Universidad Rey Juan Carlos

TESIS DOCTORAL

Deep Learning and Graph Analytics for Explainable Modeling of Clinical Time-Varying Data Associated with Infectious Diseases

Autor

Sergio Martínez Agüero

Directores

Cristina Soguero Ruiz

Programa de Doctorado en Tecnologías de la Información y las Comunicaciones

Escuela Internacional de Doctorado

2024

©2024 Sergio Martínez Agüero
Algunos derechos reservados
Este documento se distribuye bajo la licencia
“Atribución 4.0 Internacional” de Creative Commons,
disponible en https://creativecommons.org/licenses/by/4.0/deed.es

This work has been partly supported by the Institute of Health Carlos III, Spain (grant DTS 17/00158), by the Spanish Ministry of Economy under the grants with reference TEC2016-75361-R, PID2019-107768RA-I00 and PID2019-106623RB-C41, by Project Ref. F656 financed by Rey Juan Carlos University, by the Young Researchers R&D Project Ref. 2020-661 financed by Rey Juan Carlos University and Community of Madrid, and by the Youth Employment Initiative (YEI) R&D Project Ref. TIC-11649 financed by the Community of Madrid (Spain).

A journey will have pain and failure. It is not only the steps forward that we must accept. It is the stumbles. The trials. The knowledge that we will fail.

But if we stop, if we accept the person we are when we fail, the journey ends. That failure becomes our destination. To love the journey is to accept no such end. I have found, through painful experience, that the most important step a person can take is always the next one.

– Brandon Sanderson

Acknowledgements

Firstly, I would like to thank the University Hospital of Fuenlabrada for giving me the opportunity to work with such an extensive dataset, which has undoubtedly allowed me to learn a lot. I especially want to thank Joaquín Álvarez, the head of the Intensive Care Unit, for guiding me at all times.

This project is the final goal of four years, with many ups and downs in which I have grown not only as a professional but also as a person. One never knows how certain projects end up changing you and making fate take you even to the end of the world. The Higher Technical School of Telecommunication Engineering and its members are blamed for this. Thanks to the teachers there for always being there when I needed them and for always answering all my questions with a smile. Especially to my thesis director, Cristina Soguero, for always being concerned about me.

I could not forget my kids, who know I've been discussing the same thing for months. Thanks especially to those of you who care about me every time I have a problem. I am very lucky to be able to count on you.

Of course, thanks to my parents and my brother. You are the biggest pillar of my life and an endless source of support. You have always been there, cheering me on and showing your love every time I worry.

THANK YOU ALL! Because I went to Madrid as an irresponsible, scared, and shy kid and I have ended up being a man proud of who he is and what he has achieved.

Resumen

Esta tesis doctoral realiza una investigación de las herramientas de ciencia de datos para abordar dos problemas incipientes en entornos clínicos modernos: la Multiresistencia Antimicrobiana (MRA) y el Coronavirus 2019 (COVID-19). Al concentrarse en estas dos áreas, la investigación aborda problemas de salud pública urgentes y de gran impacto. La MRA representa un desafío creciente en la medicina global, con patógenos que se vuelven resistentes a los tratamientos antimicrobianos convencionales, lo que complica el manejo de infecciones y aumenta la mortalidad. Por otro lado, el COVID-19, es una pandemia que ha afectado a millones de personas en todo el mundo, y sigue presentando desafíos en su control y en la comprensión de sus patrones de propagación y efectos a largo plazo. La investigación se centra en las Unidades de Cuidados Intensivos (UCI), identificadas como epicentros críticos para la adquisición de enfermedades infecciosas. Este enfoque es fundamental debido a la alta vulnerabilidad de los pacientes en las UCI, quienes a menudo presentan sistemas inmunológicos debilitados y están expuestos a una variedad de procedimientos invasivos que aumentan el riesgo de infección.

Inicialmente, la tesis aborda un análisis detallado sobre la importancia de la MRA y el COVID-19, explorando el impacto de estos en aspectos sociales, económicos y en los sistemas de salud a nivel global y nacional. Se enfatiza la amenaza que suponen la evolución de amenazas existentes, ejemplificada por el incremento de la MRA y de nuevas patologías como el COVID-19. Al analizar las complejas repercusiones demográficas y económicas de estas amenazas sanitarias globales, la investigación subraya la urgencia de adoptar metodologías avanzadas de análisis de datos, indispensables para gestionar e interpretar la naturaleza compleja de los datos clínicos, crucial para entender la progresión de estas enfermedades en las UCIs.

El objetivo principal de esta tesis es el diseño de modelos basados en aprendizaje profundo y análisis de grafos para predecir la aparición de enfermedades infecciosas y extraer conocimiento clínico. Para llevar a cabo este objetivo, se persigue como primer objetivo específico la creación de bases de datos para la realización de esta tesis doctoral, implementando diversas técnicas de modelado e integración de datos. La siguiente contribución específica de esta tesis es el desarrollo de modelos de datos multimodales interpretables, específicamente diseñados para la predicción temprana de la MRA. Estos modelos integran datos clínicos tanto estáticos como dinámicos, y han sido específicamente diseñados para conseguir buenas prestaciones, sin comprometer la interpretabilidad, aspecto clave en entornos clínicos. Se logra este equilibrio mediante un modelado adecuado de los datos, incorporando una fase previa de selección de características y mecanismos específicos para mejorar la interpretabilidad de los modelos. Los resultados obtenidos han mostrado como utilizando modelos multimodales basados en series

temporales pueden mejorar las prestaciones obtenidas para predecir la aparición de MRA en la ICU. Además, se han identificado aspectos claves en la aparición de gérmenes multirresistentes como la ventilación mecánica, el número de vecinos MRA o ciertas familias de antibióticos. Estos modelos no solo son útiles en la detección y manejo temprano de la MRA, sino que también podrían servir para futuras aplicaciones en otros problemas clínicos.

Finalmente, esta tesis introduce modelos basados en grafos para extraer conocimiento sobre el COVID-19. Estos modelos de grafos, que representan los datos clínicos como estructuras nodales interconectadas, son aplicados para descrubir y analizar las complejas relaciones e interdependencias presentes en los datos clínicos. Empleando técnicas sofisticadas de análisis de redes, como el análisis de centralidad para identificar nodos clave, estos modelos ofrecen insights profundos y detallados sobre las dinámicas de transmisión, los patrones de morbilidad y las respuestas a los tratamientos del COVID-19. Mediante el uso de representaciones basadas en grafos, este estudio proporciona una perspectiva innovadora en la visualización y análisis de la interconexión de variables clínicas, revelando patrones y asociaciones complejas. En concreto, se ha identificado la prevalencia de comorbilidades como hipertensión, diabetes u obesidad entre otras, así como fiebre o tos como síntomas predominantes. Además algunos tratamientos como la combinación de lopinavir/ritonavir, hydroxycloroquina, y corticosteroides fueron identificados como tratamientos frecuentes durante la primera ola del COVID-19.

En conclusion, esta tesis doctoral constituye un trabajo pionero que aplica la ciencia de datos en el campo de la epidemiología, ofreciendo métodos novedosos y efectivos para el análisis y la predicción de enfermedades infecciosas. A través de una metodología que incorpora aprendizaje profundo y análisis de grafos, se ha logrado no solo la creación de modelos predictivos con buenas prestaciones, sino también una comprensión más profunda de estas enfermedades, que además pueden ser aplicables en otros casos de uso. La integración de tecnologías de ciencia de datos en el cuidado de la salud, como se demuestra en esta investigación muestra un prometedor futuro en la mejora continua de los sistemas de salud global.

Abstract

This doctoral thesis conducts an investigation of data science tools to address two emerging problems in modern clinical environments: Antimicrobial Multidrug Resistance (AMR) and Coronavirus Disease 2019 (COVID-19). The research addresses urgent and high-impact public health issues by focusing on these two areas. AMR represents a growing challenge in global medicine, with pathogens becoming resistant to conventional antimicrobial treatments, complicating infection management and increasing mortality. On the other hand, COVID-19 is a pandemic that has affected millions of people around the world and continues to present challenges in its control and understanding of its patterns of spread and long-term effects. The research focuses on Intensive Care Units (ICUs), which are identified as critical epicenters for the acquisition of infectious diseases. This approach is fundamental due to the high vulnerability of patients in ICUs, who often have weakened immune systems and are exposed to a variety of invasive procedures that increase the risk of infection.

Initially, the thesis provides a detailed analysis of the importance of AMR and COVID-19, exploring their impact on social, economic, and health systems at both global and national levels. The thesis emphasizes the threat posed by the evolution of existing threats, exemplified by the increase in AMR and new pathologies such as COVID-19. By analyzing the complex demographic and economic repercussions of these global health threats, the research underscores the urgency of adopting advanced data analysis methodologies, indispensable for managing and interpreting the complex nature of clinical data and crucial for understanding the progression of these diseases in ICUs.

The main objective of the research is to construct models based on deep learning to predict the onset of infectious diseases and to extract knowledge through graph analysis. To accomplish this goal, the initial step involves developing comprehensive databases designed to carry out this dissertation, developing various data modeling and integration techniques. The following specific contribution of this thesis is the development of interpretable multimodal data models specifically designed for the early prediction of AMR. These models integrate static and dynamic clinical data and have been specifically designed to achieve high performance without compromising interpretability, a key aspect in clinical environments. This balance is achieved through meticulous data modeling, incorporating an initial feature selection phase and specific mechanisms to enhance the interpretability of the models. The results have shown that using multimodal models based on time series can improve the performance obtained in predicting the occurrence of AMR in the ICU. Furthermore, key aspects in the emergence of multidrug-resistant germs have been identified, such as mechanical ventilation, the number of

AMR neighbors, and certain families of antibiotics. These models are useful in the early detection and management of AMR and could also serve for future applications in other clinical problems.

Finally, this thesis introduces graph-based models to extract knowledge about COVID-19. These graph models, which represent clinical data as interconnected nodal structures, are applied to decipher the complex relationships and interdependencies present in clinical data. Employing sophisticated network analysis techniques, such as centrality analysis, to identify critical nodes, these models offer deep and detailed insights into the dynamics of transmission, morbidity patterns, and treatment responses of COVID-19. Through the use of graph-based representations, this study provides an innovative perspective on the visualization and analysis of the interconnection of clinical variables, revealing complex patterns and associations. Specifically, the prevalence of comorbidities such as hypertension, diabetes, or obesity, among others, as well as fever or cough as predominant symptoms have been identified. Additionally, some treatments, such as the combination of lopinavir/ritonavir, hydroxychloroquine, and corticosteroids, were identified as common treatments during the first wave of COVID-19.

In conclusion, this doctoral thesis constitutes pioneering work that applies data science in the field of epidemiology, offering novel and effective methods for analyzing and predicting infectious diseases. Through a rigorous methodology incorporating deep learning and graph analysis, a deeper understanding of these diseases has been achieved, as well as the creation of predictive models applicable in other use cases. As demonstrated in this research, the integration of data science technologies in healthcare shows a promising future in the continuous improvement of global health systems.

List of acronyms and abbreviations

AMG Aminoglycosides

AMR Antimicrobial Multidrug Resistance

APACHE-II Acute Physiology and Chronic Health Evaluation II

ATF Antifungals

BBCE Balanced Binary Cross-Entropy

BCE Binary Cross-Entropy

Bi-LSTM Bidirectional Long Short-Term Memory

CF1 Cephalosporins 1th Generation

CF2 Cephalosporins 2th Generation

CF3 Cephalosporins 3th Generation

CF4 Cephalosporins 4th Generation

CIB Confidence Intervals with Bootstrapping

CI Confidence Interval

CMI Conditional Mutual Information

COVID-19 Coronavirus Disease 2019

DL Deep Learning

DNN Deep Neural Networks

EHR Electronic Health Record

FHSI First Hidden State Initializer

FS Feature Selection

GCC Glycyclines

GLASSO Group Least Absolute Shrinkage Selection Operator

GLI Glycopeptides

GRN Gated Residual Network

GRU Gated Recurrent Unit

HAM Hadamard Attention Matrix

HIV Human Immunodeficiency Virus

ICD-9 International Classification of Diseases, Ninth Revision

ICU Intensive Care Unit

JHF Joint Heterogeneous Fusioner

LASSO Least Absolute Shrinkage Selection Operator

LFCO Late Fusion Convex Optimization

LFLR Late Fusion Logistic Regression

LIN Lincosamides

LIP Lipopeptides

LR Logistic Regression

LSTM Long Short-Term Memory

MAC Macrolides

MLP Multilayer Perceptron

ML Machine Learning

MON Monobactamas

MTS Multivariate Time Series

MV Mechanical Ventilation

NLHA Non-Linear Hadamard Attention

NN Neural Networks

NTI Nitroimidazolics

OTR Miscellaneous

OXA Oxazolidinones

PAP Broad-Spectrum Penicillins

PEN Penicillins

PFI Permutation Feature Importance

POL Polypeptides

QUI Quinolones

RNN Recurrent Neural Networks

SAPS-3 Simplified Acute Physiology Score

SE Static Encoder

SHAP SHapley Additive exPlanations

SUL Sulfamides

TFT Temporal Fusion Transformer

TPI Time Perturbation Importances

TTC Tetracyclines

UHF University Hospital of Fuenlabrada

Notation

I	Total number of patients.
i	Index to identify a specific patient, varying from 1 to I.
$\mathbf{z}_i$	Vector representing the static featuers for the i-the patient.
G	Number of features in the $\mathbf{z}_i$ vector.
z_i^g	Scalar value for the g-th feature and the i-th patient.
$\mathbf{X}_i$	MTS matrix for the i-th patient.
D	Number of MTS in $\mathbf{X}_i$.
T_i	Length of every MTS in $\mathbf{X}_i$.
$\bar{\mathbf{x}}_i^t$	Vector representing the D MTS in the time step t-th.
$x_i^{(t,d)}$	Scalar for the MTS d-th in the time step t-th for the i-th patient.
$\underline{\mathbf{x}}_i^d$	Vector representing the d-th MTS for the patient i.
y_i	The true label for the i-th patient.
$\hat{y}_i$	The prediction generated by the ML model for the ith patient.
R	Number of replacements in the bootstrap resampling algorithm.
S_d y S_{nd}	Set of deceased and non-deceased patients.
$\{S_d^{(r)}\}_{r=1}^R$	Resampled sets of positive patients, for each resampling r and S_{nd}.
μ_d	Mean of the variable in the positive patient population S_d.
$\{\mu_d^{(r)}\}_{r=1}^R$	Mean values calculated for each resampling of the S_d patients.
ΔP	Difference in means between the two populations, calculated as $\Delta P = \mu_d - \mu_{nd}$.
$\Delta P^{(r)}$	Difference in mean values for the positive and negative groups for each resampling.
$\mathrm{CI}_{\Delta P}$	Confidence Interval for the mean difference (ΔP).
H_0 y H_1	Null and Alternative Hypotheses.
$\mathscr{G}$	Graph.
$\mathscr{N}$, $\mathscr{E}$	Node set $n_1, ..., n_I$, and edge set of the graph.
$\mathbf{A}$	Adjacency matrix.
$A_{i,j}$	Matrix entry, relationship between the nodes n_i y n_j.
d_i	Normalized weighted degree of node.
$\eta(\mathscr{G})$	Graph edge density.
$H(\mathscr{G})$	Graph edge entropy.
$\mathbf{x}_i$	Mean statistics vector for a patient in $\mathbb{R}^I$.
K	Number of klusters in K-means.
$\mathbf{f}_i$	Vector to test the graph smoothness.

w^g, b	Weights and bias of the MLP architecture.
$J(w,b)$	Cost function.
$\nabla J(w,b)$	Gradient of the cost function.
η	Learning rate.
$\mathbf{h}^t$	Hidden state of the RNNs
. $\mathbf{w}_f^{(t,D)}$ y b_f	Weights and bias of the LSTM's forget layer.
$\mathbf{w}_i^{(t,D)}$ y b_i	Weights and bias of the LSTM's input layer.
i^t	Output of the LSTM's input gate layer.
$\mathbf{o}^t$	Output of the LSTM's output gate layer.
$\mathbf{z}^t$	Output of the GRU's update gate layer.
$\tilde{r}^t$	Output of the GRU's reset gate layer.
X	A specific random variable.
$\mathscr{X}$	Set of values of X.
$p(x)$	is $Pr\{X = x\}$
$\mathbb{H}(X)$	Shannon entropy for the random variable X.
$\mathbb{I}(X)$	Mutual Information for the random variable X.
$\mathbb{I}(X,Y\|Z)$	Conditional Mutual Information for the random variable X and Y given Z.
$\mathscr{D}'$	Subset of Features Selected.
α	Coefficient vector in feature space optimized by LASSO.
α^d	Coefficient for I samples of feature d.
$\|\alpha\|_1$	L1 norm.
$\|\alpha\|_2$	L2 norm.
λ	Regularization parameter in LASSO.
$\bar{\mathbf{z}}_i^{cont}$	Context Vector of the FHSI.
$\mathbf{z}_i^{cat}$ $\mathbf{z}_i^{bin}$ y $\mathbf{z}_i^{num}$	Vector representing the categorical, binary, and numeric features of the i-th patient.
θ_i	Attention matrix for the i-th patient.
$\underline{\theta}$	Attention matrix employed in the HAM architecture
$\tilde{\mathbf{X}}_i$	Weighted input matrix generated by attention architectures.

$\mathbf{M}$	Matrix representing relevance/salience scores.
$\mathbf{X}_i^P$	Perturbed input.
π	Perturbation operator that using $\mathbf{M}$ create the perturbed input.
$\hat{y}_i^P$	Output of the perturbed input.
$\mu_i^{(t,d)}$	Moving average corresponding to the time series of patient i-th, feature d up to time t
$m^{(t,d)}$	coefficient that weighs the relative importance of the current value of the time series compared to the moving average.
W_{MAW}	Moving average window width.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

An outline of the thesis is provided in this chapter. First, the motivation for the conducted research is explained. The primary and specific goals are then explained in detail, along with the proposed methodology for achieving them. Additionally, it provides an overview of the subsequent chapters.

1.1 Context and Motivation

Infectious diseases have posed a significant threat to human populations for centuries [5]. The outbreak of Coronavirus Disease 2019 (COVID-19) serves as a reminder of how new or mutating infectious diseases can have a profound impact [6]. The evolving landscape of infectious diseases impacts global health and requires continual vigilance and innovation in healthcare approaches, specifically in the Intensive Care Units (ICUs). The complexity of ICU patients renders infectious diseases a significant concern in these units.

In recent years, Antimicrobial Multidrug Resistance (AMR) has become a more pressing worldwide health concern [7], becoming one of the most critical threats in ICUs. AMR is the phenomenon where microorganisms evolve to withstand the effects of multiple antimicrobial drugs designed to eliminate them [8] and can arise both naturally and as a consequence of the overuse and misuse of antibiotics [9]. The implications of AMR are far-reaching, making previously treatable infections potentially untreatable and increasing the complexity, cost, and risk of treatments [10]. In a detailed study from 2019, researchers revealed concerning statistics about the impact of AMR [1]. According to the study, nearly 5 million deaths worldwide were linked to AMR, with approximately 1.27 million deaths directly caused by drug-resistant

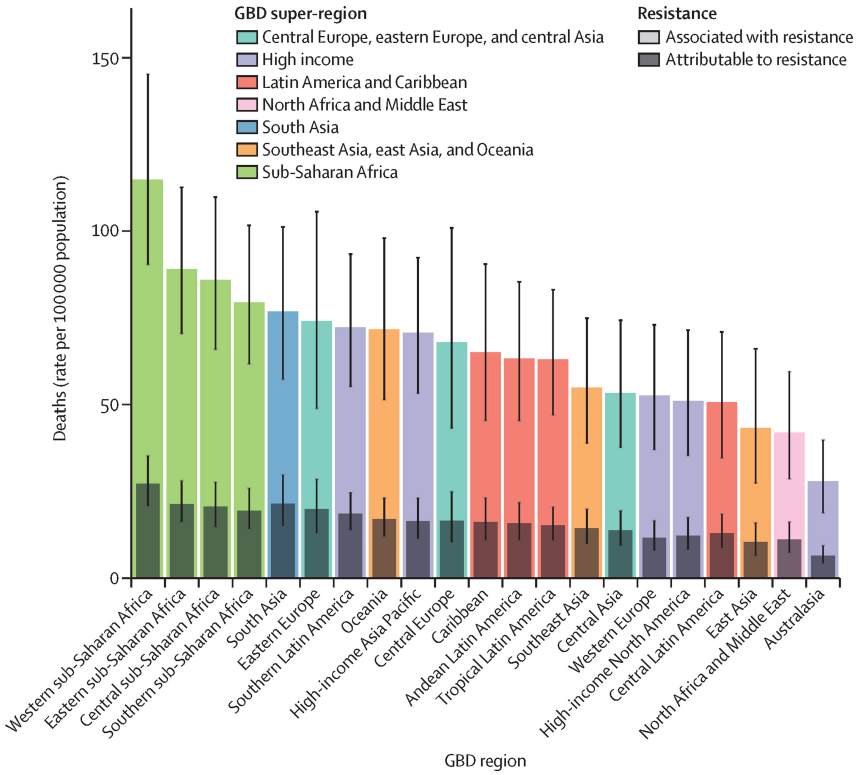


Figure 1.1: The mortality rate associated with AMR by Global Burden of Disease region, 2019. Figure obtained from [1].

infections [1]. Therefore, AMR emerged as the 12th leading cause of death in the Global Burden of Disease Study, surpassing The Human Immunodeficiency Virus (HIV) and Malaria [1]. Furthermore, mortality rates from AMR for all ages were higher in developing countries (see Figure 1.1). Looking into the future, Figure 1.2 predicts that in 2050, the number of fatalities from AMR-related causes will surpass that of cancer deaths, with an alarming 10 million deaths per year if no decisive actions are taken to tackle AMR. This prediction paints a grim scenario of a burgeoning health catastrophe that requires urgent and effective global interventions.

AMR not only impacts health but also has a substantial economic impact. In the European Union alone, the annual costs associated with AMR amount to approximately 1.5 billion euros, arising from healthcare expenses and productivity losses [11]. Globally, the economic damage by 2050 due to AMR is predicted to be comparable to those of the 2008 financial crisis [12]. That impact could result in a 100 trillion loss in Gross Domestic Product [11, 12], emphasizing the need for economic and medical strategies to tackle AMR. Moreover, the confluence of AMR and the development of new infectious diseases such as COVID-19 pose a severe risk for global health [13].

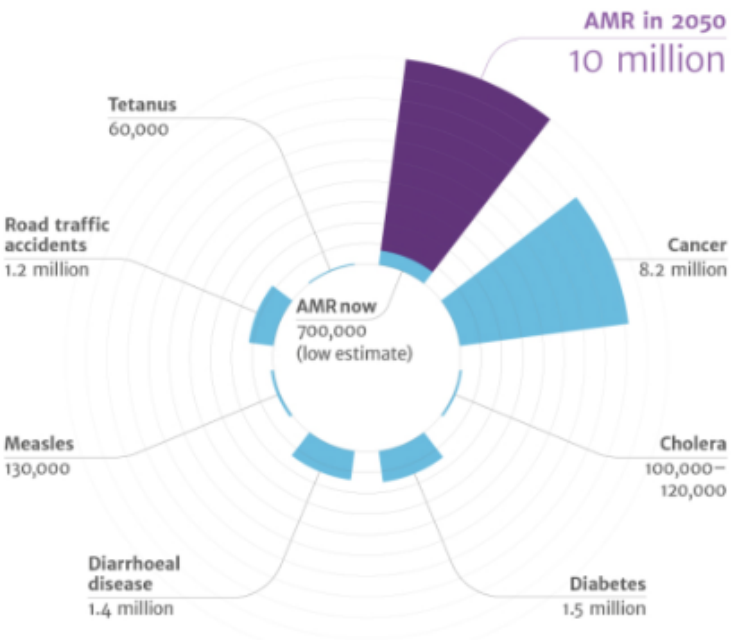


Figure 1.2: Estimated number of deaths worldwide in 2050. Figure obtained from [2].

The COVID-19 pandemic has added another layer of complexity to the landscape of infectious diseases. Spain, one of the countries most severely impacted during the initial wave, experienced an overwhelming strain on its healthcare system, particularly its ICUs [14]. This strain has challenged the capacity of healthcare services and instigated significant economic burdens.

In an economic context, the burden of COVID-19 on the Spanish healthcare system has been substantial [14, 15]. The overall cost associated with ICU care for COVID-19 patients has been high due to several factors: the necessity for highly skilled medical staff, the utilization of sophisticated and costly medical equipment, and prolonged hospital admissions for COVID-19 patients [16]. Additionally, the disruption caused in the ICUs had cascading effects on the treatment of other medical conditions, leading to an indirect but significant economic impact. Evaluating the economic toll of the pandemic's initial wave is complex, but some studies have estimated it. For instance, a detailed analysis focusing on Spain's economic indicators in 2020 indicated a sharp decline in the Gross Domestic Product by approximately 11.41%, a 9.37% decrease in business turnover, and a notable 11.9% rise in unemployment rates during that period [3]. Figure 1.3 illustrates the economic impact of COVID-19 on European countries, particularly noting Spain's significant Gross Domestic Product decline. The situation is particularly relevant in the context of ICUs, where the patient population is often at a higher risk and where the cost - both in terms of human life and economic resources - can be substantial.

Given these difficulties, it is clear that a rigorous and comprehensive approach is needed to understand and manage infectious diseases effectively, especially in high-risk settings like ICUs. The cost, both in human lives and economic terms, requires a focused effort on developing innovative solutions. Addressing the complex challenges of infectious diseases, such as AMR and COVID-19, requires the adoption of innovative and technologically advanced approaches. Central to this endeavor is the exploitation of the rapidly evolving field of machine learning (ML), and particularly Deep Learning (DL) [17]. Artificial Intelligence, particularly ML and DL, demonstrates remarkable efficacy in discerning intricate patterns and correlations within extensive datasets. This capability holds paramount importance in enhancing disease detection, forecasting disease advancement, and formulating therapeutic strategies [17, 18].

Implementing ML and DL in healthcare is enhanced by using electronic health records (EHRs). EHRs serve as comprehensive longitudinal data repositories, encompassing a broad spectrum of patient information, including demographics, medical history, test results, and treatment strategies. The longitudinal nature of EHRs, tracking patient data over time, can be modeled to act as Multivariate Time Series (MTS). This diverse and comprehensive data collection is a valuable resource for ML models, offering useful data necessary for robust models [19].

Analyzing MTS data from EHRs is vital in addressing the complexities of infectious diseases in healthcare [20]. MTS data includes measurements of various variables over time, offering a detailed view of the dynamic aspects of disease progression, how diseases spread, and the effectiveness of treatments. Integrating MTS analysis with ML frameworks allows for developing robust predictive models capable of capturing complex interactions in infectious disease dynamics. These methods enable the extraction of valuable insights regarding the interplay between various clinical variables, environmental factors, and disease dynamics, facilitating more accurate forecasting and early detection of outbreaks.

This thesis aims to address the challenges of infectious diseases by employing data-driven methodologies. The objective is to enhance our understanding and management of clinical time-varying data, thereby contributing to the broader effort of combating infectious diseases such as AMR and COVID-19. By integrating healthcare and data science expertise, this research aims to devise robust models against this health challenge.

1.2 Objectives

This dissertation aims to design DL and graph analytical models using the heterogeneity of EHR data to facilitate the early detection of infectious diseases within the ICU.

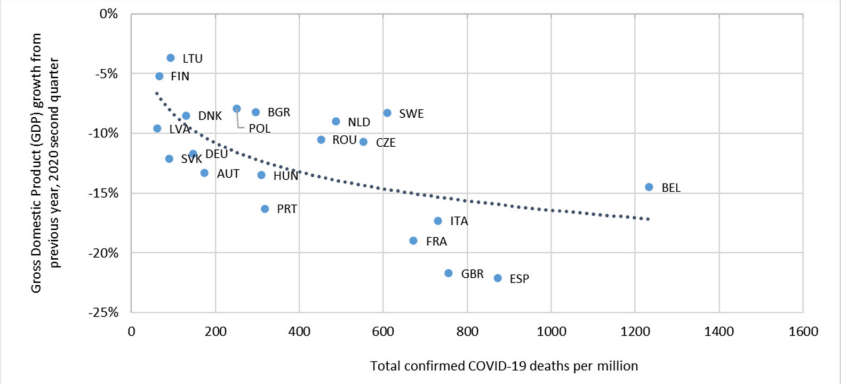


Figure 1.3: The COVID-19 pandemic's impact on health and economic growth in European nations through the second quarter of 2020. Figure obtained from [3].

In addressing the challenges inherent to the research objectives, this work employs established methodologies, adapting existing techniques to fulfill the specified tasks. This approach is practical, prioritizing real-world applications over theoretical exploration. The intention is to apply ML in real clinical scenarios to demonstrate its benefits. Moreover, the thesis includes the development of ML algorithms tailored to the specific tasks at hand. Additionally, the dissertation investigates the potential applicability and acceptability of explainable methods within the healthcare system. In pursuit of the principal objective, the following specific goals are methodically addressed:

- **(O1)**: To develop and preprocess two datasets related to infectious diseases within the ICU setting. In the first dataset, AMR in the complex ICU setting of the University Hospital of Fuenlabrada (UHF) is analyzed and modeled using MTS data in addition to other static variables from 2004 to 2022. The second dataset compiles data associated with antibiotics and their treatment protocols used during the initial COVID-19 outbreak gathered from the UHF. Both datasets will incorporate patient health information and have undergone a data cleaning and modeling procedure to ensure high-quality data. These datasets play a pivotal role in understanding infectious diseases in the ICU and facilitating potential future research endeavors in this critical area.

- **(O2)**: To develop multimodal architectures to predict the early detection of AMR bacteria in the ICU. This involves using static features to model patients' initial health status and tracking the evolution of their conditions with MTS data. Additionally, several feature selection (FS) techniques and interpretable ML models have been considered for extracting knowledge. This dual approach is designed to improve decision-making in ICU settings

by offering a comprehensive view of patient health and contributing to enhanced medical practices.

- **(O3)**: To comprehensively analyze COVID-19 patient outcomes in the ICU. Toward that end, we undertook an exploratory analysis, identifying statistically significant differences between patients who passed away due to the virus and those who were discharged. This analysis included demographic variables, symptoms, comorbidities, and medication records. Additionally, we established a methodological framework for applying graph-based network analytics to static and dynamic (including MTS) variables within EHR. Our approach is aimed at deepening our understanding of the key factors that affect patient outcomes in the ICU.

1.3 Summary of Contributions

This section enumerates the publications associated with the objectives previously outlined.

The goal of O1 and O2 were accomplished in [21, 22, 23, 24, 25]. In this dissertation, we will primarily focus on [21], as it represents a key study where the author of this thesis served as the main contributor. In [21], we employ interpretable DL and signal processing methodologies to analyze MTS data collected from the ICU at the UHF (Madrid, Spain). Given the prevalence of AMR bacteria as a significant threat to health systems and particularly to ICU patients, early detection of antibiotic resistance is crucial for patient prognosis. To facilitate the adoption and implementation of data-based processing and learning schemes in ICUs, our research develops trustworthy, interpretable models for early AMR prediction by integrating meaningful FS with Recurrent Neural Networks (RNNs). The models developed are designed to handle the challenges of class imbalance, limited patient data, and data irregularity while maintaining a balance between accuracy and user comprehension. Our models incorporate SHapley Additive exPlanations (SHAP) post-hoc interpretability, and their understandability and trustworthiness have been validated by clinicians for practical application. Furthermore, we employ linguistic fuzzy systems to generate natural language explanations that are easy to understand.

Building upon the framework presented in [21], we have developed an improved methodology, which we propose in [26]. This new methodology involves a novel approach for analyzing and modeling complex EHR data by combining static features and MTS to predict the emergence of AMR in the ICU. To effectively characterize the patient's initial health status and its evolution, we developed multimodal deep neural network (DNN) architectures, with the

most promising results obtained from the "First Hidden State Initializer" model. This sample-dependent variable selection framework generates an encoding vector to supplement the MTS context. Complementarily, we employed two approaches to knowledge extraction. Initially, classical FS methods were examined, followed by the implementation of a permutation multimodal FS approach. Both procedures were assessed in terms of performance and interoperability. Subsequently, various interpretable mechanisms were applied to discern hidden patterns within the dataset. Overall, the proposed interpretable multimodal DNNs demonstrate efficacy in predicting AMR while concurrently providing explanations for AMR prediction in the ICU. Furthermore, the methodology proposed could be used in a range of clinical issues involving EHR data, thereby broadening its impact and usefulness.

The contributions and novelties of the works previously presented are clearly articulated to emphasize specific advancements in the field of AMR within ICUs:

- This research introduces a comprehensive analysis and modeling of MTS and static features pertinent to AMR. The study covers a substantial dataset of 3,470 ICU patients, preprocessed and modeled to extract clinical knowledge. We proposed a novel methodology to overcome challenges inherent in AMR classification, such as addressing class-imbalance, managing irregularities in MTS, and handling high-dimensional data.
- The development of multimodal architectures is a significant novelty in our work. These architectures integrate the static and MTS data to characterize both the initial health status and the progression of each patient's condition.
- Our approach to knowledge extraction encompasses both traditional and novel methods, considering both FS strategies and the application of interpretable mechanisms.
- Finally, the validation of our models' interpretability by clinicians underscores the practical relevance and applicability of our findings in real-world clinical settings. This step ensures that our models are not only statistically valid but also clinically meaningful, making them valuable tools for predicting AMR in ICUs.

The publications related with O1, O2 are listed below:

- [21] **S. Martínez-Agüero**, C. Soguero-Ruiz, J. M. Alonso-Moral, I. Mora-Jiménez, J. Álvarez-Rodríguez, and A. G. Marques, "Interpretable clinical time-series modeling with intelligent feature selection for early prediction of antimicrobial multidrug resistance," Future Generation Computer Systems, vol. 133, pp. 68-83, 2022, Elsevier.

- [22] **S. Martínez Agüero**, A. G. Marques, J. M. Alonso-Moral, I. Mora-Jiménez, J. Álvarez-Rodríguez, and C. Soguero-Ruiz, "Multimodal Interpretable Data-Driven Models for Early Prediction of AMR". arXiv preprint arXiv:2402.06295, 2023

- [23] À. Hernández-Carnerero, M. Sànchez-Marrè, I. Mora-Jiménez, C. Soguero-Ruiz, **S. Martínez-Agüero**, and J. Álvarez-Rodríguez, "Dimensionality reduction and ensemble of LSTMs for antimicrobial resistance prediction," Artificial Intelligence in Medicine, vol. 138, p. 102508, 2023, Elsevier.

- [24] L. Pascual-Sánchez, I. Mora-Jiménez, **S. Martínez-Agüero**, J. Álvarez-Rodríguez, and C. Soguero-Ruiz, "Predicting multidrug resistance using temporal clinical data and machine learning methods," in 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2826-2833, 2021, IEEE.

- [25] À. Hernández-Carnerero, M. Sànchez-Marrè, I. Mora-Jiménez, C. Soguero-Ruiz, **S. Martínez-Agüero**, and J. Álvarez-Rodríguez, "Antimicrobial resistance prediction in intensive care unit for pseudomonas aeruginosa using temporal data-driven models," International Journal of Interactive Multimedia and Artificial Intelligence, 2021.

The COVID-19 related tasks (O1, and O3) are addressed in [27]. The work in [27] aims to develop a graph-based methodology to identify connections between comorbidities, previous medications, symptoms, and COVID-19 treatments for patients admitted to a Spanish ICU during the pandemic's initial surge. By examining the trajectory of these patients, who either passed away due to the virus or were discharged from the ICU, we employ hypothesis testing via bootstrap methods to discriminate between the deceased and non-deceased populations. Subsequently, graph-based representations and network analytics are utilized to identify associations among clinical features. This analytical approach reveals that deceased patients typically had multiple comorbidities with solid connections and received a more comprehensive range of drugs during their ICU stay. Furthermore, the most common treatment was the concurrent administration of lopinavir/ritonavir and hydroxychloroquine, independent of patient outcomes. By employing graph tools and representations, this study provides insights into the connections among comorbidities, pharmacological treatments, and patient trajectories. In conclusion, the methodology presented constitutes a novel data-analysis tool for clinicians, with potential applicability for inspecting post-COVID symptoms or patient progression. The publications related with O1, O3 are listed below:

- [27] **S. Martínez-Agüero**, C. Soguero-Ruiz, J. M. Alonso-Moral, I. Mora-Jiménez, J. Álvarez-Rodríguez, and A. G. Marques, "Data and Network Analytics for COVID-19

ICU Patients", IEEE Journal of Biomedical and Health Informatics. 25(12):4340–4353, 2021.

1.4 Outline of the Dissertation

This Thesis consists of seven chapters, which are described as follows:

- Chapter 1: Introduction and Objectives. In this chapter, the motivation behind this dissertation, the objectives, and the methodology used are presented. Furthermore, the structure of the document and the scientific contributions are shown.
- Chapter 2: Preliminaries. This chapter establishes a solid foundation for understanding the complex terms and behaviors associated with AMR. An explanation of the COVID-19 pandemic follows it. The chapter aims to provide clarity on these subjects by examining EHRs and illustrating how data-driven models are applied in the healthcare sector.
- Chapter 3: Dataset Description and Exploratory Analysis. It provides a detailed overview of the datasets utilized, along with an initial exploration of their characteristics.
- Chapter 4: Interpretable Data-Driven Modeling for Early Prediction of Antimicrobial Multidrug Resistance. This chapter delves into the development of clinical time-series models aimed at early prediction of AMR.
- Chapter 5: Multimodal Interpretable Models for Early Prediction of Antimicrobial Multidrug Resistance. In this chapter, we expand the scope to include multimodal, interpretable, data-driven approaches, further enhancing the predictive capabilities of AMR.
- Chapter 6: Data and Network Analytics for COVID-19 ICU Patients. In this chapter, the focus narrows to the application of data and network analytics, specifically within the context of COVID-19 patients in ICUs.
- Chapter 7: Conclusions and Future Work. This chapter summarizes the essential findings and outlines potential directions for future research in this rapidly evolving field.

Chapter 2

Preliminaries

In this chapter, we conduct an analysis of infectious diseases, with a particular focus on AMR and the COVID-19 pandemic. This exploration is undertaken through the prism of EHRs and data-centric models in healthcare. Our objective is to establish a comprehensive understanding of these diseases within the context of modern data analytics and ML methodologies. Firstly, the chapter presents a description of AMR, a complex and growing challenge in modern medicine. This description will focus on the complex dynamics of AMR, emphasizing its multifaceted nature. Following this, the chapter shifts focus to COVID-19, offering a comprehensive analysis of its impact on global health systems and patient outcomes. Then, we explore the role of EHRs in modern healthcare, underscoring their potential as a rich source of data for epidemiological and clinical research. Furthermore, the chapter discusses the application of data-driven models within the healthcare environment, highlighting their applicability in decision-making and patient care. Lastly, we present the methodology and mathematical notation used throughout this study. This section aims to present an overview of the scientific methods used, ensuring the integrity and reproducibility of our findings. This chapter serves as the foundation for the discussions that follow, establishing a multidisciplinary framework essential for comprehending how the healthcare problems at hand are connected.

2.1 Infectious Diseases

The study of infectious diseases is a critical and ever-evolving field within medical research. This study involves the investigation of illnesses caused by organisms such as bacteria, viruses, fungi, and parasites. These diseases pose significant public health challenges globally. This

section is dedicated to unraveling the complex nature of two infectious diseases: AMR and COVID-19.

Antimicrobial Multidrug Resistance

AMR is a biological process where microorganisms, including bacteria, viruses, fungi, and parasites, evolve to resist the effects of medications designed to kill them [9]. This phenomenon is particularly alarming in healthcare because it renders standard treatments less effective, prolongs infections, and increases the likelihood of these resistant infections spreading to other patients. The development of AMR can be attributed to various factors, including genetic changes in microorganisms and the pressure from the usage of antimicrobial drugs [28]. In clinical settings, this often results from the overuse, misuse, or inappropriate prescribing of antibiotics and other antimicrobial agents.

