligence Research Centre (IDEAI-UPC), Dept. of Computer Science, Universitat Politècnica de Catalunya, Spain, email: miquel@cs.upc.edu

³ Department of Signal Theory and Communications, Telematics and Computing Systems, Rey Juan Carlos University, Madrid 28943, Spain, email: inmaculada.mora@urjc.es

⁴ Department of Signal Theory and Communications, Telematics and Computing Systems, Rey Juan Carlos University, Madrid 28943, Spain, email: cristina.soguero@urjc.es

⁵ Department of Signal Theory and Communications, Telematics and Computing Systems, Rey Juan Carlos University, Madrid 28943, Spain, email: sergio.martinez@urjc.es

⁶ Intensive Care Department, University Hospital of Fuenlabrada, Madrid 28942, Spain, email: joaquin.alvarez@salud.madrid.org

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). This volume is published and copyrighted by its editors. Advances in Artificial Intelligence for Healthcare, September 4, 2020, Virtual Workshop.

bacteria in the ICU and their resistance plays a key role. Since ICU patients have a critical health status and the antibiogram result can take from 24h to 48h [12], it is of major importance to develop tools which can help to anticipate this result. This would contribute not only to save patient’s lives but also to prevent the spread of a resistant bacteria.

Because of the aforementioned reasons, in the current study we propose to use a Data Mining (DM) technique, and specific temporal data processing to get a quick estimation of the antibiogram result. In this sense, many of the state-of-the-art studies regarding the use of DM methods to predict antimicrobial susceptibility are using whole genome sequencing [8, 15, 14, 1]. Despite it is a very promising technique, it involves very significant costs. As an alternative, we propose to use information from the ICU health records and demographic data of patients, along with historic antibiogram results to train a DM model, aiming to predict resistant bacteria in new cultures. As opposed to the whole genome sequencing, our approach intends to use data which are already available in the vast majority of hospitals, in order to speed up the process of identifying positive cases. Similar strategies have been analyzed in the past [12, 20, 19].

The remainder of this paper is structured as follows. Section 2 describes the dataset and the procedure to create new features. The experimental setup is established in Section 3, and results in Section 4. Finally, conclusions, limitations of the study, and suggested future research are presented in Section 5.

2 DATA DESCRIPTION AND PREPROCESSING

Data considered in this work is a unified and anonymized dataset specifically collected for the study of antimicrobial resistance in the ICU of the University Hospital of Fuenlabrada (UHF) in Spain. The data set covers the years from 2004 to 2013. During this time interval, 1914 patients were admitted to the ICU, and 22142 cultures were carried out from 2186 admissions. It has a number of 257 different types of bacteria and 26 antimicrobial families.

The dataset contains the results of antibiograms carried out to patients in the ICU, that is, the results of the sensitivity tests (susceptible (s); or resistant (r)) for certain pairs of bacteria and family of antibiotic used in the test. It also includes demographic data of the patients and information of their ICU admission: age, gender, date of ICU admission, clinical origin of the patient before ICU admission, reason for admission, patient category, comorbidities and pluripathology (it indicates whether a patient has more than two comorbidities).

As already mentioned, in this study we focused on just one type of bacteria among the multiple available in the data set: *P. aeruginosa*. This bacterium is considered MDR if it is resistant to three or more of the following antimicrobial families within the same culture: *Aminoglycosides*, *Carbapenems*, *4th Generation Cephalosporins*, *Extended-spectrum penicillins*, *Polymyxins* and *Quinolones*.

In this work, we analyze all instances containing the bacterium and antimicrobial families of interest. Then, each instance is represented by features described in Table 1. Features **c&car** and **c&cf4** represent the target to be predicted, this is, the result of the sensibility test for *P. aeruginosa* to the antimicrobial families of *Carbapenems* and *4th Generation Cephalosporins*, respectively. We consider only these antimicrobial families since this is the first approach to analyze the problem and allows us to reduce the scope of this study.

Table 1. Feature names and their description. Marked in bold the target features.

Feature name	Description
c&car	Result of the sensibility test to the <i>Carbapenems</i> family (r/s).
c&cf4	Result of the sensibility test to the <i>4th Generation Cephalosporins</i> family (r/s).
days_to_culture	Number of days elapsed from admission to the date of the culture.
date_culture	Date of the culture.
number_antibiotics	Number of antibiotics tested in the antibiogram.
culture_type	Type of culture performed (pharynx, urine, blood, etc.).
culture_type_grouped	Type of culture performed grouped (respiratory, urine, surface, etc.).
culture_type_grouped_2	Type of culture performed grouped (clinical sample/surface).
day_month_culture	Day of the month on which culture is carried out.
month_culture	Month on which culture is carried out.
year_culture	Year on which culture is carried out.
origin	Clinical origin before ICU admission.
reason_admission	Reason of admission at ICU.
goi_A	Group of illness A. Considers cardiovascular events.
goi_B	Group of illness B. Considers kidney failure, arthritis.
goi_C	Group of illness C. Considers respiratory problems.
goi_D	Group of illness D. Considers pancreatitis, endocrine.
goi_E	Group of illness E. Considers epilepsy, dementia.
goi_F	Group of illness F. Considers diabetes, arteriosclerosis.
goi_G	Group of illness G. Considers neoplasms.
pluripathology	Number of groups of illness to which patient belongs.
patient_category	Patient’s clinical category.
age	Patient age.
gender	Patient’s gender.
start_date	Date on which the patient’s admission begins.
day_week_admission	Day of the week on which the patient’s admission begins.
day_month_admission	Day of the month on which the patient’s admission begins.
month_admission	Month on which the patient’s admission begins.
year_admission	Year on which the patient’s admission begins.

