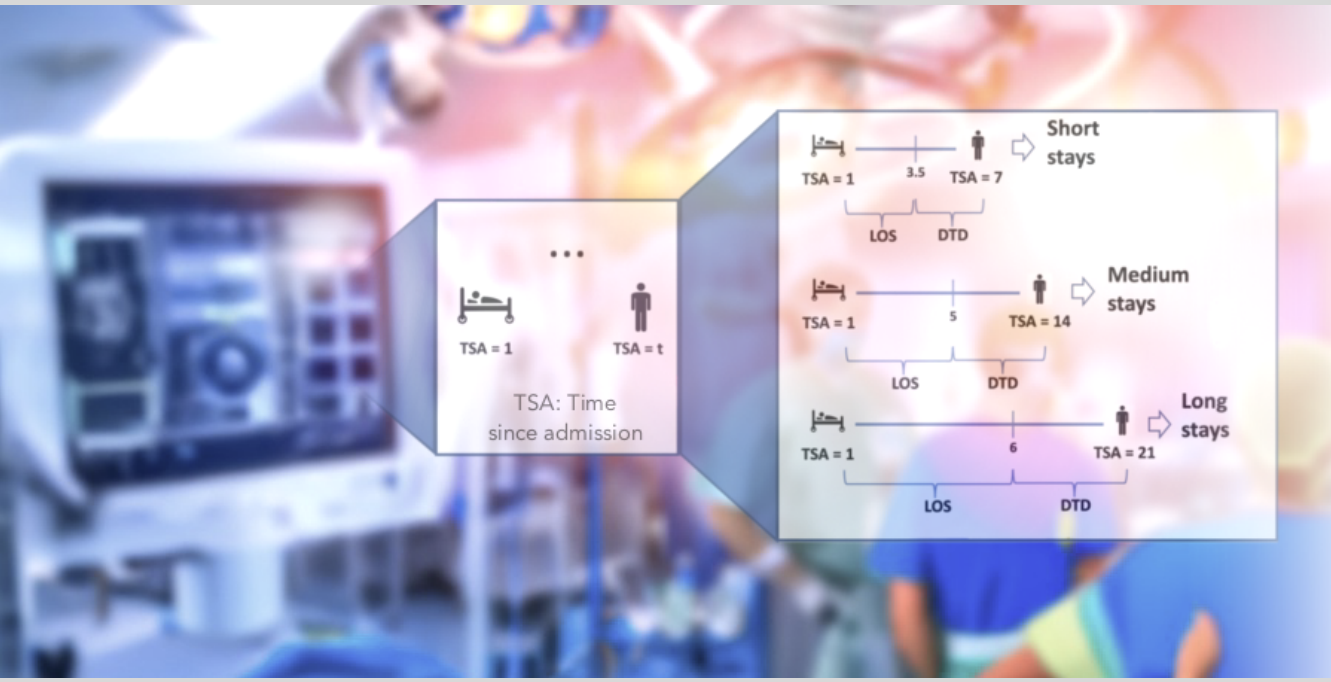
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS

David Cuadrado Gómez

UNIVERSITAT ROVIRA i VIRGILI

Machine learning methods for predicting days to discharge in Intensive Care Units patients

DAVID CUADRADO GÓMEZ



DOCTORAL THESIS
2023

DOCTORAL THESIS

Machine learning methods for predicting days to discharge in Intensive Care Units patients

Author:
David Cuadrado Gómez

Supervisors:
Dr. David Riaño
Dr. Aïda Valls

UNIVERSITAT
ROVIRA i VIRGILI

Departament d'Enginyeria Informàtica i Matemàtiques

Escola Tècnica Superior d'Enginyeria

UNIVERSITAT ROVIRA I VIRGILI

Tarragona, Spain

2023

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

UNIVERSITAT
ROVIRA i VIRGILI

I STATE that the present study, entitled "Machine learning methods for predicting days to discharge in Intensive Care Units" patients, presented by David Cuadrado Gómez for the award of the degree of Doctor, has been carried out under the main supervision of Dr. David Riaño Ramos, in memoriam, and continued under my supervision at the Department of Computer Science and Mathematics of Universitat Rovira i Virgili.