Analyzing clinical procedures—particularly the antibiogram—along with patient cultures is a crucial strategy in managing AMR. Cultures involve isolating and growing microorganisms from samples (e.g., blood, urine, wound swabs) on a growth medium. These microorganisms are then tested for antibiotic susceptibility or resistance through antibiograms, where they are incubated with antibiotics to assess their response. Antibiograms are pivotal in addressing AMR, aiding in the selection of the most appropriate treatment for individual patients. They also play a crucial role in more significant public health strategies aimed at preventing the spread of drug-resistant bacteria in hospitals. Regularly performed antibiograms help to identify resistance trends, track new resistant strains, and refine prescribing practices. These tools are also vital for antimicrobial stewardship programs, promoting responsible antibiotic use to mitigate AMR development. However, controlling AMR spread remains an increasingly challenging task in modern healthcare.

The rise of AMR has significant implications for clinical practice. It has led to an increasing number of cases where standard antibiotic therapies, once effective against infections, are now failing [29]. As a result, conditions that were previously considered manageable are becoming significant health concerns. This change requires more resources, more extended hospital stays, and the use of last-resort, often more toxic and expensive antibiotics [1]. In extreme cases, some infections have become completely untreatable, posing grave risks to patient health and public health at large [30].

Moreover, the clinical challenges of AMR extend beyond the treatment of severe infections. The effects of AMR are very severe in the ICU, since they are critical areas where severely ill patients are at a heightened risk of acquiring infections, including those caused by multidrug-resistant organisms. The presence of AMR in these settings not only complicates treatment

strategies but also significantly increases morbidity and mortality rates [31]. For patients in the ICU, infections with resistant pathogens mean longer hospital stays, increased medical costs, and higher risks of unfavorable outcomes. The escalated use of broad-spectrum antibiotics in ICUs, necessary to address severe and complex cases, further exacerbates the problem by potentially promoting the development and spread of resistant germs [32]. Consequently, ICUs often serve as epicenters for AMR, making rigorous infection control protocols and antimicrobial stewardship programs essential [33]. It also complicates the management of less critical health issues, such as post-surgical recovery and the treatment of minor wounds or cuts [34]. Such cases, which were once easily handled with antibiotic treatments, now carry the risk of developing into severe complications due to the reduced efficacy of these drugs [30]. This scenario underscores the urgent need to employ data-driven models and knowledge acquisition methodologies in clinical settings.

In summary, from a clinical standpoint, AMR represents a significant shift in the landscape of infectious disease treatment and management. It necessitates a reevaluation of current prescribing practices, increased vigilance in monitoring and controlling infections, and an effort in research and development to stay ahead of this evolving challenge [35]. Efforts to mitigate the impact of AMR include global initiatives like the World Antibiotic Awareness Week, a focus on improved hygiene practices, and the reduction of antibiotic usage in the agriculture sector, particularly with food-producing animals [36].

COVID-19

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), commonly known as COVID-19, is a highly infectious respiratory illness first identified in December 2019. It is caused by a novel coronavirus belonging to the same family as SARS-CoV and MERS-CoV (Middle East Respiratory Syndrome Coronavirus), which have been responsible for previous epidemics. Symptoms of COVID-19 range from mild to severe and include fever, cough, shortness of breath, fatigue, and loss of taste or smell [37]. Severe cases can lead to pneumonia, acute respiratory distress syndrome, multi-organ failure, and death, particularly in older adults and those with pre-existing health conditions [38]. COVID-19 spreads primarily through respiratory droplets and aerosols when an infected person coughs, sneezes, talks, and touches surfaces contaminated with the virus [39]. The disease's rapid transmission rate and potential severity have required global public health interventions, including social distancing, mask-wearing, and extensive vaccination campaigns.

The COVID-19 virus has emphasized the importance of effective treatments for infectious diseases. The pandemic led to the intensive use of antibiotics in hospitals, particularly in

COVID-19 patients during its initial stage. The overuse and misuse of antibiotics during the pandemic have increased resistance in bacteria, posing a threat to the treatment of future infections [36].

During the early stages of the COVID-19 pandemic, Spain experienced one of the highest mortality rates globally, significantly exceeding those reported in many other countries [40]. This situation led to considerable challenges in the Spanish healthcare system, particularly over several months as the health system struggled to cope with the effects of the pandemic [41]. The severity of some of the infected patients reduced the capacity of the Spanish healthcare system to manage routine medical services and emergencies. It also highlighted the limitations in the availability of intensive care resources, placing them under extreme pressure [42]. This was of notable concern given the increased need for intensive care beds, medical equipment such as ventilators, and specialized staff. It was further exacerbated by the lengthy ICU stays required for severe COVID-19 cases, often extending for weeks [43].

Among the patients admitted to the ICU, a high mortality rate of over 50% was observed [44, 41]. This figure stands in contrast with the overall mortality rate associated with COVID-19, which was estimated to be around 2.3% during the initial stages of the pandemic [45]. This high ICU mortality rate was also significantly more significant than the 22% mortality rate seen in ICU patients with other viral pneumonia [46]. These statistics underscore the exceptional severity of COVID-19, especially in critical care settings, and highlight the need for focused efforts in ICUs during pandemics. It also reiterates the critical role of ICUs in the healthcare system, particularly in managing patients suffering from severe infectious diseases such as COVID-19. Therefore, there is a clear need for thorough research into the epidemiology, emergence, prevalence, and burden of infectious diseases.

Additionally, some treatments being developed for COVID-19, such as monoclonal antibodies, have potential implications for the rise of AMR [47]. These monoclonal antibodies are produced in laboratories and are designed to target specific proteins on the virus's surface. However, if the virus mutates and changes these proteins, the monoclonal antibodies may lose their effectiveness, potentially leading to the development of resistant strains of the virus. Moreover, the COVID-19 pandemic has disrupted healthcare infrastructures across the globe, causing delays in identifying and treating bacterial infections. This disruption, combined with an increased reliance on antibiotics, could inadvertently create an environment where AMR can develop more efficiently. The complexity of this issue requires investigation and consideration of diverse solutions, as will be elaborated upon in the Future Work Section 7.2 of this dissertation.

Developing new solutions demands a coordinated effort from governments, industry, and medicine. Over the past years, the development of modern, data-driven healthcare systems has been pivotal. This approach relies on data to classify, diagnose, treat diseases, and provide patient care [48]. In the pre-digital era, patient data was recorded and stored manually. This process, while essential for maintaining medical records, came with significant challenges such as potential errors, difficulties in data retrieval, and limited accessibility. The digitization of clinical data, however, has brought about a transformation, offering enhanced scalability and reusability of data, and in turn, providing novel opportunities for innovation and improved patient care [49]. The current digital revolution is changing the way that patients are treated by bringing technology with transformative potential into a variety of medical sectors, including general medical practices, hospitals, and research institutes. The conversion of analog data sources into digital formats has optimized the storage and retrieval of patient data and enabled the application of advanced data analysis techniques [50]. Clinical medicine has a tendency to accept technology later than other industries. Despite efforts to integrate algorithms into everyday practice, clinical decision support systems are not widely adopted [51]. This situation must change, as clinical data is now a valuable resource for improving patient care, identifying disease patterns, and developing new treatments. Building upon this digital transformation, EHRs have revolutionized how healthcare providers manage and access patient information.

2.2 Electronic Health Records

EHRs provide healthcare professionals with immediate access to longitudinal patient health information, enhancing the precision of health history analysis [52]. This real-time data access enables the identification of health trends and patterns, which is crucial for crafting more targeted and effective treatment strategies and ultimately boosting patient health outcomes. EHRs also streamline collaboration among healthcare teams while minimizing the risk of errors stemming from incomplete or outdated patient information. Overall, EHRs have significantly increased the efficiency and efficacy of clinical data analysis, directly benefiting patient care.

EHRs provide healthcare professionals with immediate access to patient data, allowing for a more accurate analysis of a patient's health history. EHRs also enable the identification of trends and patterns in patient data, which can be used to develop more effective treatment plans and improve patient outcomes [53]. Furthermore, EHRs facilitate collaboration between healthcare institutions and reduce errors caused by incomplete or inaccurate information [54]. With the use of EHRs, clinical data analysis has become an efficient and effective practice, leading

to improved patient care [55]. The data from EHR can be used to develop powerful predictive algorithms. By analyzing large datasets, these algorithms can identify patterns and make predictions. These algorithms fall under the broad category of ML [56]. The data contained within EHRs is heterogeneous, constituting multimodal information such as diagnosis codes, medication prescriptions, and laboratory test results; semi-structured data such as clinical notes; and unstructured data such as images from radiology, pathology, and other medical imaging modalities. An illustration of the multimodal EHR data used in this dissertation is provided in Figure 2.1.

The utilization of EHR for the development of data-driven models could enhance clinical decision-making and operational management offers. Research based on EHR data has already shown promising results in various areas, including patient mortality prediction, hospital readmission prediction, and early detection of clinical events, indicating a shift towards more proactive, personalized medicine [57, 58, 59]. This shift could lead to more efficient resource use and cost optimization [60]. However, several challenges inherent in EHR data complicate its use:

- Data heterogeneity: EHRs contain diverse data types from various clinical environments. Integrating these data sources requires significant efforts, resources, and expert domain knowledge [61].
- High dimensionality: The data recorded in EHRs, such as diagnoses and drug prescriptions, result in a high-dimensional feature space. With over 13,000 diagnosis codes and 3,430 drug codes, the feature space becomes increasingly complex.
- Data quality: The success of models driven by EHR data depends on the quality of the data. EHR data often includes incomplete records, outliers, and inconsistencies, potentially leading to unreliable outcomes. Effective pre-processing of data is therefore crucial to mitigate these issues [62].
- Temporal characteristics: EHRs data track patient information over time, usually recording visits across various clinical departments. This longitudinal nature of the data adds complexity, particularly in accurately interpreting the sequence and timing of these interactions. This characteristic presents unique challenges for traditional ML approaches, a concern raised by [62].
- Imbalanced classes: For ML applications in healthcare, having a representative sample size is essential. Class imbalance in EHRs poses a significant challenge in getting a

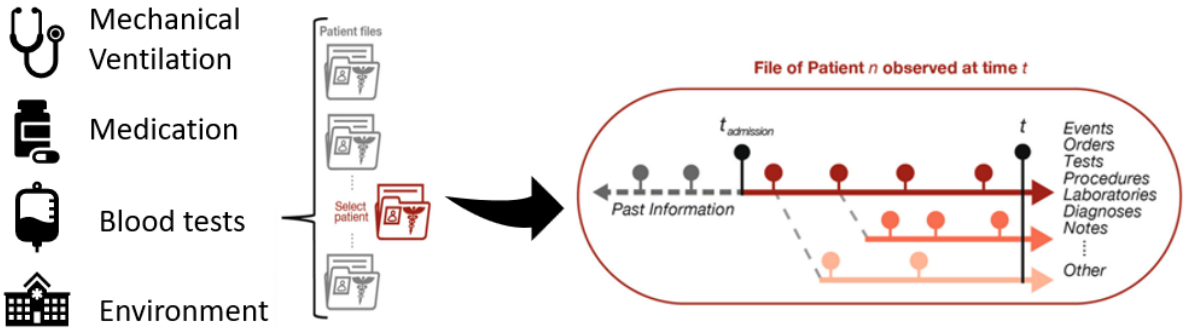


Figure 2.1: Illustration of the multimodal EHR data used in this dissertation.

representative sample. This occurs when there are far more records of one class compared to the other class. This imbalance can lead to models that perform well on the majority class but poorly on the more clinically significant minority class [63].

- Data privacy: The sensitive nature of clinical data, protected by laws like the Health Insurance Portability and Accountability Act in the US and the General Data Protection Regulation in the EU, limits its use. This legal framework requires institutional approvals for data sharing, which can be restrictive and time-consuming [64, 65].

In conclusion, integrating EHRs into healthcare systems represents a significant improvement in patient data management and clinical decision-making. EHRs provide an invaluable repository of multimodal, longitudinal data, enabling a more accurate and holistic understanding of patient health. These records also set the groundwork for the development of sophisticated, data-driven models in healthcare. This exploration will delve into how ML and other computational techniques can use EHR data to predict outcomes, personalize treatments, and enhance the efficiency and effectiveness of healthcare services.

2.3 Data-driven Models in the Healthcare Environment

ML techniques are becoming vital in analyzing clinical data, primarily because they excel at uncovering patterns and trends in EHR [17]. These algorithms are particularly adept at predicting clinical outcomes, such as forecasting the likelihood of a patient developing certain diseases or their response to treatments. This predictive power comes from ML's ability to process and learn from a patient's medical history and various data types.

One significant advantage of ML in healthcare is its capability to manage complex, multi-faceted data [66]. Medical data often includes diverse formats like clinical notes, lab results, and time series, all of which ML can integrate and analyze efficiently. This integration is crucial because it enables ML models to discern relevant features from vast data sets, leading to more precise predictions.

The use of ML techniques in clinical data analysis has become increasingly important in recent years because ML offers a powerful tool for identifying patterns and trends in EHR data [67]. Additionally, ML algorithms can be trained to predict clinical outcomes, such as the likelihood of a patient developing a particular disease, based on the patient's medical history and other available data [68]. One of the critical advantages of ML in the clinical domain is its ability to handle complex and high-dimensional data. As was previously highlighted, medical data is challenging to work with, as it often includes a wide range of different data types (e.g., clinical notes, laboratory results, time series) that must be integrated and analyzed. ML models can automatically learn relevant features from this data and use them to make accurate predictions. In this context, the use of ML in the analysis of clinical data has the potential to transform the way that healthcare is delivered by enabling more accurate predictions and personalized treatment plans. For example, the Manifal Hospital in Bangalore employed IBM's Watson for Oncology in 2015, revealing a disparity in diagnostic consensus rates among cancers, with 85% for rectal and 17.8% for lung cancer [69]. Similarly, the Johns Hopkins University Hospital utilized AI technologies in collaboration with GE Healthcare in 2016, improving operational efficiency in areas such as patient bed assignment and inter-hospital transfers [70]. The Moorfields Eye Hospital in London successfully incorporated AI solutions to identify eye diseases, demonstrating 94% accuracy, thereby addressing the growing demand for eye scans [71].

One of the defining characteristics of clinical data is its temporal nature. It is essential to consider the temporal aspect of clinical data when developing ML models. One of the critical reasons why considering the temporal aspect of clinical data is essential is that healthcare data is inherently dynamic. Patients' health statuses change over time, and the data collected during each visit reflects these changes [72]. If we ignore the temporal aspect of clinical data, we risk overlooking significant trends or patterns that could be critical for accurate predictions or diagnoses. For example, if we are predicting the likelihood of a patient developing a particular disease, we need to consider how their health status changes over time. Medical records typically contain a sequence of events, such as diagnoses, procedures, and medications, that unfold over time [72]. This temporal structure can provide valuable insights into disease progression and treatment efficacy but also poses significant challenges for data analysis.

Recognizing this, recent ML advancements have focused on techniques that exploit this temporal aspect of clinical data. These techniques, often categorized as temporal or sequential models, can model dynamic relationships between events over time and make predictions about future events based on past observations. RNNs are a common type of temporal model that is applied to a wide range of clinical prediction tasks, such as predicting patient outcomes and identifying disease patterns [73]. RNNs are particularly good at modeling time-varying data, as they can process sequences of variable length and capture long-term dependencies between events.

In clinical settings, it is crucial to consider static features that characterize the initial health status and MTS to model the patient's health status evolution. In such cases, fusion models are employed to unify these two types of data [74]. Fusion models aim to combine multiple information sources to maximize the predictive power of the resulting model [74]. Different data fusion models can be employed to combine static variables and MTS and extract complementary, accurate, and comprehensive knowledge [75, 76]. Clinical decision-making in the healthcare arena is greatly aided by the integration of static data, such as age or comorbidities, with MTS [77]. As a result, a number of data-fusion architectures have lately been suggested for use in medical environments. Cheng et al., for instance, used a collection of deep early fusion neural networks (NNs) to forecast hospitalizations for gastrointestinal bleeding based on various multimodal data entered into the EHR. [78]; Li et al. created a joint fusion model to combine data on demographics, notes from doctors, and clinical time series features [79]. Shuai et al. employed an attentional joint fusion classifier to predict the illness risk using text notes and time series [80].

While the fusion models provide a practical approach to integrating heterogeneous health data, they do not fully utilize the intricate relational structures within the data. To address this gap, graph-based models emerge as a promising direction, enabling a more comprehensive representation of the data. These models can encapsulate various relationships among data points, such as temporal dependencies in MTS and intrinsic connections among static features. Furthermore, they can capture complex interactions and dependencies that standard fusion models may overlook [81]. Graphs, consisting of a set of nodes and edges connecting them, represent a mathematical structure that combines versatility with a wealth of analytical results derived from disciplines such as graph theory and complex systems [81]. This versatility is evidenced by the adoption of graphs across various data-science-related fields, such as ML [82], signal processing [83], and statistics [84], to represent intricate dataset structures and integrate them into data-science processing and learning pipelines [84]. In addition to their mathematical benefits, graphs are a comprehensible tool for representing high-dimensional data and facilitate

the visualization of the information in question [82, 83, 85]. Prior research has demonstrated the value of network-based approaches in visualizing collaborative EHR usage for heart failure patients [86], modeling disease graphs [87], and predicting graphs for previously unknown adverse drug reactions [88]. Despite the comparatively limited application of graph-based representations for MTS in the literature, the preliminary findings are encouraging. Such analytical methods hold the potential to detect outbreaks in their emerging stages or to characterize the clinical progression of patients [89, 90, 91].

Although research has demonstrated the effectiveness of DL models and graph modeling in various fields, including medicine [92, 93], their complexity often makes it difficult to understand their underlying mechanisms. This leads to a situation where, rather than unraveling the inner workings of the models, researchers are often constrained to indirect methodologies to decipher the impact of input features [94]. This lack of interpretability has been identified as the primary barrier to applying DL models in clinical decision-making, which requires understandable relationships between input data and predictions [95]. To overcome this challenge, fundamental advances in ML interpretability are necessary [96]. In recent years, several interpretable models have been developed in the healthcare domain using different methods, such as feature importance methods [97, 98], feature interaction attribution [99, 100], neuron layer attribution [101, 102], and explanation with high-level concepts [21, 103].

Chapter 3

Data Description and Exploratory Analysis

This chapter describes the two datasets (the AMR dataset and the COVID-19 dataset) that the UHF gathered, and then it presents the exploratory data analysis that was done for this thesis.

3.1 Antimicrobial Multidrug Resistance Dataset

This thesis was developed using real clinical data from 3,470 patients who were registered in the ICU of UHF between 2004 and 2022. This dataset contains patient information, including demographic, microbiology laboratory culture data, temporal features that reveal patient health progression, and temporal features reflecting ICU occupancy and antibiotic pressure. The integration of these heterogeneous data provides a view of the evolving landscape of ICU management over nearly two decades. This dataset, therefore, stands as a valuable resource for in-depth analysis in medical research, offering insights into long-term trends and the effectiveness of care strategies. To extract new insights from this data , we will next conduct an exploratory data analysis of the features examined.

Demographic features

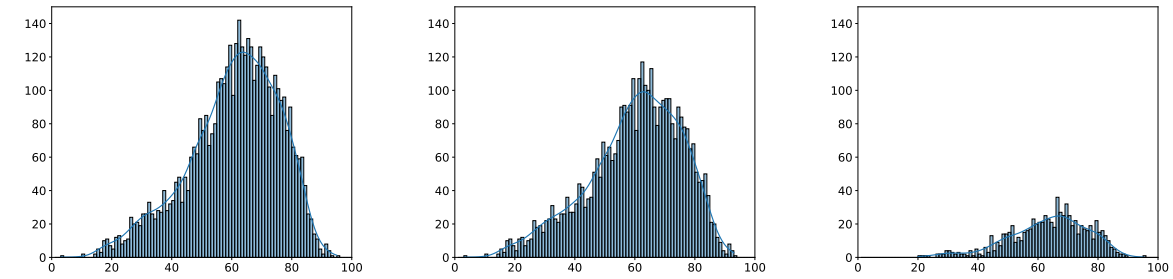
Demographic features are a vital set of patient information, typically collected at the onset of their stay in the ICU. This information forms the foundation for understanding the patient's baseline health status and potential care needs. These features not only include features like age and gender but also extend to more detailed aspects such as the patient's clinical origin before ICU admission, which provides insight into their pre-ICU health trajectory, and the reason

for ICU admission, which is coded according to the International Classification of Diseases, Ninth Revision (ICD-9) [104]. Furthermore, we included the presence of pluripathology or the coexistence of multiple diseases or conditions in a single patient. This aspect is crucial in the ICU as it can complicate patient management and diagnosis.

Furthermore, we modeled critical care indices such as Acute Physiology and Chronic Health Evaluation II (Apache-II) or Simplified Acute Physiology Score (SAPS-3) [105, 106]. These indices are scoring systems designed to assess disease severity and predict mortality risk in ICU patients. They incorporate a range of clinical parameters and have been validated in numerous studies [106, 105]. These scoring systems assist in clinical decision-making and facilitate comparative studies and research in critical care medicine. By including these indices in the demographic profile, healthcare providers can better understand the patient's condition and align their care approaches with the severity of the patient's illness.

In order to understand the behavior of the demographic features previously presented we performed an exploratory analysis of all of them, including Age, Gender, Origin before ICU, and Reason for Admission, as well as the Apache-II and SAPS-3 scoring systems.

Regarding age, Figure 3.1 shows the age distribution for (a) all patients, (b) non-AMR patients, and (c) AMR patients. The subfigures (a), (b), and (c) suggest a prevalence of middle-aged individuals, with fewer representations from younger and older age groups. Specifically, the average ages are as follows: 60.52 ± 15.15 years for subfigure (a), 59.93 ± 15.50 years for subfigure (b), and 63.31 ± 13.02 years for subfigure (c).

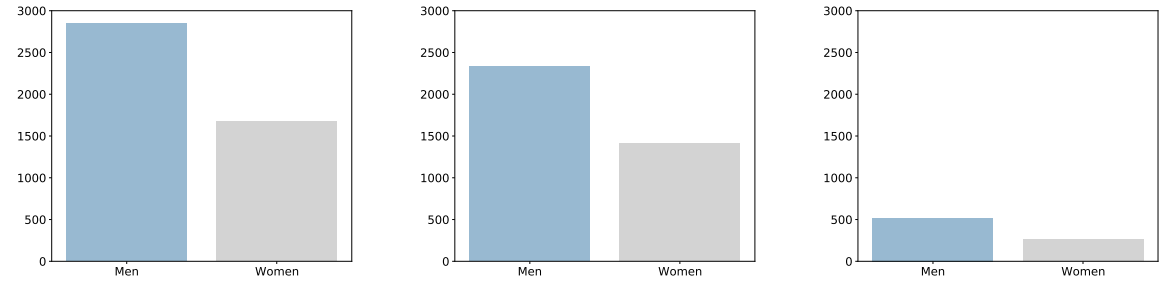


(a) (b) (c)

Figure 3.1: Histograms of age for (a) all patients; (b) non-AMR patients; and (c) AMR patients.

Figure 3.2 shows the gender distribution for (a) all patients, (b) non-AMR patients, and (c) AMR patients. The subfigures (a), (b), and (c) demonstrate a consistent predominance of male patients in ICU admissions across the groups. Notably, there is a marginally higher proportion

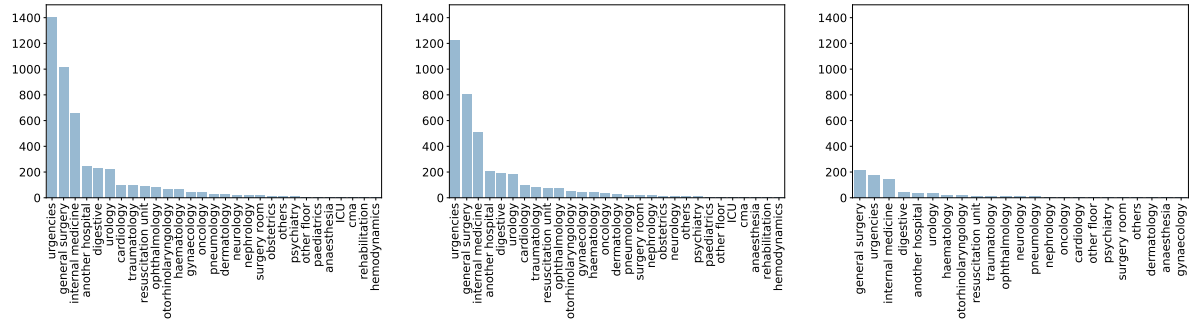
of male patients in the non-AMR group than the others.



(a) (b) (c)

Figure 3.2: Bar plots comparing the gender distributions for (a) all patients; (b) non-AMR patients; and (c) AMR patients.

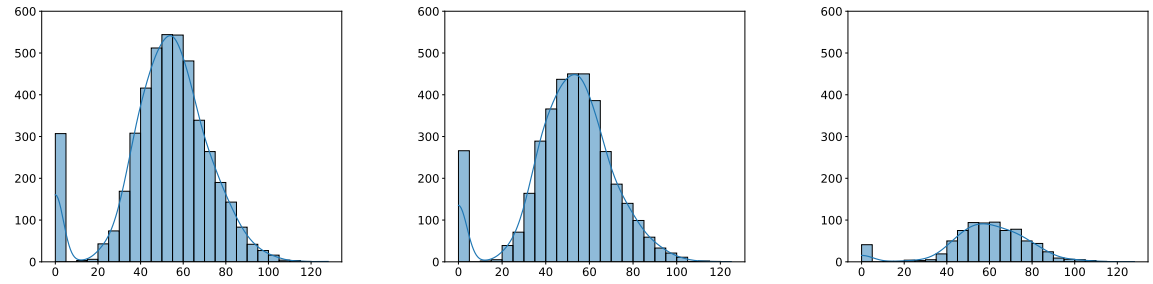
Figure 3.3 presents bar plots of the origin before ICU: (a) all patients, (b) non-AMR patients, and (c) AMR patients. Most patients generally come from the emergency room, especially in subfigure (a) and (b), indicating that urgent conditions are common reasons for ICU admissions. In contrast, subfigure (c) shows a more evenly distributed distribution across units, with a continued but less dominant emergency room frequency.



(a) (b) (c)

Figure 3.3: Bar plots comparing the origin before ICU for (a) all patients; (b) non-AMR patients; and (c) AMR patients.

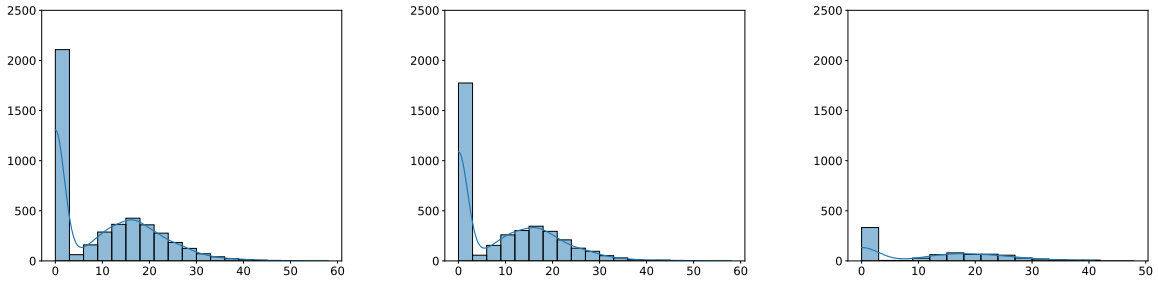
Subsequently, Figure 3.4 presents histograms comparing the SAPS-3 values for three patient groups: (a) all patients, (b) non-AMR patients, and (c) AMR patients. The histogram for all patients (a) indicates a clustering around the median, showing that most of the patients have a similar score. The non-AMR group (b) follows a similar pattern with a slight shift towards higher scores. The AMR group (c) also shows a similar distribution as the other subfigures. Specifically, the average SAPS-3 score are as follows: 52.02 ± 20.52 for subfigure (a), 50.70 ± 20.33 for subfigure (b), and 58.46 ± 20.20 for subfigure (c).



(a) (b) (c)

Figure 3.4: Histograms comparing the SAPS-3 for (a) all patients; (b) non-AMR patients; and (c) AMR patients.

Following the analysis, Figure 3.5 offers a visual comparison of the APACHE-II scores across three patient groups: (a) all patients, (b) non-AMR patients, and (c) AMR patients. The histogram for all patients (a) is right-skewed, showing most have lower APACHE-II scores and, thus, a lower mortality risk, with fewer high-risk patients. The non-AMR group (b) displays a similar pattern but slightly shifted towards higher scores, indicating slightly more severe cases. The AMR group (c) follows a similar distribution as the other subfigures. Specifically, the average APACHE-II score is as follows: 9.43 ± 10.38 for subfigure (a), 9.00 ± 10.11 for subfigure (b), and 11.53 ± 11.34 for subfigure (c).

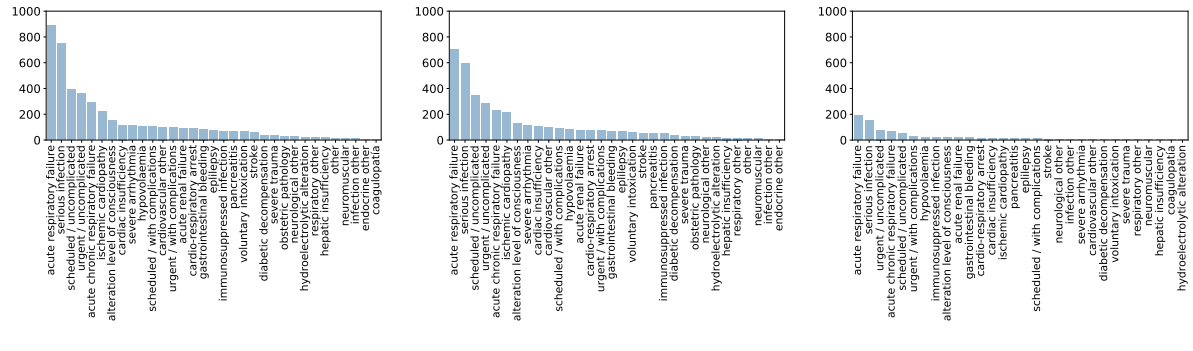


(a) (b) (c)

Figure 3.5: Histograms comparing Apache-II for (a) all patients; (b) non-AMR patients; and (c) AMR patients.

Figure 3.6 shows bar plots that compare the reasons for ICU admission across three distinct patient groups: (a) all patients, (b) non-AMR patients, and (c) AMR patients. The bar plots visually represent the frequency of each reason for ICU admission within these populations. All subfigures exhibit similar right-skewed distributions, showing common ICU admission reasons

like "acute respiratory failure" and "serious infection" as most prevalent, with others being less frequent.



(a) (b) (c)

Figure 3.6: Bar plots comparing the reason why the patients got into the ICU for (a) all patients; (b) non-AMR patients; and (c) AMR patients.

Microbiology laboratory results

These variables detail the cultures the UHF microbiology laboratory conducts to detect possible infections. To contextualize the information in these features, it is crucial to emphasize that the type of germ identified (if any) and the type of culture performed are essential. This distinction is critical as having a superficial skin infection is different from having one in a lung. However, the UHF laboratory does more than just culture to identify possible infections; they also perform antibiograms on the identified germs. Antibiotic susceptibility tests, or antibiograms, are tests that detail the effectiveness of a set of antibiotics used against the germs detected in the cultures. This information is crucial in an era where antibiotic resistance is a growing concern, as it aids in selecting the most appropriate and effective treatment regimen.

Processing these variables was a vital step in the study, particularly in differentiating between AMR and non-AMR patients. The dataset categorizes patients based on their AMR status, making it a valuable tool for studying resistance trends and contributing to broader efforts in infection control and antibiotic utilization.

Temporal features

These features track the treatments of the patient under study. We can group these MTS into different groups of features: (i) the antibiotics taken by the patient under study, (ii) the mechanical ventilation (MV), and (iii) the results of previous cultures performed on the patient.

The antibiotics taken by the patient were grouped into families of antibiotics such as Aminoglycosides (AMG), Antifungals (ATF), various generations of Cephalosporins (CF1, CF2, CF3,

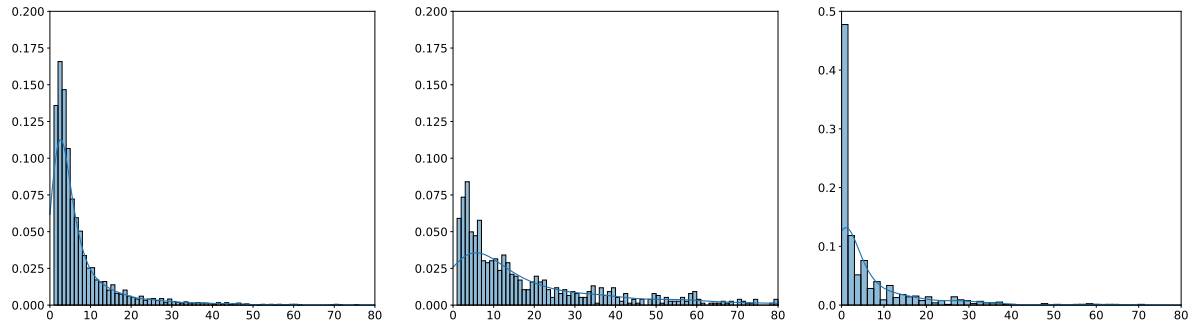
CF4), Glycopeptides (GLI), Lincosamides (LIN), Glycyclines (GCC), Lipopeptides (LIP), Macrolides (MAC), Monobactamas (MON), Nitroimidazolics (NTI), unclassified antibiotics (Others), Sulfamides (SUL), Oxazolidinones (OXA), Penicillins (PEN), Broad-Spectrum Penicillins (PAP), Polypeptides (POL), Quinolones (QUI) and Tetracyclines (TTC). Each was administrated based on their specific antimicrobial properties and the patient's health needs. Thus, for a specific patient (denoted as the i-th patient), the characteristic associated with each treatment (denoted as the d-th treatment) is represented by a matrix of binary variables $\underline{\mathbf{x}}_i^d \in \{0,1\}^{D\times T_i}$. This matrix indicates whether or not the patient received the specified treatment during each of the T_i time slots, which correspond to 24-hour periods during the patient's stay in the ICU.

The MV metric reflects the severity of the patient's respiratory condition and has implications for potential complications, such as ventilator-associated pneumonia. The length of mechanical ventilation can also indicate the overall severity of the patient's illness and their progress in the ICU.

The results of previous cultures' features are vital for understanding the patient's microbiological background. They offer insights into any infections the patient may have had and their response to treatments. This historical perspective is essential for identifying antibiotic resistance patterns and tailoring future antimicrobial therapies to the patient's needs.

These temporal features — antibiotic administration, mechanical ventilation, and previous culture results — provide a comprehensive view of the patient's treatment journey in the ICU. In order to get information about the trends of those features, we performed an exploratory data analysis.

First, to illustrate MTS length, Figure 3.7 shows three histograms, each representing the frequency distribution of days of stay for patients, divided into two categories: AMR patients and non-AMR patients. Additionally, the figure includes a subfigure indicating the specific days when the first multi-resistant culture was identified in patients. This information is very informative, providing a clear-cut-off point for our analysis. The histograms clearly show that patients with AMR infections generally have more extended hospital stays compared to those without AMR infections. The longer tail on the right side of the AMR patient distribution indicates a greater frequency of extended hospitalizations. However, the distribution for non-AMR patients declines more sharply, suggesting that these patients typically have shorter stays. Additionally, we can observe an early peak in the days to culture histogram for AMR patients, indicating that the cultures are performed quickly, which is crucial for treating their serious infections effectively. These histograms highlight AMR's significant healthcare burden, evidenced by longer hospital stays and the need for rapid diagnostic efforts.



(a) (b) (c)

Figure 3.7: Histograms displaying the length of stay for (a) non-AMR patients and (b) AMR patients. Additionally, subfigure (c) illustrates the distribution of days on which the first multidrug-resistant organism was detected in patients.

After analyzing the duration of the patients' stays and the length of the MTS, we examined the antibiotics intake. In this dissertation, we treat the families of antibiotics as binary features. A binary feature, by definition, includes only two distinct categories encoded as '0' or '1', representing the absence or presence of a specific attribute, respectively. Specifically, a '1' indicates that the patient received an antibiotic from that family during the time step, while a '0' indicates that the patient did not receive any antibiotic from that family. Heatmaps in Figure 3.8 illustrate the proportion of AMR and non-AMR who received treatment with each antibiotic family over time. Each column represents a different time step during the patient's stay (only 14 days are considered), while each row corresponds to a specific family of antibiotics. The first heatmap (Figure 3.8 (a)) details drug consumption for the AMR patient population, while the second heatmap (Figure 3.8 (b)) represents drug intake for non-AMR patients. The third heatmap (Figure 3.8 (c)) shows the difference in drug use between these two groups (subtracting the antibiotic consumption values of the AMR group from the non-AMR group). From these heatmaps, we can discern several patterns:

- Variability in drug usage: There is a variability in drug usage between the two patient populations. This could indicate an adaptation of the treatment based on the patients' resistance profiles, suggesting personalized treatment in the ICU.

- Time-dependent trends: The data show how drug intake varies over time, with some drugs being administered more frequently at certain times during the patient's stay. This could reflect the evolving nature of treatments and the response to therapy. Families such as ATF, CAR, or GLI are more frequently given towards the end of the stay of the patients, while others keep the same behavior.

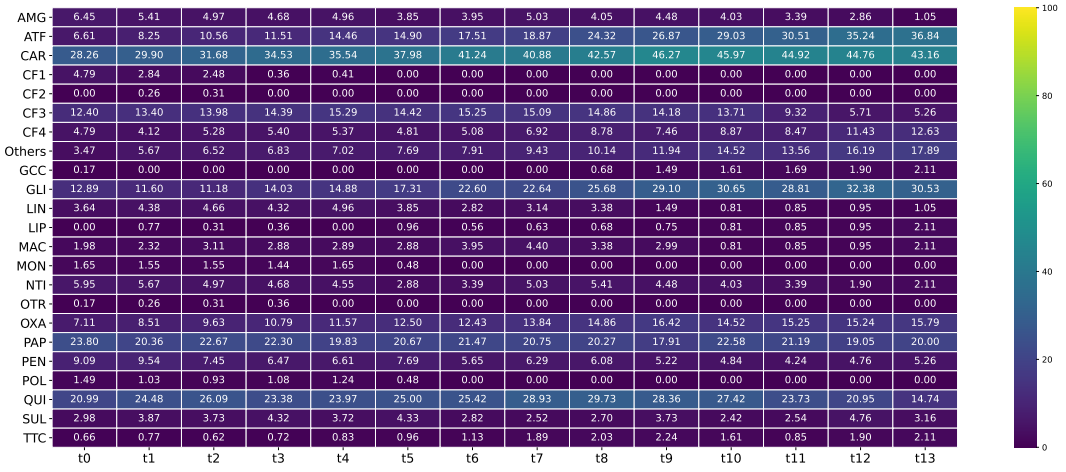
- Specific drugs of interest: Certain drugs are used differently between the two groups, as evidenced by the differences in the heatmaps. Drugs represented by larger positive differential values may indicate those that are preferred or more effective in treating AMR infections, such as ATF, CAR, or GLI. Conversely, drugs with negative values may be less used in AMR patients, perhaps due to ineffectiveness against resistant germs.

