Antimicrobial Resistance Prediction in Intensive Care Unit for *Pseudomonas Aeruginosa* using Temporal Data-Driven Models

Àlvar Hernàndez-Carnerero[1]*, Miquel Sànchez-Marrè[1], Inmaculada Mora-Jiménez[2], Cristina Soguero-Ruiz[2], Sergio Martínez-Agüero[2], Joaquín Álvarez-Rodríguez[3]

[1] Intelligent Data Science and Artificial Intelligence Research Centre (IDEAI-UPC), Dept. of Computer Science, Universitat Politècnica de Catalunya, 08034 Barcelona (Spain)
[2] Department of Signal Theory and Communications, Telematics and Computing Systems, Rey Juan Carlos University, Madrid 28943 (Spain)
[3] Intensive Care Department, University Hospital of Fuenlabrada, Madrid 28942 (Spain)

Received 4 January 2021 | Accepted 29 January 2021 | Published 25 February 2021

Abstract

One threatening medical problem for human beings is the increasing antimicrobial resistance of some microorganisms. This problem is especially difficult in Intensive Care Units (ICUs) of hospitals due to the vulnerable state of patients. Knowing in advance whether a concrete bacterium is resistant or susceptible to an antibiotic is a crux step for clinicians to determine an effective antibiotic treatment. This usual clinical procedure takes approximately 48 hours and it is named antibiogram. It tests the bacterium resistance to one or more antimicrobial families (six of them considered in this work). This article focuses on cultures of the *Pseudomonas Aeruginosa* bacterium because is one of the most dangerous in the ICU. Several temporal data-driven models are proposed and analyzed to predict the resistance or susceptibility to a determined antibiotic family previously to know the antibiogram result and only using the available past information from a data set. This data set is formed by anonymized electronic health records data from more than 3300 ICU patients during 15 years. Several data-driven classifier methods are used in combination with several temporal modeling approaches. The results show that our predictions are reasonably accurate for some antimicrobial families, and could be used by clinicians to determine the best antibiotic therapy in advance. This early prediction can save valuable time to start the adequate treatment for an ICU patient. This study corroborates the results of a previous work pointing that the antimicrobial resistance of bacteria in the ICU is related to other recent resistance tests of ICU patients. This information is very valuable for making accurate antimicrobial resistance predictions.

Keywords

Antimicrobial Resistance, Intensive Care Unit, Prediction, *Pseudomonas Aeruginosa*, Temporal Data-Driven Modeling.

DOI: 10.9781/ijimai.2021.02.012

I. Introduction

Antimicrobial resistance occurs when a germ develops the capacity to not respond to the drugs designed to combat them [1]. Nowadays, antimicrobial resistance is one of the greatest threats to the global health system [2]. Apart from the health consequences, the economic impact deriving from antimicrobial resistance is not a trivial issue, resulting in a 7% reduction in the Gross Domestic Product by 2050 [3]. Indeed, it has become more acute in recent years due to the excessive use of antibiotics in many facets of daily life [4].

The acquisition of antimicrobial resistance is favoured in hospital environments, being even worsened for patients admitted to the Intensive Care Unit (ICU). This could be motivated by the duration and intensity of the drug treatment, as well as by the use of life support devices. The critical health status of ICU patients pushes actions to anticipate the result of the cultures provided by the microbiology laboratory, which usually takes 48 hours. A culture is a biological sample collected to isolate a bacterium, aiming to analyze its susceptibility to different antibiotics. The test used to measure this susceptibility is called antibiogram, and its result (susceptible/resistant) is commonly used by clinicians to determine the antibiotic treatment [5]. It is interesting to note that several families of antibiotics may have similar susceptibility when tested on a given germ species [6]. There are several species with high prevalence, for example, Acinetobacter spp.; *Enterococcous fecalis* and *Enterococcus faecium*; *Escherichia coli*; *Klebsiella pneumoniae*; *Pseudomonas aeruginosa*; and *Staphylococcus aureus*, among others. In this paper, we focus on *Pseudomonas aeruginosa* for the following reasons: (1) its virulence, specially in the ICU;(2) its ability to cause chronic infectious diseases; and (3) its ability to develop multi-drug resistance [7], [8].

For all these reasons, anticipation to the culture result in case of resistance, is vital to isolate the patient and control the spread of

* Corresponding author.
E-mail address: alvar.hernandez@upc.edu

antimicrobial resistance among other ICU patients. Computational tools inspired on data-driven models may be supportive to clinical decisions previous to the antibiogram result. The article [6] introduces the *concept drift* observed in antimicrobial resistance data sets, and it uses a windowing scheme together with dynamic classifiers to perform resistance prediction. It classifies cultures as susceptible or resistant to some antibiotics using a database of EHR which includes years from 2002 to 2004, considering cases of meningitis. A high number of the state-of-the-art studies use whole genome sequencing [9]–[12]. Because of its considerable cost, in this study we propose to predict resistant bacteria based on Electronic Health Records (EHR) data from ICU, together with historic antibiogram results. This data is already available in most hospitals, and therefore the methodology proposed in this paper can be straightforward extrapolated. Comparable approaches are studied in previous works [6], [13]–[18]. In [17], bacterial infection in the ICU using EHR data is predicted (binary classification task) by applying a set of machine learning (ML) methods. The prediction is carried out at the patient level in order to determine which patients no longer require more antimicrobial treatment. Longitudinal data from 2001 to 2012, extracted over the 24-hour, 48-hour or 72-hour window following their first antibiotic dose, are considered. No temporal modelling was explicitly taken into account. The work in [18] presents an study for predicting bacterial resistance also using EHR data, from 2013 to 2015. An ensemble of ML methods is used to classify isolated bacterial cultures as susceptible or resistant to a particular antibiotic. The temporal relation among instances is considered here, with features indicating the proportion of past antibiotic resistance infections identified as having the highest average impact. This study also concludes that the feature encoding the date of the culture has some effect on the prediction, probably due to the fluctuating resistance frequencies through time.

Owing to the dynamics of antimicrobial resistance, we analyze in this paper electronic health records collected during 15 years, from 2004 to 2019, by the University Hospital of Fuenlabrada (UHF) in Madrid, Spain. This data have been partly considered in previous studies carried out by the authors [14], [15], [16],[19]. In particular, authors in [14] used a reduced dataset taking into account two years less (from 2004 to 2017) than in the current work. All patients admitted in the ICU in this period were considered in [14], regardless of their length stay. Additionally, authors in [14] used ML to determine whether a *Pseudomona Aeruginosa* bacterium will be resistant or not (binary target) to different families of antimicrobials without considering information about historic antibiogram results. In [15], we analyzed for the first time the dynamics on *Pseudomona Aeruginosa* by considering incremental time windows on a period of time from 2004 to 2013, with two families of antibiotics. It was also our first incursion on the use of features taking into account the result provided by previous antibiograms of other ICU patients. This current paper extends the work in [15] while considering the predictive window length (one month) that best results provided in [15]. Specifically, to carry out predictions, the Random Forest (RF) method has been added to previously considered method, Logistic Regression (LR). We have increased both the number of years under study and the number of antibiotics (from 2 to 6). We have also considered as features the result provided by previous antibiograms of each patient, weighted by a factor depending on the time elapsed since the last antibiogram was tested. Furthermore, two approaches have been explored to analyze the dynamic of antimicrobial resistance by evaluating the models in several time horizons.

The rest of the paper is as follows. In Section II, we describe the data set analyzed in this paper and provide a graphical exploration of it. Section III introduces the data preprocessing as well as the methods used for temporal modelling. Results and discussion and provided in IV. Finally, the conclusions are presented in Section V.

II. Materials

A. Data Set Description

Data considered in this work correspond to 3812 admissions of 3346 ICU patients, collected at the UHF during a period of 15 consecutive years (from July 2004 to May 2019). Note that, since the number of ICU admissions exceeds the number of patients, there are patients with more than one ICU admission during this period. A total of 43658 cultures were collected. Although there are more than 290 different types of bacteria and 27 antimicrobial families, we only take into account here the cultures where *Pseudomonas* have been detected, ending up in a total of 764 cultures. For this bacterium, the antibiograms considered in this work test the response (encoded as susceptible (s) or resistant (r)) against the following set of family of antibiotics $a = \{\texttt{amg}, \texttt{car}, \texttt{cf4}, \texttt{pap}, \texttt{pol}, \texttt{qui}\}$. Elements in the set a refer to: Aminoglycosides (AMG), Carbapenems (CAR), 4th Generation Cephalosporins (CF4), Extended-spectrum penicillins (PAP), polymyxins (POL) and Quinolones (QUI), respectively.

Since data-driven models are based on learning from instances, we consider here the target $\texttt{c\&a}_i$, as the antibiogram result for a specific family of antibiotic a_i, for every culture collected to any patient. The feature vector associated with each target is represented by the 40 features described in Table I. We define here the instance as the pair composed by the feature vector (input features to the data-driven models) and the target (outcome of the data-driven models).

TABLE I. Name and Description of the Features Characterizing Each Instance for Every Family of Antibiotics (AMG, CAR, CF4, PAP, POL, QUI), Tested on a particular Patient P. The Result for the Antibiogram Family a_i Is Encoded in the Binary Target Feature $\texttt{c\&a}_i$ (Not Presented in this Table)

Feature name	Description
age	age of the patient
gender	gender of the patient
origin	clinical origin before ICU admission
goi_*	7 features, each linked to a different group of
pluripathology	illness *: A, B, C, D, E, F, G
	number of groups of illness
patient_category	clinical category of the patient
reason_admission	reason of admission at ICU
start_date	date the patient was admitted
day_week_admission	day of the week the patient was admitted to the ICU
day_month_admission	day of the month the patient was admitted to the ICU
month_admission	month the patient was admitted
year_admission	year the patient was admitted
date_culture	date of the culture
day_week_culture	weekday the culture was collected
day_month_culture	day of month the culture was collected
month_culture	month the culture was collected
year_culture	year the culture was collected
culture_type	type of culture
culture_type_group1	1st level grouping for the culture type
culture_type_group2	2nd level grouping for the culture
days_to_culture	type number of days elapsed from start_date to date_culture
$\texttt{p\&a}_i$	6 features, each linked to one family a_i of previous antibiograms of patient p: amg, car, cf4, pap, pol and qui
$\texttt{r\&a}_i$	6 features, each linked to one family a_i of previous antibiograms for other recent patients different from p: amg, car, cf4, pap, pol and qui

As for the input features, we first analyze **demographic data**: age, gender, group of illness A (cardiovascular events), B (kidney failure, arthritis), C (respiratory problems), D (pancreatitis, endocrine), E (epilepsy, dementia), F (diabetes, arteriosclerosis) and G (neoplasms), and pluripathology (indicating whether the patient has more than two comorbidities). The median age of patients admitted to the ICU was 64 years (interquartile range 55-73, range 18-87), with a majority of men (70%). Pluripathological patients are 40.6% of the patients, with comorbidities mostly related to respiratory problems (33.4%), diabetes (26.3%) and neoplasms (33.1%).

We then focus on the **information about the ICU admission**: date of admission to the ICU, department of origin before ICU admission (surgery, internal medicine, urology,...), reason for admission (serious infection, acute respiratory failure, hypovolaemia,...) and patient category (medical or surgical). The clinical origin before the ICU admission more common was surgery (31.1% patients) and emergency department (18.4%). The reason of admission more common was serious infection (22.5% patients) and acute respiratory failure (18.4% patients). The most common patient category was medical (52.2 %).

This work also analyses the information related to the **cultures**. Specifically, we consider the culture type (exudate, drainage, biopsy, sputum, bronchoaspirate, etc.); first level grouping for the type of culture, which classifies the cultures into surface, liquids, respiratory, etc.; and the second level grouping for the type of culture, used to identify a clinical sample or a surface culture. Besides, the date of the culture, the weekday the culture was collected , as well as the month and the year.

Finally, to collect temporal information in each instance associated to patient p, the current study proposes to generate two kind of features linked to previous resistant antibiograms. In particular, we consider: (1) previous resistant results of the same patient, and (2) previous resistant results of all patients who recently stayed in the ICU.