2.1 Generation of new features

In addition to the selected features, we propose to generate a new kind of features based on the temporal information of cultures recorded in the data set.

The purpose of these features is to capture the presence of resistant bacteria in the ICU along time and the "intensity" of that presence. By "intensity", we consider the number of patients infected and the number of days since resistant bacterium was detected. For a specific instance of the data set, which represents a culture C_p of patient P_p , cultures containing *P. aeruginosa* have been collected for patients P_i ($C_i = \{C_{ij}\}$) between 21 days and 48 hours before the date of the culture C_p . These cultures exclude those associated to the patient P_p under analysis. Note that, since the results of the test usually takes 48 hours to be provided, it is not possible to use cultures taken, for instance, one hour ago. Apart from that, from a clinical viewpoint, if the culture result is positive, it is kept as positive for the next 21 days. For this reason, we consider cultures collected 21 days before the date D_p of the current culture C_p of patient P_p .

In total, six features are created using the information of the past cultures, one for each type of the antimicrobial families mentioned above: $r_{\&amg}$, $r_{\&car}$, $r_{\&cf4}$, $r_{\&pap}$, $r_{\&pol}$ and $r_{\&qui}$. For instance, feature $r_{\&amg}$ identifies only cultures tested for the *Aminoglycosides* family. The value of this feature is obtained taking into account the set of past cultures for all other patients (P_i), and excluding the patient under analysis (P_p). Each culture C_{ij} on date d_{ij} for patient P_i has the sensibility test result r_{ij} , which is 0 or 1 depending on whether the bacterium is susceptible or resistant to a specific family of antibiotics. In order to give more emphasis to the most recent cultures, we use a negative exponential function [2] to weight the culture results associated to each patient P_i as follows:

$$f_{C_{ij}}(D_p) = \begin{cases} 0 & \text{if } r_{ij} = 0 \\ n^{-(D_p - d_{ij})} & \text{if } r_{ij} = 1 \end{cases} \quad (1)$$

where, n is a real number experimentally set to 1.1. Then, to compute the value of each of the six features linked to patient P_p on date D_p , for each patient excepting P_p the maximum outcome in Eq. (1) is determined and added up according to Eq. (2).

$$FV_{C_p}(D_p) = \sum_{\forall P_i \neq P_p} \max_{C_i} f_{C_{ij}}(D_p) \quad (2)$$

2.2 Data preprocessing

To proceed with the model design for $c_{\&car}$ and $c_{\&cf4}$, we created two data subsets, one associated to each target. These subsets will be used to train a Logistic Regression (LR) model for each target. Training two different classifiers instead of, for instance a multi-class classifier, allows each classifier to be specialized in predicting its particular target, therefore tuning classifier's weights individually. In order to limit the study to MDR acquired in the ICU, only instances of patients admitted in the ICU for more than 48h are considered.

The final data set for the $c_{\&car}$ was composed by 450 cultures and 34 features including the target one, and the final data set for the $r_{\&cf4}$ was composed by 556 cultures and the same 34 features including the target one. In both data sets there are only missing values in six features. The percentages of missing values in those features are depicted in the Table 2.

Before training the models, we deal with missing data associated to the proposed features $r_{\&amg}$, $r_{\&car}$, $r_{\&cf4}$, $r_{\&pap}$, $r_{\&pol}$ and $r_{\&qui}$. Note that there may not be any culture in the ICU for *P.*

Table 2. Features with missing values (expressed in %) in both data sets.

Dataset	$r_{\&amg}$	$r_{\&car}$	$r_{\&cf4}$	$r_{\&pap}$	$r_{\&pol}$	$r_{\&qui}$
$c_{\&car}$	11.33	15.56	11.33	11.33	39.33	11.78
$c_{\&cf4}$	10.79	17.99	10.79	10.79	35.97	11.15

aeruginosa and a particular antimicrobial family for some time intervals, and therefore, a missing value is considered for that feature. This fact can be addressed following different approaches, such as deleting instances with missing data or imputing missing values [22]. In this study we propose an strategy based on the clinical meaning of the generated features. The smaller the value of these features, the fewer patients will have been infected with resistant *P. aeruginosa* bacteria and the greater the length of time since they were infected. As a result, if $r_{\&*}$ features do not have any value, it suggests that no *P. aeruginosa* has been recently detected in the ICU. Therefore, very likely no patients would have been recently infected with a resistant bacteria, and the value provided by Eq. (2) should be very small. In this case, missing values are replaced by a 0.