Further research that combines these drug usage patterns with patient outcomes and microbiological data is vital for a deeper understanding of these patterns and for developing effective treatment strategies for both AMR and non-AMR patients. The study notes that certain drug families, like Carbapenems (CAR), Glycopeptides (GLI), and Antifungals (AFT), are more frequently used in treating AMR patients. In contrast, non-AMR patients are often treated with antibiotics like Penicillins. However, some drug families, such as Quinolones, Lipopeptides, and broad-spectrum Penicillins, are used similarly in both AMR and non-AMR patient groups.

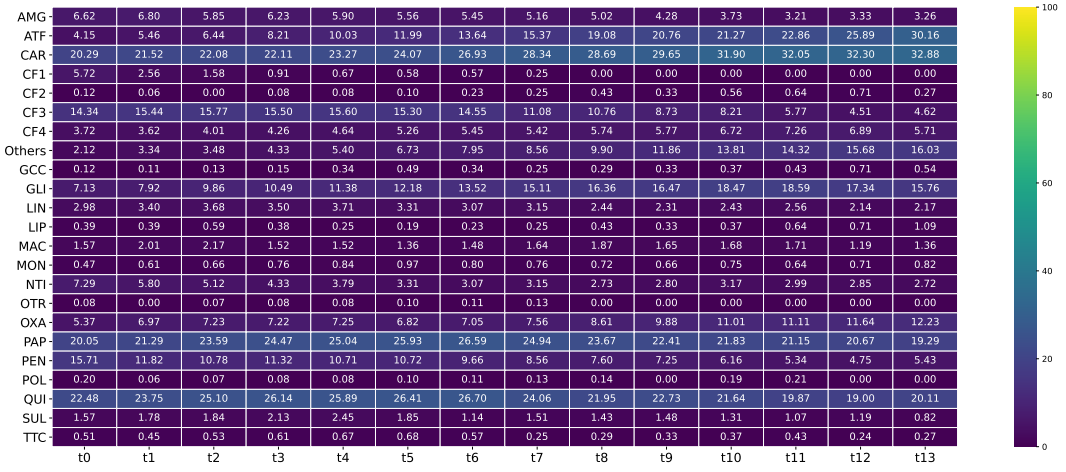
We continue analyzing the temporal dynamics of MV usage among patients, focusing on showing patterns and trends over time. In Figure 3.9, we present a detailed comparison of MV usage over time through two distinct boxplot diagrams complemented by a line plot. The left boxplot of subfigure (a) corresponds to the AMR population, while the right boxplot of subfigure (b) shows the non-AMR population data. Interpreting these plots, the AMR boxplots' skewed medians and variable box sizes indicate an inconsistent need for MV during a patient's ICU stay, hinting at an increased requirement for intensive MV starting from day 6. In contrast, the right plot's uniformity suggests consistent median values over time, implying regular MV usage for most patients upon treatment initiation. This consistent pattern potentially masks variations, necessitating the inclusion of Figure 3.9 (c). This additional plot illustrates the daily ratio of patients receiving MV. It reveals a consistently higher mean for AMR patients, with a progressively steepening trend over time. Therefore, after analyzing both figures, we can conclude that AMR patients have a higher distribution of MV usage. Additionally, this analysis incorporates Table 3.1, which provides statistical evidence of the increased likelihood of AMR patients receiving prolonged MV contained in the 14-time steps window, corroborating the observations made in Figure 3.9.

Finally, we study the patient's cultural features to identify specific time steps where particular germs have been detected. Out of the multiple germs identifiable through cultures, six are prone to multidrug resistance: *Stenotrophomonas*, *Pseudomonas*, *Enterobacter*, *Staphylococcus Aureus*, *Acinetobacter*, and *Enterococcus*. Consequently, we developed six features, each tracking the appearance of one of the previous germs in the cultures performed in the patient.

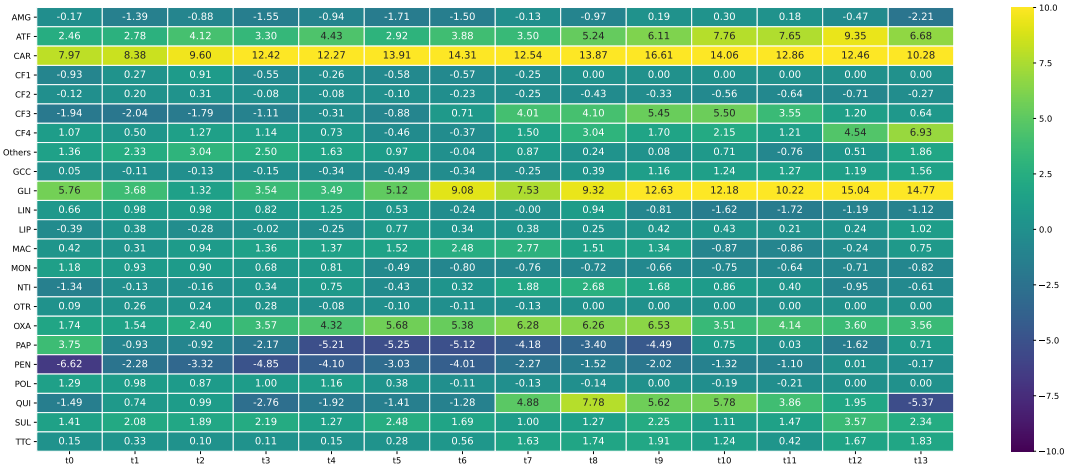
In subsequent sections, we reference these germs with the suffix *pc* (previous cultures), il-



(a)

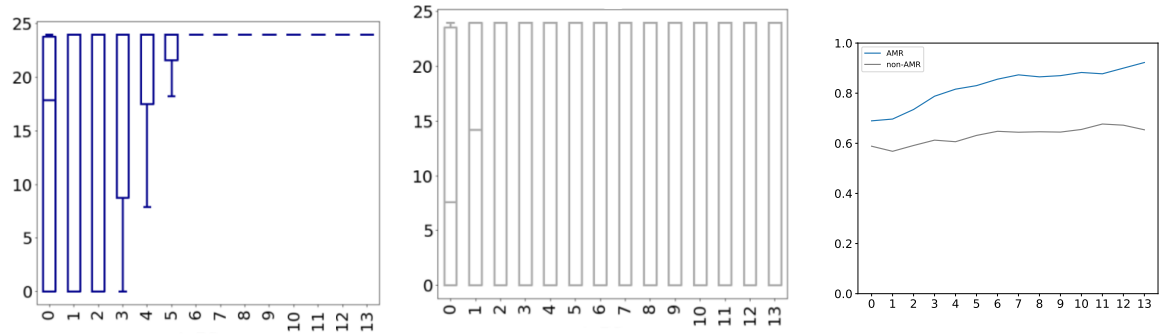


(b)



(c)

Figure 3.8: Heatmaps representing the percentage of patients taking each family of antibiotics over time for (a) AMR patients; (b) non-AMR patients; and (c) the comparison between AMR and non-AMR groups. Rows represent antibiotic families, and columns represent time steps.



(a) (b) (c)

Figure 3.9: Boxplots depicting the distribution of mechanical ventilation duration for patients in each time step: (a) AMR patients and (b) non-AMR patients. Additionally, subfigure (c) illustrates the percentage of AMR and non-AMR patients requiring mechanical ventilation during each time step.

Feature	Statistic	Total Patients	AMR patients	Non-AMR patients
Mechanical Ventilation	Mean	14.13	17.50	13.40
	Median	23.93	24.00	23.42
	σ	11.34	10.09	11.46
	Minimum	0.00	0.00	0.00
	Maximum	24.00	24.00	24.00

Table 3.1: Statistics for the time the patients are assisted with mechanical ventilation, presented over a period of 14 time steps.

lustrating the features discussed herein. For instance, the feature representing the occurrence of *Pseudomonas* is labeled $Pseudomonas_{pc}$. Additionally, the MTS incorporates a variable $Others_{pc}$, accounting for germs outside the six groups identified in earlier cultures. The inclusion of $Others_{pc}$ acknowledges the potential role of non-resistant germs as precursors to multidrug-resistant strains. It is important to note that the germs referenced in the six primary variables of previous cultures were not AMR. This aligns with our objective to forecast the initial AMR infection in patients.

Figure 3.10 presents a set of histograms representing the specific germ detection frequency ratio in cultures. These cultures are taken before developing an AMR culture or before ICU discharge in non-AMR patients across 14 time steps. The data is segregated into AMR (represented by blue bars) and non-AMR (gray bars). Figures 3.10 (a), (b), and (f) show a pattern of germ detection exclusively at the beginning of the hospital stay for AMR patients. This observation suggests a probable acquisition of pathogens such as Acinetobacter, Enterobacter, and Stenotrophomonas by AMR patients before their ICU admission, possibly from other clinical

departments. Figure 3.10 (c) shows a higher frequency of Enterococcus in AMR patients' cultures, especially during the early and middle phases of their ICU stay, as opposed to non-AMR patients. Contrasting, Figure 3.10 (d) reveals a consistent pattern, with no significant change over time, in the prevalence of Pseudomonas, more frequently found in AMR patients' cultures. Figure 3.10 (e) highlights a tendency for Staphylococcus to be more commonly found in the initial stages of AMR patients' hospitalization. Finally, Figure 3.10 (g) shows the emergence of various germs in cultures, with both AMR and non-AMR patients demonstrating similar ratios of germ presence throughout the observed time frame.

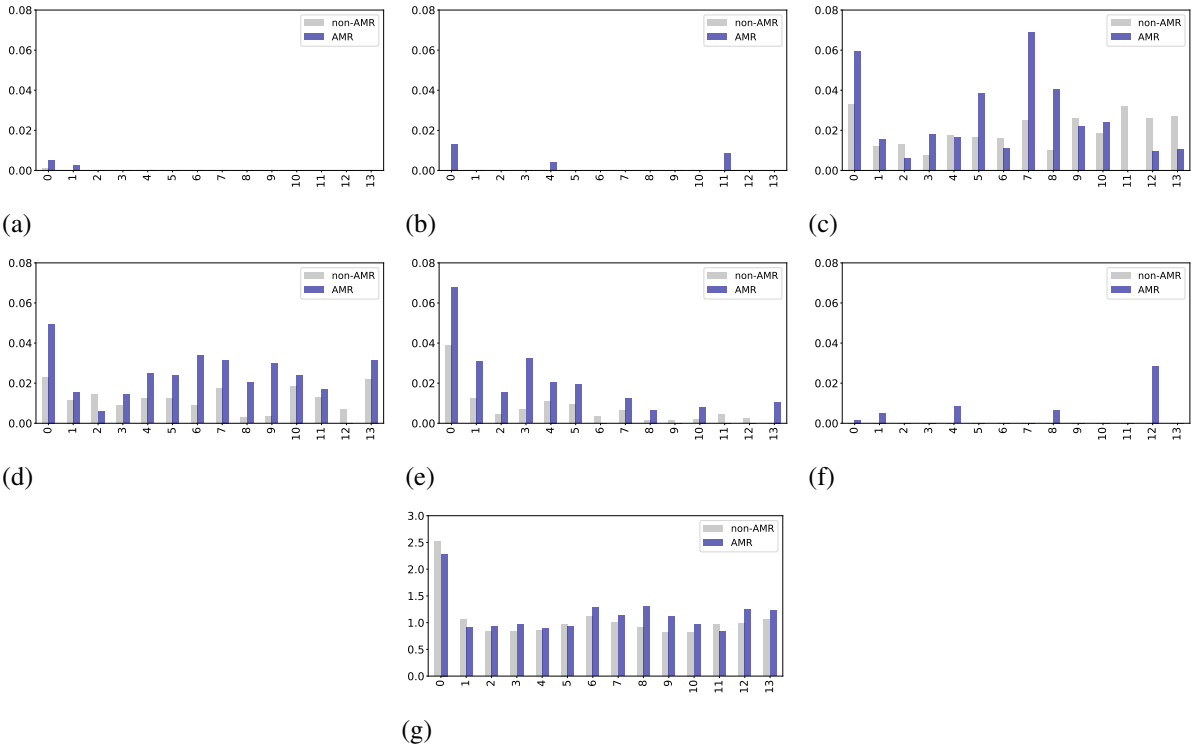


Figure 3.10: Histograms representing the ratio of specific germ detection frequency in previous cultures for: (a) Acinetobacter (b) Enterobacter; (c) Enterococcus; d) Pseudomonas; e) Staphylococcus; f) Stenotrophomonas; and g) other types of germs. Blue bars represent the cultures of AMR patients, and gray ones represent the cultures of non-AMR patients.

Temporal environmental features

We capture both the occupancy of the ICU and the antimicrobials administered to other patients in the ICU (referred to as 'neighbors') during the same periods as the patient under study. These features are crucial to understanding the patient's health status being examined. By considering other patients' conditions sharing the same clinical environment, we gain a comprehensive perspective of the ICU's dynamics and its impact on individual patients.

A total of 17 numeric variables were generated to characterize the ICU occupancy and antibiotic pressure. These variables include the total number of patients in the unit at any given time, indicating the unit's occupancy and potential workload for healthcare staff. Additionally, the number of patients/neighbors with AMR is specifically noted, reflecting the potential risk of cross-infection and the overall burden of AMR pathogens within the unit. Furthermore, the dataset records the use of the 15 different antibiotic families administered to these neighboring patients. This information is crucial as it sheds light on the spectrum of antimicrobial treatments used around the patient under study. These features help in understanding the potential for AMR development within the ICU and can provide insights into the prevailing microbial trends and challenges within the unit.

By integrating these temporal environmental features into the dataset, researchers and healthcare professionals can comprehensively understand the factors influencing patient health in the ICU. This includes not only the direct medical interventions and conditions of individual patients but also the collective impact of the surrounding clinical environment. Such an approach can lead to more informed and effective strategies for infection control, resource allocation, and patient care in these critical settings. The first step to get that information is to study the distributions of the features by an exploratory data analysis.

We start the exploratory data analysis of these features by studying the quantity and characteristics of patients' neighbors and AMR neighbors in the ICU, as detailed in Table 3.2. These numerical variables initially displayed similar distributions for AMR and non-AMR patients. Temporal dynamics were subsequently introduced, as shown in Figure 3.11 through boxplots spanning a 14-day timeline. AMR patient information is depicted in dark blue, while non-AMR information is in gray. In Figures 3.11 (a) and (b), the spread and median values suggest a relatively stable number of neighbors throughout the 14 days, with no significant outliers or day-to-day variability. This could imply a consistent admission rate in the ICU. However, Figures 3.11 (c) and (d), focused on AMR neighbors, show a much lower count and less variation, indicating that AMR cases are less frequent than the overall patient count. The day 0 value in Figure 3.11 (c) suggests that new AMR patients may have a lower likelihood of having an AMR neighbor upon admission, or it could be an outlier.

In conclusion, while the general ICU population around each patient remains consistent, the AMR population is notably lower and does not follow the same distribution. These visualizations underscore the importance of monitoring overall neighbor counts and AMR-specific neighbor counts to manage infection control in a hospital setting.

We also studied the antibiotics administered to the patient's neighbors in the ICU. The

Feature	Statistic	Total Patients	AMR patients	Non-AMR patients
# of neighbors	Mean	8.75	9.13	8.68
	Median	7.00	8.00	7.00
	σ	2.35	2.36	2.34
	Minimum	0.00	0.00	0.00
	Maximum	15.00	15.00	15.00
# of AMR-neighbors	Mean	0.80	0.93	0.78
	Median	0.00	1.00	0.00
	σ	1.09	1.19	1.04
	Minimum	0.00	0.00	0.00
	Maximum	8.00	7.00	8.00

Table 3.2: Statistics for the number of neighbors and the number of AMR-neighbors within a 14-time-step window.

heatmaps shown in Figure 3.12 show the drug intake in the ICU, comparing the consumption patterns of neighboring patients with and without AMR over a 14-day period. Figure 3.12 (a) shows the drug intake of the neighbors of non-AMR patients, and Figure 3.12 (b) presents the drug intake of the neighbors of AMR patients, revealing a heterogeneous pattern of medication use, suggesting that treatments are highly tailored to the individual conditions of each patient. They also show that certain medications are becoming more or less prevalent over time, possibly indicating common changes in treatment approaches in response to patients' changing health situations.

Figure 3.12 (c) highlights the differential impact of AMR on medication strategies by contrasting the data from (a) and (b). This figure reveals positive values indicating increased antibiotic use among neighbors of AMR patients, thereby reflecting a higher usage of antibiotics in the ICU when AMR patients were admitted.

Additionally, we used boxplots (see Figure 3.13) to analyze the usage of antibiotics from a numerical point of view. These plots effectively demonstrate the variation in antibiotic usage. A closer examination of Figures 3.13 (a) and (b) reveals comparable patterns in the usage of all antibiotics under consideration. Notably, there is a consistently higher median usage of antibiotics like CAR and PAP, suggesting a more intensive treatment regimen of those antibiotics in the ICU under study.

Following the database presentation, it should be noted that it will be used in Chapters 4 and 5, which will also go into depth on the specific modeling that was done to carry out the particular experiments.

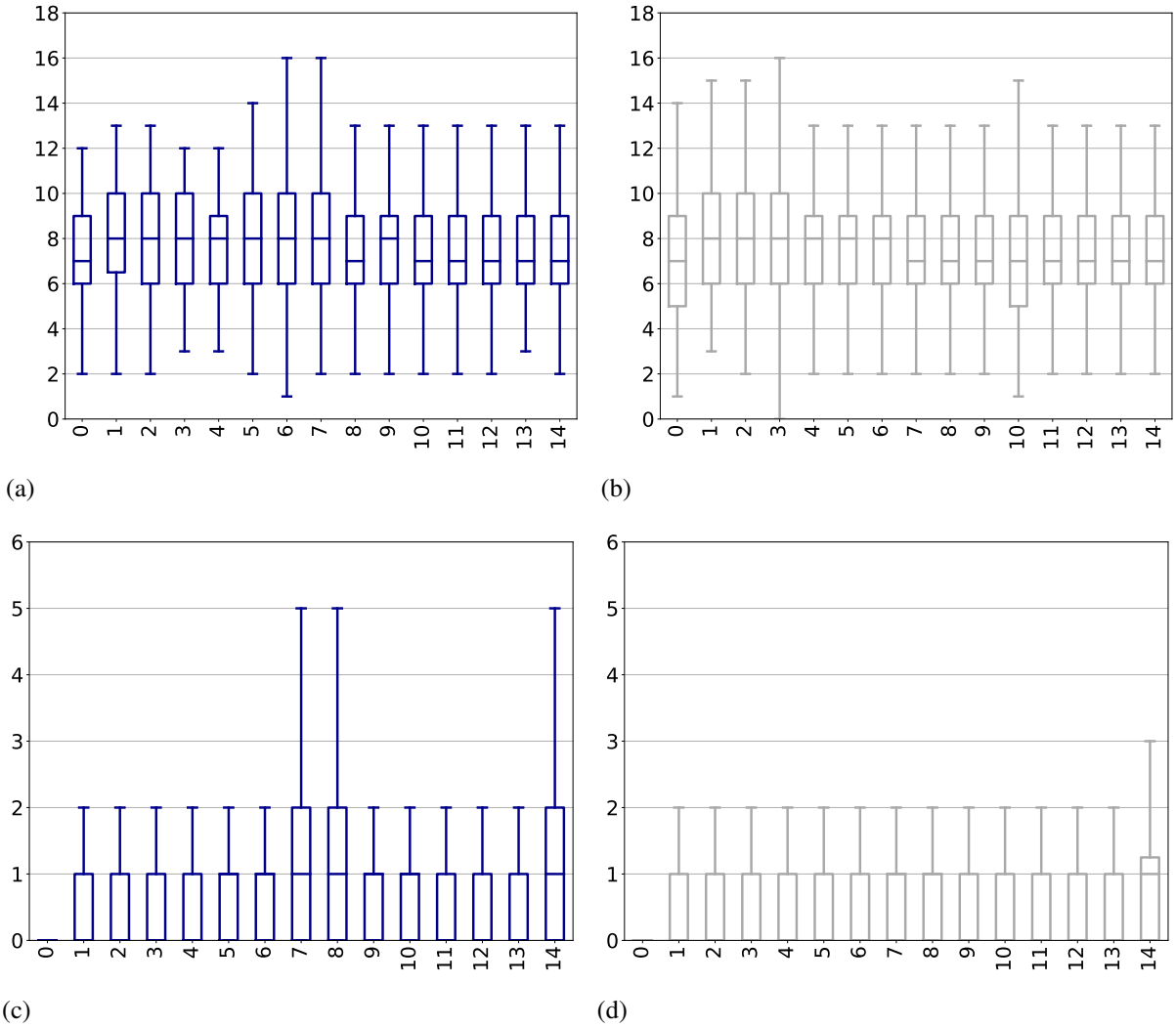


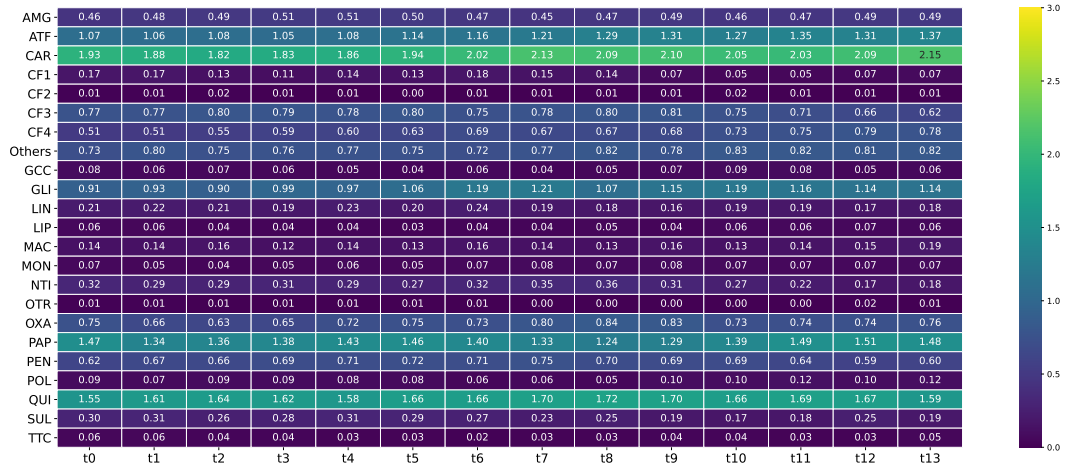
Figure 3.11: Boxplot of number of (a) neighbors of AMR patients; (b) neighbors of non-AMR patients; (c) AMR neighbors of AMR patients; (d) AMR neighbors of non-AMR patients.

3.2 COVID-19 Dataset

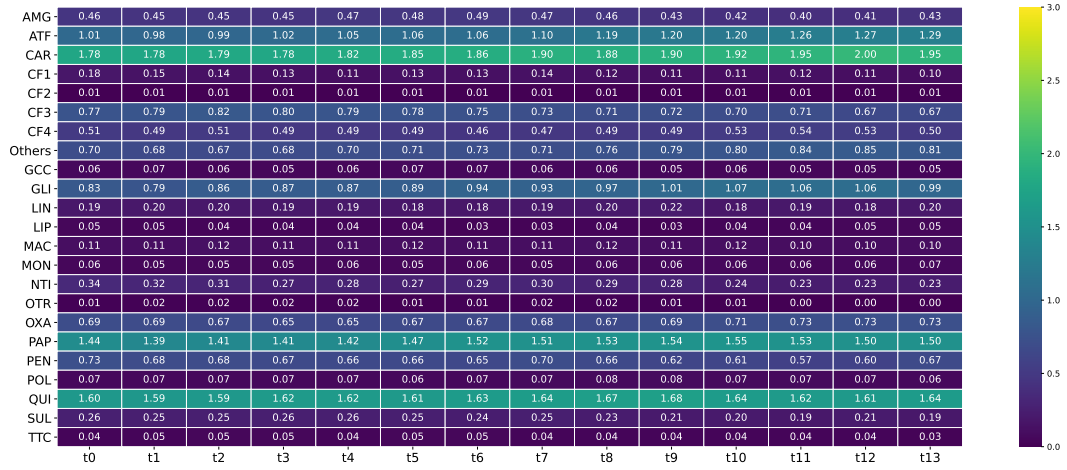
In the present dissertation, a dataset was constructed to examine the EHR of 97 patients diagnosed with COVID-19 who were admitted to the ICU at the UHF. The data includes hospitalizations occurring between March 5, 2020, and July 15, 2020, with all patients having tested positive[1] for COVID-19. 39 patients (40.2%) died in this unit, and 58 patients (59.8%) survived.

The dataset incorporates diverse EHR features clinically chosen by UHF's medical staff.

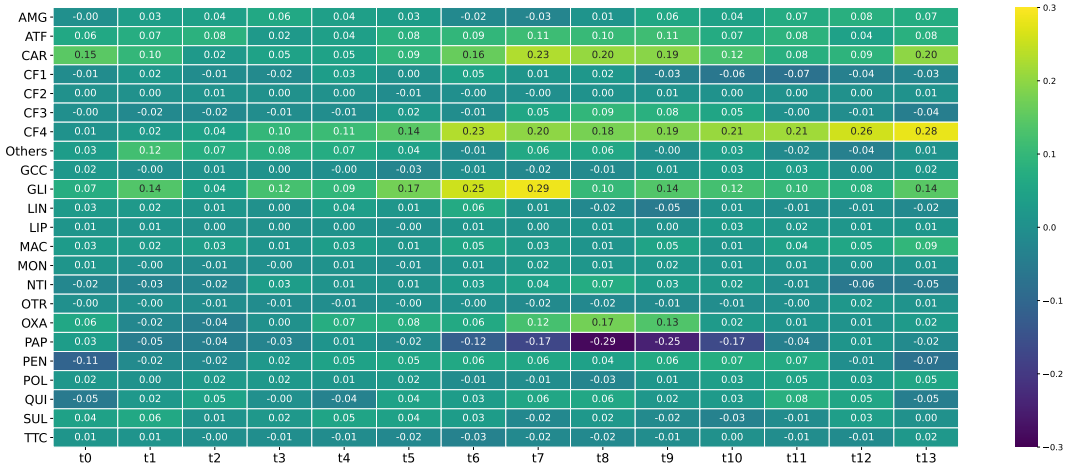
[1]Tests were based on nasopharyngeal samples and used nucleic acid transcription-mediated amplification SARS-CoV-2 assay (Procleix® assay in Procleix Panther® system, Grifols Diagnostic Solutions Inc., San Diego, CA).



(a)

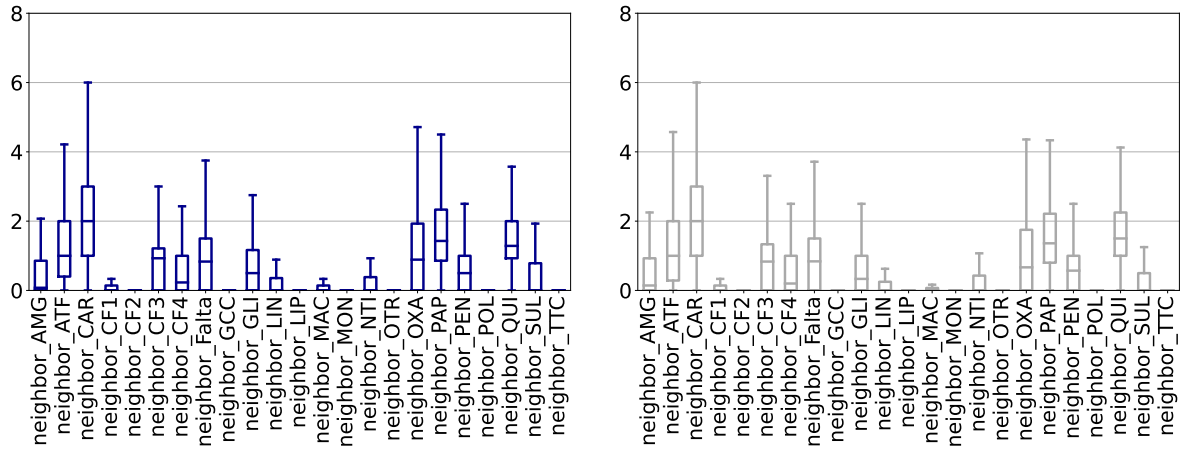


(b)



(c)

Figure 3.12: Heatmaps representing the percentage of neighbors taking each family of antibiotics over time for (a) AMR patients, (b) non-AMR patients, and (c) the comparison between AMR and non-AMR groups. Rows represent antibiotic families, and columns represent time steps.



(a) (b)

Figure 3.13: Boxplot with the drug intake of neighbors of AMR patients represented in the left panel, and (b) AMR neighbors of non-AMR patients represented in the right panel.

These features include four static variables: demographics, comorbidities, prior regular medication, and initial symptoms. Additionally, dynamic variables concerning drug administration during hospitalization are included. A schematic representation of this data for an individual patient is provided in Fig. 3.14. Demographics include age, gender, and the SAPS-3. Comorbidities related to high mortality, such as - smoking, obesity, diabetes, hypertension, chronic obstructive pulmonary disease (COPD), hypothyroidism, HIV, heart disease, transplantation, metastatic cancer, and blood cancer - are included. Pre-infection regular medications cover angiotensin-converting-enzyme inhibitors (AC:R), angiotensin II receptor blocker (AR:R), insulin (IN:R), corticosteroids (CO:R), and immunosuppressants (IM:R). Recorded symptoms before hospital admission include fever, cough, diarrhea, and dyspnea.

Table 3.3 shows statistics of the previously presented features. Numeric features (age and SAPS-3) are expressed as mean $\pm$ standard deviation. Binary features are quantified by patient counts (with percentages in brackets) where the feature is present. The initial row details the total, deceased, and surviving patient counts. A notable observation among non-survivors is the elevated incidence of smoking and heart disease comorbidities, critical as many developed acute respiratory distress syndrome. Also, symptoms like dyspnea, diarrhea, and cough were more prevalent in the deceased than in the surviving patients.

We follow this section by examining the drug administration for COVID-19 patients, focusing on daily patterns over 24-hour cycles. We use MTS to track the administration of several specific drugs at daily intervals. For each 24-hours, we record whether patients received any of

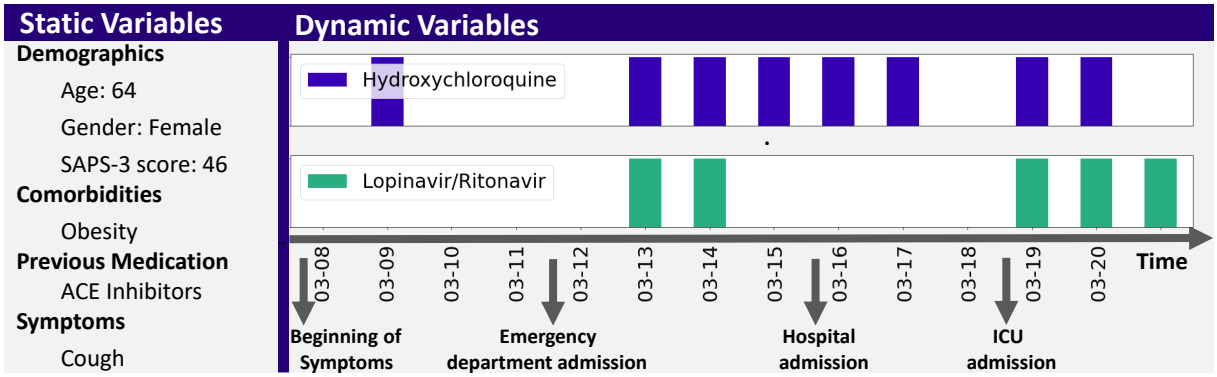


Figure 3.14: Description of the data available for a particular patient.

the following drugs: anakinra (AN:D), azithromycin (AZ:D), baricitinib (BA:D), chloroquine (CH:D), corticosteroid (CO:D), human immunoglobulin (HU:D), hydroxychloroquine (HY:D), imatinib (IM:D), interferon beta-1b (IN:D), lopinavir/ritonavir (LO:D), remdesivir (RE:D), and tocilizumab (TO:D). These drugs can be grouped into immunosuppressors (anakinra, baricitinib, and tocilizumab); antivirals (lopinavir/ritonavir and remdesevir); antimalarials licensed by the Food and Drug Administration during the first wave of the pandemic (chloroquine and hydroxychloroquine); corticosteroids (dexamethasone); antitumor drug (imatinib); immunostimulant (interferon beta-1b); and macrolide antibiotic (azithromycin).

The heatmaps presented in Figure 3.15 provide a comprehensive view of a 30-day treatment period for patients. Subfigures (a) and (b) begin with the onset of symptoms and extend to the 29th day. Each heatmap row represents a different drug administered, with the final row indicating the daily patient count. The data reveals that a significant portion of patients are hospitalized for less than three weeks. By the 29th day ($t = 29$), many have either succumbed to their condition or been released from the hospital, resulting in fewer ICU patients. Heatmaps also show significant differences in medication usage — corticosteroids, hydroxychloroquine, and lopinavir/ritonavir — between deceased and non-deceased patients. As pointed out earlier, it is apparent that the number of active patients decreases with time. The first drops (days 12 to 18) are likely due to patients who die, while later reductions can be attributed to patients who die and those who recover and are discharged to the ICU.

Subfigures 3.15 (c)-(j) offers a detailed view of the drug treatments given at various stages of hospitalization for both patients who survived and those who did not. Here, $t = 0$ marks the start of each interval, reflected in the corresponding heatmaps. More concretely, subfigure (c) and (d) specifically illustrate the medication regimes from symptom onset to emergency department admission for both deceased and non-deceased patients. Survived patients typically experienced a longer duration between symptom onset and hospital admission. Even

	Abbreviation	All Pat.	Dec. Pat.	Non-Dec. Pat.
Demographic				
Gender	GE:S	66 (68.0%)	28 (71.8%)	38 (65.5%)
Age	AG:S	62.6 ± 8.8	64.5 ± 7.5	61.4 ± 9.3
SAPS-3	SA:S	54.1 ± 11.1	57.7 ± 9.9	51.8 ± 11.3
Comorbidities				
Diabetes	DI:S	29 (29.9%)	10 (25.6%)	19 (32.8%)
Smoking	SM:S	13 (13.4%)	8 (20.5%)	5 (8.6%)
Metastatic cancer	MC:S	6 (6.2%)	2 (5.1%)	4 (6.9%)
Hypertension	HY:S	49 (50.5%)	19 (48.7%)	30 (51.7%)
Transplanted	TR:S	1 (1.0%)	1 (2.5%)	0 (0.0%)
Hem. Cancer	HC:S	4 (4.1%)	4 (10.3%)	0 (0.0%)
Obesity	OB:S	32 (33.0%)	14 (35.9%)	18 (31.0%)
Hypothyroidism	HP:S	13 (13.4%)	5 (12.8%)	8 (13.8%)
COPD	CO:S	18 (18.6%)	7 (17.9%)	11 (18.9%)
Heart disease	HD:S	20 (20.6%)	13 (33.3 %)	7 (12.1%)
HIV	HI:S	2 (2.1%)	0 (0.0%)	2 (3.4%)
None / Others	NC:S	18 (18.6%)	6 (15.4%)	12 (20.7%)
Symptoms				
Cough	CU:S	68 (70.1%)	29 (74.3%)	39 (67.2%)
Fever	FE:S	80 (82.5%)	31 (79.5%)	49 (84.5%)
Dyspnea	DY:S	61 (62.9%)	28 (71.8%)	33 (56.9%)
Diarrhoea	DR:S	20 (20.6%)	11 (28.2%)	9 (15.5%)
None / Others	NS:S	4 (4.1%)	1 (2.6%)	3 (5.2%)
Regular Medication				
ARA2	AR:R	13 (13.4%)	8 (20.5%)	5 (8.6%)
ACE inhibitors	AC:R	27 (27.8%)	7 (18.0%)	20 (34.5%)
Corticosteroid	CO:R	5 (5.2%)	3 (7.7%)	2 (3.4%)
Insulin	IN:R	8 (8.2%)	3 (7.7%)	5 (8.6%)
Immunosuppressants	IM:R	7 (7.2%)	2 (5.1%)	5 (8.6%)
None / Others	NM:R	51 (52.6%)	21 (53.8%)	30 (51.7%)

Table 3.3: Statistics and abbreviations for demographic variables, comorbidities, regular medication, and symptoms across all patients, deceased patients, and non-deceased patients. For numeric features, the mean $\pm$standard deviation is provided, while for binary features, the count of patients and percentage (in parentheses) are displayed.

though only a few patients received medication in this phase, two clear trends stand out: 1) non-deceased patients often received corticosteroids, and 2) the extended use of hydroxychloroquine

and lopinavir/ritonavir was noticeable. Subfigures 3.15 (e)-(f) show the medications given during the emergency department phase. Here, the volume of drug administration is substantially higher than during the initial "Symptoms Interval", despite patients spending less time in the emergency department. The drug treatments administered during the hospitalization phase are shown in Figs.3.15(g)-(h). Patients continue to receive corticosteroids, hydroxychloroquine, and lopinavir/ritonavir, but the prescription of chloroquine drops off. This reflects a shift in treatment strategies early in the pandemic, with hydroxychloroquine gradually replacing chloroquine. Finally, subfigures 3.15 (i)-(j) illustrate the pattern of drug administration during the "ICU Stay Interval". The diversity and frequency of medication use increase significantly, indicating more aggressive treatment strategies for critically ill patients. A wide variety of drugs, including antibiotics like anakinra, baricitinib, imatinib, and tocilizumab, are used. Interestingly, chloroquine is absent in ICU treatments, likely because it was only used in the pandemic's early stages. Notably, patients who did not survive tended to receive more intensive drug treatments such as anakinra, corticosteroids, hydroxychloroquine, and lopinavir/ritonavir, suggesting a link between the severity of the illness and the intensity of the treatment.

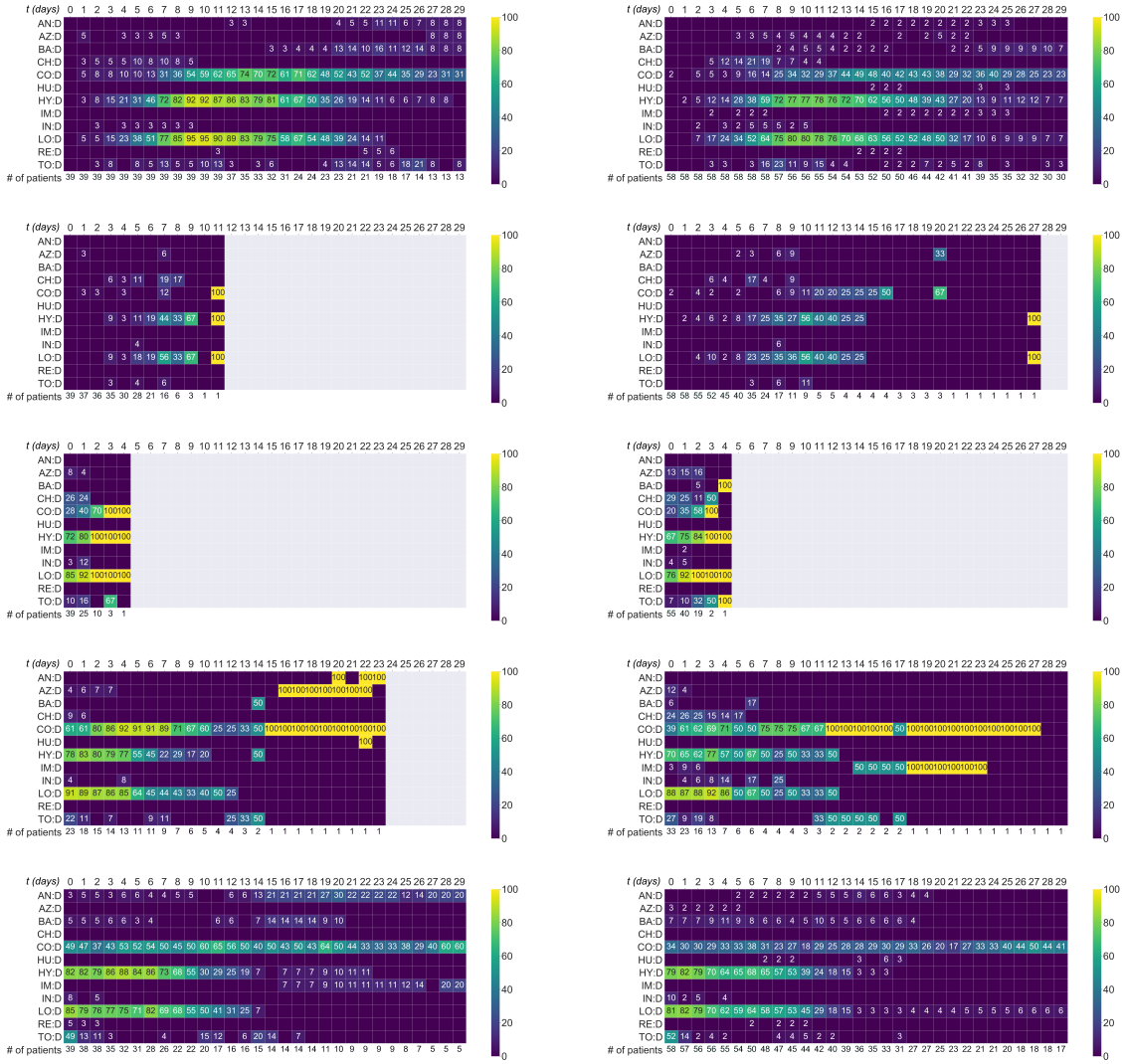


Figure 3.15: Heatmaps illustrating drug treatment patterns during different intervals: (a)-(b) from symptom onset to discharge within the first 30 days; (c)-(d) during the "Symptoms Interval"; (e)-(f) within the "Emergency-Department Interval"; (g)-(h) during the "Hospital Stay Interval"; and (i)-(j) within the "ICU Stay Interval" for both deceased patients (left panels) and non-deceased patients (right panels). In each heatmap, t=0 marks the beginning of the corresponding interval. The bottom row displays the number of patients per day, while the remaining cells show the percentage of patients receiving the indicated drug relative to the total number of patients on the specified day.