Own past cultures features. The first kind of features is associated with the detection of resistant bacteria in previous antibiograms for a specific patient p, and aims to quantify the current "*intensity*" of these bacteria. These features consider the result of antibiograms of *Pseudomonas Aeruginosa* during an interval between 21 days and 48 hours previous to the current culture being studied for patient p, $c^{(p)}$. The 48-hour limit is considered since it is usually the time the results of the antibiogram take to be available. Furthermore, cultures are gathered until 21 days before the date d of current culture c, because if the antibiogram result is positive, from a clinical point of view, it is kept as positive for the following 21 days.

Thus, when a culture is collected, a total of six features, one per antimicrobial family, are generated: `p&amg`, `p&car`, `p&cf4`, `p&pap`, `p&pol` and `p&qui`. Each feature takes into account the antibiogram results for the corresponding antimicrobial family, e.g. `p&pap` just consider previous results associated with patient p for the family of antibiotics PAP. Because of that, the group of own past cultures of patient p, named $C^{(p)}$, is divided into six data sets $C_{a_i}^{(p)}$. To illustrate how the value for each feature $\texttt{p\&a}_i$, i = 1, 2, ⋯, 6 is obtained, let us consider that the subset $C_{a_i}^{(p)}$ has $n_{a_i}^{(p)}$ cultures, i.e. $C_{a_i}^{(p)} = \{ c_{a_i,k}^{(p)} \}_{k=1}^{n_{a_i}^{(p)}}$. Each culture $c_{a_i,k}^{(p)} \in C_{a_i}^{(p)}$ has associated: (1) a date $d_{a_i,k}^{(p)}$ when it was collected ; and (2) a susceptibility test result $r_{a_i,k}^{(p)}$, which is *susceptible* or *resistant* depending on whether the bacterium is susceptible or resistant to a_i, respectively. To calculate the potential contribution of a culture $C_{a_i}^{(p)}$ to the feature $\texttt{p\&a}_i$, i = 1, 2, … , 6, the Negative Exponential Function (NEF) is applied as follows:

$$\text{NEF}\left(c_{a_i,k}^{(p)}, c^{(p)}\right) = \begin{cases} 0 & \text{if } r_{a_i,k}^{(p)} = \text{susceptible} \\ e^{-\lambda\left(d - d_{a_i,k}^{(p)}\right)} & \text{if } r_{a_i,k}^{(p)} = \text{resistant} \end{cases} \quad (1)$$

where the value of parameter λ is experimentally set to 0.095. To compute the feature value $\texttt{p\&a}_i$ for the instance associated with culture $c^{(p)}$ of patient p, the maximum outcome in Equation (1) is obtained according to Equation (2):

$$\texttt{p\&a}_i = \max_{k=1,\cdots,n^{(p)}} \text{NEF}\left(c_{a_i,k}^{(p)}, c^{(p)}\right) \quad (2)$$

ICU-patients past cultures features. The second kind of features are named `r&amg`, `r&car`, `r&cf4`, `r&pap`, `r&pol` and `r&qui`. These features aim to encode the "*intensity*" of resistant bacteria in the ICU during the time previous to the date d of the current instance and culture. Differently from the previous set of six features $\texttt{p\&a}_i$, now the "*intensity*" takes into account the number of patients (different from current patient p) that were infected by a resistant bacterium and, for each of them, the time elapsed since the bacterium was detected. For a particular feature, a single value is calculated by considering the result of past susceptibility tests of *Pseudomonas Aeruginosa* for the P patients, denoted as p_j with $j = 1 \cdots P$, in the ICU during the time interval between 21 days and 48 hours previous to date d of culture $c^{(p)}$ of patient p. An exponential decay is again considered to weight the result of each susceptibility test.

The group $C^{(p')}$ of past cultures of other patients is divided into six subsets $C_{a_i}^{(p')}$ too. Every particular subset $C_{a_i}^{(p')}$ is split into n disjoint subsets, as many as patients:

$$C_{a_i}^{(p')} = \bigcup_{j=1}^{P} C_{a_i}^{(p_j)}, \qquad C_{a_i}^{(p_j)} = \left\{ c_{a_i,k}^{(p_j)} \right\}_{k=1}^{n_{a_i}^{(p_j)}} \quad (3)$$

where $C_{a_i}^{(p_j)}$ is composed of the $n_{a_i}^{(p_j)}$ antibiogram results for a_i in patient p_j. As previously mentioned, the set of cultures of patient p are excluded from $C_{a_i}^{(p')}$.

Since every culture $c_{a_i,k}^{(p_j)}$ has a susceptibility test result $r_{a_i,k}^{(p_j)}$ and a date $d_{a_i,k}^{(p_j)}$, the application of the NEF expression equivalent to that in Equation (1), just replacing $c_{a_i,k}^{(p_j)}$, $d_{a_i,j}^{(p)}$ and $r_{a_i,k}^{(p)}$ by $c_{a_i,k}^{(p_j)}$, $d_{a_i,k}^{(p_j)}$ and $r_{a_i,k}^{(p_j)}$, respectively. Then, each feature $\texttt{r\&a}_i$ is obtained by adding up the maximum value of Equation (1) for each patient p_j, as indicated in Equation (4).

$$\texttt{r\&a}_i = \sum_{j=1}^{P} \max_{k=1,\cdots,n_{a_i}^{(p_j)}} \text{NEF}\left(c_{a_i,k}^{(p_j)}, c^{(p)}\right) \quad (4)$$

B. Graphical Exploration

Owing to the high number of features, we start by identifying the most relevant features per family of antibiotics. For this purpose, we consider a filter approach with the Mutual Information (MI) score [20]. Thus, for each family of antibiotics, Fig. 1 shows the five features with the highest MI values, comprising among them the date of culture and the information about the previous cultures both for the own patient and for the UCI environment. According to the mutual information score, the most relevant feature is `date_culture` for each of the antimicrobial families considered. This results supports the importance of the antimicrobial resistance dynamics, which is common for all families of antibiotics.

To get a deeper insight on this issue, Fig. 2 graphically illustrates the evolution of the number of susceptible antibiograms (a) and resistant antibiograms (b) for each family of antimicrobials tested on *Pseudomonas* along time. Not all families of antibiotics were tested during the whole period considered. Specifically, clinicians first agreed to modify the range of tested antibiotics in the ICU of the UHF, first by including POL in 2007 and then by stop susceptibility testing antibiograms of QUI in 2018, due to its high resistance. Furthermore, there is a very noticeable fall in the number of resistant and susceptible

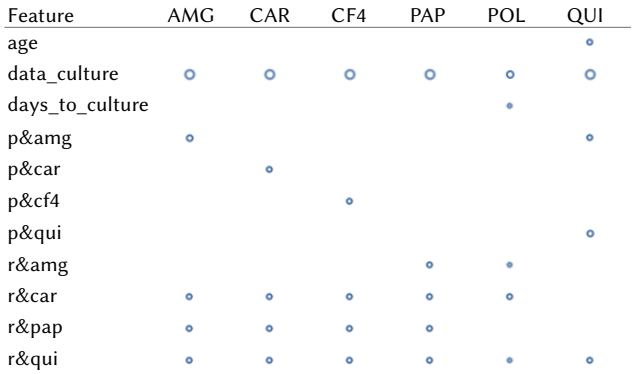
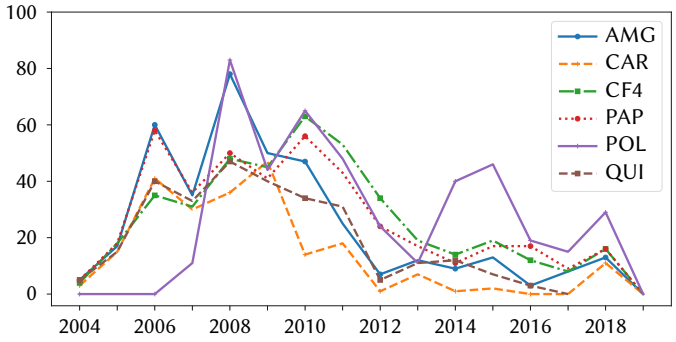
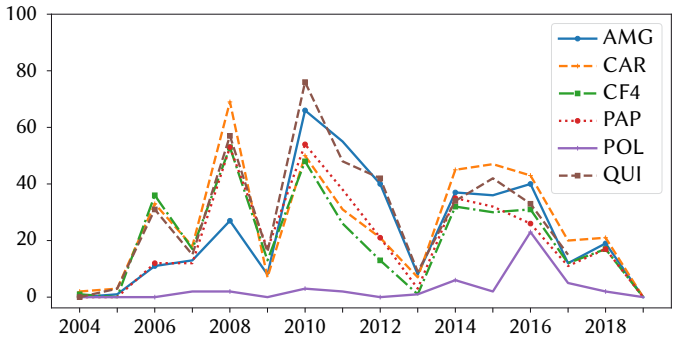


Fig. 1. For each antimicrobial family, the five features with the highest MI scores, indicated by the circle size from MI=0.56 (biggest size, pair date_culture-AMG) to MI=0.09 (smallest size, pair r&amg-POL).



(a)



(b)

Fig. 2. Temporal evolution for the number of annual susceptible (a) and resistant (b) antibiograms when tested on *Pseudomonas* cultures for each family of antimicrobials.

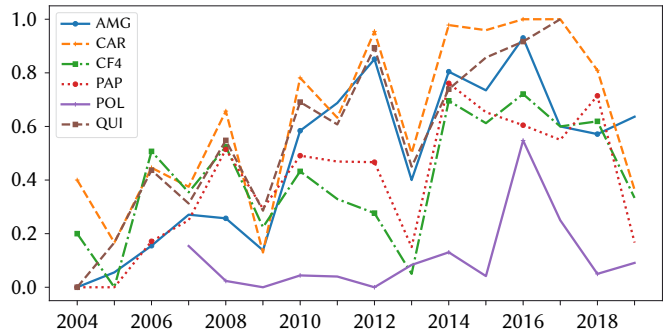


Fig. 3. Temporal evolution of the ratio between the number of annual resistant antibiograms tested on *Pseudomonas* cultures and the total annual number of cultures on *Pseudomonas* for each family of antimicrobials.

antibiograms in 2013. This decrease is probably motivated because of integration problems due to software update in the ICU health information system in 2013. As stated in the literature, the number of susceptible antibiograms tend to decrease in the most recent years.

In this line, we also analyze the annual ratio of resistant antibiograms results for each family of antimicrobials. To obtain this ratio, the number of resistant cultures per year has been divided by the number of total cultures per year (both resistant and susceptible cultures). The general trend is that, as time progresses (and therefore the value of `date_culture` increases), a higher percentage of instances tend to be resistant.

The second most relevant feature for the antimicrobial families AMG, CAR and QUI are `p&amg`, `p&car` and `p&qui`, respectively. This shows the importance of the outcome of previous antibiograms of the same patient for the family under consideration. In the case of CF4, `p&cf4` is the 4th most important feature. Though not presented in Fig. 1, `p&pap` is ranked on the 7th position for PAP, and `p&pol` in the 11th position for POL. It is interesting to remark here that, in all cases, the MI score for a particular family of antibiotics is higher for the $p\&a_i$ feature corresponding to that particular family than to any of the other five $p\&a_i$ features. This points out the relevance of considering the particular antimicrobial family when using results of previous antibiograms.

Fig. 4 shows the boxplots for each of the six features named $p\&a_i$, associated to the antibiogram results of the same patient for each family of antibiotics (in rows). Blue boxplots refer to $p\&a_i$ for resistant results, while black ones refer to $p\&a_i$ for susceptible results. In general, we observe that the median of $p\&a_i$ is higher when the culture c was resistant than when it was susceptible. The results shown in Fig. 4 for CAR and QUI are particularly interesting for susceptible cultures (black boxplots) for all the families, with most of the previous antibiogram results being susceptible. However, for CF4 and PAP, most of antibiogram results are susceptible for `p&cf4`, `p&pap` and `p&pol`, whereas for POL it only happens for `p&pol`. Note that, regardless the family of antibiotics tested, the boxplot of `p&car` and `p&qui` for resistant cultures (blue bloxplots) is very similar to the boxplot associated to the corresponding family of antibiotics considered (e.g, see `p&amg`, `p&car` and `p&qui` in Fig. 4 for AMG, or `p&pap`, `p&car` and `p&qui` for PAP.

