UNIVERSITÀ DEGLI STUDI DI CATANIA

DIPARTIMENTO DI MATEMATICA E INFORMATICA

International PhD in Informatics

XXXIV cycle

Giuliana Favara

PhD Thesis

Application of Public Health Informatics to monitor and prevent Healthcare Associated Infections and related outcomes in Intensive Care Units

Tutor: Prof.ssa Antonella Agodi

PhD Coordinator: Prof. Sebastiano Battiato

academic year 2020-2021

Table of content

List of publications **5**

Acknowledgments **6**

Abbreviations **7**

1. Introduction **11**

1.1 Intensive Care Unit **11**

1.1.1 Description of Intensive Care Unit patient 11

1.1.2 Infection, Prevention and Control practices 12

1.1.3 Surveillance 13

1.1.4 Scoring systems 14

1.2 Risk factors in Intensive Care Unit **16**

1.2.1 Patients and Hospital Characteristics 16

1.2.2 Invasive devices, Antibiotic use, and Multi Drug Resistance 17

1.3 Adverse outcomes in Intensive Care Unit **19**

1.3.1 Healthcare Associated Infections 19

1.3.2 Sepsis 21

1.3.3 Mortality 22

1.4 Health Informatics **23**

1.4.1 Precision Public Health 23

1.4.2 Predictive models 24

1.4.3 Targeted selective treatments 25

1.5 Novel tools in healthcare **26**

1.5.1 Artificial intelligence 27

1.5.2 Machine learning algorithms 28

1.5.3 Deep learning 30

2. Objectives **32**

2.1 Rationale and specific objectives **32**

3. Methods **34**

3.1 Study design and setting **34**

3.2 Participants **34**

3.3 Data collection **35**

3.4 Data processing and cleaning **35**

3.5 Definitions of exposure variables and outcomes **36**

3.6 Statistical analysis **37**

3.7 Cluster analysis **38**

3.8 Visual Analysis **39**

3.9 Learning Model generation **39**

3.9.1 Machine Learning algorithm 39

3.9.2 Support Vector Machine algorithm 41

3.10 Tables and figures **44**

4. Results 46

4.1 Analysis by gender of risk factors and outcomes of patients admitted to Intensive Care Unit 46

4.1.1 Study population 46

4.1.2 Characteristics of patients 46

4.1.3 Survival analysis 47

4.2. Application of Support Vector Machine to predict Healthcare-Associated Infections 48

4.2.1 Study population 48

4.2.2 Characteristics of patients 48

4.2.3 ROC curve analysis using conventional statistical approach 49

4.2.4 ROC curve analysis using the Support Vector Machine model 49

4.2.5 Table and figures 51

4.3 Application of cluster analysis to identify patients at risk of catheter-associated urinary tract infections 60

4.3.1 Study population 60

4.3.2 Characteristics of clusters 60

4.3.3. Duration of urinary catheterization 61

4.3.4 Incidence of catheter-associated urinary tract infections and sepsis **61**

4.3.5 Tables and Figures 63

4.4 Application of cluster analysis to identify patients at risk of pneumonia 65

4.4.1 Study population 65

4.4.2 Characteristics of clusters 65

4.4.3 Outflow of patients from Intensive Care Unit admission to diagnosis of pneumonia 66

4.4.4 Association of pneumonia with the risk of sepsis and death 66

4.4.5 Table and Figures 68

4.5 Analysis of risk factors and outcomes in patients with sepsis 70

4.5.1 Study population 71

4.5.2 Characteristics of patients 71
4.5.3 ROC curve analysis using conventional statistical approach 71
4.6. Application of Support Vector Machine to predict seven-day mortality 73
4.6.1. Study population 73
4.6.2 ROC curve analysis using conventional statistical approach 73
4.6.3. ROC curve analysis using the Support Vector Machine model 74
4.6.4. Support Vector Machine maintained its predictive ability among patients who did not develop Healthcare Associated Infection 75
4.6.6 Tables and figures 77
4.7. Application of Cluster analysis to compare incidence of Healthcare-Associated Infections among Intensive Care Units 91
4.7.1 Characteristics of Intensive Care Units 91
4.7.2 Characteristics of clusters at hospital level 91
4.7.3 Characteristics of clusters at Intensive Care Unit level 92
4.7.4 Incidence of Healthcare Associated Infections across clusters 92
4.8. Analysis of hospital and Intensive Care Unit characteristics associated with the incidence of Healthcare Associated Infections 93
4.8.1 Characteristics of Intensive Care Units 93
4.8.2 Comparison of Intensive Care Units' characteristics between Italian regions 93
4.8.3 Association between Intensive Care Units characteristics and Healthcare Associated Infections . 93
5. Discussion 95
6. Conclusion and future perspectives 100
7. References 102

List of publications

This PhD thesis is based on the following original publications:

- "Gender-differences in risk factors and outcomes among patients admitted to Italian Intensive Care Units: findings from the SPIN-UTI network". Manuscript in preparation.
- Barchitta M, Maugeri A, Favara G, Riela PM, Gallo G, Mura I, Agodi A; SPIN- UTI Network. A machine learning approach to predict healthcare-associated infections at intensive care unit admission: findings from the SPIN-UTI project. J Hosp Infect. 2021 Jun; 112:77-86. doi: 10.1016/j.jhin.2021.02.025. Epub 2021 Mar 5. PMID: 33676936.
- Barchitta M, Maugeri A, Favara G, Riela PM, La Mastra C, La Rosa MC, San Lio RM, Gallo G, Mura I, Agodi A; SPIN-UTI Network. Cluster analysis identifies patients at risk of catheter-associated urinary tract infections in intensive care units: findings from the SPIN-UTI Network. J Hosp Infect. 2021 Jan;107:57-63. doi: 10.1016/j.jhin.2020.09.030. Epub 2020 Oct 2. PMID: 33017617.
- G. Favara, P. M. Riela, A. Maugeri, M. Barchitta, G. Gallo and A. Agodi, "Risk of Pneumonia and Associated Outcomes in Intensive Care Unit: An Integrated Approach of Visual and Cluster Analysis," 2019 IEEE World Congress on Services (SERVICES), 2019, pp. 289-294, doi: 10.1109/SERVICES.2019.00083.
- "Risk assessment and outcomes in patients with sepsis in ICU: results of the SPIN-UTI project". Manuscript in preparation
- Barchitta M, Maugeri A, Favara G, Riela PM, Gallo G, Mura I, Agodi A. Early Prediction of Seven-Day Mortality in Intensive Care Unit Using a Machine Learning Model: Results from the SPIN-UTI Project. J Clin Med. 2021 Mar 2;10(5):992. doi: 10.3390/jcm10050992. PMID: 33801207; PMCID: PMC7957866.
- "The Association of Hospital and Intensive Care Unit characteristics with the Incidence of Healthcare associated infections: findings from the SPIN-UTI project". Manuscript in preparation

- "Incidence of healthcare associated infections in Italian intensive care units: a clustering approach on data from the SPIN-UTI network". Manuscript in preparation

The publications have been adapted with the permission of the copyright owners.

Acknowledgments

The research was carried out at the Department of Medical and Surgical Sciences and Advanced Technologies "GF Ingrassia", University of Catania (Catania, Italy) in collaboration with the Department of Mathematics and Informatics, University of Catania (Catania, Italy). I'm also grateful to the Italian Study Group of Hospital Hygiene (GISIO) of the Italian Society of Hygiene, Preventive Medicine and Public Health (SItI), and to all participants and collaborators of the "Italian Nosocomial Infections Surveillance in Intensive Care Units" SPIN-UTI project, at local and national level.

Abbreviations

Infection prevention and control **(IPC)**

Healthcare-associated infection **(HAI)**

Antimicrobial resistance **(AMR)**

Sustainable Development Goals **(SDGs)**

Water, sanitation, and hygiene **(WASH)**

World Health Organization **(WHO)**

Hand Hygiene Self-Assessment Framework **(HHSAF)**

Point prevalence survey **(PPS)**

Prognostic scoring systems **(PSS)**

Acute Physiology and Chronic Health Evaluation **(APACHE)**

Simplified Acute Physiology Score **(SAPS)**

Mortality Probability Model **(MPM)**

Glasgow Coma Scale **(GCS)**

Length of hospital and ICU stay **(LOS)**

Endotracheal tubes **(ETT)**

Urinary catheters **(UC)**

Central venous catheters **(CVC)**

Disability-adjusted life-years **(DALY)**

Multidrug Resistance **(MDR)**

Medical errors **(MEs)**

Adverse events **(AEs)**

Central line-associated bloodstream infections **(CLABSIs)**

Catheter-associated urinary tract infections **(CAUTIs)**

Intubator-associated pneumonia **(IAP)**

European prevalence of infection in intensive care **(EPIC I)**

European Center for Disease Control and Prevention **(ECDC)**

Systemic inflammatory response syndrome **(SIRS)**

Altered white blood cell count **(WBC)**

C-reactive protein **(CRP)**

Erythrocyte sedimentation rate **(ESR)**

White blood cells **(WBC)**

Procalcitonin **(PCT)**

Electronic health records **(EHRs)**

Artificial intelligence **(AI)**

Organization for Economic Co-operation and Development **(OECD)**

Electronic medical records **(EMR)**

Machine learning **(ML)**

Natural language processing **(NLP)**

Multi- layer perceptron **(MLP)**

Convolutional Neural Network **(CNN)**

Self-organizing map **(SOM)**

Autoencoder **(AE)**

Restricted Boltzmann machines **(RBM)**

Deep belief network **(DBN)**

Backpropagation neural network **(BPNN)**

Generative adversarial network (**GAN)**

Positive Predictive Value **(PPV)**

Decision tree **(DT)**

Random forest **(RF)**

k-nearest neighbour classifier **(k-NN)**

Naïve Bayes classifier **(NB)**

Artificial neural networks **(ANN)**

Adaptive boosting **(adaBoost)**

t-distributed stochastic neighbour embedding **(t-SNE)**

Uniform Manifold Approximation and Projection **(UMAP)**

Positive airway pressure **(CPAP)**

Blood urea nitrogen **(BUN)**

Health informatics **(HI)**

Precision public health **(PPH)**

Information technology **(IT)**

Principal Component Analysis **(PCA)**

Integrated Clustering of Multi-dimensional biomedical data **(ICM)**

Hospitals in Europe Link for Infection Control through Surveillance **(HELICS)**

Healthcare-Associated Infections in Intensive Care Units **(HAIICU)**

'Not only Structured Query Language' **(NoSqL)**

Intubation-associated pneumonia **(IAP)**

Interquartile range **(IQR)**

Standard error **(SE)**

Receiver operating characteristic **(ROC)**

Area Under the Curve **(AUC)**

Cluster Features **(CF)**

Schwarz's Bayesian Information Criterion **(SBIC)**

Akaike Information Criterion **(AIC)**

Support Vector Classification **(SVC)**

Support Vector Machine **(SVM)**

K-Nearest Neighbor **(K-NN)**

Synthetic Minority Over-sampling Technique **(SMOTE)**

Confidence interval **(CI)**

Severe Acute Respiratory Syndrome Coronavirus 2 **(SARS-CoV-2)**

1. Introduction

1.1 Intensive Care Unit

Hospitals play a critical role in managing and improving health status and well-being of individuals worldwide [1]**.** For the first time, in 1952, the term intensive-care unit (ICU) was used in Copenhagen during the polio epidemic, aiming at reducing the mortality rate of polio patients using respiratory device [1-3]. Over the recent years, ICUs which provide care for critical patients, began to be considered the core component of the hospital with specialist staff and complex equipment [1,4,5]. Since severity of patients and their long disease history make difficult diagnosis, prognosis and risk management, ICUs receive more attention than other hospital wards. Patients staying in ICUs, indeed, are mainly older and highly heterogeneous, and came from different medical and surgical wards. This in turn, involves the use of medical devices, specialized professionals, as well as of invasive diagnostic and therapeutic procedures [6,7].

1.1.1 Description of Intensive Care Unit patient

Although patient healthcare has improved, it still represents a pressing public health challenge which fall within the reduction of clinical risk in ICU, in particular, and in hospital in general [2,8,9].

Commonly, patients are admitted in ICU when are at higher risk of severe complications or in critical ill conditions. By contrast, when these patients have no choice to return to a better quality of life, in general they should not be admitted to the ICU [7]. Moreover, this patients' group is characterized by large clinical heterogeneity, especially referred to prolonged hospital and ICU stays and to the need for invasive procedures and treatments [8]. Indeed, they are often immune-compromised, and more likely to be intubated and catheterized than those staying in other hospital wards [10,11]. These factors, in turn, are associated with lower survival and adverse outcomes, including healthcare-associated infection (HAIs), sepsis and mortality [8,12,13]. Patients admitted to ICUs have a worse clinical prognosis and prolonged hospital stays [14], suggesting the need of novel approaches to monitor their disease severity and to early predict health deterioration [15-19].

1.1.2 Infection, Prevention and Control practices

Infection prevention and control (IPC) is a crucial strategy of all health systems useful both for patients using care and for those providing care [20]. This strategy aims at preventing the spread of HAIs and to ensure healthcare quality among patients and healthcare workers. Specifically, HAIs prevention is crucial to guarantee patients care quality, to contain the spread of antimicrobial resistance (AMR) and to avoid adverse outcomes [21]. IPC programs are also crucial in achieving of the Sustainable Development Goals (SDGs), especially those related to universal health coverage and the reduction of neonatal and maternal mortality, as well as to universal access to water, sanitation, and hygiene (WASH). For these reasons, most recent World Health Organization (WHO) guidelines, identified eight core components of effective IPC programs at national and facility levels, to support the prevention of HAI and to counteract AMR. Moreover, these evidence-based guidelines take into account the balance between benefits and costs, the resource availability, and the patient preferences [20,22]. Establishing IPC programs is a core component to achieve safe high-quality healthcare practices. Moreover, a good built environment-constituted by appropriate infrastructures, materials, equipment, bed occupancy and human resources- is crucial at facility level to support the implementation of IPC guidelines.

Among the most common practices, the WHO considered hand hygiene as a simply reliable essential action to tackle infectious diseases in all healthcare facilities. For these reasons, the WHO launched a Global Patient Safety Challenge with the slogan of "*Clean care is safer care*" in 2005, whose world day is celebrated every year on May 5 [23]. In 2010, WHO developed an efficient tool, Hand Hygiene Self-Assessment Framework (HHSAF) to evaluate the application of hand hygiene practices in health care facilities [24]. To date, proper hand hygiene is of crucial importance for the major public health challenges, such as reducing the spread of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), preventing HAIs, and combating AMR.

1.1.3 Surveillance

Surveillance is defined as *"the ongoing, systematic collection, analysis, and interpretation of health data"*, which is crucial to plan and evaluate public health practices. Of note, the WHO considered surveillance as one of the core components for effective IPC programs against HAIs [20]. In 1965, the Director of the WHO established the epidemiological surveillance unit in WHO's Division of Communicable Diseases. In 1968, the 21st World Health Assembly affirmed the three main characteristics of surveillance: a) the systematic collection of data, b) the evaluation of these data, and c) the prompt dissemination of results to take actions. Surveillance also involves the continuous gathering of health data needed to monitor the population's health status, aiming at timely identifying patients needed. In general, surveillance model includes four major pillars: i) data collection; ii) data analysis and interpretation; iii) dissemination of results; iv) targeted interventions [25].

Surveillance is intended as "providing *information* to respond to public health threats and opportunities, or to organize future actions", having a key role in identifying efforts and priorities and providing a practical benchmark to analyze trends over the years [26]. The large amount of surveillance data needs to be shared and analyzed in order to monitor trends in risk factors and to protect population from risk exposures and adverse health events. For instance, working on epidemiological data of patients staying in ICU implies explaining how data is obtained and how should be interpreted. Moreover, surveillance of particular health-related condition should be repeated periodically to ensure that it is meeting its proposed objectives and to identify those elements and limitations that should be ameliorate [27,28].

Although surveillance can have many purposes, it usually begins by identifying stakeholders who benefit from, use, and contribute to surveillance, and by collecting documents, forms, and reports that aim at providing public health recommendations [29].

Passive collection of data generated from automated patient records remain the most common less demanding surveillance, but also more suitable to low sensitivity and to misclassification and underreporting biases. By contrast, active surveillance typically had higher specificity and sensitivity

than passive one, even if it requires a greater engagement for healthcare professionals. Regarding to the timing of data collection, surveillance may be also classified as retrospective - when based on previously recorded routine data after patient discharge - and prospective - considered the gold standard for the collection of reliable and timely information.

Moreover, the point prevalence surveys (PPSs) provide an overall framework with point estimates of the prevalence of HAIs. Although prevalence studies have some intrinsic limitations, they require much less resources than prospective studies to estimate the burden of HAIs and the efficacy of IPC programs at national and international level [28,30].

In 2016, the European Centre for Disease Prevention and Control (ECDC) coordinated the second European PPS of HAIs and antimicrobial use in acute care hospitals, representing the major survey of its kind performed in Europe. Its estimates indicated that - on a given day - approximately 6% of all patients who stayed in European acute care hospitals were infected with at least one HAI [31,32].

1.1.4 Scoring systems

Prognostic scoring systems (PSS) are routinely used in clinical practice to measure health status or illness severity of patients admitted in ICU. Moreover, these PPS are widely used to assess disease severity organ dysfunction at ICU admission, and to predict adverse outcomes. These scores represent a practical strategy in healthcare settings for early prediction of adverse outcomes, typically mortality, and for characterization of disease severity and organ dysfunction [33]. Indeed, patients staying in hospital or ICU show signs of clinical deterioration within 24 hours prior to an adverse clinical outcome, requiring timely interventions [34]. In fact, PSS are used by hospital care staff to identify early signs of health worsening and to trigger healthcare worker attention or medical emergency team [35,36].

In general, PSS are classified into generic – if used in critically ill patients admitted to ICUs- and specific, if used for assessing an organ disfunction or disease. Among generic scores, the Acute Physiology and Chronic Health Evaluation (APACHE), the Simplified Acute Physiology Score

(SAPS), and the Mortality Probability Model (MPM) can assess disease severity and predict the risk for future adverse outcomes in ICU settings [33]. In 1985 and 1991, APACHE II and APACHE III scores were established for the management of critical patients in ICU presenting different disease types. These two scores, respectively, included twelve and seventeen physiological, also differing in the total score obtained and in the way in which how chronic health status is assessed. Furthermore, the APACHE III is more difficult and requires a longer time [37].

In 1984, the SAPS -which consider thirteen physiological variables and age- was developed and validated in France to predict risk of death in ICU patients [38]. In 1993, Le Gall and colleagues used a logistic regression model to develop the SAPS II. In lined, this validated score is computed including the following seventeen components: age, heart rate, systolic blood pressure, body temperature, Glasgow Coma Score, continuous positive airway pressure, PaO2, FiO2, urinary output, serum urea nitrogen level, sodium, potassium, bicarbonate, bilirubin, white blood Cell, chronic diseases, type of admission. Each component was assessed within 24 h from ICU admission and the worst value was recorded. The total SAPS II was finally computed as the sum of weighted values for each component [39]. To date, SAPS II still constitute the most widely used tool for the prediction of adverse outcomes in ICU [37,40-43].

Although it is not used in all countries, the SAPS III was developed using more complex statistical techniques and including twenty variables divided into three sub-groups, to predict mortality from clinal parameters recorded within one hour from admission [44-46].

Unlike those mentioned above, MPM is based on less variables collected within the first hour of ICU admission, making its usefulness also when laboratory resources are constrained. Moreover, MPM computation does not produce an intermediate physiology score, but instead a probability of mortality [47]. Finally, Glasgow Coma Scale (GCS) is reliable tool to describe the level of consciousness following a severe brain injury. It is based on three parameters like eye opening, verbal response and motor response with total value ranging from1 to 15 [48].

Of note, PSS are also useful for standardizing public health practices and comparing different ICUs in terms quality of patient care. To do this, data prospectively collected from a large number of ICUs patients were to be considered. Since, PSS vary in their efficacy and prediction validity, it is crucial to choose a PSS that accurately predicts the outcomes of interest [37].

1.2 Risk factors in Intensive Care Unit

Patients in ICUs are more likely to had prolonged hospital stays, sepsis and mortality than those staying in other hospital wards, suggesting the need to develop targeted strategies of prevention and control [14]. For these reasons, uncovering the main risk factors of adverse events in ICUs still represents a pressing issue in public health [49,50]. Beyond these, several patients' characteristics- i.e., use of invasive devices and procedures, severity status, type of infection, antibiotic therapy, and pathogen virulence characteristics – are considered the most common factors associated with worse clinical prognosis [49-54] [14,55-57]. Moreover, healthcare indicators at hospital and ICU level (i.e., type and size of ICU and hospital, types of protocol, the presence of healthcare workers) might affect the quality of healthcare system and, in turn, the incidence of adverse outcomes [49-54].

1.2.1 Patients and Hospital Characteristics

Intrinsic patient characteristics constitute some of the leading causes of HAIs on all hospital wards, and especially in ICUs. Patients admitted in ICUs have the major risk factors for adverse outcomes. In line, the frequency of HAIs is higher among patients in ICUs because they have more severe clinical conditions, they are often immunocompromised, and they are more likely to be intubated and catheterized than patients in other hospital wards [10,11]. Moreover, exposure to invasive devices and procedures, general health status and comorbidities, infections with MDR bacteria still pose a challenge for patients' safety [14,58-62].

Beyond characteristics of patients, several indicators at hospital and ICU level might be associated with higher incidence of HAIs and adverse outcomes [49-54]. For instance, the size and type of hospital

and ICU, the type of care protocol adopted, the number of healthcare professionals, and the number of ICU and hospital beds are crucial in guaranteeing good practices in all healthcare settings [63-67] [49-52,58,63-65,67-72]. Indeed, structural, process, and outcome indicators contribute to measure quality and performance of healthcare systems [63-67]. Among these, regional administrations of hospital and ICUs play a crucial role.

For instance, in Italy the comparison between and within regions might represent a useful approach to identify what hospital and ICU indicators are associated with a higher prevalence of HAIs [73-77]. It's worth mentioning that in Italy - a country with decentralized healthcare system - regions of Northern and Southern depict a worrying gap in terms of availability of healthcare assistances [78,79]. For this reason, it is crucial identifying those patient, hospital and ICU characteristics that are mostly associated with the incidence of adverse outcomes in ICUs. Yet, since these factors might act together determining groups of patients, hospitals and wards with specific characteristics, innovative approaches are becoming crucial for designing targeted prevention strategies and for improving decision-making processes [80,81]. However, the evidence on this field is still lacking and guidelines for improving structural and infrastructural characteristics that might affect the risk of HAIs are currently vague [40].

1.2.2 Invasive devices, Antibiotic use, and Multi Drug Resistance

The baseline conditions of critically ill patients pose, per se, leading cause of prolonged length of hospital and ICU stay (LOS), increased medical costs, and death. Although IPC measures have decreased incidence of HAIs, the use of invasive devices still pose represents a risk factor for patients staying in ICU. Indeed, the exposure to invasive devices- endotracheal tubes (ETT), urinary catheters (UC), and central venous catheters (CVC)- is associated with increased HAIs rates in ICU [82,83]. With this in mind, several care bundles, grouping a list of best practices when using an invasive device, have been proposed to counteract HAIs spread and to improve patient outcomes. In addition,

the duration of exposure to invasive devices represents a risk factor for HAIs, suggesting limiting the number of device-days as a successful strategy for infection control [84-87].

Although antibiotics still represent the most common strategy to prevent and treat infectious disease in humans, their misuse and overuse has made them ineffective, as well as drivers in the spread of drug-resistant pathogens [88]. In fact, pathogenic bacteria not being adequately treated by existing antibiotics poses a considerable challenge to the effectiveness and efficiency of healthcare systems [89-95] and a limit for the Agenda 2030 [96]. The AMR has an alarming impact in various setting, also suggesting the worrying impact of HAIs due to antimicrobial resistant bacteria in terms of cases, attributable deaths and disability-adjusted life-years (DALY) [97-99]. The lack of awareness on the AMR itself and on the prudent use of antimicrobials remain the leading causes for overuse and misuse of antimicrobials and inappropriate IPC practices [100-103]. Although no standardized definitions have been made by medical community, multidrug resistance (MDR) occurs when a bacterial infection is resistant to more than one appropriate antimicrobial treatment [89,90,104-108]. Currently, MDR is considered as bacterial non-susceptibility to at least one agent in three antimicrobial classes [89,90]. Every year, about 700.000 deaths due to infections caused by MDR bacteria occurs worldwide. It has been estimated that this situation could exponentially increase in 2050, because of resistance pathways evolution and novel antibiotic discovery [89,96,105,106].

Due to the inefficacy of available treatments, resistant infections, especially those acquired in hospital or in ICU, are associated to a wide range of adverse outcomes and death [89]. Several definitions are used to identify patterns of MDR both in Gram-positive and Gram-negative bacteria, making different data and study protocols difficult to be compared [90]. An alternative method is characterized by the resistance to one key antimicrobial agent. This often reflect the bacteria cross- or co- resistance to multiple classes or subclasses of antimicrobials, thus considering bacteria as MDR [90].

1.3 Adverse outcomes in Intensive Care Unit

Patients' cares undergo the high prevalence of preventable medical errors (MEs) and adverse events (AEs). This in turns depend on the non - complete adherence to best practices and clinical guidelines, which can lead to a higher patient morbidity, mortality, resource utilization and prolonged hospital and ICU stay [109]. Generally, an AE is considered a preventable injury caused by medical staff management in the acute care settings, resulting in prolonged hospitalization and disability at discharge. The AE is considered "preventable" when determined by a MEs, and "negligent" when the standard of care is not met by healthcare providers or it's below the standard required [16,109]. Notably, patients who require transfer from medical ward to the ICU are more prone to experience an AE than those in other hospital wards, resulting in addition life-sustaining interventions [109-112]. Yet, the amount of MEs and related AEs among patients in ICU remains unclear[109].

1.3.1 Healthcare Associated Infections

There has been a large consensus that in the modern medicine, clinical practice and healthcare settings could assume a twofold aspect. More than 2,000 years ago, Hippocrates stated *"I will use treatments for the benefit of the ill in accordance with my ability and my judgment, but from what is to their harm and injustice I will keep them"*. Accordingly, Semmelweis, discussing about puerperal fever, recognized that coming into hospitals can be threatening. This idea is described as unintentional physical injury resulting from medical assistance, requiring additional treatment or hospitalization, or resulting in mortality [113]. Semmelweis is considered as the first one recognized that healthcare providers had a critical role in disease transmission. In 1847, Semmelweis found a higher rate of maternal mortality among patients treated by obstetricians and medical interns than those treated by midwives, thus identifying the way of transmission, and spread of puerperal sepsis. He also noted that a pathologist died of sepsis after getting hurt with a scalpel while doing an autopsy of a patient with puerperal sepsis. Since scalpel and uncleaned hands began to be considered a way to transmit organisms, Semmelweis introduced chlorinated lime hand washing for the healthcare staff [114].

Semmelweis' hypothesis of disease transmission from healthcare providers to patients, was considered valid after Koch's postulates in 1890. For these reasons, Semmelweis is considered the first describing a HAI and suggesting hand hygiene as a preventive strategy in healthcare settings [115]. HAIs are considered as those infections acquired by patients while receiving health care in hospital or in various healthcare settings (i.e. home care, long-term care and ambulatory care facilities, senior nursing home) [116,117].

This term- which recently replaced the terms "nosocomial" or "hospital" infection- also referred to those infections acquired in the hospital, but appearing after discharge, as well as occupational infections among healthcare staff. Specifically, HAI appear two days or more after hospital admission, or until 30 days after receiving care. Moreover, several studies suggest HAIs among the most common adverse events of patients staying in hospital, also including adverse drug events, and post-operative complications [116,118-120]. In fact, HAIs are a pressing safety concern for public health, due to their significant impact on patients' mortality, length of hospital stay, and healthcare costs [116,121]. According to the ECDC, nearly 8% of patients admitted in ICU for more than two days presented with at least one HAI on a given day [10,11].

Among the HAIs, central line-associated bloodstream infections (CLABSIs), catheter-associated urinary tract infections (CAUTIs), and intubator-associated pneumonia (IAP) are the most frequent. [116]. In 1992, the European Prevalence of Infection in Intensive Care (EPIC I) study was conducted in western European ICUs [122]. On the study day, 45% of patients had suspected or proven infection and 62% were receiving prophylactic antibiotics, therapeutic antibiotics, or both types of antibiotics. In 2007, a study of similar design but extending inclusion to ICUs worldwide (Extended Prevalence of Infection in Intensive Care [EPIC II]) was conducted [123]. On the study day, 51% of the patients had suspected or proven infection and 71% were receiving prophylactic antibiotics, therapeutic antibiotics, or both types of antibiotics.

The current EPIC III study (Extended Study on Prevalence of Infection in Intensive Care III) was conducted in 2017 using a similar design to the earlier studies, but also included questions related to

the availability of specific resources for the diagnosis and treatment of infection. It was hypothesized that the prevalence of infection and the associated outcomes would vary among geographic regions [124].

1.3.2 Sepsis

Although novel protocols, antibiotics and innovations have been made, sepsis still remains a critical threat for patient health worldwide [125]. According to the WHO estimates, each year sepsis caused about 11 million of deaths and disabilities. In line, the ECDC estimates that each year in the USA at least 1.7 million of individuals are affected by sepsis. Sepsis is the result of systemic inflammatory response syndrome (SIRS) following an infection, which in turn causes tissues injury and organs dysfunction [125-127]. SIRS is described by fever (>38°C) or hypothermia (<36°C); increased heart rate (>90 beats/minute), tachypnea (>20 breaths/minute), or hyperventilation (PaCO2 < 32 mmHg); and altered white blood cell count (WBC) (>12,000 cells/mm3 or <4,000 cells/mm3) or presence of >10% immature neutrophils [125].