Tarragona, July 30th 2023.

Doctoral Thesis Supervisor

Dr. Aida Valls Mateu

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

Acknowledgements

Esta tesis no sería posible sin la ayuda de todos aquellos con los que he compartido vivencias durante el desarrollo de este doctorado.

Muchas gracias a mis padres, no solo por sus ánimos y soporte continuo, sino también por enseñarme a seguir aprendiendo y a no rendirme nunca.

Gracias a mi hermano por sus ánimos desde la distancia y por permitirme parar de vez en cuando para ganarle al Fifa.

A Isabel, por ayudarme en el día a día y cuidar de lo que más me importa, muchas gracias.

Gracias a mi familia por su apoyo en todos estos años. A mis abuelos, mis tíos y mis primos, con los que hemos compartido momentos muy felices y otros no tanto, pero siempre juntos.

Gracias a mis amigos por las risas y por recibirme siempre con los brazos abiertos a pesar de esos largos periodos de tiempo desaparecido mientras trabajaba en esta tesis.

A mis compañeros de trabajo, con los que he compartido muchas sesiones de café y de historias sobre artículos y tesis.

A Aïda, por su enorme soporte en este último año. Esta tesis no existiría si no fuera por su ayuda y enorme trabajo.

Dejo para el final a las dos personas más importantes de esta tesis:

A David, la persona que me animó a ampliar horizontes, que me ayudó a salir de mi zona de confort y con la que mi vida dio un giro de 180ª. Parte de lo que soy hoy en día es gracias a ti y te lo estaré agradecido siempre. Te echo de menos.

Y por supuesto a Bàrbara, la persona que ha estado siempre a mi lado, con quien he disfrutado en los mejores momentos y que me ha animado en los peores. La persona que un día me animó a estudiar aquello que quería y que ha acabado en la creación de esta tesis. Gracias por estar siempre a mi lado.

Muchas gracias a todos.

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

Contents

List of Figures

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

List of Tables

Abstract

The Intensive Care Units (ICUs) are specialized hospital services that provide intensive and intricate care to critically ill patients. The patients admitted in the ICU require constant monitoring and care from a dedicated team of healthcare professionals, primarily comprising physicians and nurses. The patients are treated in the ICU until their medical condition and parameters stabilize to a level where they can be safely moved to other hospital services or discharged.

Multiple factors are involved in ICU planning, including bed management, staff availability, medical equipment and organization of support services, among others. Each of these factors are crucial for the care of patients in the ICU and the lack of adequate planning can lead to a decrease in the quality of care, delays in care and increased morbidity and mortality of patients.

This thesis proposes a hybrid model for the prediction of the date of discharge of ICU patients, based on a combination of Machine Learning techniques. The model can be used to support the planning activities of the intensive care personnel, as it identifies which patients have a higher probability of early hospital discharge and which ones will still require the use of ICU services for more time. The main novelty of this prediction model is the ability to make a daily adjustment of the number of days to discharge, instead of the current models that are only applicable at admission time. This updated information is crucial to improve the efficiency of the ICU services as well as the quality of patient's care.

The first study in this thesis was done with a private dataset of critical patients from Joan XXIII Hospital in Tarragona. Later, we extended the analysis to an open public dataset called eICU, which includes patients from multiple centers at USA. Therefore, it has been possible to train and test the model with a very heterogeneous groups of patients from different populations, obtaining much reliable conclusions.