The $r\&a_i$ features are also among the most relevant features according to the MI score. In this case there is no clear distinction on the ranking depending on the antimicrobial family. It supports the importance of taking into account the existence of any resistant germ in the ICU. The feature `r&pol` (not included among the top five features in Table I) seems to be the one providing less information, probably because of low number of antibiograms with a resistant result for this family. Fig. 5 presents the bloxplots for the $r\&a_i$ features. In comparison with boxplots in Fig. 4, note that boxplots of the $r\&a_i$ features are not limited to a maximum of one, since the number of patients contributing in Equation (4) is n (usually greater than 1). For each antibiogram a_i, the median values of the $r\&a_i$ features resistant and susceptible results is much closer between them than when comparing the $p\&a_i$ features. It is also remarkable that boxplots associated with `r&pol` show a median value very close to zero both for resistant and susceptible cultures, in line with previous comments. Furthermore, when analyzing POL, the median value is higher for susceptible than for resistant cultures, excepting for `r&pol`, showing a different behavior of this antibiotic.

Finally, among the features in the top five with a higher MI score, we also find `days_to_culture` (for POL) and *age* (for QUI). Both features are also among the top ten for the rest of the antimicrobial families. From a clinical viewpoint, it is known that both age and a longer ICU stay are risk factors to become infected [14].

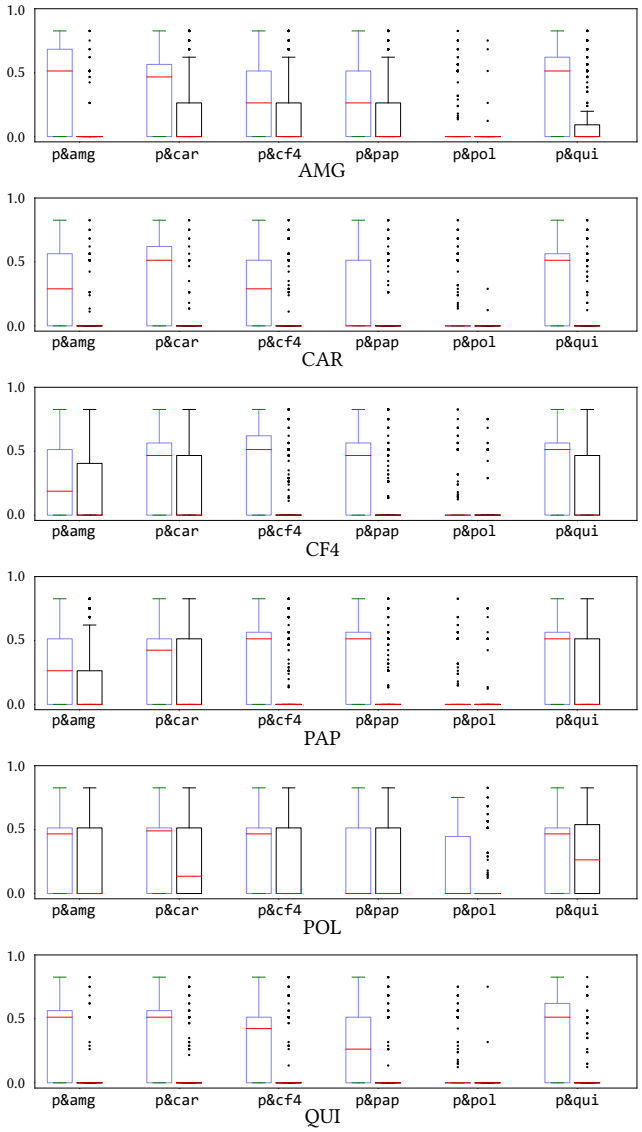


Fig. 4. Boxplots for the six features named p&a$_i$ when considering both resistant (left boxpplot, in blue) and susceptible (right bloxplot, in black) antibiograms for culture *c*.

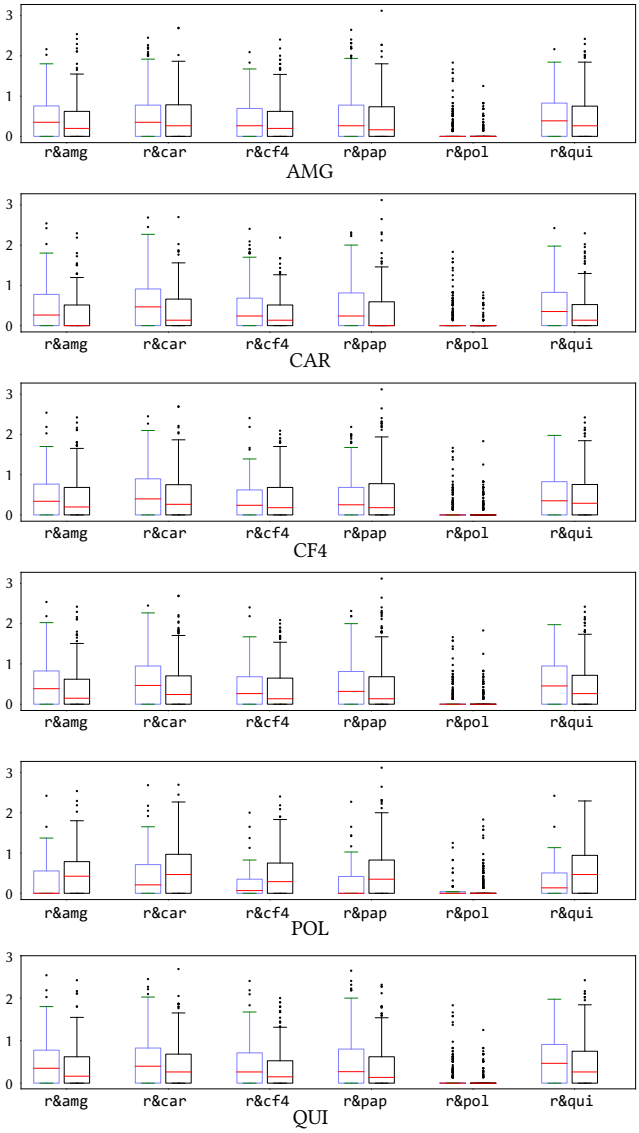


Fig. 5. Boxplots for the six features named r&a$_i$ when considering both resistant (left boxpplot, in blue) and susceptible (right bloxplot, in black) antibiograms for culture c.

III. Methods

A. Data Preprocessing

Before using the data set to predict the result of the susceptibility test, a previous stage of preprocessing is needed. The first aspect to be considered is that six binary classifiers are going to be built in order to predict whether a culture is susceptible or resistant to each of the six different antimicrobial families. A different approach to tackle this problem would be to train a multi-class classifier. However, generating different classifiers allows to individually tune the hyperparameters of each of them and also makes the interpretation and analysis of results easier. To train them, the main data set is divided in six smaller data sets, each of them just considering one binary target c&a$_i$. After that, all the instances representing cultures from patients that have stayed less than 48 hours in the ICU, are removed from every of the six data sets.

As indicated in Table I, the number of features is 40 for every data set, considering the respective target feature. The number of instances are 755, 643, 749, 749, 483 and 708 for AMG, CAR, CF4, PAP, POL and QUI data sets, respectively. Since instances represent cultures, and cultures have an intrinsic temporal ordering, instances are sorted in a temporal manner, with older instances at the beginning of the data set and the newer ones towards the end.

The missing values of the data sets are found in the 12 generated features (r&a$_i$ and p&a$_i$). The percentages of missing values for each of the data sets and features are detailed in Table II and Table III.

It is remarkable that the percentages of missing values for p&a$_i$ features are higher than those of r&a$_i$ features. This happens because, in general, during the same time interval the number of cultures associated to a group of patients will be higher than the number cultures associated to just one patient. It is also notable that, overall, p&pol and r&pol have a high percentage of missing values with respect to the rest of the features of their respective type. This is caused by the very few resistant instances there are for POL family , probably because POL started to be tested in 2007 and the rest of antimicrobial families in 2004.

TABLE II. PERCENTAGE OF MISSING VALUES OF THE p&a$_i$ FEATURES FOR EACH OF THE ANTIMICROBIAL FAMILIES

Fam	p&amg	p&car	p&cf4	p&pap	p&pol	p&qui
AMG	34.97	41.72	34.83	34.83	51.39	38.15
CAR	33.28	36.86	33.13	33.13	48.68	37.01
CF4	35.38	42.06	35.25	35.25	51.80	38.58
PAP	35.25	41.92	35.11	35.11	51.67	38.45
POL	29.81	34.78	29.61	29.61	34.16	34.58
QUI	34.75	41.95	34.75	34.75	52.40	35.03

TABLE III. PERCENTAGE OF MISSING VALUES OF THE r&a$_i$ FEATURES FOR EACH OF THE ANTIMICROBIAL FAMILIES

Fam	r&amg	r&car	r&cf4	r&pap	r&pol	r&qui
AMG	15.36	20.66	15.23	15.23	33.91	17.88
CAR	16.64	19.60	16.49	16.49	36.24	19.60
CF4	15.35	20.69	15.22	15.22	34.05	18.02
PAP	15.35	20.56	15.22	15.22	33.78	18.02
POL	15.11	19.67	14.91	14.91	20.91	19.05
QUI	14.41	20.20	14.41	14.41	34.32	14.69

In the clinical setting, dealing with missing values is an interesting and challenging topic which may have different implications. In this study, missing values are replaced by zeros because of the clinical meaning of p&a$_i$ and r&a$_i$ features. The reason for a p&a$_i$ feature not having a value is that, for the particular patient and time interval considered, it is not found a resistance test result for the specific antimicrobial family studied. If that is the case, it means that, probably, clinicians have considered that the patient may not be infected by a bacterium resistant to the antimicrobial family. Therefore, it can be inferred that likely, in the time prior of the culture being analyzed, the patient was not infected with a resistant bacterium. It seems reasonable to assign a zero in this case, since the feature gets a higher value the more recent a resistant bacterium was detected. Regarding r&a$_i$ features, a similar reasoning can be done. If in the time interval observed, none of the patients in the ICU were tested for resistance to the particular antimicrobial family, it implies clinicians considered it was unlikely to find this kind of resistant bacterium. Thus, it is probable that, prior to the culture, there were no patients infected with a bacterium resistant to the feature's antimicrobial family, causing zero to be an appropriate value.

The categorical features in the data sets are converted into numerical before using them with the machine learning methods considered in this work. The two features representing dates (`date_culture` and `start_date`) are categorical and ordered. Because of that, dates are encoded with integers, assigning lower values to older dates, and higher values to recent dates, indicating, in that way, the ordering among them. The value of a particular date is calculated as the difference, in number of days, between the particular date to be encoded and the first date in the data sets of the specific feature.

Having all features expressed as numerical, Pearson correlation is applied to detect the most correlated ones. If two features (both different from the target feature) are highly correlated, they are adding redundant information to the prediction, and therefore one of them should be removed. In this study it is considered that two features are highly correlated if their correlation coefficient is higher than 0.9 or lower than -0.9. In all of the six data sets, the same four features (`date_culture`, `year_culture`, `start_date` and `year_admission`) are highly correlated among them. Because of that, just `date_culture` is maintained and the other three are removed from the data sets. After that, the number of features in every data set is 37 including the target feature.

B. Predictive Methods

In this section, we describe briefly the data-driven classifier considered in this work. Specifically, LR is tested as base line method, and it was also used in our previous work [15]. In this study, RF has been added to carry out predictions since its interpretability capabilities.

The LR method, very common in the clinical literature, allows us to conduct a linear analysis when the dependent variable is binary. It was used in our previous study [15] because of its simplicity to serve as a baseline, and to evaluate the feasibility of learning from data. In this work, it is again used to classify the instances, now with a greater amount of data and a higher number of antimicrobial families to be analyzed. This is done in order to have a more solid insight on whether the target can be predicted with the available features and the performance this method can provide. Before using LR, each feature is standardized by removing the mean and scaling to unit variance.

The another data-driven method explored here is RF, a machine learning approach commonly used for regression and classification [21], [22]. It is an ensemble method, that is, a RF model is built from multiple decision trees named estimators, which are able to generate individual predictions. RF combines the different predictions of its decision trees (which, individually, tend to over fitting to the training set) to provide a better prediction, providing a better generalization to data not considered in training. The RF method is very robust, since it can handle data sets with an extensive number of features, high dimensionality and heterogeneous features, while having very few hyperparameters. Because of this, RF is often used as a first approach to develop machine learning systems, as it enables to get an overview of the performance on a particular task.