Chapter 4

Interpretable Data-Driven Modeling for Early Prediction of Antimicrobial Multidrug Resistance

4.1 Introduction

This chapter introduces methodologies that integrate interpretable DL with signal processing to predict AMR early using MTS data. We have employed trustworthy models for AMR prediction to address the critical need for both accuracy and interpretability in ICUs. These models integrate FS with interpretable RNNs developed to handle irregular clinical MTS, including patient treatments and ICU environmental factors.

Considering the challenge of data imbalance in ICUs, where AMR cases are infrequent, our models incorporate novel techniques to balance the data. We have also used advanced solutions for handling missing values in MTS. Additionally, we enhance our models with SHAP-based post-hoc interpretability, which has received positive validation from clinicians for its understandability and trustworthiness [107]. SHAP helps clarify how different features interact in the model, boosting clinician confidence and supporting better decision-making in controlling AMR in ICUs.

4.2 Methods

The methods section of this study is designed to describe the approaches used to analyze and use our data. Subsection 4.2.1 describes the criteria and techniques employed to identify the most relevant variables for our analysis. Following this, Subsection 4.2.2 outlines the methodologies we applied to handle MTS, emphasizing the use of advanced DL techniques to ensure robust modeling. Lastly, Subsection 4.2.3 explains how surrogate models were utilized to extract meaningful insights from the complex DL models, enhancing the interpretability of the results. These subsections provide a detailed view of the framework employed and support our research findings.

4.2.1 Feature Selection

The use of FS techniques is a useful strategy in ML, especially when working with limited patient datasets. In medical applications, such as disease diagnosis, the volume of patient data available for model training can often be limited due to various factors including privacy concerns, the rarity of certain conditions, or logistical challenges in data collection [108]. In these contexts, FS techniques play an indispensable role.

FS methods aid in reducing the input dimensions by identifying and retaining the most relevant features from the dataset. This process is critical because high-dimensional data can lead to model overfitting [109], where the model becomes excessively adapted to the training data, losing its ability to generalize to new, unseen data [110]. By reducing the number of input features, FS not only simplifies the computational requirements but also enhances the model's generalizability. This is particularly important in medical applications where models need to perform accurately and reliably across diverse patient populations and conditions [111].

Moreover, the dimensionality reduction achieved through FS does not necessarily imply a significant loss of information. Advanced FS techniques are designed to preserve the most informative aspects of the data, ensuring that the essential characteristics that contribute to accurate predictions are retained. This aspect is crucial in medical settings where every piece of data can be vital for accurate diagnosis and treatment planning.

In this chapter, three FS techniques have been implemented, including Conditional Mutual Information (CMI), Group Least Absolute Shrinkage Selection Operator (GLASSO), and Confidence Intervals with Bootstrap (CIB).

Conditional Mutual Information

Mutual information, a technique that originated from information theory, is increasingly recognized as a powerful tool in data analysis, particularly in FS and ML [112]. Fundamentally, MI quantifies the amount of information that one variable holds about another, thereby measuring the degree of dependence between them [113, 114]. This metric, captures both linear and non-linear relationships, making it a versatile tool for analyzing complex datasets. Moreover, mutual information's non-parametric nature makes it applicable to a wide range of data types, including continuous, discrete, and mixed features. Overall, mutual information serves as a robust tool for uncovering intricate relationships within data, enhancing the interpretability and performance of ML models.

In this dissertation, the strategy is to apply a feature selection (FS) scheme that maximizes the Conditional Mutual Information (CMI) between the chosen features and the target label y. The notion of CMI stems from the principles of Shannon entropy. To be mathematically precise, with $\mathscr{X}$ denoting the set of values the (discrete) random variable X can take, the entropy of X is defined as $\mathbb{H}(X) = -\sum_{x\in\mathscr{X}} p(x)log(p(x))$, where $p(x)$ is $Pr\{X = x\}$. When two random variables (X and Y) are present, two different generalizations of entropy can be defined. One is the joint entropy, which is defined as $\mathbb{H}(X,Y) = -\sum_{x\in\mathscr{X}}\sum_{y\in\mathscr{Y}} p(x,y)log(p(x,y))$, with $p(x,y) = Pr\{X = x, Y = y\}$. The second one, which is the most relevant one in the context of FS, is conditional entropy, which is defined as

$$\mathbb{H}(X|Y) = -\sum_{x\in\mathscr{X}}\sum_{y\in\mathscr{Y}} p(x,y)log(p(x|y)), \tag{4.1}$$

with $p(y|x) = Pr\{Y = y|X = x\} = Pr\{X = x, Y = y\}/Pr\{X = x\}$. The MI between X and Y measures the shared information between both variables, and is expressed as

$$\mathbb{I}(X,Y) = \mathbb{H}(X) - \mathbb{H}(X|Y) = \mathbb{H}(Y) - \mathbb{H}(Y|X) = \mathbb{I}(Y,X). \tag{4.2}$$

More specifically, the MI above measures how much information the variable X contains about variable Y.

With all this notation and hand, we are ready to define the CMI as the expected value of the MI of two random variables given a third random variable [115, 116], so that

$$\mathbb{I}(X,Y|Z) = \mathbb{H}(X,Z) - \mathbb{H}(Y|Z) - \mathbb{H}(X,Y,Z) - \mathbb{I}(Z). \tag{4.3}$$

CMI is a widely-used metric for carrying out FS. The goal of CMI-based FS is to obtain the set $\mathscr{D}' \subseteq \{1,2,...,D\}$ of D' features that maximize the CMI between the reduced input $\mathbf{X}^{\mathscr{D}'}$

and the associated label y. To avoid the extremely complex process of solving the optimization problem in its full form, we use a simpler, step-by-step method. In this method, we employ an iterative unidimensional optimization of the CMI metric, selecting the most informative feature that is not already included in $\mathscr{D}'$ at each step. Additionally, when calculating the value of $\mathbb{I}\left(y, \underline{\mathbf{x}}^d \,\middle|\, \{\underline{\mathbf{x}}^{d'}\}_{d' \in \mathscr{D}'}\right)$ from the data, it is crucial to consider that the variables $\underline{\mathbf{x}}^d$ are multidimensional. This means that $\mathscr{X}$ is the Cartesian product of the value sets for each of the entries of $\underline{\mathbf{x}}^{d'}$).

Group LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) is a statistical method for regularizing regression and classification problems, notably performing FS as detailed in Fonti's 2017 study [117]. At its core, LASSO implements a regularization process by introducing a penalty term to the traditional least squares method. This penalty is proportional to the absolute value of the regression coefficients, compelling the model to compress some coefficients towards zero. Consequently, this compression effect not only aids in reducing model complexity but also inherently performs FS. The features corresponding to the coefficients shrunk to zero are effectively eliminated from the model, simplifying the feature space and potentially improving model performance.

LASSO's efficacy in FS lies in its capacity to handle high-dimensional data, where the number of features significantly exceeds the number of observations [117]. In such scenarios, traditional regression models often suffer from overfitting, losing predictive power. LASSO counters this by selectively retaining features that contribute most significantly to the model's predictive accuracy while discarding redundant or irrelevant features. Moreover, the technique's flexibility allows it to be adapted across various domains, including finance, biostatistics, and ML [118, 119]. Having established the fundamental principles behind LASSO's effectiveness in FS, especially in high-dimensional data scenarios, let's delve into how LASSO identifies the most informative features through its optimization process.

Let's initially concentrate on static variables $\mathbf{z}_i \in \mathbb{R}^G$ and that all the entries of $\mathbf{z}_i$ are numerical. The objective of LASSO is then to determine the optimal value of $\alpha \in \mathbb{R}^G$ to minimize the cost

$$\min_{\alpha \in \mathbb{R}^G} \frac{1}{2} \sum_{i=1}^{I} \left(y_i - \mathbf{z}_i^\top \alpha \right)^2 + \lambda \sum_{g=1}^{G} |\alpha_g|, \tag{4.4}$$

where $\|\alpha\|_1 = \sum_{d=1}^{D} |\alpha_d|$ is the ℓ_1 norm of α, and $\lambda > 0$ serves as a regularization parameter. The cost function integrates a data-fitting term alongside a regularization term that penalizes the coefficients, effectively reducing some to zero. LASSO will automatically select the most

informative features by minimizing the cost function while eliminating those that are redundant or unnecesary. Consequently, the idea of using LASSO for FS involves fitting the model and then considering only the features g with a coefficient α_g different from 0.

The LASSO technique is applicable to static data; however, in this dissertation, we also address MTS, necessitating a specialized variant of LASSO designed for handling matrices, known as Group LASSO [120, 121]. This modification is essential for dealing with grouped input features, where each group is either retained entirely as relevant or completely discarded. Initially, we define vectors $\alpha^d = [\alpha_1^d, \alpha_2^d, ..., \alpha_I^d]$, each corresponding to the T time steps recorded for feature d. With D such vectors in total, we aim to determine DT coefficients. The optimal regularized regressor for MTS variables is obtained by solving

$$\min_{\{\alpha^d \in \mathbb{R}^T\}_{d=1}^D} \frac{1}{2} \sum_{i=1}^{I} \left(y_i - \sum_{d=1}^{D} (\underline{\mathbf{x}}_i^d)^\top \alpha^d \right)^2 + \lambda \sum_{d=1}^{D} \|\alpha^d\|_2, \tag{4.5}$$

where we recall that $\underline{\mathbf{x}}_i^d$ is the vector collecting the entries of the d-th row of $\bar{\mathbf{X}}_i$, and $\|\alpha^d\|_2 = ((\alpha_1^d)^2 + ... + (\alpha_W^d)^2)^{1/2} \geq 0$ is the ℓ_2 norm of α^d. The optimization process described aligns with that in Eq. (4.4), but adjusted to accommodate the multidimensional input by replacing $|\alpha_d|$ with $\|\alpha^d\|_2$. This way, if the optimal solution sets $\alpha_*^d = [0, 0, ..., 0]^\top$, then the d-th row of matrices $\{\bar{\mathbf{X}}_i\}_{i=1}^I$ is not selected [120]. By incorporating a binary cross entropy cost and employing a logistic regressor, the formulations in (4.4) and (4.5) can be modified to address classification challenges.

Confidence Intervals with Bootstrap

Bootstrap resampling, a non-parametric statistical technique, plays a pivotal role in estimating the distribution of a statistic, such as the mean, from a given sample [122, 123]. This method involves random sampling with replacement, as detailed by Hastie et al. (2009) [124], simulating the process of drawing multiple samples from the same population. By resampling from the original dataset repeatedly, Bootstrap generates numerous simulated samples, enabling the estimation of a statistic's distribution. This approach is invaluable for small sample sizes or when the underlying population distribution is unknown.

Moreover, Bootstrap resampling facilitates the computation of the standard error and the Confidence Interval (CI) for the estimated statistic, providing a measure of precision and reliability of the estimate [125]. The CI provides a range that likely contains the true parameter, offering a numerical indication of how uncertain the estimated statistic is.

In the context of data science and FS, the flexibility of the Bootstrap method becomes par-

ticularly valuable [126]. One of its applications is in hypothesis testing, where it is employed to assess the importance of features, especially in scenarios where traditional assumptions about data distribution (e.g., normality) may not hold. Bootstrap resampling can test hypotheses on feature importance or relevance without relying on parametric assumptions about the data distribution. This is particularly beneficial when dealing with complex or high-dimensional data where the underlying distributions are either intractable or unknown. The non-parametric nature of Bootstrap allows it to be applied across a wide range of data types and structures, making it a versatile tool in the FS process.

Consider two populations, S_d representing positive patients and S_{nd} representing negative ones. Our objective is to ascertain if the difference between their mean values, μ_d for S_d and μ_{nd} for S_{nd}, is statistically significant. Instead of directly calculating $\Delta P = \mu_d - \mu_{nd}$ and comparing this value to a predetermined threshold, we employ a resampling bootstrap method. Firstable, we resample both S_d and S_{nd} populations R times with replacement. This process yields R new sets for each population, represented as $\{S_d^{(r)}\}_{r=1}^{R}$ and $\{S_{nd}^{(r)}\}_{r=1}^{R}$. Following the resampling, the mean of each variable for every resampled set is computed, resulting in R mean values each for both $\{\mu_d^{(r)}\}_{r=1}^{R}$ and $\{\mu_{nd}^{(r)}\}_{r=1}^{R}$. The next step involves calculating the difference between the means for each resample, generating a series of differences $\Delta P^{(r)} = \mu_d^{(r)} - \mu_{nd}^{(r)} \quad \forall r = 1, \cdots, R$. Using these differences, a histogram of ΔP is constructed, from which the 95% CI for ΔP, denoted as $\text{CI}_{\Delta P}$ is empirically determined. The final stage of the analysis involves interpreting the results within the framework of hypothesis testing. If the value 0 is within the computed $\text{CI}_{\Delta P}$, it indicates that there is no significant difference between the mean values of the two populations, thereby supporting the null hypothesis H_0. Conversely, if 0 does not fall within $\text{CI}_{\Delta P}$, it suggests a significant difference between the means, thereby endorsing the alternative hypothesis H_1, which posits that the variable in question is relevant and informative.

Enhancing Classical Feature Selection with a Multi-Method Voting Strategy

In predictive models, selecting relevant variables is critical for the development of robust and accurate models. Traditionally, this process relies on a single method, often leading to biased or suboptimal selections due to the limitations inherent in any technique [127]. To improve this process, a multi-method approach, combined with a voting strategy has gained popularity. This approach involves using multiple FS techniques, each applied independently to the dataset. The key is in the aggregation of results: each technique 'votes' for or against the inclusion of each feature. Variables that receive votes above a predetermined threshold are then selected as relevant.

The advantages of this strategy are manifold [128]: i) It significantly reduces the bias that

might be present in a single-method approach, as it integrates multiple perspectives; ii) the consensus-driven nature of the voting strategy ensures that only variables with strong evidence across different methods are chosen, leading to increased stability in model performance; and iii) often identifies a more informative subset of variables, enhancing the model's predictive accuracy. Additionally, it is flexible and adaptable to various types of data and modeling objectives.

Empirical studies have supported the effectiveness of this approach, demonstrating that a multi-method FS with a voting strategy can lead to more robust and accurate predictive models [129]. This approach presents a compelling alternative to traditional methods, offering a balanced and comprehensive pathway to identifying the most relevant predictors for robust modeling.

4.2.2 Processing and Modeling of Time Series Using Deep Learning

In the realm of statistical analysis, MTS presents a unique challenge due to its temporal dimensionality and potential non-stationarity. Traditional statistical models, while effective for certain applications, often fail to capture the complex, dynamic relationships inherent in time-dependent data. This shortcoming has led to the investigation of more sophisticated computing methods, particularly in the area of DL. These techniques, a subset of ML, have garnered considerable attention for their ability to process and learn from vast amounts of data. Its applications in time series analysis are particularly promising, offering a paradigm shift in how we approach, model, and interpret temporally structured data.

The superiority of DL models in handling time series data lies in their inherent architecture, which allows for the learning of temporal dependencies and patterns that traditional methods might overlook. These models can adaptively learn from the temporal structure of data, making them particularly adept at forecasting, anomaly detection, and feature extraction in time series datasets.

This subsection of the dissertation delves into the application of various DL architectures in time series analysis, each offering unique advantages and considerations: the Multilayer Perceptron (MLP), the RNNs, the Long Short-Term Memory (LSTM), and the Gated Recurrent Unit (GRU). Also, we present methodologies to deal with the complexity of MTS using the previous methods listed.

Multilayer Perceptron

The Multilayer Perceptron is a type of feed-forward neural network (NN) characterized by

its layered structure [130]. It typically includes three distinct types of layers: an input layer, one or more hidden layers, and an output layer. The input layer receives the initial data, which is then processed through subsequent hidden layers before reaching the output layer. The key computational unit in these layers is the neuron or node.

Each neuron in the hidden layers computes an output using a scalar non-linear activation function, which transforms a weighted sum of its inputs. The input to each neuron is a linear combination of the outputs from the preceding layer's neurons, represented as $\sum_{g=1}^{G} w^g z^g + b$, where w^g denotes the weight associated with the g-th neuron in the previous layer, z^g is the output of the g-th neuron, and b is the bias term. The non-linear function could be a Sigmoid, Hyperbolic Tangent, or Rectified Linear Unit, among others [131].

During the learning process, the weights and biases are adjusted to minimize a predefined cost function, which measures the difference between the network's prediction and the actual target values. This optimization is usually performed using stochastic gradient descent or its variants like Adam and RMSprop [132, 133]. The cost function $J(w,b)$ is generally non-convex, and the gradient descent update rule can be expressed as $w^{\text{new}} = w^{\text{old}} - \eta \nabla J(w,b)$, where η is the learning rate and $\nabla J(w,b)$ is the gradient of the cost function for the weights and biases.

MLPs are fully connected networks, meaning each neuron in one layer is connected to all neurons in the subsequent layer. This comprehensive connectivity allows MLPs to capture complex patterns and relationships in data. Theoretically, MLPs are universal approximators, as they can approximate any continuous function to a desired degree of accuracy given a sufficient number of neurons and layers [134]. This property is formally expressed in the Universal Approximation Theorem, which underscores the potential of MLPs in modeling non-linear mappings between input and output spaces, making them particularly suitable for a wide range of tasks, from regression to classification [135].

Overall, the MLP's architecture, consisting of multiple layers of neurons with non-linear activation functions, provides a powerful framework for learning from data. Its ability to learn non-linear relationships and its flexibility in architecture design make it a cornerstone model in the field of NNs and DL.

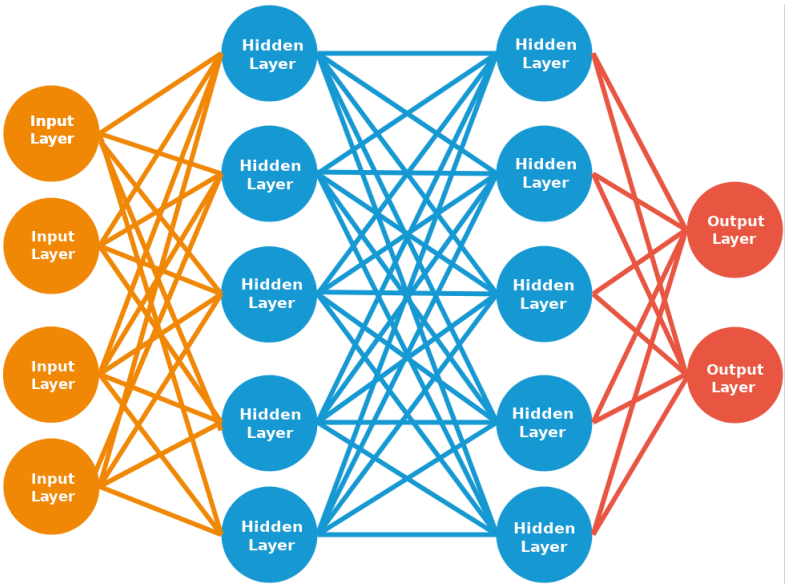


Figure 4.1: MLP architecture classic representation.

RNN Based Architectures

RNNs are a class of NNs designed for time series data, as they utilize internal states to maintain an "artificial memory" of previous inputs [136]. They stand apart from MLPs in their unique ability to form cyclical connections among neurons, a feature illustrated in Figure 4.2. In contrast to MLPs, which are limited to static mappings from input to output vectors, RNNs generate outputs that consider the entire history of previous inputs. This advanced functionality is enabled by the presence of internal cycles within the network's neurons, functioning as a form of 'artificial memory.' This memory component is critical, enabling the RNN to preserve information from past inputs in its internal state, thereby enhancing its sequential data processing capabilities. The main problem in large RNN networks is how the effect of an input on the internal layers, and eventually on the output of the network, can greatly decrease or increase as it goes through the network's recurrent connections. This phenomenon, widely recognized as the vanishing gradient problem, severely restricts the network's ability to effectively process long sequences of data. The vanishing gradient issue arises from the inherent properties of the backpropagation algorithm used in neural networks. As this algorithm propagates errors back through the network's layers over lengthy sequences, the gradients responsible for updating network weights either shrink to negligible levels (vanish) or grow excessively large (explode). This results in the network being unable to learn long-range dependencies effectively, and experiencing unstable, erratic training in the latter.

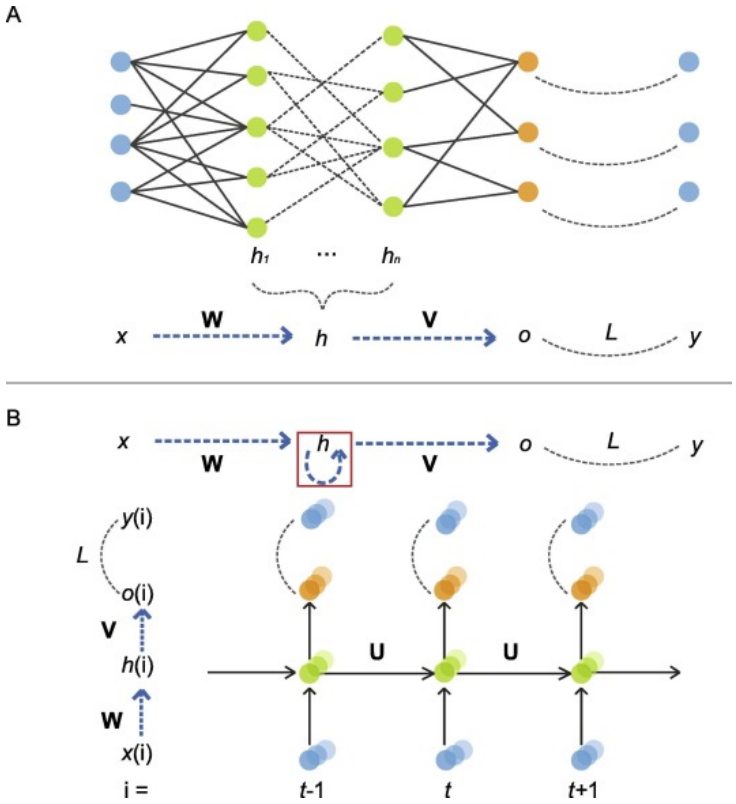


Figure 4.2: Schema of RNN: A. Common neural network unfold forms (top) and schema (bottom); B. An example of an RNN unfold form (top) and schema (bottom). One-time step delay is represented by the red square. Figure extracted from [4].

Long Short Term Memory Networks

LSTMs are an advanced type of RNNs designed specifically to process sequential data, playing a pivotal role in DL, particularly for time series analysis [137]. Developed to overcome the limitations of traditional RNNs, LSTMs are exceptionally well-suited for making predictions based on MTS, which makes them invaluable in various applications including speech recognition, language modeling, and even in the medical field for patient data analysis [138].

The effectiveness of LSTMs resides in their unique architecture that allows them to remember long-term dependencies in data sequences. As it was previously commented, traditional RNNs struggle with the vanishing gradient problem, where the network becomes unable to learn and retain information from earlier inputs as the sequence progresses [136]. LSTMs address this issue with their distinctive memory cell structure, which consists of a cell state (C^t) and three gates: the input gate (i^t), the forget gate (f^t), and the output gate (o^t) [139]. Central to LSTM functionality is the hidden state (h^t), a component of the LSTM cell that interacts with these to equip the network with short-term memory. The hidden state works together with the

cell state, that allows the network to remember or forget information over long periods. Both of them contribute to the cell's decision-making process about what to store, discard, and output at each time step, thereby effectively capturing temporal information that traditional RNNs often miss.

The forget gate layer is responsible for determining which information from previous states should be discarded. To get useful information from previous states the forget gate layer performs linear combination $f^t = \mathbf{w}_f^{(t,D)}[\mathbf{h}^{t-1} * \mathbf{X}^t] + b_f$, where $\mathbf{h}^{t-1}$ is the output from the previous state, $\mathbf{X}^t$ the current input, and $\mathbf{w}_f^{(t,D)}$ and b_f internal weights of the forget layer. This combination passes through a sigmoid activation, generating a mask that modulates the previous cell state $\mathbf{C}^{t-1}$ by element-wise multiplication.

The input gate layer selects information from the current input $\mathbf{X}^t$ to update the cell state. This gate operates through two mechanisms: one generating a candidate vector $\tilde{C}^t = \mathbf{w}_c^{(t,D)}[\mathbf{h}^{t-1} * \mathbf{X}^t] + b_c$, and the other, $i^t = \mathbf{w}_i^{(t,D)}[\mathbf{h}^{t-1} * \mathbf{X}^t] + b_i$, that determines how much of this new information will be used. As with the previous layer $\mathbf{w}_c^{(t,D)}$, $\mathbf{w}_i^{(t,D)}$, b_c, and b_i are internal weights of the input gate layer. The product of $\mathbf{i}^t$ and $\tilde{\mathbf{C}}^t$ is then added to the modified previous state, $\tilde{C}^{t-1} * \mathbf{i}^t$.

The output gate layer computes the neuron's final output, a filtered version of the cell state. This process involves two activation functions. Initially, a sigmoid function selects portions of the cell state to include in the output, denoted as $\mathbf{o}^t$. Subsequently, the cell state, processed through a *tanh* function, is element-wise multiplied with $\mathbf{o}^t$, yielding the final output $\mathbf{h}^t$.

In healthcare and clinical data analysis, LSTMs have shown significant promise. They are used to analyze patient data over time, predict disease progression, and assist in decision-making for treatment strategies. For instance, LSTMs can process a patient's medical history and provide insights into their future health risks or the likelihood of disease recurrence [140]. This capability is valuable in the context of chronic diseases or conditions that require ongoing monitoring and analysis [141].

Moreover, LSTMs' ability to handle MTS data is a significant advantage in clinical settings. They can analyze data from multiple sources – such as lab results, vital signs, and patient-reported symptoms – to create a comprehensive view of a patient's health status over time [142]. This multi-sourced analysis can lead to more accurate diagnoses, personalized treatment plans, and better patient outcomes.

In terms of interpretability, a critical aspect in healthcare, LSTMs offer both challenges and opportunities [143]. While their complex architecture can make understanding their decision-making process difficult, various techniques have been developed to interpret LSTM models.

Techniques such as attention mechanisms can highlight parts of the input sequence that are most influential in the model's predictions, providing valuable insights into how the model is processing the data.

In this thesis, we also explore Bidirectional Long Short-Term Memory (Bi-LSTM) networks, a variant of RNN architectures, as delineated in Schuster and Paliwal [144]. Bi-LSTMs employ a dual LSTM structure, with the first LSTM processing the MTS in a forward sequence and the second in reverse. This architecture mirrors classical time-varying stochastic process smoothing methods by utilizing past and future data for enhanced estimation accuracy. However, it's noteworthy that the increased parameter count in Bi-LSTMs necessitates a substantial volume of training data to capitalize on their potential performance advantages.

Gated Recurrent Unit

GRUs are another form of RNNs, similar to LSTMs, but with a simplified structure that often enables faster training and efficient learning, particularly in cases where the amount of data is limited [145]. Introduced as a variant of the LSTM, GRUs have been successfully applied in various domains like natural language processing, speech recognition, and time series prediction.

The key innovation in GRU's design is the integration of the forget and input gates into a single "update gate" [146]. This simplification reduces the complexity of the model without significant loss in performance, especially in tasks where long-term dependencies are less critical. The update gate ($\mathbf{z}^t$) decides how much of the past information needs to be passed along to the future. It functions similarly to LSTM's forget and input gates, determining the balance between the information transferred from previous states and new information added from the current input [147]. Another feature of GRUs is the "reset gate" ($\mathbf{r}^t$), which decides how much of the past information to forget. This gate works by modulating the impact of the previous hidden state ($\mathbf{h}^{t-1}$) on the current state's content. When the reset gate is close to 0, the model effectively forgets the previously computed state, allowing the GRU to adapt rapidly to changes in the input sequence's pattern.

The GRU's architecture also simplifies the process of updating the cell state. In each timestep, the hidden state ($\mathbf{h}^t$) is a combination of the previous hidden state ($\mathbf{h}^{t-1}$) and a candidate hidden state ($\tilde{\mathbf{h}}^t$). The update gate controls this combination, enabling the GRU to capture temporal dependencies effectively while mitigating the vanishing gradient problem common in standard RNNs [136].

In practical applications, GRUs have shown substantial promise. For instance, in natural language processing, GRUs have been utilized for tasks like language modeling and machine

translation, demonstrating capabilities comparable to LSTMs but often with less computational overhead [148]. Their simplified structure makes them particularly attractive for deployment in systems where computational resources are a limiting factor. Moreover, in the healthcare domain, GRUs offer the potential for analyzing time-sensitive patient data. While they may not capture long-term dependencies as effectively as LSTMs, their efficiency in learning patterns over shorter sequences can be advantageous in scenarios where rapid decision-making based on recent data is critical [147].

Despite their simpler architecture, interpreting GRUs remains a challenge, similar to LSTMs. Techniques such as visualization of activation patterns and attention mechanisms are often employed to gain insights into the network's decision-making process. These methods help in understanding the influence of different parts of the input data on the model's predictions, enhancing the interpretability of the GRUs in practical applications.

Addressing Data Imbalance in Classification Models

In many binary classification models, an equal number of samples for each class is often assumed [149]. This assumption, however, does not hold in numerous real-world scenarios, particularly in healthcare, leading to an imbalance where one class (e.g., AMR patients) is underrepresented compared to another (non-AMR patients). Such imbalances can result in models biased towards the majority class, affecting the generalization performance in the models trained [150].

To mitigate class imbalance, various strategies have been proposed in past works [63]. This study explores two effective methods: i) undersampling the majority class, and ii) employing asymmetric misclassification costs. The undersampling approach involves randomly discarding samples from the majority class to equalize the representation of both classes. In this dissertation, when undersampling is applied, we use the Binary Cross-Entropy (BCE) cost function during model training [151].

The cost-sensitive method differs by assigning greater penalties to errors in the minority class. This is achieved via the Balanced Binary Cross-Entropy (BBCE) function, an adaptation of the standard BCE. The BBCE cost, influenced by the weight parameter $\beta \in (0,1)$, is defined as:

$$-\frac{1}{I'}\sum_{i=1}^{I'}(\beta y_i \log(\hat{y}_i) + (1-\beta)(1-y_i)\log(1-\hat{y}_i)) \tag{4.6}$$

where I' denotes the patient count in the training set. A balanced dataset leads to $\beta = 0.5$,

reducing Eq.(4.6) to the BCE function. The value of β aligns with the proportion of majority class samples within the total dataset, aiding in adjusting cost sensitivity according to class representation. Following this approach, the value of β in this dissertation has been set as the number of samples of the majority class divided by the number of total samples.

Approaches to Deal with Missing Values in Multivariate Time Series

Missing values are a prevalent issue in real-world datasets, particularly in MTS. This challenge is especially notable in clinical settings where data collection is irregular, varying over time. Such missing values are often non-random, reflecting factors like patient health status or healthcare provider decisions [152]. Additionally, when utilizing windowed data, there are situations when the window is larger than the patient's record. This requires deciding how to fill the beginning or end of the record.

Typical methods to address missing values include zero-filling, linear interpolation, and statistical imputation [153]. Our approach, influenced by binary data characteristics and methodologies like those proposed by Lipton et al. [154] for RNN-based clinical data predictions, encompasses three strategies for managing missing values in $\bar{\mathbf{X}}_i$:

1. **Removing:** Exclusion of patients with missing data from the dataset. While this strategy simplifies the issue, it reduces training sample size, potentially affecting generalization. It is thus preferable in situations with a high number of training examples.

2. **Zero Padding:** Filling missing values with zeros, including the beginning or end of the record. This method is particularly prevalent with binary data, where a zero value often signifies a default state, such as the non-presence of a medical condition or absence of medication prescription.

3. **Masking:** Implementing advanced ML architectures that use a masking scheme to explicitly account for missing values. This approach is compatible with RNN-based architectures like GRU, LSTM, and Bi-LSTM. We adopt a modified version where each input sample is accompanied by a mask, indicating the positions of missing values in the input vector [155].

4.2.3 Surrogated Models to Gain Interpretability

The field of computational modeling has made it possible to examine complex systems with a high level of detail and precision. However, the increased complexity of these models often makes them difficult to interpret, which is essential for their reliability and usefulness in making

important decisions. This subsection focuses on surrogate models, a method developed to make complex models more understandable. Surrogate models are simplified versions of complex models that retain enough accuracy for practical use, making it easier to grasp, explain, and forecast the workings of complex systems. This subsection explains how surrogate models are created, used, and how effective they are, emphasizing their ability to maintain interpretability without substantially losing predictive accuracy. Through discussing various surrogate modeling techniques and their uses in different fields, this subsection highlights the importance of making complex computational models both sophisticated and accessible, aiming to strike a balance that aids in advancing our understanding.

SHapley Additive exPlanations

This dissertation presents a comprehensive evaluation of post-hoc interpretability methods for DNNs, with a primary focus on SHAP, as pioneered by Lundberg et al. in their paper [156]. SHAP's based on cooperative game theory, specifically the allocation of Shapley values to individual features of a dataset underscores its unique approach to model interpretability. This approach transcends the limitations of specific model architectures, positioning SHAP as a versatile, model-agnostic tool that can be applied to a wide range of ML models, including the often complex and opaque DNNs.

The essence of SHAP lies in its capacity to create local explanations for model predictions, achieved through a meticulous linear combination of binary variables. Those variables represent the presence or absence of each feature in the model. This process involves a detailed computation where the effect of adding or removing a particular feature is observed and measured [157]. This measurement reflects the contribution of each feature to the final prediction, akin to determining each player's contribution to the overall outcome in a cooperative game. This analogy is particularly apt for DNNs, as it allows for a granular analysis of how individual features affect the model's decision pathways [158].

One of the remarkable aspects of SHAP is its ability to provide a clear ranking of feature importance. This ranking is not arbitrary but is derived from the computed Shapley values, offering a quantifiable measure of each feature's influence on model predictions [159]. Such detailed insights are invaluable in domains where understanding the 'why' behind a model's decision is as crucial as the decision itself. In DNNs, where layers of computations and non-linear interactions often obscure the rationale behind outputs, SHAP's interpretability framework is a powerful tool for demystifying these complexities [160].

Additionally, the application of SHAP extends beyond mere academic interest. In practical scenarios, especially in sensitive fields like healthcare or finance, where decisions have signifi-

cant consequences, understanding the factors driving these decisions is essential. SHAP enables this understanding by not only highlighting the influential features but also by providing a pathway to scrutinize potentially biased or irrelevant features influencing the model. This level of transparency is crucial for building trust in AI systems and for ensuring that decisions made by these systems are fair, accountable, and aligned with ethical standards [98].

4.3 Experiments and Results

This section begins by outlining the specific design and procedures of the experiment. It then presents and analyzes the results from the FS process. Next, it evaluates the accuracy of different ML models tested in this research. The section ends with an analysis of the interpretability attributes of the developed models.

4.3.1 Experimental Setup

The database employed in these experiments was detailed in Section 3.1; therefore, this subsection will focus solely on describing the specific modeling of the experiments conducted in this chapter of the thesis. We define modeling as the set of decisions implemented to adapt the original dataset to the experiments at hand. The data modeling process encompasses several critical steps, such as conceptual design, logical design, and implementation.

The first decision in the modeling performed was to limit the years employed, from 2004 to 2020, inclusive, encompassing a total of 3,158 patients, with 433 identified as having AMR. Although the dataset extended to 2022, analyses were deliberately confined to the 2020 threshold, due to the significant impact of the COVID-19 pandemic on subsequent data, rendering post-2020 records less representative for the purposes of this study. The dataset presented in this chapter exclusively encompasses MTS data. It includes information on the medications administered to the patient, the MV, and features associated with the patient's neighbors. Notably, it excludes static features and does not cover previous cultures performed on the patient under study. The dataset comprises MTS with variable lengths, however necessitating a uniform input size for the ML models. To address this, a windowing technique was employed. This technique involves setting a window length, W, and defining a time interval $[t_i^{ini}, t_i^{end}]$ for each patient i, where $t_i^{end} = t_i^{ini} + W - 1$. The specific time interval varies per patient due to the asynchronous nature of the data.

A graphical representation of the temporal windowing process (with $W = 5$) for two patients

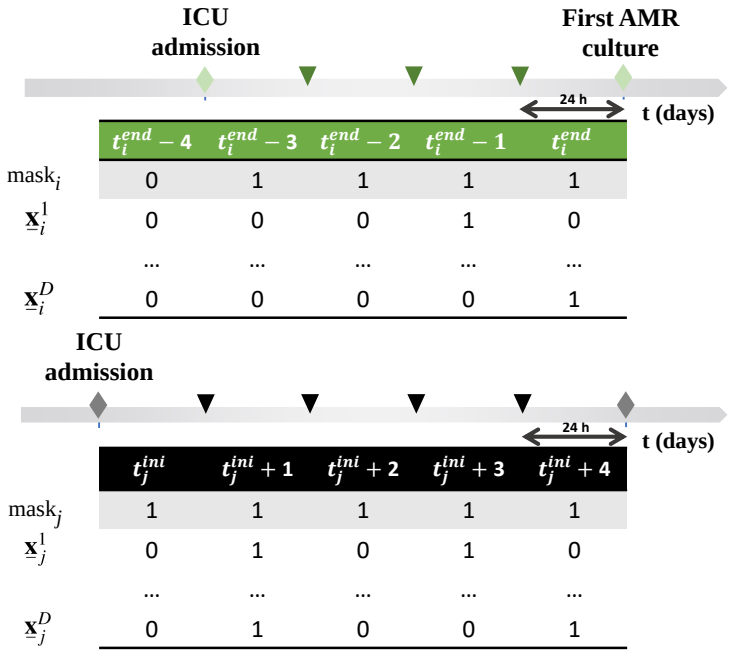


Figure 4.3: The construction of a temporal feature matrix within a specified time frame, which is segmented into five consecutive intervals, each spanning 24 hours. In the upper section of the figure, representing the AMR patient cohort, the term t_i^{end} signifies the time step associated with the first AMR culture for the i-th patient. Conversely, in the lower section, which illustrates the non-AMR patient cohort, t_j^{ini} indicates the admission time for the j-th patient.