The first contribution addresses the difficulties in prediction of the days to discharge (DTD) for patients in intensive care due to their high heterogeneity. Both healthcare practitioners and computers struggle with making accurate predictions when dealing with such diverse medical conditions. To improve DTD prediction accuracy, it is crucial to have tools for analyzing patient heterogeneity. This thesis proposes four measures to quantify patient heterogeneity. These measures were tested on patients admitted in Hospital Joan XXIII over four years. The results provide a deeper understanding of ICU patients and serve as a foundation for improving DTD predictors.

The second contribution presents two approaches to the analysis of the DTDs of ICU patients from different perspectives: biomarker identification and phenotype recognition. Several machine learning methods are constructed for each approach

and tested with the data patients admitted at ICU of Hospital Joan XXIII. The results in of this study confirm the complexity of calculating the days to discharge with intelligent data analysis methods.

The last contribution addresses the challenge of predicting the date of discharge of ICU patients accurately. Currently, a single prediction of the duration of stay is made at admission time, known as Length of Stay (LOS) measure. However, depending on the patient's evolution at ICU, the final number of days deviates from the initial LOS value. This thesis proposes the use of Machine Learning methods to calculate a new measure called DTD (Days to Discharge), which takes into account the patient's state at each of the days of stay. Given the previous findings about heterogeneity, three groups were identified for studying the DTD. Three algorithms (Random forest, XGBoost and lightGBM) were used to generate DTD and LOS predictive models on the large and diverse and public eICU database. A comparative study shows that combining these models into a hybrid model can improve the accuracy of the predictions. The LOS model is effective at the beginning of the stay, while after some initial days, the DTD model becomes much more effective till the end of the stay. The results achieved a root mean square error (RMSE) and mean average error (MAE) below one day on the eICU dataset.

Chapter 1

Introduction

1.1 Motivation

According to the World Federation of Societies of Intensive and Critical Care Medicine, intensive care units (ICU) are "organized systems for the provision of care to critically ill patients that provides intensive and specialized medical and nursing care, an enhanced capacity for monitoring and multiple modalities of physiologic organ support to sustain life during a period of acute organ system insufficiency" (Marshall et al., 2017). Due to a serious illness, injury or medical condition that threatens his life and/or vital organs, patients attended in ICU require continuous monitoring and the need for a multidisciplinary team of medical professionals, including nurses, doctors and other specialists.

Patients cases in ICUs use to be heterogeneous and it may include diverse type of patient inside an ICU. For example, we can find critical patients, who are often intubated and mechanically ventilated, trauma patients who have experienced a severe injury, such as a head injury or multiple fractures, post-operative patients that have undergone major surgery or neurological patients with severe neurological conditions, such as a stroke or brain injury and many other types.

Due to their critical conditions, these patients require constant medical attention and monitoring to ensure that their vital signs remain stable. To achieve this, healthcare professionals connect them to multiple monitoring devices, such as ECG machines, arterial and central venous catheters and pulse oximeters among others. These devices allow the healthcare team to keep a close eye on the patient's heart function, blood pressure, oxygen saturation and other critical vital signs. By continuously monitoring the patient's condition, healthcare professionals can adjust their treatment to maintain stable respiratory and hemodynamic status. This helps to ensure that the patient receives the best possible care and has the best chance of a successful recovery.

Patients in the ICU also require constant medication to help manage their medical conditions and maintain stable hemodynamic and respiratory status. These medications can range from painkillers and sedatives to antibiotics and blood thinners, as well as vasopressors or neuromuscular blocking agents, which help maintain blood pressure or muscle relaxation respectively. The goal of medication management in the ICU is to provide the necessary treatment while minimizing any potential side effects and optimizing the patient's chances of recovery.

It was estimated that in the US the mean ICU cost and length of stay were $31,574 ± 42,570 and 14.4 days ± 15.8 ($2,193 per day in average) for patients requiring mechanical ventilation and $12,931 ± 20,569 and 8.5 days ± 10.5 ($1,521 per day in average) for those not requiring mechanical ventilation (J. Dasta, 2005). A multi-country study in Europe showed that ICU direct costs ranged, in average, from €1168 to €2025 per patient and day (Tan, 2012). The European Hospital and Health Care Federation parencite that, in 2014, the average length of stay in acute care hospitals was 6.4 bed days in the EU-28 countries, ranging from 5.2 to 7.0 days. These important daily costs urge governments and ICU managers in a effort to reduce the length of stay of ICU patients without compromising healthcare quality.