Afterwards, all categorical features are converted to binary following a one-hot encoding strategy, except for the features representing dates. Since dates have an intrinsic ordering, smaller numerical values are assigned to further dates in the past, and greater values correspond to more recent dates.

Finally, Pearson correlation between features (without considering the targets) is calculated in order to discard the most correlated features, since they provide very similar information. The methodology is as follows. When two features have a correlation coefficient higher than 0.9 or lower than -0.9, just one of them is randomly selected, discarding the rest. A visual representation of Pearson correlation between features for Carbapenems and 4th Generation Cephalosporins subsets is shown in Fig. 1. We can conclude that a similar correlation in patterns is found for both subsets. In both, there is just one group of features that are correlated more than 0.9, which are the following: `date_culture`, `year_culture`, `start_date` and `year_admission`. From these four correlated features, `date_culture` is selected because it is the most representative among them and the rest are discarded. Finally, both data sets had 31 features after this discarding.

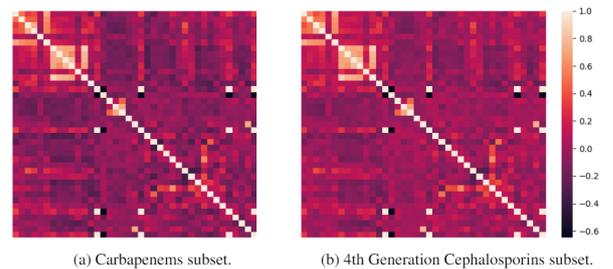


Figure 1. Visual representation of Pearson correlation between features for: (a) Carbapenems and (b) 4th Generation Cephalosporins subset.

3 EXPERIMENTAL SETUP

The first experiment, once data is processed, is to evaluate the relevance of the set of features selected and proposed, regarding the two different target features to be predicted. This is going to be analyzed

by using *Mutual Information* (MI) [5]. It is a quantity that measures the mutual dependence of the two variables, that is, it quantifies the amount of information that a random variable provides about other random variable. In terms of the probabilities, the MI of two jointly discrete random variables X and Y is calculated as:

$$I(X; Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p_{(X,Y)}(x, y) \log \left(\frac{p_{(X,Y)}(x, y)}{p_X(x) p_Y(y)} \right) \quad (3)$$

where $p_{(X,Y)}$ is the joint probability mass function of X and Y , and p_X and p_Y are the marginal probability mass functions of X and Y respectively.

Looking now at the prediction of the target, the type of the problem proposed have associated a series of special characteristics that have to be considered in order to properly address it.

First property in health records have an inherent *temporal ordering*. This forces, to use as training only instances that belong to a time prior to the test antibiograms that are to be predicted. In addition a margin of time has to be respected between train and test windows, since results of antibiograms are not immediately available after they are carried out. As before, in this particular case, a time margin of 48h needs to be considered.

The second particularity encountered when trying to predict is the *concept drift*. It is the fact that the concept of interest may depend on some hidden context, not given explicitly in the form of predictive features. Changes in the hidden context over time can induce more or less radical changes in the target concept. Changes in hidden context may not only result in a change of the target concept, but may also cause a change of the underlying data distribution [20]. In the particular domain of this study, antimicrobial resistance, the hidden context that changes over time are the mutations of bacteria, that allow them to be more resistant to antibiotics, as time passes by. In [20] it is proposed to use *instance selection* to handle concept drift as it is the technique that is most commonly used and has been found to offer good results. More specifically it is proposed to use a technique based on instance selection that consists in generalizing from a window that moves over recently arrived instances and uses the learnt concepts for prediction only in the immediate future. This represents a very good approach to apply to the resolution of the problem analyzed in this study except for the third particularity of the data set.

The data scarcity makes it difficult to learn from temporal windows containing several months, even years. In the first data set considered, the one with target feature `c&car` there are a total of 450 instances, and for the second data set with target feature `c&cf4` there are 556. Taking into account those data sets represent cultures from 10 years (from 2004 to 2013), the average of cultures per year is 45 and 56 respectively, which is a relatively small number of instances considering how fast ICU bacteria is able to mutate and change its sensitivity patterns. For this reason, we propose to use an *incremental window for training*, which increases its size as test window moves towards more recent instances. That is, the training window will be fixed from the first temporal instances, which are the oldest, and it will gradually increment in size, containing more instances, as more recent instances of test are predicted. The *test window*, on the other hand, will have a *fixed size* and it will progressively slide to select more recent instances.

The DM technique used in experiments is *Logistic Regression* (LR). It is chosen because of its simplicity to be used as a baseline. A baseline is a model that is both simple to set up and has a reasonable chance of providing acceptable results. LR is a technique which is used for the classification, and is used in this study to evaluate the

feasibility of learning the data. The classifier has to decide whether the target is sensitive or resistant. Therefore it is a binary decision.