C. Temporal Modeling

Analyzing the problem to be solved, some special characteristics have to be considered when designing the experiments.

The first one is the temporal ordering among instances of the data sets. Since instances are associated with cultures with a susceptibility test, they have an inherent order marked by the date when they were collected. This forces to maintain this same order when predicting instances, that is, past instances cannot be predicted with instances in their respective future. This particularity arises from the fact that, in the real world, when predicting an antibiogram result, future results are not available.

Antimicrobial resistance is a phenomenon that changes over time as bacteria mutates. It allows bacteria to be more resistant to antibiotics as time progresses. As previously mentioned, the features considered include demographic data, information about the patient's admission, and information about the culture and antibiogram results. Since bacteria's mutations are not among the available features, the feature's values telling apart one class from another may change along time. This fact has been previously described as the *concept drift* in which the concept being studied depends on some hidden context, not explicitly given in the form of predictive features [6]. An approach that is normally used to tackle this type of problems is the so called *windowing*, which generalizes from a *sliding window* that moves over the data set instances and applies the knowledge gathered to predict only in the immediate future.

The other particularity is the data scarcity. As previously mentioned, the maximum number of cultures (755) is observed for the AMG antimicrobial family. With the time interval considered (15

years, from 2004 to 2019), there is at most an average of 50 cultures per year. Data scarcity is a trouble spot when using windowing, because in this paradigm, usually, just a small fraction of the data set (the one considered by the sliding window at each particular time) is used for training.

A solution proposed in the previous work [15] was to build an *incremental training window* as the one depicted in panel (d) of Fig. 6. This type of window, which grows in length, contains instances that are as temporarily close as possible to the test instances. Then, the concept drift can be avoided by predicting temporarily close instances to the training set, but it also contains instances far in the past, so that the number of available instances for training is higher than when using *sliding window*. In addition to the *incremental training window*, this work considers a more commonly used *sliding training window with fixed size* to compare their prediction performance. Below we first describe the characteristics of the test window, which is the same for both types of training windows. After that, we present the characteristics of the two types of training windows considered in this work.

The test window consists in a sliding window with a fixed size of just 1 month. Considering just a small amount of time, it is ensured that test instances are as close as possible to the training set. In the experiments of this study, this window begins just considering the first month (January) of 2016. After that, in each prediction step, the test window shifts one month towards later dates. In Fig. 6, steps are indicated at the end of each row as (1), (2), (3), ... (N) for every approach. In the last step, this window considers the last month of the data set. The test window, when shifted, does not overlap with its previous position, that is, in each step predicted instances are different from instances predicted in any other step.

The *incremental training window*, as previously mentioned, is a window of increasing size. In the experiments, this window starts containing instances from 2004 to 2015. In the following steps, the window increases in size one month at a time. In the last step, the training window includes all the instances in the data set except the last month, which is the one considered by the test window.

The *sliding training window with a fixed size* consists in a window just considering 4 years of instances. In every step, this window shifts 1 month towards last instances of the data set, in the same way as the test window does. Since the train and test windows always shift the same amount of time, the distance between them, if any, is always the same. The last step, as previously explained, is the one in which the test window considers the last month of the data set. This kind of window is tested with three different configurations, 0 years approach, 2 years approach and 4 years approach, which are represented in panels (a), (b) and (c) of Fig. 6. In the 0 years approach, the distance between the training and test windows is 0 years, that is, the training window is next to the test one. In this case, the training window considers years from 2012 to 2015 in the initial step. In the 2 years approach the distance among windows is 2 years, therefore taking into account that the test window initially contains the first month of year 2016, the training window includes years from 2010 to 2013, so that the desired distance is respected. Similarly, in the 4 years approach, the window starts considering years from 2008 to 2011, because of the same reason. These three different configurations are considered in order to observe how the prediction evolves as the windows move away from each other, and therefore, the concept drift is more noticeable.

For both types of training windows, at each step, a classifier is trained, and the performance is evaluated on a test set with each of the two methods considered (LR and RF). It is relevant to take into account that patients from training and test windows are different. That is, when predicting a particular patient's susceptibility test

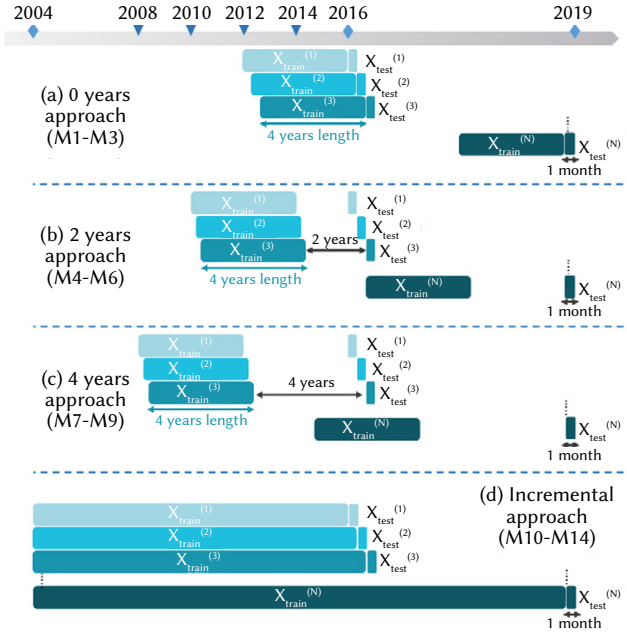


Fig. 6. Sketch for the proposed 14 models (M1 to M14). All models consider a test window of 1 month. Panels (a), (b) and (c) consider a training window of 4 years, with a 1-month sliding training and test windows. Different time slots are considered between the training and the test set: 0 (a), 2 (b) and 4 (c) years. Panel (d) shows an incremental approach for the training set (starting from an initial length of 12 years and incremental steps of 1 month), with the test set immediately after the training set.

result, it is ensured that there are not other susceptibility results of the same patient in the training set. Also, in the approaches where training and test windows are next to each other (as in the *incremental training window and the 0 years approach*), a margin of 48 hours is considered between them, since it is the time required for getting the antibiogram's results.

As the windows traverse the data set, they encounter *class imbalance*, due to the temporal evolution of bacterial resistance. This causes that, in the time interval considered by test windows, there is a higher number of instances from one class. Because of that, in order to evaluate the prediction of the classifiers, is not enough to consider the global accuracy. To get a realistic approximation of the classifier performance, the success in susceptible instances and the success in resistant instances are also calculated. The names assigned to these figures of merit are *Total Accuracy* (A_{Tot}), *Resistant Accuracy* (A_{Rst}) and *Susceptible Accuracy* (A_{Scb}), respectively. For a test window with *Ns* susceptible instances and *Nr* resistant instances, if the method succeeds in predicting *Ss* susceptible instances and *Sr* resistant instances, these figures of merit are computed as follows:

$$A_{Tot} = \frac{Ss + Sr}{Ns + Nr} \tag{5}$$

$$A_{Rst} = \frac{Sr}{Nr} \tag{6}$$

$$A_{Scb} = \frac{Ss}{Ns} \tag{7}$$

These three figures of merit are calculated for the test set of the particular approach considered. In order to get the mean value of these measurements, for every step, the values of *Ns*, *Nr*, *Ss* and *Sr* are accumulated and, at the end, the three figures of merit are obtained. This accumulation is carried out because test windows may have a different amount of instances, due to the fact that not all 1-month time intervals contain the same number of antibiograms. For that reason,

an average would not be adequate, since some instances would have more weight than others depending on the number of instances in their test window.

In addition to the experiments using the different windows, a series of experiments are carried out considering different aspects of the prediction. First, it is analyzed the prediction contribution of the most relevant features according to the MI score. In particular, the features studied are `date_culture` and the two groups of features related to `p&a`$_i$ and `r&a`$_i$. To assess their contribution, the target is predicted with and without considering these features, and the two outcomes are compared.

Secondly, since the *incremental training window* considers a high amount of instances (from the beginning of the data set) it is proposed to assign weights to its training instances. The purpose is to give a higher importance to the training instances that are temporarily closer to the test, which theoretically would have a more similar distribution to the test instances, and lower importance to instances far from the test. Equation (8) details how the weight is generated for each instance.

$$e^{-\lambda(d_l - d_c)} \quad (8)$$

where d_l represents the date of the last culture in the training window, and d_c is the culture date for the instance which weight is being calculated. In the equation, the difference of these two dates is expressed in days. The parameter λ is empirically chosen for each experiment as the one providing the best results among the following: 0, 1e-05, 1e-04, 1e-03, 1e-02, 0.1 and 1. If λ is very small, all instances get a very similar weight, regardless of how far they are from the end of the training window. For instance, for λ = 0, all instances has a weight of 1. On the other hand, if the value of λ is high, only a very few instances very close to the end of the training set get a weight close to 1, and the great majority of instances get a weight very close to 0. Note that when the value of λ is zero, it is the same case as the incremental training window without weights. In the case of high values for λ, it is more similar to the 0 *years approach of the sliding training window with a fixed size*. So, in the end, these weights allow to regulate the amount of past instances considered for prediction.

To encode the models obtained from different combinations of windowing and features, a number is assigned to each model, with the following description:

M1. *Sliding training window with a fixed size* and following the **0 years approach**. It uses **neither `r&a`$_i$ nor `p&a`$_i$** features.

M2. *Sliding training window with a fixed size* and following the **2 years approach**. It uses **neither `r&a`$_i$ nor `p&a`$_i$** features.

M3. *Sliding training window with a fixed size* and following the **4 years approach**. It uses **neither `r&a`$_i$ nor `p&a`$_i$** features.

M4. *Sliding training window with a fixed size* and following the **0 years approach**. It uses **`r&a`$_i$** features but **not `p&a`$_i$** features.

M5. *Sliding training window with a fixed size* and following the **2 years approach**. It uses **`r&a`$_i$** features but **not `p&a`$_i$** features.

M6. *Sliding training window with a fixed size* and following the **4 years approach**. It uses **`r&a`$_i$** features but **not `p&a`$_i$** features.

M7. *Sliding training window with a fixed size* and following the **0 years approach**. It uses **both `r&a`$_i$ and `p&a`$_i$** features.

M8. *Sliding training window with a fixed size* and following the **2 years approach**. It uses **both `r&a`$_i$ and `p&a`$_i$** features.

M9. *Sliding training window with a fixed size* and following the **4 years approach**. It uses **both `r&a`$_i$ and `p&a`$_i$** features.

M10. *Incremental training window*. It uses **neither `r&a`$_i$ nor `p&a`$_i$** features.

M11. *Incremental training window*. It uses **`r&a`$_i$** features but **not `p&a`$_i$** features.

TABLE IV. Prediction Accuracy Results for the **AMG** Antimicrobial Family. Column *M* Indicates the Model, Column *DC* Refers to Whether `date_culture` Is Used (✓) or Not (X). In the Three Left/Right Grouped Columns, LR/RF Is Used

		LR			RF		
M	DC	A_{Tot}	A_{Rst}	A_{Scb}	A_{Tot}	A_{Rst}	A_{Scb}
1	✓	62.5	75.76	22.73	54.55	66.67	18.18
2	✓	**75.0**	**83.33**	**50.0**	64.77	83.33	9.09
3	✓	75.0	96.97	9.09	53.41	60.61	31.82
4	✓	60.23	74.24	18.18	60.23	78.79	4.55
5	✓	73.86	81.82	50.0	70.45	90.91	9.09
6	✓	76.14	96.97	13.64	54.55	62.12	31.82
7	✓	59.09	71.21	22.73	76.14	93.94	22.73
8	✓	73.86	80.30	54.55	73.86	90.91	22.73
9	✓	81.82	96.97	36.36	73.86	81.82	50.0
10	✓	73.86	98.48	0.0	51.14	62.12	18.18
11	✓	73.86	98.48	0.0	64.77	83.33	9.09
12	✓	77.27	96.97	18.18	**76.14**	**87.88**	**40.91**
13	✓	62.5	77.27	18.18	62.5	74.24	27.27
14	✓	75.0	89.39	31.82	**81.82**	**90.91**	**54.55**
1	X	60.23	69.7	31.82	62.5	77.27	18.18
2	X	67.05	72.73	50.0	63.64	80.3	13.64
3	X	29.55	27.27	36.36	36.36	34.85	40.91
4	X	60.23	74.24	18.18	65.91	87.88	0.0
5	X	65.91	72.73	45.45	68.18	86.36	13.64
6	X	31.82	27.27	45.45	34.09	34.85	31.82
7	X	65.91	75.76	36.36	80.68	98.48	27.27
8	X	**73.86**	**77.27**	**63.64**	75.0	90.91	27.27
9	X	69.32	69.7	68.18	65.91	63.64	72.73
10	X	34.09	22.73	68.18	39.77	30.30	68.18
11	X	37.5	24.24	77.27	38.64	24.24	81.82
12	X	73.86	71.21	81.82	**77.27**	**72.73**	**90.91**
13	X	60.23	69.7	31.82	56.82	69.7	18.18
14	X	76.14	78.79	68.18	**79.55**	**75.76**	**90.91**

features.