In order to have control of medical personnel, necessary medication and availability of beds for future admissions, a proper planning is required. The availability of nurses and doctors as well as the number required can be modified depending on the workload of each patient. In the case of medication, a control of all the necessary medication is required during the day as well as for the following days depending on the patient's stay. The number of beds is also important as it requires long-term planning to know availability and possible relocations in case of shortage. For all this, the forecast of the number of days of stay of each patient plays an essential role.

In recent years, methods based on Artificial Intelligence (AI) have emerged as powerful tools for analyzing patient data and making predictions about their health outcomes. Machine Learning algorithms can identify patterns in large datasets and use them to make accurate predictions about future events. At ICUs, the constant monitoring of the critical patient plays a very important role in the discharge prediction, since it allows to know the evolution of the patient throughout his stay. Nowadays, the clinical information collected at the ICUs is saved daily for all patients, therefore, we can get a database large enough to make good predictions. These data can be categorized as demographic, constant and variable data:

- Demographic data includes information such as the patient's name, gender, ethnicity and medical history. This information can help healthcare professionals to better understand the patient's background and risk factors for certain conditions. Demographic data is typically collected at the time of admission and is used throughout the patient's stay in the ICU.
- Constant data refers to the patient's physiological characteristics that remain relatively stable over time, such as age and height.
- Variable data refers to physical characteristics that fluctuate over time, such as heart rate, blood pressure and oxygen saturation. The collection and analysis of variable data can provide valuable insights into the patient's condition and help healthcare professionals to make timely and informed decisions about their care.

The hypothesis of this doctoral thesis is that with previous processing of the available collected data, it is possible to create prediction models to predict each patient's length of stay at an ICU. Concretely, with the application of Machine Learning (ML) techniques, we aim to create prediction models for the duration of a patient's stay in the ICU.

ICUs receives patients who are severely ill and may suffer from different complications during their recovery process. As a result, it becomes a hard task for prediction models to precisely approximate the length of stay. When an ICU patient is admitted, doctors usually try to predict an approximation of how long this patient will stay. This approximation is called Length of Stay or LOS. The quality of the number of days prediction can be increased if we approach the problem from another point of view. The prediction of the LOS value is based on the data collected during the first 24-48 hours since the admission. But this method does not take into account how the patient's condition may change over time. In this thesis, we propose a new approach in which we're tracking the patient's evolution each day and recalculate our prediction accordingly thanks to all the variable data collected daily. By looking at how the patient is evolving each day, we can create a more accurately prediction model and obtain better results in order to predict the number of days of discharge or DTD. This approach gives us a more exact and personalized way to predict how long a patient will need to stay in the ICU, helping healthcare professionals plan their resources and care for the patient more effectively.

Obtaining a high accuracy in the days to discharge prediction with machine learning is a challenging task, with only a limited number of papers available on this topic. There are only a few studies that have focused on this variable, making it difficult to assess its effectiveness and reliability. One of the main reasons for the lack of research on the days to discharge variable is the complexity of predicting the discharge date of critically ill patients. Furthermore, the data required for training machine learning models to predict the days to discharge variable is not always readily available. Collecting and organizing data from electronic health records and other sources can be time-consuming and resource-intensive, making it challenging to conduct large-scale studies on this variable.

With proper data preprocessing, we can obtain a robust database for making predictions of DTD and LOS. Both parameters have clinical, management and administrative impact since they can help clinicians to detect deviation from standards and, therefore, implement corrective actions. DTD prediction brings some additional benefits in the organization, facilitating decision making for the efficient use and planning of resources in ICUs, such as bed occupancy (Ruyssinck et al., 2016), complex surgeries requiring ICU (Bing-Hua, 2014), or transfers to other centers (Droogh et al., 2015), if a saturation is expected.