To assess the performance of the proposed *incremental window framework*, experiments are carried out with different configurations. The characteristics of defined training and test windows are the following:

- For each experiment a set of LR classifiers is trained, each for a different training-test window.
- The size chosen for test windows is fixed in 3 months. Which is a relatively short time, near to the training instances, containing enough test instances.
- Different test windows do not overlap between them. That is, compared to the others, each test window contains different instances belonging to different time intervals.
- Instances in the training window do not contain antibiograms belonging to patients that also are present in the test window. For instance, if the result of an antibiogram of a particular patient is to be predicted in the test set, there are not past antibiograms of the same patient in the training set. That way, it is ensured that patients from training and test windows are different.

With each training and test pair of windows, a simple validation is performed to maintain the temporal order. The *class imbalance* due to the nature of the problem, causes that in the test window there might be more instances of one class than the other. To get a realistic approximation of the algorithm performance, the true negatives (success in sensitive instances) and the true positives (success in resistant instances) are calculated, together with the average between these two values and the general accuracy. For a particular test window with n_s sensitive instances and n_r resistant instances, if the method succeeds in predicting p_s sensitive instances and p_r resistant instances, the just mentioned values are calculated as:

$$\text{general_accuracy} = \frac{p_s + p_r}{n_s + n_r} \quad (4)$$

$$\text{resistant_success} = \frac{p_r}{n_r} \quad (5)$$

$$\text{sensitive_success} = \frac{p_s}{n_s} \quad (6)$$

$$\text{average_success} = \frac{\frac{p_r}{n_r} + \frac{p_s}{n_s}}{2} \quad (7)$$

These metrics offer a better approximation of the performance, because allow to track the success rate in the minority class label, which in many real problems is the most important one. For instance, if the test set counts with 8 sensitive and 2 resistant instances and the DM technique predicts all instances as sensitive, the general accuracy metric would pretty high value of a 80%, while it will be performing poorly in indentifying resistant antibiograms which are the ones most needed to detect.

Finally, to calculate the mean accuracy among several windows, an accumulation of the success rates is done, and the performance is evaluated using Equations 4, 5, 6 and 7. In other words, the performance of several windows is not calculated by averaging the individual performance values, but by accumulating the number of success instances in each window and using it to calculate the accuracy. This is done because test windows may have a different number of instances, due to the fact that not all 3-month time intervals have the same number of antibiograms. Therefore, making an average between their accuracy values would not be adequate since some instances would have more weight than others depending on the number of instances in their test window.

4 RESULTS

This section presents the results obtained after carrying out the experiments described in Section 3. In Section 4.1, the mutual dependence among features is analyzed by using MI. Sections 4.2 and 4.3 evaluate the impact of features, `date_culture` and `r*` respectively, on the prediction of bacteria resistance. The improvement in prediction obtained by using the incremental training window scheme is assessed in Section 4.4.

4.1 Features relevance using mutual information

The results of feature relevance using MI for *Carbapenems* and *4th Generation Cephalosporins* targets are represented in left and right columns of Table 3 respectively. The method MI computes the relevance or weight of one feature according to the co-occurrence of this feature and the target feature as described in equation 3. The MI method does not take into account the possible interaction of this feature with other ones regarding the target feature. In both cases, the feature `date_culture` is by far the variable containing most information about the feature to be predicted. Regarding *Carbapenems*, `date_culture` has a value of 0.53, while the second most important feature which is `age` just receives a relevance value of 0.18. In *4th Generation Cephalosporins* `date_culture` gets a value of 0.44 and `r&qui`, the second one 0.16. The importance of `date_culture` suggests that results of antibiograms are highly dependent on the time they were performed.

In addition, it is notable that five of the six proposed features `r*`, which consider antibiograms of other patients in the ICU, are between the eight most relevant features for both *Carbapenems* and *4th Generation Cephalosporins*. Therefore, we can infer they contain a great amount of information to predict antimicrobial resistance. The proposed feature with a considerably lowest importance is `r&pol`. This is probably caused because in the data set there are a smaller number of antibiograms for *Polymyxins* antimicrobial family (296) compared to the amount for *Aminoglycosides* (564), *Carbapenems* (450), *4th Generation Cephalosporins* (556), *Extended-spectrum penicillins* (558) and *Quinolones* (558).

Hence, one can infer that the result of the antibiogram for a particular patient is dependent on the past results of antibiograms of other patients in the ICU. To explain this fact we propose as a hypothesis that bacteria has been spreading from one patient to another in the ICU by cross-transmission. Therefore, the fact that a patient is infected with a resistant bacteria may increase the odds of another patient becoming infected as well.

4.2 Prediction contribution of `date_culture` feature

After observing that `date_culture` is the most important feature according to MI, the behavior of this significant variable is evaluated when predicting the result of antibiograms. To do that the target feature is predicted in two modes, one considering all features including `date_culture`, and a second one discarding `date_culture` from the features set. These predictions are made by using an incremental window for training and a 3-months sliding window for test as described in the experimental setup section. The training window starts containing 2004 and 2005 instances and the test window the first three months of 2006. After that, the training window increases three months and test window slides three months until the end of the database is reached. Results of the accumulated accuracy for all windows are shown in the two left columns of Table 4 for *Carbapenems*

Table 3. MI feature weighting for `c&car` and `c&cf4` target.