M12. *Incremental training window*. It uses **both `r&a`$_i$ and `p&a`$_i$** features.

M13. *Incremental training window* with instance weighting. It uses **`r&a`$_i$** features but **not `p&a`$_i$** features.

M14. *Incremental training window* with instance weighting. It uses **both `r&a`$_i$ and `p&a`$_i$** features.

Each of the above kind of models are designed with and without considering the `date_culture` feature, also with the two aforementioned machine learning methods, LR and RF.

After studying the outcomes of the different experiments, the feature relevance is calculated again, now with an embedded method from the RF model. Also, `date_culture` and the **`p&a`$_i$** set of features is analyzed in more depth by making the predictions with just one of these features at a time.

IV. Results and Discussion

The Results and Discussion section is divided in two different subsections. In the Subsection A, the performance of the predictive methods is assessed by considering different experiments. In the Subsection B, the features identified as the most relevant along the study are further analyzed.

A. Prediction

The prediction results are detailed in Tables IV, V, VI, VII, VIII and

TABLE V. Prediction Accuracy Results for the **CAR** Antimicrobial Family. Column *M* Indicates the Model, Column *DC* Refers to Whether `date_culture` Is Used (✓) or Not (X). In the Three Left/Right Grouped Columns, LR/RF Is Used

		LR			RF		
M	DC	A_{Tot}	A_{Rst}	A_{Scb}	A_{Tot}	A_{Rst}	A_{Scb}
1	✓	**93.18**	**98.78**	**16.67**	93.18	100.0	0.0
2	✓	90.91	97.56	0.0	92.05	98.78	0.0
3	✓	88.64	95.12	0.0	80.68	86.59	0.0
4	✓	93.18	98.78	16.67	93.18	100.0	0.0
5	✓	88.64	95.12	0.0	90.91	97.56	0.0
6	✓	89.77	96.34	0.0	77.27	82.93	0.0
7	✓	93.18	98.7805	16.67	93.18	100.0	0.0
8	✓	89.77	96.34	0.0	89.77	96.34	0.0
9	✓	88.64	95.12	0.0	72.73	78.05	0.0
10	✓	**93.18**	**100.0**	**0.0**	93.18	100.0	0.0
11	✓	93.18	100.0	0.0	93.18	100.0	0.0
12	✓	93.18	100.0	0.0	93.18	100.0	0.0
13	✓	**94.32**	**100.0**	**16.67**	93.18	98.78	16.67
14	✓	94.32	100.0	16.67	92.05	97.56	16.67
1	X	90.91	97.56	0.0	93.18	100.0	0.0
2	X	84.09	90.24	0.0	93.18	100.0	0.0
3	X	61.36	65.85	0.0	77.27	82.93	0.0
4	X	89.77	95.12	16.67	93.18	100.0	0.0
5	X	81.82	87.80	0.0	90.91	97.56	0.0
6	X	55.68	59.76	0.0	56.82	60.98	0.0
7	X	88.64	93.90	16.67	93.18	100.0	0.0
8	X	79.55	85.37	0.0	85.23	91.46	0.0
9	X	**68.18**	**70.73**	**33.33**	69.32	73.17	16.67
10	X	60.23	60.98	50.0	61.36	65.85	0.0
11	X	50.0	52.44	16.67	51.14	54.88	0.0
12	X	**75.0**	**74.39**	**83.33**	79.55	81.71	50.0
13	X	**94.32**	**100.0**	**16.67**	93.18	98.78	16.67
14	X	92.05	97.56	16.67	93.18	98.78	16.67

TABLE VI. Prediction Accuracy Results for the **CF4** Antimicrobial Family. Column *M* Indicates the Model, Column *DC* Refers to Whether `date_culture` Is Used (✓) or Not (X). In the Three Left/Right Grouped Columns, LR/RF Is Used

		LR			RF		
M	DC	A_{Tot}	A_{Rst}	A_{Scb}	A_{Tot}	A_{Rst}	A_{Scb}
1	✓	53.93	74.14	16.13	50.56	60.34	32.26
2	✓	52.81	48.28	61.29	46.07	39.66	58.06
3	✓	35.96	10.34	83.87	41.57	13.79	93.55
4	✓	57.30	74.14	25.81	46.07	53.45	32.26
5	✓	49.44	43.10	61.29	39.33	32.76	51.61
6	✓	34.83	5.17	90.32	37.08	5.17	96.77
7	✓	64.04	82.76	29.03	67.42	84.48	35.48
8	✓	**60.67**	**55.17**	**70.97**	50.56	53.45	45.16
9	✓	46.07	18.97	96.77	49.44	36.21	74.19
10	✓	52.81	62.07	35.48	55.06	68.97	29.03
11	✓	46.07	56.9	25.81	38.20	50.0	16.13
12	✓	**61.8**	**74.14**	**38.71**	59.55	74.14	32.26
13	✓	58.43	81.03	16.13	55.06	67.24	32.26
14	✓	61.8	74.14	38.71	**61.8**	**72.41**	**41.94**
1	X	58.43	65.52	45.16	47.19	48.28	45.16
2	X	47.19	37.93	64.52	48.31	24.14	93.55
3	X	47.19	24.14	90.32	35.96	3.45	96.77
4	X	58.43	65.52	45.16	49.44	51.72	45.16
5	X	51.69	34.48	83.87	33.71	17.24	64.52
6	X	42.7	17.24	90.32	31.46	3.45	83.87
7	X	62.92	75.86	38.71	61.8	74.14	38.71
8	X	**64.04**	**53.45**	**83.87**	50.56	41.38	67.74
9	X	49.44	25.86	93.55	47.19	31.03	77.42
10	X	35.96	24.14	58.06	40.45	10.34	96.77
11	X	37.08	31.03	48.39	33.71	13.79	70.97
12	X	52.81	46.55	64.52	**56.18**	**51.72**	**64.52**
13	X	51.69	56.9	41.94	57.30	70.69	32.26
14	X	59.55	63.79	51.61	**60.67**	**58.62**	**64.52**

IX for AMG, CAR, CF4, PAP, POL and QUI families, respectively. The best results are in bold. For each table and models considering or not the `date_culture` feature, three results are marked: the best result among models from M1 to M9, the best result from M10 to M12 and the best result from M13 to M14. Table X shows the chosen values for the λ hyperparameter (instance weighting). The prediction models, identified in column *M* in Tables from IV to IX, are analyzed in three different groups according to the type of temporal window. Firstly, the experiments with an *sliding training window with fixed size* are discussed, with the impact of the distance between training and test windows becoming manifest. Secondly, the results obtained using an *incremental training window* are studied. Finally, we evaluate whether results of the *incremental training window* can be improved by an instance weighting approach.

1. Sliding Training Windows with Temporal Distance Variation Among Training and Test Windows

The figures of merit provided by models considering the temporal distance between the training and test sets are in rows with numbers 1 to 9 in the *M* column of Tables from IV to IX.

In the case of the LR method when considering the feature `date_culture`, the evolution of the figures of merit is not consistent among antimicrobial families when analyzing the separation between training and test windows. In some families, the *Total Accuracy* increases as the training window approaches the test window, while the opposite happens for other families. The same is observed with *Resistant Accuracy* and *Susceptible Accuracy*, its behavior varies depending on the antimicrobial family being predicted.

Predicting with RF and using feature `date_culture`, the evolution of the figures of merit is more similar among the different antimicrobial families. In general, *Total Accuracy* increases, *Resistant Accuracy* increases and *Susceptible Acurracy* decreases as the training window approaches test window. When this pattern is less evident, it may be helpful to analyze when both $\texttt{r\&a}_i$ and $\texttt{p\&a}_i$ features are considered. Also, the general performance of the three figures of merit appears to be better when both $\texttt{r\&a}_i$ and $\texttt{p\&a}_i$ features are used.

For LR and **not using** the feature `date_culture`, the aforementioned pattern appears, in which *Total Accuracy* increases, *Resistant Accuracy* increases and *Susceptible Accuracy* decreases when reducing the distance between windows. Comparing these results with those provided by LR and `date_culture`, two remarks deserve to be underscored: for the families in which this pattern was not previously evident (such as AMG, CAR and QUI), now windows 4 and 2 years apart have lower *Total Accuracy* and lower *Resistant Accuracy*, with similar figures of merit in the 0 years-apart windows; on the other hand, for the families where this pattern was reasonably evident (such as CF4, PAP and POL), the figures of merit usually improve, while maintaining the same pattern. Also using both the $\texttt{r\&a}_i$ and $\texttt{p\&a}_i$ features tend to improve the performance.

Considering RF for prediction and **not using** the feature `date_culture`, the same behavior as in LR without `date_culture`, is

TABLE VII. Prediction Accuracy Results for the **PAP** Antimicrobial Family. Column *M* Indicates the Model, Column *DC* Refers to Whether `date_culture` Is Used (✓) or Not (X). In the Three Left/Right Grouped Columns, LR/RF Is Used

		LR			RF		
M	DC	A_{Tot}	A_{Rst}	A_{Scb}	A_{Tot}	A_{Rst}	A_{Scb}
1	✓	50.56	66.04	27.78	55.06	84.91	11.11
2	✓	60.67	77.36	36.11	51.69	64.15	33.33
3	✓	46.07	52.83	36.11	35.96	24.53	52.78
4	✓	50.56	66.04	27.78	46.07	69.81	11.11
5	✓	**65.17**	**62.26**	**69.44**	59.55	52.83	69.44
6	✓	47.19	49.06	44.44	37.08	20.75	61.11
7	✓	61.8	83.02	30.56	68.54	86.79	41.67
8	✓	67.42	79.25	50.0	68.54	81.13	50.0
9	✓	56.18	58.49	52.78	60.67	58.49	63.89
10	✓	64.04	98.11	13.89	52.81	67.92	30.56
11	✓	61.8	96.23	11.11	39.33	47.17	27.78
12	✓	65.17	98.11	16.67	**67.42**	**75.47**	**55.56**
13	✓	64.04	96.23	16.67	50.56	56.60	41.67
14	✓	68.54	90.57	36.11	**67.42**	**75.47**	**55.56**
1	X	55.06	64.15	41.67	50.56	67.92	25.0
2	X	58.43	64.15	50.0	43.82	37.74	52.78
3	X	47.19	45.28	50.0	40.45	22.64	66.67
4	X	52.81	66.04	33.33	46.07	66.04	16.67
5	X	57.30	64.15	47.22	47.19	39.62	58.33
6	X	49.44	47.17	52.78	34.83	16.98	61.11
7	X	61.8	73.58	44.44	67.42	86.79	38.89
8	X	**66.29**	**67.92**	**63.89**	68.54	77.36	55.56
9	X	55.06	49.06	63.89	62.92	58.49	69.44
10	X	39.33	28.30	55.56	44.94	20.75	80.56
11	X	37.08	22.64	58.33	32.58	11.32	63.89
12	X	**70.79**	**67.92**	**75.0**	69.66	71.7	66.67
13	X	53.93	62.26	41.67	51.69	56.60	44.44
14	X	**71.91**	**69.81**	**75.0**	70.79	69.81	72.22

TABLE VIII. Prediction Accuracy Results for the **POL** Antimicrobial Family. Column *M* Indicates the Model, Column *DC* Refers to Whether `date_culture` Is Used (✓) or Not (X). In the Three Left/Right Grouped Columns, LR/RF Is Used

