Environmental Pollution 367 (2025) 125620

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Measuring water pollution effects on antimicrobial resistance through explainable artificial intelligence

Alfonso Monaco [a,b,1], Mario Caruso [a,1], Loredana Bellantuono [b,c,*], Roberto Cazzolla Gatti [d], Alessandro Fania [a,b], Antonio Lacalamita [a,b], Marianna La Rocca [a,b], Tommaso Maggipinto [a,b], Ester Pantaleo [a,b], Sabina Tangaro [b,e], Nicola Amoroso [b,f,2], Roberto Bellotti [a,b,2]

[a] Università degli Studi di Bari Aldo Moro, Dipartimento Interateneo di Fisica M. Merlin, Bari, 70125, Italy
[b] Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Bari, Bari, 70125, Italy
[c] Università degli Studi di Bari Aldo Moro, Dipartimento di Biomedicina Traslazionale e Neuroscienze (DiBraiN), Bari, 70124, Italy
[d] Alma Mater Studiorum University of Bologna, Department of Biological Sciences, Geological and Environmental (BiGeA), Bologna, 40126, Italy
[e] Università degli Studi di Bari Aldo Moro, Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Bari, 70126, Italy
[f] Università degli Studi di Bari Aldo Moro, Dipartimento di Farmacia - Scienze del Farmaco, Bari, 70125, Italy

ARTICLE INFO

Keywords:
Antimicrobial resistance
Water quality
Machine learning
Explainable artificial intelligence
SHAP

ABSTRACT

Antimicrobial resistance refers to the ability of pathogens to develop resistance to drugs designed to eliminate them, making the infections they cause more difficult to treat and increasing the likelihood of disease diffusion and mortality. As such, antimicrobial resistance is considered as one of the most significant and universal challenges to both health and society, as well as the environment. In our research, we employ the explainable artificial intelligence paradigm to identify the factors that most affect the onset of antimicrobial resistance in diversified territorial contexts, which can vary widely from each other in terms of climatic, economic and social conditions. Specifically, we employ a large set of indicators identified through the One Health framework to predict, at the country level, mortality resulting from antimicrobial resistance related to *Acinetobacter baumannii*, *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Streptococcus pneumoniae*. The analysis reveals the outstanding importance of indicators related to water accessibility and quality in determining mortality due to antimicrobial resistance to the considered pathogens across countries, providing perspective as a potential tool for decision support and monitoring.

1. Introduction

Antimicrobial resistance (AMR) refers to the ability of bacteria, viruses, fungi and parasites to develop a resistance to drugs designed to eliminate them, making the infections they cause more challenging to treat and increasing the likelihood of disease spread and mortality (D'Costa et al., 2011; Allen et al., 2010; Davies and Davies, 2010).

The development of AMR is a complex phenomenon stemming from the interplay of genetic factors and selective pressures in microbial populations, triggered by the abuse and misuse of antibiotics in healthcare (Ventola, 2015; Fleming-Dutra et al., 2016; Ayukekbong et al., 2017; Michael et al., 2014) (especially for self-medication practices (Rather et al., 2017)), veterinary medicine (Wellington et al., 2013; Van Boeckel et al., 2015), breeding (Tang et al., 2017) and agriculture (Cheng et al., 2019). This practice leads to the emergence of stronger strains, facilitating the uncontrolled spread of antibiotic-resistant bacteria across different pathways. Furthermore, the insufficient surveillance and regulation of antimicrobial use in agriculture and breeding, where crops and animals are often exposed to drugs for growth promotion and disease prevention, directly impacts food safety and security, generating disastrous consequences on socio-economic scenarios and even efficacy of medical treatments of infectious diseases (Cassini et al., 2015; Murray et al., 2022; Dadgostar, 2019). As a result, surgeries and chemotherapy become riskier, determining higher mortality rates, prolonged illnesses and increased healthcare costs (Holmes et al., 2016; Partridge et al., 2018; Aminov, 2011; Andersson and Hughes, 2011; Ferri et al., 2017). Moreover, the release of antimicrobial substances into the environment, primarily through

* Corresponding author at: Università degli Studi di Bari Aldo Moro, Dipartimento di Biomedicina Traslazionale e Neuroscienze (DiBraiN), Bari, 70124, Italy.
E-mail address: loredana.bellantuono@ba.infn.it (L. Bellantuono).
[1] These authors contributed equally to this work.
[2] These authors jointly supervised this work.

https://doi.org/10.1016/j.envpol.2024.125620
Received 6 August 2024; Received in revised form 29 December 2024; Accepted 30 December 2024
Available online 7 January 2025

agricultural runoff (Chee-Sanford et al., 2009; Woolhouse et al., 2015; Pruden et al., 2012, 2006; Pei et al., 2006), improper disposal of pharmaceuticals and inadequately treated wastewater (Flandroy et al., 2018; Pallares-Vega et al., 2019; Ju et al., 2019; Szczepanowski et al., 2009; Nadimpalli et al., 2018), significantly contributes to the emergence and spread of antibiotic-resistant bacteria in environmental settings (Kraemer et al., 2019; Cavicchioli et al., 2019; Tacconelli and Pezzani, 2019; Rohr et al., 2019; Wagg et al., 2021). Contamination of water, soil and wildlife contributes to the intensification of the evolutionary pressure on bacteria in these ecosystems, creating a reservoir of resistance genes that changes microbial communities, altering the balance between species. This disruption can generate a self-perpetuating mechanism that produces cascading effects, in which the consumption of crops irrigated with contaminated water and the exposure of livestock to polluted water and soil can favor the development of resistance in animal populations, increasing the risk of zoonotic transmission (Cheng et al., 2019; Tang et al., 2017; Van Boeckel et al., 2015).

AMR's universality and its consequences on the health, environmental, political, social and economic spheres have recently led outstanding supranational agencies to recognize this phenomenon as a global threat. In May 2015, the World Health Assembly endorsed a Global Action Plan to tackle AMR. Furthermore, in September 2016, the United Nations General Assembly prioritized it on the agenda of national policymakers, international organizations and financial institutions in developed and developing countries. The complex and multifaceted nature of the AMR challenge highlighted the need to adopt a holistic, interdisciplinary approach to address it. The answer to this need is the One Health perspective, which acknowledges the interconnectedness of human health, animal health and the environment (Wellington et al., 2013). This framework, consolidated in recent years following the outbreak of the COVID-19 pandemic, promotes an integrated vision in which the scientific, medical and political fields work in synergy to understand and address the main threats to global health.

In this research work, we aim to determine the factors that mostly affect the AMR mortality rates in very diverse territorial contexts, which differ from each other in various characteristics such as climatic, economic and social conditions. Starting from a large set of indicators of interest identified through the One Health framework, we develop an *Artificial Intelligence* (*AI*) model to predict in 203 world countries AMR mortality rates related to the following pathogens: *Acinetobacter baumannii* (AB), *Escherichia coli* (EC), *Klebsiella pneumoniae* (KP), *Pseudomonas aeruginosa* (PA) and *Streptococcus pneumoniae* (SP). In the year 2019 only, these pathogens were each responsible for more than 250.000 deaths associated with AMR (Murray et al., 2022).

Acinetobacter baumannii is a gram-negative bacterium that behaves as an opportunistic pathogen in humans, affecting people with compromised immune systems. It is almost exclusively isolated from hospital environments (Antunes et al., 2014), where it causes nosocomial infections characterized by a wide range of symptoms, with occasional findings in environmental soil and water samples (Yeom et al., 2013).

Escherichia coli is a gram-negative and facultative anaerobic bacterium. The relation between humans and *Escherichia coli*, normally living in lower intestine of warm-blooded mammals and mostly propagating through fecal-oral transmission, is usually a mutualistic one, in which both species benefit from each other (Tenaillon et al., 2010). However, *Escherichia coli* can acquire genetic elements that induce virulent factors, causing diseases out of the lower intestines such as gastroenteritis, urinary tract infections, neonatal meningitis and hemorrhagic colitis. The spread of antibiotic-resistant *Escherichia coli* strains is a cross-cutting issue affecting living species, food and the environment (Pormohammad et al., 2019; Petty et al., 2014).

Klebsiella pneumoniae is a gram-negative and facultative anaerobic bacterium normally found in the bacterial flora of mouth, skin and intestines (Ryan and Ray, 2004), but becomes heavily pathogenic if it comes into contact with lungs or blood. The range of clinical diseases caused by *Klebsiella pneumoniae* includes pneumonia, thrombophlebitis, urinary tract infections, cholecystitis, diarrhea, upper respiratory tract infection, wound infection, osteomyelitis, meningitis and sepsis.

Pseudomonas aeruginosa is a common gram-negative bacterium, aerobic or facultative anaerobic, that can cause disease in plants and animals, including humans (Diggle and Whiteley, 2019). It often infects tissues that are damaged or with reduced immunity and induces generalized inflammation and sepsis. Since it thrives on moist surfaces, *Pseudomonas aeruginosa* can be found also on medical equipment, causing cross-infections in hospitals and clinics.

Streptococcus pneumoniae is a gram-positive and alpha-hemolytic bacterium that spreads by direct person-to-person contact via respiratory droplets. *Streptococcus pneumoniae* typically colonizes the upper respiratory tract with no symptom; it may become pathogenic in individuals with a weaker immune system.

Figure S1, included in the Supplementary Materials, shows the average percentages of resistant strains related to these five pathogens, calculated considering different classes of antibiotics. These averages were obtained based on the results shown by Murray and his research team in their well-known article (Murray et al., 2022). The trends in Figure S1 show that the highest resistance diffusion is mostly found in sub-Saharan Africa and South Asia (Murray et al., 2022). This could be surprising, considering that higher antibiotic use in high-resource countries would lead to a greater burden of bacterial AMR in those regions. However, the significant AMR burden in sub-Saharan Africa and South Asia is greatly influenced not only by the prevalence of resistant bacteria but also by the higher rates of severe infections like lower respiratory infections, bloodstream infections and intra-abdominal infections, which are more common in these regions (Diseases and Collaborators, 2020; Morgan et al., 2011). Other impactful factors to the higher AMR resistances in *Low and Middle Income Countries* (*LMIC*) include a limited number of laboratory facilities for microbiological testing, the antibiotic misuse due to weak regulations, the limited availability of advanced antibiotics and the poor sanitation and hygiene conditions. On the other hand, the growing presence of antibiotic-resistant strains in some more developed areas, like Europe and the Americas, creates general concern among the scientific community (Bonnin et al., 2013; Principe et al., 2014; Budia-Silva et al., 2024; Del Prete et al., 2019; European Centre for Disease Prevention and Control, 2023; Gómez-Zorrilla and Suarez, 2023; Elfadadny et al., 2024).