Feat. name	Wgts. c&car	Feat. name	Wgts. c&cf4
<code>date_culture</code>	0.5312	<code>date_culture</code>	0.4440
<code>age</code>	0.1866	<code>r&qui</code>	0.1637
<code>days_to_culture</code>	0.1713	<code>age</code>	0.1578
<code>r&qui</code>	0.1711	<code>days_to_culture</code>	0.1505
<code>r&pap</code>	0.1668	<code>r&pap</code>	0.1491
<code>r&amg</code>	0.1459	<code>r&amg</code>	0.1295
<code>r&car</code>	0.1382	<code>r&car</code>	0.1289
<code>r&cf4</code>	0.1228	<code>r&cf4</code>	0.1137
<code>day_month_admission</code>	0.1167	<code>day_month_admission</code>	0.1038
<code>origin</code>	0.0981	<code>number_antibiotics</code>	0.0799
<code>number_antibiotics</code>	0.0936	<code>origin</code>	0.0792
<code>reason_admission</code>	0.0860	<code>reason_admission</code>	0.0608
<code>day_month_culture</code>	0.0509	<code>month_admission</code>	0.0556
<code>month_culture</code>	0.0333	<code>goi_E</code>	0.0485
<code>culture_type</code>	0.0279	<code>day_month_culture</code>	0.0393
<code>month_admission</code>	0.0219	<code>day_week_admission</code>	0.0366
<code>day_week_admission</code>	0.0214	<code>culture_type</code>	0.0243
<code>culture_type_grouped</code>	0.0190	<code>r&pol</code>	0.0209
<code>r&pol</code>	0.0144	<code>pluripathology</code>	0.0204
<code>pluripathology</code>	0.0074	<code>month_culture</code>	0.0131
<code>gender</code>	0.0061	<code>culture_type_grouped</code>	0.0116
<code>goi_B</code>	0.0040	<code>patient_category</code>	0.0099
<code>goi_F</code>	0.0025	<code>gender</code>	0.0082
<code>goi_A</code>	0.0020	<code>goi_B</code>	0.0030
<code>culture_type_grouped_2</code>	0.0017	<code>goi_F</code>	0.0026
<code>goi_E</code>	0.0008	<code>goi_C</code>	0.0023
<code>goi_D</code>	0.0003	<code>goi_G</code>	0.0022
<code>patient_category</code>	4.5e-05	<code>culture_type_grouped_2</code>	0.0010
<code>goi_G</code>	2.0e-05	<code>goi_A</code>	6.1e-05
<code>goi_C</code>	2.9e-06	<code>goi_D</code>	4.1e-09

Table showing the weight given by MI to each of the features in descending order regarding its value. Left column contains values for *Carbapenems* antimicrobial family and right column for *4th Gen. Cephalosporins* family.

and in the two right columns of the same Table 4 for *4th Generation Cephalosporins*.

It is remarkable that in both cases the success percentage slightly increases when `date_culture` is not used. We state that this may be due precisely to the high influence this feature has on prediction of antimicrobial resistance. For some time intervals where the majority of instances belong to a particular label, this feature might be forcing the DM method to learn that, in this particular interval of values of `date_culture`, it is highly probable that the predicted instance belongs to the majority label. That is, it may be introducing some kind of bias towards the majority label of the time interval. This is reflected in the way *resistant success* and *sensitive success* varies when `date_culture` is removed. For *Carbapenems* target it can be seen that when using this feature, success in resistant instances increases and success in sensitive instances decreases. In *Carbapenems* data set there is a higher number of resistant instances (238) than sensitive instances (212). This seems to indicate that the majority class enhances its accuracy when using `date_culture`, and the minority class makes it worse. The same is observed for *4th generation Cephalosporins*. In *4th generation Cephalosporins* columns, the opposite situation happens, when `date_culture` is considered, Resistant success decreases and Sensitive success increases, and now the majority class are sensitive instances (350) as opposed to resistant instances (206). This apparently shows the bias `date_culture` feature introduces on classifying instances as the majority class, because of its high influence over the prediction.

In next experiments, `date_culture` is discarded since it slightly reduces the accuracy of the algorithm.

Table 4. Year 2006 to 2013 prediction for *Carbapenems* and *4th Gen. Cephalosporins* with incremental window.

Metric	Carbapenems		4th Gen. Ceph.	
	With	Without	With	Without
Accuracy	56.96	57.22	57.49	60.88
Resistant success	56.88	50.92	25.26	40.72
Sensitive success	57.06	64.97	77.85	73.62
Average	56.97	57.94	51.55	57.17

Comparison with and without using `date_culture` feature for prediction of years 2006 to 2013 for both `c&car` and `c&cf4` target features.