(patient i, who is AMR and j who is non-AMR) is illustrated in Figure 4.3. The time series data consists of daily observations, with the final day of the window varying depending on the patient population. For patient i t_i^{end} corresponds to the day an AMR culture was detected. The previous $W-1$ days are then retroactively defined.

Four different window lengths were tested: $W=3$, $W=4$, $W=5$, and $W=6$. These were chosen based on extensive data analysis, revealing that initial AMR detection typically occurs within the first few days of a patient's ICU stay. Notably, 50% of AMR patients had a positive culture within five days of ICU admission, similar to the median stay duration of non-AMR patients (four days).

4.3.2 Feature Selection Results

Figure 4.4 presents a heatmap illustrating which features were selected by each of the three FS methods across four different window lengths. A green box in the heatmap indicates a feature was selected by a method, whereas a gray box shows it was not selected. To decide which

features were included in the refined set $\mathscr{D}'$, a dual-stage selection process was utilized. The first stage was to integrate insights from FS algorithms and the second one took into account the temporal nature of the data. Initially, the selection process used a "temporal voting" approach, where a feature was selected for a method by its selection frequency across different time windows. A feature had to be selected in at least two separate window lengths to be considered temporally robust. This step ensured that the importance of features was consistent over time, not just in a single, potentially anomalous, window. Subsequently, the process applied a "majority rule" criterion, requiring a feature to be selected by at least two different FS methods for inclusion in $\mathscr{D}'$. This step provided a more comprehensive evaluation of each feature's relevance by incorporating diverse methodological perspectives, thus increasing the reliability and applicability of the final selected features. This dual-stage approach effectively balances temporal stability and methodological consensus, strengthening the confidence in the research findings and their practical implications.

Figure 4.4 reveals that the technique utilizing CI from bootstrap resampling selected a larger number of features (40 out of 50) compared to the CMI or Group LASSO methodologies, which each identified 19 features. Group LASSO predominantly selected patient-related variables, whereas CMI favored features about the ICU environment. The consistency of Group LASSO's selections across varying time window lengths is particularly noteworthy. Upon aggregating results via a voting mechanism across the methods, a total of 26 features were ultimately chosen. This set includes 14 features linked to antibiotics administered to patients (AMG, ATF, CAR, CF1, CF3, CF4, GLI, NTI, OXA, PAP, PEN, POL, QUI, and Others), the MV, and 11 features associated with the ICU environment (number of patients, number of AMR patients, CARn, CF3n, GCCn, GLIn, MONn, PAPn, POLn, TTCn, and Others$_n$).

Considering feature importance as a means of enhancing explainability for early prediction of AMR, we delve into the clinical significance of these features for model training. All antibiotic families implicated in clinical criteria for AMR emergence were included. Notably, certain antibiotics, regardless of window length and FS method (including ATF, CF3, PAP, MV, the number of AMR patients, and CAR$_n$), were consistently identified as relevant. These findings have been corroborated by clinicians, underscoring their potential utility in constructing data-driven models.

Data Source	Strategies to handle imbalance	Strategies to handle missing values	Models	Accuracy	Specificity	Sensitivity	ROC AUC
Non-FS	Undersampling	Removing	**MLP**	64.15 ± 7.76	66.1 ± 11.08	53.1 ± 14.3	59.6 ± 3.52
			GRU	61.99 ± 3.99	62.91 ± 4.44	56.08 ± 4.21	59.50 ± 3.24
			LSTM	61.98 ± 4.32	62.92 ± 4.7	55.64 ± 11.28	59.28 ± 5.97
			Bi-LSTM	63.91 ± 6.28	65.65 ± 7.48	53.33 ± 6.86	59.49 ± 4.74
		Zero Padding	**MLP**	59.36 ± 2.26	59.15 ± 2.57	61.08 ± 4.5	60.11 ± 2.56
			GRU	59.74 ± 2.66	59.48 ± 3.71	61.1 ± 4.02	60.29 ± 0.75
			LSTM	59.14 ± 2.02	59.06 ± 3.18	59.84 ± 8.32	59.45 ± 3.14
			Bi-LSTM	57.99 ± 1.84	57.41 ± 2.19	61.73 ± 3.82	59.57 ± 2.16
		Masking	**GRU**	67.38 ± 2.59	68.91 ± 3.69	57.51 ± 7.01	63.21 ± 2.48
			LSTM	65.92 ± 1.79	67.38 ± 2.19	56.3 ± 1.98	61.84 ± 0.99
			Bi-LSTM	65.34 ± 2.74	66.92 ± 2.63	54.95 ± 3.06	60.94 ± 2.81
	BBCE	Removing	**MLP**	56.33 ± 6.22	54.0 ± 7.86	**71.52 ± 5.41**	62.76 ± 2.4
			GRU	57.78 ± 7.58	57.18 ± 10.33	62.42 ± 10.62	59.8 ± 2.07
			LSTM	55.54 ± 11.97	54.26 ± 14.95	65.12 ± 11.06	59.69 ± 3.27
			Bi-LSTM	55.38 ± 8.89	53.55 ± 11.45	68.75 ± 11.77	61.15 ± 1.95
		Zero Padding	**MLP**	55.47 ± 3.24	54.29 ± 3.15	63.57 ± 7.49	58.93 ± 4.66
			GRU	57.80 ± 4.58	56.12 ± 5.98	69.58 ± 6.4	62.85 ± 2.02
			LSTM	57.02 ± 3.58	55.71 ± 3.46	65.70 ± 4.74	60.70 ± 3.98
			Bi-LSTM	59.97 ± 7.31	59.57 ± 10.64	63.88 ± 14.82	61.73 ± 3.52
		Masking	**GRU**	67.03 ± 2.74	68.52 ± 3.92	57.51 ± 7.01	63.01 ± 2.35
			LSTM	60.86 ± 3.35	60.2 ± 4.12	65.67 ± 3.71	62.93 ± 1.60
			Bi-LSTM	59.52 ± 3.90	58.75 ± 5.33	65.01 ± 6.32	61.88 ± 1.49
FS	Undersampling	Removing	**MLP**	59.92 ± 2.97	60.19 ± 2.79	58.42 ± 5.79	59.31 ± 3.93
			GRU	60.32 ± 6.07	60.52 ± 6.66	59.16 ± 6.14	59.84 ± 5.03
			LSTM	64.24 ± 3.19	65.35 ± 3.71	57.12 ± 2.12	61.23 ± 1.95
			Bi-LSTM	60.9 ± 5.45	61.1 ± 6.65	59.17 ± 6.85	60.13 ± 3.91
		Zero Padding	**MLP**	63.11 ± 5.48	63.42 ± 6.9	62.14 ± 6.69	62.78 ± 2.55
			GRU	61.95 ± 2.88	62.26 ± 4.04	60.38 ± 6.59	61.32 ± 2.23
			LSTM	65.93 ± 1.71	66.64 ± 2.78	61.72 ± 7.32	64.18 ± 2.67
			Bi-LSTM	63.1 ± 5.38	63.36 ± 6.36	61.54 ± 4.89	62.45 ± 3.53
		Masking	**GRU**	64.08 ± 4.1	64.14 ± 5.85	64.16 ± 8.29	64.15 ± 1.65
			LSTM	**69.23 ± 2.28**	**70.79 ± 3.30**	59.41 ± 6.22	65.10 ± 2.18
			Bi-LSTM	68.62 ± 2.35	70.35 ± 2.69	57.18 ± 3.76	63.76 ± 1.99
	BBCE	Removing	**MLP**	57.91 ± 7.52	57.01 ± 9.54	65.34 ± 8.24	61.17 ± 2.85
			GRU	59.11 ± 4.37	57.58 ± 5.79	69.52 ± 5.81	63.55 ± 1.74
			LSTM	57.15 ± 6.06	55.75 ± 8.35	66.92 ± 9.91	61.34 ± 1.05
			Bi-LSTM	53.84 ± 11.01	51.33 ± 14.48	70.8 ± 12.48	61.07 ± 2.19
		Zero Padding	**MLP**	66.24 ± 2.32	66.89 ± 2.82	62.37 ± 5.27	64.63 ± 2.54
			GRU	58.01 ± 4.22	56.22 ± 5.13	69.68 ± 3.92	62.95 ± 2.56
			LSTM	60.81 ± 3.83	60.43 ± 5.01	63.45 ± 6.39	61.94 ± 2.29
			Bi-LSTM	55.59 ± 3.97	53.61 ± 4.93	69.19 ± 4.35	61.40 ± 1.27
		Masking	**GRU**	63.01 ± 2.93	61.95 ± 4.17	69.94 ± 5.75	65.95 ± 1.29
			LSTM	65.40 ± 3.94	64.88 ± 5.31	68.58 ± 6.43	**66.73 ± 1.80**
			Bi-LSTM	63.33 ± 2.47	62.98 ± 3.37	65.89 ± 4.04	64.44 ± 0.78

Table 4.1: Average performance (Accuracy, Specificity, Sensitivity, and ROC AUC) presented as mean ± standard deviation across 5 test partitions. The results are shown for neural networks trained on a 5-day window under various conditions: without FS and with FS in the first row; undersampling and BBCE to manage class imbalance in the second column; handling irregular MTS with "Removing," "Zero Padding," and "Masking" techniques in the third column; and employing MLP, GRU, LSTM, and Bi-LSTM as classifiers in the fourth column. The highest values for each metric are highlighted in bold..

4.3.3 Early Prediction of Antimicrobial Multidrug Resistance Using Neural Networks

Table 4.1 displays the mean and standard deviation of performance metrics (Accuracy, Specificity, Sensitivity, and ROC AUC) across five test partitions, comparing conventional NNs (exemplified by the MLP) and RNN models (including LSTM, GRU, and Bi-LSTM). These mod-

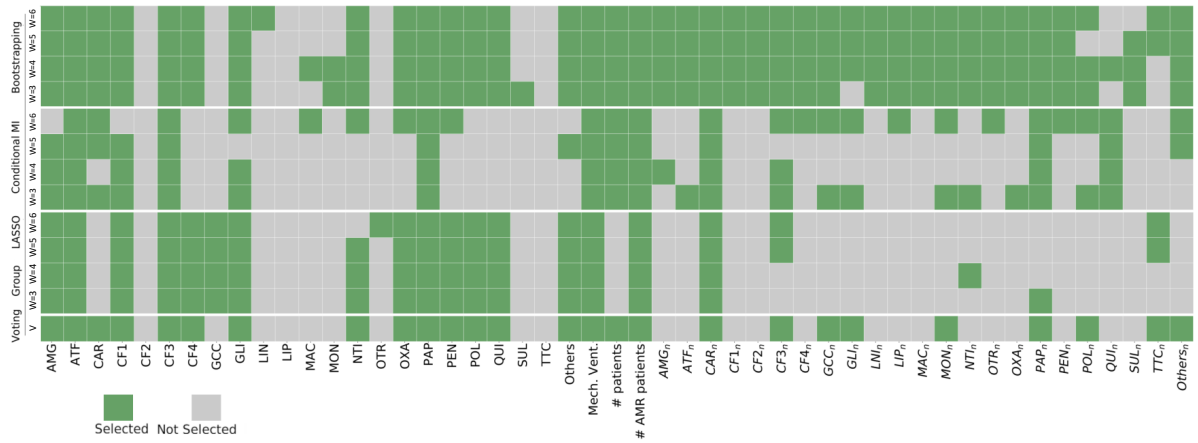


Figure 4.4: Feature matrix (columns) and FS methods (CIB, CMI and Group LASSO, distinguished by window length W). The selected features are denoted by green cells, and the non-selected features are depicted by gray cells.

els are evaluated based on: (a) different FS approaches, (b) class imbalance handling, and (c) managing missing values in MTS.

The results reveal that the FS procedure enhances model performance, with models utilizing FS data outperforming those trained on non-FS data in terms of Accuracy and Specificity. This is evidenced by the higher average ROC AUC values for FS models (62.09) as compared to non-FS models (60.84), suggesting that FS helps improve the precision and class differentiation capabilities of the models.

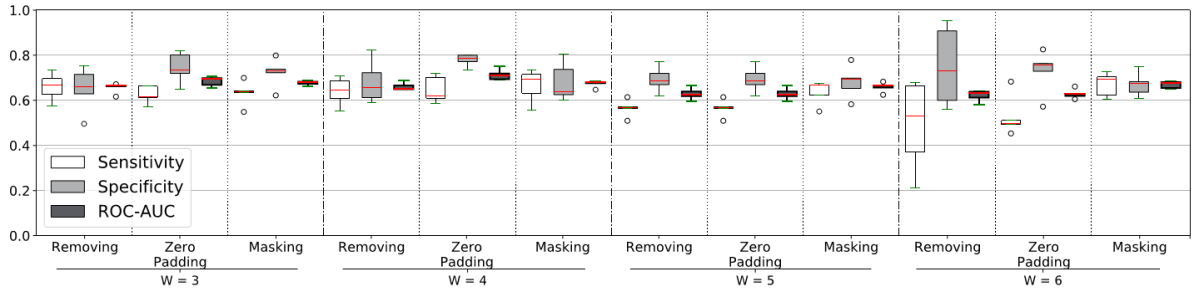
In addressing class imbalance, two strategies, Undersampling and BBCE, were compared. Undersampling achieved a higher average ROC AUC of 63.95, indicating its efficacy in some scenarios, particularly in increasing Sensitivity. Conversely, BBCE performed better in Specificity and ROC AUC in certain cases, underscoring the need to select the appropriate strategy based on the dataset characteristics and the objectives of the models.

Regarding the methods for handling missing values, the choice significantly impacts model performance, with no single approach uniformly outperforming the others across all models and metrics. However, "Masking" is the most effective achieving the highest mean ROC AUC (63.66) compared to "Removing" and "Zero Padding" (60.56 and 61.58, respectively).

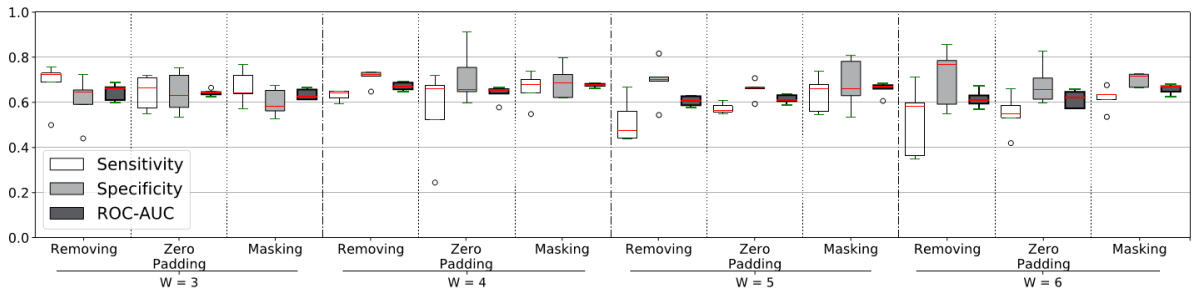
The comparative analysis of different classifiers reveals that LSTM models generally demonstrate a better performance. This is particularly evident when these models are trained with FS data, undersampling, and the "Masking" strategy. The highest overall performance is observed in LSTM models combined with BBCE and "Masking," reaching an ROC-AUC of 66.73%

and leading in Sensitivity (68.58%). However, the highest Accuracy (66.54%) and Specificity (64.88%) are achieved with the undersampling strategy, suggesting a trade-off between different performance metrics and the chosen strategies.

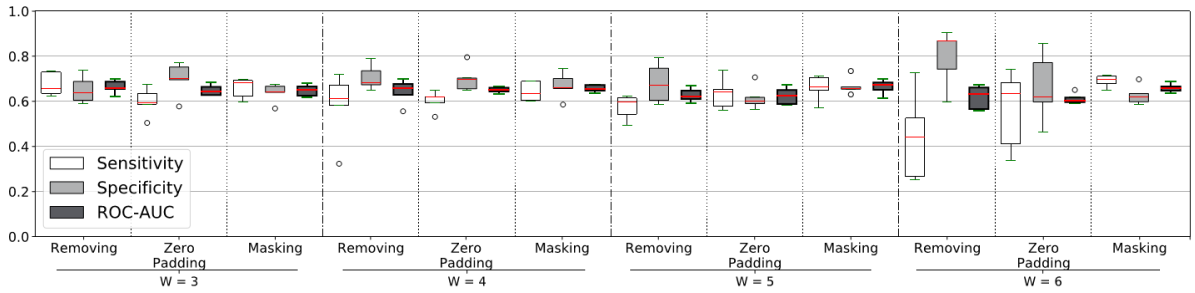
Further analysis of LSTM, GRU, and Bi-LSTM models using windowed modeling is shown in Figure 4.5, focusing on Specificity, Sensitivity, and ROC AUC. Performance varied with the window length, with four and five-day windows outperforming three and six-day ones. GRU and LSTM generally surpassed Bi-LSTM models, potentially due to the latter's complex architecture or the irrelevance of considering both past and future data for this classification task.



(a)



(b)



(c)

Figure 4.5: Boxplot analysis illustrating performance metrics (Specificity, Sensitivity, and ROC AUC) across 5 test partitions, with considerations for FS and the use of BBCE to address data imbalance. Varied window lengths ($W = 3$, $W = 4$, $W = 5$, and $W = 6$) and three MTS classifiers (a) GRU, (b) LSTM, and (c) Bi-LSTM are explored.

4.3.4 Interpreting Long Short-Term Memory Models Using SHapley Additive exPlanations Analysis

The previous section established the LSTM network with 26 input features and a window length of $W = 5$, combined with "Masking," as a high-performing model. However, the inherent complexity of LSTMs poses interpretability challenges. This section delves into a post-hoc interpretability analysis of the LSTM model using SHAP, aiming to unravel the model's working.

Initially, we examined the entire patient population using SHAP, focusing on an LSTM model incorporating $W = 5$, undersampling, and "Masking." Subsequently, we extended our analysis to individual patients across various window lengths ($W = 3,4,5,6$), maintaining the same model parameters. For each patient, we calculated the average Shapley values for all time steps, offering insights into the feature contributions to the model's predictions.

The resultant SHAP graph (Figure 4.6) visually represents these insights. In the graph, each dot symbolizes a patient, with the dot color reflecting the actual feature value and the x-axis position indicating the feature's impact on the model's output, as determined by the sum of Shapley values. A notable example is the positive correlation of high $Mech.Vent.$ values with increased model output, suggesting a higher likelihood of AMR in patients requiring $Mech.Vent.$ (as was previously indicated in Chapter 3). The top five influential features, including $Mech.Vent.$, ATF, CF1, the number of AMR patients, and GLI_n, align with clinical intuition and existing literature, underscoring the importance of controlling AMR germs and invasive devices in healthcare settings.

Further, Figure 4.7 presents a detailed SHAP analysis for four distinct patient types from our dataset: both AMR and non-AMR patients, differentiated by stays longer or shorter than five days. This analysis, considering different window lengths and the 26 selected features, highlights the variability in feature contributions across individual patients. Notably, features like the number of AMR patients, MV, and certain drugs (e.g., ATF, AMG, OXA) consistently emerge as significant across models. However, the effectiveness of model classification varies with window length, emphasizing the complexity of optimal window length selection. For instance, models with $W = 5$ and $W = 6$ accurately classify AMR patients with incomplete data, while $W = 3$ and $W = 4$ are more effective for non-AMR patients with complete data.

In conclusion, this SHAP-based interpretability analysis of the LSTM model offers valuable insights into the model's functioning, particularly in identifying key features influencing predictions. The differences observed across different patients and window lengths highlight that choosing the right model is complex and that personalized methods are crucial in medical

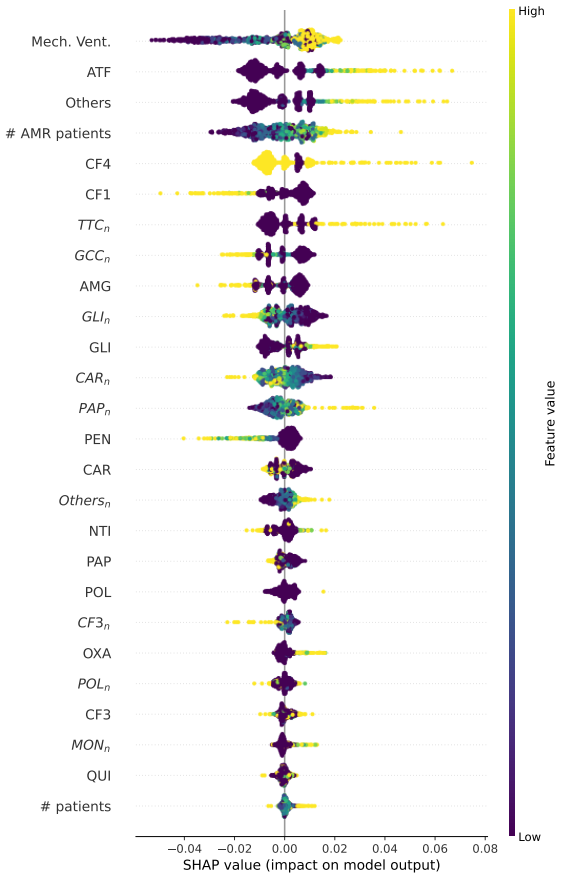


Figure 4.6: Distribution of Shapley values generated from the LSTM model with undersampling and "Masking" with the 26 features selected by FS.

settings.

4.4 Conclusions

This chapter's findings significantly contribute to the fight against AMR in ICUs. The alarming frequency of infections in ICUs, where 20-30% of admissions lead to AMR, highlights the critical need for early prediction methods. Through the development of DL models for clinical MTS data, we have taken a crucial step towards timely detection and treatment of infectious diseases, potentially reducing both mortality and healthcare costs. The complexity of clinical data has long made MTS modeling a daunting task, and while RNNs show promising classification performance, their lack of interpretability poses a significant challenge in clinical decision-making processes.

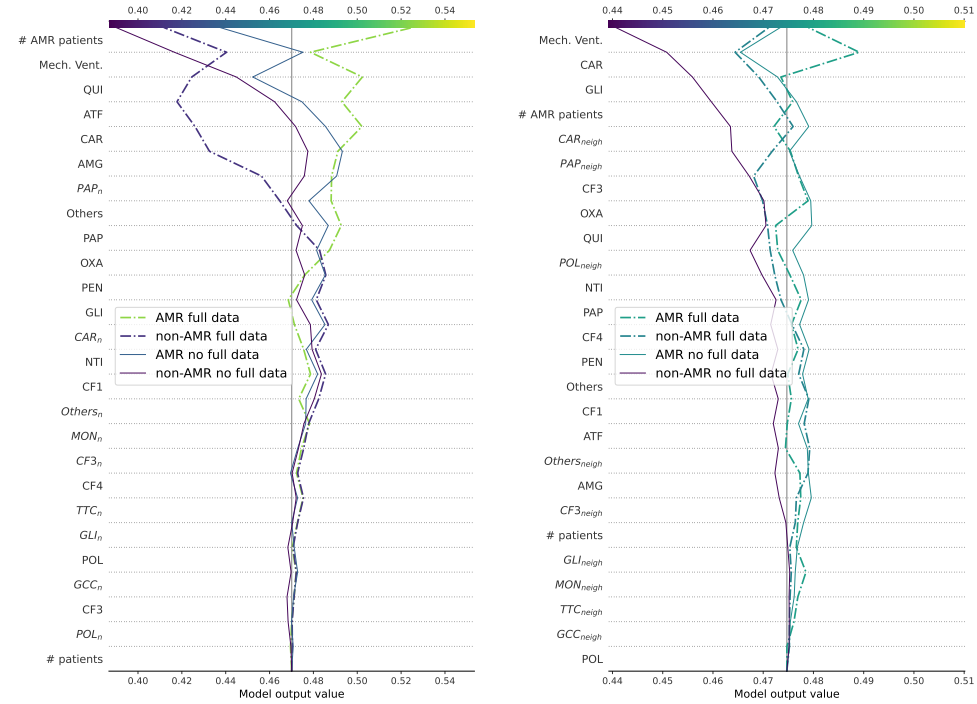
The use of FS techniques emerged as a pivotal strategy in our approach. Employing methods like Conditional Mutual Information, Group LASSO, and Confidence Intervals with Bootstrap, we efficiently managed high-dimensional clinical data, enhancing model performance and interpretability. The application of these techniques ensured that the models were not only computationally efficient but also robust and capable of generalization across diverse patient populations. We also presented a multi-method voting strategy to enhance classical feature selection, further refining the traditional FS methodology. This innovative approach, combining multiple FS techniques, led to a more comprehensive and reliable selection of relevant features.

Our innovative approach in handling data imbalance and missing values in MTS data further strengthened the reliability of our predictions. By implementing methods like undersampling, asymmetric misclassification costs, and masking strategies, we effectively addressed these common challenges in clinical MTS datasets.

The incorporation of SHAP for post-hoc interpretability provided a deeper understanding of the model's decision-making process. This aspect was crucial for gaining the trust of clinicians and ensuring the future practical applicability of our models in ICUs. The SHAP-based analysis offered insights into the relevance of features like mechanical ventilation and specific antibiotic treatments, aligning with clinical knowledge and highlighting the model's ability to capture essential predictive factors for AMR.

The performance of our models was rigorously tested across various settings, with a particular emphasis on LSTM networks. These tests confirmed the effectiveness of our methodologies in early detection of AMR, with the LSTM models exhibiting strong predictive capabilities across different scenarios.

However, to generalize these findings, future research should incorporate more MTS and demographic features. The next chapter will explore the development of interpretable neural network mechanisms that consider the significance of each time step, aiming to further increase clinician trust and model adoption in real-world healthcare settings.



(a) (b)

(c) (d)

Figure 4.7: Visualization of model output values and Shapley values for the LSTM model trained with undersampling and 'Masking,' utilizing 26 selected features, and varying window lengths: (a) $W = 3$; (b) $W = 4$; (c) $W = 5$; (d) $W = 6$. The gray vertical line signifies the base value of the SHAP models, while each colored line corresponds to an individual patient. Feature relevance is ranked and presented, with the top-ranked feature being the most relevant. 'Full data' represents patient stays longer than the respective window ($T_i > W$), while 'no full data' denotes cases where $T_i < W$.

Chapter 5

Multimodal Interpretable Models for Early Prediction of Antimicrobial Multidrug Resistance

5.1 Introduction

This chapter builds on the previous chapter's methodology by exploring the development of multimodal interpretable data-driven models and their pivotal role in the early prediction of AMR. It introduces a new framework, integrating multimodal architectures and interpretable mechanisms specifically tailored for analyzing complex MTS. By using multiple data types together, the multimodal approaches provide a more complete understanding of AMR dynamics compared to traditional unimodal models. Integrating the different data types enables a more holistic understanding of the AMR phenomenon, enhancing the accuracy and reliability of the models.

Additionally, the chapter emphasizes the critical role of model interpretability. In clinical settings, where AMR predictions can significantly influence patient treatment options, it is essential for models to not only be accurate but also transparent and understandable in their decision-making. This focus on both multimodal architecture and interpretability not only increases the effectiveness of the models but also enhances their trustworthiness, which is vital for their acceptance and use in clinical environments. Furthermore, the chapter provides an in-depth exploration of how these models are developed to handle the temporal complexities inherent in AMR data. MTS, with its sequential nature and time-dependent variations, presents

unique challenges. The chapter discusses the adaptation of these models to effectively capture and analyze temporal patterns and trends in AMR data, a crucial aspect for early and accurate predictions.

Therefore, this chapter presents a cutting-edge approach in the fight against AMR, merging multimodal data integration with interpretable modeling techniques. This approach marks a significant step forward in developing robust, accurate, and clinically applicable tools for early AMR prediction, thereby contributing to more effective antimicrobial stewardship and improved patient predictions.

5.2 Methods

This section is structured into three subsections. In subsection 5.2.1, we discuss the process of selecting appropriate features across various modalities of data. Subsection 5.2.2 explores the integration of heterogeneous data sources using state-of-the-art multimodal DL architectures to enhance model performance. Finally, subsection 5.2.3 focuses on the methodologies implemented to gain understandability in MTS analysis, a crucial aspect when dealing with DL models in a clinical environment. Each subsection is designed to provide a comprehensive overview of the methods employed, contributing to the robustness and clarity of our research outcomes.

5.2.1 Feature Selection in Multimodal Interpretable Data-Driven Models

Building upon the methodologies established in the previous chapter, this chapter explores the application of FS in the context of multimodal data. The efficacy of the FS approach previously presented, which demonstrated remarkable success, forms the foundation of the methodologies discussed herein. In this chapter, we largely reapply the established FS framework with a notable enhancement: the integration of a novel method known as Permutation Feature Importance (PFI). This method is specifically designed for handling MTS. By incorporating PFI, we aim to enhance our model's accuracy in predicting AMR.

As a reminder, our FS approach employs four distinct methods: CMI, GLASSO, CIB, and the newly introduced PFI. Each method contributes uniquely to our framework's robustness and precision. This chapter details the synergistic effects of these FS methods when applied in tandem, particularly in the realm of complex MTS. Integrating PFI into our proven FS framework enhances our model's capability to discern pivotal features. Such understanding is critical

for developing data-driven models that are not only accurate in early AMR prediction but also interpretable and applicable in real-world clinical settings.

Permutation Feature Importance

PFI is a FS methodology, which is specifically designed to work with models that have already been trained [161]. This method plays a crucial role in identifying key features that significantly influence the output of these complex models. PFI attempts to emulate the traditional recursive feature elimination methods, which require retraining the model with various feature combinations. PFI addresses this challenge not by retraining the model with different feature sets but rather by evaluating the impact of each feature on the model's performance in a more efficient manner. It does this by systematically altering or 'perturbing' individual features in the input data and observing the resultant effect on model performance. Originally, PFI was introduced in the realm of random forest classifiers, as delineated in [161, 162]. Since then, its applicability has been successfully expanded and generalized to various other model architectures, as evidenced in later studies [163, 164]. In our research, we have employed the PFI method across a diverse range of trained architectures, which we will detail in the following sections.

The implementation of PFI begins with training the ML model and then evaluating its classification performance. This evaluation is conducted using a set of validation samples and is based on a predefined performance metric. Following this, the method selects one feature at a time, for instance, the d-th feature. We then permute the value of the d-th feature with one feature chosen uniformly at random for each sample in the validation set, while all other features ($d' \neq d$) are left intact. This selective permutation strategy is critical as it alters the input data without changing the overall distribution of each feature. However, it effectively 'breaks' the relationships or patterns the model has learned, thus providing insights into the significance of each feature.

After the permutation of the d-th feature, the model's performance is reassessed using the modified validation set. This performance is then compared to the results obtained with the original, unaltered validation set. The underlying hypothesis driving this approach is that the permutation of highly relevant features will lead to substantial losses in the performance of the chosen metric, as delineated in [165]. This process is methodically repeated for each feature in the dataset ($d = 1, \ldots, D$). The outcome of this iterative process is the identification and selection of the D' most relevant features based on the extent of performance degradation observed with each feature's permutation.

5.2.2 Deep Learning Data Fusion Architectures

As previously discussed, "Data Fusion," or "Multimodality," integrates data from multiple sources, enhancing the performance and comprehensiveness in knowledge extraction [75, 76]. This process can be categorized into three approaches: early fusion, joint fusion, and late fusion [74].

Early fusion models integrate inputs from various modalities before feeding them to the model, utilizing techniques like concatenation, pooling, or gated units [166]. Conversely, joint fusion merges feature mappings from intermediate architectural layers, refining feature extraction through backpropagated loss during training [74]. Late fusion architectures, in contrast, aggregate predictions from multiple models to produce a single output. In the context of late fusion classification systems that process both dynamic and static data, the architecture can be segmented into three distinct blocks: one block generates a posteriori probabilities from static data, another computes a posteriori probabilities from MTS, and a third block merges these probabilities to produce the estimated label by the late fusion model. Integrating static data such as age or comorbidities with MTS data in the healthcare sector is critically important for clinical decision-making. This has led to the proposal of several new data fusion architectures in the clinical setting in recent years.

We present three innovative multimodal architecture designs optimized for handling MTS and static data. Two of them are based on joint fusion architectures — the Joint Heterogeneous Fusioner (JHF) and the First Hidden State Initializer (FHSI) — and two are based on late fusion — the Late Fusion Convex Optimization (LFCO) and the Late Fusion Logistic Regression (LFLR). These architectures will be further detailed later.

Joint Heterogeneous Fusioner

As previously discussed, the field of data fusion architectures is diverse. It is a common and often justified assumption that the fusion of different data types gets information from the target not in isolation but through complex cross-modality interactions. Joint fusion methodologies address this by modeling the interplay of features derived from intermediate representations. This process typically involves concatenating the marginal representations of intermediate features and then processing the resultant vector through fully connected layers, culminating in a task-specific output layer [167].

The present dissertation introduces the Joint Heterogeneous Fusioner, an innovative architecture designed to merge different data types, specifically static features and MTS. Our approach incorporates “prior knowledge” regarding the structure of the modality variables,

thereby optimizing the intermediate layer's functionality. For MTS data, we employ a GRU to capture and analyze temporal dynamics, which are then integrated into a unified feature vector. For categorical static variables, we utilize the widely-recognized entity embeddings [168], while linear transformations are applied to binary and numeric static variables to generate feature representations.

The integration of marginal representations is a critical step in data fusion, and various methodologies exist for this purpose. In our design, we opted for concatenation, given its widespread usage and interpretability advantages. The concatenated representations undergo a final transformation through a linear layer, followed by a sigmoid activation function, producing the ultimate fusion output. This architecture is designed to leverage and enhance the inherent synergies within heterogeneous data sources, offering a robust foundation for advanced analytical applications.

First Hidden State Initializer

The Temporal Fusion Transformer (TFT) is a novel model featuring numerous innovations. It has significantly outperformed existing benchmarks in MTS forecasting that incorporate both static data and MTS. Inspired by the success of the original TFT, we have adapted it to suit the specific structure of our framework, resulting in the creation of a new joint fusion multimodal architecture known as the "First Hidden State Initializer".

In the medical field, comprehending a patient's initial condition is pivotal for gauging their progression. This starting point significantly influences the medications and procedures administered to the patient throughout their treatment. Building on this principle, the FHSI architecture integrates static features to construct a context vector that enhances the initial hidden state of a GRU. To generate this context vector, the FHSI architecture incorporates an internal module called the Static Encoder (SE). Figure 5.1 illustrates the overall structure of the FHSI, with each component depicted in distinct colors. The internal workings of the FHSI are elaborated upon below:

- To construct the context vector $\bar{\mathbf{z}}_i^{cont}$, the SE initially maps static features into an embedding. Given the diversity of feature types—categorical, binary, and numerical—distinct embedding strategies are employed. Categorical variables $\mathbf{z}_i^{cat}$ utilize entity embeddings, a widely-adopted approach as outlined by Gugulothu et al. (2017)[168]. For binary and numeric variables, $\mathbf{z}_i^{bin}$ and $\mathbf{z}_i^{num}$, linear transformations are applied. This mapping process is depicted in light green in Figure5.1.

- In the next block within the SE, a variable selection mechanism is incorporated, visual-

ized in dark green in Figure 5.1. This mechanism generates a patient-specific vector by applying a Hadamard product to the initial input [169]. To enhance the model's adaptability for non-linear processing, we integrate a Gated Residual Network (GRN) within the variable selection network, drawing on established methodologies from Lim et al. (2021) and Tan et al. (2018)[170, 171].

- We employ the context vector $\bar{\mathbf{z}}_i^{cont}$ as the initial state for the GRU, (represented as $\mathbf{h}_i^0 = \bar{\mathbf{z}}_i^{cont}$ in the light blue box in Figure 5.1). The GRU block then updates the initial hidden state with the information contained in the MTS.

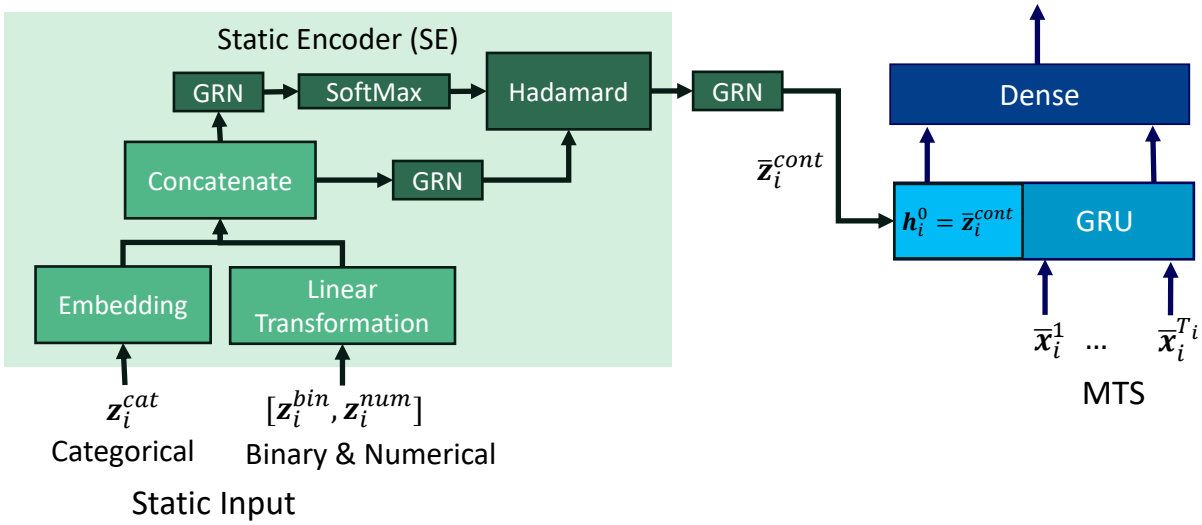


Figure 5.1: Overview of the FHSI Architecture. FHSI processes both static and time-varying inputs with distinct color-coded blocks. The SE component is depicted in various shades of green, where light green signifies the initial embedding mapping network and dark green represents the Variable Selection Network. The GRU block is denoted by a light blue box, and the final non-linear dense layer is illustrated in a dark blue box.

Late Fusion Models

Ensemble learning approaches, which aggregate outputs from multiple models, often outperform their individual components in diverse applications [172]. Typically, these methods involve training basic models, integrating their outputs through an aggregation module, and fine-tuning them to optimize aggregation parameters. Such aggregators can adopt various strategies, including parameter optimization, weighting coefficients, and error-processing techniques [173].

This dissertation focuses on late fusion techniques for handling multimodal data. Specifically, it examines a method that designates separate models for static and MTS data and then

fuses these outputs. This research explores two late fusion architectures, sharing the same basic models, but with different aggregation modules. The basic model to deal with static data we employed an MLP, while a GRU was utilized for MTS data. Both models estimate the posterior probability of AMR infection.

Late Fusion Convex Optimization

The LFCO model implements a linear combination of MLP and GRU outputs using two weights: $w_{MLP} \in [0,1]$ and $w_{GRU} \in [0,1]$. Enforcing the convexity constraint $w_{GRU} + w_{MLP} = 1$, we optimize these weights through a unidimensional exhaustive search, maximizing classification performance on validation data. The ratio w_{GRU}/w_{MLP} offers interpretative insights, indicating the relative significance of dynamic variables (processed by GRU) compared to static ones (handled by MLP) in predicting AMR infections.

Late Fusion Logistic Regression

Alternatively, the LFLR approach uses Logistic Regression (LR) to merge MLP and GRU outputs. During LR training, the MLP and GRU models remain static. The LR, a parametric method, applies a logistic function to a linear combination of inputs (outputs from MLP and GRU) to estimate the final prediction [174]. This process exclusively focuses on fusing the outputs from MLP and GRU, without modifying these underlying models.