The upper panel of Fig. 1 shows the mean mortality rates for the considered pathogens in the four classes of countries defined by the so-called *Atlas method*, employed by the World Bank since 1993 (World Bank, 2011). The member countries of each class are displayed in the lower panel of Fig. 1 and listed in descending order of income in the Supplementary Table S1. Besides observing that mortality rates are higher for lower-income countries, it is evident that the highest worldwide mean values are associated to *Klebsiella pneumoniae* and *Escherichia coli* in particular. The geographical distributions of AMR mortality rates related to the pathogens considered in this study are displayed in Fig. 2.

Besides predicting mortality rates from AMR for different pathogens and countries using *Machine Learning* models trained on territorial indicators identified within the One Health framework, we determine the most influential ones by means of *eXplainable Artificial Intelligence* (*XAI*) (Lundberg and Lee, 2017; Lundberg et al., 2020).

This approach has been successful in investigating the determinants of mortality and incidence of several diseases, such as Covid-19 (Cazzolla Gatti et al., 2020), cancer (Cazzolla Gatti et al., 2023; Lacalamita et al., 2023) and dementia (Bellantuono et al., 2022). We decided to employ this algorithm in this work since it has the potential to play a critical role also in tackling AMR.

XAI provides transparency and interpretability in the decision-making process by revealing how *Machine Learning* models make their predictions. This valuable information, which is returned for each

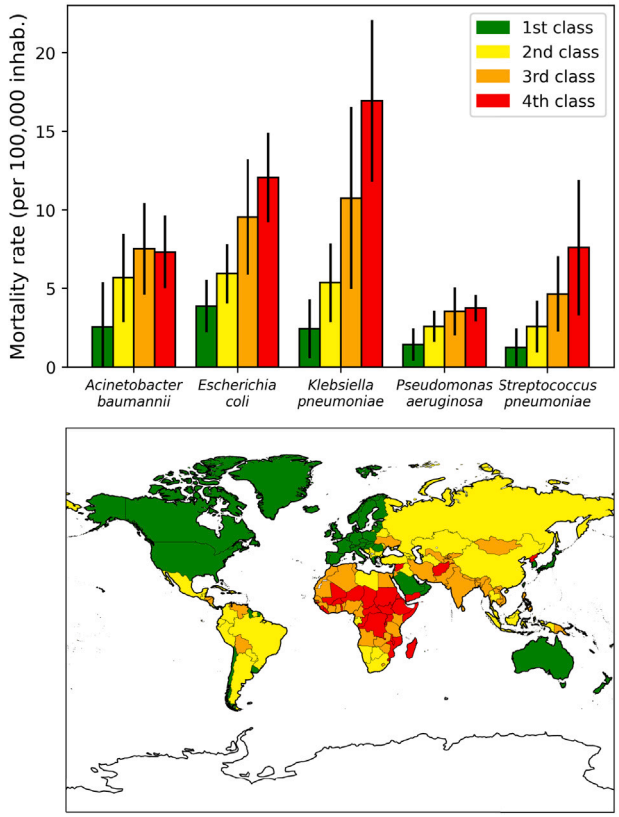

Fig. 1. AMR mortality rates related to the considered pathogens in the world country income groups. The upper panel reports the mean mortality rates (per 100,000 inhabitants) with their standard deviations, in the four classes of countries defined by the World Bank *Atlas method* (World Bank, 2011), labeled in descending order of income. The lower panel shows the countries belonging to each income class, with the same legend as in the upper panel.

country and pathogen combination, can help increase awareness in decision-making for healthcare professionals, policymakers and citizens.

The article is structured in the following manner: the *Materials and Methods* section provides a detailed account of data collection and pre-processing, as well as *Machine Learning* and *XAI* methods employed in our research; the *Results* section presents the main findings of our work; in the *Discussion* section, we analyze our outcomes in relation to previous literature, while highlighting the role of each important variable in determining AMR mortality rates related to the aforementioned pathogens, in different countries.

2. Materials and methods

In this work, we aim to investigate the association between AMR mortality due to resistant infections of several pathogens and a set of multifaceted indicators concerning environmental pollution, social and economic conditions, clinical and lifestyle factors.

In particular, we conduct a nowcasting analysis referred to the year 2019, focusing on 203 countries, each of which is defined by very heterogeneous characteristics. The target of our models is to predict AMR mortality rate, defined as the age-standardized number of deaths per 100,000 inhabitants, to ensure a fair comparison across countries. The *Machine Learning* pipeline employed in this study is shown in Fig. 3.

After a data collection and a pre-processing phase, we implement a *Machine Learning* framework based on the *Boruta* wrapper method to perform feature selection and on three different learning methods, embedded in a cross-validation procedure. Finally, we assess the impact of each feature on the models' predictions through *XAI*, by means of the SHapley Additive exPlanation (*SHAP*) values algorithm. The described pipeline is employed to predict AMR mortality rates related to resistant infections caused by *Acinetobacter baumannii*, *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Streptococcus pneumoniae*.

2.1. Data collection

In this study, we considered 357 intensive indicators referred to 203 nations. The feature dataset included different types of environmental pollution, socio-economic factors and clinical comorbidities for the year 2019. Our study was limited to one year since the dependent variable of our model was only available for that year. The independent variables were chosen based on the literature studies consulted and listed in the bibliography, concerning possible links between these variables and antibiotic resistance.

2.1.1. Independent variables

The complete list of nations considered in this study is shown in the Supplementary Materials, see Table S1. The features used as independent variables, listed in Supplementary Table S2, can be grouped into five categories:

- **Water related variables**, downloaded from AQUASTAT (FAO, 2023a), from Food and Agriculture Organization of the United Nations (FAO, 2023b) and from Our World in Data (OWID) (Our World in Data, 2023) data warehouses. These variables were related to the role played by the water pollution in the antimicrobial resistance issue (Kraemer et al., 2019; Flandroy et al., 2018; Pallares-Vega et al., 2019; Ju et al., 2019; Szczepanowski et al., 2009; Nadimpalli et al., 2018; Manaia et al., 2018) and includes the availability of clean water, the access of population to freshwater sources or facilities, the amount of renewable water and the water stress characterizing a given nation.
- **Sanitation and Hygiene related variables**, collected from OWID, World Health Organization(WHO) (World Health Organization, 2023) and *Hydrosheds* (Hydrosheds, 2023) databases. This category includes features describing the access of people to sanitation, to medical care and to hand-washing facilities; there are also variables related to faecal pollution, wastewaters management and quality of sewage treatment plants (Cani, 2018; Hu et al., 2013; Karkman et al., 2019; Hocquet et al., 2016; Folkesson et al., 2012; Reinthaler et al., 2003; Kümmerer, 2009).
- **Food variables**, downloaded from OWID and FAO databases. This data comprises meat, aquaculture, milk and eggs production and consumption. There are also features related to the total number of livestock and aquaculture sites (Ventola, 2015; Cheng et al., 2019; Tang et al., 2017; Kraemer et al., 2019; Chee-Sanford et al., 2009; Woolhouse et al., 2015; Watts et al., 2017; Zhu et al., 2013).
- **Social, economic and life-style factors**, collected from OWID. This group of features are related to urbanization, governments health-care expenditure and life-style related factors (Nadimpalli et al., 2018; Pruden et al., 2012; Hocquet et al., 2016; Berendonk et al., 2015; Bielen et al., 2017; Larsson, 2014; Larsson et al., 2007).
- **Soil pollution variables**, downloaded from AQUASTAT and OWID data warehouses. These variables regards fertilizers, pesticides, antibiotics use in livestock, worldwide human antibiotics consumption and pollution data linked to mismanaged plastic waste (Larsson et al., 2007; Ramakrishnan et al., 2019; Berendonk et al., 2015).

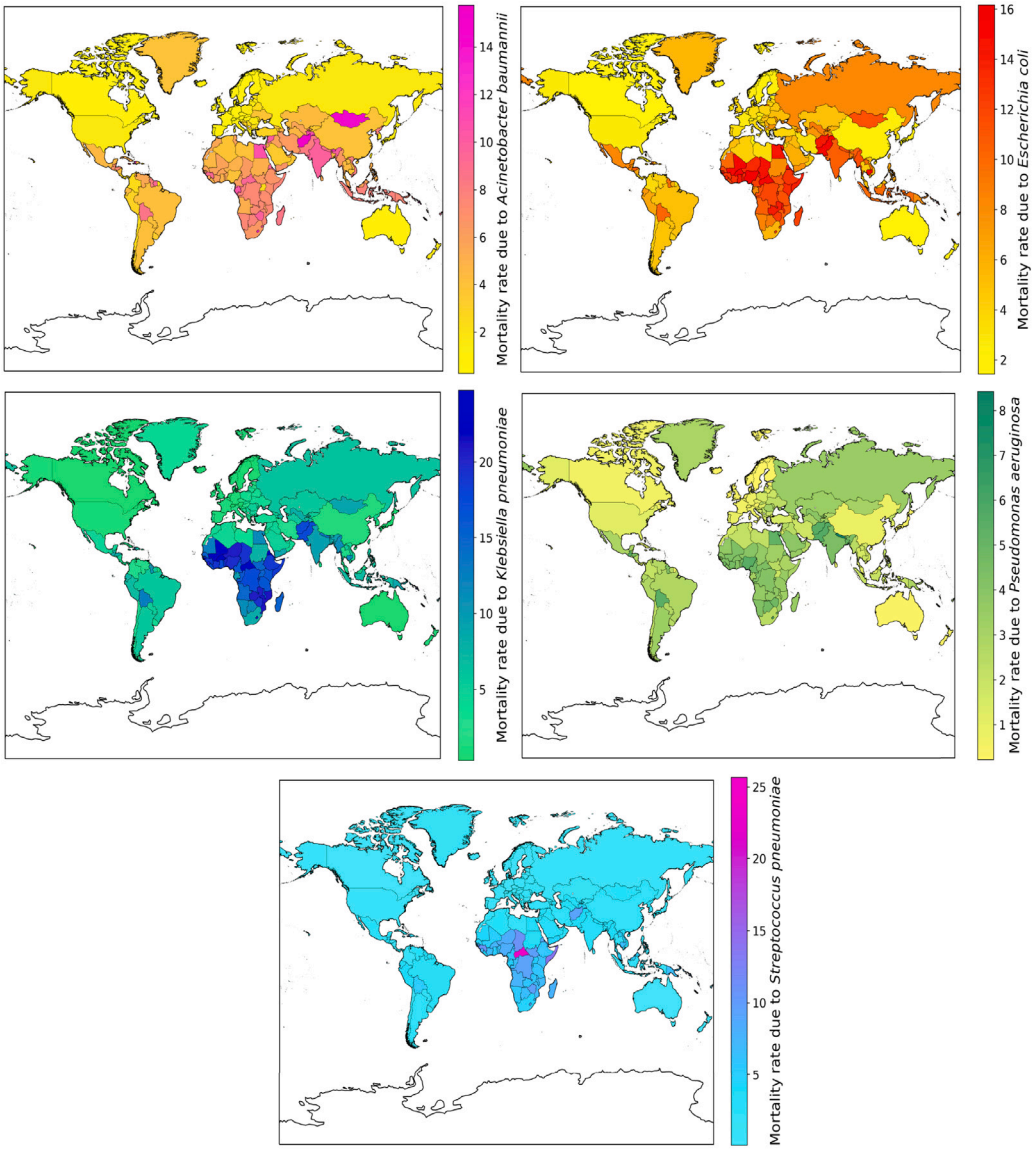

Fig. 2. World maps of mean AMR mortality rates related to infections caused by the selected pathogens. Color bars are reported for numerical reference.