4.3 Prediction contribution of $r\&*$ features

A similar experiment is carried out for proposed $r\&*$ features. Accuracy metrics are calculated with and without them, for test years between 2006 and 2013. In Table 5 the results of this experiment are represented. The accuracy remains almost the same whether $r\&*$ features are used to predict or not, for both antimicrobial families. The difference is observed in resistant success and sensitive success. When $r\&*$ features are taken into account, these two metrics are more balanced, which means that both the majority class and minority class get a similar success rate, which is a desirable effect. Moreover, one can note that in both antimicrobial families, resistant success increases, which means that cultures from other patients are helping to better discriminate the resistant instances. Sensitive success decreases which is probably caused by the LR decision boundary, which after moving to better predict resistant instances is lowering its performance in recognizing sensitive instances.

Table 5. Year 2006 to 2013 prediction for *Carbapenems* and *4th Gen. Cephalosporins* with incremental window.

Metric	Carbapenems		4th Gen. Ceph.	
	With	Without	With	Without
Accuracy	57.22	57.97	60.88	60.68
Resistant success	50.92	47.25	40.72	30.41
Sensitive success	64.97	71.19	73.62	79.80
Average	57.94	59.22	57.17	55.11

Comparison with and without using the set of $r\&*$ features for prediction of years 2006 to 2013 for both `c&car` and `c&cf4` target features.

4.4 Prediction contribution of incremental window scheme

At last, the usefulness of the incremental training window and 3-month test window scheme is evaluated to assess whether it improves success metrics with respect to using training windows furthest from the test set, and the same 3-month test windows. In this experiment years 2012 and 2013 are predicted. In Table 6 success metrics for *Carbapenems* data sets are calculated, for 2012 and 2013. Figure 2 shows the number of test instances in each 3-month test window during the mentioned years. In the two left columns of Table 6 accuracy improves from a 59%, using a fixed window of training from 2004 to 2011, to a 68% when using the *incremental window* to predict 2012 instances. In the three columns on the right of Table 6 it is observed that, when predicting 2013, accuracy raises from a 64% when using training data until 2011 to a 71% when considering year 2012 as training too. When using the incremental window, the accuracy remains the same in a 71%. The reason why, in this case using the incremental window does not increase the accuracy with respect to training data until 2012 may be that the number of instances from first months of 2013 is small as it can be seen in Figure 2, and including them in the incremental training window may not increase the knowledge of the problem.

Table 6. Year 2012 and year 2013 prediction for *Carbapenems*.

Metric	2012		2013		
	04-11	Inc.	04-11	04-12	Inc.
Accuracy	59.09	68.18	64.29	71.43	71.43
Resistant success	57.14	66.67	57.14	85.71	85.71
Sensitive success	100.0	100.0	71.43	57.14	57.14
Average	78.57	83.33	64.29	71.43	71.43

In the two left columns, comparison of the year 2012 prediction using a fixed training window with instances from 2004 to 2011 and an *incremental training window*. In the three columns on the right, comparison of 2013 prediction using a fixed training window with instances from 2004 to 2011, a fixed training window from 2004 to 2012 and an *incremental training window*. The target feature predicted in all cases is `c&car`.

Table 7 and Figure 3, represent the results of the same experiment for *4th Generation Cephalosporins* data. In the two left columns of Table 7 accuracy for predicting 2012 is the same considering training instances until 2011 and using the incremental window. As before, that can be due to instances of first months of 2012 not providing enough knowledge to improve accuracy of prediction, although in this case there is a greater number of them as it is shown in Figure 3. The three columns on the right of Table 7 show that accuracy to pre-

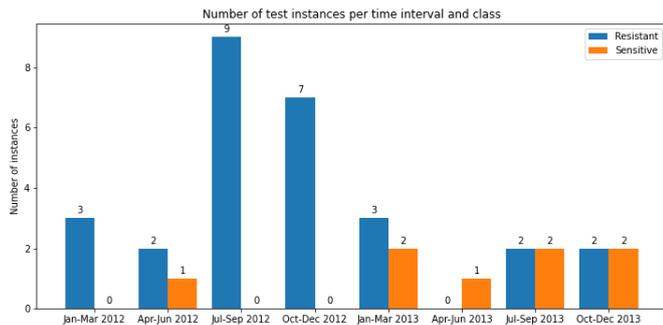


Figure 2. Number of test instances for *Carbapenems* in years 2012 to 2013 with 3 months granularity.

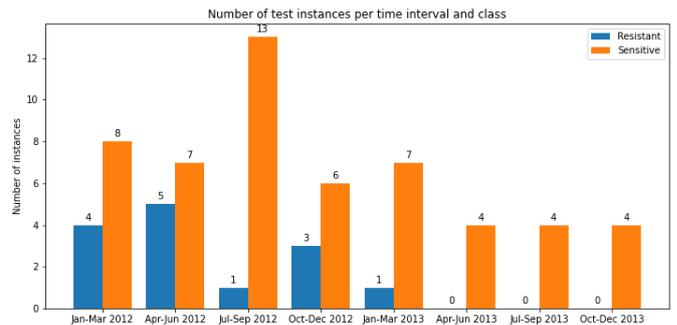


Figure 3. Number of test instances for *4th Gen. Cephalosporins* in years 2012 to 2013 with 3 months granularity.

dict 2013 is maintained in a 71% when making use of training data until 2011 and training data until 2012, which as we have just seen may be caused by 2012 instances not providing much information. Considering the *incremental window* for training, the percentage increases to an 80%.