In more severe cases sepsis could evolve into septic shock characterized by at least one major organ dysfunction, arterial hypotension or hypoperfusion, and adverse outcomes [125,126,128]. Generally, the evolution of SIRS to sepsis and/or septic shock occur rapidly, suggesting the need of novel tools to early diagnose sepsis at the onset [125]. Several biomarkers of inflammation, such as C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and white blood cells (WBC) count, are used in diagnosis and monitoring sepsis. Among these, procalcitonin (PCT)- a prohormone of calcitonin- represent the biomarker widely used worldwide [125,129]. Indeed, during the SIRS process, the PCT is secreted in proportionally response to bacterial endotoxins. In fact, a higher PCT production indicates severe sepsis. Moreover, PCT level starts to rise within 2-4 hours of inflammation process and peaks at 8-24 hours, allowing a better diagnosis, prognosis, and monitoring of patients with sepsis progression than other conventional biomarkers [125,130].

Given the nonspecific nature of early symptoms and clinical signs, sepsis early diagnosis to closely monitor all patients in the ICU, and to identify those infections developing sepsis or septic shock might result difficult [126]. Commonly, care for septic shock includes early antibiotics treatment, antifungal administration, regulation of blood volume, and guaranteeing sufficient tissue perfusion [126]. Early detection of sepsis and early administration of antibiotic treatment have been known to be the strongest modulators of outcomes in patients with sepsis [126,131]. Although measuring PCT routinely is useful in ICU setting- both in predicting adverse outcome and in the identification of patients at higher risk of septic shock- early sepsis detection remain still challenging in clinical settings [132].

1.3.3 Mortality

Critical ill patients are admitted to ICU when in critical health status, such as severe infectious diseases, multiple trauma, and organs failure [133-135]. For these reasons, these type of patients generally have an higher risk of adverse outcomes, especially in terms of mortality [14].

Indeed, mortality in ICUs is two time higher among infected patients than those not infected [123,136]. This in turn depends on several patients' factor such as the use of invasive procedures, their severity, type of infection, therapy, and microorganisms' characteristics, including clonal spread [14,55-57].

Several lines of evidence suggested that outcomes of patients staying in ICU is the cumulative results of different risk factors related both to patients' clinical conditions and personal characteristics, and to the exposure to invasive devices and procedures [137-139].

With this in mind, the complexity of HAI burden suggested the need of novel approaches aimed at early identifying patients at higher risk of death [140]. Indeed, the prediction of patients at higher risk of mortality in ICU play a key role in improving patients' survival and in implementing their management [141]. Although several conventional statistical approaches are widely used in clinical practice, modern Machine Learning (ML) models showed more accurate results in the early

identification of patients who are more likely to die during their stay in ICU, considering different risk factors [141-145].

Globally, the prevalence of mortality in ICU ranges from 9% to 61%, with the highest percentage in in the low-income countries [146].

1.4 Health Informatics

To date, health informatics (HI) represents the systematic application of informatics and technology-based innovations to public health issues [147,148]. Recent advance in HI has become so crucial in modern medicine and biomedical research, that most of public health policies and practices depend on information and learning technologies. For these reasons, HI should constitute a key strategy for the implementation of national health information and surveillance, providing a benefit both for healthcare providers and patients.

Indeed, this developing technology helps healthcare providers as follows: i) to early identify high-risk conditions and to avoid misdiagnoses; ii) to improve ward workflow efficiency by providing smart decision support; iii) to maintain a high-quality data collection, clinical documentation and storage [148,149].

1.4.1 Precision Public Health

Precision public health (PPH) is an emerging a multidisciplinary area of interest relying on big data and data science. The term "precision" focused on the implementation of population-level effort, policy, and targeted interventions. Specifically, PPH is defined as *"about using the best available data to target more effectively and efficiently interventions"*

Thanks to the great availability of data sources, PPH is considered a global strategy to ameliorate public health policies. In the last decades, health- and non-health-related data produced by EHRs (i.e., patients characteristics, biomarkers, exposure to invasive devices) increased exponentially, giving rise to the term "big data" [150-153].

Big data is complex and heterogeneous, allowing integrating multiple social and environmental risk factors, as well as to predict the progression from health to disease status. Moreover, these data provide the potential for more "precision" in clinical practice and public health, aiming at finding preventable risk factors and at allowing more precision diagnoses and prognoses. In fact, it has been estimated that each patients' hospitalization produce more than 150.000 data [154].

In this scenario, the use of information technology (IT) and data science have a crucial role in enhancing public health surveillance, which is defined as the systematic collection, management, and interpretation of data to encourage novel actions [150]. Specifically, surveillance systems use data collected from surveys, registries, real-time monitoring systems, and studies having the potential of improving early disease detection [155].

Another scientific field of application of big data in PPH is the potential for using novel predictive models. Indeed, novel data science approaches are useful to integrate different data sources
and to improve prediction of clinical outcomes, diagnosis, and treatment. To achieve the main PPH objectives, data analysis modernization in healthcare settings and public health is needed. In this scenario, bringing together existing data using digitalized approaches could be a strategy to help the application of novel data science approaches in public health. To do this, identifying patients' similarity constitutes a key strategy to address precision medicine challenge, also allowing patients stratification into clinically relevant subgroups [150,156].

1.4.2 Predictive models

In the last decades, more efforts have been made in the field of precision medicine, whose primary aim is to predict health deterioration and to monitor patient disease severity [15-19]. A large amount of patient characteristics is needed to define their similarity or diversity with other patients [156,157]. Indeed, availability of health-related data is crucial for a more precise identification of groups of patients requiring more attention, as well as preventive strategies and selective treatments.

In general, clinical, and molecular characteristics, imaging data, laboratory results and clinical outcomes represent the most useful data domains to develop targeted predictive models. Although molecular data is the most represented category in the process of patients' care, treatment and decision-making, several studies suggest the application of an integrated use of clinical and molecular data [158,159]. In this scenario, employing heterogenous types of data also requires grouping patients according to their personal and clinical similarity, suggesting the need of ad hoc techniques. Indeed, several clustering approaches (i.e., Two-Step, Hierarchical and k- means) are used for managing high dimensional types of data, aiming to group patients according to their disease status, prognosis, or response to treatment. Moreover, also methods of dimensionality reduction (i.e., Principal Component Analysis (PCA), Factor Analysis, or matrix factorization), are frequently used when aiming to select representative characteristics of specific patients' group.

Moreover, several supervised and unsupervised algorithms are commonly used when the goal is assessing a measure of patient's similarity in terms of clinical, diagnostic, treatment profiles and /or expression patterns, to make prediction of a predetermined outcome.

Finally, software tools (i.e., Integrated Clustering of Multi-dimensional biomedical data (ICM)) represent useful approach providing a set of methodologies of data combination, clustering, and visualization, with a specific graphical user interface [156,160].

1.4.3 Targeted selective treatments

One of the main concerns of PPH is to improve selective treatments targeted to specific patient diseases, both considering their similarity and improving clinical outcomes [156]. This in turn, allows clinicians to use the same targeted treatments for those patients having similar prognosis [156,157]. Although PPH approach promises highly personalized and highly effective treatment, it still poses considerable challenges.

Indeed, PPH aims to treat patients with the right drug at the right dose at the right time, also taking into accounts genetic variability of each patient [161]. However, when patients are in critical ill

condition, it is no practical collecting genotyping data. In the last century, medical communities began to adopt a novel approach for patients' treatment based on their clinical characteristics. Although this change could led to fight diseases, the cost of drug discovery and development dramatically increased, so becoming a burden for public health [162].

To date, recent strides in ML led to the development of specific treatments aimed to specific groups of patients with certain needs. Commonly, AI approaches mainly focused on eight key components: drug development, prediction of resistance, identification of appropriate doses, drugs combination, improvement of delivery systems, designing of administration routes, evaluation pharmacokinetic profile, and prediction of adverse outcome [163].

Patients in critical conditions are those who could benefit more from more specific therapies tailored to their clinical requirement. Indeed, they are generally more likely to be exposed to invasive devices and to report underlying disease. For these reasons, data from patients' electronic medical systems allow to identify critically ill patients and to promote treatments personalization. Indeed, the application of PPH approaches in the development of the best treatment plans should consider heterogeneity of patient' characteristics, risk factors, and responses to treatments [161].

Specifically, these applications have a crucial role both in the identification of more effective treatments for infectious diseases, as well as in predicting drug resistance [163].

1.5 Novel tools in healthcare

In the last decades, the collection of large amounts of patient clinical characteristics have increased worldwide [164-166]. Commonly, electronic health records (EHRs) represent a data source which provide the potential for reliable novel tools which could guide understanding patients risk factors, as well as improving their management [140,164,165,167-169].

With this in mind, identifying patients at higher risk of critical health-conditions still represent a major challenge for public health, with so many healthcare epidemiologists which continue to interpret large amount of complex data [167,170]. Although conventional statistical models are the gold standard

when predicting adverse outcomes, novel automated approaches are considered a strategy to optimize accuracy across a wide range of medical applications [171]. Indeed, novel statistical and mathematical methods are crucial to real-time identify patients or subgroups of patients at higher risk during their stay in healthcare settings. With this in mind, the use of digital technologies, particularly in the context of machine learning, artificial intelligence and deep learning, holds great promise for medical research [172].

1.5.1 Artificial intelligence

Artificial intelligence (AI) is referred to an emerging field of data analysis, involving the use of algorithms performed by computer code. These algorithms are capable to conduct rapid analyses for transforming large amount of clinical data into decisions, evidence, and outputs. During the Council on Artificial Intelligence of the Organization for Economic Co-operation and Development (OECD) AI was defined as a machine-based system enable to response to human-defined goals through early predictions and results [172].

Accordingly, the application of novel AI methods could be a strategy to achieve the health-related and wellbeing objectives under the SDG 3, as well as the universal health coverage [172,173]. In 1976, was developed the first medical system, commonly known as MYCIN, able to propose the best antibiotic therapy for several bacterial infections. However, this system had no large application both for its limited involvement into clinical work and for large number of rules needed for its application [174].

Thus, Al is considered a tool supporting the analyses of electronic medical records (EMR) data, which is commonly unstructured and heterogeneous. Since data collected from numerous sources have increased, data still constitute one of the major limitations when introducing AI in healthcare settings [175]. In high-income countries, AI methods are improving the delivery of healthcare services in terms of diagnosis, prevention, and treatments. Indeed, according to Schwabe and Wahl, the application of

AI for health is identified in four domains, as follows: prediction-based diagnosis, morbidity or mortality risk assessment, disease outbreaks and surveillance, and health policy and planning [173]. Firstly, healthcare providers routinely use AI to integrate patient characteristics and to early identify those at higher risk of adverse outcomes. Moreover, AI allows healthcare providers for a prompt risk assessment, as well as to make faster and more accurate diagnoses and to avoid clinical biases. This in turn, suggested that the application of novel AI technologies is useful to predict health outcomes before they occur, also developing preventive strategies targeted to each patient requirements [168,175,176].

A second application of AI for biomedical research is based on the use of EMR data for risk assessment and for optimization of clinical care. In fact, these heterogeneous data can be helpful to develop clinical best practices and to design novel guidelines. Moreover, in the last two decades, the role of AI evolving to facilitate drug discovery and development. In fact, testing of medicines might be based on virtual models of the human body by 2040, allowing prescription drugs tailored to individual patient characteristics [177].

Third, AI might contribute to health systems management and planning. In fact, AI can be used to assist personnel in improving medical resources, supporting decision-making, and scheduling patients' appointment. Finally, several implications about the use of AI for public health surveillance and health-promotion must be considered. In this scenario, AI could be used to early identify patients at higher risk, and to identify the underlying causes of adverse health outcomes [172,175].

1.5.2 Machine learning algorithms

In the last decades, medicine highlighted the need of alternative computational techniques to analyze large amounts of clinical data, and to address clinical decision making [142,178-180]. This novel research area was encouraged by novel devices to collect large volumes of data, by the increase in data storage and processing systems, and by a global system of computer networks to transfer data. [181]. ML, a subset of AI, was referred to those techniques and algorithms which was developed to

automatically learn from data and to interpret unknown situations [174,182]. Moreover, the ML approach represent the switch of paradigm from programming computers to perform a task without having an explicit objective. In line, this approach suggested the notion that humans learn how to solve a complex task from experience, while machines learn from structured and unstructured data using statistics and computer science [144,145,181]. In general, ML has a large field of applications such as predicting treatment targets, diagnosing bacteria causing infectious diseases, classifying antimicrobial resistance, and predicting outcomes [142].

Typically, ML models were classified into "supervised or predictive learning" and "unsupervised or descriptive learning", according to how the model learns from data [143,181].

In particular, supervised ML methods used labeled training set of data to choose the best model able to make outcomes prediction [144]. In this case, the algorithm infers a function from the data that can be used for predicting outputs from different inputs. Indeed, supervised ML methods aim to identify patterns in data and make predictions without having explicit preprogrammed rules, by separating the classes. Notably, decision tree (DT), random forest (RF), k-nearest neighbour classifier (kNN), Naïve Bayes classifier (NB), artificial neural networks (ANN), support vector machine (SVM), and adaptive boosting (adaBoost) are among the most common supervised ML methods.

By contrast, unsupervised methods aim to explore a hidden structure from an unlabeled training set of data, without a predefined outcome of interest. In both these cases, training dataset is necessary for the algorithm, in order to conduct multiple tests aiming at identifying the model with the highest accuracy. Notably, k-means and Apriori algorithms, t-distributed stochastic neighbour embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP) represent the most common unsupervised ML methods [172,174]. In this scenario, clinical data is becoming a major concern in precision medicine and public health, suggesting the need of developing innovative medical decision-making approaches. These data arise from heterogeneous sources of patient conditions and parameters collected before, during and after hospitalization. For instance, ICUs constitute the healthcare setting where the cares of critically patients have a higher data dependence than other

wards. Indeed, applying ML model on the data collect in ICUs, could be useful to identify alarming vital signs and to predict patient outcomes [143,171,183]. Interestingly, several line of evidence suggested that ML algorithms could enrich conventional statistical approaches, especially in terms of prediction of ICU prognosis, clinical deterioration and risk assessment [184].

1.5.3 Deep learning

Deep Learning, a subdomain of ML methods, has been successful to intelligently learn complex patterns and to make prediction from data [185]. Particularly, deep learning, also recognized as "neural network", shows better performance in several applications than conventional ML approaches [186]. Indeed, a deep learning algorithm is able to automatically find unexpected relationships between inputs and outputs in order to identify a hierarchy of more complex feature. However, the output of the model could vary depending on the data characteristics [187].

Commonly, deep learning is classified into supervised, unsupervised, or semi-supervised [172,186]. Although deep learning methodologies could be applied for different medical purposes, choosing that more suitable is still challenging, especially in healthcare applications [182].

For instance, deep learning promises a greater accuracy in the diagnosis based on oncology-oriented and radiology images analysis [186,188]. Moreover, deep learning is also applied for speech recognition, language translation, and natural language processing (NLP) [182].

For instance, Multi- layer Perceptron (MLP) and the Convolutional Neural Network (CNN) are among the most widely used deep learning algorithms. The MLP, also known as the feed-forward artificial neural network, is constituted by a network with one or more input layers and one output layer linked by connection nodes. Specifically, the MLP uses the "Backpropagation" technique to adjust the weight values of the model. By contrast, the CNN method is constituted by convolution layers, pooling layers and fully connected layers, improving the conventional approach [186]. In addition, also the self-organizing map (SOM), the autoencoder (AE), the Restricted Boltzmann

machines (RBM), the deep belief network (DBN), the backpropagation neural network (BPNN) and the generative adversarial network (GAN) are commonly used for various applications [186].

One of the most common conventional approaches to determining a patient's disease, is logistic regression model that weights each patient characteristic, making a weighted sum an accurate predictor of disease.

Since patient disease conditions involve different patient characteristics, deep learning could be the more appropriate method thank to its possibility of considering more characteristics in input [182]. Indeed, deep learning tools also allow the risk stratification for a broad range of patient populations [172,189].

2. Objectives

2.1 Rationale and specific objectives

In line with the current state of the art, improving the management of patients admitted to ICU remains a priority in Public Health. In particular, early identifying patients whose clinical conditions and personal characteristics could led to a higher risk of adverse outcomes in ICU, represent a challenge for clinicians and healthcare professionals worldwide [26,190-194].

To do this, several early warning scores have been proposed as useful predictors of health conditions or illness severity of ICU patients [37,40-42]. Although these conventional statistical approaches are widely used in clinical practice, recent advanced in health informatics proposed more accurate models to identify patients or subgroups of patients who more likely to die or to be affected by HAI during their stay in ICU, also taking into account different sets of risk factors [140-145,168,169]. Indeed, the convenience of large amount of healthcare data have made possible the application of ML methods for predicting specific adverse outcomes, for patient risk stratification and for improving quality of care [195].

For these reasons, firstly was evaluated the performance of the early warning score ability to predict the risk of adverse outcome in patients admitted to ICU using traditional statistical approach. Next, a ML algorithm was developed and tested, considering early warning score in combination with additional characteristics at ICU admission, aiming to further improve the predictive performance.

Moreover, specific objectives of the present pHd thesis are:

- To evaluate gender-differences in risk factors and in the probability of death among patients admitted to ICUs;
- To perform a conventional statistical method to assess the ability of SAPS II in predicting the risk of HAIs in patients admitted in ICUs. Next, the results obtained were compared applying a SVM algorithm, to evaluate its performance in distinguishing between non-infected patients and those who were diagnosed with at least one HAI during their ICU stay;

- To perform a cluster analysis to distinguish patients according to their characteristics at ICU admission, and to identify clusters of patients at higher risk for CAUTIs, pneumonia and associated sepsis. Accordingly, variability across clusters in terms of duration of urinary catheterization, incidence of CAUTIs and associated sepsis was explored. In addition, the risk of sepsis and death associated with pneumonia caused by Acinetobacter baumannii, Klebsiella pneumoniae and Pseudomonas aeruginosa was evaluated;
- To evaluate the ability of the SAPS II to predict the risk of death after seven days from their admission to ICU, using a conventional statistical method. Next, these results were compared with those obtained from an SVM algorithm, which combines the SAPS II with additional patients' characteristics, to further improve the predicting performance;
- To perform a cluster analysis to classify Italian ICUs according to their qualitative and quantitative characteristics [196-198]. Thus, the variability across clusters in terms of incidence density of HAIs was explored. Moreover, the main hospital and ICU indicators associated with HAI incidence were evaluated at national level and compared between Italian regions.

3. Methods

3.1 Study design and setting

The ongoing "Italian Nosocomial Infections Surveillance in Intensive Care Units" SPIN-UTI project was established in 2006 by the Italian Study Group of Hospital Hygiene (GISIO) of the Italian Society of Hygiene, Preventive Medicine and Public Health (SItI). The project is being launched and conducted according to the protocol of the Hospitals in Europe Link for Infection Control through Surveillance (HELICS) Network, and updated in accordance to the ECDC Healthcare-Associated Infections in Intensive Care Units (HAIICU) protocol [199,200]. The study was approved by the ethics committee of the involved institution (Ethics Committee 'Catania 1', Catania, Italy; Protocol nos. 111/2018/PO and 295/2019/EMPO), whose protocol and full details have been described elsewhere [26,43,190-194,201].

In general, each edition of the SPIN-UTI project consists of a six-month prospectively surveillance over a two-year period (i.e., from 1st October of a given year to 31st March of the following year). During the seven biennial editions (i.e., from 2006-2007 to 2018-2019), the SPIN-UTI project has surveyed 22,000 patients admitted to 108 Italian ICUs, collecting data on more than 4,000 infections and 5,000 microorganisms [26,190-194].

3.2 Participants

In brief, hospitals participation is voluntary, and data is handled confidentially. In particular, the Italian ICUs participating to the project must fit the definition established by the European Society of Intensive Care Medicine [4], while neonatal and pediatric ICUs are excluded [26,190-194,201-203].

Moreover, all patients staying more than two days in ICU are included in the SPIN-UTI surveillance (i.e., according to the following formula "*Date of discharge from the ICU" –* "*Date of admission to the ICU* + 1" > 2). By contrast, patients who stay in ICU less than three days are excluded a priori, because one of the primary outcomes of the project is to evaluate the incidence of HAIs, which develop after 48 hours of ICU stay.

3.3 Data collection

The SPIN- UTI project adopts a web-based data collection procedure using an online platform. In general, by using different electronic data forms, data regarding characteristics of hospitals and ICUs, patients, infections, and associated micro-organisms are collected. Indeed, patients with complete assessment on characteristics at admission (e.g., age, sex, SAPS II, origin of the patient, admission type), dates of insertion and removal of invasive devices (e.g., intubation), infection status (i.e., infection date, infection site and associated-microorganisms) and microorganisms (i.e., antimicrobial resistance data) were collected. The electronic questionnaires also aimed to collect information regarding structural and functional characteristics of hospital and ICU participants. In particular, ICUs and hospitals were included if reported complete assessment on the following information: hospital and ICU sizes (i.e., total number of beds in the hospital or in the ICU); hospital type (primary; secondary; tertiary); ICU type; percentage of ICU patients who were intubated in the last year (i.e., percentage of intubated patients over the past year in the ICU); total days in hospital and in ICU in the last year; area of ICU origin (Northern, Central, or Southern Italy); total hospital and ICU costs; incidence density of HAIs [199]. Since 2008, also surveillance of ICU-acquired sepsis has been included in the SPIN-UTI protocol [190,191,199].

3.4 Data processing and cleaning

The workflow of the data processing was carried out as follows: data acquisition, storage, security, anonymization, warehousing and cleaning. Although data of the different SPIN-UTI editions have initially stored in different formats (i.e., SPSS spreadsheets, CSV, etc.), data acquisition involved the use of a uniform protocol. Indeed, a preliminary work was to clean, make uniform and merge the data of the different SPIN- UTI editions. Next, data on the patients' ICU stay, on the use of invasive devices, on the infections and on the microorganisms, which were initially stored in different subsets, were aggregated in an overall dataset.

Specifically, the overall dataset was designed according to the 'not only Structured Query Language' (NoSQL) approach, which is useful when working with a wide variety of data models, including key value, document, columnar and graph formats. Furthermore, a NoSQL database takes the advantages of horizontal scaling and cloud computing environment. In particular, Python data analysis libraries, such as Pandas and Py-Mongo, were used in this study to perform data processing and analysis. Specifically, Py-Mongo contains tools for working with MongoDB software, which represents a NoSQL well- organized platform used to perform conventional statistical analyses. Finally, the online Plotly library was used for graphical data representation with a Sankey diagram [140,204].

3.5 Definitions of exposure variables and outcomes

SAPS II at ICU admission was used as one of the main predictors. Specifically, the SAPS II was computed by imputing the worst value of each component assessed within 24 h from ICU admission [39]. The components included: age; heart rate; systolic blood pressure; temperature; Glasgow Coma Scale, mechanical ventilation, or continuous positive airway pressure (CPAP), PaO2, FiO2, urine output, blood urea nitrogen (BUN), sodium, potassium, bicarbonate, bilirubin, white blood cell; chronic diseases, type of admission. The total SAPS II was finally computed as the sum of weighted values for each component [39].

Patients were considered as admitted with trauma if reporting blunt or penetrating traumatic injury, with or without surgical intervention. Moreover, data on patients who underwent non-surgical treatment for signs and symptoms related to the acute coronary syndrome were considered. Instead, impaired immunity was defined as an impairment due to treatment, diseases or <500 PMN/mm3. Finally, any antibiotic therapy administered in the 48 hours preceding ICU admission and/or during the first two days of ICU admission were collected.

According to the ECDC HAI-ICU protocol, the occurrence of HAI was assessed using a set of clinical and laboratory criteria [200]. In line with the protocol, an infection is considered as HAI acquired in the ICU - if it occurs in the ICU after more than 48 hours. In practice, if considering the day of ICU

admission as the first day, all infections with onset from day three onwards in the ICU should be considered HAIs. Specifically, a HAI was "device-associated" when a device was used in a patient within the 48-hour period before onset of infection. The term "device-associated" is only used for pneumonia, bloodstream infections, and urinary tract infections, when intubation, central vascular catheter, indwelling urinary catheter were used. Specifically, clinical diaries- reporting signs and symptoms, microbiological and laboratory parameters - are used by ICU participants for the validation of the cases of infection.

Thus, pneumonia is defined as intubator-associated pneumonia (IAP) if an invasive respiratory device was used -even intermittently- in the 48 hours preceding the onset of infection [205].

However, an HAI was considered catheter-associated urinary tract infections (CAUTI) when an indwelling urinary catheter was in place within seven days of a positive laboratory result for uropathogens (bacteria or fungi), or when signs and symptoms of urinary tract infections were manifest [200].

Among the outcomes considered, sepsis is intended as severity of HAIs, and defined according to the definition of sepsis of the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference. Mortality within seven days from ICU admission was evaluated considering patients who stayed in ICU for at least seven days and those who died within two to seven days after ICU admission. Accordingly, patients who stayed in ICU for less than two days, those who were discharged prior to 7 days, and those who died with the first two days were excluded [203].

3.6 Statistical analysis

All the statistical analyses were performed using the SPSS software (version 26.0, SPSS, Chicago, IL, USA), and p-values < 0.05 were considered statistically significant. Firstly, to check the skewed distribution of continuous variables, the Kolmogorov- Smirnov test was applied. Thus, descriptive statistics were used to characterize the hospitals and patients' participants by reporting mean or median (interquartile range, IQR) for continuous variables, and frequency (percentage values) for

those categorical. Similarly, comparisons between variables were analyzed using the Mann Whitney U-test or the Kruskal-Wallis t test for continuous variables with skewed distribution, and the two-tailed Chi-squared test for categorical variables.

Moreover, linear regression models were performed, and results were reported as β and standard error (SE). Receiver operating characteristic (ROC) curve was used to evaluate the performance of a binary classifier as its discrimination threshold is varied. Indeed, ROC curve is a useful graphical tool to evaluate the ability of the logistic regression models to accurately discriminate patients according to their outcomes. To do this, the best cut-off value which maximized the Youden Index was calculated. Predictive performance was reported in terms of sensitivity, specificity and Area Under the Curve (AUC) with their 95% Confidence Interval (95% CI).

3.7 Cluster analysis

Clustering method constitutes an exploratory tool to reveal natural clusters within a dataset that would otherwise not be apparent. In general, this approach was conducted to group similar observations into a dataset, such that observations in the same cluster were similar to each other. For these reasons, the Two-Step Cluster analysis was applied to partition the original dataset into different clusters with high within-cluster similarity and between-cluster variability. This approach involves the construction of a Cluster Features (CF) Tree by positioning the first case at the root of the tree and adding each successive case to an existing node or forming a new node, in accordance with their similarity. Next, an agglomerative clustering algorithm is employed to categorize CF tree' nodes of clustered variables and to produce a range of clustering options. This algorithm is able to handle categorical and continuous variables, and to automatically determine the exact number of clusters by comparing the values of a model-choice criterion across different clustering solutions [206].

Indeed, the optimal number of clusters was determined automatically according to Schwarz's Bayesian Information Criterion (SBIC) or the Akaike Information Criterion (AIC), which allow model selection based on the likelihood function. In particular, the Two-Step Cluster analysis

employs the log-likelihood data vectors' distance measure to manage categorical and continuous variables, assuming that variables in the cluster model are independent. Variables included in the clustering algorithm were ranked according to their predictive importance values, whose values ranged from 0.1 (i.e., low predictive ability) to 1 (i.e., high predictive ability). Moreover, the cluster solution obtained was tested by excluding variables with predictive importance <0.2 [206].

3.8 Visual Analysis

Visual Analysis represent a useful approach of data aggregation, which is applied to observe progressive sequences of events that pass through different states and bring to distinct outcomes. This approach is inspired to the Outflow Graph model and to the visual encoding of Outflow Visualization and adapted using the Sankey diagrams. In general, the use of Outflow-like approach is suggested by the nature of the phenomena under investigation. In this case, a patient is considered in an initial "state" that during the hospitalization evolves into other "states" because new events are occurred.
Here, Sankey diagrams were used to immediately visualize the flow of patients during their ICU stay, without temporal information, also maintaining the total anonymity of the information. Sankey diagrams consist of nodes, whose height represents patients staying in a particular "state," instead, the height of each edge represents the number of patients evolving into other "states". One of the strengths of this method is the independence of the diagram dimension from the number of dataset records [207].

3.9 Learning Model generation

3.9.1 Machine Learning algorithm

The supervised SVM algorithm, which can be used for binary classification and regression problems, was applied. However, our dataset was not linearly separable, not allowing to satisfy all the constraints of SVM [208]. In the case of two or three exposure variables, the functions used to classify features are lines or planes, respectively. In the case of more than three exposure variables, as in the

model used in this study, the function used to classify features is referred to as a "hyperplane". Accordingly, the rationale behind the SVM model is to find an optimal hyperplane that clearly classifies the different classes. The separating hyperplane found by the SVM algorithm provides the largest margin between the two classes. The comparison with other classifiers (i.e., DT, K-NN, NB) revealed the SVM model as that reporting the best predictive performances. However, the dataset was not linearly separable even in a feature space and did not allow all the constraints of SVM to be satisfied. For this reason, a non-linear Kernel function was applied. Gaussian RBF, also called as Radial Basis Function Kernel, is a popular kernel function used in SVM models to map data which are not originally linearly separable into a higher dimensional feature space, where they are made linearly separable. Moreover, linear kernels are less time consuming than non-linear kernel, but have lower accuracy. To assess the predictive ability of the SVM model, accuracy (i.e., proportion of total records that are correctly predicted), and AUC which represents a measurement of accuracy, with values ranging from 0.5 for no prediction to 1.0 for perfect prediction) [40,41,209,210] were calculated. In addition, two evaluation metrics for classification problems were calculated, namely precision and recall.