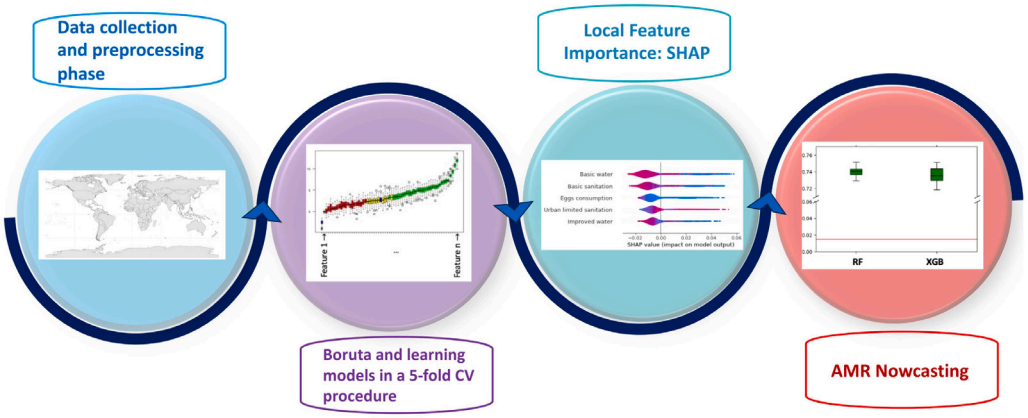

Fig. 3. Flowchart of the complete analysis. After a preliminary data collection and a pre-processing phase, we predict the AMR mortality rates through *Linear Regression*, *Random Forest* and *XGBoost*. Hence, we select the best performing algorithm and we apply a local feature importance procedure, based on a *XAI* approach, to measure the role of each variable in the model. The pipeline is implemented to predict AMR mortality due to resistant infections related to *Acinetobacter baumannii*, *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Streptococcus pneumoniae*.

- **Air pollution variables**, collected from *Google Earth Engine* (Google Earth Engine, 2023) database, and acquired by *Copernicus Sentinel-5 Precursor* mission and by *Copernicus Atmosphere Monitoring Service*, that provides the composition of the Earth's atmosphere at global and regional scales. These features include information about the concentrations of several atmospheric pollutants such as carbon dioxide, methane, ozone, particulate matter, and so on (Kraemer et al., 2019; Li et al., 2018b; Zhou et al., 2023).

2.1.2. Dependent variables

Data related to AMR mortality rates has been collected from the *The Institute for Health Metrics and Evaluation* website (Institute for Health Metrics and Evaluation, 2024), section *Antimicrobial resistance*. This data was originally reported in the work of Murray et al. (2022). We decided to investigate five pathogens: in particular, those with the highest *associated* mortality rate (Tacconelli et al., 2018; Rice, 2008). The five pathogens with the respective antibiotics categories to which they are resistant are listed below:

- ***Acinetobacter baumannii*** (AB) with *aminoglycosides*, *anti-pseudomonal penicillin* inhibitors, *beta-lactam* inhibitors, *carbapenems*, third-generation *cephalosporins*, fourth-generation *cephalosporins* and *fluoroquinolones*.
- ***Escherichia coli*** (EC) with *aminoglycosides*, *aminopenicillin*, *beta-lactam* inhibitors, *carbapenems* and *trimethoprim-sulfamethoxazole*.
- ***Klebsiella pneumoniae*** (KP) with *aminoglycosides*, *beta-lactam* inhibitors, *fluoroquinolones* and *trimethoprim-sulfamethoxazole*.
- ***Pseudomonas aeruginosa*** (PA) with *aminoglycosides*, *anti-pseudomonal penicillin* inhibitors, *carbapenems*, third-generation *cephalosporins*, fourth-generation *cephalosporins* and *fluoroquinolones*.
- ***Streptococcus pneumoniae*** (SP) with *beta-lactam* inhibitors, *carbapenems*, third-generation *cephalosporins*, *fluoroquinolones*, *macrolide* and *trimethoprim-sulfamethoxazole*.

The data we investigated represent the age-standardized estimated deaths *associated to* AMR per 100,000 inhabitants of a given nation, for each considered pathogen-drug combination. For each pathogen, we calculated the average value between different aforementioned combinations, obtaining five dependent variables to predict with our models.

2.2. Pre-processing

The feature dataset contained missing values. In particular, among the 203 nations, there were a few for which it was not possible to obtain updated data for the year 2019. We employed a two-steps strategy in order to tackle this issue.

In the first step, we decided to fill gaps using data related to years *before* 2019, but as close as possible to it, assuming that the data distributions of previous years were fairly constant in a small temporal neighborhood of 2019. For clarity, let us suppose that feature f had a missing value for country C for 2019: this vacancy was filled using the value given by feature f for country C for 2018; if also this value was missing, the 2017 value was used. This procedure of handling missing values never exceeded fillings over a maximum of three years prior to 2019.

After this step, there were still missing values affecting the dataset. In the second step, we grouped countries according to the *Atlas* criterion, developed by the World Bank (World Bank, 2011). This criterion partitions nations according to their Gross National Income (GNI) per capita and per year, expressed in *U.S.* dollars. Using this criterion, the considered 203 nations were divided into four classes: High income (1), Upper-middle income (2), Lower-middle income (3) and Low income (4) nations.

The remaining missing values in the dataset, representing cases in which a given feature was unavailable for a specific country, were filled with the average value of the same feature computed within the income group to which that country belonged. For clarity, let us consider a generic feature f whose values were missing for countries X, Y, K, Z, belonging to the income groups 1, 2, 3, and 4, respectively. We computed the averages $\bar{f}_j$ of feature f, with $j = 1, \ldots, 4$, within each class of nations. Then, we imputed the missing values of feature f for countries X, Y, K, and Z with $\bar{f}_1$, $\bar{f}_2$, $\bar{f}_3$, and $\bar{f}_4$, respectively.

2.3. Feature selection procedure

For all the selected pathogens, we implemented a feature selection technique using the so-called *Boruta* (Kursa and Rudnicki, 2010) wrapper method. *Boruta* is a robust *Random Forest*-based approach that minimizes noise and correlated features by randomizing the training set. For each feature, this algorithm generates a synthetic counterpart, called the *shadow* version, created by randomly shuffling the values of the original feature. In this way, *Boruta* internally handles two datasets: the original one and its synthetic counterpart, in which the corresponding shadow version replaces the feature of interest. A *Random Forest* algorithm is then trained on both datasets to compare the importance of original and synthetic features; this workflow is repeated for each feature, performing numerous independent shuffles. Ultimately, the *Boruta* method selects only those features for which the performance metrics returned by the *Random Forest* algorithm trained on the original dataset are significantly higher than those obtained on the corresponding synthetic versions.

In this study, the entire feature selection procedure was nested into a repeated 5-fold cross-validation framework with 20 repetitions, as detailed in the *Learning procedure* section. We implemented the *Boruta* algorithm setting 100 as the number of random shuffling iterations and p-value < 0.05 as the significance threshold to reject the null hypothesis that the performance achieved by training on the original dataset is significantly higher than the ones obtained by replacing the feature of interest with its randomly-shuffled counterparts. The *Random Forest* algorithm used to perform this comparison within the *Boruta* pipeline consisted of $D = 500$ trees of maximum depth $M = 8$, trained on $F = \sqrt{S}$ randomly chosen features at each split, with S the total number of input features. The functioning of a *Random Forest* algorithm and the meaning of its internal parameters are explained in the next section.

2.4. Learning procedure

We designed a learning framework to forecast the AMR mortality rate values for all the considered pathogens and 203 nations. The pipeline adopted in this study, schematized in Fig. 3, is based on a repeated cross validation framework. Cross-validation is a statistical method used to evaluate the predictive performance and robustness of a *Machine Learning* model. In a k-fold cross-validation, the dataset is randomly split into k disjoint and equally-sized folds, $k - 1$ of which are used for training, while the *Machine Learning* model is tested on the remaining one; the process is repeated k times, ensuring that each fold serves as a validation set exactly once. Hence, the overall performance of the model is evaluated as the average of the metrics obtained for each of the k folds. Due to the randomized nature of the splitting of the initial dataset into the k folds, this cross-validation procedure is typically repeated over many iterations, to have more reliable estimates of the performance returned by the considered *Machine Learning* algorithm.

In the present research work, we implemented a 5-fold cross-validation scheme with 20 repetitions. It is worth highlighting that the *Boruta* feature selection pipeline, described in the previous section, was implemented within each of the $5 \times 20 = 100$ nested iterations of this repeated cross-validation procedure. Specifically, the *Boruta* algorithm

was independently run on each training set, consisting of 4 out of the 5 folds determined within a given repetition of the cross-validation scheme, to extract the most important features. Then, a restricted version of the training dataset, including only the features selected by the *Boruta* algorithm within this specific iteration, was used as input to train *Machine Learning* regressors, whose performance was evaluated on the remaining fold. In this study, we predicted the AMR mortality rate using *Linear Regression Random Forest* and *XGBoost* algorithms.