		LR			RF		
M	DC	A_{Tot}	A_{Rst}	A_{Scb}	A_{Tot}	A_{Rst}	A_{Scb}
1	✓	**68.97**	**63.33**	**71.93**	63.22	0.0	96.49
2	✓	44.83	6.67	64.91	65.52	0.0	100.0
3	✓	47.13	0.0	71.93	65.52	0.0	100.0
4	✓	67.82	63.33	70.18	66.67	3.33	100.0
5	✓	49.43	6.67	71.93	65.52	0.0	100.0
6	✓	50.57	0.0	77.19	65.52	0.0	100.0
7	✓	66.67	63.33	68.42	65.52	3.33	98.25
8	✓	54.02	6.67	78.95	65.52	0.0	100.0
9	✓	52.87	0.0	80.70	65.52	0.0	100.0
10	✓	58.62	13.33	82.46	65.52	0.0	100.0
11	✓	**63.22**	**30.0**	**80.70**	65.52	0.0	100.0
12	✓	56.32	23.33	73.68	66.67	3.33	100.0
13	✓	**72.41**	**60.0**	**78.95**	73.56	46.67	87.72
14	✓	65.52	56.67	70.18	59.77	23.33	78.95
1	X	**74.71**	**63.33**	**80.70**	65.52	0.0	100.0
2	X	56.32	0.0	85.96	65.52	0.0	100.0
3	X	64.37	0.0	98.25	65.52	0.0	100.0
4	X	72.41	60.0	78.95	64.37	0.0	98.25
5	X	58.62	0.0	89.47	65.52	0.0	100.0
6	X	60.92	0.0	92.98	65.52	0.0	100.0
7	X	70.11	60.0	75.44	64.37	0.0	98.25
8	X	57.47	0.0	87.72	65.52	0.0	100.0
9	X	60.92	0.0	92.98	65.52	0.0	100.0
10	X	65.52	0.0	100.0	65.52	0.0	100.0
11	X	63.22	0.0	96.49	65.52	0.0	100.0
12	X	**64.37**	**6.67**	**94.74**	65.52	0.0	100.0
13	X	**65.52**	**56.67**	**70.18**	68.97	33.33	87.72
14	X	65.52	56.67	70.18	64.37	26.67	84.21

observed for all antimicrobial families: note the same pattern for the evolution of the figures of merit (*Total Accuracy* increases, *Resistant Accuracy* increases and *Susceptible Accuracy* decreases as the distance between train and test windows decreases). Comparing these results to previous ones of RF using `date_culture`, it is noticed that now, for all families, windows of 4 and 2 years apart have lower *Total Accuracy* and lower *Resistant Accuracy*, with similar or improved figures of merit in the 0 years-apart windows. Furthermore, using both `r&a`$_i$ and `p&a`$_i$ features tend to provide a better performance.

In the considered experiments (from model 1 to model 9), it is also noticeable how results change depending on the antimicrobial family. It is specially remarkable for the CAR and POL families. Considering CAR, it is observed that, for the majority of models, the values of *Total Accuracy* and *Resistant Accuracy* are very high, while *Susceptible Accuracy* values are very low, in most cases zero. On the other hand, for the POL family, *Total Accuracy* and *Susceptible Accuracy* are very high and *Resistant Accuracy* is low in general, with many zero values. These results suggest that the outcomes depend on the class distribution along time, for each antimicrobial family. In Fig. 3 it is noticed that CAR is the family with the highest ratio of resistant instances (almost 1 for the last years of the data set), and POL is the family with the lowest ratio of resistant instances. Although less obvious, the rest of the families also appear to be influenced by their respective class distribution.

Firstly, it is interesting to discuss the common pattern observed in almost all families, which causes *Total Accuracy* to increase, *Resistant Accuracy* to increase and *Susceptible Accuracy* to decrease as the distance between train and test windows gets smaller. The reason of this behavior is the *temporal class imbalance*, that is, in the first years of the data set, the majority of instances belong to the susceptible class, but as time progresses, the majority of instances become resistant, as it is depicted in Fig. 3. Using *sliding training windows with fixed size* and the approach with 4 years of distance between windows, the training window has to shift towards the past since the test window starts in 2016 for all experiments, therefore containing years from 2008 to 2011 for the first step of the training window, as explained in Section III.C. Being in the past, it contains a higher number of susceptible instances compared to resistant ones, which causes to perform better in predicting susceptible instances (better *Susceptible Accuracy*) and worse in predicting resistant instances (worse *Resistant Accuracy*). The opposite happens when the distance between windows is 0 years. In this case the window is near the last years of the data set, therefore it contains more resistant instances (improving *Resistant Accuracy*) and less susceptible instances (decreasing *Susceptible Accuracy*). The *Total Accuracy* improves when the distance is small because in test window the majority of instances are, mostly, resistant. If the majority class is well predicted, the *Total Accuracy* is high. We conclude that not all the three figures of merit improve as expected when distance is diminishing, in fact one of them gets worse. Applying oversampling to the minority class in this kind of fixed-size temporal windows, in order to balance the number of the two kind of instances, could improve the

TABLE IX. Prediction Accuracy Results for the **QUI** Antimicrobial Family. Column *M* Indicates the Model, Column *DC* Refers to Whether `date_culture` Is Used (✓) or Not (X). In the Three Left/Right Grouped Columns, LR/RF Is Used

		LR			RF		
M	DC	A_{Tot}	A_{Rst}	A_{Scb}	A_{Tot}	A_{Rst}	A_{Scb}
1	✓	62.26	68.75	0.0	88.68	97.92	0.0
2	✓	66.04	70.83	20.0	71.7	77.08	20.0
3	✓	90.57	97.92	20.0	50.94	50.0	60.0
4	✓	66.04	72.92	0.0	88.68	97.92	0.0
5	✓	66.04	70.83	20.0	71.7	79.17	0.0
6	✓	92.45	100.0	20.0	39.62	33.33	100.0
7	✓	84.91	93.75	0.0	90.57	100.0	0.0
8	✓	77.36	83.33	20.0	84.91	89.58	40.0
9	✓	90.57	97.92	20.0	**83.02**	**81.25**	**100.0**
10	✓	**88.68**	**97.92**	**0.0**	67.92	75.0	0.0
11	✓	88.68	97.92	0.0	83.02	91.67	0.0
12	✓	88.68	97.92	0.0	86.79	95.83	0.0
13	✓	88.68	95.83	20.0	90.57	97.92	20.0
14	✓	88.68	95.83	20.0	**92.45**	**100.0**	**20.0**
1	X	50.94	56.25	0.0	84.91	93.75	0.0
2	X	60.38	64.58	20.0	77.36	83.33	20.0
3	X	67.92	72.92	20.0	28.30	22.92	80.0
4	X	62.26	68.75	0.0	81.13	89.58	0.0
5	X	67.92	72.92	20.0	71.7	79.17	0.0
6	X	50.94	54.17	20.0	30.19	25.0	80.0
7	X	77.36	85.42	0.0	86.79	95.83	0.0
8	X	79.25	85.42	20.0	83.02	89.58	20.0
9	X	75.47	77.08	60.0	**83.02**	**81.25**	**100.0**
10	X	54.72	60.42	0.0	33.96	31.25	60.0
11	X	54.72	60.42	0.0	49.06	47.92	60.0
12	X	**79.25**	**79.17**	**80.0**	75.47	77.08	60.0
13	X	90.57	95.83	40.0	88.68	95.83	20.0
14	X	79.25	79.17	80.0	**83.02**	**83.33**	**80.0**

accuracy in the minority class.

Secondly, it is relevant the change in behavior of prediction when `date_culture` is not considered in both LR and RF methods. Overall, when using `date_culture` for prediction in the 4 years and 2 years approaches, the *Resistant Accuracy* increases and the *Susceptible Accuracy* decreases compared to models not using `date_culture`. This probably happens because `date_culture` is compensating the lack of resistant instances of training windows in 4 and 2 years approaches, by telling the classifier the most probable class in test years, which tend to be resistant, and hence *Resistant Accuracy* is high in most cases, causing *Susceptible Accuracy* to decrease. The disadvantage of using `date_culture` is that it causes the minority class to worsen its prediction, since it introduces bias towards classifying instances as the most probable class of the time interval. Since, in the 0 years approach, without considering the `date_culture` feature, the results are similar or better than when `date_culture` is taken into account, we conclude that it is convenient not to use this feature.

2. Incremental Window

The experiments concerning the results of prediction by using an incremental training window are in rows with numbers from 10 to 12 in the *M* column of Tables from IV to IX.

In the case of using the LR method and including the feature `date_culture`, adding just features `r&a`$_i$ does not generally improve figures of merit. With the addition of both features `r&a`$_i$ and `p&a`$_i$, half of the antimicrobial families (AMG, CF4 and PAP) improve their results, although this improvement is mild.

With RF and using the `date_culture` feature, the inclusion of the `r&a`$_i$ features does not improve performance. Conversely, adding `r&a`$_i$ and `p&a`$_i$ features improves results in 5 out of the 6 families (AMG, CF4, PAP, POL and QUI), with no worsening of the figures of merit of the CAR family.

For both LR and RF models without `date_culture`, it is noticed that including just the `r&a`$_i$ features does not provide an improvement in performance. However, taking into account both the `r&a`$_i$ and `p&a`$_i$ features, there is a significant improvement for almost all antimicrobial families. *Total Accuracy* and *Resistant Accuracy* are, in general, considerably lower when `r&a`$_i$ and `p&a`$_i$ features are not used together, in comparison with the results provided by including `date_culture`.

Taking into account the results with *sliding windows of fixed size* of 4 years and the current ones with an *incremental training window*, it is observed that, in general, the best results are obtained with an *incremental training window*. Though for some antimicrobial families, a specific combination of sliding windows can outperform the results of the *incremental training window*, there is not a common approach of sliding windows with better results for all families. Furthermore, when the *incremental training window* outperforms, it is for very little. The exception is the POL antimicrobial family, which achieves clearly better results with the 0 years approach. With the *incremental training window*, best results are mostly achieved by not including `date_culture`, and adding both the `r&a`$_i$ and `p&a`$_i$ features. This confirms that the use of *incremental training window* represents a useful temporal approach to tackle the task presented in this study.

It is notable that, although MI suggested that the set of `r&a`$_i$ features contain relevant information to predict the targets, its use in conjunction with other features does not appear to improve performance. On the other hand, the `p&a`$_i$ features show a great potential to predict the result of the susceptibility test, since they improve performance in almost all cases.

It is also worth to analyze the fact that, if `date_culture` is not used, *Total Accuracy* and *Resistant Accuracy* get a low value when the `r&a`$_i$ and `p&a`$_i$ features are not jointly used, in comparison with the results obtained by using `date_culture`. The reason of this behavior is similar as the one indicated in previous experiments when not using the `date_culture` feature. Without `date_culture`, classifiers tend to predict much of the test instances as susceptible, because it is usually the majority class in incremental training windows (windows starting at the beginning of the data set). The `date_culture` feature compensates this by introducing bias towards predicting the majority class in the time interval, which in test (near the end of the data set) is resistant. In any case, using `date_culture` worsens the *Susceptible Accuracy*. By adding the `p&a`$_i$ features, it is not necessary to count with `date_culture` to get a good performance. Moreover, results with `p&a`$_i$ features and without `date_culture`, improve both *Resistant Accuracy* and *Susceptible Accuracy* because this kind of features do not introduce a temporal bias towards one of the two classes.

3. Incremental Window with Weights

The prediction results using an incremental training window and instance weighting are in rows with numbers 13 and 14 in the M column of Tables from IV to IX. The λ values for each particular case are expressed in Table X.

It is observed that, using instance weighting, results improve for most of the antimicrobial families. The following are the best figures of merit of $A_{Tot} - A_{Rst} - A_{Scb}$ provided by applying instance weighting:

- **AMG**: 79.55%-75.76%-90.91%. Obtained using RF, without `date_culture` and with both the `r&a`$_i$ and `p&a`$_i$ sets of features. The weight hyperparameter is λ =1e-05.