In this thesis, Machine learning techniques have been used in the creation of predictive models for LOS and DTD, with methods such as regression models, Random Forest, Support Vector Machine and Artificial Neural Network being employed.

To evaluate the effectiveness of these predictive models, two different indicators are commonly used: coefficient of determination (R2) and root mean square error (RMSE). They are used to compare new results with previous studies to ensure that models are performing better. A literature search shows that the best R2 obtained by previous studies is 0.81 and the smallest RMSE is 1.79. After analyzing the results with medical personnel, we discovered that a forecast error of 1 or 2 days is not acceptable if the purpose is to improve ICU planning. Patients in the ICU can experience rapid fluctuations in their condition and immediate action is often necessary. Therefore, we decided to work on reducing such large forecast errors.

At this point, we defined the main challenge of the thesis: to obtain a Root Mean Squared Error (RMSE) below 1 day. Achieving an RMSE below 1 day would mean that the forecasts were highly accurate, providing valuable information for ICU planning and resource allocation.

This challenge highlights the importance of precision in ICU planning and management. Even small errors in forecasting can have significant consequences for patients and healthcare providers. Therefore, it is crucial to develop accurate and reliable models that can provide timely and relevant information for ICU planning. This goal can be achieved through rigorous research and collaboration between medical professionals and data analysts, as was done in this thesis.

Thanks to the ICU patient database provided by Hospital Joan XXIII in Tarragona, we have trained and tested our Machine Learning models for predict the number of days of stay and improved the current results. We have also used the eICU database which has enabled us to analyze data from a large number of patients with different conditions, as well as varying locations. Patients in the Joan XXIII Hospital database are Spanish, while those in the eICU databases come from the United States. By leveraging these databases, we can work with a diverse range of patients and different types of health systems, which enhances the accuracy of our predictive models.

1.2 Objectives

Predicting the number of days that patients will spend in the ICU is important for managing patient care, beds and medication. While many papers have explored the length of stay, focusing on the first 24-48 hours only, there is potential for improvement by considering each day of the patient's stay and their evolving condition. This could lead to increased accuracy in predicting how long patients will need to stay in the ICU.

The objective of this thesis is to achieve an accurate prediction of the day of discharge of an ICU patient. To attain this main goal, the following sub-objectives must be accomplished:

- This study has been initially conducted using data provided by Hospital Joan XXIII in Tarragona. Our first task is to construct a comprehensive and reliable database that will help us enhance the current results of LOS prediction and expand our investigation by including the DTD prediction. This database will contain every parameter from all patients admitted in the ICU and obtained by the constant monitoring during their stay. Once the database is completed, we must clean it by identifying any unusual or abnormal variables, fill in any missing data and remove the outliers that might affect the accuracy of the results. It is also essential to use feature importance techniques to identify the most significant parameters and exclude the less important ones from the dataset.
- Using the database collected, structured and prepared, we will examine the intricacy of ICU patients characteristics of this local hospital. New approaches will be suggested to measure the diversity among critically ill patients, with the focus on identifying groups that help to determine the length of their hospital stay.

- Finally, we will expand the analysis with the use of a larger database: eICU. By having a larger number of patients, greater heterogeneity and a completely different type of patient (mainly as they are from another country and from another health system), we will be able to study how to build Machine Learning models to achieve our ultimate goal of obtaining a good prediction of the number of days remaining till the date of discharge, during the stay of a given patient in the ICU.

1.3 Contributions

By working on the different objectives previously described, this doctoral thesis contributes mainly in the following three topics, for which we have produced the following three research publications:

1. Development of four methods and their corresponding measures to quantify the heterogeneity of the patient in the Intensive Care Unit. Published at:

   Cuadrado D, Riaño D, Gómez J, Rodríguez A, Bodí M. (2021). Methods and measures to quantify ICU patient heterogeneity. Journal of Biomedical Informatics. May;117:103768. doi: 10.1016/j.jbi.2021.103768. Epub 2021 Apr 9. PMID: 33839305.