Random Forest (Breiman, 2001) is an algorithm that consists of an ensemble of binary classification trees (*CART*). This supervised *Machine Learning* technique is highly popular due to its effectiveness in handling multi-modal datasets and its straightforward tuning process, which involves adjusting only two parameters: the number of randomly chosen features at each split (F) and the total number of trees in the ensemble (D). Furthermore, *Random Forest* exhibits strong resilience to the problem of *overfitting*, thanks to its training methodology, which incorporates a bootstrapping approach and a feature randomization strategy during the forest construction process. Another significant advantage of *Random Forest* is its ability to assess the importance of each variable included in the training phase through an internal feature importance evaluation.

Extreme Gradient Boosting (*XGBoost*) (Chen and Guestrin, 2016) is an extensively optimized iteration of the gradient boosting algorithm, characterized by the additional advantage of parallelization, which significantly accelerates the training process. Unlike the traditional method of training a single optimal model on the entire dataset, *XGBoost* adopts a unique approach: it trains multiple models in parallel, each using different subsets of the training data, and then selects the best-performing model through a voting mechanism.

The parameters of the *Machine Learning* algorithms involved in this work were adjusted on the basis of a fine-tuning procedure. In our work, the best performances were achieved by setting:

1. a *Random Forest* configuration with $D = 500$ trees, $F = \sqrt{S}$ and $M = 8$, where S is the number of input features and M is the maximum depth of each tree;
2. an *XGBoost* configuration with $D = 500$ trees, $\eta = 0.01$ and $P = 0.3$, where η is the learning rate and P is the subsample ratio of the training instances.

Performance measurements were based on the coefficient of determination (R^2) between predicted and actual values of AMR mortality rates:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(A_i - F_i)^2}{\sum_{i=1}^{n}(A_i - \bar{A})^2}, \tag{1}$$

with A_i and F_i representing, respectively, the ith actual and predicted value, and $\bar{A}$ indicating the mean actual value computed over all the n instances. Additionally, we assessed the root mean square error (RMSE), defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(A_i - F_i\right)^2}, \tag{2}$$

and the mean absolute error (MAE), defined as

$$MAE = \frac{1}{n}\sum_{i=1}^{N}\left|A_i - F_i\right|. \tag{3}$$

All the analyses were performed with *python, version 3.9.13*.

2.5. Feature importance procedure: Shapley values

The SHapley Additive exPlanation (*SHAP*) technique is a mathematical method used to clarify the predictions made by a *Machine Learning* model (Jiménez-Luna et al., 2020; Miller, 2019; Bellantuono et al., 2023; Bussmann et al., 2020). Based on concepts from cooperative game theory, this *XAI* approach examines the contribution of individual variables to the predictions of the model. In essence, *SHAP* serves as a model-agnostic interpreter of a *Machine Learning* model, that is considered as a *black box* with hidden internal mechanisms (Lundberg and Lee, 2017; Lundberg et al., 2020). Specifically, *SHAP* takes as input the features, target variables and predictions of the original model and tries to mimic its learning process while prioritizing interpretability: it quantifies the impact of each feature on the prediction of a specific observation by employing the concept of Shapley values (*SHAP*).

Considering the complete feature set S and every possible feature subset $F \subseteq S$, the *SHAP* value of a single feature is defined as the variation between the model's outputs with and without it. In particular, the *SHAP* value of the jth feature referred to the prediction of the observation x is evaluated by incorporating the jth feature into all possible subsets,

$$SHAP_j(x) = \sum_{F \subseteq S - \{j\}} \frac{|F|!(|S| - |F| - 1)!}{|S|!}[f_x(F \cup \{j\}) - f_x(F)], \tag{4}$$

where $|F|!$ and $(|S| - |F| - 1)!$ denote, respectively, the number of permutations of features that precede and follow the jth one; $|S|!$ is the number of permutations of all features; $f_x(F \cup \{j\})$ and $f_x(F)$ indicate the prediction f for the sample x returned by the *Machine Learning* model trained on feature subsets that include or not the jth feature, respectively.

In this research work, we applied the *SHAP* algorithm for local explanations to evaluate the impact of each feature on the predictions returned by both *Random Forest* and *XGBoost* models. Specifically, *SHAP* values were independently computed within each of the 100 validation sets identified in the nested iterations of the 5-fold cross validation with 20 repetitions: in each case, the *XAI* algorithm was coupled to the *Machine Learning* regressor, whose predictions served as input to quantify the impact of features on them. This *XAI* analysis was implemented with the *Python* package *shap*.

2.6. Summary of the features selected by the Boruta algorithm and visualization of their SHAP values

As explained in the previous sections, the *Machine Learning* procedure implemented in this study is based on a nested 5-fold cross validation, repeated 20 times: for each iteration, the *Boruta* algorithm operates on a specific training set and extracts the most important features, which are then used to train the regression model and predict the AMR mortality rate. Therefore, the features selected by *Boruta* within each of the 100 iterations are not necessarily always the same but may differ from one iteration to the other. On the other hand, to evaluate the impact of the different environmental, social, and health factors on mortality from antimicrobial resistance, it is worth identifying the features most frequently chosen by the *Boruta* algorithm in the totality of the iterations considered. To address this, we built a *competition table*, where each row represents a feature, and each column corresponds to one of the repeated cross-validation iterations. The *competition table* was compiled during the repeated cross-validation *Machine Learning* procedure: for each training cycle, the *Boruta* algorithm was applied to the current training set to identify the relevant features. Hence, the corresponding cells in the *competition table* were marked to keep track of the features selected in that particular iteration. This approach made it possible to count and identify *a posteriori* the features selected by *Boruta* in each iteration of the nested 5-fold cross-validation repeated 20 times. We identified for each case the features Boruta selected in at least 90% of the total iterations. This procedure was repeated for all the pathogens examined in this study, and the inherent lists of selected features were compared. Moreover, the features exceeding 90% of occurrences in a *competition table* were used as input to construct the *SHAP* summary plots displayed in the *Results* section.

3. Results

As described in the *Materials and Methods* section, firstly, we applied a pre-processing procedure in which we filled the missing entries in our dataset with the mean values of the corresponding features, computed within the income group to which the nation with missing data belonged. Then, we implemented a 5-fold cross-validation repeated 20 times using a nested feature selection strategy based on the *Boruta* algorithm. For each of the considered pathogens, a different number of features exceeded the 90% occurrence threshold in the *competition table* reporting the *Boruta* selections on all the iterations of the repeated cross validation: 41 features for *Acinetobacter Baumanni*, 48 for *Escherichia coli*, 54 for *Klebsiella pneumoniae*, 19 for *Pseudomonas aeruginosa*, and 20 for *Streptococcus pneumoniae*. The complete lists of selected features per pathogen are displayed in the Supplementary Material, Tables S3–S7. The features common to the lists referred to the different pathogens are: *Improved sanitation*, *Desalinated water*, *Safe water*, *Eggs cons./capita*, *Basic water*, CO_2*/capita*, *U. limited sanitation* and *Basic sanitation*. As we shall discuss in the following, these features are among the ones that most influence the predictive power of the considered *Machine Learning* algorithms.

Moreover, within each iteration of the cross-validation framework, we trained different learning models to predict the AMR mortality rates related to resistant infections due to the five selected pathogens. As specified before, we employed a *Linear Regression*, which served as a benchmark for the study, and two ensemble *Machine Learning* algorithms: *Random Forest* and *XGBoost*.

The performances of each learning model for the five considered pathogens are summarized in Table 1. The box-plots of the evaluated regression performance metrics are reported in the Supplementary Figures S2–S4. According to a *Kruskal–Wallis* test, the two ensemble methods provided statistically comparable performance metrics and both of them outperformed the linear model for each analyzed pathogen. The best results, as measured by the coefficient of determination R^2, were obtained for two particular bacteria: *Escherichia coli* and *Klebsiella pneumoniae*. For simplicity, we performed the *SHAP* analysis only in combination with *Random Forest* by calculating the *SHAP* values associated to AMR mortality rates. Fig. 4 shows the distribution of *SHAP* values referred to *Escherichia coli* and *Klebsiella pneumoniae* for each considered country and for the 20 most important features in terms of mean absolute *SHAP* value. Analogously, Fig. 5 displays the *SHAP* values referred to *Streptococcus pneumoniae*, evaluated for nations belonging to the first-income class. The *SHAP* summary plot were constructed using as input the *SHAP* values of the features exceeding the 90% occurrence threshold of *Boruta* selections in the *competition table*, computed from all the 100 repeated cross validation iterations of the learning procedure. These plots reveal that the features related to water quality and sanitation services are always among the most impactful factors in the prediction of AMR mortality rates.

SHAP summary plots related to other pathogens and to different classes of nations are shown in the Supplementary Figures S5–S9. Supplementary Figures S10–S11 show also the correlation matrices (*Pearson* and *Spearman* correlations) between the distributions of *SHAP* values across countries, for the five selected pathogens. *SHAP* values referred to *Escherichia coli*, *Klebsiella pneumoniae* and *Streptococcus pneumoniae* exhibit significant (p-value < 0.01) positive mutual correlations. In contrast, *SHAP* values related to *Pseudomonas aeruginosa* are weakly correlated with the ones of all the other pathogens. Finally, *SHAP* values referred to *Acinetobacter baumannii* are significantly correlated only to *Pseudomonas aeruginosa*'s ones.

Table 1
Summary table of the performance metrics in the prediction of AMR mortality rates related to resistant infections due to the all the selected pathogens. We evaluated the performances of three learning models, i.e. *Linear Regression* (LinR), *Random Forest* (RF) and *XGBoost* (XGB), using the root mean square error (RMSE), the mean absolute error (MAE) and the coefficient of determination (R^2) between actual and predicted values. All metrics were obtained using a 5-fold Cross Validation repeated 20 times. According to a *Kruskal–Wallis* test, the performances of the ensemble methods were statistically comparable and significantly better than the *Linear Regression*'s ones.