Specifically, precision or Positive Predictive Value (PPV) is the proportion of true positive divided by the total number of true positives and false positives. In particular, precision estimates the correct predictions when the classifier predicts a positive condition.

Recall, often referred to as 'Sensitivity' is the proportion of true positive divided by the total number true positives and false negatives. Specifically, recall estimates the correct predictions of positive condition when the condition is positive. A perfect classifier should have precision and recall equal to 1. The analyses were performed using Python and Support Vector Classification (SVC) from Sklearn 0.22.1 [211]. Notably, the SVM model was trained on the training set composed by synthetic records, and then tested on the test set made of "real" records. All the analyses were performed using Python and the SciPy stack.

3.9.2 Support Vector Machine algorithm

Slack variables with penalty were also introduced to satisfy all the constraints in the minimization problem of SVM [208]. Indeed, by choosing very large slack variable values, a degenerate solution leading to the model overfitting could be found. Moreover, to penalize the assignment of too large slack variables, the following penalty was introduced in the classification objective:

$$\textbf{(1)}\ C \sum_{i=1}^{N} \varepsilon_i$$

where:

- $\boldsymbol{\varepsilon_i}$ indicated "slack variables", one for each datapoint i, to allow certain constraints to be violated;
- $\boldsymbol{C}$ indicated a tuning parameter that controlled the trade-off between the penalty of slack variables ε_i and the optimization of the margin.

Specifically, high values of C penalize slack variables leading to a hard margin, whereas low values of C lead to a soft margin, which is a bigger corridor which allows certain training points inside at the expense of misclassifying some of them. In particular, C parameter sets the confidence interval range of the learning model.

The RBF function expression on two sample, $x\ and\ x'$, was defined as:

$$\textbf{(2)}\ K(x, x') = exp\left(-\gamma \left|\left|x - x'\right|\right|^2\right)$$

where:

- $\left|\left|x - x'\right|\right|^2$ was the squared Euclidean distance between the two feature vectors;
- γ was a free parameter.

Specifically, the RBF function can be applied to a dataset through the choice of two parameters (i.e., C and γ), and the classifier performance of SVM depends on the choice of these two parameters. A Grid Search method was used to find the optimal parameters of the RBF for SVM. This method considered m values in C and n values in γ, according to the M×N combination of C and γ [212], by training different SVM using a K-fold cross validation.

3.9.3 Datasets of "Real" and Synthetic Records

The dataset of "real" records consisted of those patients having complete data and meeting the inclusion criteria. This dataset was used both for conventional statistical analyses and as test set for the SVM algorithm. Since a lot of records of the SPIN-UTI dataset had missing information, a dataset made of recovered and synthetic data was created to be used as training set, and to tune the learning algorithms. The presence of missing data is one of the most common problems encountered by when analyzing real-world data. As many statistical models and machine learning algorithms rely on complete datasets, it is key to handle the missing data appropriately. Moreover, ML algorithms requires large and balanced datasets to be trained. To replace the missing values, several imputation methods, such as the replacement of missing values with 0, mean, median or mode values and regression imputation, are used.

To do this, recovered data were imputed from incomplete records of the original dataset by replacing the missing values, using the K-NN imputation method described by Malarvizhi and Thanamani [213]. This method is useful for dealing with all kind of missing values, also allowing to recover part of the missing values for both quantitative and qualitative data [213].

Specifically, the K-NN algorithm considering the Euclidean distance in the feature space- for non-binary variables- and the Jaccard distance- for dichotomic variables- were applied. In particular, the Jaccard distance is defined as the complementary of size of the intersection divided by the size of the union of the sample sets. The K-NN method, is based on the assumption that a point value can be approximated by the values of the points that are closest to it, based on the other variables [214].

$$J(A,B) = |A \cap B| \,/\, |A \cup B|$$

$$0 \leq J(A,B) \leq 1$$

$$d_j(A,B) = 1 - J(A,B)$$

Figure 1 shows the algorithm of data imputation used to replace the missing value. At the end of the procedure, the algorithm provides the reconstructed data for the binary outcome (i.e., coded as 0 and 1). In addition, the process can be repeated on the entire dataset with the reconstructed data.

Moreover, synthetic data were generated to balance the two classes of patients according to their outcome, using the Synthetic Minority Over-sampling Technique (SMOTE). In order to balance dataset, several techniques are commonly used by replicating data or by generating synthetic data. In particular, the under-sampling technique consists of random elimination from the majority class to match the numbers with the minority class [215]. Although this technique allows to reduce rows' number in the dataset, its application is not recommended when the minority class is too small, providing a model with less data for training and more likely to generate errors.

Contrary to under-sampling approaches, SMOTE is a common oversampling method useful to resample the minority class data following those in the majority class by duplicating records. While the classic over- sampling technique duplicates minority data from the minority data population, SMOTE looks at the feature space for the minority class data points and considers its k-nearest neighbors to create new synthetic points and to increase the cardinality of the class itself [216].

In simple terms, SMOTE starts by selecting random data from the minority class population, then identifies the K-NN, and finally creates synthetic data from the random data and the randomly selected K-NN (**Figure 2**) [216]. Finally, to check the goodness of the training set, the distribution of outcomes and exposure variables were compared with those obtained from the test set.

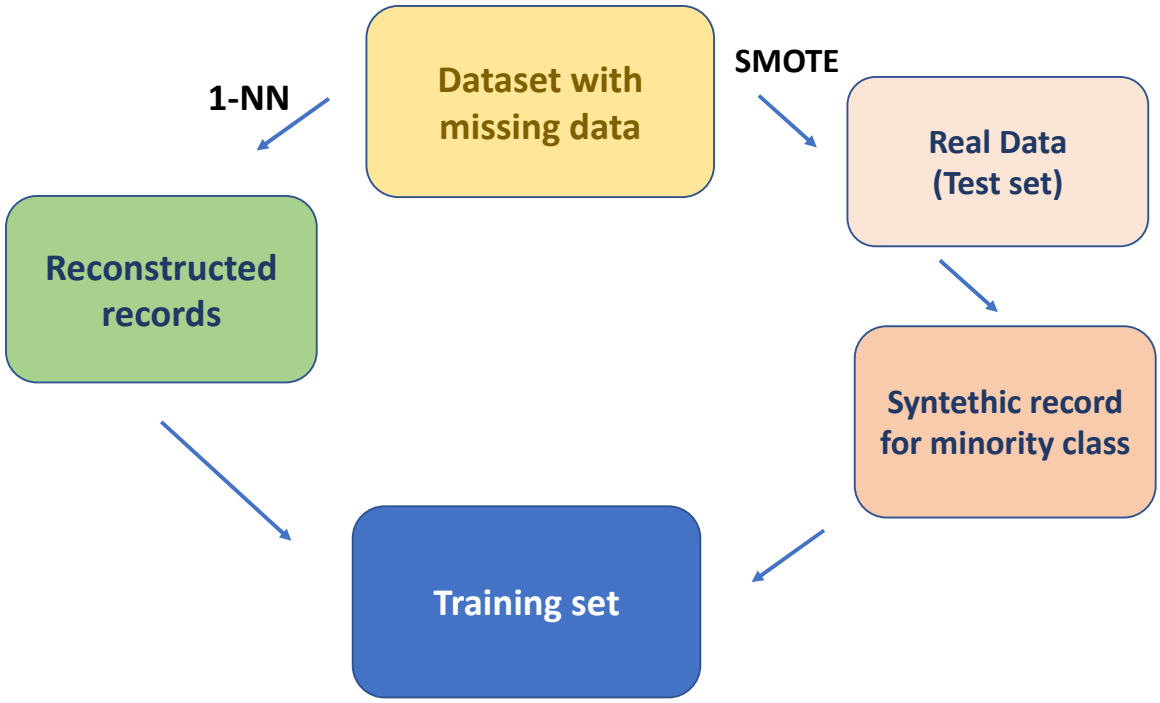
3.10 Tables and figures

Figure 1. Description of code used for data imputation

Algorithm 1 Multiple data imputation

D_S: the dataset with missing data
O: the outcome variable in D_S
kNN: the k-NN imputer with $k = 1$ and $metrics = (Euclidean, Jaccard)$
procedure DATAIMPUTATION(D_S, O)
    $m_{O_0} \leftarrow$ list of variables for class 0
    $m_{O_1} \leftarrow$ list of variables for class 1
    for each variable v_0 in m_{O_0} **do**
        $I_0 \leftarrow kNN(D_S[v_0])$
    end for each
    for each variable v_1 in m_{O_1} **do**
        $I_1 \leftarrow kNN(D_S[v_1])$
    end for each
    return (I_0, I_1)
end procedure

Figure 2. Flowchart of the data imputation techniques



4. Results

4.1 Analysis by gender of risk factors and outcomes of patients admitted to Intensive Care Unit

4.1.1 Study population

On a total of 20,060 SPIN-UTI patients, a sub-sample of 12,534 SPIN-UTI patients (median age 70 years, 60.6% males) who stayed in ICU for more than 48 hours were included. In this sub-sample, median SAPS II at admission was 40 (IQR 26.0), length of ICU stay was 7 days (IQR 11), and length of hospital stay was 17 days (IQR 21). In general, 4.3% of patients were traumatized and 5.7% were immunocompromised. Moreover, 70.5% of patients were from other wards or other hospitals and 57% had a medical type of ICU admission. With respect to medical cares, 16.6% and 34.0% of patients underwent non-surgical treatment for acute coronary disease or surgical intervention, respectively. With respect to the use of antibiotic treatments, 63.9%, 83.4% and 56.8% were on Antibiotic therapy in 48 hours before or during ICU stay, respectively. In particular, 73.9% of patients required a urinary catheter, 56.2% intubation and 48.4% central venous catheter. Finally, the percentage of patients with at least one HAI was 16.0%, with 21.2% and 25.7% of patients died at ICU or hospital discharge, respectively.

4.1.2 Characteristics of patients

Overall, male (n=7698) and female patients (n=4836) were compared for their characteristics at ICU admission. Specifically, female patients were older, more likely to come from other wards or healthcare facilities and to undergo a surgical intervention than those males (p-values<0.05). By contrast, female patients comprised less patients with trauma, who were less likely to be intubated and to undergo non-surgical treatment for acute coronary disease (p-values<0.05).

Interestingly, female patients reported a higher proportion of deaths (22.8%) at ICU discharge than those males (20.1%; p<0.001). No differences were evident for SAPS II, type of ICU admission,

impaired immunity, antibiotic therapy in 48 hours before or after ICU admission, presence of urinary catheter at ICU admission, presence of central venous catheter at ICU admission, status at hospital discharge, antibiotic therapy during ICU stay or at ICU admission, and length of hospital and ICU stay.

4.1.3 Survival analysis

After adjusting for covariates, female patients reported 13% higher odd of dying (OR=1.13; 95%CI=1.00-1.28; p=0.046) than males. Accordingly, females also reported lower survival in ICU than males (median= 32.0 days vs. median= 34.0 days; p<0.001). However, after applying a cox regression model no difference was evident in survival between male and female patients.

4.2. Application of Support Vector Machine to predict Healthcare-Associated Infections

4.2.1 Study population

On a total of 20,060 SPIN-UTI patients, a sub-sample of 7827 patients (median age 69 years, 60.6% males) enrolled from 2006 to 2019 was included. The remaining 12,233 patients (61%) were excluded because of missing data. In this sub-sample, 73.9% of patients came from other wards or other hospitals and 52.4% had a surgical purpose for ICU admission. In general, median SAPS II at admission was 40 (IQR 28) and length of ICU stay was 5 days (IQR 10). With respect to medical cares, 10.2% and 40.9% of patients underwent non-surgical treatment for acute coronary disease or surgical procedure, respectively. Moreover, 3.4% of patients were traumatized, 8.6% reported impaired immunity, and 59% were on antibiotic therapy. In particular, 77.5% of patients required a urinary catheter, 59.8% intubation and 41% central venous catheter. Finally, the percentage of patients with ICU- acquired sepsis was 6.1%, and mortality in ICU was 23.2%.

4.2.2 Characteristics of patients

Table 1 reports the comparison between infected (n=1225, 15.7%) and non-infected patients (n=6602, 84.3%) for their characteristics at ICU admission. Infected patients were more likely to come from the community and to report a medical type of ICU admission than those non-infected. Infected patients included more patients with trauma and who were more likely immunocompromised. For this reason, SAPS II was higher among infected patients than those non-infected. With respect to the presence of invasive devices at ICU admission, infected patients were more likely to be intubated and less likely to be catheterized compared to non-infected patients. As expected, infected patients reported a higher length of ICU stay (20 vs 4 days; $p<0.001$) than those non-infected. Accordingly, also mortality was higher in infected patients (35.1%) than in those non-infected (21.0%; $p<0.001$). No differences were evident for sex, age, non- surgical treatment for acute

coronary disease, antibiotic therapy in 48 h before ICU admission, and presence of central venous catheter at ICU admission.

4.2.3 ROC curve analysis using conventional statistical approach

The performance of SAPS II at ICU admission in predicting HAIs was evaluated using a conventional statistical approach. **Figure 1** reported the ROC curve with an AUC of 0.612 [95% confidence interval (CI) 0.60 0.63; $p<0.001$]. Although these results were significant, the accuracy of SAPS II for the prediction of HAIs risk was 56%.

4.2.4 ROC curve analysis using the Support Vector Machine model

Next, additional exposure variables were used to develop a ML algorithm for the prediction of HAIs acquired in ICUs. Specifically, the ML algorithm combined SAPS II with all variables collected at ICU admission but not included in the SAPS II computation.

Since the dataset was strongly imbalanced especially in terms of infected and not-infected classes, SMOTE algorithm was applied on the complete records to avoid a low performance in recall score. This algorithm provided new synthetic records for the infected class, also discarding the duplicated data generated by the algorithm. Thus, the test set included only included 39% of patients (n=7827), with complete assessment of variables considered in this study (**Figure 2**), while the training set included recovered (n=7758), and synthetic records (n=2544).

Specifically, the test set was composed of real data for patients with complete assessment of the following variables at ICU admission: sex (dichotomous); patient origin (categorical: other ward/healthcare facility, community); non-surgical treatment for acute coronary disease (dichotomous); surgical intervention (dichotomous); SAPS II at admission (continuous); presence of invasive devices at ICU admission (three dichotomous variables for urinary catheter, intubation and central venous catheter, respectively); trauma (dichotomous); impaired immunity (dichotomous); and antibiotic therapy in 48 h preceding ICU admission (dichotomous).

To evaluate potential discrepancies between the training and the test set, the distribution of each variable was compared between them. The distribution of infected and non-infected patients between the training and test sets is summarized in **Table 2**. As reported in **Figure 3,** SAPS II and age followed the same distribution in the training and test sets. Likewise, **Figures 4 and 5** show that the distributions of categorical variables were similar between the training and test sets.

To improve the accuracy for the prediction of HAI risk, SAPS II along with other characteristics at ICU admission (i.e., sex, patient origin, non- surgical treatment for acute coronary disease, surgical intervention, presence of intubation, presence of urinary catheter, presence of central vascular catheter, trauma, impaired immunity, antibiotic therapy in 48 h preceding ICU admission) was used in the SVM algorithm. **Figure 6** reported the ROC curve for the SVM prediction model of the test set, whose accuracy was of 88%. Particularly, precision and recall values were 0.95 and 0.91 for non-infected patients and 0.60 and 0.73 for those who were diagnosed with at least one HAI during their ICU stay, respectively. Predictive ability was assessed using the ROC curve, which had an AUC of 0.90 (95% CI 0.88-0.91; $p<0.001$). Next, this predictivity was compared with that obtained using the same SVM model without accounting for SAPS II in the test set. **Figure 7** reported the ROC curve for the SVM prediction model of the test set, having an accuracy of 78%. In line, precision and recall values were 0.87 and 0.87 for non-infected patients and 0.31 and 0.32 for those infected, respectively. Moreover, the AUC value provided by the ROC curve was 0.66 (95% CI 0.65- 0.68; $p<0.001$), indicating lower predictive ability.

4.2.5 Table and figures

Table 1. Characteristics of patients according to their infectious status

Characteristics	Patients (*n*=7827)	Infected patients (*n*=1225)	Non- infected patients (n=6602)	*p*-value
Age, years	69.0 (21.0)	69.0 (21.0)	69.0 (21.0)	0.064
Sex (% men)	60.6%	62.8%	60.1%	0.084
Patient's origin				
Other ward/healthcare facility	73.9%	67.7%	75.1%	**<0.001**
Community	26.1%	32.3%	24.9%	
SAPS II score at admission	40.0 (28.0)	47.0 (27.0)	38.0 (27.0)	**<0.001**
Type of ICU admission				
Medical	47.6%	53.6%	46.5%	**<0.001**
Surgical	52.4%	46.4%	53.5%	
Trauma	3.4%	5.0%	3.2%	**0.001**
Impaired immunity	8.6%	10.4%	8.2%	**0.015**
Non-surgical treatment for acute coronary disease	10.2%	8.9%	10.4%	0.109
Surgical intervention	40.9%	36.7%	41.7%	**<0.001**
Antibiotic therapy in 48 hours before ICU admission	59%	59.8%	58.9%	0.579
Presence of urinary catheter at ICU admission	77.5%	74.4%	78.0%	**0.006**
Presence of intubation at ICU admission	59.8%	63.8%	59.1%	**0.002**
Presence of central venous catheter at ICU admission	41%	39.7%	41.3%	0.295
ICU-acquired sepsis (%yes)	6.1%	37.6%	-	-
Outcome (%death)	23.2%	35.1%	21.0%	**<0.001**
Length of ICU stay, days	5.0 (10.0)	20.0 (20.0)	4.0 (6.0)	**<0.001**

*Results are reported as median (interquartile range) for continuous variables, or percentage for categorical variables. Statistical analyses were performed using the Mann-Whitney or the Chi-squared test.

Table 2. Composition of training and test datasets

Outcome	Training Set	Test set
Class 0 (Non-infected)	6702 recovered	6602
Class 1 (Infected)	3600 (of which 1056 recovered and 2544 synthetic)	1225
Total	10302	7827

Figure 1. ROC curve of the SAPS II for predicting HAIs. The figure shows the ability of SAPS II to identify patients who developed at least one HAI during their stay in an intensive care unit. The blue curve represents the ability of SAPS II to distinguish between patients who developed at least one HAI and those who did not [AUC= 0.612, 95% confidence interval 0.60-0.63; p<0.001]. The black dotted line is the reference for no predictive ability (AUC= 0.500)

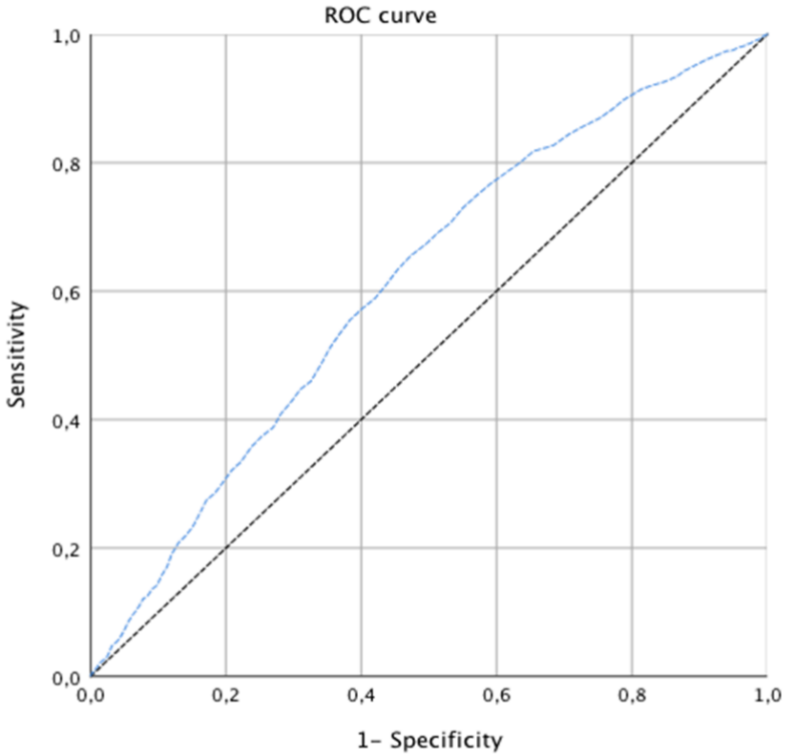


Figure 2. Matrix of missing values

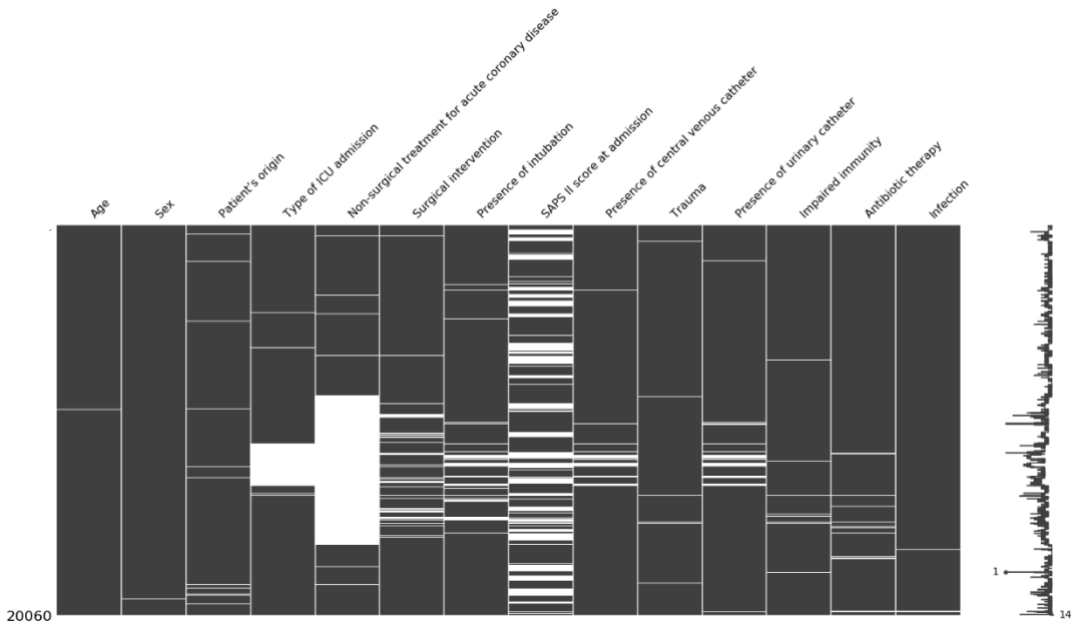


Figure 3. Comparison of Age (A) and SAPS II score (B) distributions between training and test datasets

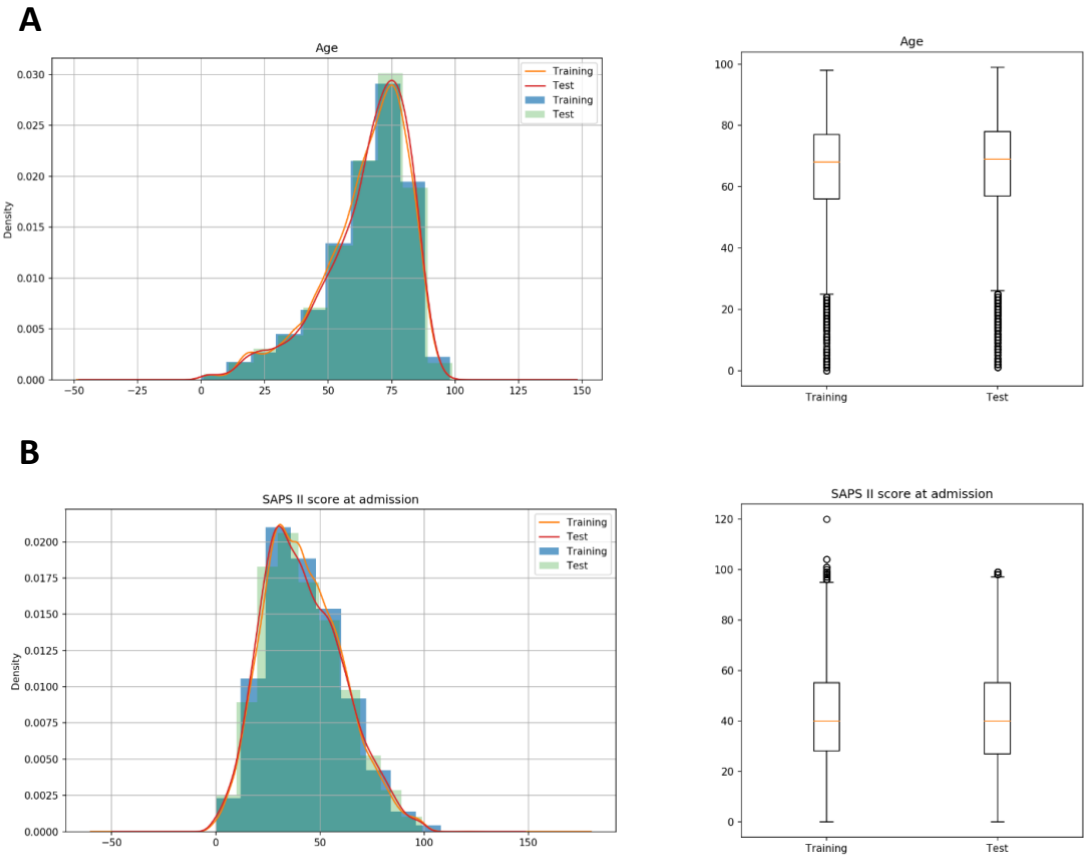


Figure 4. Comparison of dichotomous variables between training and test sets

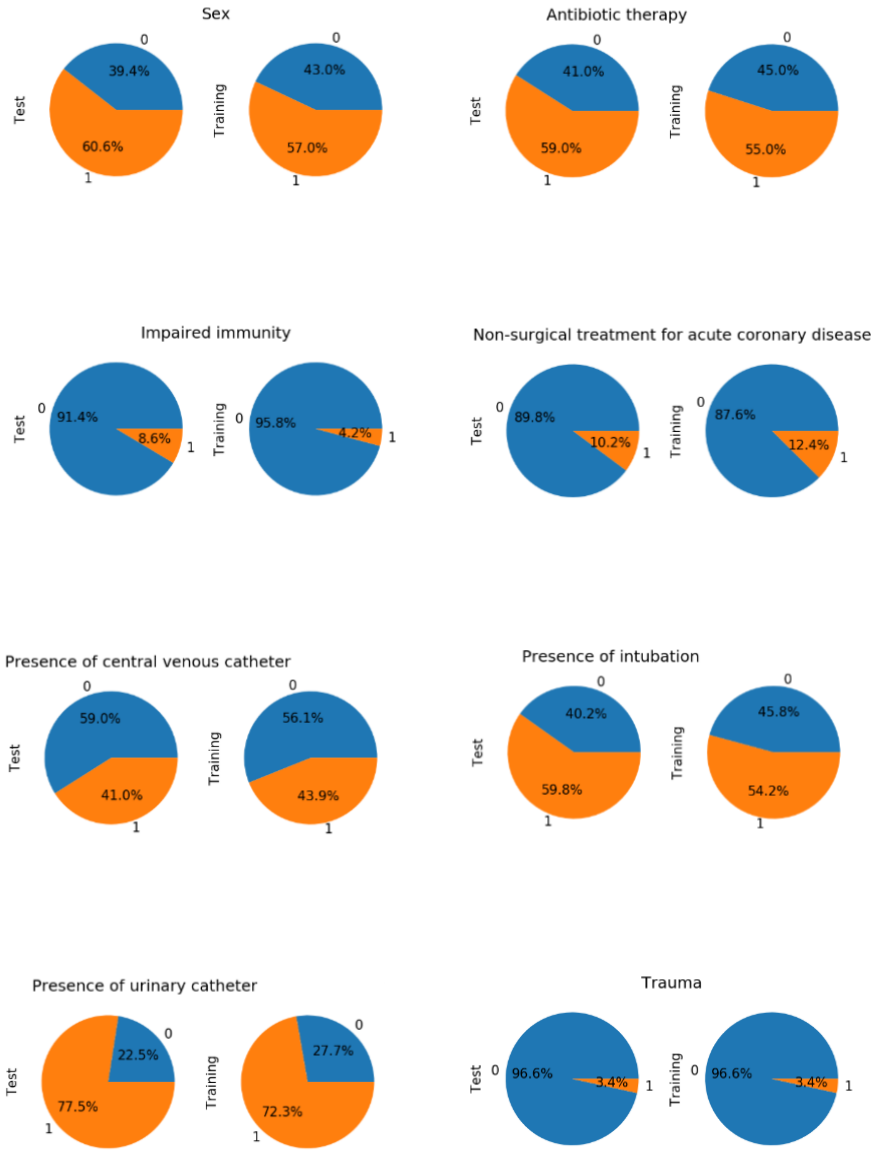


Figure 5. Comparison of categorical variables between training and test datasets

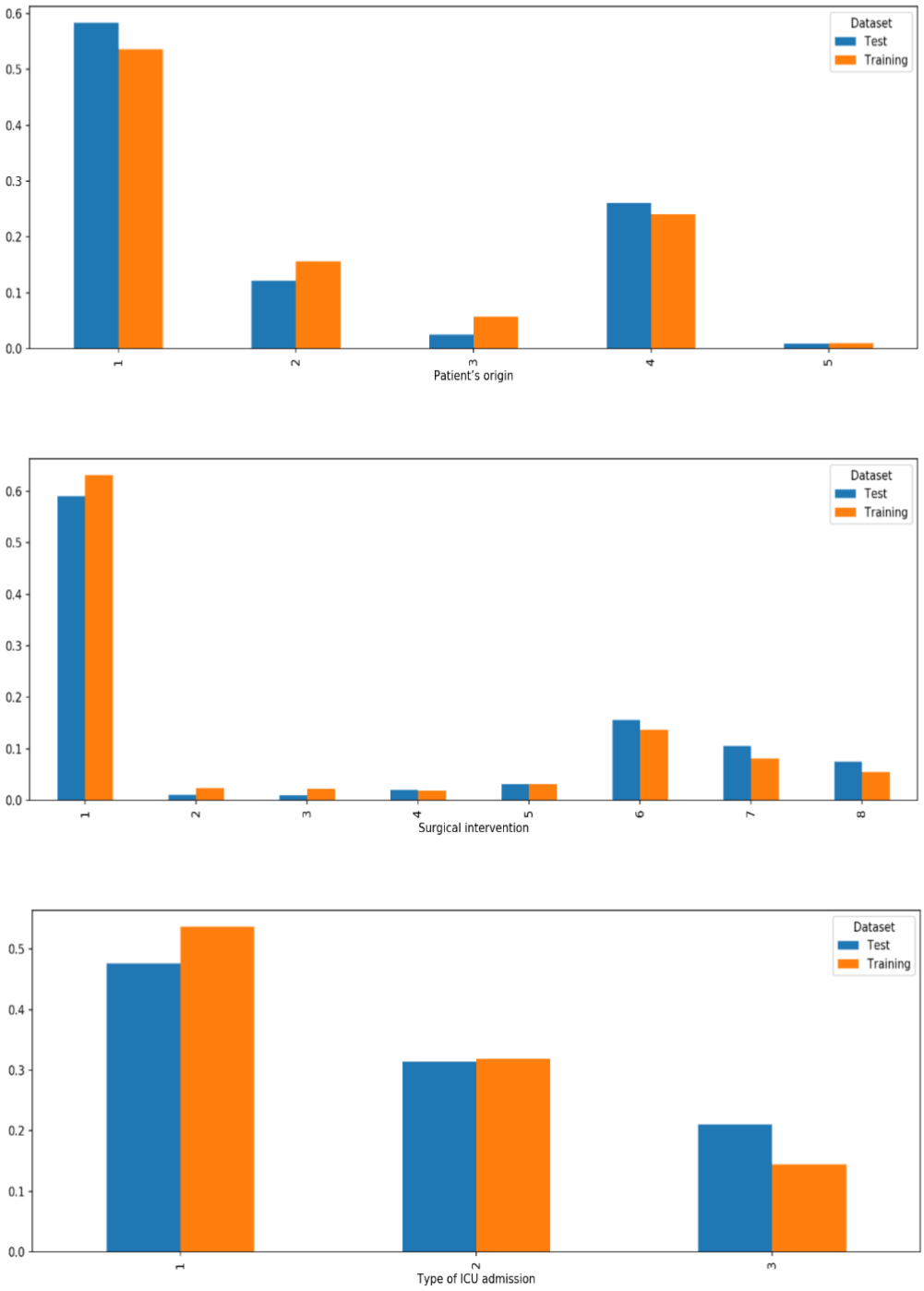


Figure 6. ROC curve of the SVM algorithm predicting HAISs. The model is based on the SVM algorithm, which combines the SAPS II with additional features at intensive care unit admission. The blue curve represents the ability of the SVM algorithm to distinguish between patients who developed at least one HAI and those who did not [AUC= 0.90, 95% confidence interval 0.88 and 0.91; p<0.001]. The black dotted line is the reference for no predictive ability (AUC =0.500)