5.2.3 Interpretability in Time Series

In the previous chapter, we emphasized the crucial role of interpretability in deep learning, particularly for clinical applications where understanding model decisions is vital. This chapter advances our exploration by refining methods to enhance interpretability in MTS-based models. Given the complexity of MTS, conventional interpretability techniques often fall short in extracting knowledge, requiring more sophisticated approaches.

We introduce three innovative methodologies to improve the interpretability of time series models: attention mechanisms, dynamic masking, and Time Perturbation Importances (TPI). Each technique provides distinct benefits for uncovering patterns and relationships in MTS data. Attention mechanisms help to pinpoint and highlight the most informative parts of the time series, making it easier to understand which data segments are crucial for the model's predictions. Dynamic masking, or Dynamask, involves deliberately hiding parts of the data to assess how these omissions affect the model's output, thereby identifying which time steps or features are most significant. Lastly, TPI evaluates how sensitive the model is to changes in the time series, illustrating how data modifications impact predictions.

By integrating these methodologies, this chapter aims to push the boundaries of interpretability in time series analysis, moving beyond conventional approaches to offer a deep understanding of how models process and respond to MTS data. This advancement is particularly significant in clinical contexts, where comprehending the nature of time-dependent data is crucial for accurate diagnosis and treatment planning. The subsequent sections will thoroughly examine each methodology, including its theoretical foundations, implementation strategies, and potential applications, culminating in a comprehensive framework for interpreting complex time series models.

Attention Mechanisms

Attention mechanisms in DNNs have emerged as indispensable tools, significantly enhancing interpretability, a critical aspect in ML, particularly in domains requiring precise and explainable outcomes [175, 176]. These mechanisms assign variable 'attention' or 'weight' to different elements within an input sequence during the DNN's computational process. This dynamic allocation process is key in determining the model's focus, selectively emphasizing certain data components over others. Such a system of weighted importance plays a pivotal role in how a DNN processes information and influences its ultimate decision or output. The ability of these mechanisms to allocate higher relevance to specific components means that the network, in its complex series of calculations, gives precedence to certain data features. This discriminatory allocation is not merely a computational convenience; it forms the basis of interpretability by shedding light on the decision-making process of the DNN.

The intrinsic value of attention mechanisms extends beyond their operational functionality; they provide a window into the otherwise opaque inner workings of DNNs. By showing which components or features are prioritized, these mechanisms reveal the aspects of data allowed most critical for generating specific outputs [177]. This revelation is important in fields where the rationale behind a model's decision is as crucial as the decision itself, such as in clinical decision support systems in healthcare. Understanding which features a DNN focuses on can lead to more informed interpretations and trust in the model's outputs, especially in scenarios where decisions have significant consequences. The attention mechanism was originally developed for machine translation models [178], although it has been successfully applied to very different problems, like medical computer vision tasks [179], ECG analysis [180], and blood pressure response [181].

The ability of attention mechanisms to discern key features within data not only enhances the interpretability of models but also boosts their overall performance. These mechanisms allow DNNs to focus on the most relevant features, optimizing processing efficiency. By pri-

oritizing significant data points, attention mechanisms prevent the network from overloading with irrelevant information [182]. This selective focus ensures that computational resources are allocated to the most impactful aspects of the data, leading to the development of more precise and operationally efficient models.

Furthermore, attention mechanisms significantly enhance the capabilities of DL and AI by offering a practical balance between computational intensity and clear, transparent decision-making processes. These mechanisms enable models to selectively emphasize and process key features in data, leading to the development of more advanced and understandable machine learning systems [183]. They represent a bridge between the raw computational power of DNNs and the need for understandable, transparent decision-making processes in AI systems.

Drawing on the approach presented in [175], our attention mechanism operates at the input variable level by employing a dense layer with a Softmax activation function. To elucidate, for each patient i, an attention matrix $\theta_i \in \mathbb{R}^{D\times T_i}$ is created by postulating and training an MLP that takes $\mathbf{X}_i \in \mathbb{R}^{D\times T_i}$ as input and produces θ_i as output. This matrix θ_i is used then to weight the original input $\mathbf{X}_i$ through a Hadamard product, resulting in the weighted (attention modulated) input $\tilde{\mathbf{X}}_i$. The Hadamard product and the learnable matrix θ_i serve the purpose of enabling the architecture to focus on specific feature-time instant pairs that are deemed more pertinent for patient i.

Our research introduces two novel modifications to the attention mechanism described in [175]. The first model, known as the Non-Linear Hadamard Attention (NLHA) model, is an adaptation of the concept introduced in [175]. It substitutes the MLP layer with a GRN. In contrast, the second model, termed as the Hadamard Attention Matrix (HAM) model, diverges more significantly from the attention mechanism in [175]. Instead of creating a unique attention matrix θ_i for each patient $i = 1,...,I$, HAM learns a single matrix $\underline{\theta}$ which is then applied uniformly ($\theta_i = \underline{\theta}$) to all MTS $\mathbf{X}_i$ with $i = 1,...,I$.

To ensure clinical interpretability of θ_i, our models apply the attention matrix directly to the raw input data $\mathbf{X}_i$, before any transformation or embedding is applied to the input. Consequently, the entries of θ_i can be directly interpreted to get the global contribution of each (d,t) feature-time instant pairs within the classification architecture.

Dynamic Mask

Dynamask, a groundbreaking perturbation-based post-hoc methodology developed for MTS architectures, is a significant advancement in ML interpretability. This method is designed to work with pre-trained black-box models, enabling researchers to identify and understand the relevance and influence of individual entries within the MTS. This approach is deeply rooted in

the foundational research conducted by Crabbe et al. [100], and it adapts and extends the principles of post-hoc masks. These principles were initially conceptualized for image classification, as detailed in works by Fong et al. [184, 185]. In image classification, the technique involves the application of perturbations to selected pixels and subsequently observing the resultant effects on the output of a black-box classifier. This process is vital as it sheds light on specific image regions that play a pivotal role in the classifier's decision-making mechanism.

Expanding upon this, Dynamask introduces an innovative twist by incorporating time-aware perturbations for MTS data analysis. This method modifies the value of a particular feature at a specific time step by replacing it with a value derived from an average of past feature values. This is particularly crucial in analyzing MTS, where temporal dynamics play a key role. In a practical setting, consider a classifier working with an input MTS matrix, represented as $\mathbf{X}_i \in \mathbb{R}^{D \times T_i}$. This classifier generates a label, denoted as $\hat{y}_i$. Concurrently, Dynamask captures and records saliency scores in a separate matrix, $\mathbf{M} \in \mathbb{R}^{D \times T_i}$. The perturbation operator, denoted as π, utilizes this matrix $\mathbf{M}$ to create a perturbed version of the input, labeled $\mathbf{X}_i^P$. This perturbed input is then processed by the classifier to produce an altered label, $\hat{y}_i^P$. The comparison and analysis of $\hat{y}_i^P$ and the original $\hat{y}_i$ are crucial; the error generated from this comparison is utilized in a back-propagation process. This process is iteratively applied, adjusting the values in $\mathbf{M}$ through numerous inputs and training epochs, thereby facilitating the learning of these values and improving the interpretability of the classifier's decisions.

Dynamask's flexibility is further exemplified by its compatibility with a variety of time-averaging operators. In our application, we use a simple moving average approach. This approach is showed in equation (5.1), which defines the (d,t) entry of the perturbed input matrix $\mathbf{X}_i^P$ as follows:

$$m^{(t,d)} x_i^{(t,d)} + (1 - m^{(t,d)}) \mu_i^{(t,d)}, \text{with} \mu_i^{(t,d)} = \frac{1}{W_{MAW} + 1} \sum_{t'=t-W_{MAW}}^{t} x_i^{(t',d)}, \tag{5.1}$$

Here, W_{MAW} is the width of the moving average window. It's important to highlight that for the initial time steps (specifically when $t \leq W_{MAW}$), the computation of $\mu_i^{(t,d)}$ is adapted to account for the limited number of input values available for averaging. The values of $m^{(t,d)}$ in equation (5.1) are of significant interest; values approaching one indicate the critical importance of the current value $x_i^{(t,d)}$, whereas values closer to zero suggest that consideration should be given to the average of previous time steps values. This mechanism offers an understanding of the temporal dynamics within the MTS data.

To further enhance the interpretability capabilities of Dynamask, we have introduced an

additional component in the training cost function. This component comprises a penalty designed to promote the generation of mask values that are sparse and bounded within a range of one. This strategy, advocated in the original Dynamask research by Crabbe et al. [100], is key in ensuring that the resulting mask values are not only effective in elucidating the classifier's decision-making process but also including the constraints that render the interpretation insightful. This penalty encourages the model to focus on the most salient features, avoiding overfitting and ensuring that the interpretation remains focused and relevant. This approach to interpretability is valuable in complex data environments like healthcare, where understanding the reasoning behind a model's decision is key.

Time Perturbation Importances

Time Perturbation Importances stands as a novel inspection methodology designed to highlight the most significant time steps in a dataset, focusing on those that hold the greatest relevance based on models that have already been trained [186, 187]. The approach of TPI is closely aligned with the PFI method previously presented, but it introduces modifications to adapt to the specific characteristics of MTS.

TPI's core mechanism involves a detailed analysis of how the performance of a model, which has already been trained, deteriorates when the information linked to a specific time step is deliberately distorted. This performance degradation is typically measured in terms of a specific metric. The fundamental distinction between TPI and PFI lies in the manner in which data is handled. Instead of permuting data across different time steps, which could potentially disrupt the inherent temporal structure of the dataset, TPI opts for a different route. It introduces modifications to the original data by adding a layer of white Gaussian noise. This approach ensures that the data's temporal sequence remains intact while allowing for an analysis of the importance of each time step.

To apply the TPI method effectively, one must initially focus on training a specific model. Once the model is trained, the next step is to evaluate it using a chosen figure of merit, such as accuracy, precision, or recall, on a set of samples from the validation dataset. This step is critical as it establishes a baseline performance level for the model. After establishing this baseline, the process involves selecting a particular time step, referred to as the t-th step, and perturbing the information related to this time step across all patients and features within the validation dataset. This perturbation is achieved by adding noise to the data at this time step while ensuring that the data at other time steps, denoted as $t' \neq t$, remains unaffected. This selective perturbation allows for an analysis of how changes at a single time step impact the model's overall performance.

The final step in the TPI methodology involves a comparative analysis. Here, the figures of

merit obtained from the model after the perturbation of each time step are compared with those derived from the original validation dataset. This comparison is essential because it demonstrates how the model's performance changes in response to modifications at each time step. The larger the observed degradation in the figure of merit, the more critical that specific time step is deemed to be for the problem at hand. This insight is particularly valuable as it helps in understanding the dynamics of the dataset and the role of temporal information in model performance, thereby aiding in the refinement and optimization of the model for better performance and reliability in predictions.

5.3 Experiments and Results

In the following section, we describe the detailed experiments and the specific results obtained, presenting a clear view of our research process. We begin with the 'Modeling' subsection, where we discuss the development of our models and data modeling. Next, in the 'Full Feature Set Analysis Results' subsection, we present the results using all available features in our model. We provide a detailed examination of these results, highlighting key findings and their implications for the broader research context. The following subsection, 'Feature Selection and Interpretability in Knowledge Extraction', focuses on the methodologies for selecting the most important features and discusses how these selections enhance the interpretability of the extracted knowledge. Lastly, the 'Selected Feature Set: Analysis Results' subsection presents a detailed analysis of the results obtained by applying only these chosen features, allowing for a comparative evaluation against the full feature set findings. This structure aims to provide a clear and systematic exploration of our research methods and findings, underlining the rigor and depth of our analysis.

5.3.1 Experimental Setup

The database employed in these experiments was detailed in Section 3.1; therefore, this subsection will focus solely on describing the specific modeling of the experiments conducted in this chapter of the thesis.

Specifically, we extend our data analysis to a 16-year period, from January 2004 to February 2020, encompassing 3,158 ICU stays. Although the dataset extended to 2022, analyses were deliberately confined to the 2020 threshold due to the significant impact of the COVID-19 pandemic on subsequent data, rendering post-2020 records less representative for the purposes of

this study. The dataset presented in this chapter exclusively encompasses all the features previously presented in Section 3, both static and MTS. The MTS employed includes information on the medications administered to the patient, the MV, the previous cultures of the patient, and features associated with the patient's neighbors. This study uniquely focuses on the first instance of AMR in each patient's culture, with AMR cultures identified in 605 cases. This scenario presents a classification challenge due to the significant imbalance between AMR and non-AMR cases.

Our modeling approach for the MTS differs from the previous chapter, employing a different time window model. In Chapter 4, the initial time step depended on the patient population. However, this chapter defines the initial time slot ($t = 0$) as the patient's admission day to the ICU for both populations. The final day ($t = T_i$) varies depending on the patient's AMR status: for non-AMR patients ($y_i = 0$), $t = T_i$ marks their ICU discharge, whereas for AMR patients ($y_i = 1$), it denotes the day their culture is identified as AMR. Given the variability in MTS lengths, we adopt a 14-day temporal window, informed by literature and clinical expertise [188, 189]. For each patient, if $T_i < 14$, we use their original MTS ($\mathbf{X}_i$) as input; if $T_i \geq 14$, only the first 14 columns of $\mathbf{X}_i$ are utilized.

This windowing choice is supported by previous studies, which suggest that longer windows in models handling irregular MTS lengths are more effective in predicting AMR onset compared to shorter windows or models that standardize MTS lengths through imputation [21]. The 14-day window is clinically significant: it represents a critical period for AMR germ emergence in the ICU and is the standard quarantine duration for patients identified with AMR infections [188]. These modeling decisions are thus deeply rooted in both empirical research and practical clinical protocols [189].

5.3.2 Early Prediction of Antimicrobial Multidrug Resistance Emergence Using All Features

This subsection focuses on predicting the early emergence of AMR using a range of data recorded in EHRs. We employ various data-driven models, non-multimodal (MLP, GRU) and multimodal (JHF, FHSI, LFCO), utilizing all the features set. The mean and standard deviation of Accuracy, Specificity, Sensitivity, and ROC AUC are calculated across three test splits, as shown in Table 5.1. The same three test sets were considered in all the experimental work to maintain fairness with all methods.

The table reveals that the MLP model, limited to static variables, exhibits suboptimal per-

Method	Accuracy	Specificity	Sensitivity	ROC AUC
MLP	58.60 ± 0.52	58.62 ± 0.48	58.37 ± 4.64	62.29 ± 2.34
GRU	63.19 ± 2.47	59.91 ± 4.17	77.83 ± 5.83	75.50 ± 0.36
FHSI	62.76 ± 3.25	59.17 ± 4.45	78.98 ± 3.56	$\mathbf{76.74 \pm 1.36}$
JHF	65.14 ± 1.55	62.58 ± 1.29	76.55 ± 1.80	76.20 ± 1.17
LFLR	$\mathbf{67.25 \pm 2.29}$	$\mathbf{65.90 \pm 3.56}$	73.75 ± 3.76	76.21 ± 1.31
LFCO	60.92 ± 3.14	56.39 ± 4.38	$\mathbf{81.38 \pm 3.53}$	76.18 ± 1.31

Table 5.1: Performance summary with mean $\pm$ standard deviation for accuracy, specificity, sensitivity, and ROC AUC on three test partitions, considering all features. Highest performances are highlighted in bold.

formance with an ROC AUC of 62.29%, likely attributed to its exclusion of temporal patient data. In contrast, LFLR excels in Accuracy and Specificity, recording the highest mean scores at 67.25% and 65.90%, respectively. Notably, LFCO surpasses its counterparts in Sensitivity, achieving a mean score of 81.38%, indicating its superior ability to identify true positives. Regarding ROC AUC, a key indicator of a model's discriminatory power between classes, FHSI and LFLR show nearly equivalent efficacy, with FHSI marginally ahead at 76.74%. This suggests that while both models proficiently classify both classes, FHSI holds a slight edge. Additionally, the GRU and multimodal models, incorporating both MTS and static variables, demonstrate similar results, with multimodal models showing a modest enhancement.

Building on the methodology of Martinez et al. (2022) [21], further experiments will integrate FS methods and interpretable mechanisms. This approach aims to enhance our understanding and improve model performance beyond the preliminary results presented in this section.

5.3.3 Feature Selection and Interpretability for Knowledge Extraction

The preceding subsection detailed preliminary experiments yielding unfavorable results. To enhance our understanding and improve model performance, we delved deeper into our dataset using FS processes and various interpretable models.

We initiated our analysis using the FS process. Figure 5.2 illustrates a matrix aligning variables in columns against FS techniques in rows (see Sec. 5.2.1). Blue cells denote selected features. This matrix segregates classical FS methods (CIB, CMI, GLASSO) at the top, with PFI results applied on each of the models presented at the bottom. A majority voting scheme among the classical methods was employed, selecting features endorsed by at least two methods. The PFI analysis highlighted key features such as patient age, SAPS-3 score, and year

of admission for static variables, as well as MV and AMR neighbors for MTS. Notably, CF1 and PEN antibiotics were frequently selected by PFI methods. Classical methods displayed a broader feature selection range, particularly in MTS, with less consensus compared to PFI methods. Shared selections across all methods included CAR and PEN antibiotics, patient age, gender, SAPS-3 score, and admission year.

Clinically, the selected features align well with existing literature, as confirmed by the UHF staff. The relevance of the SAPS-3 score, MV, and AMR neighbors corroborates clinical expectations. Once the FS approaches have been applied, we examine these features through interpretable mechanisms outlined in Sec. 5.2.3.

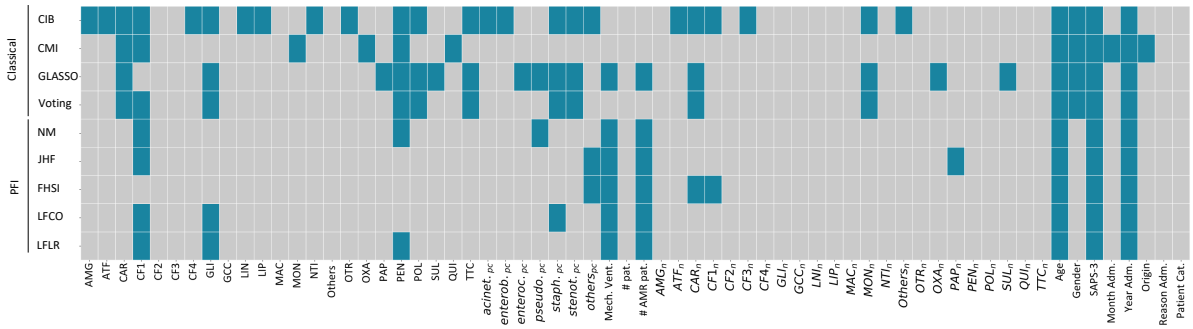


Figure 5.2: Feature Selection Matrix for Static and MTS data (in columns) and FS approaches (classified as PFI and classical techniques in rows). Dark blue cells indicate selected features. NM results encompass both MLP for static data and RNN for MTS.

Figure 5.3 showcases a heatmap for NLHA scores using the FHSI model, which demonstrated optimal performance. This heatmap, depicting feature importance across time slots (with '0' signifying ICU admission day), focused solely on MTS. Mechanical ventilation emerged as a crucial variable, particularly in early patient stay days.

Following the FS results, we analyze the attention scores obtained when applying the NLHA and HAM mechanisms. Figure 5.3 showcases a heatmap for NLHA scores using the FHSI model, which demonstrated optimal performance. The heatmap's columns correspond to features, while rows to time-slots of the MTS under study ('0' denotes the day of the ICU admission). This visualization focuses on MTS, excluding static variables, to emphasize the importance scores of features and time slots. Since NLHA generates an attention matrix θ_i for each sample, the Figure 5.3 represents the average across all the attention matrices. Notably, MV emerges as the most significant variable, followed by the patient's number of AMR neighbors, aligning with prior PFI technique findings. Early days of hospitalization show elevated scores for the MV feature, underscoring its criticality in patient care during initial ICU stay.

The scores of attention corresponding to the matrix $\underline{\mathbf{A}}$ of the HAM architecture using the FHSI black-box model are presented in Figure 5.4. The representation is similar to the one presented for Figure 5.3. The importance of MV and the number of AMR neighbors is also evidenced here, particularly during the initial phase of a patient's hospitalization. Antibiotics such as CAR, GLI, or PEN also have high scores on the first day of the patient's stay.

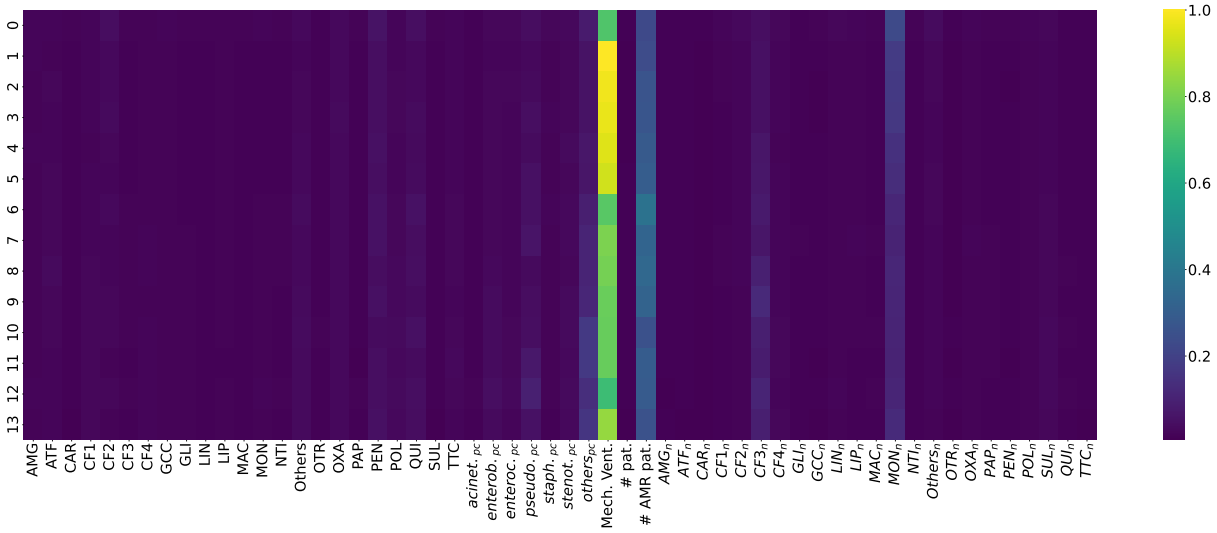


Figure 5.3: Heatmap of average θ_i matrices for the NLHA model. Displaying feature importance scores over time steps as rows (with '0' denoting ICU admission day) and features as columns.

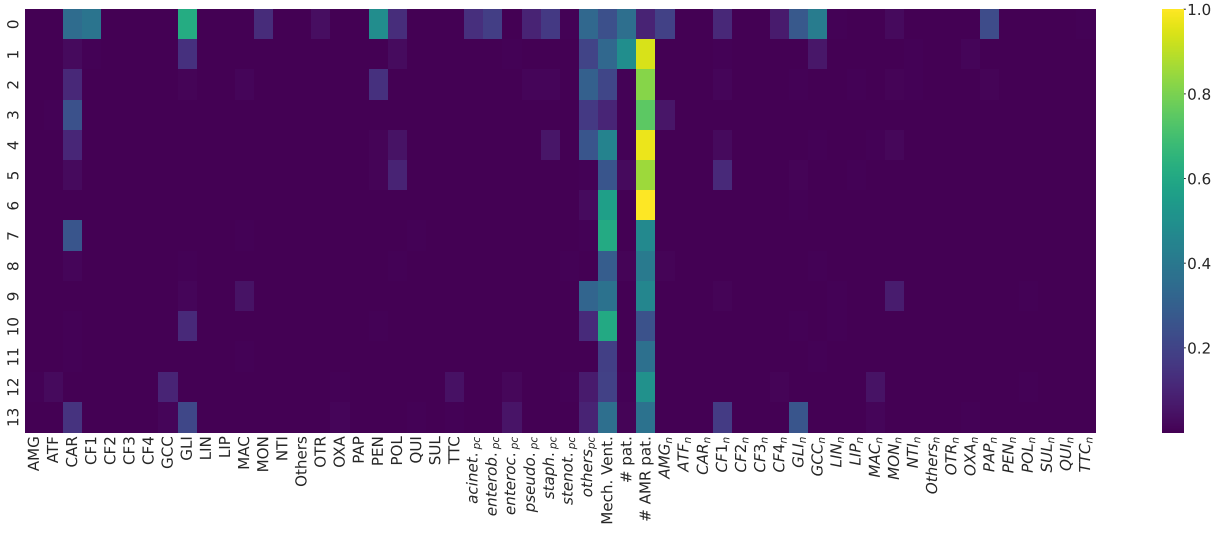


Figure 5.4: Heatmap of the matrix $\underline{\mathbf{A}}$ for the Ham model. Displaying feature importance scores over time steps as rows (with '0' denoting ICU admission day) and features as columns.

Figure 5.5 displays the scores of the Dynamask mechanism when applied in conjunction

with the FHSI model, chosen for its marginally superior ROC AUC compared to alternative models. The figure's columns correspond to features and rows to distinct time slots of the MTS under study, starting from ICU admission (denoted as '0'). Similar to Figures 5.3 and 5.4, Figure 5.5 focuses solely on MTS, highlighting the significance of both features and time steps. Notably, the figure underscores the relevance of MV and the number of AMR neighbors, consistent with their high importance in Figures 5.3 and 5.4. Additionally, the Dynamask mechanism attributes considerable importance to factors such as the CAR antibiotic family, results from prior cultures identifying non-AMR germs, and the number of neighbors of the patients, indicating their significant roles in the model's analysis.

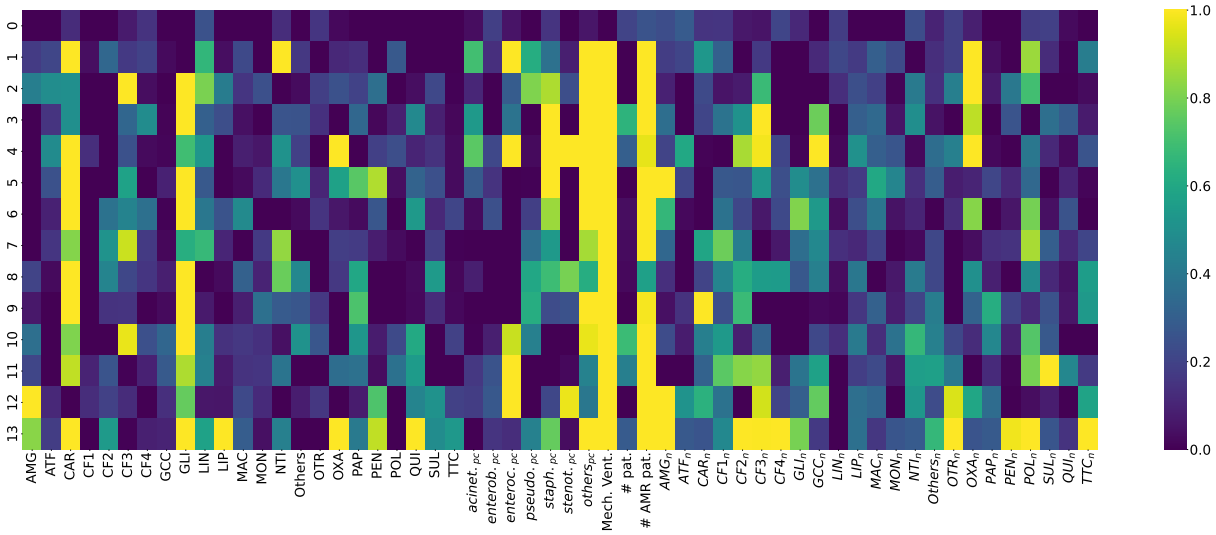


Figure 5.5: Importance score heatmap generated by the Dynamask model applied to a pre-trained FHSI model. Columns represent features, rows indicate time-steps ('0' Marks the ICU Admission Day).

The LFCO model previously presented can also give us knowledge about the task to solve. The weights w_{MLP} and w_{GRU}, representing the significance of static and MTS variables, respectively, are pivotal for this analysis. A w_{MLP} value above 0.5 signifies greater relevance of static variables, whereas lower values highlight the predominance of MTS variables. This dichotomy is predicated on the LFCO's sole constraint: $w_{GRU} + w_{MLP} = 1$. Our experiments show that the mean value of w_{MLP}, is 0.34 (standard deviation of 0.03). Therefore, we can conclude that considering the LFCO scores, the MTS are more important than the static variables. This statement is aligned with the comparative analysis of MLP and GRU outcomes in the previous section (see Table 5.1).

To investigate the temporal aspects of our dataset, we conducted an analysis focusing on the patient's stay. This research is clinically significant, as it identifies specific time frames when

heightened vigilance is necessary to prevent the emergence of AMR germs.

Utilizing the TPI methodology applied to the previously established FHSI model , we analyzed the significance of various time steps. The results revealed that the initial days of the hospital stay (specifically the first, second, third, and fourth), together with the eleventh and fourteenth-time steps, hold the greatest importance, as evidenced by their high TPI scores (0.26, 0.91, 1.00, 0.38, 0.29, and 0.41 respectively). In contrast, the other time steps demonstrated relatively lower significance, averaging a score of 0.14. These findings suggest two key insights: firstly, the initial days of hospitalization are crucial for predicting AMR development, a hypothesis supported by many patients developing resistance within the first 72 hours. Secondly, the significance of the last day in patients with windowed MTS implies a potential need for extended observation windows. However, subsequent exploratory analysis indicated that extending the observation window may not yield better predictive performance.

5.3.4 Analysis of the Selected Features

To assess the effectiveness of using FS strategies in AMR prediction models, we compared the performance of models trained with different FS techniques. The results, detailed in Table 5.2 focus on four critical figures of merit: Accuracy, Specificity, Sensitivity, and ROC AUC. The Table is organized based on two categories of FS strategies. The first category includes classical FS strategies, employing methods such as CIB, CMI, GLASSO, and a voting procedure. In contrast, the second category focuses on features selected through PFI techniques. While various features were initially tested, the optimal results emerged from using 3, 4, and 5 MTS alongside 3 static features (patient age, SAPS-3 score, and year of admission). Key findings indicate that PFI strategies outperform classical FS methods, with an average ROC-AUC of 78.77 compared to 70.46 for the latter. Within classical FS techniques, GLASSO emerges as the most effective, while CMI shows the weakest performance.

Analyzing the performance of individual classifiers, the MLP underperforms, as was previously discussed in Sec. 5.3.2, potentially due to its limited use of static variables. The GRU classifier tends towards higher sensitivity, especially when paired with CIB features. This characteristic could be advantageous when the cost of missing a positive classification is significant. In contrast, the FHSI classifier exhibits robustness and versatility, showing consistently high performance across various metrics, particularly in terms of accuracy and ROC AUC. This consistency points to the FHSI classifier's potential as a reliable tool for generalizable predictive modeling. Meanwhile, the LFLR and LFCO classifiers, though showing moderate performance, do not distinctly excel in any specific metric, indicating a more balanced but less specialized

performance profile.

The variability in the performance metrics, as indicated by the standard deviation values, is crucial for understanding the consistency and reliability of the classifier-feature pairings. This aspect is particularly important in predictive modeling, where stability and predictability of outcomes are key. The multimodal techniques, which include static variables like the age of the patient, SAPS-3 score, and year of admission, demonstrate the effectiveness of integrating both dynamic and static factors in classification tasks with an average ROC AUC of 76.85.

Finally, the results indicate a particularly strong performance from the FHSI classifier using 3 MTS + 3 features, which stands out across all metrics. This finding suggests an optimal interplay between the classifier and the specific combination of features, highlighting the importance of FS in enhancing classifier performance. This configuration achieved the highest Accuracy (73.89 ± 3.55), Specificity (72.63 ± 5.43), and ROC AUC (84.33 ± 1.38). The higher balance achieved between sensitivity and specificity in this configuration underscores its potential applicability in scenarios requiring high-precision classification.

In conclusion, the results from the table underscore the significance of selecting appropriate classifiers and feature sets in predictive modeling, especially in complex, multimodal contexts. The standout performance of the FHSI classifier with a specific feature combination offers a valuable model for future studies. This analysis not only highlights the nuanced interplay between classifiers and features but also emphasizes the necessity for careful, informed choices in the design of predictive models to optimize performance and reliability.

5.4 Conclusions

This chapter improved the previous DL models by integrating multimodal architectures and interpretable models, leveraging the synergistic potential of diverse data types. This marked a significant advancement in the realm of AMR prediction. The application of multimodal approaches addressed the inherent limitations of unimodal systems, allowing for a more comprehensive understanding of AMR dynamics and enhancing the predictive accuracy and reliability of AMR detection models.

Key contributions included the development of innovative multimodal architectures, such as the JHF and the FHSI, alongside late fusion models like the LFCO and the LFLR. These models were adeptly tailored to process static data and MTS, proving vital in understanding patient conditions and predicting AMR emergence.

Furthermore, the chapter emphasized the importance of interpretability in these models, especially given their application in critical clinical settings. The integration of attention mechanisms, dynamic masking, and TPI significantly enhanced the interpretability of the models. These methods provided insights into the most informative parts of the time series and revealed the impact of individual time points or features on model output, thereby facilitating a deeper understanding of the model's decision-making processes. The extensive suite of experiments and the analysis of results presented in this chapter demonstrated the efficacy of these models and interpretability methods. Also, improving the previous FS process with a new methodology, such as PFI, substantially improved model performance. These results underlined the necessity of careful FS in building robust and accurate predictive models. Both interpretable mechanisms and FS implementation showed that MV and the # AMR patients are key factors in AMR development. Certain antibiotics like Carbapenems, Cephalosporins, Glycopeptides, and Penicillins are also important. This aligns with existing research, which notes the common use of these antibiotics and links invasive procedures like mechanical ventilation to increased infection and resistance risks.

Data Source	Method	Features	Accuracy	Specificity	Sensitivity	ROC AUC
Classical FS	MLP	**CIB features**	58.12 ± 4.46	58.95 ± 6.74	54.38 ± 6.58	61.92 ± 1.40
		CMI features	49.74 ± 9.55	45.14 ± 13.53	70.96 ± 10.19	62.23 ± 1.23
		GLASSO features	58.12 ± 4.46	58.95 ± 6.74	54.38 ± 6.58	61.92 ± 1.40
		Voting features	58.12 ± 4.46	58.95 ± 6.74	54.38 ± 6.58	61.92 ± 1.40
	GRU	**CIB features**	64.14 ± 3.71	59.96 ± 3.78	**82.75 ± 3.31**	78.83 ± 3.39
		CMI features	37.29 ± 4.12	27.07 ± 7.64	81.84 ± 8.02	60.22 ± 3.24
		GLASSO features	68.72 ± 1.65	67.31 ± 2.50	75.38 ± 3.93	78.80 ± 1.62
		Voting features	47.68 ± 2.98	42.85 ± 3.18	69.43 ± 0.28	60.99 ± 2.79
	JHF	**CIB features**	68.57 ± 2.50	67.04 ± 3.92	75.64 ± 4.38	79.05 ± 1.39
		CMI features	58.02 ± 3.30	55.70 ± 5.13	67.78 ± 3.86	65.11 ± 1.64
		GLASSO features	69.41 ± 2.47	67.97 ± 3.25	76.21 ± 3.33	80.07 ± 2.30
		Voting features	58.23 ± 0.93	56.76 ± 1.61	64.76 ± 4.08	65.12 ± 3.06
	FHSI	**CIB features**	71.52 ± 3.23	70.43 ± 4.68	76.82 ± 3.96	81.01 ± 0.21
		CMI features	56.80 ± 1.46	54.07 ± 1.89	68.79 ± 2.36	66.66 ± 1.27
		GLASSO features	68.83 ± 4.13	65.95 ± 5.73	82.46 ± 4.13	81.76 ± 2.43
		Voting features	62.82 ± 2.49	63.44 ± 3.31	60.23 ± 1.58	66.95 ± 1.62
	LFLR	**CIB features**	70.04 ± 1.81	68.78 ± 2.02	75.73 ± 0.74	78.49 ± 1.82
		CMI features	54.22 ± 3.14	50.49 ± 4.53	71.44 ± 4.40	65.55 ± 0.25
		GLASSO features	69.99 ± 0.49	68.51 ± 1.00	76.89 ± 2.91	79.34 ± 0.42
		Voting features	54.96 ± 5.53	51.36 ± 6.60	71.02 ± 2.09	65.72 ± 3.23
	LFCO	**CIB features**	67.62 ± 1.38	65.49 ± 1.91	77.37 ± 1.81	78.47 ± 1.23
		CMI features	50.21 ± 5.18	43.66 ± 7.57	79.81 ± 7.14	65.55 ± 0.34
		GLASSO features	68.93 ± 2.76	67.19 ± 4.13	77.30 ± 4.57	79.99 ± 0.66
		Voting features	52.58 ± 5.75	48.85 ± 6.70	69.66 ± 2.33	65.57 ± 2.78
PFI	MLP	**3 features**	46.04 ± 3.77	39.54 ± 6.97	74.17 ± 7.98	62.09 ± 1.07
		4 features	62.29 ± 0.97	64.00 ± 1.56	54.88 ± 1.57	62.16 ± 0.71
		5 features	52.85 ± 2.02	50.50 ± 2.68	63.33 ± 2.27	62.60 ± 1.19
	GRU	**3 MTS**	67.51 ± 3.03	64.47 ± 3.22	81.16 ± 1.37	81.85 ± 1.43
		4 MTS	68.78 ± 2.90	66.42 ± 3.66	79.62 ± 1.35	80.88 ± 1.90
		5 MTS	67.14 ± 2.57	64.09 ± 2.64	80.93 ± 3.16	80.68 ± 2.44
	JHF	**3 MTS + 3 feat.**	71.89 ± 1.74	70.02 ± 2.10	80.49 ± 6.32	82.94 ± 2.01
		4 MTS + 3 feat.	69.36 ± 1.90	67.18 ± 2.41	79.35 ± 1.81	81.61 ± 1.06
		5 MTS + 3 feat.	69.78 ± 2.45	68.61 ± 2.72	75.23 ± 4.34	80.97 ± 2.24
	FHSI	**3 MTS + 3 feat.**	**73.89 ± 3.55**	**72.63 ± 5.43**	79.47 ± 5.62	**84.33 ± 1.38**
		4 MTS + 3 feat.	71.94 ± 3.03	69.49 ± 4.53	82.27 ± 5.49	83.48 ± 2.68
		5 MTS + 3 feat.	71.84 ± 1.81	69.86 ± 3.48	80.01 ± 4.82	82.92 ± 2.08
	LFLR	**3 MTS + 3 feat.**	68.88 ± 2.87	66.82 ± 3.68	78.56 ± 4.25	81.83 ± 1.69
		4 MTS + 3 feat.	68.93 ± 1.55	66.60 ± 2.44	79.84 ± 4.17	82.07 ± 1.28
		5 MTS + 3 feat.	68.09 ± 1.20	66.06 ± 1.10	77.28 ± 4.02	81.32 ± 1.85
	LFCO	**3 MTS + 3 feat.**	69.78 ± 1.71	68.26 ± 2.13	76.88 ± 2.98	82.25 ± 1.37
		4 MTS + 3 feat.	69.41 ± 1.47	67.61 ± 1.52	77.65 ± 2.40	81.50 ± 1.45
		5 MTS + 3 feat.	69.25 ± 1.30	66.42 ± 1.42	81.72 ± 1.67	82.32 ± 1.04

Table 5.2: Performance summary with mean ± standard deviation for Accuracy, Specificity, Sensitivity, and ROC AUC across three test partitions, This experiments consider: classical-FS and PFI methods (first column); various classifiers including MLP, GRU, JHF, FHSI, LFLR, and LFCO (second column); and different feature sets identified by each approach (third column). All the multimodal methods (JHF, FHSI, LFLR, and LFCO) utilize the same static variables (patient age, SAPS-3 score, and year of the admission). The highest values for each figure of merit are highlighted in bold.