Acinetobacter baumannii - AB				
MODEL	RMSE	MAE	R^2	p-value
LinR	0.281 ± 0.064	0.172 ± 0.013	−0.559 ± 0.866	> 0.01
RF	0.133 ± 0.001	0.094 ± 0.001	0.641 ± 0.010	< 0.01
XGB	0.133 ± 0.002	0.095 ± 0.002	0.643 ± 0.014	< 0.01
Escherichia coli - EC				
MODEL	RMSE	MAE	R^2	p-value
LinR	0.261 ± 0.073	0.151 ± 0.011	0.001 ± 0.687	> 0.01
RF	0.134 ± 0.002	0.104 ± 0.002	0.742 ± 0.008	< 0.01
XGB	0.133 ± 0.002	0.105 ± 0.002	0.734 ± 0.009	< 0.01
Klebsiella pneumoniae - KP				
MODEL	RMSE	MAE	R^2	p-value
LinR	0.214 ± 0.087	0.122 ± 0.014	0.337 ± 0.761	> 0.01
RF	0.105 ± 0.002	0.078 ± 0.001	0.835 ± 0.005	< 0.01
XGB	0.105 ± 0.002	0.080 ± 0.002	0.833 ± 0.006	< 0.01
Pseudomonas aeruginosa - PA				
MODEL	RMSE	MAE	R^2	p-value
LinR	0.137 ± 0.005	0.103 ± 0.003	0.412 ± 0.043	> 0.01
RF	0.121 ± 0.002	0.090 ± 0.002	0.539 ± 0.015	< 0.01
XGB	0.125 ± 0.003	0.093 ± 0.002	0.509 ± 0.020	< 0.01
Streptococcus pneumoniae - SP				
MODEL	RMSE	MAE	R^2	p-value
LinR	0.082 ± 0.005	0.052 ± 0.002	0.541 ± 0.053	> 0.01
RF	0.070 ± 0.001	0.041 ± 0.010	0.669 ± 0.012	< 0.01
XGB	0.069 ± 0.001	0.042 ± 0.010	0.677 ± 0.014	< 0.01

4. Discussion

As shown in Table 1, *Random Forest* and *XGBoost* outperformed the *Linear Regression* model in predicting AMR mortality related to the five pathogens analyzed. Specifically, the two *Machine Learning* models provided mean absolute errors that are less than or equal to 10% and statistically significant R^2 measured at the 1% level in all considered cases. These results highlight that *Random Forest* and *XGBoost* were able to capture the non-linear relationships present in the input data, which were instead neglected by *Linear Regression*.

The choice of using two ensemble models is motivated by the fact that they are easy to tune. Moreover, it is worthwhile noticing that two different ensemble approaches, namely boosting and bagging, have returned statistically comparable performances.

As far as we know, this is the first research work that employs *Machine Learning* techniques and *XAI* methods to predict and explain mortality associated with antibiotic-resistant infections on a global scale. The scientific literature includes several studies concerning the use of *AI* methods in the context of antibiotic resistance. These approaches are increasingly employed for synthesizing new antibiotics (Popa et al., 2022; Cherkasov et al., 2008; Melo et al., 2021) or identifying novel resistant bacterial strains through genomic analyses (Ali et al., 2023; Noman et al., 2023; Anahtar et al., 2021). Furthermore, some studies specifically addressed the use of *AI* in investigating the link between environmental conditions and antibiotic resistance (Sakagianni et al., 2023; Baker et al., 2023; Li et al., 2018a; Wang et al., 2024; Jiang et al., 2024), focusing on more localized geographical targets than the one examined in our research work. Although considering more restricted territorial areas, these studies highlighted aspects that we have also analyzed in the present work, such as the hazards related to intensive farming and improper disposal of wastewater.

The results of the *XAI* analysis, summarized in Figs. 4–5, reveal the outstanding importance of features related to accessibility and quality of water. This outcome confirms the deep connection of the actions to contrast AMR with *Sustainable Development Goals* (*SDGs*) (United

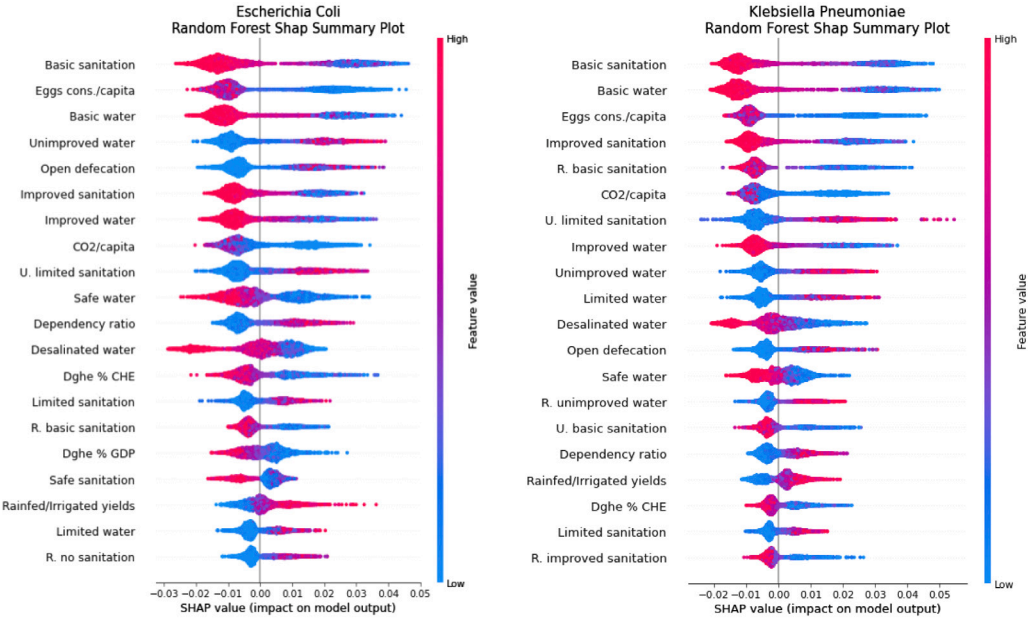

Fig. 4. *SHAP* values distribution of the most influential features in the *Random Forest* prediction of AMR mortality rates related to *Escherichia coli* (left) and *Klebsiella pneumoniae* (right). Each point of a given row corresponds to a different country.

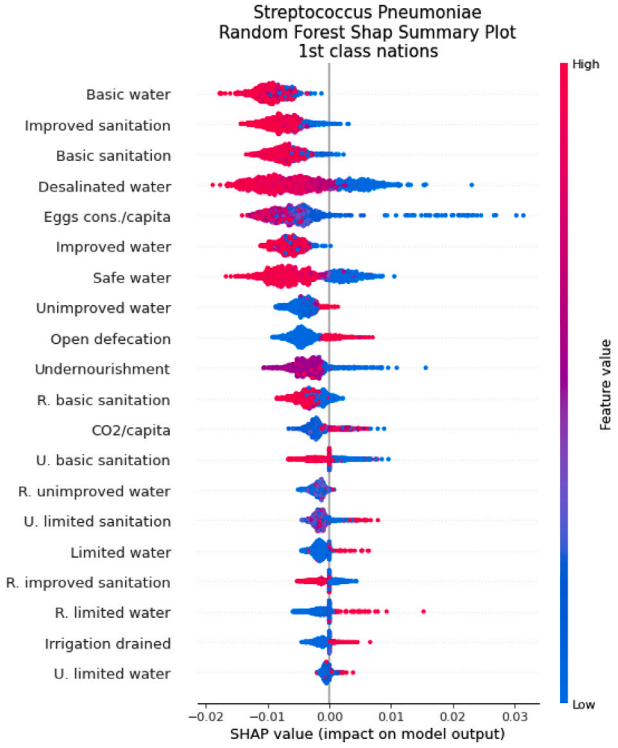

Fig. 5. *SHAP* values distribution of the most influential features in the *Random Forest* prediction of AMR mortality rate related to *Streptococcus pneumoniae*. Each point in a given row corresponds to a different country belonging to the first class of income, identified through the World Bank *Atlas method* (World Bank, 2011).

Nations, 2015), especially because improving clean water and sanitation is the subject of *Goal 6*. Moreover, the features identified as most impactful by *XAI* are related to health (*Goal 3*), conservation of life below water (*Goal 14*), hunger elimination and food sustainability (*Goal 2*).

The *SHAP* summary plots in Fig. 4 reveal a pronounced similarity between the rankings of the factors with the highest impact on the predictions for the AMR mortality rates related to *Escherichia coli* and *Klebsiella pneumoniae*. Many of these features belong to the categories of water and sanitation (see *Materials and Methods* section for more details), indicating the fundamental role of water availability and hygiene in the emergence represented by AMR. This result is consistent with established knowledge since aquatic environments serve as storage for genes that are resistant to antimicrobial agents and allow the transfer of these genes across different bacterial species, disrupting the microbial communities in these ecosystems. Moreover, Fig. 4 shows that higher values of factors indicating favorable conditions are typically associated with lower AMR mortality rates. Actually, immediate access to good quality water is essential to minimize deaths from infections caused by antibiotic-resistant bacteria in a given country's population.

This phenomenon is particularly evident when considering, as an example, the features *Basic water*, *Improved water* and *Safe water*. In these cases, the blue points in Fig. 4 correspond to positive *SHAP* values and represent countries in the 4th group (low-income countries), where these features generally assume lower values. This implies an increase in mortality due to resistant infections, as consistently illustrated by the *XAI* outcomes. On the other hand, high values of these features (corresponding to the red points in the *SHAP* summary plot) influence the model's predictions by contributing to a decrease in the predicted AMR (negative *SHAP* values); these points represent countries in the 1st group (high-income countries), where these features generally have higher values. This reasoning is reversed for other features, such as *Unimproved water*, *Open defecation* or *Urban* (U.) *limited sanitation*, where positive and negative *SHAP* values are consistently associated with countries in the 4th and 1st groups, respectively. The *XAI* approach allows us to interpret the predictive power of the algorithm and can be applied to all other features to quantify their impact on the outcomes of *Machine Learning* models, as discussed below.

Our algorithm has indeed identified also other features that are related to clean water: they play a pivotal role in predicting mortality rates from antibiotic-resistant strains of *Escherichia coli* and *Klebsiella pneumoniae*. For instance, higher values of *Desalinated water* are associated with reduced mortality rates, while higher values of *Dependency ratio* are linked to increased mortality rates. These trends are reasonable, considering the relationship of the aforementioned features to

clean water. *Desalinated water* represents the total amount of desalinated water produced by a given country in a year, which its population can use as clean fresh water. *Dependency ratio* indicates the share of the total renewable water resources originating outside the country: this factor is related to the possible lack of good quality water, a well-known factor impacting the development of antibiotic resistance (Liguori et al., 2022).