These last experiments reveal that using training data temporarily as close as possible to the test set, always maintain the same success rates or improve them than using more distant data. Therefore, we conclude that the *incremental window* is a good scheme for the training set to make predictions of this particular problem. Also in these last experiments, in which test years used are 2012 and 2013, the accuracy values achieved are higher than the ones in previous experiments where the accuracy value was accumulated from 2006 to 2013. The fact that accuracy improves as the training set is bigger implies that the algorithm is able to learn the model and generalize from the training data.

Table 7. Year 2012 and year 2013 prediction for *4th Gen. Cephalosporins*.

Metric	2012		2013		
	04-11	Inc.	04-11	04-12	Inc.
Accuracy	59.57	59.57	75.0	75.0	80.0
Resistant success	15.38	15.38	0.0	0.0	0.0
Sensitive success	76.47	76.47	78.95	78.95	84.21
Average	45.93	45.93	39.47	39.47	42.11

In the two left columns, comparison of the year 2012 prediction using a fixed training window with instances from 2004 to 2011 and an *incremental training window*. In the three columns on the right, comparison of 2013 prediction using a fixed training window with instances from 2004 to 2011, a fixed training window from 2004 to 2012 and an *incremental training window*. The target feature predicted in all cases is `c&c&f4`.

5 CONCLUSIONS AND FUTURE WORK

In this paper we suggest to use health records and past antibiogram data to predict antimicrobial resistance in the ICU. We propose to use information about recently detected resistant germs in ICU as features of the data set. To handle changes in data distribution over time caused by progressive mutation of bacteria, we suggest to use an *incremental window for training set*, and a *test window with fixed size* such that training and test instances are temporarily as close as possible.

Experiments show that information of recent resistant bacteria detected in patients of the ICU, contains a relatively high amount of in-

formation to predict bacteria resistance in other patients, which could indicate that bacteria is spreading among ICU patients. It has also been observed that features providing the specific temporal ordering between all instances in the data set tend to decrease prediction accuracy. Lastly, experiments indicate that using an incremental window for training, maintain success rates or improve them. Therefore, we can conclude it is a scheme that improves the algorithm performance.

As future work we consider including further patient’s details about their admission, such as the antibiotics they have been administered, whether they have required intubation, if they have needed mechanical ventilation, among others, which are indicators that can have an impact on the appearance of resistant bacteria. Also, including past antibiogram information, for each particular patient whose sensibility test has to be predicted, would be an interesting approach to evaluate. To extend this study we propose to use other kind of DM algorithms different to LR, to assess whether they can improve success rates seen in this study. In addition, it would be advantageous to predict the resistance to the six antimicrobial families relevant to *P. aeruginosa* mentioned in the current study, so it would be possible to detect when a bacteria may become multidrug resistant.

ACKNOWLEDGEMENTS

We are thankful to University Hospital of Fuenlabrada, Madrid, Spain for providing the database used in this research.

This work has been partly supported by the Spanish Thematic Network “Learning Machines for Singular Problems and Applications (MAPAS)” (TIN2017-90567-REDT, MINECO/FEDER EU), by the IDEAI-UPC Consolidated Research Group Grant from Catalan Agency of University and Research Grants (AGAUR, Generalitat de Catalunya) (2017 SGR 574), by the Spanish Ministry of Economy, Industry and Competitiveness under the Research Project Klinilycs (TEC2016-75361-R), by the Science and Innovation Ministry Grants AAVis-BMR (PID2019-107768RA-I00) and BigTheory (PID2019-106623RB-C41), by the Spanish Institute of Health Carlos III (grant DTS 17/00158), by Project Ref. F656 financed by Rey Juan Carlos University, by the Young Researchers R&D Project Ref. 2020-661, financed by Rey Juan Carlos University and Community of Madrid (Spain), and by the Youth Employment Initiative (YEI) R&D Project Ref. TIC-11649 financed by the Community of Madrid (Spain).