Chapter 6

Data and Network Analytics for COVID-19 Intensive Care Unit Patients

6.1 Introduction

In the wake of the COVID-19 pandemic, understanding patient health trajectories has become crucial, particularly in the context of treatment strategies and symptom progression [190]. Traditional models have often been inadequate, leaving gaps in effective patient management and treatment protocols. To address these limitations, our study introduces a novel graph-based data science approach uniquely tailored to analyze COVID-19 patient trajectories more accurately and informally. This method, diverging significantly from conventional linear or less dynamic models, allows for a more comprehensive understanding of the interplay between patient comorbidities, prior medications, and symptomatology. It is designed to track patient progress through various critical stages of the disease, from symptom onset to emergency department arrival, hospitalization, and ICU admission. It offers a dynamic perspective on disease progression and treatment responses [191].

We classified the patient population into two groups: those who succumbed to the condition (deceased) and those who survived ICU admission (non-deceased). This classification is essential for understanding the different trajectories and outcomes of COVID-19 patients, allowing for a comparative analysis that can reveal critical trends and information essential for enhancing future treatments. To unravel the complexities within the patient data, our methodology first employs a CIB method to identify significant characteristics and differences between the deceased and ICU survivor groups. Following this, we employ graph-based models and network analysis

techniques to explore simple pairwise and more intricate relationships among clinical features. These analytical tools give clinicians novel perspectives and insights into patient evolution and symptom development. Overall, the approach not only offers an innovative analytic tool for the clinical setting but also improves more informed decision-making, potentially enhancing patient care and treatment outcomes [192].

6.2 Time Series Processing Based on Graph Modeling

Clinical data, with its high-dimensional and heterogeneous characteristics, poses more significant analysis challenges than traditional datasets. This complexity is evident in clinical records, which include variables of diverse nature, some of which are static while others are dynamic. This variability and the often limited number of patient samples limit the use of many traditional statistical methods, which rely on large datasets. Addressing these issues, this dissertation employs robust non-parametric data science techniques to analyze and gain insights from EHR in the COVID-19 dataset. We also utilize graphs representing pairwise associations among heterogeneous features in the EHR to model the data. Graphs, as mathematical structures, offer versatility and a well-established analytical framework [81]. Their utility spans various data science domains, such as ML [82], signal processing [83], and statistics to structure complex datasets and integrate into data processing and learning pipelines [84]. Graphs also provide relative ease of understanding and effective visualization of high-dimensional data [82, 83, 85].

Previous research has illustrated the efficacy of network approaches in various healthcare contexts, including visualizing collaborative EHR use in heart failure [86], modeling disease graphs [87], and predicting unknown adverse drug reactions [88]. While graph-based methodologies for MTS in EHR are less common, prior research findings have demonstrated their potential efficacy in early outbreak detection and in enhancing our comprehension of the clinical trajectory of COVID-19 patients [89, 90, 91].

This chapter aims to demonstrate the application of data-based and graph-based network analytics on static, dynamic, and MTS variables in the EHRs of COVID-19 patients with varying conditions, treatments, and outcomes. We conduct a case study on COVID-19 patients in a Spanish ICU, focusing on relational insights from their EHRs. Our approach, centered on network analytics via graphs, seeks to understand the progression of COVID-19 and associated drug treatments from symptom onset to ICU discharge.

Given the clinical variables available in our dataset and the constraints imposed by our sample size, our analytical approach primarily adopts a descriptive framework. We focus on

the identification of correlations and associations among these variables. It is important to note that this descriptive analysis establishes the basis for future predictive studies. However, the viability and accuracy of predictive modeling depend on the acquisition of supplementary data in subsequent phases of the research.

6.3 Graph Modeling Using Correlation Coefficients

Analyzing MTS in clinical settings is challenging due to data sparsity, irregular sampling, and noise. Graphs have become an effective tool for data analysis, offering an intuitive framework for visualizing datasets with irregular structures [84]. Graphs efficiently represent, analyze, and visualize data over irregular domains [84]. A graph $\mathscr{G}$ comprises a node set $\mathscr{N} = \{n_1, ..., n_I\}$ and an edge set $\mathscr{E}$, where an edge $(n_i, n_j) \in \mathscr{E}$ signifies a relationship between nodes n_i and n_j. These relationships can be binary or weighted, with the weight indicating the relationship's strength. In data analytics, graph nodes often correspond to variables, and edges are constructed using various methods, including correlation or influence measures [84, 85].

In this research, we compute associations between variable pairs using three correlation coefficients due to the variables' heterogeneous nature (numerical/binary). The Pearson coefficient assesses pairwise correlations between numerical features [193], while the Phi coefficient is used for binary features and the Point-Biserial coefficient for binary-numerical feature pairs [194]. To manage large edge counts in constructed graphs, we apply a thresholding scheme using K-means clustering ($K = 2$) to categorize edges as relevant or irrelevant [124]. This technique also accommodates the variables' heterogeneity by allowing different thresholds for each edge type.

Once the surviving edges (and their strength) have been obtained, the graph's global connectivity is represented by the adjacency matrix $\mathbf{A}$, size $J \times J$ for static and $I \times I$ for dynamic features. In the last case, the matrix entry $A_{i,j}$ is zero if nodes n_i and n_j are unrelated. The adjacency matrix allows us to measure the connectivity of a specific node using the normalized weighted degree $d_i = \frac{1}{I}\sum_{j=1}^{I} |A_{i,j}|$ definition.

We propose a methodology that involves defining temporal intervals to address the temporal nature of many variables within our dataset. Subsequently, we construct graphs associated with each of these intervals separately for both deceased and non-deceased patients. In modeling MTS, various data-processing tools are available for consideration [195]. Previous research has utilized a feature engineering approach, drawing on clinical expertise to create tailored features from existing data [196, 153]. In contrast, our approach follows a more conventional

approach. We represent each of the temporal signals describing drug intake for patient m (i.e., each row of $\mathbf{X}_m$) using a statistical summary. Specifically, we calculate the mean, such that the matrix $\mathbf{X}_m \in \mathbb{R}^{I \times T_m}$ is transformed into the vector $\mathbf{x}_m \in \mathbb{R}^I$. Each entry of $\mathbf{x}_m$ corresponds to the aggregated number of doses of drug i administered to patient m, divided by the number of columns in $\mathbf{X}_m$. Significantly, our analysis often focuses on specific periods of time, such as the duration a patient spends in the ICU. In such instances, we condition the mean calculation to the relevant interval by selecting only the columns of $\mathbf{X}_m$ corresponding to that period. We then divide by the number of days that patient m spent within that interval. Once we have generated the signals $\mathbf{x}_m$ for all patients ($m = 1, ..., M$) within the designated interval of interest, we proceed to construct graphs separately for the deceased and non-deceased populations. This method allows us to effectively explore the temporal dynamics and relationships within our data.

Beyond constructing graphs, we analyze their topological properties using graph-theoretical tools to understand feature relationships, including their temporal evolution. Key metrics in this analysis are edge density $\eta(\mathscr{G}) = \frac{1}{I-1}\sum_{i=1}^{I} d_i$ and edge entropy $H(\mathscr{G}) = -\sum_{i=1}^{I} d_i \ln d_i$, which help in assessing network complexity and graph complexity [84, 81]. These metrics facilitate understanding of the global dataset properties, with edge entropy offering insights into graph structures and motifs [197]. The overarching goal of our graph approach is to gain a deeper understanding of the COVID-19 dataset.

Static Feature Analysis

This analysis begins by examining static features such as symptoms, comorbidities, regular medication, and demographic variables of COVID-19 patients. In the subfigures of these subsections, we construct two distinct graphs to elucidate the disparities in disease outcomes: one for deceased patients ($\mathscr{G}_d$) and another for survivors ($\mathscr{G}_{nd}$), as illustrated in Figure 6.1. These figures also incorporate the difference in adjacency matrices ($\bar{\mathbf{A}} = \mathbf{A}_d - \mathbf{A}_{nd}$), with node sizes indicating the prevalence of symptoms and comorbidities, and edge widths reflecting the correlation strength between features.

The graphical analysis shown in Figure 6.1 reveals the prevalence of comorbidities, such as hypertension, diabetes, and obesity. Moreover, it underscores the prominence of fever, cough, and dyspnea as dominant symptoms.

It is worth noting that within $\mathscr{G}d$ (the graph associated with deceased patients), nodes representing heart disease and smoking exhibit larger sizes compared to $\mathscr{G}nd$ (the graph linked to non-deceased patients). Furthermore, there is a distinction in the usage of medications, with deceased patients more frequently using ACE inhibitors, while survivors predominantly used

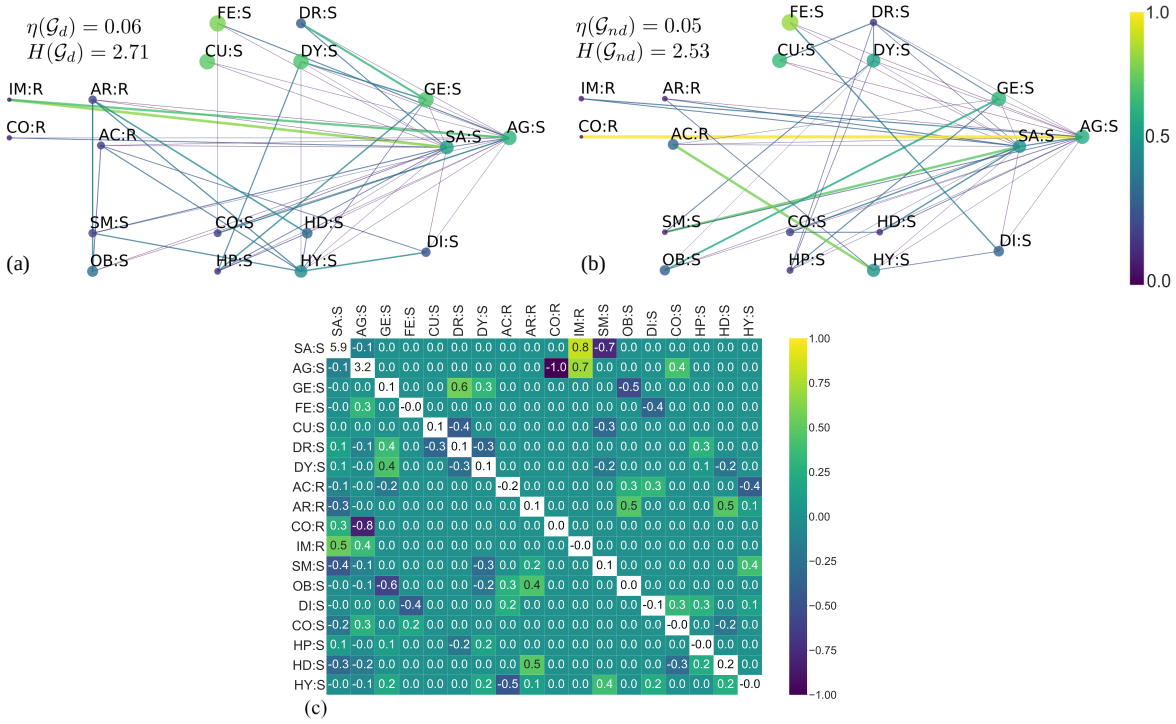


Figure 6.1: A visual representation of the network for demographic variables, symptoms, comorbidities, and regular medication is presented as follows: (a) Graph for deceased patients; (b) Graph for non-deceased patients; (c) Matrix representation of the difference between the graphs in (a) and (b). In particular, the non-diagonal entries of the matrix in (c) A matrix representation illustrating the differences between the graphs in (a) and (b). Additionally, the diagonal elements within (c) represent the differences in means of the corresponding variables between deceased and non-deceased patients.

ARA2. Additionally, $\mathscr{G}d$ demonstrates a slightly higher edge density of 0.06 compared to $\mathscr{G}nd$, which has an edge density of 0.05. This observation implies that deceased patients possessed more intricate medical histories prior to contracting COVID-19. This complexity is further supported by the edge entropy values, with $H(\mathscr{G}d) = 2.71$ and $H(\mathscr{G}nd) = 2.53$.

Specific differences include stronger associations in $\mathscr{G}_d$ between ARA2 and heart-related comorbidities and a notable disparity in the immunosuppressants-age and immunosuppressants-SAPS-3 edges between the graphs. Additionally, $\mathscr{G}_d$ demonstrates a stronger link between diarrhea and gender, while the smoking-SAPS-3 relationship is more pronounced in $\mathscr{G}_{nd}$. Certain medication nodes in $\mathscr{G}_{nd}$ exhibit strong connections with comorbidities.

The study identifies distinct pairwise relationships among demographic factors, comorbidities, medications, and symptoms. However, it is imperative to note that this descriptive approach does not establish causality nor attribute differences solely to COVID-19 treatment strategies.

Dynamic Feature Analysis

This section explores the dynamic aspects of COVID-19 treatment, focusing on the MTS of drug administration. We aim to analyze how drug treatments interact for both deceased and surviving patients over time. To achieve this objective, we have created eight distinct families of graphs by partitioning the patient journey into separate intervals, including symptom onset, emergency department stay, hospital stay, and ICU stay. For each of these intervals, we construct and compare graphs for deceased ($\mathscr{G}d$) and surviving ($\mathscr{G}nd$) patients. These graphs enable us to assess disparities in drug administration frequency (node size) and the strength of inter-drug relationships (edge strength) between the two patient groups.

Figure 6.2 presents the network analysis of drug treatments across four intervals: "Symptoms," "Emergency-Department Stay," "Hospital Stay," and "ICU Stay," each depicted over three columns representing deceased patients, survivors, and the adjacency matrix differences between these groups. The graphs demonstrate a time-varying pattern of drug use, with key observations for each interval. In the "Symptoms Interval," both $\mathscr{G}_d^{(1)}$ and $\mathscr{G}_{nd}^{(1)}$ show limited drug variety, with a pronounced correlation between hydroxychloroquine and lopinavir/ritonavir, particularly in deceased patients. The "Emergency-Department Interval" exhibits increased complexity, with disparities in edge density and entropy between $\mathscr{G}_d^{(2)}$ and $\mathscr{G}_{nd}^{(2)}$, indicating a more intricate drug network for survivors. The "Hospital Stay Interval" reveals a further expansion in the drug variety, with lopinavir/ritonavir, hydroxychloroquine, and corticosteroid emerging as the most common drugs. Here, $\mathscr{G}_d^{(3)}$ shows greater complexity compared to $\mathscr{G}_{nd}^{(3)}$, with significant correlations involving anakinra and tocilizumab in the former. The "ICU Stay Interval" presents the most connected drug network, with all nine drugs in play. Deceased patients' graphs ($\mathscr{G}_d^{(4)}$) exhibit higher edge density and entropy, indicating an increased variety in treatment attempts, with tocilizumab-remdesivir being a notable connection.

For a more granular analysis, we segment the 28-day "ICU Stay Interval" into four non-overlapping 7-day periods, creating a series of weekly graphs ($\mathscr{G}_d^{(\tau)}$ and $\mathscr{G}_{nd}^{(\tau)}$ for $\tau = 1, ..., 4$). This breakdown, shown in Fig. 6.3, reveals that the initial two weeks witnessed similar treatment complexities for both patient groups, with various drugs administered. However, in the subsequent weeks, the complexity of deceased patients' treatment regimes remains elevated, potentially indicating unsuccessful treatments.

Throughout the study, the drugs most commonly administered across all intervals were lopinavir/ritonavir, hydroxychloroquine, and corticosteroids. It was observed that the complexity of treatment regimens for deceased patients tended to be higher, especially in the later stages of their hospitalization. A significant trend identified was the gradual decrease in the usage of

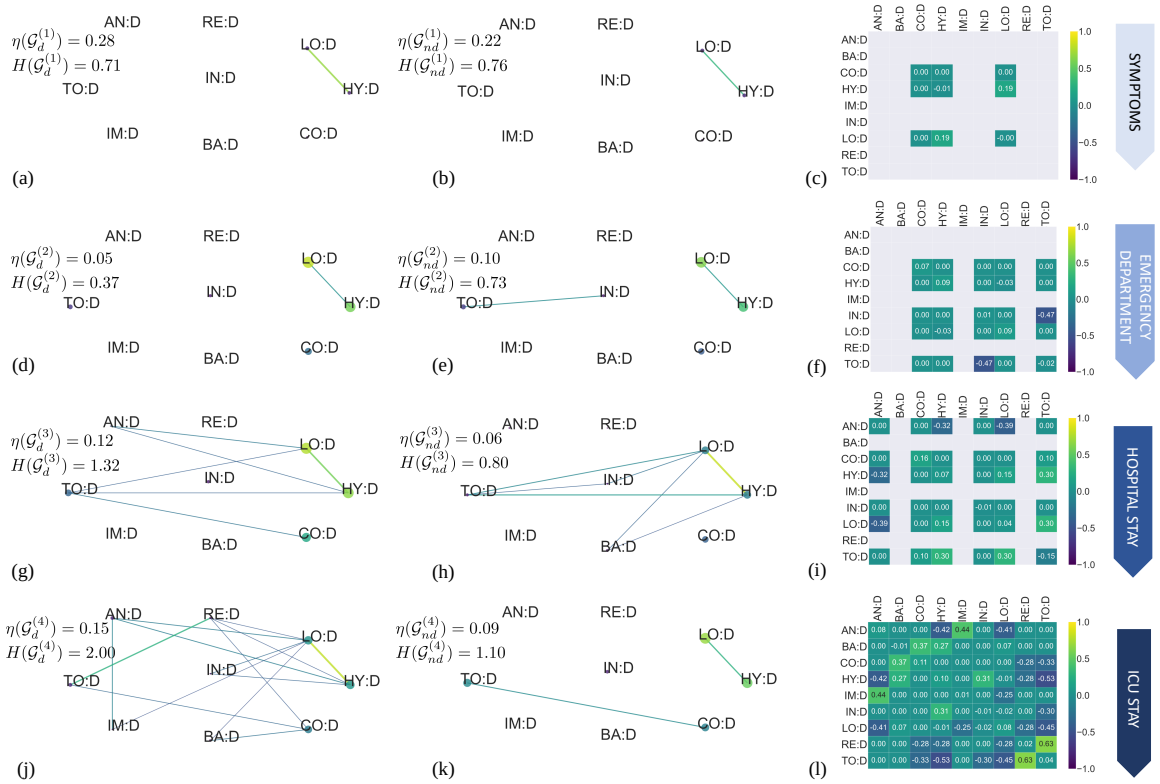


Figure 6.2: Network visualization of the drugs when considering non-deceased patients (central column) and deceased patients (left column). Each row corresponds to: the "Symptoms Interval" (a, b and c); the "Emergency-Department Stay Interval" (d,e and f); the "Hospital Stay Interval" (g,h and i); and the "ICU Stay Interval" (j,k and l). The matrices within each cell display the differences between the adjacency matrices for the deceased and non-deceased patient graphs (non-diagonal entries), while the diagonal entries depict the disparities in the average values of the variables between deceased and non-deceased patients for the corresponding intervals.

lopinavir/ritonavir and hydroxychloroquine after the first two weeks. Notably, this increased complexity in the treatment protocols for deceased patients during the latter half of their ICU stays did not appear to correlate with improved outcomes.

Integrating Static and Dynamic Features

The culmination of our graph analysis involves integrating static and dynamic features into a comprehensive graph. The bipartite graph links static variables (symptoms, comorbidities, medications) with dynamic drug administration variables, providing a holistic view of the patient data during the ICU stay. This integrated graph explores correlations between static patient characteristics and dynamic drug treatment patterns. The connections within this graph are categorized into three types: links between nodes within $\mathcal{N}^J$, those within $\mathcal{N}^I$, and links

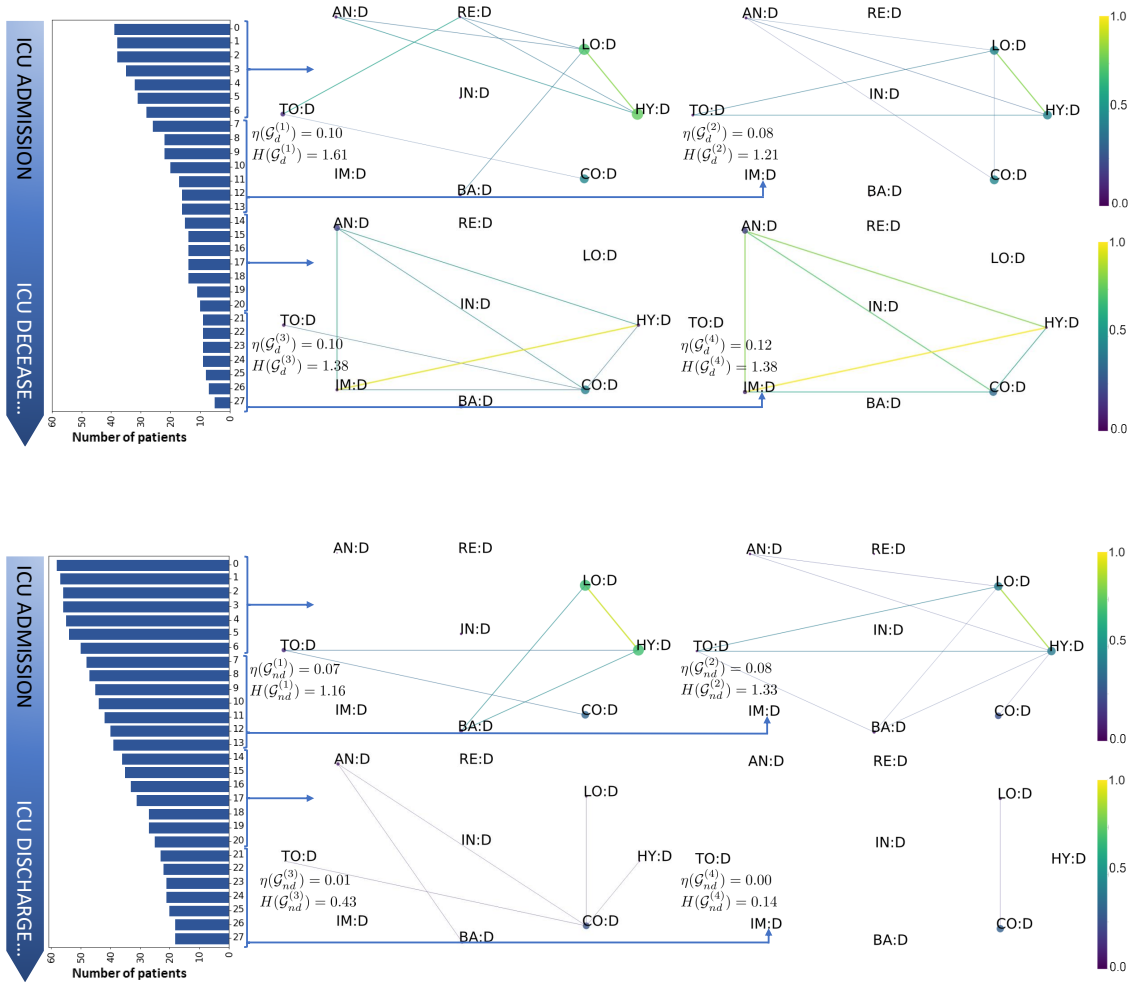


Figure 6.3: Visual network analysis for MTS drug treatments using non-overlapping 7 day intervals over a span of 28 days, with day $t = 0$ representing admission to the ICU) for (a) deceased who died and (b) patients who survived.

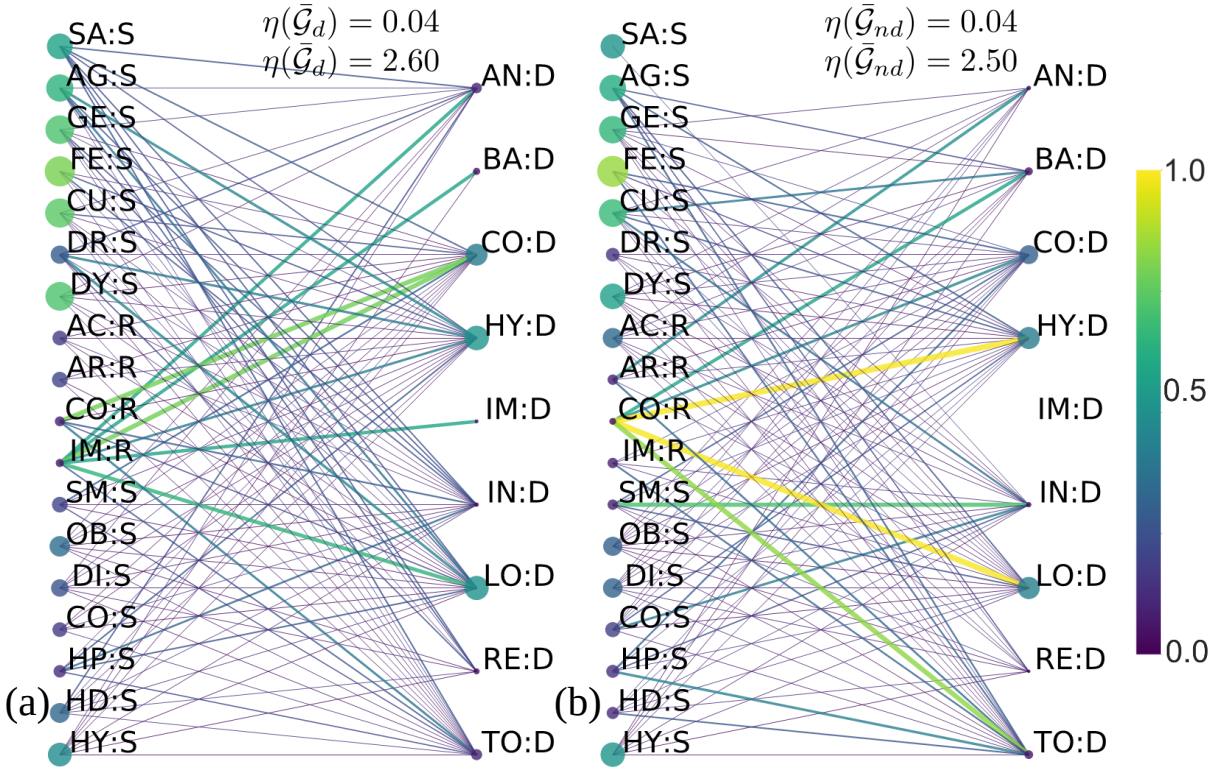


Figure 6.4: Visual network analytics illustrating the connections between the static variables (left-hand side nodes) and MTS variables (right-hand side nodes) during the "ICU Stay Interval" for (a) deceased patients and (b) non-deceased patients.

connecting nodes from $\mathcal{N}^J$ to $\mathcal{N}^I$. Essentially, this graph is a union of three subgraphs: the static variable graph, the dynamic variable graph, and a new subgraph depicting interactions between static and dynamic variables.

In focusing on the "ICU Stay Interval," we compute and illustrate the integrated graphs for both deceased ($\bar{\mathcal{G}}d$) and non-deceased ($\bar{\mathcal{G}}nd$) patients in Fig. 6.4. This representation shows $J = 18$ static variables on the left and $I = 9$ dynamic variables (drugs) on the right. The resulting subgraph is bipartite, as each link connects a node from $\mathcal{N}^J$ to $\mathcal{N}^I$, with no overlaps within these sets, simplifying the graph's interpretation. The edge weights are determined using Pearson correlation for links between age and SAPS-3 nodes in $\mathcal{N}^J$, drug nodes in $\mathcal{N}^I$, and Point-Biserial correlation for the remaining connections.

A comparative visual analysis of $\bar{\mathcal{G}}d$ and $\bar{\mathcal{G}}nd$ indicates structural similarities but highlights stronger connections in $\bar{\mathcal{G}}nd$, particularly between smoking and interferon beta-1b, and corticosteroids with regular medication variables. Notably, the complexity measures of these graphs, such as edge density (η) and entropy (H), are quite similar for both groups, with $\eta(\bar{\mathcal{G}}d) = 0.04$, $\eta(\bar{\mathcal{G}}nd) = 0.04$, $H(\bar{\mathcal{G}}d) = 2.6$, and $H(\bar{\mathcal{G}}_{nd}) = 2.5$. This similarity in graph structure indicates

that the main differences between the patient groups lie in the specific strengths of certain connections rather than in the overall topology of the graphs.

6.4 Results of Graph-Predictive Models

This section focuses on using graph-based methods to analyze complex interactions between variables in datasets of deceased and non-deceased COVID-19 patients. These graphs pave the way for developing schemes to predict patient outcomes based on their clinical data. Despite this approach's innovative potential, the small size of our dataset poses challenges to its reliability. However, sharing preliminary findings is beneficial for evaluating the effectiveness of such approaches. Nevertheless, sharing preliminary early results is beneficial for evaluating the potential and effectiveness of such approaches.

To clarify how to design of a graph-based predictive model, consider a scenario where $\mathbf{z}_i$ represents a vector denoting the proportion of drug intakes for a given patient i. Utilizing the graphs $\mathscr{G}_d$ and $\mathscr{G}_{nd}$, we can construct a two-feature vector $\mathbf{f}_i = [\mathbf{z}_i^T(\mathbf{I}+\mathbf{A}_d)^{-1}\mathbf{z}_i;\ \mathbf{z}_i^T(\mathbf{I}+\mathbf{A}_{nd})^{-1}\mathbf{z}_i]$ for each patient i, as proposed in Marques et al. (2017) [198]. In this framework, the elements of $\mathbf{f}_i$ evaluate the smoothness of $\mathbf{z}_i$ across the respective graphs. This assessment is crucial, as it quantifies the degree of variance in the patient's drug intake relative to the inherent variability encapsulated within the adjacency matrices $\mathbf{A}_d$ and $\mathbf{A}_{nd}$ [85].

The efficacy of such classifiers is related to their specific architecture. However, initial experiments employing elementary classifiers, such as the nearest centroid classifier and a 3-nearest neighbors classifier [124], have demonstrated promising results, with accuracy and specificity exceeding 70%. Furthermore, when employing a rudimentary leave-one-out cross-validation approach, the performance only decreases marginally (approximately 5%), indicating robustness in this preliminary methodology [124].

These early results should be viewed as an exploratory step into the potential application of graph-based predictive models in clinical settings. Additional details into this methodology are provided in the online appendix [199].

Results of Feature Importance within Bootstrap Confidence Intervals

We conducted a hypothesis test using CIB (see subsection 4.2.1 for the theoretical explanation), resampling 3,000 times to establish the statistical significance of various parameters related to COVID-19 treatment.

The analysis examined demographic variables, comorbidities, medications, and symptoms.

They determined the CIs for both binary and numerical features, as illustrated in Fig. 6.5 (a) and (b) of their bootstrapping results. Significantly, heart disease comorbidity and SAPS-3 scores emerged as statistically significant variables. This was evidenced by a considerable divergence of the empirical distribution's mass from zero when comparing the means of these variables between the two populations. Notably, certain variables like dyspnea, hematological cancer, transplantation, ACE inhibitors, and HIV presented borderline statistical significance. Their 95% CIs marginally overlapped zero, suggesting that a lower confidence threshold (e.g., 90%) might render these variables significant.

The research then shifted to analyzing the MTS, correlating the usage of various drugs in treating COVID-19 with patient outcomes across the previously presented stages, as delineated in Fig. 6.5 (c). As we work with MTS variables, we begin by computing the ratio of drug intake days to the total duration of the interval for each patient within a given interval. Subsequently, we employ a bootstrapping procedure on the patient data, wherein we evaluate $CI_{\Delta P}$ of each drug and interval. It is essential to note that our statistic of interest in this context is the mean value.

During the "Symptoms Interval," no significant differences were observed in drug administration between deceased and surviving patients. However, in the "Emergency-Department Stay Interval," deceased patients more frequently received lopinavir/ritonavir and chloroquine. Intriguingly, during the "Hospital Interval," chloroquine was more associated with surviving patients, alongside tocilizumab, interferon beta-1b, and imatinib. In contrast, deceased patients more often received hydroxychloroquine, corticosteroids, and azithromycin. The "ICU Stay Interval" analysis revealed that deceased patients were more likely to be treated with tocilizumab, lopinavir/ritonavir, imatinib, hydroxychloroquine, corticosteroids, and anakinra.

6.5 Conclusions

This chapter presents a groundbreaking approach to understanding the complex trajectories of COVID-19 patients in critical care. Employing graph-based network analytics provides a holistic view of patient data, a significant advancement over traditional linear models.

The research makes several key contributions. It introduces a novel methodological framework, employing graph-based techniques to process and analyze static and dynamic clinical features. This approach offers a more comprehensive understanding of patient trajectories in ICU settings. The study's bifurcation analysis divides the patient population into deceased and non-deceased groups and allows for a nuanced comparative analysis that reveals critical patterns

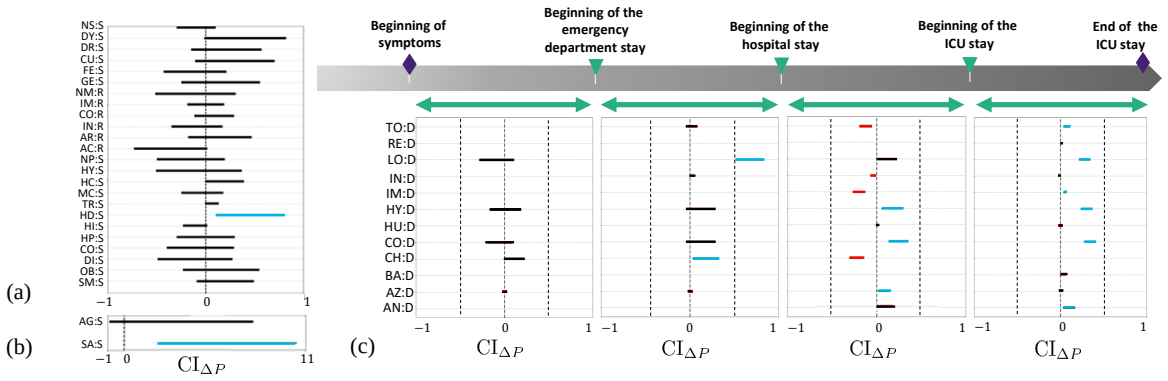


Figure 6.5: $CI_{\Delta P}$ when employing bootstrapping techniques on both deceased and non-deceased patient cohorts, with the mean as the chosen statistic. The figure presents the results for three distinct feature categories: (a) binary static features encompassing demographic variables, comorbidities, medication, and symptoms; (b) numerical static features, specifically age and SAPS-3 (Simplified Acute Physiology Score); and (c) the ratio denoting the patient's drug intake days relative to the total duration of the patient's interval. Features exhibiting no statistically significant difference ($0 \in CI_{\Delta P}$) are depicted in black. Blue and red bars indicate features with statistically higher and lower average values, respectively, for deceased patients.

and differences in COVID-19 patient outcomes. A significant aspect of the chapter is its use of various correlation coefficients to construct detailed graphs, shedding light on the intricate relationships between clinical variables. This enhanced interpretability of high-dimensional data is crucial for understanding the complex interplay of patient comorbidities, medications, and symptomatology.

The chapter also makes pioneering strides in exploring graph-based predictive models. Although preliminary, this exploration sets a foundation for future research utilizing network analytics for outcome prediction in clinical settings. The implications of this research are far-reaching. It offers the potential for enhanced patient care by providing healthcare providers with a deeper understanding of patient data, enabling more tailored and effective treatment strategies. Additionally, the approach detailed in this thesis can be adapted to other diseases and clinical settings, broadening its impact and applicability.

In conclusion, this thesis represents a significant advancement in the application of data and network analytics in healthcare, particularly in the critical care of COVID-19 patients. It not only enhances our understanding of complex patient data but also paves the way for future innovations in healthcare analytics. The potential of this research extends beyond COVID-19, offering a new approach to tackling various healthcare challenges with sophisticated analytical techniques.

Chapter 7

Conclusions and Future Work

In this concluding chapter of the dissertation, we synthesize the main findings of the research and discuss their impact and contributions to the field. This chapter is structured into three subsections. The first, "Conclusions," recaps the main results and their contributions to the field, linking them to the research questions and objectives outlined in Chapter 1. The subsequent section, "Limitations and Future Work," acknowledges the challenges faced during the research and suggests possible areas for further investigation that could improve or expand the current work. The chapter concludes with a "Concluding Remark," which briefly summarizes the study's impact and relevance, suggesting how it could influence future research and practice.

7.1 Conclusions

This thesis aims to advance DL and graph analytics for modeling clinical data of infectious diseases by addressing challenges such as the limited number of patients, the irregularity in the clinical MTS, and explainability.

The current global health situation is impacted by a rising wave of infectious diseases. That situation is a challenge with economic and social impacts, which could lead to societal changes. Consequently, it is essential to formulate and execute innovative strategies to tackle this challenge effectively. Among the potential solutions, applying data-driven methodologies is promising, providing healthcare professionals with advanced tools and capabilities to manage this critical issue. However, constructing data-driven models trained with real clinical data introduces different challenges. These encompass the restricted number of patient data, irregularities inherent in clinical MTS, and lack of interpretability, which altogether compromise the

efficacy of these models. Such complexity demands careful attention to ensure the successful use and efficacy of data-driven initiatives in the fight against infectious illnesses.

Firstly, in order to deal with the limited number of patients, we focus on two simple but effective methods to address it: i) undersampling the majority class and ii) defining asymmetric misclassification costs. When following an undersampling strategy, samples from the majority class are randomly discarded until the number of elements in the majority and minority populations is similar. In the cost-sensitive approach, errors in a sample from the minority class are penalized more than those from the majority class.

Secondly, given the time series' irregular lengths and sampling intervals, it was necessary to employ various modelling approaches to address these discrepancies. We have employed different modellings: i) statistical modeling, ii) time-slot windowed modeling, iii) time-slot masked modeling, and iv) graph modeling. Initially, statistical modeling—characterized by summarizing the time series into a suite of statistics such as the mean, median, and standard deviation—is employed to extract patterns and relationships within the data, serving as a foundational tool for preliminary analysis. This method, however, may inadequately capture complex temporal dynamics. However, time-slot windowed modeling facilitates the data analysis within specific temporal windows, thus accounting for temporal fluctuations. However, it may falter when handling missing data points. To rectify this, time-slot masked modeling is introduced. This approach strategically masks specific time steps to handle missing data, providing a more robust analysis in the face of incomplete data sets. Despite its strengths, this model alone may not fully capture the interrelations between different variables. Hence, the final approach, correlation-based graph modeling, is implemented. Correlation-based graphs are particularly useful for analyzing irregular MTS because they do not rely on uniform time intervals to examine relationships among MTS. These graphs evaluate the strength of statistical relationships between different observations, regardless of their occurrence time. Collectively, these diverse modeling techniques provide a multi-faceted approach to address the complexities and irregularities present in clinical MTS data, thereby bolstering the integrity and reliability of the ensuing analyses.

Thirdly, enhancing model explainability, a crucial aspect of data-driven methodologies, demands the implementation of a gamut of DL methods, each distinctive in its nature and functionality. To address explainability, we use a set of explainable DL methods of different nature: i) white-box explainable architectures, ii) post-hoc models, and iii) interpretable mechanisms integrated into black-box models. In ML, a set of white-box models prioritizes transparency. This transparency allows a clear understanding of their internal operations and decision-making

processes. However, these models tend to be simpler and often fail to achieve the performance levels of more complex models. In some situations, the use of those complex models known as black-box models is necessary, which compromises their transparency. In such scenarios, post-hoc models are introduced. These models understand the decision-making process of complex, frequently black-box models and offer explanations after the fact. They reveal the "logic" behind the model's outputs, providing insights into its functionality. While post-hoc models offer explanations for decisions after they are made, there is a need for a more proactive approach. Thus, interpretable mechanisms integrated directly into black-box models are proposed. This integration allows the model to offer immediate explanations, making its decision-making process transparent. Together, these three approaches create a robust toolkit for addressing explainability in DL, which is crucial for engendering user trust and facilitating effective utilization of these models in healthcare contexts.