TABLE X. VALUES OF THE HYPERPARAMETER λ FOR RESULTS OF M13 AND M14 IN TABLES FROM IV TO IX. THE COLUMN *Fam* SPECIFIES THE FAMILY BEING PREDICTED, AND COLUMN DC WHETHER `date_culture` IS TAKEN INTO ACCOUNT. THE TWO LEFT/RIGHT COLUMNS REFER TO THE LR/RF METHODS. COLUMNS M13 AND M14 INDICATE THE MODEL FOR WHICH λ IS CHOSEN

		LR		RF	
FAM	DC	M13	M14	M13	M14
AMG	✓	1e-03	1e-03	1e-04	1e-05
AMG	X	1e-03	1e-03	1	1e-05
CAR	✓	1e-02	1e-02	1	1
CAR	X	1e-02	0.1	1	1
CF4	✓	1	0	1	1e-05
CF4	X	1e-03	1e-03	1	1e-05
PAP	✓	1e-04	1e-03	1	0
PAP	X	1e-03	1e-05	1	1e-04
POL	✓	1e-03	0.1	1	1
POL	X	0.1	0.1	1	1
QUI	✓	0.1	0.1	1e-02	1e-02
QUI	X	0.1	0	1	1e-02

- **CAR**: 94.32%-100.0%-16.67%. Obtained using LR, with or without `date_culture` and with the `r&a`$_i$ set of features. The weight hyperparameter is λ =1e-02.
- **CF4**: 60.67%-58.62%-64.52%. Obtained using RF, without `date_culture` and with both the `r&a`$_i$ and `p&a`$_i$ sets of features. The weight hyperparameter is λ =1e-05.
- **PAP**: 71.91%-69.81%-75.0%. Obtained using LR, without `date_culture` and with both the `r&a`$_i$ and `p&a`$_i$ sets of features. The weight hyperparameter is λ =1e-05.
- **POL**: 72.41%-60.0%-78.95%. Obtained using LR, with `date_culture` and with just the `r&a`$_i$ set of features. The weight hyperparameter is λ =1e-03.
- **QUI**: 83.02%-83.33%-80.0%. Obtained using RF, without `date_culture` and with both the `r&a`$_i$ and `p&a`$_i$ sets of features. The weight hyperparameter is λ =1e-02.

Our results show that M13 and M14 performance, in the majority of families, improves or is maintained when the `p&a`$_i$ set of features is taken into account, confirming what was observed in the two previous groups of experiments. The only exception to that is the POL antimicrobial family. When the `date_culture` feature is used, just the POL family gets better results; in any other case, it is better to not consider this feature. The substantially different behavior of POL is probably due to the very small number of resistant instances for this family, which makes it very dependent on the `date_culture` feature. Besides that, for half of the families (CAR, PAP and POL), the best method is LR, while for the other half (AMG, CF4 and QUI), RF gets the best results.

It is also important to analyze the hyperparameter λ used to assign weights to instances. As previously explained, when the value of λ is small, a greater number of instances get a similar high weight (close to 1); otherwise, when λ is high, just a few instances, temporally close to the test set, get a high weight and the rest of instances get very small weights. For AMG, CF4 and PAP, λ is very small and results are very similar to those of the respective incremental window without weights. This happens because almost all instances are being considered. On the other hand, families CAR, POL and QUI, with a greater λ, show results that are, mostly, more similar to the respective sliding training window with a fixed size than to the incremental window.

Comparing the results of the incremental window with the performance for the rest of experiments, it is noticed that it improves the results for 3 of the 6 families, which are AMG, PAP and QUI. In the case of CAR, the whole incremental training window achieves better results than the version with weights. As before, the family CF4 gets better performance with a specific combination of sliding windows, probably because some particularity of its distribution; POL notably gets its best result with the 0 years approach windows, without `date_culture` and with neither the `r&a`$_i$ nor `p&a`$_i$ sets of features.

B. Relevant Features Analysis

Taking into account previous results, it seems that some features with high MI score, such as `r&a`$_i$, do not help to predict the target feature. The feature `date_culture`, which has the highest MI score, increases the performance in some particular cases, but also introduces bias, and the best results in previous experiments are achieved when this feature is not used. On the other hand, the set of features `p&a`$_i$, also with high MI scores, appears to improve performance in almost all antimicrobial families.

Our analysis reveals the inconsistency between features ranked as relevant according to MI and those that actually increase prediction performance. In order to contrast feature relevance, they are now obtained with an embedded method. Since RF has been used as classifier, tree-based estimators have been selected to compute the new feature importance, with Fig. 7 showing the ranking in relevance. Now, the most relevant feature for AMG, CAR, CF4, PAP and QUI are `p&amg`, `p&car`, `p&cf4`, `p&pap` and `p&qui`, respectively. In the case of POL, `p&pol` is ranked on the 7th position. Regarding `date_culture`, it is still very important. In the case of POL, `date_culture` is the most important one. The set of features `r&a`$_i$ are not considered important overall.

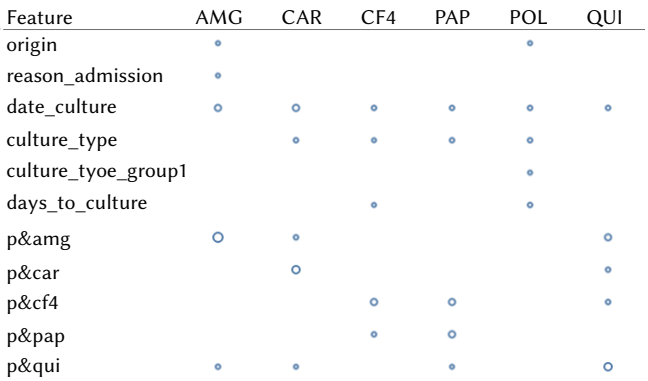


Fig. 7. For each antimicrobial family, the five features with the highest RF relevance scores, indicated by the circle size, from relevance=0.19 (biggest size, pair `p&amg`-AMG) to relevance=0.03 (smallest size, pair reason_admission-AMG).

The new ranking in feature relevance agrees to a greater extent with the prediction performance observed. The set of `p&a`$_i$ features are the most important ones, except for the POL family, where the most relevant feature is `date_culture`. These results make sense, since `date_culture` was the only feature improving performance in the POL family, due to small number of resistant instances. Also, the `r&a`$_i$ features get low relevance values, as expected. The reason why this method provides more insightful results is probably because it takes into account all other features in the data set, while in MI the feature relevance is calculated separately for each feature.

To further analyze the impact of the most relevant features, the antibiogram result has been predicted using just one feature. Two experiments have been carried out, each for one of the most important features in the data set (the `p&a`$_i$ features and `date_culture`). Results with the respective `p&a`$_i$ features are detailed in Table XI, showing that the performance of both LR and RF is very

similar and the figures of merit are relatively high for most of the families. This evidences the high prediction power of this kind of features, even when using for prediction just one of them. Table XII presents the results with just `date_culture`. We observe that the prediction is dramatically biased towards the majority class when the LR method is considered, which in most cases is resistant due to the fact that test instances are in the future with respect to training instances. In the case of the POL antimicrobial family, results are biased towards the susceptible class since it generally is the majority class. Using RF, prediction is also biased, although to a lesser extent. As expected, the only family improving its performance when using just `date_culture` feature is POL.

TABLE XI. Results Using Just the Respective $\texttt{p\&a}_i$ Feature When Predicting the Antibiogram Result for Every Antimicrobial Family (Column *Fam*). For Instance, Just `p&amg` Is Used to Predict Resistance to the AMG Family. In the Three Left/Right Grouped Columns, the LR/RF Method Is Applied

	LR			RF		
Fam	A_{Tot}	A_{Rst}	A_{Scb}	A_{Tot}	A_{Rst}	A_{Scb}
AMG	80.68	74.24	100.0	80.68	74.24	100.0
CAR	63.64	62.2	83.33	62.5	60.98	83.33
CF4	65.17	65.52	64.52	64.04	63.79	64.52
PAP	70.79	66.04	77.78	70.79	66.04	77.78
POL	62.07	0.0	94.74	63.22	0.0	96.49
QUI	73.58	70.83	100.0	73.58	70.83	100.0

TABLE XII. Results Using Just the `date_culture` Feature When Predicting the Antibiogram Result for Every Antimicrobial Family (Column *Fam*). In the Three Left/Right Grouped Columns, the LR/RF Method Is Applied

	LR			RF		
Fam	A_{Tot}	A_{Rst}	A_{Scb}	A_{Tot}	A_{Rst}	A_{Scb}
AMG	75.0	100.0	0.0	56.82	66.67	27.27
CAR	93.18	100.0	0.0	90.91	96.34	16.67
CF4	65.17	100.0	0.0	51.69	60.34	35.48
PAP	59.55	100.0	0.0	57.3	54.72	61.11
POL	65.52	0.0	100.0	66.67	56.67	71.93
QUI	90.57	100.0	0.0	88.68	95.83	20.0

V. Conclusions

One important and increasing problem in daily operation of worldwide health systems, and in particular, of hospitals is antimicrobial resistance. This resistance in some microorganisms (bacterium, viruses, etc.) appears when these microorganisms become to be resistant to antimicrobial drugs to which they were susceptible before. This change is due to a mutation of the microorganism or to the acquisition of the resistance gen. This problem is even more difficult in hospital ICUs, due to the critical condition of those patients. Therefore, a reliable and anticipated prediction for a given bacterium of being resistant or not to one or more antimicrobial families in a patient culture would greatly help physicians in their fight against those microorganisms.

In this study, a real anonymized data set with information about patients staying at the ICU in the University Hospital of Fuenlabrada (UHF) has been used. The data set is related to 3812 admissions of 3346 ICU patients, collected at the UHF during a period of 15 consecutive years (from July 2004 to May 2019). The collected data set from UHF was browsed to generate the final data set under study with the information regarding the patients and their different cultures. Originally there were 40 features, but after the application of some pre-processing techniques they were reduced to 37 to avoid the use of high correlated features.

The analysis have been focused on the *Pseudomonas Aeruginosa* bacteria because is one of the most dangerous bacteria in the ICU and its proved ability to develop multi-drug resistance. Furthermore, six antimicrobial families were considered: *Aminoglycosides* (AMG), *Carbapenems* (CAR), *4th Generation Cephalosporins* (CF4), *Extended-spectrum Penicillins* (PAP), *Polymyxins* (POL) and *Quinolones* (QUI).

Logistic Regression and Random Forest models were tested. Different temporal modeling strategies were proposed based on different windowing schemes (sliding training window, incremental training window) to capture the concept drift phenomenon related to the resistance process of microorganisms. In addition, some new temporally-oriented features ($\texttt{p\&a}_i$ and $\texttt{r\&a}_i$ features) capturing the resistance/susceptibility information regarding past cultures of the same patient or regarding the other patients were proposed and evaluated to improve the prediction accuracy of the different models. A temporal weighting scheme of the instances was proposed and it improved the prediction accuracy. Using or not some important features, according to the MI score, like `date_culture`, $\texttt{p\&a}_i$ features and $\texttt{r\&a}_i$ features were tested in fourteen models (M1 to M14). The results show that the Random Forest method with an incremental win-dow approach, using temporal weighting of the instances and the temporally-oriented features of past cultures is better, especially because both the accuracy for resistant bacteria and susceptible bacteria is more balanced.

Regarding previous studies such as [6], [17] and [18], some similarities and differences are observed with this study. There are many differences between [6] and our work, such as the time interval considered in the data set, the number of instances, the generation of new longitudinal features or the methods used, but the *concept drift* is observed in both works. It is even more noticeable in our work due to the long time interval considered, with the windowing approach showing great benefits when applied to this problem. Unlike the work in [17], our study applies temporal modelling with windowing, including data from the 21 days previous to the antibiogram result to be predicted. In this line, authors in [18] also consider the date of culture and apply a temporal modelling, but without windowing.