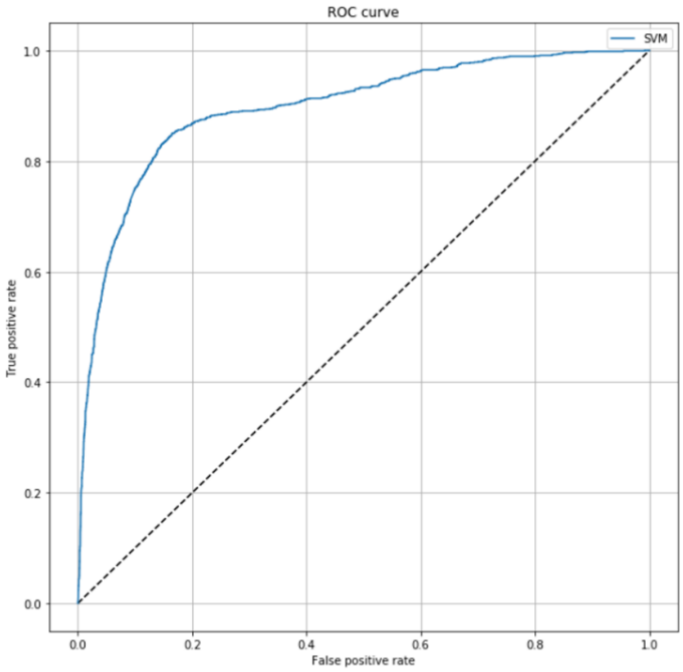
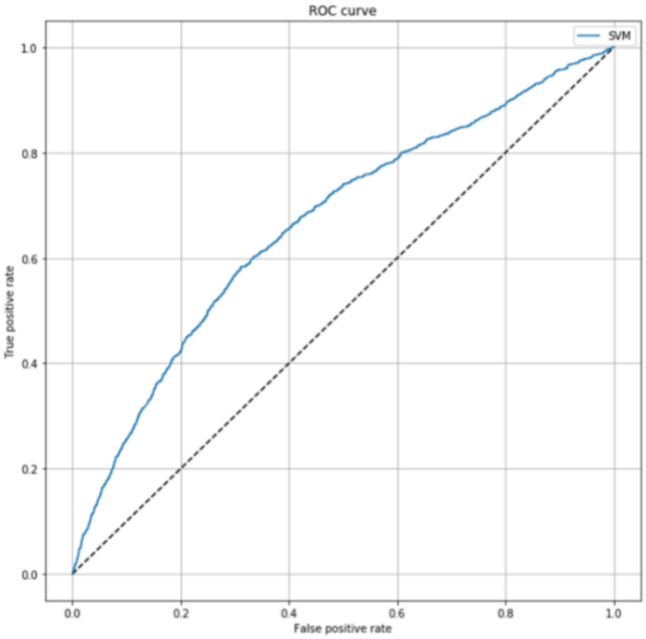


Figure 7. ROC curve of a SVM algorithm predicting HAIs by excluding the SAPS II. The model is based on the SVM algorithm which combines patients' characteristics collected at intensive care unit admission. The blue curve represents the ability of the SVM algorithm to distinguish between patients who developed at least one HAI and those who did not [AUC= 0.66, 95% confidence interval 0.65 and 0.68; p<0.001]. The black dotted line is the reference for no predictive ability (AUC= 0.500)



These tables and figures were adapted from Barchitta et al., Journal of Hospital Infection 2021

4.3 Application of cluster analysis to identify patients at risk of catheter-associated urinary tract infections

4.3.1 Study population

On a total of 13,512 patients, 9656 SPIN-UTI participants admitted - from 2008 to 2017 - to 76 ICUs of 55 hospitals were included. The remaining 3856 participants (28.5%) were excluded because of missing data for at least one variable imputed in the cluster analysis. In this subsample, 264 patients acquired at least one CAUTI, resulting in a cumulative incidence of 2.7 CAUTIs per 100 patients and an incidence density of 2.7 CAUTIs per 1000 patient-days. Moreover, female sex ($p=0.033$), higher SAPS II score ($p<0.001$), medical type of ICU admission ($p<0.001$) and being a traumatized patient ($p=0.011$) were those characteristics at ICU admission positively associated with CAUTI acquisition.

4.3.2 Characteristics of clusters

Performing a Two-step cluster analysis three different clusters of patients were distinguished according to their characteristics at ICU admission. The Two-step clustering method was performed to identify different clusters of patients based on age, sex, SAPS II score at admission, patient origin, type of admission, trauma, and administration of antibiotics in 48 h before or after ICU admission. Notably, administration of antibiotics in 48 h before or after ICU admission, type of ICU admission, and patient origin were the top three variables with higher predictive importance, followed by SAPS II at admission, trauma, age, and sex.

Table 1 reported the characteristics of patients with relative within-cluster homogeneity and between-cluster variability. Patients in Cluster 1 (n=2143) were mostly with a medical type of ICU admission and came from the community. Cluster 1 was also characterized by an intermediate percentage of patients who received antibiotics in 48 h before or after ICU admission, lower median age, higher proportion of trauma patients and higher SAPS II score. Cluster 2 (n=5854) comprised patients who were more likely to come from other wards/hospitals, and to report administration of antibiotics 48 h

before or after ICU admission. Cluster 2 included older patients with an intermediate SAPS II score, and approximately 50% reported a surgical type of ICU admission. Although patients in Cluster 3 (n=1659) reported a lower percentage of patients with administration of antibiotics 48 h before or after ICU admission, and lower SAPS II score, they were similar to those in Cluster 2 in terms of patient origin, type of admission and age.

4.3.3. Duration of urinary catheterization

In general, the duration of urinary catheterization was 85,8 days, which in turns corresponded to a urinary catheter utilization rate of 84.6 urinary-catheter-days per 100 patient-days. Although length of ICU stay was similar across clusters, **Figure 1** visually revealed differences in terms of urinary catherization and its duration. Indeed, patients in Clusters 1 or 2 were less likely to be catheterized (82.9% and 84.1%, respectively) than those in Cluster 3 (85.6%; $p<0.001$).

Nevertheless, patients in Clusters 1 or 2 had a longer duration of urinary catheterization (median 7 days, IQR 12 days for Cluster 1; median 7 days, IQR 11 days for Cluster 2) compared with patients in Cluster 3 (median 6 days, IQR 8 days; $p<0.001$).

4.3.4 Incidence of catheter-associated urinary tract infections and sepsis

In general, patients with urinary catheterization showed a higher incidence of UTIs than patients who were without urinary catheterization (3.0 per 100 patients vs 1.2 per 100 patients; $p=0.004$). The rate of CAUTIs was 3.2 per 1000 catheter-days, with an incidence of 0.4 per 100 patients in those catheterized for <5 days, 0.8 per 100 patients in those catheterized for >=5 days and <=10 days, and 7.2 per 100 patients in those catheterized for >10 days ($p<0.001$). Interestingly, patients belonging to Cluster 1 had a higher incidence of CAUTIs (3.5 per 100 patients) than those in Clusters 2 or 3 (2.5 per 100 patients in both clusters; $p=0.033$).

Finally, 37.0% of patients with CAUTIs developed sepsis, but no difference was evident in the incidence of sepsis across three clusters (p=0.238). However, the percentage of sepsis among patients with CAUTIs increased with increasing duration of catheterization (p=0.010).

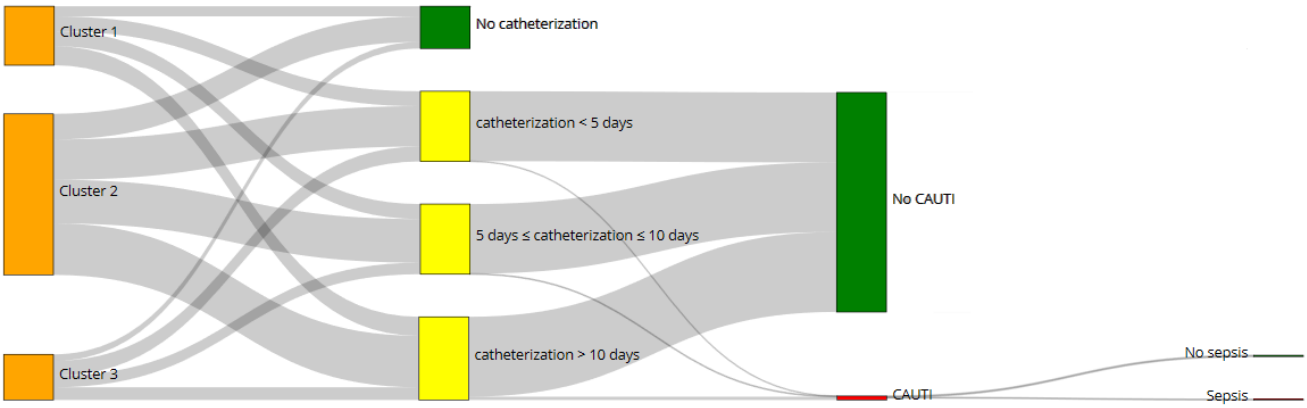
4.3.5 Tables and Figures

Table 1. Characteristics of clusters of patients at ICU admission

Characteristics	Cluster 1 (n=2143)	Cluster 2 (n=5854)	Cluster 3 (n=1659)	p-value
Age, years	69.0 (24.0)	70.0 (20.0)	70.0 (20.0)	0.028
Sex (% men)	62.8%	61.0%	60.5%	0.263
Patient's origin				
Other ward/healthcare facility	41.5%	87.3%	86.8%	**<0.001**
Community	58.5%	12.7%	13.2%	
SAPS II score at admission	40.0 (27.0)	38.0 (26.0)	37.0 (23.0)	**<0.001**
Type of ICU admission				
Medical	63.2%	47.8%	52.8%	**<0.001**
Surgical	36.8%	52.2%	47.2%	
Trauma	5.7%	4.4%	4.4%	**0.043**
Impaired immunity	5.8%	7.4%	3.6%	**<0.001**
Antibiotics within 48 hours of admission	67.9%	87.0%	32.9%	**<0.001**
Length of ICU stay, days	5.0 (10.0)	5.0 (9.0)	4.0 (8.0)	0.134
Presence of urinary catheter during ICU stay	82.9%	84.1%	85.6%	**<0.001**
Duration of urinary catheterization, days	7.0 (12.0)	7.0 (11.0)	6.0 (8.0)	**<0.001**

*Results are reported as median (interquartile range) for continuous variables, or percentage for categorical variables. Statistical analyses were performed using the Kruskal-Wallis or the Chi-squared test.

Figure 1. Sankey diagram describes the flow of patients from their admission to ICU, their urinary catheter utilization and incidence of catheter urinary tract infections and sepsis



These tables and figures were adapted from Barchitta et al., Journal of Hospital Infection 2021

4.4 Application of cluster analysis to identify patients at risk of pneumonia

4.4.1 Study population

Overall, 9656 SPIN-UTI patients staying in ICU for more than 2 days- enrolled in 92 ICUs of 62 hospitals- were included. Specifically, these patients *had no* missing values in information related to their characteristics at ICU admission (e.g., age, sex, SAPS II at admission, origin, admission type), dates of intubation, and their outcomes (e.g., pneumonia, sepsis, death).

4.4.2 Characteristics of clusters

Using a Two-step cluster analysis, three different clusters of patients were distinguished according to their characteristics at ICU admission. Specifically, the following variables were standardized and imputed in the cluster analysis: age, SAPS II score at admission, patient origin and administration of antibiotics within 48 hours of admission. Of note, age, SAPS II score at admission, patient origin and administration of antibiotics within 48 hours of admission were the variables with higher predictive importance. **Table 1** reported the characteristics of patients with relative within-cluster homogeneity and between-cluster variability. In general, from cluster 1 to cluster 3, SAPS II and percentage of patients admitted for unscheduled surgery decreased (p-trends <0.001), while percentage of patient who came from long-term care facility increased (p-trend <0.001). Accordingly, patients in Cluster 1 (n=2143) were younger and exhibited higher SAPS II score than those in cluster 2 or 3. In addition, they were more likely to be traumatized, to come from the community, and to report a medical type of admission in ICU. By contrast, patients in cluster 2 (n=5854) were older and exhibited an intermediate SAPS II score. In particular, cluster 2 comprised a high percentage of patients who came from other ward or other hospital, and approximately half of them reported a medical type of admission in ICU. Compared to clusters 1 and 3, patients in cluster 2 were more likely to report impaired immunity and administration of antibiotics within 48 hours of admission. Finally, patients

in cluster 3 (n=1659) had an age similar to those belonged to cluster 2, even they reported lower SAPS II. Similar to cluster 2, cluster 3 comprised a high percentage of patients who came from other ward/hospital, with approximately half of them having a medical admission. However, patients in cluster 3 showed lower percentages of patients with impaired immunity and administration of antibiotics within 48 hours of admission than those in Clusters 1 and 2.

4.4.3 Outflow of patients from Intensive Care Unit admission to diagnosis of pneumonia

Figure 1 reported a Sankey diagram illustrating the course of patients from their admission to ICU, suggesting how belonging to a cluster and duration of intubation could contribute to the diagnosis of pneumonia. Patients belonging to cluster 1 and 2 had higher length of intubation in days (median= 3, IQR= 9 and median= 3, IQR= 8, respectively) and in days/100 days of ICU stay (median= 66.7, IQR= 102.7 and median=70.0, IQR=75.0, respectively) than those in cluster 3 (median= 2 days, IQR= 7 days, p- trend<0.001 and median= 60.1 days/100 days of ICU stay, IQR= 100.0 days/100 days of ICU stay, p-trend=0.001, respectively). In line, **Figure 1** showed that a higher percentage of patients in cluster 1 or cluster 2 were intubated for more than 66 days/100 days of ICU stay, compared to cluster 3 (p-trend<0.001). Moreover, patients intubated for more than 66 days/100 days of ICU stay also reported a higher incidence of pneumonia (11.4%) than those who were intubated for 34-66 or 0-33 days /100 days of ICU stay (5.6% and 2.9%, respectively; p-trend<0.001).

4.4.4 Association of pneumonia with the risk of sepsis and death

Thus, **Figure 2** reported a Sankey diagram illustrating how patients with Acinetobacter baumannii or Klebsiella pneumoniae-associated pneumonia were more likely to exhibit sepsis than those infected by Pseudomonas aeruginosa. Moreover, patients with Acinetobacter baumannii or Klebsiella pneumoniae-associated pneumonia (38.9% and 38.8%, respectively) reported a higher incidence of sepsis than those infected by Pseudomonas aeruginosa (29.1%; p=0.025). In turn, sepsis was

associated with higher mortality (45.6%) than infection (32.2%; $p<0.001$). Accordingly, mortality was higher in patients with Acinetobacter baumannii and Klebsiella pneumoniae-associated pneumonia (20.6% and 29.1%, respectively) than in those with Pseudomonas aeruginosa-associated pneumonia (13.4%; $p<0.001$).

4.4.5 Table and Figures

Table 1. Characteristics of clusters of patients at ICU admission

Characteristics	Cluster 1 (n=2143)	Cluster 2 (n=5854)	Cluster 3 (n=1659)	p-value
Age, years	69.0 (24.0)	70.0 (20.0)	70.0 (20.0)	**0.028**
Sex (% men)	62.8%	61.0%	60.5%	0.263
Origin				
Other ward of this/other hospital	39.6%	82.0%	82.2%	**<0.001**
Other ICU	1.1%	3.7%	2.7%	
Community (home)	58.5%	12.7%	13.2%	
Long-term care facility	0.7%	1.5%	1.9%	
SAPS II score at admission	40.0 (27.0)	38.0 (26.0)	37.0 (23.0)	**<0.001**
Type of ICU admission				
Medical	63.2%	47.8%	52.8%	**<0.001**
Scheduled surgery	18.8%	35.6%	36.6%	
Unscheduled surgery	18.0%	16.7%	10.6%	
Trauma	5.7%	4.4%	4.4%	**0.043**
Impaired immunity	5.8%	7.4%	3.6%	**<0.001**
Administration of antibiotics within 48 hours of admission	67.9%	87.0%	32.9%	**<0.001**
Length of stay in ICU, days	5.0 (10.0)	5.0 (9.0)	4.0 (8.0)	0.134

Results are reported as median (interquartile range) for continuous variables, or percentage for bivariate or categorical variables. Statistical analyses were performed using the Kruskal-Wallis or the Chi-squared test.

Figure 1. Sankey diagram describes the flow of patients from their admission to ICU, suggesting how their belonging to a cluster duration of intubation could contribute to the diagnosis of pneumonia

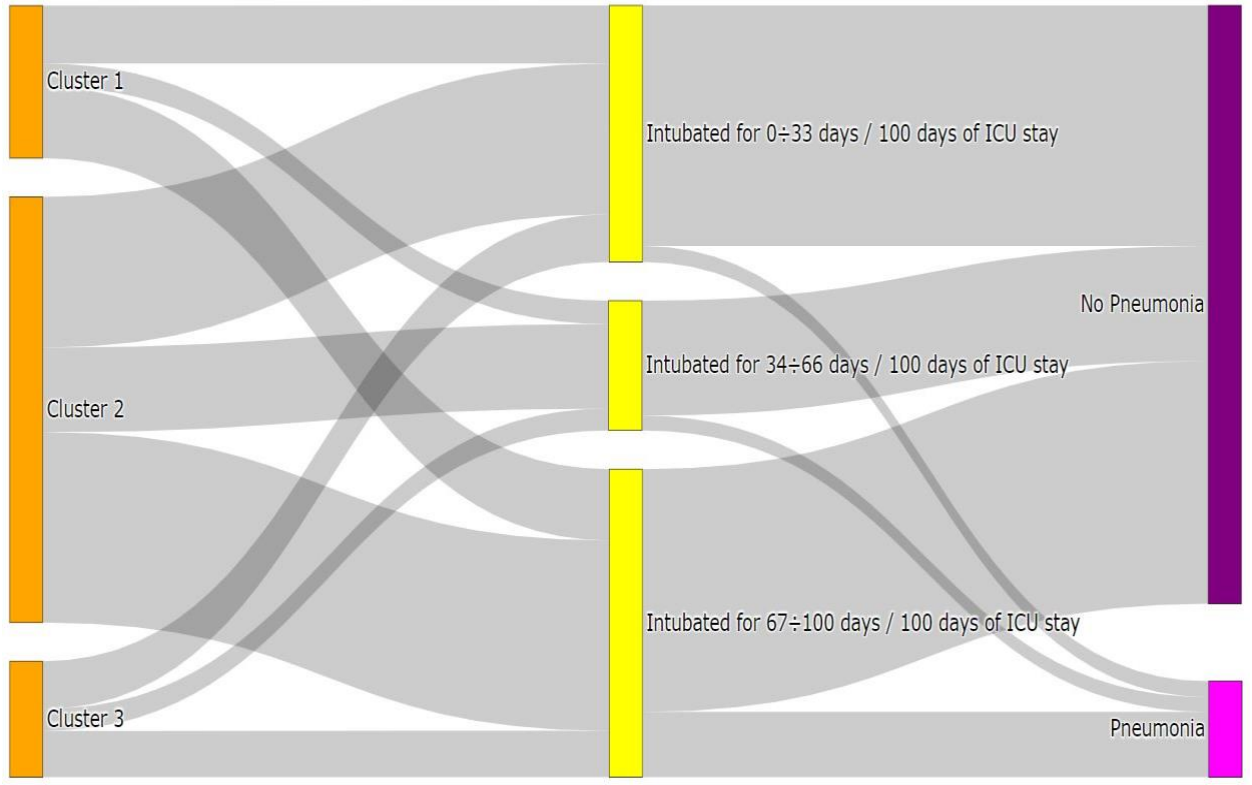
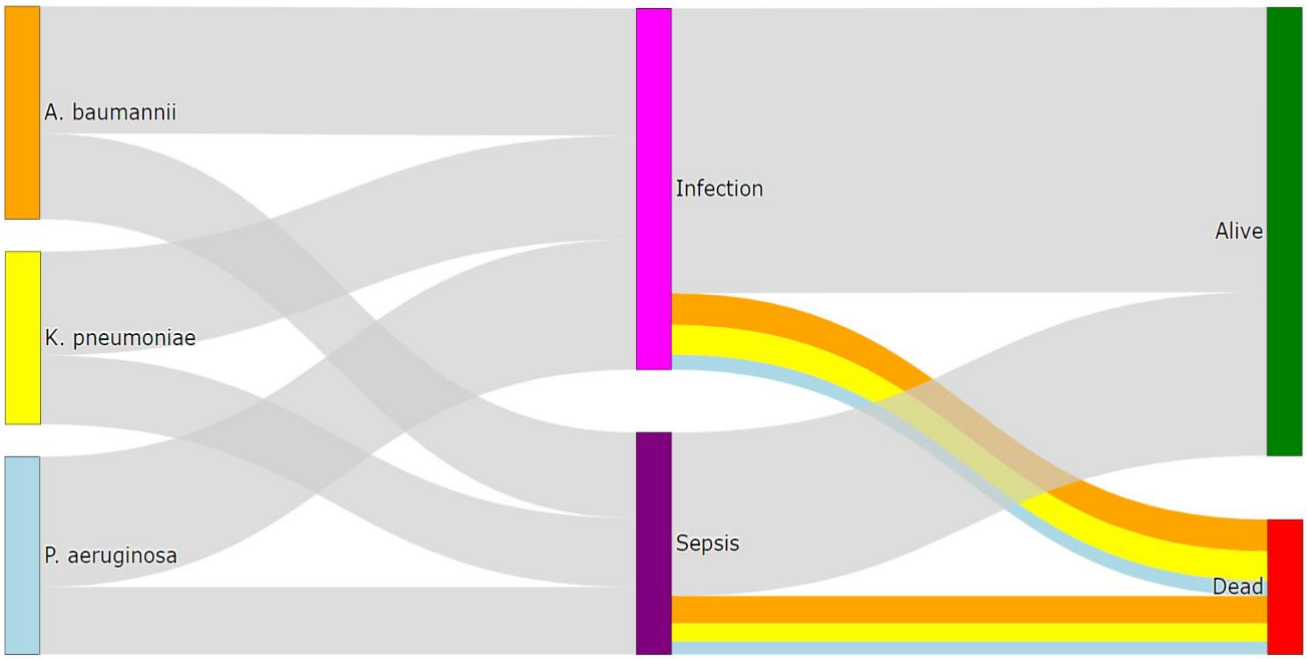


Figure 2. Sankey diagram describes the flow of patients with Acinetobacter baumannii-, Klebsiella pneumoniae- and Pseudomonas aeruginosa- associated pneumonia and their associated outcomes



These tables and figures were adapted from Favara et al., IEEE World Congress on Services, 2019

4.5 Analysis of risk factors and outcomes in patients with sepsis

4.5.1 Study population

On a total of 20,060, a sub-sample of 16,278 SPIN-UTI patients who stayed in ICU for more than 48 hours, enrolled from 2008 to 2019, were included. In this sample, 2393 (14.7%) reported at least one HAI of which 954 (5.9% of the total) developed at least one episode of sepsis. The analysis of trends in the different SPIN-UTI editions showed a significant decreasing trend in the proportion of sepsis across all UTIs participants (from 51.4% in 2008 to 35.5% in 2018; $p<0.001$). However, there was no significant trend considering the incidence rate of sepsis for 1000 patient-days.

4.5.2 Characteristics of patients

Patients who developed sepsis were younger, more likely to be traumatized and immunodeficient and to report a higher SAPS II than those who did not ($p<0.001$). In addition, they are more likely to come from other wards or healthcare facilities and to be exposed to invasive devices, and to receive antibiotics at ICU admission ($p<0.001$). By contrast, patients who developed sepsis are less likely to undergo non-surgical treatment for acute coronary disease. Although no difference was evident for the type of ICU admission, patients who developed sepsis were more likely to undergo a surgical intervention (38.8% vs 32.7%). Interestingly, patients with sepsis showed higher length of ICU stay and proportion of deaths than those without sepsis ($p<0.001$).

4.5.3 ROC curve analysis using conventional statistical approach

Next, the performance of those variables tested as significant in predicting sepsis was evaluated using a logistic regression model. In particular, the model included the following predictors: age, SAPS II, trauma, immunodeficiency, non-surgical treatment, surgical intervention, antibiotic treatment, and presence of invasive devices at the time of hospitalization. The ROC curve with an AUC of 0.64

[95% CI= 0.611 – 0.659; p<0.001], good sensitivity values (81.6%) and discrete specificity values (38.1%).

4.6. Application of Support Vector Machine to predict seven-day mortality

4.6.1. Study population

Overall, 3782 SPIN-UTI patients without missing data (60.2% males), surveyed from 2006 to 2019, were included. In this sample, the median age was 70.0 years (IQR = 20) and median SAPS II at admission was 49 (IQR = 27). Specifically, 70.9% of them originated from other wards or other hospitals and 56.9%, had a medical type of ICU admission. Moreover, 4.7% and 11.4% of patients were with trauma and/or immunocompromised, respectively. Patients who underwent antibiotic treatment, surgical procedure, or non-surgical treatment for acute coronary disease were 62.6%, 34.8%, and 9.0%, respectively. With regards to invasive devices, the presence of urinary catheter, intubation and central venous catheter was reported in 77.0%, 62.4%, and 40.5% patients, respectively.