As the *SHAP* analysis highlights, another influent feature is the *Rainfed/Irrigated yields*, which represents the ratio between the yields obtained by means of rainfed farming and those obtained from irrigated farming. This ratio is higher in low-income countries where irrigation systems are less present or even absent: rainfed agriculture accounts for more than 95% of the cultivated area in Sub-Saharan Africa, 90% in Latin America, 75% in the Middle East and North Africa, 65% in East Asia and 60% in South Asia (International Water Management Institute, 2010). From *SHAP* values distributions reported in Fig. 4, it can be noticed that lower AMR mortality rates are associated with lower values of this ratio. This result highlights the burden of environmental pollution on AMR mortality. In fact, airborne transport of resistant bacteria and antimicrobial residues by dust particles may contribute to their dispersion over long distances (Li et al., 2018b; Christner et al., 2008; McEachran et al., 2015; Zhou et al., 2023). In addition, rain could carry the pollutants to the ground, where conducive conditions could lead to the proliferation and transmission of *ARGs* (*antimicrobial resistant genes*) among bacteria.

Furthermore, some of the most relevant factors to predict AMR mortality rates returned by the *SHAP* algorithm are related to sanitation and hygiene. Specifically, these indicators are: the general, rural (R.) and urban (U.) *Limited sanitation*, *Basic sanitation* and *Improved sanitation*, *Dghe % CHE* (share of general health expenditure in Current Health Expenditure), *Dghe % GDP* (share of general health expenditure in Gross Domestic Product) and *Open defecation*, representing the percentage of a country's population that practice it. *SHAP* values consistently suggest that higher levels of sanitation utilities (*Basic* and *Improved sanitation*) and more government funds dedicated to health services are associated with lower AMR mortality rates related to *Escherichia coli* and *Klebsiella pneumoniae*. On the other hand, higher values of other indicators, such as *Limited sanitation* and *Open defecation*, lead to higher AMR mortality rates. In particular, *Open defecation* is closely related to AMR (Hu et al., 2013; Karkman et al., 2019) and especially to infections caused by *Escherichia coli* and *Klebsiella pneumoniae*, since these bacteria typically colonizes the human guts. Actually, AMR mortality rates are generally higher in countries where the population can more easily contract difficult-to-treat infections transmitted by bacteria contained in feces.

Although the *SHAP* model returns *Eggs consumption/capita* as an impactful feature related to infection-related deaths due to antibiotic-resistant bacteria (Adesiyun et al., 2006), its nontrivial trend is remarkably different from expectations. Scientific literature (Adesiyun et al., 2006) suggests a possible link between eggs consumption and AMR. Actually, eggs are one of the products of breeding, which is highly affected by the abuse of antimicrobials. However, the *SHAP* values distribution indicates that higher egg consumption values correspond to lower AMR mortality rates related to *Escherichia coli* and *Klebsiella pneumoniae*. This result may be explained by observing that egg consumption per capita can be seen as a proxy for prosperity. Indeed, this indicator is higher in wealthier countries, where hygiene and sanitation conditions are much more favorable than in low-income countries, and hence the mortality rate due to hard-to-treat infections is lower.

Another noteworthy feature that *XAI* approach highlights is the role of CO_2. Results reported in Fig. 4 suggest that higher values of CO_2*/capita* correspond to lower AMR mortality rates. As in the case of *Egg consumption/capita*, this association may be interpreted by observing that higher values of CO_2*/capita* are primarily recorded in high-income countries, where AMR mortality rates are generally lower.

However, further insights emerge from Fig. 5, showing the *XAI* outcomes for predictions of mortality from resistant *Streptococcus pneumoniae* infections in the case of high-income countries, corresponding to the best performance in terms of *RMSE* and *MAE* (see Table 1). This *SHAP* plot suggests that higher values of CO_2*/capita* increase the AMR mortality rate related to *Streptococcus pneumoniae*. This remarkable result may be explained by considering that the presence of CO_2 promotes the conjugative transmission of multi-drug resistance genes (Liao et al., 2019) and that CO_2 in particular stimulates the growth of *Streptococcus pneumoniae* (Kempner and Schlayer, 1942). These two effects seem to combine, leading to a direct link between air pollution and antibiotic resistance.

Moreover, regarding Fig. 5, two further important features have emerged: *Undernourishment* and *Irrigation drained*. The *Undernourishment* indicator measures the proportion of the population whose regular food consumption is adequate to provide the necessary dietary energy required for a healthy and active life. *SHAP* values distribution indicates that higher values of this feature are associated with lower AMR mortality rates. This result seems to be logical, since malnutrition can lead to weaker immune defenses, making individuals more susceptible to infections (Bourke et al., 2016). *Irrigation drained* is an indicator that quantifies the percentage of land areas equipped with surface irrigation, where drainage is used to prevent water-logging. According to the *AQUASTAT* glossary (refer to the *Materials and Methods* section for more information), surface irrigation systems rely on gravity moving water across the land to moisten the soil. As previously noted, surface water can be significantly contaminated by human activities, providing an ideal environment for the spread and transmission of *ARGs* among bacteria. These considerations could explain the trend reported in Fig. 5, showing a positive association between the increase of the *Irrigation drained* indicator and the higher mortality rates caused by antibiotic-resistant *Streptococcus pneumoniae* infections.

Several studies have found a strong correlation between the prevalence of AMR and the consumption of meat (Cheng et al., 2019; Tang et al., 2017). This association is due to the excessive administration of drugs to animals in breeding (Van Boeckel et al., 2015) and the widespread consumption of meat by humans (Ventola, 2015; Fleming-Dutra et al., 2016; Ayukekbong et al., 2017; Michael et al., 2014). Although our study has examined several features related to the consumption of farmed foods and antibiotics (see *Materials and Methods*), they are not among the most important variables returned by the *XAI* analysis (see Figs. 4, 5), or they are counterintuitively linked to AMR mortality, as already observed in the case of egg consumption (see also Supplementary Figures S5–S9). This result can be explained by considering the nature of our study design. Firstly, the variable we have chosen as a target, namely the mortality rates due to infections caused by antibiotic-resistant bacteria, strongly depends on the general conditions of territorial development and is more severe in low-income countries. Secondly, our research aimed to predict AMR mortality rates based on territorial indicators in 203 countries worldwide, characterized by differing economic, social and health conditions. For this reasons, the *SHAP* values distribution reveals as the main predictors positively correlated with AMR mortality all the risk factors of contracting an infection. These risk factors are namely the variables related to water contamination and poor hygienic conditions. On the other hand, variables such as the consumption of meat, eggs and medicines, which can be interpreted as proxies for economic well-being, are positively correlated with the presence of water purification systems and health facilities. These assets can contribute to hindering the spread of antibiotic-resistant infections and enhancing their treatment, respectively. A study focused on more homogeneous socio-economic contexts, in which the variability of hygienic-sanitary factors would be considerably limited, could be the ideal framework to study in a more targeted way the influence of meat and antibiotic consumption on the spread and severity of antibiotic-resistant infections. This type of

investigation would complement the wet lab analyses conducted on in-vitro bacterial colonies originating from different environments, which provide a biological framework to investigate the relationship between antibiotic intake and the development of resistance strains.

Finally, the matrices displayed in Figure S10 and S11 show a strong correlation between the mean absolute *SHAP* values of four pathogens: *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Streptococcus pneumoniae*. The result is also confirmed by the scatter plots in Figure S12, in which the nations belonging to the four different groups are shown with different colors. Our finding could indicate a commonality of causes and characteristics underlying AMR associated with the four pathogens and could provide useful guidance for the implementation of uniform health prevention strategies.

In conclusion, our work develops a decision support tool that provides a global snapshot of the problem of AMR as it becomes more widespread. It is worth highlighting that the present study could be generalized to other pathogens and extended by considering further environmental and social factors. However, this research work also presents some critical points. In fact, the considered geographical scale is very large, since a national territorial unit is employed. This may lead our model to miss specific territorial differences, especially in the case of the largest nations. To develop a tool capable of analyzing the problem at a finer scale, it is necessary to consider more homogeneous datasets at spatial and social levels and enhance the data collection procedures through ad-hoc campaigns. To this end, two possible perspectives consist of extending this model to a sub-national scale, adding further indicators that better characterize local conditions, and intensifying the data collection stage, as data gathered at different times would allow for the implementation of monitoring procedures. Furthermore, to prevent misinterpretation and erroneous policy recommendations, we remark that the relevance of socio-environmental factors in the prediction of AMR, highlighted by *XAI*, should always be validated by domain experts.

CRediT authorship contribution statement

Alfonso Monaco: Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Formal analysis, Conceptualization. **Mario Caruso:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Loredana Bellantuono:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Roberto Cazzolla Gatti:** Writing – review & editing. **Alessandro Fania:** Writing – review & editing. **Antonio Lacalamita:** Writing – review & editing. **Marianna La Rocca:** Writing – review & editing. **Tommaso Maggipinto:** Writing – review & editing. **Ester Pantaleo:** Writing – review & editing. **Sabina Tangaro:** Writing – review & editing. **Nicola Amoroso:** Writing – review & editing, Supervision. **Roberto Bellotti:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Authors were supported by the Italian Ministry of University and Research funding within the “Budget MIUR - Dipartimenti di Eccellenza 2023 - 2027” (Law 232, 11 December 2016) - Quantum Sensing and Modelling for One-Health (QuaSiModO), CUP: H97G23000100001. Authors were supported by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021 of Italian Ministry of University and Research, funded by the European Union - NextGenerationEU; Project code: CN00000013, Concession Decree No. 1031 of 17 February 2022 adopted by the Italian Ministry of University and Research, CUP: H93C22000450007, Project title: National Centre for HPC, Big Data and Quantum Computing. This work was also funded by the Italian Ministry of Enterprises and Made in ITaly (MIMIT) with the “Project CALLIOPE - Casa dell’Innovazione per il One Health” (FSC 2014–2020, CUP: E53C22002800001).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.envpol.2024.125620.

Data availability

Data will be made available on request.