REFERENCES

- [1] Gustavo Arango-Argoty, Emily Garner, Amy Pruden, Lenwood S Heath, Peter Vikesland, and Liqing Zhang, ‘Deeparg: a deep learning

- approach for predicting antibiotic resistance genes from metagenomic data', *Microbiome*, **6**(1), 1–15, (2018).
- [2] K Balakrishnan, *Exponential distribution: theory, methods and applications*, Routledge, 2018.
- [3] Nele Brusselselaers, Dirk Vogelaers, and Stijn Blot, 'The rising problem of antimicrobial resistance in the intensive care unit', *Annals of intensive care*, **1**(1), 47, (2011).
- [4] Rafael Cantón, 'Lectura interpretada del antibiograma: una necesidad clínica', *Enfermedades Infecciosas y microbiología clínica*, **28**(6), 375–385, (2010).
- [5] Thomas M Cover and Joy A Thomas. Elements of information theory, 2012.
- [6] C Defez, P Fabbro-Peray, N Bouziges, A Gouby, A Mahamat, JP Daurès, and A Sotto, 'Risk factors for multidrug-resistant pseudomonas aeruginosa nosocomial infection', *Journal of Hospital Infection*, **57**(3), 209–216, (2004).
- [7] Milislav Demerec, 'Origin of bacterial resistance to antibiotics', *Journal of bacteriology*, **56**(1), 63, (1948).
- [8] MJ Ellington, O Ekelund, Frank Møller Aarestrup, R Canton, M Doumith, Christian Giske, Hajo Grundman, Henrik Hasman, MTG Holden, Katie L Hopkins, et al., 'The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the eucast subcommittee', *Clinical microbiology and infection*, **23**(1), 2–22, (2017).
- [9] Håkan Hanberger, José-Angel Garcia-Rodríguez, Miguel Gobernado, Herman Goossens, Lennart E Nilsson, Marc J Struelens, et al., 'Antibiotic susceptibility among aerobic gram-negative bacilli in intensive care units in 5 european countries', *Jama*, **281**(1), 67–71, (1999).
- [10] S Joshi et al., 'Hospital antibiogram: a necessity', *Indian journal of medical microbiology*, **28**(4), 277, (2010).
- [11] Lisa L Maragakis, Eli N Perencevich, and Sara E Cosgrove, 'Clinical and economic burden of antimicrobial resistance', *Expert review of anti-infective therapy*, **6**(5), 751–763, (2008).
- [12] Sergio Martínez-Agüero, Inmaculada Mora-Jiménez, Jon Lérida-García, Joaquín Álvarez-Rodríguez, and Cristina Soguero-Ruiz, 'Machine learning techniques to identify antimicrobial resistance in the intensive care unit', *Entropy*, **21**(6), 603, (2019).
- [13] Scott T Micek, Ann E Lloyd, David J Ritchie, Richard M Reichley, Victoria J Fraser, and Marin H Kollef, 'Pseudomonas aeruginosa bloodstream infection: importance of appropriate initial antimicrobial treatment', *Antimicrobial agents and chemotherapy*, **49**(4), 1306–1311, (2005).
- [14] Marcus Nguyen, S Wesley Long, Patrick F McDermott, Randall J Olsen, Robert Olson, Rick L Stevens, Gregory H Tyson, Shaohua Zhao, and James J Davis, 'Using machine learning to predict antimicrobial mics and associated genomic features for nontyphoidal salmonella', *Journal of clinical microbiology*, **57**(2), (2019).
- [15] Mitchell W Pesesky, Tahir Hussain, Meghan Wallace, Sanket Patel, Saadia Andleeb, Carey-Ann D Burnham, and Gautam Dantas, 'Evaluation of machine learning and rules-based approaches for predicting antimicrobial resistance profiles in gram-negative bacilli from whole genome sequence data', *Frontiers in microbiology*, **7**, 1887, (2016).
- [16] Paz Revuelta-Zamorano, Alberto Sánchez, José Luis Rojo-Álvarez, Joaquín Álvarez-Rodríguez, Javier Ramos-López, and Cristina Soguero-Ruiz, 'Prediction of healthcare associated infections in an intensive care unit using machine learning and big data tools', in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, pp. 840–845. Springer, (2016).
- [17] AD Russell, 'Antibiotic and biocide resistance in bacteria: introduction', *Journal of applied microbiology*, **92**, 1S–3S, (2002).
- [18] Jean-Francois Timsit, Stephan Harbarth, and Jean Carlet. De-escalation as a potential way of reducing antibiotic use and antimicrobial resistance in icu, 2014.
- [19] ML Tlachac, Elke A Rundensteiner, Kerri Barton, Scott Troppy, Kirthana Beaulac, and Shira Doron, 'Predicting future antibiotic susceptibility using regression-based methods on longitudinal massachusetts antibiogram data.', in *HEALTHINF*, pp. 103–114, (2018).
- [20] Alexey Tsymbal, Mykola Pechenizkiy, Padraig Cunningham, and Seppo Puuronen, 'Handling local concept drift with dynamic integration of classifiers: Domain of antibiotic resistance in nosocomial infections', in *19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06)*, pp. 679–684. IEEE, (2006).
- [21] Lindsey M Weiner, Amy K Webb, Brandi Limbago, Margaret A Dudeck, Jean Patel, Alexander J Kallen, Jonathan R Edwards, and Dawn M Sievert, 'Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the national healthcare safety network at the centers for disease control and prevention, 2011–2014', *infection control & hospital epidemiology*, **37**(11), 1288–1301, (2016).
- [22] Shichao Zhang, Xindong Wu, and Manlong Zhu, 'Efficient missing data imputation for supervised learning', in *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*, pp. 672–679. IEEE, (2010).