Finally, from a clinical standpoint, the implications drawn from this thesis are in line with established literature, potentially validating the data-driven models in our specific clinical scenarios. For predicting the onset of AMR, our analysis underscores the pivotal role of mechanical ventilation and the number of AMR patients in the ICU. Certain families of antibiotics, such as Carbapenems, Cephalosporins, Glycopeptides, and Penicillins, emerge also as significant features. This corroborates previous findings, highlighting the widespread use of these antibiotics and the association of invasive procedures, like mechanical ventilation, with heightened infection and resistance risks.

Transitioning to the domain of COVID-19, our graphs confirm the prevalence of hypertension, diabetes, and obesity as primary comorbidities, while heart disease and smoking are more common among deceased patients. Additionally, fever, cough, and dyspnea emerge as predominant symptoms. Moreover, deceased patients exhibit a more intricate medical history compared to their non-deceased counterparts. Patient complexity intensifies during hospital and ICU stays, reflecting critical conditions necessitating more complex antibiotic treatments. Furthermore, during the initial phase of the pandemic, the combination of lopinavir/ritonavir, hydroxychloroquine, and corticosteroids emerged as the standard treatment. Interestingly, the association between age and hydroxychloroquine/lopinavir/ritonavir was slightly more potent among deceased patients. However, it is essential to note that the World Health Organization has reported that the hydroxychloroquine-and-lopinavir/ritonavir combination does not influence mortality rates among hospitalized COVID-19 patients [200].

It is important to remember that although these models provide encouraging insights, more validation by clinical professionals through rigorous controlled trials is necessary. This is an

important point that this work does not address.

7.2 Limitations and Future Work

While this thesis has addressed critical challenges for modeling and analyzing clinical data associated with infectious diseases using DL and graph analytics methods, some limitations and potential future work remain.

Focusing on the AMR dataset, the methodology outlined in this dissertation shows potential for enhancing the administration of antibiotic treatments and the management of patient care within ICUs. These strategies could prevent the spread of pathogens in these units. Additionally, it may facilitate the prevention of potential AMR outbreaks within the ICU environment. However, there is potential for further improvement of our future research trajectory through more precise methodologies and enhanced analytical techniques. From a clinical perspective, we intend to integrate novel features from supplemental sources, such as artificial nutrition and hematological tests, into our current model. The integration of these variables could enhance model performance and clinical interpretability. Furthermore, following encouraging results reported by Hernandez et al. [25], we aspire to create unique models for each type of AMR bacterial emergence. From an ML standpoint, we plan to explore alternative NN architectures along with distance and similarity measures explicitly tailored for MTS and multimodal data [201]. Another relevant line of work is to generalize the proposed model, which currently focuses on predicting the first AMR, to provide a score of the risk of acquiring AMR infections daily. Such an advancement could significantly aid real-time clinical decision-making, allowing the dynamic adaptation of patient treatment strategies and swift isolation procedures, thereby impeding AMR transmission within the ICU. Lastly, given the extensive presence of MTS and multimodal data across various sectors, we are keen to adapt and apply our multimodal architecture to other domains, including finance, marketing, and transportation, thereby broadening the applicability of our methodology beyond healthcare.

Focusing on the COVID-19 dataset, our future work encompasses several directions. First, from a clinical perspective, we plan to collect and analyze a significantly larger dataset. This expansion will include additional data from current patients, including post-COVID-19 symptoms, and data gathered across various healthcare institutions. We also intend to incorporate new features like artificial nutrition and blood test results. Enriching the dataset will help validate our preliminary findings, enhance the reliability of our models, and support the development of advanced predictive algorithms. Additionally, we aim to rigorously investigate the

impact of COVID-19 on the increase in AMR. In the context of ML, the current research has only employed a hypothesis test for feature selection, drawing upon bootstrap and confidence intervals. We propose exploring alternative methodologies that foster a more dynamic interaction with the model. We shall also integrate advanced tools to preprocess MTS that may enhance robustness in assessing temporal dependencies. Our efforts shall encompass two separate yet complementary facets of graph representation: i) the use of alternative algorithms to discern graph structures from the data and ii) the examination of an expansive range of network-based metrics. A crucial area of focus in our future work will be the construction of prediction models poised to assist clinicians in making informed decisions. This encompasses models based on MTS, graphs, or both. With the enlargement and diversification of our dataset, we anticipate developing more intricate mechanisms, such as graph neural networks. These sophisticated tools will ensure that our models are robust and adaptable to the evolving landscape of healthcare data.

7.3 Concluding Remark

In conclusion, this dissertation highlights the critical need for effective collaboration between ML specialists and domain experts in the clinical context to predict the emergence of infectious diseases. It emphasizes the necessity of interdisciplinary approaches to address the complex real-world challenges presented. As DL and infectious disease prediction continue to develop and converge, this dissertation will serve as a foundational work, stimulating further advancements in this critical area of research. By establishing robust collaborations and implementing cutting-edge research techniques, the scientific community can effectively contribute to addressing this urgent challenge.

Bibliography

[1] Christopher J.L. Murray et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet, 2022.

[2] Jim O'Neill. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. Technical report, Review on Antimicrobial Resistance, 2014.

[3] Jaime Pinilla, Patricia Barber, Laura Vallejo-Torres, Silvia Rodríguez-Mireles, Beatriz G López-Valcárcel, and Luis Serra-Majem. The economic impact of the SARS-COV-2 (COVID-19) pandemic in Spain. International Journal of Environmental Research and Public Health, 18(9):4708, 2021.

[4] Chensi Cao et al. Deep learning and its applications in biomedicine. Genomics, proteomics & bioinformatics, 16(1):17–32, 2018.

[5] Rachel E. Baker et al. Infectious disease in an era of global change. Nature Reviews Microbiology, 20(4):193–205, 2022.

[6] H.C. Yashavantha Rao and Chelliah Jayabaskaran. The emergence of a novel coronavirus (SARS-CoV-2) disease and their neuroinvasive propensity may affect in COVID-19 patients. Journal of Medical Virology, 92(7):786–790, 2020.

[7] World Health Organization et al. Antimicrobial resistance surveillance in europe 2020–2022 data. Technical report, World Health Organization. Regional Office for Europe, 2022.

[8] Md Abdus Salam et al. Antimicrobial resistance: a growing serious threat for global public health. In Healthcare, volume 11, page 1946. MDPI, 2023.

[9] Francesca Prestinaci, Patrizio Pezzotti, and Annalisa Pantosti. Antimicrobial resistance: a global multifaceted phenomenon. Pathogens and global health, 109(7):309–318, 2015.

[10] Karen O'Leary. The global burden of antimicrobial resistance. Nature Medicine, 2022.

[11] World Health Organization. Global Action Plan on Antimicrobial Resistance. Health & Medical Publishing Group, page 28, 2015.

[12] Jim O'Neill. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. Review on Antimicrobial Resistance, 2016.

[13] David E. Bloom and Daniel Cadarette. Infectious disease threats in the twenty-first century: strengthening the global response. Frontiers in immunology, 10:549, 2019.

[14] Rafael Garcia-Carretero, Oscar Vazquez-Gomez, Ruth Gil-Prieto, and Angel Gil-de Miguel. Hospitalization burden and epidemiology of the covid-19 pandemic in spain (2020–2021). BMC Infectious Diseases, 23(1):476, 2023.

[15] Mariana G. López et al. The first wave of the COVID-19 epidemic in Spain was associated with early introductions and fast spread of a dominating genetic variant. Nature genetics, 53(10):1405–1414, 2021.

[16] Robert L. Ohsfeldt, Casey Kar-Chan Choong, Patrick L Mc Collam, Hamed Abedtash, Kari A. Kelton, and Russel Burge. Inpatient hospital costs for covid-19 patients in the united states. Advances in therapy, 38:5557–5595, 2021.

[17] Andre Esteva et al. A guide to deep learning in healthcare. Nature medicine, 25(1):24–29, 2019.

[18] Gunjan Pahuja and T.N. Nagabhushan. A comparative study of existing machine learning approaches for parkinson's disease detection. IETE Journal of Research, 67(1):4–14, 2021.

[19] Jenna Wong, Mara Murray Horwitz, Li Zhou, and Sengwee Toh. Using machine learning to identify health outcomes from electronic health record data. Current epidemiology reports, 5:331–342, 2018.

[20] Hrayr Harutyunyan, Hrant Khachatrian, David Kale, Greg Ver Steeg, and Aram Galstyan. Multitask learning and benchmarking with clinical time series data. Scientific data, 6(1):96, 2019.

[21] Sergio Martínez-Agüero, Cristina Soguero-Ruiz, Jose M Alonso-Moral, Inmaculada Mora-Jiménez, Joaquín Álvarez-Rodríguez, and Antonio G Marques. Interpretable clin-

ical time-series modeling with intelligent feature selection for early prediction of antimicrobial multidrug resistance. Future Generation Computer Systems, 133:68–83, 2022.

[22] Sergio Martínez-Agüero, Antonio G Marques, Inmaculada Mora-Jiménez, Joaquín Alvárez-Rodríguez, and Cristina Soguero-Ruiz. Multimodal interpretable data-driven models for early prediction of antimicrobial multidrug resistance using multivariate time-series. arXiv preprint arXiv:2402.06295, 2024.

[23] Àlvar Hernàndez-Carnerero, Miquel Sànchez-Marrè, Inmaculada Mora-Jiménez, Cristina Soguero-Ruiz, Sergio Martínez-Agüero, and Joaquín Álvarez-Rodríguez. Dimensionality reduction and ensemble of lstms for antimicrobial resistance prediction. Artificial intelligence in medicine, 138:102508, 2023.

[24] Lidia Pascual-Sánchez, Inmaculada Mora-Jiménez, Sergio Martínez-Agüero, Joaquín Álvarez-Rodríguez, and Cristina Soguero-Ruiz. Predicting multidrug resistance using temporal clinical data and machine learning methods. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pages 2826–2833. IEEE, 2021.

[25] Àlvar Hernàndez-Carnerero, Miquel Sànchez-Marrè, Inmaculada Mora-Jiménez, Cristina Soguero-Ruiz, Sergio Martínez-Agüero, and Joaquín Álvarez-Rodríguez. Antimicrobial resistance prediction in intensive care unit for pseudomonas aeruginosa using temporal data-driven models. 2021.

[26] Sergio Martínez-Agüero, Antonio G. Marques, Inmaculada Mora-Jiménez, Joaquín Alvárez-Rodríguez, and Cristina Soguero-Ruiz. Multimodal interpretable data-driven models for early prediction of antimicrobial multidrug resistance using multivariate time-series, 2024.

[27] Sergio Martínez-Agüero, Antonio G Marques, Inmaculada Mora-Jiménez, Joaquín Álvarez-Rodríguez, and Cristina Soguero-Ruiz. Data and network analytics for covid-19 icu patients: a case study for a spanish hospital. IEEE Journal of Biomedical and Health Informatics, 25(12):4340–4353, 2021.

[28] Tanvir Mahtab Uddin et al. Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. Journal of infection and public health, 14(12):1750–1766, 2021.

[29] Ruchita Balasubramanian, Thomas P Van Boeckel, Yehuda Carmeli, Sara Cosgrove, and Ramanan Laxminarayan. Global incidence in hospital-associated infections resistant to

antibiotics: An analysis of point prevalence surveys from 99 countries. Plos Medicine, 20(6):e1004178, 2023.

[30] Carolyn Michael, Dale Dominey-Howes, and Maurizio Labbate. The antimicrobial resistance crisis: causes, consequences, and management. Frontiers in public health, 2:145, 2014.

[31] Gamze Kalın, Emine Alp, Arthur Chouaikhi, and Claire Roger. Antimicrobial multidrug resistance: Clinical implications for infection management in critically ill patients. Microorganisms, 11(10):2575, 2023.

[32] Jane Wairimu Maina, Frank Gekara Onyambu, Peter Shikuku Kibet, and Abednego Moki Musyoki. Multidrug-resistant gram-negative bacterial infections and associated factors in a kenyan intensive care unit: a cross-sectional study. Annals of Clinical Microbiology and Antimicrobials, 22(1):85, 2023.

[33] Charles-Edouard Luyt, Nicolas Bréchot, Jean-Louis Trouillet, and Jean Chastre. Antibiotic stewardship in the intensive care unit. Critical care, 18:1–12, 2014.

[34] Gabriel Birgand, Puneet Dhar, and Alison Holmes. The threat of antimicrobial resistance in surgical care: the surgeon's role and ownership of antimicrobial stewardship. British Journal of Surgery, 110(12):1567–1569, 2023.

[35] Cristina Muñoz Madero. Primer año del plan estratégico y de acción para reducir el riesgo de selección y diseminación de resistencia a los antibióticos. Albéitar: publicación veterinaria independiente, pages 4–6, 2016.

[36] Eric Pelfrene, Radu Botgros, and Marco Cavaleri. Antimicrobial multidrug resistance in the era of covid-19: a forgotten plight? Antimicrobial Resistance & Infection Control, 10(1):1–6, 2021.

[37] Diana Rofail et al. Patient experience of symptoms and impacts of covid-19: a qualitative investigation with symptomatic outpatients. BMJ open, 12(5):e055989, 2022.

[38] Ben Hu, Hua Guo, Peng Zhou, and Zheng-Li Shi. Characteristics of sars-cov-2 and covid-19. Nature Reviews Microbiology, 19(3):141–154, 2021.

[39] Mahesh Jayaweera, Hasini Perera, Buddhika Gunawardana, and Jagath Manatunge. Transmission of covid-19 virus by droplets and aerosols: A critical review on the unresolved dichotomy. Environmental research, 188:109819, 2020.

[40] Working group for the surveillance, control of COVID-19 in Spain, et al. The first wave of the covid-19 pandemic in spain: characterisation of cases and risk factors for severe outcomes, as at 27 april 2020. Eurosurveillance, 25(50):2001431, 2020.

[41] WHO. COVID-19 weekly epidemiological update - 1 December 2020. Technical report, WHO, 2020.

[42] Sarah Wahlster et al. The coronavirus disease 2019 pandemic's effect on critical care resources and providers: A global survey. Chest, 31(10):1336–1344, 2020.

[43] Michael C. Blayney et al. Prevalence, characteristics, and longer-term outcomes of patients with persistent critical illness attributable to covid-19 in scotland: a national cohort study. British Journal of Anaesthesia, 128(6):980–989, 2022.

[44] Napier House and Herbert Holborn. ICNARC report on COVID-19 in critical care. Technical report, ICNARC, 2020.

[45] Zunyou Wu and Jennifer M. McGoogan. Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention. jama, 323(13):1239–1242, 2020.

[46] Qingyun Wang et al. COVID-19 literature knowledge graph construction and drug repurposing report generation. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations, pages 66–77. Association for Computational Linguistics, 2021.

[47] Pérez de la Lastra et al. Antimicrobial resistance in the covid-19 landscape: is there an opportunity for anti-infective antibodies and antimicrobial peptides? Frontiers in Immunology, page 2698, 2022.

[48] Richard Harrison Shryock. The development of modern medicine: an interpretation of the social and scientific factors involved. University of Pennsylvania Press, 2017.

[49] Ayesha Amjad, Piotr Kordel, and Gabriela Fernandes. A review on innovation in healthcare sector (telehealth) through artificial intelligence. Sustainability, 15(8):6655, 2023.

[50] Angelos I. Stoumpos, Fotis Kitsios, and Michael A. Talias. Digital transformation in healthcare: Technology acceptance and its applications. International journal of environmental research and public health, 20(4):3407, 2023.

[51] Sophie Isabelle Lambert et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digital Medicine, 6(1):111, 2023.

[52] Christine Sinsky et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Annals of internal medicine, 165(11):753–760, 2016.

[53] Amy Sitapati et al. Integrated precision medicine: the role of electronic health records in delivering personalized treatment. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 9(3):e1378, 2017.

[54] Edmond Li, Jonathan Clarke, Hutan Ashrafian, Ara Darzi, and Ana Luisa Neves. The impact of electronic health record interoperability on safety and quality of care in high-income countries: Systematic review. Journal of Medical Internet Research, 24(9):e38144, 2022.

[55] Kasaw Adane, Mucheye Gizachew, and Semalegne Kendie. The role of medical data in efficient patient care delivery: a review. Risk Management and Healthcare Policy, pages 67–73, 2019.

[56] Iqbal H Sarker. Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3):160, 2021.

[57] Edward Choi, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2):361–370, 2016.

[58] Benjamin Shickel, Tyler J. Loftus, Lasith Adhikari, Tezcan Ozrazgat-Baslanti, Azra Bihorac, and Parisa Rashidi. Deepsofa: a continuous acuity score for critically ill patients using clinically interpretable deep learning. Scientific Reports, 9(1):1–12, 2019.

[59] Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh. Predicting healthcare trajectories from medical records: A deep learning approach. Journal of Biomedical Informatics, 69:218–229, 2017.

[60] Cao Xiao, Edward Choi, and Jimeng Sun. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. Journal of the American Medical Informatics Association, 25(10):1419–1428, 2018.

[61] Tsipi Heart, Ofir Ben-Assuli, and Itamar Shabtai. A review of phr, emr and ehr integration: A more personalized healthcare and public health policy. Health Policy and Technology, 6(1):20–25, 2017.

[62] Kathleen M. Brelsford, Susan E. Spratt, and Laura M. Beskow. Research use of electronic health records: patients' perspectives on contact by researchers. Journal of the American Medical Informatics Association, 25(9):1122–1129, 2018.

[63] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9):1263–1284, 2009.

[64] Anat Reiner Benaim et al. Analyzing medical research results based on synthetic data and their relation to real data results: Systematic comparison from five observational studies. JMIR Medical Informatics, 8(2):e16492, 2020.

[65] Andrew Yale, Saloni Dash, Ritik Dutta, Isabelle Guyon, Adrien Pavao, and Kristin P. Bennett. Generation and evaluation of privacy preserving synthetic health data. Neurocomputing, 416:244–255, 2020.

[66] Steven M. Williamson and Victor Prybutok. Balancing privacy and progress: A review of privacy challenges, systemic oversight, and patient perceptions in ai-driven healthcare. Applied Sciences, 14(2):675, 2024.

[67] Jionglin Wu, Jason Roy, and Walter F. Stewart. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Medical Care, pages 106–113, 2010.

[68] K. Shailaja, Banoth Seetharamulu, and M.A. Jabbar. Machine learning in healthcare: A review. In 2018 Second international conference on electronics, communication and aerospace technology (ICECA), pages 910–914. IEEE, 2018.

[69] Eliza Strickland. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum, 56(4):24–31, 2019.

[70] Diego A. Martinez et al. An electronic dashboard to monitor patient flow at the Johns Hopkins Hospital: communication of key performance indicators using the donabedian model. Journal of Medical Systems, 42:1–8, 2018.

[71] DonHee Lee and Seong No Yoon. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. International Journal of Environmental Research and Public Health, 18(1):271, 2021.

[72] Zakhriya Alhassan, A. Stephen McGough, Riyad Alshammari, Tahani Daghstani, David Budgen, and Noura Al Moubayed. Type-2 diabetes mellitus diagnosis from time series clinical data using deep learning models. In Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27, pages 468–478. Springer, 2018.

[73] Feng Xie et al. Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. Journal of biomedical informatics, 126:103980, 2022.

[74] Shih-Cheng Huang, Anuj Pareek, Saeed Seyyedi, Imon Banerjee, and Matthew P. Lungren. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ digital medicine, 3(1):1–9, 2020.

[75] Jing Gao, Peng Li, Zhikui Chen, and Jianing Zhang. A survey on deep learning for multimodal data fusion. Neural Computation, 32(5):829–864, 2020.

[76] Tong Meng, Xuyang Jing, Zheng Yan, and Witold Pedrycz. A survey on machine learning for data fusion. Information Fusion, 57:115–129, 2020.

[77] Nasir Hayat, Krzysztof J. Geras, and Farah E. Shamout. MedFuse: Multi-modal fusion with clinical time-series data and chest X-ray images. arXiv preprint arXiv:2207.07027, 2022.

[78] Chen-Ying Hung, Ching-Heng Lin, Chi-Sen Chang, Jeng-Lin Li, and Chi-Chun Lee. Predicting gastrointestinal bleeding events from multimodal in-hospital electronic health records using deep fusion networks. In Proceedings of 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 2447–2450. IEEE, 2019.

[79] Rui Li, Fenglong Ma, and Jing Gao. Integrating multimodal electronic health records for diagnosis prediction. In Proceedings of AMIA Annual Symposium Proceedings, volume 2021, page 726. American Medical Informatics Association, 2021.

[80] Shuai Niu, Qing Yin, Yunya Song, Yike Guo, and Xian Yang. Label dependent attention model for disease risk prediction using multimodal electronic health records. In Proceedings of 2021 IEEE International Conference on Data Mining (ICDM), pages 449–458. IEEE, 2021.

[81] Mark Newman. Networks. Oxford University Press, 2018.

[82] William L. Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. Bulletin of the Technical Committee on Data Engineering, 40(3):52–74, 2017.

[83] Antonio G. Marques et al. Graph signal processing: Foundations and emerging directions. IEEE Signal Process. Mag., 37(6):11–13, 2020.

[84] Eric D. Kolaczyk and Gábor Csárdi. Statistical analysis of network data with R, volume 65. Springer, 2014.

[85] Gonzalo Mateos, Santiago Segarra, Antonio G. Marques, and Alejandro Ribeiro. Connecting the dots: Identifying network structure via graph signal processing. IEEE Signal Process. Mag., 36(3):16–43, 2019.

[86] Nicholas D. Soulakis, Matthew B. Carson, Young Ji Lee, Daniel H. Schneider, Connor T. Skeehan, and Denise M. Scholtens. Visualizing collaborative electronic health record usage for hospitalized patients with heart failure. J. of the Amer. Medical Informat. Association, 22(2):299–311, 2015.

[87] Anis Yousefi, Negin Mastouri, and Kamran Sartipi. Scenario-oriented information extraction from electronic health records. In 22nd IEEE International Symposium on Computer-Based Medical Systems, pages 1–5, 2009.

[88] Daniel M. Bean et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. Scientific Rep., 7(1):1–11, 2017.

[89] Xiangxiang Zeng et al. Repurpose open data to discover therapeutics for COVID-19 using deep learning. J. Proteome Research, 19(11):4624–4636, 2020.

[90] Antonio Gomez-Exposito, Jose A. Rosendo-Macias, and Miguel A. Gonzalez-Cagigal. Monitoring and tracking the evolution of a viral epidemic through nonlinear kalman filtering: Application to the covid-19 case. IEEE J. of Biomed. and Health Informat., pages 1–1, 2021.

[91] Jayanthi Devaraj et al. Forecasting of covid-19 cases using deep learning models: Is it reliable and practically significant? Results in Physics, 21:103817, 2021.

[92] Benjamin Shickel, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics, 22(5):1589–1604, 2017.

[93] Francesco Piccialli, Vittorio Di Somma, Fabio Giampaolo, Salvatore Cuomo, and Giancarlo Fortino. A survey on deep learning in medicine: Why, how and when? Information Fusion, 66:111–137, 2021.

[94] Garrett B. Goh, Nathan O. Hodas, Charles Siegel, and Abhinav Vishnu. Smiles2vec: An interpretable general-purpose deep neural network for predicting chemical properties. arXiv preprint arXiv:1712.02034, 2017.

[95] Alex John London. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Center Report, 49(1):15–21, 2019.

[96] Mark Sendak et al. "The human body is a black box" supporting clinical decision-making with deep learning. In Proceedings of the 2020 conference on fairness, accountability, and transparency, pages 99–109, 2020.

[97] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through propagating activation differences. In Proceedings of International conference on machine learning, pages 3145–3153. PMLR, 2017.

[98] Christoph Molnar. Interpretable machine learning. Bookdown, 2020.

[99] Joseph D. Janizek, Pascal Sturmfels, and Su-In Lee. Explaining explanations: Axiomatic feature interactions for deep networks. J. Mach. Learn. Res., 22:104–1, 2021.

[100] Jonathan Crabbé and Mihaela Van Der Schaar. Explaining time series predictions with dynamic masks. In Proceedings of International Conference on Machine Learning, pages 2166–2177. PMLR, 2021.

[101] Kedar Dhamdhere, Mukund Sundararajan, and Qiqi Yan. How important is a neuron? arXiv preprint arXiv:1805.12233, 2018.

[102] Avanti Shrikumar, Jocelin Su, and Anshul Kundaje. Computationally efficient measures of internal neuron importance. arXiv preprint arXiv:1807.09946, 2018.

[103] Amirata Ghorbani, James Wexler, James Y. Zou, and Been Kim. Towards automatic concept-based explanations. Advances in Neural Information Processing Systems, 32, 2019.

[104] Karel Kupka. International classification of diseases: ninth revision. WHO chronicle, 32(6):219–225, 1978.

[105] W. A. Knaus, E. Draper, D. Wagner, and J. Zimmerman. Apache ii: a severity of disease classification system. Critical care medicine, 13(10):818–829, 1985.

[106] D. Ledoux, J. Canivet, J. Preiser, J. Lefrancq, and P. Damas. Saps 3 admission score: an external validation in a general intensive care population. Intensive care medicine, 34(10):1873, 2008.

[107] Hugues Turbé, Mina Bjelogrlic, Christian Lovis, and Gianmarco Mengaldo. Evaluation of post-hoc interpretability methods in time-series classification. Nature Machine Intelligence, 5(3):250–260, 2023.

[108] Nicola K. Dinsdale, Emma Bluemke, Vaanathi Sundaresan, Mark Jenkinson, Stephen M. Smith, and Ana I.L. Namburete. Challenges for machine learning in clinical translation of big data imaging studies. Neuron, 110(23):3866–3881, 2022.

[109] Xue Ying. An overview of overfitting and its solutions. In Journal of physics: Conference series, volume 1168, page 022022. IOP Publishing, 2019.

[110] Yongqiang Dai, Lili Niu, Linjing Wei, and Jie Tang. Feature selection in high dimensional biomedical data based on bf-sfla. Frontiers in Neuroscience, 16:854685, 2022.

[111] Beatriz Remeseiro and Veronica Bolon-Canedo. A review of feature selection methods in medical applications. Computers in biology and medicine, 112:103375, 2019.

[112] G. Chandrashekar and F. Sahin. A survey on feature selection methods. Computers & Electrical Engineering, 40(1):16–28, 2014.

[113] J. R. Vergara and P. A. Estévez. A review of feature selection methods based on mutual information. Neural Computing and Applications, 24(1):175–186, 2014.

[114] T. M. Cover and J. A. Thomas. Elements of Information Theory. John Wiley & Sons, 2012.

[115] Shuyang Gao, Greg Ver Steeg, and Aram Galstyan. Efficient estimation of mutual information for strongly dependent variables. In Proceedings of Artificial Intelligence and Statistics, pages 277–286. PMLR, 2015.

[116] François Fleuret. Fast binary feature selection with conditional mutual information. Journal of Machine Learning Research, 5(9), 2004.

[117] V. Fonti and E. Belitser. Feature selection using lasso. VU Amsterdam Research Paper in Business Analytics, 30:1–25, 2017.

[118] Annette Spooner et al. Ensemble feature selection with data-driven thresholding for alzheimer's disease biomarker discovery. BMC bioinformatics, 24(1):9, 2023.

[119] Georgios Sermpinis, Serafeim Tsoukas, and Ping Zhang. Modelling market implied ratings using lasso variable selection techniques. Journal of Empirical Finance, 48:19–35, 2018.

[120] Christophe Chesneau and Mohamed Hebiri. Some theoretical results on the grouped variables LASSO. Mathematical Methods of Statistics, 17(4):317–326, 2008.

[121] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 68(1):49–67, 2006.

[122] Bradley Efron. The jackknife, the bootstrap and other resampling plans. SIAM, 1982.

[123] Bradley Efron and Robert J. Tibshirani. An introduction to the bootstrap. CRC press, 1994.

[124] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning: Data mining, inference, and prediction. Springer, 2009.

[125] Michael R. Chernick, Wenceslao González-Manteiga, Rosa M. Crujeiras, and Erniel B. Barrios. Bootstrap methods. John Wiley & Sons, 2011.

[126] A. C. Davison and D. V. Hinkley. Bootstrap Methods and Their Application. Cambridge University Press, 2009.

[127] L. Nelson Sanchez-Pinto et al. Comparison of variable selection methods for clinical predictive modeling. International Journal of Medical Informatics, 116:10–17, 2018.

[128] Lily Chamakura and Goutam Saha. An instance voting approach to feature selection. Information Sciences, 504:449–469, 2019.

[129] Vitor R. Carvalho and William W. Cohen. Single-pass online learning: Performance, voting schemes and online feature selection. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 548–553, 2006.

[130] Eman A. Atta, Ahmed F. Ali, and Ahmed A. Elshamy. A modified weighted chimp optimization algorithm for training feed-forward neural network. PloS one, 18(3):e0282514, 2023.

[131] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. arXiv preprint arXiv:1811.03378, 2018.

[132] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. nature, 521(7553):436–444, 2015.

[133] P. Diederik and J. Lei. Adam: A method for stochastic optimization. In Proceedings International Conference on Learning Representations, May 2015.

[134] B.D Ripley. Pattern Recognition and Neural Networks. Cambridge University Press, 2008.

[135] K. Hornik et al. Multilayer feedforward networks are universal approximators. Neural networks, 2(5):359–366, 1989.

[136] A. Graves. Supervised sequence labelling with recurrent neural networks. Springer, 2012.

[137] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, 1997.

[138] Kamilya Smagulova and Alex Pappachen James. A survey on lstm memristive neural network architectures and applications. The European Physical Journal Special Topics, 228(10):2313–2324, 2019.

[139] Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A review of recurrent neural networks: Lstm cells and network architectures. Neural computation, 31(7):1235–1270, 2019.

[140] Zachary C. Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with lstm recurrent neural networks. arXiv preprint arXiv:1511.03677, 2015.

[141] Lu Men, Noyan Ilk, Xinlin Tang, and Yuan Liu. Multi-disease prediction using lstm recurrent neural networks. Expert Systems with Applications, 177:114905, 2021.

[142] Dongdong Zhang, Changchang Yin, Jucheng Zeng, Xiaohui Yuan, and Ping Zhang. Combining structured and unstructured data for predictive models: a deep learning approach. BMC medical informatics and decision making, 20:1–11, 2020.

[143] Tian Guo, Tao Lin, and Nino Antulov-Fantulin. Exploring interpretable lstm neural networks over multi-variable data. In International conference on machine learning, pages 2494–2504. PMLR, 2019.

[144] Mike Schuster and Kuldip K. Paliwal. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11):2673–2681, 1997.

[145] Kyunghyun Cho et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In Conference on Empirical Methods in Natural Language Processing, 2014.

[146] K. Cho, B. Merrienboer, D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. In Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, 2014.

[147] Shudong Yang, Xueying Yu, and Ying Zhou. Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example. In 2020 International workshop on electronic communication and artificial intelligence (IWECAI), pages 98–101. IEEE, 2020.

[148] Muhammad Zulqarnain, Rozaida Ghazali, Muhammad Ghulam Ghouse, and Muhammad Faheem Mushtaq. Efficient processing of gru based on word embedding for text classification. JOIV: International Journal on Informatics Visualization, 3(4):377–383, 2019.

[149] José F. Díez-Pastor, Juan J. Rodríguez, César García-Osorio, and Ludmila I. Kuncheva. Random balance: ensembles of variable priors classifiers for imbalanced data. Knowledge-Based Systems, 85:96–111, 2015.

[150] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy. Training deep neural networks on imbalanced data sets. In 2016 International Joint Conference on Neural Networks, pages 4368–4374, 2016.

[151] Daniel Ramos, Javier Franco-Pedroso, Alicia Lozano-Diez, and Joaquin Gonzalez-Rodriguez. Deconstructing cross-entropy for probabilistic binary classifiers. Entropy, 20(3):208, 2018.

[152] Karl Øyvind Mikalsen, Cristina Soguero-Ruiz, Filippo Maria Bianchi, Arthur Revhaug, and Robert Jenssen. Time series cluster kernels to exploit informative missingness and incomplete label information. Pattern Recognition, 115:107896, 2021.

[153] Cristina Soguero-Ruiz et al. Data-driven temporal prediction of surgical site infection. In Proceedings of the AMIA Annual Symposium, pages 1164–1173, 2015.

[154] Zachary C. Lipton et al. Modeling missing data in clinical time series with RNNs. Machine Learning for Healthcare, 56, 2016.

[155] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. Scientific Reports, 8(1):1–12, 2018.

[156] Scott M. Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In Advances in Neural Information Processing Systems, pages 1–10, 2017.

[157] Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. Knowledge and information systems, 41:647–665, 2014.

[158] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pages 1135–1144, 2016.

[159] Anupam Datta, Shayak Sen, and Yair Zick. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In 2016 IEEE symposium on security and privacy (SP), pages 598–617. IEEE, 2016.

[160] Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888, 2018.

[161] Leo Breiman. Random forests. Machine learning, 45(1):5–32, 2001.

[162] Leo Breiman. Statistical modeling: The two cultures. Statistical science, 16(3):199–231, 2001.

[163] Nantian Huang, Guobo Lu, and Dianguo Xu. A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. Energies, 9(10):767, 2016.

[164] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res., 20(177):1–81, 2019.

[165] Jaime Gómez-Ramírez, Marina Ávila-Villanueva, and Miguel Ángel Fernández-Blázquez. Selecting the most important self-assessed features for predicting conversion to mild cognitive impairment with random forest and permutation-based methods. Scientific reports, 10(1):1–15, 2020.

[166] Yu Dong Zhang et al. Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation. Information Fusion, 64:149–187, 2020.

[167] Sören Richard Stahlschmidt, Benjamin Ulfenborg, and Jane Synnergren. Multimodal deep learning for biomedical data fusion: a review. Briefings in Bioinformatics, 23(2), 2022.

[168] Narendhar Gugulothu, Vishnu Tv, Pankaj Malhotra, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. Predicting remaining useful life using time series embeddings based on recurrent neural networks. arXiv preprint arXiv:1709.01073, 2017.

[169] Roger A. Horn. The Hadamard product. In Proceedings of Matrices: Theory and Applications, volume 40, pages 87–169, 1990.

[170] Bryan Lim, Sercan O. Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4):1748–1764, 2021.

[171] Ke Tan, Jitong Chen, and DeLiang Wang. Gated residual networks with dilated convolutions for supervised speech separation. In Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 21–25. IEEE, 2018.

[172] Ismael Sánchez. Adaptive combination of forecasts with application to wind energy. International Journal of Forecasting, 24(4):679–693, 2008.

[173] Chao Ren, Ning An, Jianzhou Wang, Lian Li, Bin Hu, and Duo Shang. Optimal parameters selection for bp neural network based on particle swarm optimization: A case study of wind speed forecasting. Knowledge-based systems, 56:226–239, 2014.

[174] Juliana Tolles and William J. Meurer. Logistic regression: relating patient characteristics to outcomes. Jama, 316(5):533–534, 2016.

[175] Deepak A. Kaji et al. An attention based deep learning model of clinical events in the intensive care unit. PloS one, 14(2), 2019.

[176] Philippe Rémy. Keras attention mechanism. GitHub repository, 2017.

[177] Ashish Vaswani et al. Attention is all you need. Advances in neural information processing systems, 30, 2017.

[178] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

[179] Ashish Sinha and Jose Dolz. Multi-scale self-guided attention for medical image segmentation. IEEE journal of biomedical and health informatics, 25(1):121–130, 2020.

[180] Yue Zhang and Jie Li. Application of heartbeat-attention mechanism for detection of myocardial infarction using 12-lead ECG records. Applied Sciences, 9(16):3328, 2019.

[181] Ilaria Gandin, Arjuna Scagnetto, Simona Romani, and Giulia Barbati. Interpretability of time-series deep learning models: A study in cardiovascular patients admitted to intensive care unit. Journal of Biomedical Informatics, 121:103876, 2021.

[182] Gianni Brauwers and Flavius Frasincar. A general survey on attention mechanisms in deep learning. IEEE Transactions on Knowledge and Data Engineering, 2021.

[183] Derya Soydaner. Attention mechanism in neural networks: where it comes and where it goes. Neural Computing and Applications, 34(16):13371–13385, 2022.

[184] Ruth C. Fong and Andrea Vedaldi. Interpretable explanations of black boxes by meaningful perturbation. In Proceedings of the IEEE international conference on computer vision, pages 3429–3437, 2017.

[185] Ruth Fong, Mandela Patrick, and Andrea Vedaldi. Understanding deep networks via extremal perturbations and smooth masks. In Proceedings of the IEEE/CVF international conference on computer vision, pages 2950–2958, 2019.

[186] Maksims Ivanovs, Roberts Kadikis, and Kaspars Ozols. Perturbation-based methods for explaining deep neural networks: A survey. Pattern Recognition Letters, 150:228–234, 2021.

[187] Nate Silver. The signal and the noise: Why so many predictions fail-but some don't. Penguin, 2012.

[188] Alan R. Hinman, James M. Hughes, Dixie E. Snider Jr., and Mitchell L. Cohen. Meeting the challenge of multidrug-resistant tuberculosis: summary of a conference. Morbidity and Mortality Weekly Report: Recommendations and Reports, pages 49–57, 1992.

[189] Melisa L. Thombley and Daniel D. Stier. Menu of suggested provisions for state tuberculosis prevention and control laws. US Department of Health and Human Services. Centers for Disease Control and Prevention, Atlanta, 2010.

[190] Parminder Raina et al. A longitudinal analysis of the impact of the covid-19 pandemic on the mental health of middle-aged and older adults from the canadian longitudinal study on aging. Nature Aging, 1(12):1137–1147, 2021.

[191] Matthias Keicher et al. Multimodal graph attention network for covid-19 outcome prediction. Scientific Reports, 13(1):19539, 2023.

[192] Haohui Lu and Shahadat Uddin. A weighted patient network-based framework for predicting chronic diseases using graph neural networks. Scientific reports, 11(1):22607, 2021.

[193] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In Noise Reduction in Speech Process. Springer, 2009.

[194] Sachin Malik and Rajesh Singh. A family of estimators of population mean using information on point bi-serial and phi correlation coefficient. International J. of Science Education, 10(1):75–89, 2013.

[195] Junfeng Wu, Li Yao, and Bin Liu. An overview on feature-based classification algorithms for multivariate time series. In Proceedings of the IEEE Intl. Conf. on Cloud Computing and Big Data Analysis, pages 32–38, 2018.

[196] Cristina Soguero-Ruiz et al. Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. IEEE J. Biomed. and Health Informat., 61(3):87–96, 2016.

[197] Matthias Dehmer and Abbe Mowshowitz. A history of graph entropy measures. Informat. Sciences, 181(1):57–78, 2011.

[198] Antonio G. Marques et al. Modelling cardiovascular condition evolution in hypertensive population using graph signal processing. In Proceedings of the Computing in Cardiology, pages 1–4, 2017.

[199] Sergio Martinez-Aguero et al. Data and network analytics for COVID-19 ICU patients: A case study for a Spanish hospital (online appendix). https://tsc.urjc.es/~amarques/papers/Covid19ICUJBHI21.pdf. (Apr. 22, 2021). Access: Apr. 25, 2021.

[200] WHO. WHO discontinues hydroxychloroquine and lopinavir/ritonavir treatment arms for COVID-19. https://www.who.int/news. (Jul. 4, 2020). Access: Apr. 25, 2021.

[201] Óscar Escudero-Arnanz, Antonio G. Marques, Cristina Soguero-Ruiz, Inmaculada Mora-Jiménez, and Gregorio Robles. dtwParallel: A Python package to efficiently compute dynamic time warping between time series. SoftwareX, 22:101364, 2023.