References

Adesiyun, A., Offiah, N., Seepersadsingh, N., Rodrigo, S., Lashley, V., Musai, L., 2006. Frequency and antimicrobial resistance of enteric bacteria with spoilage potential isolated from table eggs. Food Res. Int. 39, 212–219.
Ali, T., Ahmed, S., Aslam, M., 2023. Artificial intelligence for antimicrobial resistance prediction: Challenges and opportunities towards practical implementation. Antibiotics 12, 523.
Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J., Handelsman, J., 2010. Call of the wild: antibiotic resistance genes in natural environments. Nat. Rev. Microbiol. 8, 251–259.
Aminov, R.I., 2011. Horizontal gene exchange in environmental microbiota. Front. Microbiol. 2 (158).
Anahtar, M.N., Yang, J.H., Kanjilal, S., 2021. Applications of machine learning to the problem of antimicrobial resistance: an emerging model for translational research. J. Clin. Microbiol. 59, e0126020.
Andersson, D.I., Hughes, D., 2011. Persistence of antibiotic resistance in bacterial populations. FEMS Microbiol. Rev. 35, 901–911.
Antunes, L.C., Visca, P., Towner, K.J., 2014. *Acinetobacter baumannii*: evolution of a global pathogen. Pathog. Dis. 71, 292–301.
Ayukekbong, J.A., Ntemgwa, M., Atabe, A.N., 2017. The threat of antimicrobial resistance in developing countries: causes and control strategies. Antimicrob. Resist. Infect. Control. 6 (47).
Baker, M., et al., 2023. Machine learning and metagenomics reveal shared antimicrobial resistance profiles across multiple chicken farms and abattoirs in China. Nat. Food 4, 707–720.
Bellantuono, L., Monaco, A., Amoroso, N., Lacalamita, A., Pantaleo, E., Tangaro, S., Bellotti, R., 2022. Worldwide impact of lifestyle predictors of dementia prevalence: An explainable artificial intelligence analysis. Front. Big Data 5, 1027783.
Bellantuono, L., Palmisano, F., Amoroso, N., Monaco, A., Peragine, V., Bellotti, R., 2023. Detecting the socio-economic drivers of confidence in government with explainable artificial intelligence. Sci. Rep. 13 (839).
Berendonk, T.U., et al., 2015. Tackling antibiotic resistance: the environmental framework. Nat. Rev. Microbiol. 13, 310–317.
Bielen, A., et al., 2017. Negative environmental impacts of antibiotic-contaminated effluents from pharmaceutical industries. Water Res. 126, 79–87.
Bonnin, R.A., Cuzon, G., Poirel, L., Nordmann, P., 2013. Multidrug-resistant *Acinetobacter baumannii* clone, France. Emerg. Infect. Diseases 19, 3004–3010.
Bourke, C.D., Berkley, J.A., Prendergast, A.J., 2016. Immune dysfunction as a cause and consequence of malnutrition. Trends Immunol. 37, 386–398.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 32–45.
Budia-Silva, M., et al., 2024. International and regional spread of carbapenem-resistant *Klebsiella pneumoniae* in Europe. Nat. Commun. 15 (5092).
Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J., 2020. Explainable AI in fintech risk management. Front. Artif. Intell. 3 (26).
Cani, P.D., 2018. Human gut microbiome: hopes, threats and promises. Gut 67, 1716–1725.
Cassini, A., et al., 2015. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the Eu and the European economic area in 2015: a population-level modelling analysis. Lancet Infect. Dis. 19, 56–66.
Cavicchioli, R., et al., 2019. Scientists’ warning to humanity: microorganisms and climate change. Nat. Rev. Microbiol. 17, 569–586.
Cazzolla Gatti, R., Di Paola, A., Monaco, A., Velichevskaya, A., Amoroso, N., Bellotti, R., 2023. The spatial association between environmental pollution and long-term cancer mortality in Italy. Sci. Total. Environ. 855, 158439.
Cazzolla Gatti, R., Velichevskaya, A., Tateo, A., Amoroso, N., Monaco, A., 2020. Machine learning reveals that prolonged exposure to air pollution is associated with SARS-CoV-2 mortality and infectivity in Italy. Environ. Pollut. (267).

Chee-Sanford, J.C., Mackie, R.I., Koike, S., Krapac, I.G., Lin, Y.F., Yannarell, A.C., Maxwell, S., Aminov, R.I., 2009. Fate and transport of antibiotic residues and antibiotic resistance genes following land application of manure waste. J. Environ. Q 38, 1086–1108.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, pp. 785–794.

Cheng, G., Ning, J., Ahmed, S., Huang, J., Ullah, R., An, B., Hao, H., Dai, M., Huang, L., Wang, X., Yuan, Z., 2019. Selection and dissemination of antimicrobial resistance in agri-food production. Antimicrob. Resist. Infect. Control. 8 (158).

Cherkasov, A., Hilpert, K., Jenssen, H., Fjell, C.D., Waldbrook, M., Mullaly, S.C., Volkmer, R., Hancock, R.E.W., 2008. Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. ACS Chem. Biol. 4, 65–74.

Christner, B.C., Morris, C.E., Foreman, C.M., Cai, R., Sands, D.C., 2008. Ubiquity of biological ice nucleators in snowfall. Science 319 (1214).

Dadgostar, P., 2019. Antimicrobial resistance: Implications and costs. Infect. Drug Resist. 12, 3903–3910.

Davies, J., Davies, D., 2010. Origins and evolution of antibiotic resistance. Microbiol. Mol. Biol. Rev. 74, 417–433.

D'Costa, King, Kalan, Morar, Sung, Schwarz, Froese, Zazula, Calmels, Debruyne, Golding, Poinar, Wright, 2011. Antibiotic resistance is ancient. Nature 477, 457–461.

Del Prete, R., Ronga, L., Addati, G., Magrone, R., Abbasciano, A., Decimo, M., Mosca, A., Miragliotta, G., 2019. Trends in klebsiella pneumoniae strains isolated from the bloodstream in a teaching hospital in southern Italy. Infez. Med. 27, 17–25.

Diggle, S.P., Whiteley, M., 2019. Microbe profile: *Pseudomonas aeruginosa*: opportunistic pathogen and lab rat. Microbiology 166, 30–33.

Diseases, G., Collaborators, I., 2020. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396, 1204–1222.

Elfadadny, A., et al., 2024. Antimicrobial resistance of pseudomonas aeruginosa: navigating clinical impacts, current resistance trends, and innovations in breaking therapies. Front. Microbiol. 15, 1374466.

European Centre for Disease Prevention and Control, 2023. Carbapenem- and/or colistin-resistant klebsiella pneumoniae in Greece: molecular follow-up survey 2022. https://www.ecdc.europa.eu/sites/default/files/documents/k-pneumoniae-greece-report-november-2023.pdf. (Accessed: 27 December 2024).

FAO, 2023a. AQUASTAT. FAO's Global Information System on water and agriculture. http://www.fao.org/aquastat/en/. (Accessed: 28 February 2024).

FAO, 2023b. FAOSTAT. FAO's global information system on food and agriculture data. https://www.fao.org/faostat. (Accessed: 28 February 2024).

Ferri, M., Ranucci, E., Romagnoli, P., Giaccone, V., 2017. Antimicrobial resistance: A global emerging threat to public health systems. Crit. Rev. Food Sci. Nutr. 57, 2857–2876.

Flandroy, L., Poutahidis, T., Berg, G., Clarke, G., Dao, M.C., Decaestecker, E., Furman, E., Haahtela, T., Massart, S., Plovier, H., Sanz, Y., Rook, G., 2018. The impact of human activities and lifestyles on the interlinked microbiota and health of humans and of ecosystems. Sci. Total. Environ. 627, 1018–1038.

Fleming-Dutra, K.E., et al., 2016. Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010–2011. JAMA 315, 1864–1873.

Folkesson, A., Jelsbak, L., Yang, L., Johansen, H.K., Ciofu, O., Hoiby, N., Molin, S., 2012. Adaptation of pseudomonas aeruginosa to the cystic fibrosis airway: An evolutionary perspective. Nat. Rev. Microbiol. 10, 841–851.

Gómez-Zorrilla, S., Suarez, C., 2023. Antibiotics special issue: Multidrug resistance in pseudomonas aeruginosa: Resistance mechanisms, epidemiology, and impact of resistance on virulence, clinical outcomes, and treatment. https://www.mdpi.com/journal/antibiotics/special_issues/Pseudomonas_Aeruginosa_Antibiotics. (Accessed: 27 December 2024).

Google Earth Engine, 2023. Google earth engine. https://earthengine.google.com. (Accessed: 28 February 2024).

Hocquet, D., Muller, A., Bertrand, X., 2016. What happens in hospitals does not stay in hospitals: Antibiotic-resistant bacteria in hospital wastewater systems. J. Hosp. Infect. 93, 395–402.

Holmes, A.H., Moore, L.S.P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P.J., Piddock, L.J.V., 2016. Understanding the mechanisms and drivers of antimicrobial resistance. Lancet 387, 176–187.

Hu, Y., et al., 2013. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. Nat. Commun. 4 (2151).

Hydrosheds, 2023. Hydrosheds, seamless hydrographic data for global and regional applications. https://www.hydrosheds.org. (Accessed: 28 February 2024).

Institute for Health Metrics and Evaluation, 2024. Antimicrobial resistance (AMR). https://www.healthdata.org/. (Accessed: 28 February 2024).

International Water Management Institute, 2010. Managing Water for Rainfed Agriculture. IWMI Water.

Jiang, P., Sun, S., Goh, S.G., Tong, X., Chen, Y., Yu, K., He, Y., Gin, K.Y.H., 2024. A rapid approach with machine learning for quantifying the relative burden of antimicrobial resistance in natural aquatic environments. Water Res. 262, 122079.

Jiménez-Luna, J., Grisoni, F., Schneider, G., 2020. Drug discovery with explainable artificial intelligence. Nat. Mach. Intell. 2, 573–584.

Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C.S., Singer, H.P., Johnson, D.R., Zhang, T., Bürgmann, H., 2019. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. ISME J. 13, 346–360.

Karkman, A., Parnanen, K., Larsson, D.G.J., 2019. Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. Nat. Commun. 10 (80).

Kempner, W., Schlayer, C., 1942. Effect of CO2 on the growth rate of the pneumococcus. J. Bacteriol. (43).

Kraemer, S.A., Ramachandran, A., Perron, G.G., 2019. Antibiotic pollution in the environment: From microbial ecology to public policy. Microorganisms 7 (180).

Kümmerer, K., 2009. Antibiotics in the aquatic environment - A review - part II. Chemosphere 75, 435–441.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. J. Stat. Softw. 36, 1–13.