In **Table 1** characteristics of patients who died (n = 875; 23.1%) within seven days from ICU admission with those who were alive (n = 2907; 76.9%) were compared. Specifically, patients who died were older, more likely male, and with a higher SAPS II than those who were alive. Moreover, they were also more likely to come from other ward or other healthcare facility and to report a medical type of ICU admission than those who did not die. The first group also consisted more of patients who reported impaired immunity and less traumatic events. Instead, no differences were evident for surgical intervention, non-surgical treatment for acute coronary disease, antibiotic therapy on admission and presence of invasive devices at ICU admission.

4.6.2 ROC curve analysis using conventional statistical approach

The dataset of "real" records consisted of a total of 3782 patients with a complete assessment of the following information: sex, patient's origin, type of ICU admission, non-surgical treatment for acute coronary disease, surgical intervention, SAPS II, presence of invasive devices at ICU admission, trauma, impaired immunity, antibiotic therapy in 48 h before or after ICU admission and onset of

HAI. The matrix of missing values and the selection of "real" records are showed in **Figures 1 and 2.** Performing a logistic regression model on the dataset of "real" records, SAPS II was considered as the independent variable and 7-day mortality as the dependent. For this reason, a logistic regression model was applied to evaluate the association of SAPS II (continuous) with death. Next, a logistic regression model was applied, also including sex (dichotomous), patient's origin (categorical: Other ward/healthcare facility, community), type of ICU admission (categorical: Medical, surgical), non-surgical treatment for acute coronary disease (dichotomous), surgical intervention (dichotomous), presence of invasive devices at ICU admission (three dichotomous variables for urinary catheter, intubation and central venous catheter, respectively), trauma (dichotomous), impaired immunity (dichotomous), antibiotic therapy in 48 h before or after ICU admission (dichotomous).

Accordingly, **Figure 3A** highlighted the accuracy of SAPS II for predicting the risk of 7-day mortality for all patients staying in ICU. SAPS II was able to distinguish patients who died from those who did not, with AUC value of 0.678 (95% CI = 0.657–0.700; $p < 0.001$) and accuracy of 69.3% (95% CI = 67.8–70.8%). ROC curve parameters are reported in **Table 2**. Specifically, the best cut-off value of SAPS II was of 54.5, resulting in sensitivity value of 61.9% (95% CI = 60.4–63.4%) and specificity of 67.1% (95% CI = 65.6–68.7%).

Figure 3B reported the ROC curve obtained from a logistic regression model, combining SAPS II with additional patients' characteristics at ICU admission. Notably, both AUC and accuracy of this model remained adequate (AUC = 0.637; 95% CI = 0.616–0.659; Accuracy = 65.2%; 95% CI = 63.7–66.7%). In line, the values of sensitivity and specificity for death were 49.0% (95% CI = 47.5–50.5%), and 70.0% (95% CI = 68.5–71.5%), respectively.

4.6.3. ROC curve analysis using the Support Vector Machine model

Applying two cycles of 1-NN imputation separately to the two classes of data (i.e., alive or died patients), we recovered 3258 records, approximately the 73% of the incomplete ones. After applying the SMOTE to 3782 "real" records, we obtained 1131 synthetic records for the class of died patients.

Given that, the dataset of synthetic records, which was used as the training set, included a total of 4389 records. To confirm the goodness of the training set, the distributions of primary outcome and exposure variables with those obtained from the test set were compared **(Table 3 and Figures 4–6).** Thus, a ML algorithm was developed to improve the prediction of 7-day mortality in ICU. In particular, SVM predictive model combining SAPS II with the following variables collected at ICU admission: sex, patient's origin, type of ICU admission, non-surgical treatment for acute coronary disease, surgical intervention, presence of intubation, presence of urinary catheter, presence of central vascular catheter; trauma, impaired immunity, and antibiotic therapy in 48 h before or after ICU admission. Interestingly, the ROC curve of SVM algorithm (**Figure 7**) reached an AUC of 0.896 (95% CI = 0.881–0.910; $p < 0.001$) and an accuracy of 83.5% (95% CI = 82.4–84.7%). In line, sensitivity was of 81.0% (95% CI = 79.9–82.1%) and specificity of 84.0% (95% CI = 82.9–85.1%).

4.6.4. Support Vector Machine maintained its predictive ability among patients who did not develop Healthcare Associated Infection

Since patients who developed HAIs during their ICU stay are generally at higher risk of death, it was tested the predictive performance of the SVM model among patients who did not acquire HAIs within 7 days from ICU admission. To do that, 520 patients with at least one HAIs were excluded from the test set. Notably, the SVM model did not depend on the onset of HAI, indeed both AUC (0.903; 95% Confidence Interval = 0.881–0.912; $p <0.001$) and accuracy (83.8%; 95% CI = 82.6–85.0%) remained stable **(Figure 8).** Accordingly, sensitivity (82.0%; 95% CI = 80.8–83.2%) and specificity (84.0%; 95% CI = 82.8–85.2%) were similar to those obtained in the overall analysis.

4.6.5. ROC curve analysis using the Support Vector Machine model by removing SAPS II

Since SAPS II was the best predictor in the model, the predictive performance of the classifier was evaluated, after removing SAPS II. For this reason, the Shapley plot reported in **Figure 9** shows the contribution of each predictor to the SVM model. Notably, the SVM model without SAPS II showed

an AUC of 0.653 (95% CI = 0.632–0.675; $p < 0.001$), with an accuracy of 68.4% (95% CI = 66.9–69.8%) on the test set **(Figure 10).** Accordingly, sensitivity and specificity decreased to 32.0% (95% CI = 30.5–33.5%) and 74.0% (95% CI = 72.5–75.5%), respectively.

4.6.6 Tables and figures

Table 1. Characteristics of patients according to their outcome status

Characteristics	Patients (n=3,782)	Died patients (n=875)	Alive patients (n=2,907)	p-value
Age, years	70.0 (20.0)	74.0 (17.0)	69.0 (21.0)	<0.001
Sex (% men)	60.2%	55.0%	61.7%	<0.001
Patient's origin				
Other ward/healthcare facility	70.9%	70.1%	71.1%	<0.001
Community	29.1%	29.9%	28.9%	
SAPS II at admission	49.0 (27.0)	59.0 (27.0)	46.0 (25.0)	<0.001
Type of ICU admission				
Medical	56.9%	59.3%	56.1%	<0.001
Surgical	43.1%	40.7%	43.9%	
Trauma	4.7%	2.4%	5.4%	<0.001
Impaired immunity	11.4%	15.0%	10.3%	<0.001
Non-surgical treatment for acute coronary disease	9.0%	10.2%	8.7%	0.174
Surgical intervention	34.8%	32.5%	35.5%	0.306
Antibiotic therapy in 48 hours before or after ICU admission	62.6%	62.2%	62.8%	0.744
Presence of urinary catheter at ICU admission	77.0%	75.9%	77.4%	0.351
Presence of intubation at ICU admission	62.4%	61.7%	62.6%	0.646
Presence of central venous catheter at ICU admission	40.5%	38.5%	41.0%	0.182

*Results are reported as median (interquartile range) for continuous variables, or percentage (%) for categorical variables. Statistical analyses were performed using the Mann-Whitney or the Chi-squared test.

Table 2. Coordinates of the ROC curve of logistic regression model with SAPS II alone

SAPS II values	Sensitivity	1-Specificity	SAPS II values	Sensitivity	1-Specificity	SAPS II values	Sensitivity	1-Specificity
1	0.997	0.998	34	0.899	0.751	67	0.333	0.148
2	0.995	0.998	35	0.895	0.729	68	0.318	0.135
3	0.995	0.997	36	0.879	0.708	69	0.303	0.125
4	0.994	0.997	37	0.869	0.688	70	0.293	0.115
5	0.994	0.996	38	0.857	0.672	71	0.272	0.108
6	0.993	0.994	39	0.848	0.649	72	0.258	0.099
7	0.993	0.992	40	0.839	0.628	73	0.247	0.092
8	0.992	0.990	41	0.827	0.604	74	0.224	0.085
9	0.992	0.987	42	0.814	0.582	75	0.211	0.079
10	0.992	0.986	43	0.806	0.561	76	0.197	0.073
11	0.990	0.984	44	0.779	0.539	77	0.178	0.065
12	0.990	0.982	45	0.765	0.521	78	0.161	0.057
13	0.987	0.980	46	0.750	0.501	79	0.152	0.052
14	0.985	0.973	47	0.731	0.479	80	0.144	0.048
15	0.983	0.971	48	0.718	0.461	81	0.131	0.042
16	0.981	0.967	49	0.704	0.440	82	0.121	0.035
17	0.976	0.961	50	0.693	0.422	83	0.104	0.030
18	0.975	0.958	51	0.678	0.404	84	0.095	0.025

19	0.971	0.952	52	0.667	0.387	85	0.087	0.024
20	0.969	0.946	53	0.649	0.369	86	0.072	0.021
21	0.967	0.941	54	0.634	0.347	87	0.064	0.017
22	0.965	0.933	55	0.619	0.329	88	0.055	0.015
23	0.962	0.924	56	0.583	0.306	89	0.050	0.014
24	0.958	0.915	57	0.563	0.289	90	0.048	0.011
25	0.955	0.899	58	0.541	0.272	91	0.039	0.010
26	0.953	0.890	59	0.512	0.258	92	0.035	0.009
27	0.950	0.875	60	0.486	0.245	93	0.031	0.008
28	0.947	0.859	61	0.471	0.231	94	0.025	0.007
29	0.937	0.845	62	0.446	0.212	95	0.025	0.006
30	0.934	0.829	63	0.413	0.200	96	0.023	0.004
31	0.927	0.810	64	0.394	0.187	97	0.018	0.003
32	0.919	0.791	65	0.377	0.172	98	0.016	0.003
33	0.909	0.772	66	0.353	0.159	99	0.011	0.002

Table 3. Composition of training and test sets

Outcome	Training Set	Test set
Class 0 (Alive patients)	2,596 with imputation of missing data	2,907 real records
Class 1 (Dead patients)	1193 total (662 with imputation of missing data and 1,131 after class balancing)	875 real records
Total	4,589 synthetic records	3,782 real records

Figure 1. Matrix of missing values

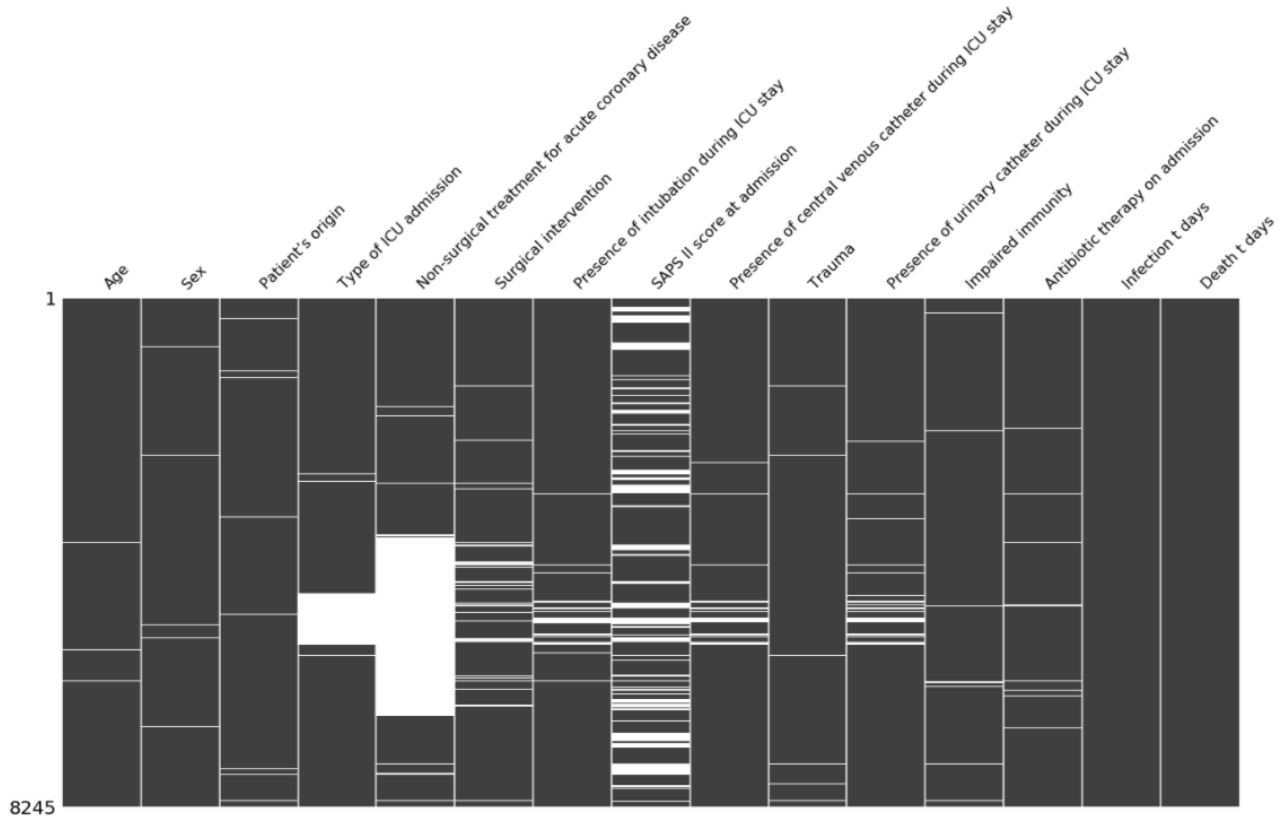


Figure 2. Selection of records with complete data satisfying inclusion criteria

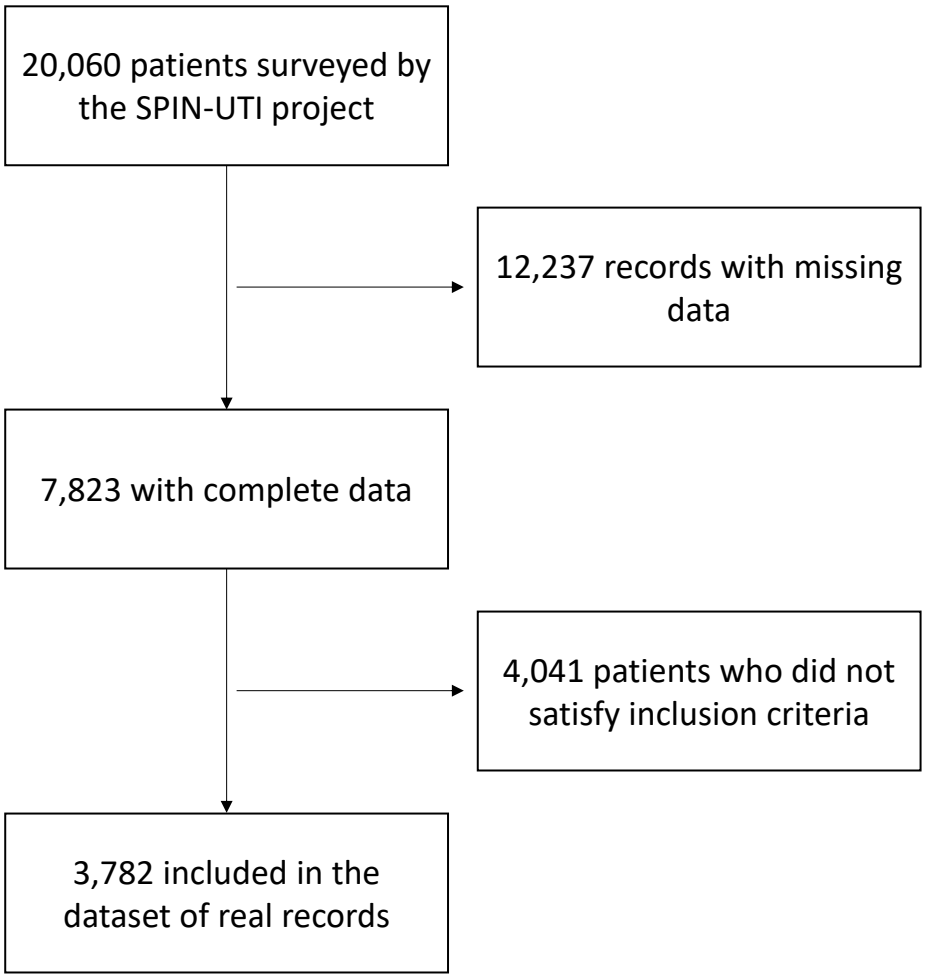
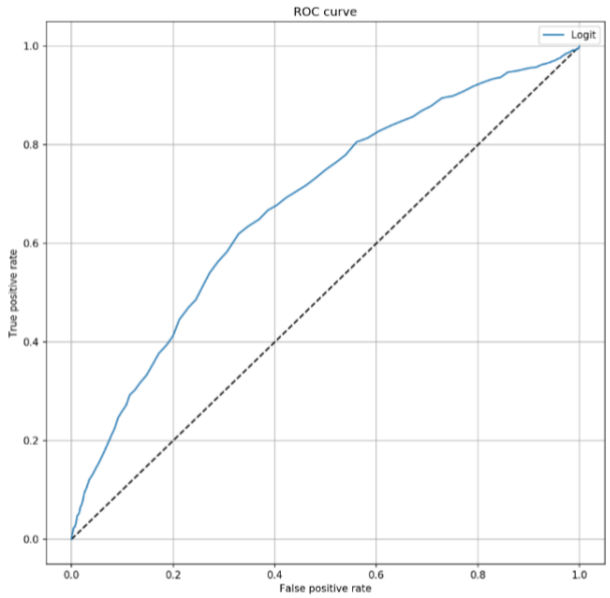


Figure 3. ROC curves of logistic regression models to predict 7-day mortality. (**A**) This curve shows the predictive performance of a logistic regression model using SAPS II alone. (**B**) This curve shows the predictive performance of a logistic regression model including additional characteristics

A



B

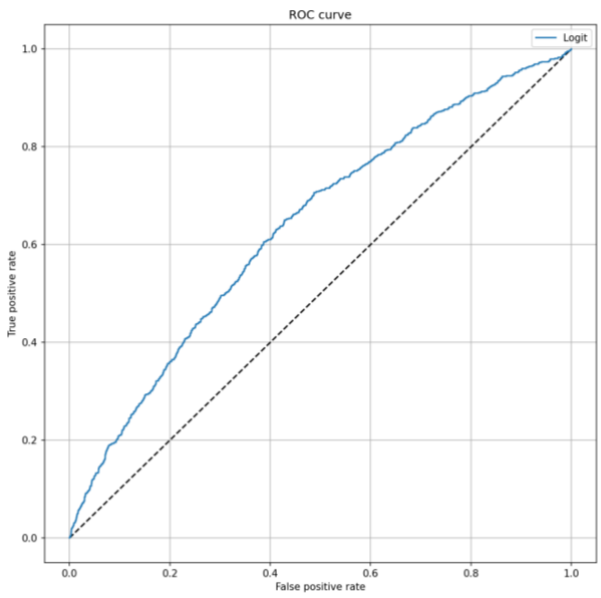
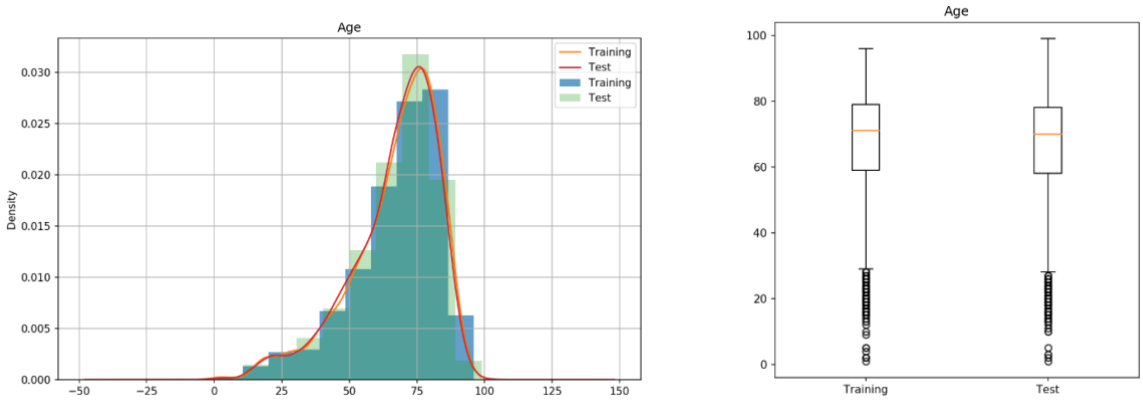


Figure 4. Comparison of Age (A) and SAPS II score (B) distributions between Training and Test sets

A



B

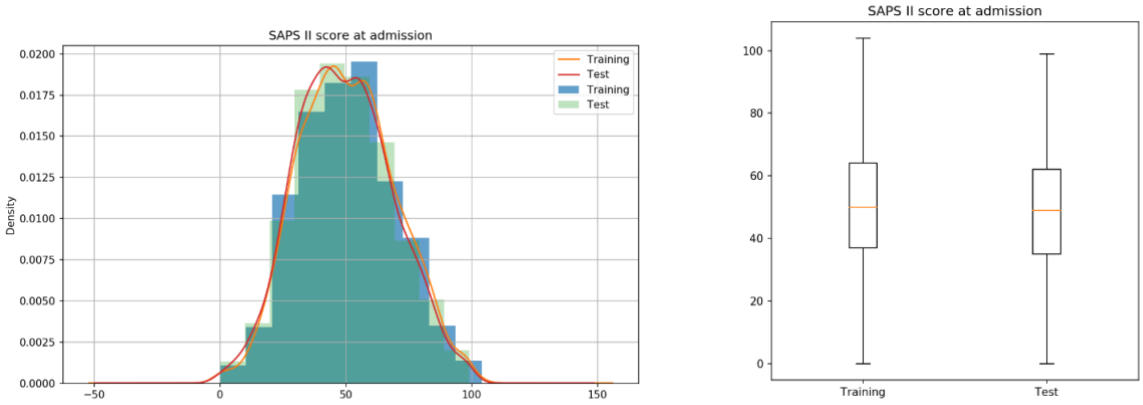


Figure 5. Comparison of dichotomous variables between Training and Test sets

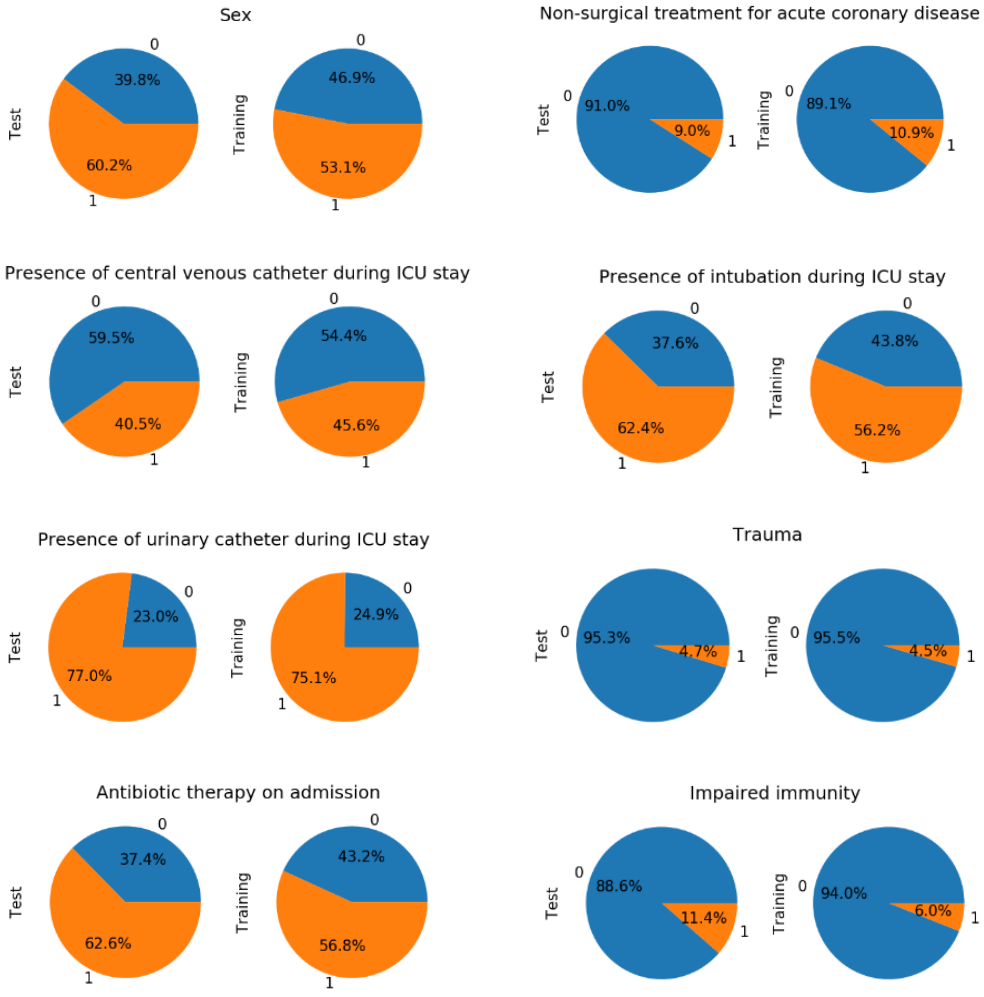


Figure 6. Comparison of categorical variables between training and test datasets

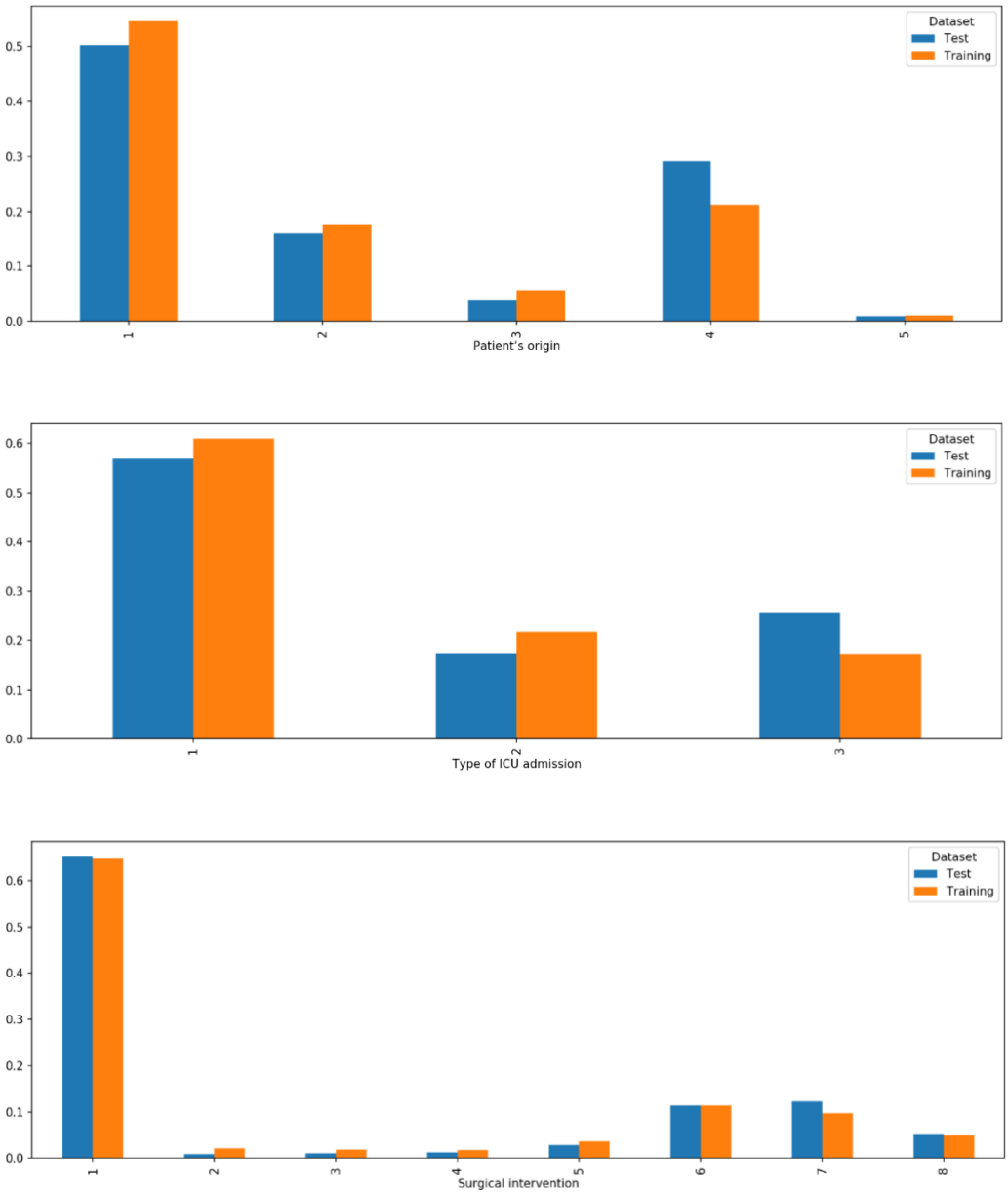


Figure 7. ROC curve of the SVM algorithm to predict 7-day mortality. This curve shows the predictive performance of the SVM algorithm including SAPS II, sex, patient's origin, type of ICU admission, non-surgical treatment for acute coronary disease, surgical intervention, presence of invasive devices at ICU admission, trauma, impaired immunity, antibiotic therapy in 48 h before or after ICU admission. The parameters applied to the SVM algorithm were C = 2 and γ = 0.3