   ISI-JCR journal, 1st decile. (2021)

2. Proposal of four methods to analyze DTD from different perspectives: quantification of heterogeneity based on the previous article, identification of biomarkers, phenotype recognition and an initial basic prediction of DTD based on a small set of ICU patients. Published at:

   Cuadrado, D., Riaño, D. (2021). ICU Days-to-Discharge Analysis with Machine Learning Technology. In: Tucker, A., Henriques Abreu, P., Cardoso, J., Pereira Rodrigues, P., Riaño, D. (eds) Artificial Intelligence in Medicine. AIME 2021. Published at: Lecture Notes in Computer Science, vol 12721. Springer, Cham. https://doi.org/10.1007/978-3-030-77211-6-11

   CORE A conference. (2021)

3. Proposal of a hybrid model that integrates the predictive models carried out with Random Forest, LightGBM and XGBoost. This hybrid model improves the predictions due to the effectiveness of the prediction of LOS at the beginning of a patient's stay and of DTD at the end of your stay. Submitted at the following journal:

   Mathematics. Special Issue "Advances of Applied Probability and Statistics", MDPI: Cuadrado, D., Riaño, D., Valls, A. (2023). Predicting ICU patients discharge with a hybrid model that combines Length of Stay and Days to Discharge.

   ISI-JCR journal, 1st quartile. (2022)

1.4 Document organization

After this introduction that presented the motivation of this work and the goals, the rest of the PhD thesis document is structured as follows.

Chapter 2 presents some background knowledge in the topic of this thesis. First, we revise the previous related works on prediction of length of stay.

The databases used for this study are detailed, with a characterization of the patients parameters available, along with the data pre-treatment. The creation of the database is a critical component of the study as it is the foundation upon which the research is built. In the second chapter, the thesis will provide a more comprehensive explanation of how the database was constructed, including the variables that were included and the methods used for data collection.

Next, we refer to a previous study that we conducted in 2018. This preliminary study focused on predicting the discharge of patients from an ICU, was the first approach to the main objective of this thesis. The chapter presents the initial results obtained from the study, which have formed the basis for the development of the current thesis. By examining the previous study in detail, the thesis aims to build upon its findings and contribute further insights into the prediction of patient discharge from ICU.

The Chapter 3 of the thesis focuses on the first contribution. This article deals with a common problem in the field of intensive care: the great heterogeneity of patients. In other words, patients in the ICU vary significantly in terms of their medical history, underlying conditions and severity of illness, among other factors. This heterogeneity makes it difficult to predict outcomes and plan treatment strategies.

The chapter proposes a set of methodologies to address this problem, with the aim of identifying subsets of patients that are more homogeneous and therefore easier to study and predict outcomes for. By creating more homogeneous groups of patients, it becomes possible to develop more accurate predictive models and to tailor treatment strategies to individual patient needs.

The specific methodologies proposed in the chapter are likely to be discussed in detail in the chapter. Machine learning methods were used to cluster patients based on certain variables, such as age, gender, commodities, or vital signs.

Chapter 4 focuses on the second scientific contribution of this thesis, which focuses on developing four different approaches to analyze the prediction of the days to discharge and improve the accuracy of their prediction.

Two different kinds of data are formalized and studied in order to detect possible relations with the duration of the stay at an ICU. First, Biomarker Identification, which involves identifying specific biomarkers (indicators) that are associated with the disease being studied.

Second, Phenotype Recognition, which is based on identifying patterns or clusters within the DTD data that are indicative of particular phenotypes (characteristics) of the disease.