Remarkable contributions of our study are the new generated sets of features that consider temporal data contained along the data set, which regards the previous resistance of bacteria for the patient under study ($\texttt{p\&a}_i$), and the resistance of bacteria previously detected in the ICU ($\texttt{r\&a}_i$). In line with [18], our work also reveals that data from past cultures contain a relatively high amount of information to predict antimicrobial resistance. Particularly, the $\texttt{p\&a}_i$ set of features showed to be the most useful for correct prediction when used in combination with some other features or even, in the case of some antimicrobial families, when used alone. Another relevant contribution of our study is the *incremental training window* scheme applied together with instance weighting. It allows to accurately classify cultures when the underlying data distribution dramatically changes along time. Our method introduces a more general and robust solution than those previously proposed, since it can be applied to heterogeneous data sets either with just a few or many years to be predicted, which is able to evolve along time and tackle the scarcity problem. Furthermore, it is able to provide high performance results for the majority of families, similar to the ones in other studies despite not using many of the most important risk factors identified in the literature, such as the antibiotics administered to patients. In addition, the thorough analysis of the relevance and interaction of different features will largely help in the development of future works.

There are different challenges to be addressed for future work.

On the one hand, *oversampling* techniques on training can be tested to check their influence on the model performance. On the other hand, we also consider including other features that could have some influence on the appearance of resistance bacteria in the ICU, like some additional patients' details about their admission, whether they required intubation or not and whether they needed mechanical ventilation or not. It would also be interesting to consider the inclusion of features encoding the antibiotic usage in a temporal context, at a patient level and ICU level. In order to properly tackle the different resistant phenotypes observed in this study, the non-uniform distribution of genotypic resistance mechanisms could be considered. It is also relevant to analyze in a different manner (such as assigning particular weights) cultures isolated from some specific sites such as tracheostomy or environmental water sources, because of their ability to generate aerosols close to patients, increasing the probability of nosocomial bacterial transmission.

Acknowledgment

We are thankful to the University Hospital of Fuenlabrada in Madrid, Spain, for providing us 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 Science and Innovation Ministry Grants Klinilycs (TEC2016-75361-R), 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, and by the Youth Employment Initiative R&D Project (TIC-11649) financed by the Community of Madrid (Spain). Funded action by the Community of Madrid in the framework of the Multiannual Agreement with the Rey Juan Carlos University in line of action 1, "Encouragement of Young Phd students investigation" Project Ref. F661 Acronym Mapping-UCI.

References

[1] W. H. Organization, et al., "Antimicrobial resistance," *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire*, vol. 75, no. 41, pp. 336–336, 2000.

[2] Infectious Diseases Society of America (IDSA), "Combating antimicrobial resistance: policy recommendations to save lives," *Clinical Infectious Diseases*, vol. 52, no. suppl_5, pp. S397–S428, 2011.

[3] M. Mendelson, M. P. Matsoso, "The world health organization global action plan for antimicrobial resistance," *SAMJ: South African Medical Journal*, vol. 105, no. 5, pp. 325–325, 2015.

[4] B. K. English, A. H. Gaur, "The use and abuse of antibiotics and the development of antibiotic resistance," in *Hot topics in infection and immunity in children VI*, Springer, 2010, pp. 73–82.

[5] S. Joshi, et al., "Hospital antibiogram: a necessity," *Indian journal of medical microbiology*, vol. 28, no. 4, p. 277, 2010.

[6] A. Tsymbal, M. Pechenizkiy, P. Cunningham, S. 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)*, 2006, pp. 679–684, IEEE.

[7] A. Lorenz, M. Preuße, S. Bruchmann, V. Pawar, N. Grahl, M. C. Pils, L. M. Nolan, A. Filloux, S. Weiss, S. Häussler, "Importance of flagella in acute and chronic pseudomonas aeruginosa infections," *Environmental microbiology*, vol. 21, no. 3, pp. 883–897, 2019.

[8] G. Meletis, M. Bagkeri, "Pseudomonas aeruginosa: Multi-drug-resistance development and treatment options," *Infection Control*, pp. 33–56, 2013.

[9] M. W. Pesesky, T. Hussain, M. Wallace, S. Patel, S. Andleeb, C.-A. D. Burnham, G. 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*, vol. 7, p. 1887, 2016.

[10] M. Ellington, O. Ekelund, F. M. Aarestrup, R. Canton, M. Doumith, C. Giske, H. Grundman, H. Hasman, M. Holden, K. 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*, vol. 23, no. 1, pp. 2–22, 2017.

[11] G. Arango-Argoty, E. Garner, A. Pruden, L. S. Heath, P. Vikesland, L. Zhang, "Deeparg: a deep learning approach for predicting antibiotic resistance genes from metagenomic data," *Microbiome*, vol. 6, no. 1, pp. 1–15, 2018.

[12] M. Nguyen, S. W. Long, P. F. McDermott, R. J. Olsen, R. Olson, R. L. Stevens, G. H. Tyson, S. Zhao, J. J. Davis, "Using machine learning to predict antimicrobial mics and associated genomic features for nontyphoidal salmonella," *Journal of clinical microbiology*, vol. 57, no. 2, 2019.

[13] M. Tlachac, E. A. Rundensteiner, K. Barton, S. Troppy, K. Beaulac, S. Doron, "Predicting future antibiotic susceptibility using regression-based methods on longitudinal massachusetts antibiogram data.," in *HEALTHINF*, 2018, pp. 103–114.

[14] S. Martínez-Agüero, I. Mora-Jiménez, J. Lérida-García, J. Álvarez-Rodríguez, C. Soguero-Ruiz, "Machine learning techniques to identify antimicrobial resistance in the intensive care unit," *Entropy*, vol. 21, no. 6, p. 603, 2019.

[15] À. Hernàndez-Carnerero, M. Sànchez-Marrè, I. Mora-Jiménez, C. Soguero-Ruiz, S. Martínez-Agüero, J. Álvarez Rodríguez, "Modelling temporal relationships in pseudomonas aeruginosa antimicrobial resistance prediction in intensive care unit.," in *Proc. of Singular Problems for Health Care (SP4HC) Workshop at the 24th European Conference on Artificial Intelligence (ECAI 2020)*, 2020.

[16] S. Martínez-Agüero, I. Mora-Jiménez, A. García-Marqués, J. Álvarez Rodríguez, C. Soguero-Ruiz, "Applying lstm networks to predict multi-drug resistance using binary multivariate clinical sequences," in *Proc. of Starting AI Researchers' Symposium (STAIRS) at the 24th European Conference on Artificial Intelligence (ECAI 2020)*, 2020.

[17] G. Eickelberg, L. N. Sanchez-Pinto, Y. Luo, "Predictive modeling of bacterial infections and antibiotic therapy needs in critically ill adults," *Journal of Biomedical Informatics*, vol. 109, p. 103540, 2020.

[18] O. Lewin-Epstein, S. Baruch, L. Hadany, G. Y. Stein, U. Obolski, "Predicting antibiotic resistance in hospitalized patients by applying machine learning to electronic medical records," *Clinical Infectious Diseases*.

[19] Ó. Escudero-Arnanz, I. Mora-Jiménez, S. Martínez-Agüero, J. Álvarez Rodríguez, C. Soguero-Ruiz, "Temporal feature selection for characterizing antimicrobialmultidrug resistance in the intensive care unit," in *Proc. of Singular Problems for Health Care (SP4HC) Workshop at the 24th European Conference on Artificial Intelligence (ECAI 2020)*, 2020.

[20] T. M. Cover, J. A. Thomas, "Elements of information theory," 2012.

[21] A. T. Azar, H. I. Elshazly, A. E. Hassanien, A. M. Elkorany, "A random forest classifier for lymph diseases," *Computer methods and programs in biomedicine*, vol. 113, no. 2, pp. 465–473, 2014.

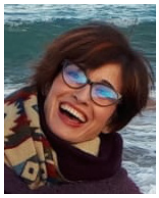
[22] P. Revuelta-Zamorano, A. Sánchez, J. L. Rojo-Álvarez, J. Álvarez-Rodríguez, J. Ramos-López, C. Soguero-Ruiz, "Prediction of healthcare associated infections in an intensive care unit using machine learning and big data tools," in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, 2016, pp. 840–845, Springer.

Àlvar Hernàndez-Carnerero

Àlvar Hernàndez-Carnerero received a B.Sc. degree in Computer Science from the Barcelona School of Informatics (FIB), 2019, and a M.Sc. degree in Intelligent Interactive Systems from the Pompeu Fabra University (UPF) in Barcelona, 2020. He is currently pursuing a Ph.D. in Artificial Intelligence at the Universitat Politècnica de Catalunya (UPC). He is a research support technician at the Computer Science Departament of UPC since 2019. He worked as an Artificial Intelligence developer at LactApp Women Health, in Barcelona, from 2018 to 2019. He also assisted to the 2019 "Apple entrepreneur camp" at the Apple headquarters in Cupertino, California, in 2019. His main research topics are machine learning, application of AI to health care and research data management and case-based reasoning.

Miquel Sànchez-Marrè

Miquel Sànchez-Marrè received a B.Sc. and M.Sc. degrees in Computer Science both from the Barcelona School of Informatics (FIB), in 1988 and 1991, respectively. He received a Ph.D. in Computer Science, in Artificial Intelligence (AI) field, in 1996 from the Universitat Politècnica de Catalunya (UPC). He is Associate Professor in the Computer Science Department (CS) of UPC since 1997 (tenure). He is member of the Intelligent Data Science Research Centre (IDEAI-UPC), and member of the Knowledge Engineering and Machine Learning Group (KEMLG). He is a member of the Editorial Board of International Journal of Applied Intelligence and Modelling journal, and Associate Editor of the Environ-mental Modelling Software journal. He is a Fellow of the International Environmental Modelling and Software Society (iEMSs) since 2005. His main research topics are case-based reasoning, machine learning, intelligent decision support sys-tems, recommender systems, dynamic learning, data science and application of AI to the Environment, Industry and Health.



Inmaculada Mora Jiménez

Inmaculada Mora Jiménez received the degree in Telecommunication Engineering from the Polytechnic University of Valencia, Spain, in 1998, and the Ph.D. degree in Telecommunication from the Carlos III University of Madrid, Spain, in 2004. She is currently an Associate Professor with the Department of Signal Theory and Communications, Telematics and Computing Systems, in Rey Juan Carlos University, Spain. She has conducted her research mainly in data analytic and biomedical engineering. She is co-author of more than 40 JCR indexed papers and 50 contributions to international conferences. She has participated in 18 competitive research projects (principal investigator of 4) and collaborated in more than 20 projects with private funding entities. Her main research interests include data science and machine learning with application to image processing, bioengineering, and wireless communications.

Cristina Soguero Ruiz

Cristina Soguero Ruiz received the Telecommunication Engineering and the Bussiness Degree from Rey Juan Carlos University, Spain, in 2011, and the Ph.D. Degree in machine learning with applications in healthcare, with the Joint Doctoral Program in Multimedia and Communications in conjunction with Rey Juan Carlos University and the University Carlos III of Madrid, in 2015. She was supported by FPU Spanish Research and Teaching Fellowship (granted in 2012). She won the Orange Foundation Best Ph.D. Thesis Award by the Spanish Official College of Telecommunication Engineering. She has published more than 30 papers in JCR journals and international conference communications. She has participated in several research projects (with public and private fundings) related to healthcare data-driven machine learning systems, being principal investigator of 4. Her current research interests include machine learning, data science, and statistical learning theory.

Sergio Martínez Agüero

Sergio Martínez Agüero received the Degree in Telecommunications Technologies Engineering in 2018 at Rey Juan Carlos University and, in 2020 the University Master's Degree in Telecommunication Engineering at Rey Juan Carlos University. He is currently working on his PhD. Thesis, entitled Deep Learning and Network Analytics for extracting knowledge from infectious diseases in the ICU, which extends the research line of the masters and degree thesis. He has several contributions in national and international congresses and published one paper in a JCR journal. He is currently part of two competitive projects funded by the Spanish Government related to healthcare data-driven machine learning models. He is interested in data science, machine learning, data visualization and network analytics.

Joaquín Álvarez-Rodríguez

Joaquín Álvarez-Rodríguez received the Ph.D. in Medicine from the Complutense University of Madrid in 1996. Since 2003 he is the head of the Intensive Care Medicine Department at the Hospital Universitario de Fuenlabrada. His main lines of work have been the quality and safety of patients, medical information systems in Intensive Care Medicine and healthcare-related infections in the ICU. In this last field he has actively participated in the national and regional coordination of the Zero Projects, projects that aim to reduce the main infections acquired in ICU and the emergence of multiresistant bacteria in the ICU. His main research areas are related to the collection of data recorded in the electronic medical record, and try to find a clinical application to them.