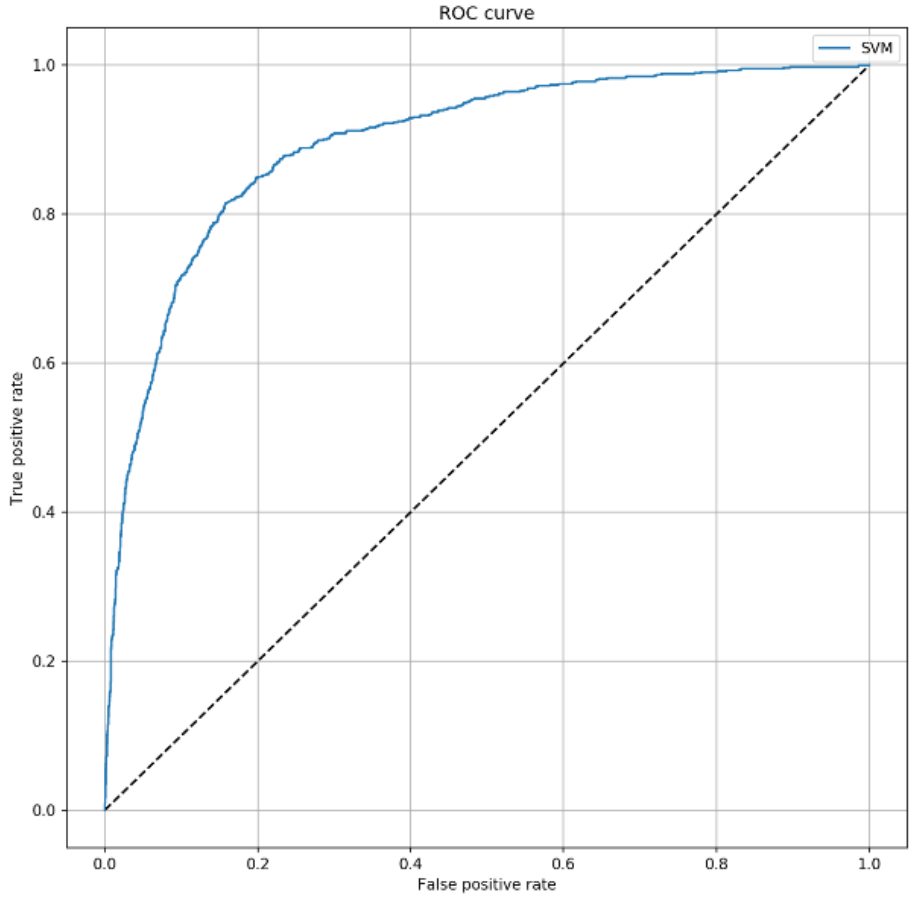


Figure 8. ROC curve of the SVM algorithm to predict 7-day mortality, by excluding infected patients. This curve shows the predictive performance of the SVM algorithm including SAPS II, sex, patient's origin, type of ICU admission, non-surgical treatment for acute coronary disease, surgical intervention, presence of invasive devices at ICU admission, trauma, impaired immunity, antibiotic therapy in 48 h before or after ICU admission. The parameters applied to the SVM algorithm were C = 2 and γ = 0.4

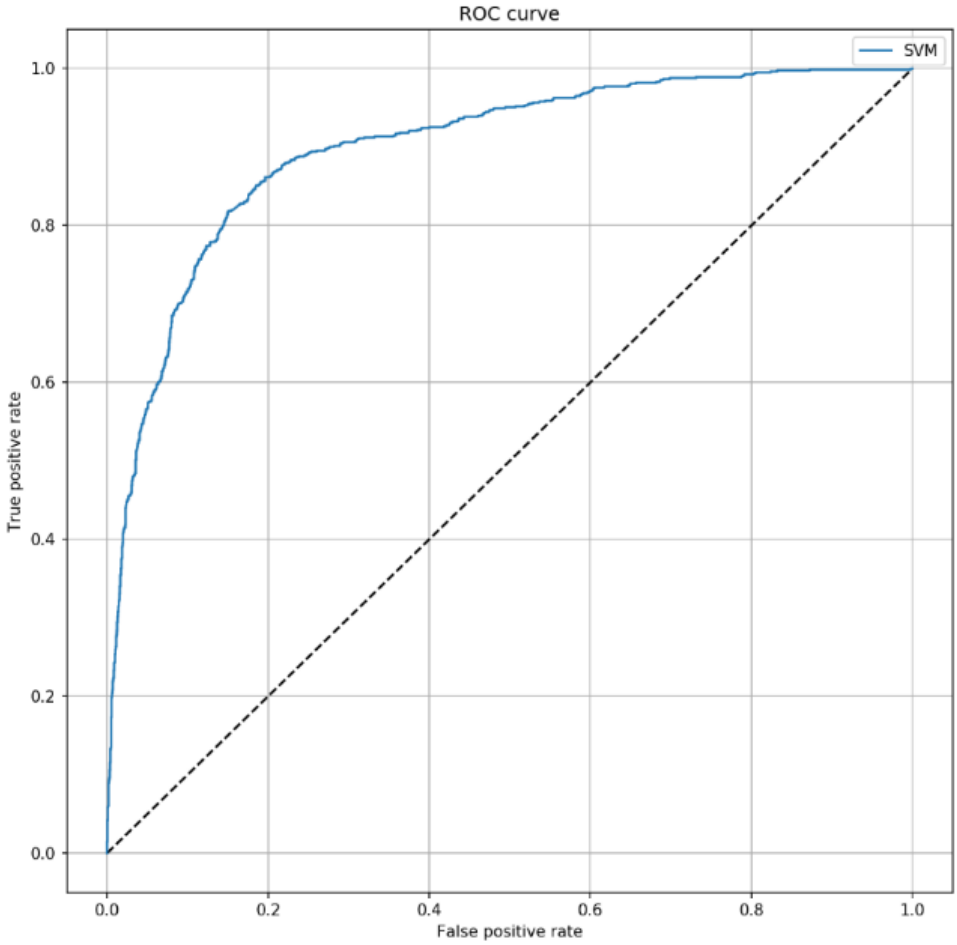


Figure 9. Shapley plot showing the contribution of each predictor to the SVM model output

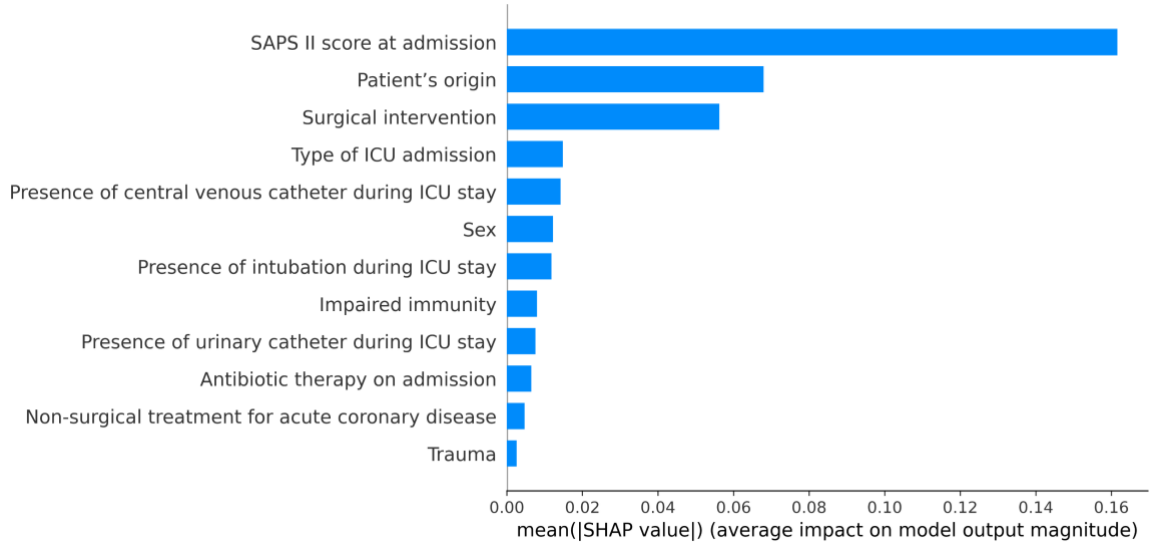
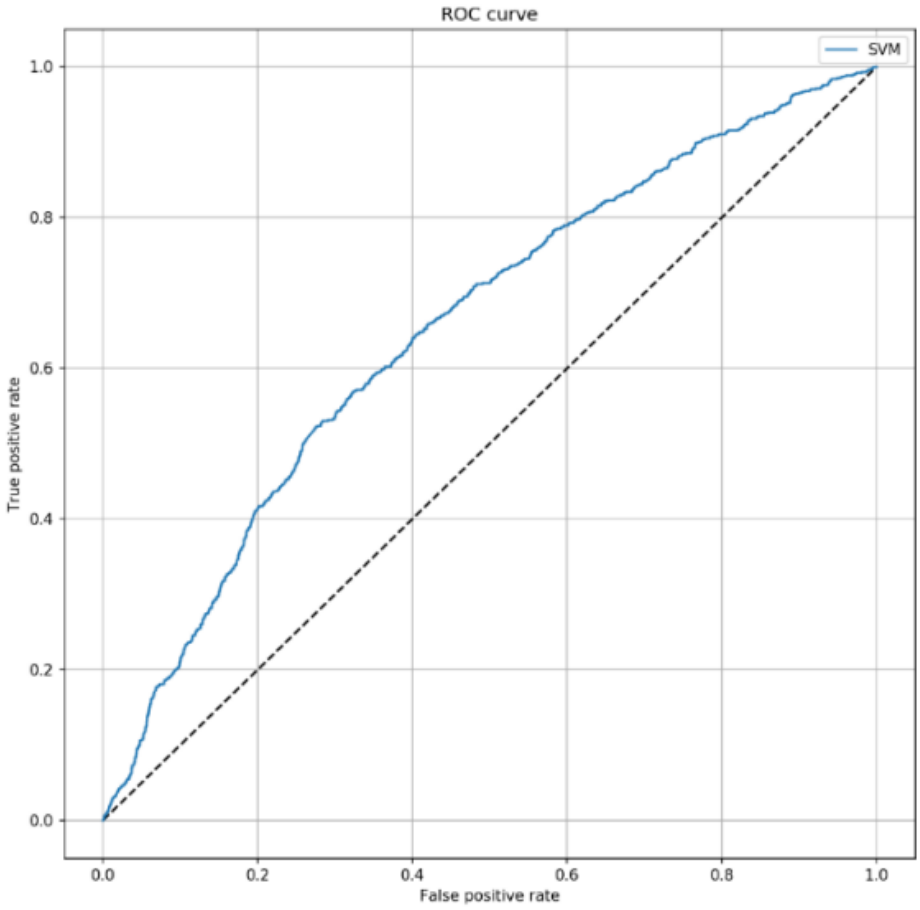


Figure 10. ROC curve of SVM algorithm predicting 7-day mortality, by excluding SAPS II score. This curve shows the predictive performance of the SVM algorithm including sex, patient's origin, type of ICU admission, non-surgical treatment for acute coronary disease, surgical intervention, presence of invasive devices at ICU admission, trauma, impaired immunity, antibiotic therapy in 48 h before or after ICU admission. The parameters applied to the SVM algorithm were C = 2 and γ = 0.1.



These tables and figures were adapted from Barchitta et al., Journal of Clincal Medicine 2021

4.7. Application of Cluster analysis to compare incidence of Healthcare-Associated Infections among Intensive Care Units

4.7.1 Characteristics of Intensive Care Units

On a total of 226 ICUs participating to the SPIN-UTI, 139 ICUs without missing information on hospital type, area of origin and total days in ICU were included. Among them, 43 ICUs joined to only one edition of the SPIN-UTI, while the remaining took part in more than one edition. With respect to characteristics at hospital level, approximately 45% of ICUs belonged to first-level hospitals and 95% to hospitals with 0-999 beds. Interestingly, 61% of them were from regions of Southern Italy. With respect to characteristics at ICU level, 79.1% ICUs were of mixed type with a median size of 8.0 beds.

4.7.2 Characteristics of clusters at hospital level

Performing a Two-Steps cluster analysis six different clusters of ICUs were distinguished according to their hospital type, area of origin and total annual days in ICU [140,206]. The remaining variables were not included in the algorithm because they did not provide additional information for improving the cluster solution.

Specifically, Cluster 1 comprised (n= 38) ICUs belonging to first-level hospitals from Southern Italy with 0-999 beds. ICUs in cluster 2 (n=22) were of first-level hospitals from Northern Italy with 0-999 beds. Cluster 3 (n=8) comprised ICUs of first- and second-level hospitals from Center Italy, while cluster 4 (n=29) included those of second-level hospitals from Southern Italy with 0-999 beds. Finally, ICUs in cluster 5 (n=20) were of second-level hospitals from Northern Italy with 0-999 beds or more (70% and 30%, respectively), while those in cluster 6 (n=22) were of third-level hospitals from Southern or Northern Italy.

In line, cluster 5 was characterized by the highest number of total annual days in hospital (median=232,298) while clusters 1 and 3 had the lowest (median= 66,457 and 69,900, respectively). Cluster 5 was also characterized by the highest total hospital costs (median= 245,411,000 €).

4.7.3 Characteristics of clusters at Intensive Care Unit level

Although most of the ICUs were of mixed type, ICUs in cluster 6 represent a significant proportion of specialized ICUs type (36.4% of surgical, medical or coronary ICUs). Although no difference across clusters was evident for the total ICU costs, the median number of beds ranged from 6 in clusters 1 and 3 to 10.5 in cluster 6. In line, ICUs in cluster 6 reported the highest total number of annual days in ICU (median=3,644).

4.7.4 Incidence of Healthcare Associated Infections across clusters

Finally, the incidence density of HAIs was compared across ICUs. Specifically, the incidence density of HAIs assumed a median value of 15.9 per 1,000 patient-days. However, the incidence of HAIs was higher in cluster 1 (median=24.2 HAIs per 1,000 patient-days), followed by cluster 6 (median= 23.5) and cluster 4 (median= 20.5), with lowest values in cluster 3 (median=10.4), cluster 2 (median= 10.5), and cluster 5 (median=12.1) (p=0.003).

4.8. Analysis of hospital and Intensive Care Unit characteristics associated with the incidence of Healthcare Associated Infections

4.8.1 Characteristics of Intensive Care Units

On a total of 226 ICUs participating to the SPIN-UTI, 62 ICUs joined to only one edition of the project, while the remaining ICUs joined more than one edition. Overall, 70.8% of ICUs were of mixed type with a median size of 8.0 beds (IQR=5.0). Moreover, 90.7% ICUs belonged to hospitals with 0-999 beds and 45.6% to 1st level hospitals**.** Median incidence density of HAIs was 16.5 per 1000 patient-days (IQR= 21.5).

4.8.2 Comparison of Intensive Care Units' characteristics between Italian regions

Overall, ICUs characteristics were compared between Northern (*n* = 78; 34.5%), Center (*n* = 14; 6.2 %) and Southern Italy (*n* = 134; 59.3%). Particularly, ICUs of Center Italy were more likely to belong to 2nd level hospitals than those of Northern and Southern Italy ($p = 0.049$). Specifically, these ICUs also reported lower total days in hospital ($p < 0.001$) and were more likely to be mixed ICUs ($p = 0.007$) than those in Northern and Southern Italy.

4.8.3 Association between Intensive Care Units characteristics and Healthcare Associated Infections

Although no associations between hospital and ICU characteristics and incidence density of HAIs were evident, incidence density of HAIs varied across Italian regions, so that ICUs of Southern Italy reported the highest value ($p < 0.001$). For these reasons, the association of hospital and ICU characteristics with HAI incidence density was also evaluated across Italian regions.

Specifically, the association of ICUs' characteristics with incidence density of HAIs was examined by linear regression models and the results were reported as β and its standard error (SE).

A positive association between total days in ICU and incidence density of HAIs in Northern regions was found ($p = 0.002$). By contrast, ICUs of Center Italy reported a positive association with ICU size ($p = 0.020$), total days in hospital ($p = 0.037$) and total days in ICU ($p = 0.006$). Finally, in regions from Southern Italy, a positive association between hospital size and incidence density of HAIs was found ($p = 0.033$)**.**

5. Discussion

Identifying patients at higher risk of HAIs represents a major challenge for public health, suggesting the need for novel tools that can guide patient management in ICUs [126,171,183]. Assessing gender-differences in risk factors and outcomes among patients admitted to Italian ICUs, is one of the most critical Public Health issues. In line, gender differences pose per se a higher risk for death among patients admitted to ICUs, also considering other risk factors. Here, gender-differences in patients' characteristics at ICU admission resulted into a higher risk of death and lower survival among females, underlining that is still important to guarantee gender equality in ICUs and other hospital wards. Beyond gender, other factors have been associated with the risk of HAIs and related outcomes [10]. With this in mind, several EWSs have been developed to evaluate disease severity and to predict the risk of adverse outcomes during ICU stay [15-18]. Among these, SAPS II represents one of the most widely used in the ICU setting and, therefore, is the most represented in the SPIN-UTI dataset. Interestingly, patients who developed at least one HAI had a higher SAPS II on ICU admission than patients who remained uninfected [37,41,43]. However, ROC curve analysis discouraged a predictive application of SAPS II because both the AUC and accuracy were very low, despite being significant. Moreover, also the prolonged use of invasive devices, impaired immunity, surgical intervention, and comorbidity have been reported to be the main risk factors for HAIs in ICUs [10,217]. Indeed, it was hypothesized that combining SAPS II with other patient characteristics could improve the predictive performance of the model. In this scenario, ML approaches represent a possible strategy for healthcare facilities, making it possible to build specific prediction models targeted at the demographics and clinical characteristics of patients [168,169]. Accordingly, ML systems have been developed in many fields of medicine, including infectious disease control and clinical decision support [218].

In this thesis, a ML model that combined SAPS II with additional patient characteristics routinely collected at ICU admission was developed and tested, showing a higher performance (AUC= 0.90). Accuracy on the test set was 88%, with precision and recall values of 95% and 91% for non-infected patients, and 60% and 73% for those who developed at least one HAI, respectively. Although SAPS II was the predictor that had the greatest weight in the model as demonstrated by the sensitivity analysis, the inclusion of additional characteristics improved the prediction of patients who developed HAIs in ICUs significantly.

Notably, the ML algorithm showed a better predictive performance than conventional statistical approaches, suggesting the SVM as a tool to identify and predict patients at higher risk of HAIs at ICU admission. Therefore, this novel approach could provide clinicians with sufficient time to potentially prevent HAIs and mitigate their severity by targeting specific IPC interventions at high-risk groups.

Moreover, understanding modifiable and non-modifiable risk factors is crucial for controlling incidence of CAUTI and preventing adverse outcomes [219]. Thus, cluster analysis was applied to identify patients at higher risk of CAUTIs, according to their characteristics at ICU admission. Specifically, being women, having higher SAPS II, medical type of ICU admission, and being trauma patients were positively associated with CAUTIs. Moreover, patients with urinary catheterization exhibited higher incidence of UTIs, while the incidence of CAUTI increased with increasing duration of catheterization.

Specifically, three clusters of patients based on their characteristics at ICU admission were identified. While cluster 1 consisted more of patients with a medical type of ICU admission who came from the community, the other two clusters were mostly characterized by patients who came from other wards or hospitals and with various types of ICU admission. However, the proportion of patients who received antibiotics 48 hours before or after ICU admission was higher in cluster 3. Interestingly, differences in terms of urinary catherization and its duration across clusters were

noted, certainly the main risk factors for CAUTI [220-223]. Across different clusters, patients in cluster 1 showed higher incidence of CAUTIs than those in other clusters. If untreated, uropathogens can also cross the tubular epithelial cell barrier and cause sepsis [224], associated with increase mortality rate [225]. In line, 37.0% of patients with CAUTI was aggravated by sepsis. Moreover, this percentage increased with increasing duration of catheterization.

Next, an integrated approach of cluster, visual and conventional statistical Analyses was applied to identify determinants of pneumonia, sepsis, and death in ICU patients. Cluster analysis demonstrated that data could be reasonably represented by three clusters of patients based on their characteristics at ICU admission. Particularly, Sankey diagrams showed flows of patients from their admission to ICU and how each cluster and duration of intubation contribute to the diagnosis of pneumonia. Moreover, these graphs represent the contributions of *Acinetobacter baumannii*, *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*-associated pneumonia to the risk of sepsis and death. Next, it has been estimated incidence of sepsis, risk factors and the associated mortality impact. In particular, the ability of patient characteristics in predicting the risk of sepsis has been tested and displayed. Specifically, sepsis was found as one of the main causes of death and increased length of stay in UTI.

Next, the accuracy of SAPS II alone to identify patients who died within seven days from their admission in ICUs was evaluated. Although AUC obtained was statistically significant, the low accuracy of nearly 69% discouraged the routinely application of SAPS II to achieve this purpose. Hence, was hypothesized that combining SAPS II with other variables collected at ICU admission could improve the prediction of 7-day mortality. To do that, the SVM algorithm, which combined SAPS II with the following patients' characteristics at ICU admission, was applied. This model exhibited an AUC of 0.90 with an accuracy of 83.5% on the test set. Interestingly, its predictive performance was higher than SAPS II alone and even than a logistic regression model including additional patients' characteristics collected at ICU admission. Interestingly, the performance for

predicting 7-day mortality was also similar in only patients who did not acquired HAIs during their hospitalization [48,168,169,226,227]**.**

Moreover, hospital and ICUs characteristics had also an impact on patient outcomes. Thus, a clustering approach was applied to categorize ICUs according to their characteristics. In particular, structural indicators could be crucial to effective infection control in hospitals and ICUs. With this in mind, the incidence of HAIs was compared between different clusters [228,229]. According to their area of origin, hospital type, and total days in ICUs, six clusters of Italian ICUs were identified. Notably, three clusters were completely or mostly characterized by ICUs from Southern Italy, with an incidence density of HAIs higher than in the other ones. This in turn, reflects the possible trend from Northern to Southern Italy, previously reported by several studies [78,79]. Accordingly, ICUs from Center and Northern Italy were included in the remaining clusters, which exhibited a lower incidence of HAIs.

It is worth mentioning that the relatively similar incidences of HAIs between clusters of ICUs from Southern Italy pointed out further considerations. In fact, the incidence density was around 20-24 HAIs per 1,000 patient-days regardless of whether ICUs were in first-, second-, or third-level hospitals. By contrast, the comparison between those clusters characterized by ICUs of first-level hospital with similar sizes showed an excess of 14 HAIs per 1,000 patient-days in ICUs from Southern Italy. This, in turns, suggested the crucial role of additional components to structural indicators to effective infection control in hospitals and ICUs. These components were not only structural characteristics of the ICUs, but they also regarded the organization of infection control, the use of guidelines, education, and training, and the development of multimodal strategies [51].

Specifically, incidence density of HAIs ranged from 12.4 per 1000 patient-days in Northern regions to 21.4 HAIs per 1000 patient-days in Southern regions [78,79]. Next, was further investigated what hospital and ICU indicators might affect the incidence of HAIs. In general, the number of hospital beds seemed to be associated with HAIs in Southern regions, while the number

of ICU beds showed a similar association in Center regions. Moreover, the total number of patient-days in hospital and in ICU were positively associated with the incidence density of HAIs. Given that it is not possible to understand whether increased incidence of HAIs was a cause or a consequence of higher bed occupancy and longer length of stay in hospital and ICU, these results suggest that should be necessary to evaluate additional factors that represent manageable ways to reduce HAI incidence [51].

6. Conclusion and future perspectives

The application of ML algorithms for predictive purposes may help to solve many public health issues, including those related to critically ill patients. In addition, these algorithms could overcome limitations of existing conventional tools, such as early warning scores [15-19,142-145,226,230]. One of the main strengths of this thesis includes some methodological aspects, such as the fact that the model was developed and tested on large datasets obtained through patient-based prospective surveillance across Italy.

The present thesis, indeed, could be useful to allow an Italian benchmark for planning preventive strategies in the future, and to compare and validate these findings with those obtained in other European countries. Moreover, the use of ML algorithms suggested a better predictive performance than conventional statistical approach. However, some aspects and limitations are to be considered. The first one is that ML algorithms should not be seen as substitutes for existing EWSs but could support clinicians in the decision- making process, such as the identification of those patients who need more attention because of their considerable risk of adverse outcomes. Indeed, by removing the SAPS II, the model significantly reduced its predictive performance. Moreover, this novel approach must not be seen as a fixed model, and it could be integrated and/or modified according to specific needs.

The second limitation is that ML algorithms require a lot of data and complete records. Although data used can be easily collectible at ICU admission, approximately 60% of SPIN-UTI patients had both structural missing and missing at random data. Although this did not completely exclude potential bias related to the high proportion of missing data, incomplete records were used to generate recovered data.

The third limitation is that ML is frequently referred as a ‘black box’ for clinicians, who expect to become familiar with it and to be able to pinpoint why a decision is suggested. For this reason,

data scientists are trying to develop more interpretable algorithms in medical fields, to allow a better understanding and communication of specific Public Health issues [164,231].

Overall, the ML algorithms provide clinicians the benefit of making prevention and diagnosis as early as possible, in a context of precision medicine applicable to all settings. In particular, applying SVM algorithms could support medical professionals in parameter optimization, clustering and classification problems [164,196,218,232-236].

Indeed, in a broader context of personalized prevention and medicine against infections, combining patient's characteristics and drug history data could guide future preventive interventions tailored to specific subgroups of patients at the highest risk [237]. However, more efforts are needed to overcome limitations of current research and to bring benefits in clinical practice, especially for the management of communicable diseases [238].

With respect to the indicators that might affect the incidence of HAIs, regional distribution of ICUs included should be taken into accounts.

Although differences between hospitals and ICUs across Italian regions might at least partially explain the different rates of HAIs between Northern, Central, and Southern Italy, the majority of ICUs included in this study were from Southern regions. Moreover, potential historical biases due to the fact that some ICUs participated in only one edition of the SPIN-UTI project are to be considered.

In conclusion, this thesis suggests again how HAIs remained an important threat for Public Health in Italy, also suggesting the need to further investigate the relationship of patients, hospitals, and ICU characteristics with incidence of HAIs. Although our approach seems to be promising, further studies should be encouraged to confirm our findings and to understand better the advantages of applying ML models. Moreover, further research involving interdisciplinary professionals– such as Epidemiologists, Data Scientists, Statisticians and Public Health professionals - should be recommended to highlight relationships not-known *a priori*.

7. References

1. Homauni, A.; Zargar Balaye Jame, S.; Hazrati, E.; Markazi-Moghaddam, N. Intensive Care Unit Risk Assessment: A Systematic Review. *Iran J Public Health* **2020**, *49*, 1422-1431, doi:10.18502/ijph.v49i8.3865.
2. Asefzadeh, S.; Yarmohammadian, M.H.; Nikpey, A.; Atighechian, G. Clinical risk assessment in intensive care unit. *Int J Prev Med* **2013**, *4*, 592-598.
3. Smith, G.; Nielsen, M. ABC of intensive care. Criteria for admission. *BMJ* **1999**, *318*, 1544-1547, doi:10.1136/bmj.318.7197.1544.
4. Guidelines for the utilisation of intensive care units. European Society of Intensive Care Medicine. *Intensive Care Med* **1994**, *20*, 163-164, doi:10.1007/BF01707674.
5. Marshall, J.C.; Bosco, L.; Adhikari, N.K.; Connolly, B.; Diaz, J.V.; Dorman, T.; Fowler, R.A.; Meyfroidt, G.; Nakagawa, S.; Pelosi, P.; et al. What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine. *J Crit Care* **2017**, *37*, 270-276, doi:10.1016/j.jcrc.2016.07.015.
6. Nielsen, A.B.; Thorsen-Meyer, H.C.; Belling, K.; Nielsen, A.P.; Thomas, C.E.; Chmura, P.J.; Lademann, M.; Moseley, P.L.; Heimann, M.; Dybdahl, L.; et al. Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the Danish National Patient Registry and electronic patient records. *Lancet Digit Health* **2019**, *1*, e78-e89, doi:10.1016/S2589-7500(19)30024-X.
7. Christiansen, C.F.; Christensen, S.; Johansen, M.B.; Larsen, K.M.; Tønnesen, E.; Sørensen, H.T. The impact of pre-admission morbidity level on 3-year mortality after intensive care: a Danish cohort study. *Acta Anaesthesiol Scand* **2011**, *55*, 962-970, doi:10.1111/j.1399-6576.2011.02480.x.
8. Roedl, K.; Amann, D.; Eichler, L.; Fuhrmann, V.; Kluge, S.; Müller, J. The chronic ICU patient: Is intensive care worthwhile for patients with very prolonged ICU-stay (≥ 90 days)? *Eur J Intern Med* **2019**, *69*, 71-76, doi:10.1016/j.ejim.2019.08.024.
9. Vincent, J.L.; Singer, M. Critical care: advances and future perspectives. *Lancet* **2010**, *376*, 1354-1361, doi:10.1016/S0140-6736(10)60575-2.
10. Marcel, J.P.; Alfa, M.; Baquero, F.; Etienne, J.; Goossens, H.; Harbarth, S.; Hryniewicz, W.; Jarvis, W.; Kaku, M.; Leclercq, R.; et al. Healthcare-associated infections: think globally, act locally. *Clin Microbiol Infect* **2008**, *14*, 895-907, doi:10.1111/j.1469-0691.2008.02074.x.
11. Suetens, C.; Latour, K.; Kärki, T.; Ricchizzi, E.; Kinross, P.; Moro, M.L.; Jans, B.; Hopkins, S.; Hansen, S.; Lyytikäinen, O.; et al. Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017. *Euro Surveill* **2018**, *23*, doi:10.2807/1560-7917.ES.2018.23.46.1800516.