Lacalamita, A., Serino, G., Pantaleo, E., Monaco, A., Amoroso, N., Bellantuono, L., Piccinno, E., Scalavino, V., Dituri, F., Tangaro, S., Bellotti, R., Giannelli, G., 2023. Artificial intelligence and complex network approaches reveal potential gene biomarkers for hepatocellular carcinoma. Int. J. Mol. Sci. 24 (15286).

Larsson, D.G.J., 2014. Pollution from drug manufacturing: Review and perspectives. Phil. Trans. R. Soc. B 369, 20130571.

Larsson, D.G.J., de Pedro, C., Paxeus, N., 2007. Effluent from drug manufactures contains extremely high levels of pharmaceuticals. J. Hazard. Mater. 148, 751–755.

Li, L.G., Yin, X., Zhang, T., 2018a. Tracking antibiotic resistance gene pollution from different sources using machine-learning classification. Microbiome 6 (93).

Li, J., et al., 2018b. Global survey of antibiotic resistance genes in air. Environ. Sci. Technol. 52, 10975–10984.

Liao, J., Huang, H., Chen, Y., 2019. CO2 promotes the conjugative transfer of multiresistance genes by facilitating cellular contact and plasmid transfer. Environ. Int. 129, 333–342.

Liguori, K., Keenum, I., Benjamin, C.D., Calarco, J., Milligan, E., Harwood, V.J., Pruden, A., 2022. Antimicrobial resistance monitoring of water environments: A framework for standardized methods and quality control. Environ. Sci. Technol. 56, 9149–9160.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, pp. 4768–4777.

Manaia, C.M., et al., 2018. Antibiotic resistance in wastewater treatment plants: tackling the black box. Environ. Int. 115, 312–324.

McEachran, A.D., Blackwell, B.R., Hanson, J.D., Wooten, K.J., Mayer, G.D., Cox, S.B., Smith, P.N., 2015. Antibiotics, bacteria, and antibiotic resistance genes: aerial transport from cattle feed yards via particulate matter. Environ. Health Perspect. 123, 337–343.

Melo, M.C.R., Maasch, J.R.M.A., de la Fuente-Nunez, C., 2021. Accelerating antibiotic discovery through artificial intelligence. Commun. Biol. 4 (1050).

Michael, C.A., Dominey-Howes, D., Labbate, M., 2014. The antimicrobial resistance crisis: Causes, consequences, and management. Front. Public Heal. 2, 145.

Miller, T., 2019. Explanation in artificial intelligence: insights from the social sciences. Artificial Intelligence 267, 1–38.

Morgan, D.J., Okeke, I.N., Laxminarayan, R., Perencevich, E.N., Weisenberg, S., 2011. Non-prescription antimicrobial use worldwide: a systematic review. Lancet Infect. Dis. 11, 692–701.

Murray, C.J.L., et al., 2022. Global burdern of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399, 629–655.

Nadimpalli, M., et al., 2018. Combating global antibiotic resistance: emerging one health concerns in lower and middle income countries. Clin. Infect. Dis. 66, 963–969.

Noman, S.M., et al., 2023. Machine learning techniques for antimicrobial resistance prediction of pseudomonas aeruginosa from whole genome sequence data. Comput. Intell. Neurosci. 2023, 5236168.

Our World in Data, 2023. Our world in data. https://ourworldindata.org. (Accessed: 28 February 2024).

Pallares-Vega, R., Blaak, H., van der Plaats, R., de Roda Husman, A.M., Leal, L.H., van Loosdrecht, M.C.M., Weissbrodt, D.G., Schmitt, H., 2019. Determinants of presence and removal of antibiotic resistance genes during WWTP treatment: a cross-sectional study. Water Res. 161, 319–328.

Partridge, S.R., Kwong, S.M., Firth, N., Jensen, S.O., 2018. Mobile genetic elements associated with antimicrobial resistance. Clin. Microbiol. Rev. 31, e00088–17.

Pei, R., Kim, S.C., Carlson, K.H., Pruden, A., 2006. Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). Water Res. 40, 2427–2435.

Petty, N.K., et al., 2014. Global dissemination of a multidrug resistant escherichia coli clone. Proc. Natl. Acad. Sci. USA 111, 5694–5699.

Popa, S.L., et al., 2022. Deep learning and antibiotic resistance. Antibiotics 11 (1674).

Pormohammad, A., Nasiri, M.J., Azimi, T., 2019. Prevalence of antibiotic resistance in *Escherichia coli* strains simultaneously isolated from humans, animals, food, and the environment: a systematic review and meta-analysis. Infect. Drug Resist. 12, 1181–1197.

Principe, L., et al., 2014. Epidemic diffusion of OXA-23-producing *Acinetobacter baumannii* isolates in Italy: results of the first cross-sectional countrywide survey. J. Clin. Microbiol. 52, 3004–3010.

Pruden, A., Arabi, M., Storteboom, H.N., 2012. Correlation between upstream human activities and riverine antibiotic resistance genes. Environ. Sci. Technol. 46, 11541–11549.

Pruden, A., Pei, R., Storteboom, H., Carlson, K.H., 2006. Antibiotic resistance genes as emerging contaminants: Studies in northern Colorado. Environ. Sci. Technol. 40, 7445–7450.

Ramakrishnan, B., Venkateswarlu, K., Sethunathan, N., Megharaj, M., 2019. Local applications but global implications: Can pesticides drive microorganisms to develop antimicrobial resistance? Sci. Total Environ. 654, 177–189.

Rather, I.A., Kim, B.C., Bajpai, V.K., Park, Y.H., 2017. Self-medication and antibiotic resistance: Crisis, current challenges, and prevention. Saudi J. Biol. Sci. 24, 808–812.

Reinthaler, F.F., Posch, J., Feierl, G., Wüst, G., Haas, D., Ruckenbauer, G., Mascher, F., Marth, E., 2003. Antibiotic resistance of E. Coli in sewage and sludge. Water Res. 37, 1685–1690.

Rice, L.B., 2008. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE. J. Infect. Dis. 197, 1079–1081.

Rohr, J.R., et al., 2019. Emerging human infectious diseases and the links to global food production. Nat. Sustain. 2, 445–456.

Ryan, K., Ray, G., 2004. Sherris Medical Microbiology, fourth ed. McGraw-Hill Medical.

Sakagianni, A., Koufopoulou, C., Feretzakis, G., Kalles, D., Verykios, V.S., Myrianthefs, P., Fildisis, G., 2023. Using machine learning to predict antimicrobial resistance-a literature review. Antibiotics 12 (452).

Szczepanowski, R., Linke, B., Krahn, I., Gartemann, K.H., Gützkow, T., Eichler, W., Pühler, A., Schlüter, A., 2009. Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria, showing reduced susceptibility to selected antibiotics. Microbiol. (Reading) 155, 2306–2319.

Tacconelli, E., Carrara, E., Savoldi, A., 2018. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect. Dis. 18, 318–327.

Tacconelli, E., Pezzani, M.D., 2019. Public health burden of antimicrobial resistance in Europe. Lancet Infect. Dis. 19, 4–6.

Tang, K.L., Caffrey, N.P., Nóbrega, D.B., Cork, S.C., Ronksley, P.E., Barkema, H.W., Polachek, Alicia J., Ganshorn, H., Sharma, N., Kellner, J.D., Ghali, W.A., 2017. Restricting the use of antibiotics in food-producing animals and its associations with antibiotic resistance in food-producing animals and human beings: a systematic review and meta-analysis. Lancet Planet Heal. 1, e316–e327.

Tenaillon, O., Skurnik, D., Picard, B., Denamur, E., 2010. The population genetics of commensal *Escherichia coli*. Nat. Rev. Microbiol. 8, 207–217.

United Nations, 2015. The global goals. https://www.globalgoals.org/goals/. (Accessed: 28 February 2024).

Van Boeckel, T.P., Brower, C., Gilbert, M., Grenfell, B.T., Levin, S.A., Robinson, T.P., Teillant, A., Laxminarayan, R., 2015. Global trends in antimicrobial use in food animals. Proc. Natl. Acad. Sci. USA 112, 5649–5654.

Ventola, C.L., 2015. The antibiotic resistance crisis: part 1: causes and threats. P. T 40, 277–283.

Wagg, C., Hautier, Y., Pellkofer, S., Banerjee, S., Schmid, B., van der Heijden, M.G., 2021. Diversity and asynchrony in soil microbial communities stabilizes ecosystem functioning. Elife 23, e62813.

Wang, J., Huang, R., Liang, Y., Long, X., Wu, S., Han, Z., Liu, H., Huangfu, X., 2024. Prediction of antibiotic sorption in soil with machine learning and analysis of global antibiotic resistance risk. J. Hazard. Mater. 466, 133563.

Watts, J.E.M., Schreier, H.J., Lanska, L., Hale, M.S., 2017. The rising tide of antimicrobial resistance in aquaculture: Sources, sinks and solutions. Mar. Drugs 15, 158.

Wellington, E.M.H., et al., 2013. The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria. Lancet Infect. Dis. 13, 155–165.

Woolhouse, M., Ward, M., Van Bunnik, B., Farrar, J., 2015. Antimicrobial resistance in humans, livestock and the wider environment. Philos. Trans. R. Soc. B Boil. Sci. 370, 20140083.

World Bank, 2011. World bank atlas method. https://web.archive.org/web/20160303202552/http://econ.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,contentMDK:20452009~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html. (Accessed: 28 February 2024).

World Health Organization, 2023. World health organization. https://www.who.int/data. (Accessed: 28 February 2024).

Yeom, J., Shin, J.H., Yang, J.Y., Kim, J., Hwang, G.S., 2013. (1)h NMR-based metabolite profiling of planktonic and biofilm cells in *Acinetobacter baumannii* 1656-2. PLOS ONE 8, e57730.

Zhou, Z., Shuai, X., Lin, Z., Yu, X., Ba, X., Holmes, M.A., Xiao, Y., Gu, B., Chen, H., 2023. Association between particulate matter pm2.5 air pollution and clinical antibiotic resistance: a global analysis. Lancet Planet. Heal. 7, e649–e659.

Zhu, Y.G., Johnson, T.A., Su, J.Q., Qiao, M., Guo, G.X., Stedtfeld, R.D., Hashsham, S.A., Tiedje, J.M., 2013. Diverse and abundant antibiotic resistance genes in Chinese swine farms. Proc. Natl. Acad. Sci. USA 110, 3435–3440.