12. Delle Karth, G.; Meyer, B.; Bauer, S.; Nikfardjam, M.; Heinz, G. Outcome and functional capacity after prolonged intensive care unit stay. *Wien Klin Wochenschr* **2006**, *118*, 390-396, doi:10.1007/s00508-006-0616-z.
13. Carson, S.S.; Garrett, J.; Hanson, L.C.; Lanier, J.; Govert, J.; Brake, M.C.; Landucci, D.L.; Cox, C.E.; Carey, T.S. A prognostic model for one-year mortality in patients requiring prolonged mechanical ventilation. *Crit Care Med* **2008**, *36*, 2061-2069, doi:10.1097/CCM.0b013e31817b8925.
14. Garrouste-Orgeas, M.; Timsit, J.F.; Tafflet, M.; Misset, B.; Zahar, J.R.; Soufir, L.; Lazard, T.; Jamali, S.; Mourvillier, B.; Cohen, Y.; et al. Excess risk of death from intensive care unit-acquired nosocomial bloodstream infections: a reappraisal. *Clin Infect Dis* **2006**, *42*, 1118-1126, doi:10.1086/500318.
15. Gerry, S.; Bonnici, T.; Birks, J.; Kirtley, S.; Virdee, P.S.; Watkinson, P.J.; Collins, G.S. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. *BMJ* **2020**, *369*, m1501, doi:10.1136/bmj.m1501.
16. Brennan, T.A.; Leape, L.L.; Laird, N.M.; Hebert, L.; Localio, A.R.; Lawthers, A.G.; Newhouse, J.P.; Weiler, P.C.; Hiatt, H.H.; I, H.M.P.S. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. 1991. *Qual Saf Health Care* **2004**, *13*, 145-151; discussion 151-142, doi:10.1136/qshc.2002.003822.
17. America, I.o.M.U.C.o.Q.o.H.C.i. To Err is Human: Building a Safer Health System. 2000.
18. Vincent, C.; Neale, G.; Woloshynowych, M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ* **2001**, *322*, 517-519, doi:10.1136/bmj.322.7285.517.
19. Hillman, K.M.; Bristow, P.J.; Chey, T.; Daffurn, K.; Jacques, T.; Norman, S.L.; Bishop, G.F.; Simmons, G. Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med* **2002**, *28*, 1629-1634, doi:10.1007/s00134-002-1496-y.
20. Storr, J.; Twyman, A.; Zingg, W.; Damani, N.; Kilpatrick, C.; Reilly, J.; Price, L.; Egger, M.; Grayson, M.L.; Kelley, E.; et al. Core components for effective infection prevention and control programmes: new WHO evidence-based recommendations. *Antimicrob Resist Infect Control* **2017**, *6*, 6, doi:10.1186/s13756-016-0149-9.
21. World Health Organization. "About infection prevention and control". Available online: https://www.who.int/teams/integrated-health-services/infection-prevention-control/about
22. Price, L.; MacDonald, J.; Melone, L.; Howe, T.; Flowers, P.; Currie, K.; Curran, E.; Ness, V.; Waddell, D.; Manoukian, S.; et al. Effectiveness of national and subnational infection prevention and control interventions in high-income and upper-middle-income countries: a systematic review. *Lancet Infect Dis* **2018**, *18*, e159-e171, doi:10.1016/S1473-3099(17)30479-6.
23. World Health Organization. "5 May, World Hand Hygiene Day". Available online: https://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministero&id=5472
24. Sharma, R.; Sharma, M.; Koushal, V. Compliance to hand hygiene world health organization guidelines in hospital care. *Int J Prev Med* **2014**, *5*, 127-128.

25. European Center for Desease Prevention and Control. The Components of Surveillance. Available online: https://wiki.ecdc.europa.eu/fem/Pages/The%20Components%20of%20Surveillance.aspx
26. Agodi, A.; Auxilia, F.; Barchitta, M.; Brusaferro, S.; D'Alessandro, D.; Montagna, M.T.; Orsi, G.B.; Pasquarella, C.; Torregrossa, V.; Suetens, C.; et al. Building a benchmark through active surveillance of intensive care unit-acquired infections: the Italian network SPIN-UTI. *J Hosp Infect* **2010**, *74*, 258-265, doi:10.1016/j.jhin.2009.08.015.
27. Pearce, N. Classification of epidemiological study designs. *Int J Epidemiol* **2012**, *41*, 393-397, doi:10.1093/ije/dys049.
28. Barchitta, M.; Matranga, D.; Quattrocchi, A.; Bellocchi, P.; Ruffino, M.; Basile, G.; Agodi, A. Prevalence of surgical site infections before and after the implementation of a multimodal infection control programme. *J Antimicrob Chemother* **2012**, *67*, 749-755, doi:10.1093/jac/dkr505.
29. Center for Desease Prevention and Control. Evaluating and Improving Surveillance Available online: https://www.cdc.gov/csels/dsepd/ss1978/lesson5/section7.html (accessed on
30. Mayon-White, R.T.; Ducel, G.; Kereselidze, T.; Tikomirov, E. An international survey of the prevalence of hospital-acquired infection. *J Hosp Infect* **1988**, *11 Suppl A*, 43-48, doi:10.1016/0195-6701(88)90164-8.
31. European Center for Desease Prevention and Control." *Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals 2011–2012*; Stockholm, 2013.
32. Eurosurveillance Editorial Team. Erratum for Euro Surveill. 2018;23(46). *Euro Surveill* **2018**, *23*, doi:10.2807/1560-7917.ES.2018.23.47.181122e1.
33. Vincent, J.L.; Moreno, R. Clinical review: scoring systems in the critically ill. *Crit Care* **2010**, *14*, 207, doi:10.1186/cc8204.
34. McGaughey, J.; Alderdice, F.; Fowler, R.; Kapila, A.; Mayhew, A.; Moutray, M. Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. *Cochrane Database Syst Rev* **2007**, CD005529, doi:10.1002/14651858.CD005529.pub2.
35. Smith, M.E.; Chiovaro, J.C.; O'Neil, M.; Kansagara, D.; Quiñones, A.R.; Freeman, M.; Motu'apuaka, M.L.; Slatore, C.G. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. *Ann Am Thorac Soc* **2014**, *11*, 1454-1465, doi:10.1513/AnnalsATS.201403-102OC.
36. Whittington, J.; White, R.; Haig, K.M.; Slock, M. Using an automated risk assessment report to identify patients at risk for clinical deterioration. *Jt Comm J Qual Patient Saf* **2007**, *33*, 569-574, doi:10.1016/s1553-7250(07)33061-4.
37. Sadaka, F.; EthmaneAbouElMaali, C.; Cytron, M.A.; Fowler, K.; Javaux, V.M.; O'Brien, J. Predicting Mortality of Patients With Sepsis: A Comparison of APACHE II and APACHE III Scoring Systems. *J Clin Med Res* **2017**, *9*, 907-910, doi:10.14740/jocmr3083w.

38. Le Gall, J.R.; Loirat, P.; Alperovitch, A.; Glaser, P.; Granthil, C.; Mathieu, D.; Mercier, P.; Thomas, R.; Villers, D. A simplified acute physiology score for ICU patients. *Crit Care Med* **1984**, *12*, 975-977, doi:10.1097/00003246-198411000-00012.
39. Le Gall, J.R.; Lemeshow, S.; Saulnier, F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* **1993**, *270*, 2957-2963, doi:10.1001/jama.270.24.2957.
40. Gilani, M.T.; Razavi, M.; Azad, A.M. A comparison of Simplified Acute Physiology Score II, Acute Physiology and Chronic Health Evaluation II and Acute Physiology and Chronic Health Evaluation III scoring system in predicting mortality and length of stay at surgical intensive care unit. *Niger Med J* **2014**, *55*, 144-147, doi:10.4103/0300-1652.129651.
41. Mungan , I.b.; Bektaş , S.e.; Çavuş , M.A.; Sarı , S.; Turan , S. The predictive power of SAPS-3 and SOFA scores and their relations with patients outcomes in the Surgical Intensive Care Unit. **2019**.
42. Haddadi, A.; Ledmani, M.; Gainier, M.; Hubert, H.; Tagne, J.; De Micheaux, P. Comparing the APACHE II, SOFA, LOD, and SAPS II scores in patients who have developed a nosocomial infection.
43. Agodi, A.; Barchitta, M.; Auxilia, F.; Brusaferro, S.; D'Errico, M.M.; Montagna, M.T.; Pasquarella, C.; Tardivo, S.; Arrigoni, C.; Fabiani, L.; et al. Epidemiology of intensive care unit-acquired sepsis in Italy: results of the SPIN-UTI network. *Ann Ig* **2018**, *30*, 15-21, doi:10.7416/ai.2018.2247.
44. Metnitz, P.G.; Moreno, R.P.; Almeida, E.; Jordan, B.; Bauer, P.; Campos, R.A.; Iapichino, G.; Edbrooke, D.; Capuzzo, M.; Le Gall, J.R.; et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. *Intensive Care Med* **2005**, *31*, 1336-1344, doi:10.1007/s00134-005-2762-6.
45. Moreno, R.P.; Metnitz, P.G.; Almeida, E.; Jordan, B.; Bauer, P.; Campos, R.A.; Iapichino, G.; Edbrooke, D.; Capuzzo, M.; Le Gall, J.R.; et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med* **2005**, *31*, 1345-1355, doi:10.1007/s00134-005-2763-5.
46. Mungan, İ.; Bektaş, Ş.; Altınkaya Çavuş, M.; Sarı, S.; Turan, S. The predictive power of SAPS-3 and SOFA scores and their relations with patient outcomes in the Surgical Intensive Care Unit. *Turk J Surg* **2019**, *35*, 124-130, doi:10.5578/turkjsurg.4223.
47. Higgins, T.L. Quantifying risk and benchmarking performance in the adult intensive care unit. *J Intensive Care Med* **2007**, *22*, 141-156, doi:10.1177/0885066607299520.
48. Strand, K.; Flaatten, H. Severity scoring in the ICU: a review. *Acta Anaesthesiol Scand* **2008**, *52*, 467-478, doi:10.1111/j.1399-6576.2008.01586.x.
49. Stiller, A.; Salm, F.; Bischoff, P.; Gastmeier, P. Relationship between hospital ward design and healthcare-associated infection rates: a systematic review and meta-analysis. *Antimicrob Resist Infect Control* **2016**, *5*, 51, doi:10.1186/s13756-016-0152-1.

50. Despotovic, A.; Milosevic, B.; Milosevic, I.; Mitrovic, N.; Cirkovic, A.; Jovanovic, S.; Stevanovic, G. Hospital-acquired infections in the adult intensive care unit-Epidemiology, antimicrobial resistance patterns, and risk factors for acquisition and mortality. *Am J Infect Control* **2020**, *48*, 1211-1215, doi:10.1016/j.ajic.2020.01.009.
51. Zingg, W.; Holmes, A.; Dettenkofer, M. Corrections. Hospital organisation, management, and structure for prevention of health-care-associated infection: a systematic review and expert consensus. *Lancet Infect Dis* **2015**, *15*, 263, doi:10.1016/S1473-3099(15)70069-1.
52. Verburg, I.W.M.; Holman, R.; Dongelmans, D.; de Jonge, E.; de Keizer, N.F. Is patient length of stay associated with intensive care unit characteristics? *J Crit Care* **2018**, *43*, 114-121, doi:10.1016/j.jcrc.2017.08.014.
53. Bilimoria, K.Y. Facilitating Quality Improvement: Pushing the Pendulum Back Toward Process Measures. *JAMA* **2015**, *314*, 1333-1334, doi:10.1001/jama.2015.12470.
54. Pitches, D.W.; Mohammed, M.A.; Lilford, R.J. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* **2007**, *7*, 91, doi:10.1186/1472-6963-7-91.
55. Alexopoulos, E.C.; Batzi, E.; Messolora, F.; Jelastopulu, E. Wide range of point prevalences of healthcare-associated infections in Western Greece. *Epidemiol Infect* **2011**, *139*, 1734-1739, doi:10.1017/S0950268810002670.
56. Barchitta, M.; Maugeri, A.; Favara, G.; Riela, P.M.; La Mastra, C.; La Rosa, M.C.; San Lio, R.M.; Gallo, G.; Mura, I.; Agodi, A.; et al. Cluster analysis identifies patients at risk of catheter-associated urinary tract infections in intensive care unit: findings from the SPIN-UTI network. *J Hosp Infect* **2020**, doi:10.1016/j.jhin.2020.09.030.
57. Zarrilli, R.; Di Popolo, A.; Bagattini, M.; Giannouli, M.; Martino, D.; Barchitta, M.; Quattrocchi, A.; Iula, V.D.; de Luca, C.; Scarcella, A.; et al. Clonal spread and patient risk factors for acquisition of extensively drug-resistant Acinetobacter baumannii in a neonatal intensive care unit in Italy. *J Hosp Infect* **2012**, *82*, 260-265, doi:10.1016/j.jhin.2012.08.018.
58. Kołpa, M.; Wałaszek, M.; Gniadek, A.; Wolak, Z.; Dobroś, W. Incidence, Microbiological Profile and Risk Factors of Healthcare-Associated Infections in Intensive Care Units: A 10 Year Observation in a Provincial Hospital in Southern Poland. *Int J Environ Res Public Health* **2018**, *15*, doi:10.3390/ijerph15010112.
59. Prin, M.; Li, G. Complications and in-hospital mortality in trauma patients treated in intensive care units in the United States, 2013. *Inj Epidemiol* **2016**, *3*, 18, doi:10.1186/s40621-016-0084-5.
60. Gordts, B.; Vrijens, F.; Hulstaert, F.; Devriese, S.; Van de Sande, S. The 2007 Belgian national prevalence survey for hospital-acquired infections. *J Hosp Infect* **2010**, *75*, 163-167, doi:10.1016/j.jhin.2010.01.006.

61. Rodríguez-Acelas, A.L.; de Abreu Almeida, M.; Engelman, B.; Cañon-Montañez, W. Risk factors for health care-associated infection in hospitalized adults: Systematic review and meta-analysis. *Am J Infect Control* **2017**, *45*, e149-e156, doi:10.1016/j.ajic.2017.08.016.
62. Kim, J.H.; Kwon, Y.S.; Baek, M.S. Machine Learning Models to Predict 30-Day Mortality in Mechanically Ventilated Patients. *J Clin Med* **2021**, *10*, doi:10.3390/jcm10102172.
63. Copnell, B.; Hagger, V.; Wilson, S.G.; Evans, S.M.; Sprivulis, P.C.; Cameron, P.A. Measuring the quality of hospital care: an inventory of indicators. *Intern Med J* **2009**, *39*, 352-360, doi:10.1111/j.1445-5994.2009.01961.x.
64. Groene, O.; Skau, J.K.; Frølich, A. An international review of projects on hospital performance assessment. *Int J Qual Health Care* **2008**, *20*, 162-171, doi:10.1093/intqhc/mzn008.
65. Mainz, J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* **2003**, *15*, 523-530, doi:10.1093/intqhc/mzg081.
66. Ngantcha, M.; Le-Pogam, M.A.; Calmus, S.; Grenier, C.; Evrard, I.; Lamarche-Vadel, A.; Rey, G. Hospital quality measures: are process indicators associated with hospital standardized mortality ratios in French acute care hospitals? *BMC Health Serv Res* **2017**, *17*, 578, doi:10.1186/s12913-017-2534-3.
67. Brook, R.H.; McGlynn, E.A.; Shekelle, P.G. Defining and measuring quality of care: a perspective from US researchers. *Int J Qual Health Care* **2000**, *12*, 281-295, doi:10.1093/intqhc/12.4.281.
68. Gruenberg, D.A.; Shelton, W.; Rose, S.L.; Rutter, A.E.; Socaris, S.; McGee, G. Factors influencing length of stay in the intensive care unit. *Am J Crit Care* **2006**, *15*, 502-509.
69. Walther, S.M.; Jonasson, U. Outcome of the elderly critically ill after intensive care in an era of cost containment. *Acta Anaesthesiol Scand* **2004**, *48*, 417-422, doi:10.1111/j.0001-5172.2004.00355.x.
70. Zimmerman, J.E.; Alzola, C.; Von Rueden, K.T. The use of benchmarking to identify top performing critical care units: a preliminary assessment of their policies and practices. *J Crit Care* **2003**, *18*, 76-86, doi:10.1053/jcrc.2003.50005.
71. Rosenberg, A.L.; Zimmerman, J.E.; Alzola, C.; Draper, E.A.; Knaus, W.A. Intensive care unit length of stay: recent changes and future challenges. *Crit Care Med* **2000**, *28*, 3465-3473, doi:10.1097/00003246-200010000-00016.
72. Niskanen, M.; Reinikainen, M.; Pettilä, V. Case-mix-adjusted length of stay and mortality in 23 Finnish ICUs. *Intensive Care Med* **2009**, *35*, 1060-1067, doi:10.1007/s00134-008-1377-0.
73. Barchitta, M.; Quattrocchi, A.; Maugeri, A.; Rosa, M.C.; Mastra, C.; Basile, G.; Giuffrida, G.; Rinaldi, F.M.; Murolo, G.; Agodi, A. The "Obiettivo Antibiotico" Campaign on Prudent Use of Antibiotics in Sicily, Italy: The Pilot Phase. *Int J Environ Res Public Health* **2020**, *17*, doi:10.3390/ijerph17093077.
74. Furmenti, M.F.; Rossello, P.; Bianco, S.; Olivero, E.; Thomas, R.; Emelurumonye, I.N.; Zotti, C.M.; Group, H.I.C. Healthcare-associated infections and antimicrobial use in long-term care facilities (HALT3): an overview of the Italian situation. *J Hosp Infect* **2019**, *102*, 425-430, doi:10.1016/j.jhin.2019.02.007.

75. Tardivo, S.; Moretti, F.; Agodi, A.; Appignanesi, R.; Baldovin, T.; Barchitta, M.; Brusaferro, S.; Canino, R.; Carli, A.; D'Errico, M.M.; et al. Essential strategies in HAI prevention and control: performance assessment through the implementation of the HAI-CoSIP tool of the GISIO-SItI group. A pilot study in a sample of Italian Organizations. *Ann Ig* **2018**, *30*, 70-85, doi:10.7416/ai.2018.225.
76. Brusaferro, S.; Arnoldo, L.; Finzi, G.; Mura, I.; Auxilia, F.; Pasquarella, C.; Agodi, A.; Board; Group. Hospital Hygiene and Infection Prevention and Control in Italy: state of the art and perspectives. *Ann Ig* **2018**, *30*, 1-6, doi:10.7416/ai.2018.2245.
77. Tardivo, S.; Moretti, F.; Nobile, M.; Agodi, A.; Appignanesi, R.; Arrigoni, C.; Baldovin, T.; Brusaferro, S.; Canino, R.; Carli, A.; et al. Definition of criteria and indicators for the prevention of Healthcare-Associated Infections (HAIs) in hospitals for the purposes of Italian institutional accreditation and performance monitoring. *Ann Ig* **2017**, *29*, 529-547, doi:10.7416/ai.2017.2183.
78. Cicchetti, A.; Gasbarrini, A. The healthcare service in Italy: regional variability. *Eur Rev Med Pharmacol Sci* **2016**, *20*, 1-3.
79. Mangano, A. An analysis of the regional differences in health care utilization in Italy. *Health Place* **2010**, *16*, 301-308, doi:10.1016/j.healthplace.2009.10.013.
80. Bisaso, K.R.; Anguzu, G.T.; Karungi, S.A.; Kiragga, A.; Castelnuovo, B. A survey of machine learning applications in HIV clinical research and care. *Comput Biol Med* **2017**, *91*, 366-371, doi:10.1016/j.compbiomed.2017.11.001.
81. Dhillon, A.; Singh, A. Machine Learning in Healthcare Data Analysis: A Survey. **2019**, doi:10.15412/J.JBTW.01070206
82. Bennett, E.E.; VanBuren, J.; Holubkov, R.; Bratton, S.L. Presence of Invasive Devices and Risks of Healthcare-Associated Infections and Sepsis. *J Pediatr Intensive Care* **2018**, *7*, 188-195, doi:10.1055/s-0038-1656535.
83. Miller, M.R.; Griswold, M.; Harris, J.M.; Yenokyan, G.; Huskins, W.C.; Moss, M.; Rice, T.B.; Ridling, D.; Campbell, D.; Margolis, P.; et al. Decreasing PICU catheter-associated bloodstream infections: NACHRI's quality transformation efforts. *Pediatrics* **2010**, *125*, 206-213, doi:10.1542/peds.2009-1382.
84. Blot, S.I.; Peleman, R.; Vandewoude, K.H. Invasive devices: no need? No use! *Intensive Care Med* **2007**, *33*, 209-211, doi:10.1007/s00134-006-0465-2.
85. van der Kooi, T.I.; de Boer, A.S.; Manniën, J.; Wille, J.C.; Beaumont, M.T.; Mooi, B.W.; van den Hof, S. Incidence and risk factors of device-associated infections and associated mortality at the intensive care in the Dutch surveillance system. *Intensive Care Med* **2007**, *33*, 271-278, doi:10.1007/s00134-006-0464-3.
86. Safdar, N.; Dezfulian, C.; Collard, H.R.; Saint, S. Clinical and economic consequences of ventilator-associated pneumonia: a systematic review. *Crit Care Med* **2005**, *33*, 2184-2193, doi:10.1097/01.ccm.0000181731.53912.d9.

87. Myny, D.; Depuydt, P.; Colardyn, F.; Blot, S. Ventilator-associated pneumonia in a tertiary care ICU: analysis of risk factors for acquisition and mortality. *Acta Clin Belg* **2005**, *60*, 114-121, doi:10.1179/acb.2005.022.
88. WHO. Antimicrobial resistance. Available online: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance (accessed on
89. Serra-Burriel, M.; Keys, M.; Campillo-Artero, C.; Agodi, A.; Barchitta, M.; Gikas, A.; Palos, C.; López-Casasnovas, G. Impact of multi-drug resistant bacteria on economic and clinical outcomes of healthcare-associated infections in adults: Systematic review and meta-analysis. *PLoS One* **2020**, *15*, e0227139, doi:10.1371/journal.pone.0227139.
90. Magiorakos, A.P.; Srinivasan, A.; Carey, R.B.; Carmeli, Y.; Falagas, M.E.; Giske, C.G.; Harbarth, S.; Hindler, J.F.; Kahlmeter, G.; Olsson-Liljequist, B.; et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect* **2012**, *18*, 268-281, doi:10.1111/j.1469-0691.2011.03570.x.
91. Cosgrove, S.E.; Carmeli, Y. The impact of antimicrobial resistance on health and economic outcomes. *Clin Infect Dis* **2003**, *36*, 1433-1437, doi:10.1086/375081.
92. Friedman, N.D.; Temkin, E.; Carmeli, Y. The negative impact of antibiotic resistance. *Clin Microbiol Infect* **2016**, *22*, 416-422, doi:10.1016/j.cmi.2015.12.002.
93. Hidron, A.I.; Edwards, J.R.; Patel, J.; Horan, T.C.; Sievert, D.M.; Pollock, D.A.; Fridkin, S.K.; Team, N.H.S.N.; Facilities, P.N.H.S.N. NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006-2007. *Infect Control Hosp Epidemiol* **2008**, *29*, 996-1011, doi:10.1086/591861.
94. Livermore, D.M. Bacterial resistance: origins, epidemiology, and impact. *Clin Infect Dis* **2003**, *36*, S11-23, doi:10.1086/344654.
95. Nathwani, D.; Raman, G.; Sulham, K.; Gavaghan, M.; Menon, V. Clinical and economic consequences of hospital-acquired resistant and multidrug-resistant Pseudomonas aeruginosa infections: a systematic review and meta-analysis. *Antimicrob Resist Infect Control* **2014**, *3*, 32, doi:10.1186/2047-2994-3-32.
96. Jasovský, D.; Littmann, J.; Zorzet, A.; Cars, O. Antimicrobial resistance-a threat to the world's sustainable development. *Ups J Med Sci* **2016**, *121*, 159-164, doi:10.1080/03009734.2016.1195900.
97. Cassini, A.; Högberg, L.D.; Plachouras, D.; Quattrocchi, A.; Hoxha, A.; Simonsen, G.S.; Colomb-Cotinat, M.; Kretzschmar, M.E.; Devleesschauwer, B.; Cecchini, M.; et al. 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* **2019**, *19*, 56-66, doi:10.1016/S1473-3099(18)30605-4.
98. Drug-Resistant Infections: A Threat to Our Economic Future. **2017**

99. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations **2016**.

100. A European One Health Action Plan against Antimicrobial Resistance (AMR). **2017**.

101. Paget, J.; Lescure, D.; Versporten, A.; Goossens, H.; Schellevis, F.; van Dijk, L. Antimicrobial resistance and causes of non-prudent use of antibiotics in human medicine in the UE.

102. Charani, E.; Edwards, R.; Sevdalis, N.; Alexandrou, B.; Sibley, E.; Mullett, D.; Franklin, B.D.; Holmes, A. Behavior change strategies to influence antimicrobial prescribing in acute care: a systematic review. *Clin Infect Dis* **2011**, *53*, 651-662, doi:10.1093/cid/cir445.

103. Pinder, R.; Sallis, A.; Berry, D.; Chadborn, T. Behaviour Change and Antibiotic Prescribing in Healthcare Settings: Literature Review and Behavioural Analysis **2015**.

104. Nikaido, H. Multidrug resistance in bacteria. *Annu Rev Biochem* **2009**, *78*, 119-146, doi:10.1146/annurev.biochem.78.082907.145923.

105. Alekshun, M.N.; Levy, S.B. Molecular mechanisms of antibacterial multidrug resistance. *Cell* **2007**, *128*, 1037-1050, doi:10.1016/j.cell.2007.03.004.

106. Blair, J.M.; Webber, M.A.; Baylay, A.J.; Ogbolu, D.O.; Piddock, L.J. Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol* **2015**, *13*, 42-51, doi:10.1038/nrmicro3380.

107. Higgins, C.F. Multiple molecular mechanisms for multidrug resistance transporters. *Nature* **2007**, *446*, 749-757, doi:10.1038/nature05630.

108. Baym, M.; Stone, L.K.; Kishony, R. Multidrug evolutionary strategies to reverse antibiotic resistance. *Science* **2016**, *351*, aad3292, doi:10.1126/science.aad3292.

109. Ahmed, A.H.; Giri, J.; Kashyap, R.; Singh, B.; Dong, Y.; Kilickaya, O.; Erwin, P.J.; Murad, M.H.; Pickering, B.W. Outcome of adverse events and medical errors in the intensive care unit: a systematic review and meta-analysis. *Am J Med Qual* **2015**, *30*, 23-30, doi:10.1177/1062860613514770.

110. Liu, V.; Kipnis, P.; Rizk, N.W.; Escobar, G.J. Adverse outcomes associated with delayed intensive care unit transfers in an integrated healthcare system. *J Hosp Med* **2012**, *7*, 224-230, doi:10.1002/jhm.964.

111. Bracco, D.; Favre, J.B.; Bissonnette, B.; Wasserfallen, J.B.; Revelly, J.P.; Ravussin, P.; Chioléro, R. Human errors in a multidisciplinary intensive care unit: a 1-year prospective study. *Intensive Care Med* **2001**, *27*, 137-145, doi:10.1007/s001340000751.

112. Escobar, G.J.; Greene, J.D.; Gardner, M.N.; Marelich, G.P.; Quick, B.; Kipnis, P. Intra-hospital transfers to a higher level of care: contribution to total hospital and intensive care unit (ICU) mortality and length of stay (LOS). *J Hosp Med* **2011**, *6*, 74-80, doi:10.1002/jhm.817.

113. McCannon, C.J.; Berwick, D.M.; Massoud, M.R. The science of large-scale change in global health. *JAMA* **2007**, *298*, 1937-1939, doi:10.1001/jama.298.16.1937.

114. Noakes, T.D.; Borresen, J.; Hew-Butler, T.; Lambert, M.I.; Jordaan, E. Semmelweis and the aetiology of puerperal sepsis 160 years on: an historical review. *Epidemiol Infect* **2008**, *136*, 1-9, doi:10.1017/S0950268807008746.

115. Sydnor, E.R.; Perl, T.M. Hospital epidemiology and infection control in acute-care settings. *Clin Microbiol Rev* **2011**, *24*, 141-173, doi:10.1128/CMR.00027-10.
116. Haque, M.; Sartelli, M.; McKimm, J.; Abu Bakar, M. Health care-associated infections - an overview. *Infect Drug Resist* **2018**, *11*, 2321-2333, doi:10.2147/IDR.S177247.
117. Hughes, R.G. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. 2008.
118. Revelas, A. Healthcare - associated infections: A public health problem. *Niger Med J* **2012**, *53*, 59-64, doi:10.4103/0300-1652.103543.
119. Garrouste-Orgeas, M.; Philippart, F.; Bruel, C.; Max, A.; Lau, N.; Misset, B. Overview of medical errors and adverse events. *Ann Intensive Care* **2012**, *2*, 2, doi:10.1186/2110-5820-2-2.
120. Parameswaran Nair, N.; Chalmers, L.; Peterson, G.M.; Bereznicki, B.J.; Castelino, R.L.; Bereznicki, L.R. Hospitalization in older patients due to adverse drug reactions -the need for a prediction tool. *Clin Interv Aging* **2016**, *11*, 497-505, doi:10.2147/CIA.S99097.
121. Alp, E.; Damani, N. Healthcare-associated infections in intensive care units: epidemiology and infection control in low-to-middle income countries. *J Infect Dev Ctries* **2015**, *9*, 1040-1045, doi:10.3855/jidc.6832.
122. Vincent, J.L.; Bihari, D.J.; Suter, P.M.; Bruining, H.A.; White, J.; Nicolas-Chanoin, M.H.; Wolff, M.; Spencer, R.C.; Hemmer, M. The prevalence of nosocomial infection in intensive care units in Europe. Results of the European Prevalence of Infection in Intensive Care (EPIC) Study. EPIC International Advisory Committee. *JAMA* **1995**, *274*, 639-644.
123. Vincent, J.L.; Rello, J.; Marshall, J.; Silva, E.; Anzueto, A.; Martin, C.D.; Moreno, R.; Lipman, J.; Gomersall, C.; Sakr, Y.; et al. International study of the prevalence and outcomes of infection in intensive care units. *JAMA* **2009**, *302*, 2323-2329, doi:10.1001/jama.2009.1754.
124. Hsu, V. Prevention of health care-associated infections. *Am Fam Physician* **2014**, *90*, 377-382.
125. Malik, M.; Sreekantan Nair, A.; Illango, J.; Siddiqui, N.; Gor, R.; Fernando, R.W.; Hamid, P. The Advancement in Detecting Sepsis and Its Outcome: Usefulness of Procalcitonin in Diagnosing Sepsis and Predicting Fatal Outcomes in Patients Admitted to Intensive Care Unit. *Cureus* **2021**, *13*, e14439, doi:10.7759/cureus.14439.
126. Yee, C.R.; Narain, N.R.; Akmaev, V.R.; Vemulapalli, V. A Data-Driven Approach to Predicting Septic Shock in the Intensive Care Unit. *Biomed Inform Insights* **2019**, *11*, 1178222619885147, doi:10.1177/1178222619885147.
127. Fleischmann, C.; Scherag, A.; Adhikari, N.K.; Hartog, C.S.; Tsaganos, T.; Schlattmann, P.; Angus, D.C.; Reinhart, K.; Trialists, I.F.o.A.C. Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. *Am J Respir Crit Care Med* **2016**, *193*, 259-272, doi:10.1164/rccm.201504-0781OC.
128. Martin, G.S. Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes. *Expert Rev Anti Infect Ther* **2012**, *10*, 701-706, doi:10.1586/eri.12.50.

129. Peters, R.P.; van Agtmael, M.A.; Danner, S.A.; Savelkoul, P.H.; Vandenbroucke-Grauls, C.M. New developments in the diagnosis of bloodstream infections. *Lancet Infect Dis* **2004**, *4*, 751-760, doi:10.1016/S1473-3099(04)01205-8.
130. Becker, K.L.; Snider, R.; Nylen, E.S. Procalcitonin assay in systemic inflammation, infection, and sepsis: clinical utility and limitations. *Crit Care Med* **2008**, *36*, 941-952, doi:10.1097/CCM.0B013E318165BABB.
131. Dellinger, R.P.; Levy, M.M.; Rhodes, A.; Annane, D.; Gerlach, H.; Opal, S.M.; Sevransky, J.E.; Sprung, C.L.; Douglas, I.S.; Jaeschke, R.; et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med* **2013**, *39*, 165-228, doi:10.1007/s00134-012-2769-8.
132. Singer, M.; Deutschman, C.S.; Seymour, C.W.; Shankar-Hari, M.; Annane, D.; Bauer, M.; Bellomo, R.; Bernard, G.R.; Chiche, J.D.; Coopersmith, C.M.; et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **2016**, *315*, 801-810, doi:10.1001/jama.2016.0287.
133. Bellani, G.; Laffey, J.G.; Pham, T.; Fan, E.; Brochard, L.; Esteban, A.; Gattinoni, L.; van Haren, F.; Larsson, A.; McAuley, D.F.; et al. Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA* **2016**, *315*, 788-800, doi:10.1001/jama.2016.0291.
134. Cardoso, L.T.; Grion, C.M.; Matsuo, T.; Anami, E.H.; Kauss, I.A.; Seko, L.; Bonametti, A.M. Impact of delayed admission to intensive care units on mortality of critically ill patients: a cohort study. *Crit Care* **2011**, *15*, R28, doi:10.1186/cc9975.
135. Gayat, E.; Cariou, A.; Deye, N.; Vieillard-Baron, A.; Jaber, S.; Damoisel, C.; Lu, Q.; Monnet, X.; Rennuit, I.; Azoulay, E.; et al. Determinants of long-term outcome in ICU survivors: results from the FROG-ICU study. *Crit Care* **2018**, *22*, 8, doi:10.1186/s13054-017-1922-8.
136. Wang, W.; Zhu, S.; He, Q.; Zhang, R.; Kang, Y.; Wang, M.; Zou, K.; Zong, Z.; Sun, X. Developing a Registry of Healthcare-Associated Infections at Intensive Care Units in West China: Study Rationale and Patient Characteristics. *Clin Epidemiol* **2019**, *11*, 1035-1045, doi:10.2147/CLEP.S226935.
137. Chen, Y.C.; Lin, S.F.; Liu, C.J.; Jiang, D.D.; Yang, P.C.; Chang, S.C. Risk factors for ICU mortality in critically ill patients. *J Formos Med Assoc* **2001**, *100*, 656-661.
138. Mendez-Tellez, P.A.; Dorman, T. Predicting patient outcomes, futility, and resource utilization in the intensive care unit: the role of severity scoring systems and general outcome prediction models. *Mayo Clin Proc* **2005**, *80*, 161-163, doi:10.4065/80.2.161.
139. Abate, S.M.; Assen, S.; Yinges, M.; Basu, B. Survival and predictors of mortality among patients admitted to the intensive care units in southern Ethiopia: A multi-center cohort study. *Ann Med Surg (Lond)* **2021**, *65*, 102318, doi:10.1016/j.amsu.2021.102318.

140. Favara, G.; Riela, P.; Maugeri, A.; Barchitta, M.; Gallo, G.; Agodi, A. Risk of Pneumonia and associated outcomes in Intensive Care Unit: an integrated approach of Visual and Cluster Analysis. **2019**, doi:10.1109/SERVICES.2019.00083.
141. Kong, G.; Lin, K.; Hu, Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Med Inform Decis Mak* **2020**, *20*, 251, doi:10.1186/s12911-020-01271-2.
142. Scardoni, A.; Balzarini, F.; Signorelli, C.; Cabitza, F.; Odone, A. Artificial intelligence-based tools to control healthcare associated infections: A systematic review of the literature. *J Infect Public Health* **2020**, *13*, 1061-1077, doi:10.1016/j.jiph.2020.06.006.
143. Parreco, J.P.; Hidalgo, A.E.; Badilla, A.D.; Ilyas, O.; Rattan, R. Predicting central line-associated bloodstream infections and mortality using supervised machine learning. *J Crit Care* **2018**, *45*, 156-162, doi:10.1016/j.jcrc.2018.02.010.
144. Deo, R.C. Machine Learning in Medicine. *Circulation* **2015**, *132*, 1920-1930, doi:10.1161/CIRCULATIONAHA.115.001593.
145. Deo, R.C. Machine Learning in Medicine: Will This Time Be Different? *Circulation* **2020**, *142*, 1521-1523, doi:10.1161/CIRCULATIONAHA.120.050583.
146. Vincent, J.L.; Jones, G.; David, S.; Olariu, E.; Cadwell, K.K. Frequency and mortality of septic shock in Europe and North America: a systematic review and meta-analysis. *Crit Care* **2019**, *23*, 196, doi:10.1186/s13054-019-2478-6.
147. Yasnoff, W.A.; O'Carroll, P.W.; Koo, D.; Linkins, R.W.; Kilbourne, E.M. Public health informatics: improving and transforming public health in the information age. *J Public Health Manag Pract* **2000**, *6*, 67-75.
148. Lee, S.; Mohr, N.M.; Street, W.N.; Nadkarni, P. Machine Learning in Relation to Emergency Medicine Clinical and Operational Scenarios: An Overview. *West J Emerg Med* **2019**, *20*, 219-227, doi:10.5811/westjem.2019.1.41244.
149. Schnipper, J.L.; Linder, J.A.; Palchuk, M.B.; Einbinder, J.S.; Li, Q.; Postilnik, A.; Middleton, B. "Smart Forms" in an Electronic Medical Record: documentation-based clinical decision support to improve disease management. *J Am Med Inform Assoc* **2008**, *15*, 513-523, doi:10.1197/jamia.M2501.
150. Khoury, M.J.; Armstrong, G.L.; Bunnell, R.E.; Cyril, J.; Iademarco, M.F. The intersection of genomics and big data with public health: Opportunities for precision public health. *PLoS Med* **2020**, *17*, e1003373, doi:10.1371/journal.pmed.1003373.
151. Khoury, M.J.; Iademarco, M.F.; Riley, W.T. Precision Public Health for the Era of Precision Medicine. *Am J Prev Med* **2016**, *50*, 398-401, doi:10.1016/j.amepre.2015.08.031.
152. Big hopes for big data. *Nat Med* **2020**, *26*, 1, doi:10.1038/s41591-019-0740-8.
153. Chowkwanyun, M.; Bayer, R.; Galea, S. "Precision" Public Health - Between Novelty and Hype. *N Engl J Med* **2018**, *379*, 1398-1400, doi:10.1056/NEJMp1806634.
154. Banks, M.A. Sizing up big data. *Nat Med* **2020**, *26*, 5-6, doi:10.1038/s41591-019-0703-0.

155. Dolley, S. Big Data's Role in Precision Public Health. *Front Public Health* **2018**, *6*, 68, doi:10.3389/fpubh.2018.00068.
156. Parimbelli, E.; Marini, S.; Sacchi, L.; Bellazzi, R. Patient similarity for precision medicine: A systematic review. *J Biomed Inform* **2018**, *83*, 87-96, doi:10.1016/j.jbi.2018.06.001.
157. Jameson, J.L.; Longo, D.L. Precision medicine--personalized, problematic, and promising. *N Engl J Med* **2015**, *372*, 2229-2234, doi:10.1056/NEJMsb1503104.
158. Serra, A.; Fratello, M.; Fortino, V.; Raiconi, G.; Tagliaferri, R.; Greco, D. MVDA: a multi-view genomic data integration methodology. *BMC Bioinformatics* **2015**, *16*, 261, doi:10.1186/s12859-015-0680-3.
159. Ow, G.S.; Tang, Z.; Kuznetsov, V.A. Big data and computational biology strategy for personalized prognosis. *Oncotarget* **2016**, *7*, 40200-40220, doi:10.18632/oncotarget.9571.
160. Wang, F. Adaptive semi-supervised recursive tree partitioning: The ART towards large scale patient indexing in personalized healthcare. *J Biomed Inform* **2015**, *55*, 41-54, doi:10.1016/j.jbi.2015.01.009.
161. Pirracchio, R.; Cohen, M.J.; Malenica, I.; Cohen, J.; Chambaz, A.; Cannesson, M.; Lee, C.; Resche-Rigon, M.; Hubbard, A.; Group, A.R. Big data and targeted machine learning in action to assist medical decision in the ICU. *Anaesth Crit Care Pain Med* **2019**, *38*, 377-384, doi:10.1016/j.accpm.2018.09.008.
162. Patel, L.; Shukla, T.; Huang, X.; Ussery, D.W.; Wang, S. Machine Learning Methods in Drug Discovery. *Molecules* **2020**, *25*, doi:10.3390/molecules25225277.
163. Groft, S.C.; Posada, M.; Taruscio, D. Progress, challenges and global approaches to rare diseases. *Acta Paediatr* **2021**, *110*, 2711-2716, doi:10.1111/apa.15974.
164. Luz, C.F.; Vollmer, M.; Decruyenaere, J.; Nijsten, M.W.; Glasner, C.; Sinha, B. Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. *Clin Microbiol Infect* **2020**, *26*, 1291-1299, doi:10.1016/j.cmi.2020.02.003.
165. Adler-Milstein, J.; Holmgren, A.J.; Kralovec, P.; Worzala, C.; Searcy, T.; Patel, V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. *J Am Med Inform Assoc* **2017**, *24*, 1142-1148, doi:10.1093/jamia/ocx080.
166. Rüping, S. [Big data in medicine and healthcare]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **2015**, *58*, 794-798, doi:10.1007/s00103-015-2181-y.
167. Wiens, J.; Shenoy, E.S. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clin Infect Dis* **2018**, *66*, 149-153, doi:10.1093/cid/cix731.
168. Komorowski, M. Artificial intelligence in intensive care: are we there yet? *Intensive Care Med* **2019**, *45*, 1298-1300, doi:10.1007/s00134-019-05662-6.
169. Rajkomar, A.; Dean, J.; Kohane, I. Machine Learning in Medicine. Reply. *N Engl J Med* **2019**, *380*, 2589-2590, doi:10.1056/NEJMc1906060.

170. Kaye, K.S.; Anderson, D.J.; Cook, E.; Huang, S.S.; Siegel, J.D.; Zuckerman, J.M.; Talbot, T.R. Guidance for infection prevention and healthcare epidemiology programs: healthcare epidemiologist skills and competencies. *Infect Control Hosp Epidemiol* **2015**, *36*, 369-380, doi:10.1017/ice.2014.79.
171. Churpek, M.M.; Yuen, T.C.; Winslow, C.; Meltzer, D.O.; Kattan, M.W.; Edelson, D.P. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med* **2016**, *44*, 368-374, doi:10.1097/CCM.0000000000001571.
172. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization;. **2021**.
173. Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S.D.; Tegmark, M.; Fuso Nerini, F. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat Commun* **2020**, *11*, 233, doi:10.1038/s41467-019-14108-y.
174. Peiffer-Smadja, N.; Rawson, T.M.; Ahmad, R.; Buchard, A.; Georgiou, P.; Lescure, F.X.; Birgand, G.; Holmes, A.H. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect* **2020**, *26*, 584-595, doi:10.1016/j.cmi.2019.09.009.
175. Yuan, K.C.; Tsai, L.W.; Lee, K.H.; Cheng, Y.W.; Hsu, S.C.; Lo, Y.S.; Chen, R.J. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int J Med Inform* **2020**, *141*, 104176, doi:10.1016/j.ijmedinf.2020.104176.
176. Majnarić, L.T.; Babič, F.; O'Sullivan, S.; Holzinger, A. AI and Big Data in Healthcare: Towards a More Comprehensive Research Framework for Multimorbidity. *J Clin Med* **2021**, *10*, doi:10.3390/jcm10040766.
177. Fleming, N. How artificial intelligence is changing drug discovery. *Nature* **2018**, *557*, S55-S57, doi:10.1038/d41586-018-05267-x.
178. Fleuren, L.M.; Klausch, T.L.T.; Zwager, C.L.; Schoonmade, L.J.; Guo, T.; Roggeveen, L.F.; Swart, E.L.; Girbes, A.R.J.; Thoral, P.; Ercole, A.; et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* **2020**, *46*, 383-400, doi:10.1007/s00134-019-05872-y.
179. Beam, A.L.; Kohane, I.S. Big Data and Machine Learning in Health Care. *JAMA* **2018**, *319*, 1317-1318, doi:10.1001/jama.2017.18391.
180. Yu, K.H.; Beam, A.L.; Kohane, I.S. Artificial intelligence in healthcare. *Nat Biomed Eng* **2018**, *2*, 719-731, doi:10.1038/s41551-018-0305-z.
181. Goodswen, S.J.; Barratt, J.L.N.; Kennedy, P.J.; Kaufer, A.; Calarco, L.; Ellis, J.T. Machine learning and applications in microbiology. *FEMS Microbiol Rev* **2021**, *45*, doi:10.1093/femsre/fuab015.
182. Hinton, G. Deep Learning-A Technology With the Potential to Transform Health Care. *JAMA* **2018**, *320*, 1101-1102, doi:10.1001/jama.2018.11100.

183. Chen, L.; Dubrawski, A.; Wang, D.; Fiterau, M.; Guillame-Bert, M.; Bose, E.; Kaynar, A.M.; Wallace, D.J.; Guttendorf, J.; Clermont, G.; et al. Using Supervised Machine Learning to Classify Real Alerts and Artifact in Online Multisignal Vital Sign Monitoring Data. *Crit Care Med* **2016**, *44*, e456-463, doi:10.1097/CCM.0000000000001660.

184. Linnen, D.T.; Escobar, G.J.; Hu, X.; Scruth, E.; Liu, V.; Stephens, C. Statistical Modeling and Aggregate-Weighted Scoring Systems in Prediction of Mortality and ICU Transfer: A Systematic Review. *J Hosp Med* **2019**, *14*, 161-169, doi:10.12788/jhm.3151.

185. Fitzpatrick, F.; Doherty, A.; Lacey, G. Using Artificial Intelligence in Infection Prevention. *Curr Treat Options Infect Dis* **2020**, 1-10, doi:10.1007/s40506-020-00216-7.

186. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput Sci* **2021**, *2*, 160, doi:10.1007/s42979-021-00592-x.

187. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436-444, doi:10.1038/nature14539.

188. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Healthc J* **2019**, *6*, 94-98, doi:10.7861/futurehosp.6-2-94.

189. Naylor, C.D. On the Prospects for a (Deep) Learning Health Care System. *JAMA* **2018**, *320*, 1099-1100, doi:10.1001/jama.2018.11103.

190. Agodi, A.; Barchitta, M.; Quattrocchi, A.; Spera, E.; Gallo, G.; Auxilia, F.; Brusaferro, S.; D'Errico, M.M.; Montagna, M.T.; Pasquarella, C.; et al. Preventable proportion of intubation-associated pneumonia: Role of adherence to a care bundle. *PLoS One* **2017**, *12*, e0181170, doi:10.1371/journal.pone.0181170.

191. Agodi, A.; Auxilia, F.; Barchitta, M.; Brusaferro, S.; D'Errico, M.M.; Montagna, M.T.; Pasquarella, C.; Tardivo, S.; Mura, I.; SPIN-UTI network of the GISIOWorking Group of the Italian Society of Hygiene, P.e.M.a.P.H.S. Antibiotic consumption and resistance: results of the SPIN-UTI project of the GISIO-SItI. *Epidemiol Prev* **2015**, *39*, 94-98.

192. Agodi, A.; Auxilia, F.; Barchitta, M.; Brusaferro, S.; D'Alessandro, D.; Grillo, O.C.; Montagna, M.T.; Pasquarella, C.; Righi, E.; Tardivo, S.; et al. Trends, risk factors and outcomes of healthcare-associated infections within the Italian network SPIN-UTI. *J Hosp Infect* **2013**, *84*, 52-58, doi:10.1016/j.jhin.2013.02.012.

193. Agodi, A.; Auxilia, F.; Barchitta, M.; D'Errico, M.M.; Montagna, M.T.; Pasquarella, C.; Tardivo, S.; Mura, I.; GISIO-SItI. [Control of intubator associated pneumonia in intensive care unit: results of the GISIO-SItI SPIN-UTI Project]. *Epidemiol Prev* **2014**, *38*, 51-56.

194. Agodi, A.; Barchitta, M.; Mura, I.; Pasquarella, C.; Torregrossa, M.V.; SItI, G. The commitment of the GISIO-SItI to contrast Healthcare-Associated Infections and the experience of prevalence studies in Sicily. *Ann Ig* **2018**, *30*, 38-47, doi:10.7416/ai.2018.2233.

195. Michard, F.; Teboul, J.L. Predictive analytics: beyond the buzz. *Ann Intensive Care* **2019**, *9*, 46, doi:10.1186/s13613-019-0524-9.

196. Barchitta, M.; Maugeri, A.; Favara, G.; Riela, P.M.; La Mastra, C.; La Rosa, M.C.; San Lio, R.M.; Gallo, G.; Mura, I.; Agodi, A.; et al. Cluster analysis identifies patients at risk of catheter-associated urinary tract infections in intensive care units: findings from the SPIN-UTI Network. *J Hosp Infect* **2021**, *107*, 57-63, doi:10.1016/j.jhin.2020.09.030.

197. Castela Forte, J.; Perner, A.; van der Horst, I.C.C. The use of clustering algorithms in critical care research to unravel patient heterogeneity. *Intensive Care Med* **2019**, *45*, 1025-1028, doi:10.1007/s00134-019-05631-z.

198. Vranas, K.C.; Jopling, J.K.; Sweeney, T.E.; Ramsey, M.C.; Milstein, A.S.; Slatore, C.G.; Escobar, G.J.; Liu, V.X. Identifying Distinct Subgroups of ICU Patients: A Machine Learning Approach. *Crit Care Med* **2017**, *45*, 1607-1615, doi:10.1097/CCM.0000000000002548.

199. ECDC. European surveillance of healthcare-associated infections in intensive care units- HAI-Net ICU protocol- Protocol version 1.02. **2015**.

200. ECDC. European Centre for Disease Prevention and Control. European surveillance of healthcare-associated infections in intensive care units. ECDC HAIICU protocol V1.01 Standard and Light. Stockholm. **2010**.

201. Masia, M.D.; Barchitta, M.; Liperi, G.; Cantù, A.P.; Alliata, E.; Auxilia, F.; Torregrossa, V.; Mura, I.; Agodi, A.; (GISIO), I.S.G.o.H.H. Validation of intensive care unit-acquired infection surveillance in the Italian SPIN-UTI network. *J Hosp Infect* **2010**, *76*, 139-142, doi:10.1016/j.jhin.2010.05.013.

202. Barchitta, M.; Maugeri, A.; Favara, G.; Riela, P.M.; Gallo, G.; Mura, I.; Agodi, A.; network, S.-U. A machine learning approach to predict healthcare-associated infections at intensive care unit admission: findings from the SPIN-UTI project. *J Hosp Infect* **2021**, doi:10.1016/j.jhin.2021.02.025.

203. Barchitta, M.; Maugeri, A.; Favara, G.; Riela, P.M.; Gallo, G.; Mura, I.; Agodi, A. Early Prediction of Seven-Day Mortality in Intensive Care Unit Using a Machine Learning Model: Results from the SPIN-UTI Project. *J Clin Med* **2021**, *10*, doi:10.3390/jcm10050992.

204. Czygan, M.; Phuong, V. *Getting started with Python data analysis*; 2015.

205. Horan, T.C.; Emori, T.G. Definitions of key terms used in the NNIS System. *Am J Infect Control* **1997**, *25*, 112-116, doi:10.1016/s0196-6553(97)90037-7.

206. Devlin, U.M.; McNulty, B.A.; Nugent, A.P.; Gibney, M.J. The use of cluster analysis to derive dietary patterns: methodological considerations, reproducibility, validity and the effect of energy mis-reporting. *Proc Nutr Soc* **2012**, *71*, 599-609, doi:10.1017/S0029665112000729.

207. Wongsuphasawat, K.; Gotz, D. Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization. *IEEE Trans Vis Comput Graph* **2012**, *18*, 2659-2668, doi:10.1109/TVCG.2012.225.

208. Cortes, C.; Vapnik, V. Support-Vector Networks.

209. Martos-Benítez, F.D.; Larrondo-Muguercia, H.; León-Pérez, D.; Rivero-López, J.C.; Orama-Requejo, V.; Martínez-Alfonso, J.L. Performance of three prognostic models in critically ill patients with cancer: a prospective study. *Int J Clin Oncol* **2020**, doi:10.1007/s10147-020-01659-0.

210. D'Arrigo, G.; Provenzano, F.; Torino, C.; Zoccali, C.; Tripepi, G. I test diagnostici e l'analisi della curva ROC.
211. Cortes, C.; Vapnik, V. Support-Vector Networks. **1995**.
212. Han, S.; Qubo, C.; Meng, H. Parameter selection in SVM with RBF kernel function. **2012**.
213. Malarvizhi, R.; Thanamani, A. K-nearest neighbor in missing data imputation. **2012**.
214. Obadia, Y. The use of KNN for missing values. **2017**.
215. Japkowicz, N. The Class Imbalance Problem: Significance and Strategies. **2000**.
216. Chawla, N.; Bowyer, K.; Hall, L.; Kegelmeyer, W. SMOTE: Synthetic Minority Over-sampling Technique.
217. Tan, X.; Rolls, K.; Wiseman, T.; Betihavas, V. Risk factors for Healthcare Associated Infections (HAI) or sepsis in trauma patients : an integrative literature review.
218. Peiffer-Smadja, N.; Rawson, T.M.; Ahmad, R.; Buchard, A.; Georgiou, P.; Lescure, F.X.; Birgand, G.; Holmes, A.H. Corrigendum to 'machine learning for clinical decision support in infectious diseases: a narrative review of current applications' clinical microbiology and infection (2020) 584-595. *Clin Microbiol Infect* **2020**, *26*, 1118, doi:10.1016/j.cmi.2020.05.020.
219. Stamm, W.E.; Hooton, T.M.; Johnson, J.R.; Johnson, C.; Stapleton, A.; Roberts, P.L.; Moseley, S.L.; Fihn, S.D. Urinary tract infections: from pathogenesis to treatment. *J Infect Dis* **1989**, *159*, 400-406.
220. Chenoweth, C.; Saint, S. Preventing catheter-associated urinary tract infections in the intensive care unit. *Crit Care Clin* **2013**, *29*, 19-32, doi:10.1016/j.ccc.2012.10.005.
221. Nicolle, L.E. Catheter associated urinary tract infections. *Antimicrob Resist Infect Control* **2014**, *3*, 23, doi:10.1186/2047-2994-3-23.
222. Nicolle, L.E. Urinary catheter-associated infections. *Infect Dis Clin North Am* **2012**, *26*, 13-27, doi:10.1016/j.idc.2011.09.009.
223. Meddings, J.; Saint, S. Disrupting the life cycle of the urinary catheter. *Clin Infect Dis* **2011**, *52*, 1291-1293, doi:10.1093/cid/cir195.
224. Tenke, P.; Kovacs, B.; Bjerklund Johansen, T.E.; Matsumoto, T.; Tambyah, P.A.; Naber, K.G. European and Asian guidelines on management and prevention of catheter-associated urinary tract infections. *Int J Antimicrob Agents* **2008**, *31 Suppl 1*, S68-78, doi:10.1016/j.ijantimicag.2007.07.033.
225. Foxman, B. Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. *Infect Dis Clin North Am* **2014**, *28*, 1-13, doi:10.1016/j.idc.2013.09.003.
226. Lovejoy, C.A.; Buch, V.; Maruthappu, M. Artificial intelligence in the intensive care unit. *Crit Care* **2019**, *23*, 7, doi:10.1186/s13054-018-2301-9.
227. Meiring, C.; Dixit, A.; Harris, S.; MacCallum, N.S.; Brealey, D.A.; Watkinson, P.J.; Jones, A.; Ashworth, S.; Beale, R.; Brett, S.J.; et al. Optimal intensive care outcome prediction over time using machine learning. *PLoS One* **2018**, *13*, e0206862, doi:10.1371/journal.pone.0206862.
228. Böcker, A.; Derksen, S.; Schmidt, E.; Teckentrup, A.; Schneider, G. A hierarchical clustering approach for large compound libraries. *J Chem Inf Model* **2005**, *45*, 807-815, doi:10.1021/ci0500029.

229. Varshavsky, R.; Horn, D.; Linial, M. Global considerations in hierarchical clustering reveal meaningful patterns in data. *PLoS One* **2008**, *3*, e2247, doi:10.1371/journal.pone.0002247.
230. Desautels, T.; Calvert, J.; Hoffman, J.; Jay, M.; Kerem, Y.; Shieh, L.; Shimabukuro, D.; Chettipally, U.; Feldman, M.D.; Barton, C.; et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med Inform* **2016**, *4*, e28, doi:10.2196/medinform.5909.
231. The Lancet Respiratory Medicine. Opening the black box of machine learning. *Lancet Respir Med* **2018**, *6*, 801, doi:10.1016/S2213-2600(18)30425-9.
232. Maugeri, A.; Barchitta, M.; Agodi, A. A Clustering Approach to Classify Italian Regions and Provinces Based on Prevalence and Trend of SARS-CoV-2 Cases. *Int J Environ Res Public Health* **2020**, *17*, doi:10.3390/ijerph17155286.
233. Maugeri, A.; Barchitta, M.; Battiato, S.; Agodi, A. Modeling the Novel Coronavirus (SARS-CoV-2) Outbreak in Sicily, Italy. *Int J Environ Res Public Health* **2020**, *17*, doi:10.3390/ijerph17144964.
234. Maugeri, A.; Barchitta, M.; Battiato, S.; Agodi, A. Estimation of Unreported Novel Coronavirus (SARS-CoV-2) Infections from Reported Deaths: A Susceptible-Exposed-Infectious-Recovered-Dead Model. *J Clin Med* **2020**, *9*, doi:10.3390/jcm9051350.
235. Maugeri, A.; Barchitta, M.; Battiato, S.; Agodi, A. Estimation of unreported SARS-CoV-2 cases in Italy using a Susceptible-Exposed-Infectious-Recovered-Dead model. *J Glob Health* **2020**, *10*, 021105, doi:10.7189/jogh.10.021105.
236. Ripoli, A.; Sozio, E.; Sbrana, F.; Bertolino, G.; Pallotto, C.; Cardinali, G.; Meini, S.; Pieralli, F.; Azzini, A.M.; Concia, E.; et al. Personalized machine learning approach to predict candidemia in medical wards. *Infection* **2020**, doi:10.1007/s15010-020-01488-3.
237. Gastmeier, P. From 'one size fits all' to personalized infection prevention. *J Hosp Infect* **2020**, *104*, 256-260, doi:10.1016/j.jhin.2019.12.010.
238. Jensen, S.O.; van Hal, S.J. Personalized Medicine and Infectious Disease Management. *Trends Microbiol* **2017**, *25*, 875-876, doi:10.1016/j.tim.2017.09.006.