In Chapter 5, the knowledge obtained from the two previous chapters is used to develop a final proposal. This proposal takes the form of a hybrid model, which combines two different predictions: the prediction of LOS (length of stay) and DTD (time to discharge). These predictions are particularly effective depending on when they are made during the patient's stay, i.e. at the beginning or the end.

The development of this hybrid model is an important contribution to the field of study, as it represents an innovative and practical application of the knowledge gained from the previous research articles. The chapter goes into detail about the specific methods used to develop the hybrid model, as well as the results of testing and validation to demonstrate its effectiveness.

Finally, chapter 6 summarizes the conclusions of this work and propose several interesting lines of research for future work.

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

Chapter 2

Medical background and data preparation

2.1 Related Work

The daily management of an Intensive Care Unit (ICU) is a complex task for its daily planning due to the diverse range of critically ill patients that require constant medical attention. ICU patients may have a large variety of pathologies with an affectation of one or more threatened vital functions, which are potentially reversible. The Working Group on Quality Improvement of the European Society of Intensive Care Medicine classified these patients (Valentin, Ferdinande, and Quality Improvement, 2011) into two groups: those ones "requiring monitoring and treatment because one or more vital functions are threatened by an acute (or an acute on chronic) disease [...] or by the sequel of surgical or other intensive treatment [...] leading to life-threatening conditions" and those with "already having failure of one of the vital functions [...] but with a reasonable chance of a meaningful functional recovery". Patients in their end-stages of untreatable terminal diseases were left out of these groups.

More specific classifications distinguish between ICU patients requiring close monitoring, patients facing critical lung issues, patients with severe cardiac problems, and patients with serious infections. All this variability in admission extends throughout the patient's stay in the ICU, and is reflected in the great disparity in patient's evolution, treatments (Kross et al., 2014), outcomes, and costs (Rossi et al., 2006; Jacobs et al., 2022). Moreover, ICU resources (i.e. beds) are usually quite limited and when there is an unexpected increase of demand (i.e. in COVID-19 pandemics, in earthquakes or other catastrophes) a good knowledge and planning of the patient's occupation (i.e. days to discharge) becomes crucial for a good health service.

Constant monitoring is crucial not only for planning purposes, but also to ensure that patients receive appropriate medication and nursing care and that their vital signs remain stable until they are discharged to another ward.

In this context, it is desirable to have reliable computer tools to predict the duration of patients in ICU, in order to accurately estimate the availability of beds, medications, and health personnel required to care for each patient. In the literature, there are two approaches to such predictive tools: static and dynamic.

- Static tools predict the length of stay (LOS). The LOS is a measure of the number of days a patient stays in the ICU from the moment they are admitted until

the day of discharge. This indicator is considered static because its value remains constant once it is calculated and it is obtained from data collected during the first 24-48 hours of the patient's admission. Some studies found that physicians are not good at predicting ICU-LOS statically and they are poor at predicting stays longer than five days (Nassar and Caruso, 2016; Gusmão Vicente, 2004). Moreover, a systematic review analyzed 31 ICU-LOS predictive models and concluded that they suffer from serious limitations (Verburg et al., 2017). Statistical and machine learning approaches (e.g., (Kramer, 2010; Livieris, 2018)) provide moderate predictions of short term LOS (1-5 days), but are unable to correctly predict long-term LOS ($>$ 5 days). Therefore, while LOS is essential for resource allocation during a patient's initial admission, it becomes less informative as their stay progresses and their medical needs change.

- Dynamic tools make predictions in a daily basis. These tools predict the days to discharge (DTD), a dynamic indicator that changes over time as the patient is treated in the ICU. DTD represents the number of days remaining until the patient is considered ready for transfer to another hospital or unit. This indicator differs from length of stay (LOS) in that it takes into account the patient's evolving medical condition and progress towards recovery. DTD is an important tool for ICU staff because it provides daily information on the expected time of transfer for each patient. This information can be used to plan for bed availability, medication needs and ensuring that resources are allocated appropriately and that patients receive the highest quality of care. By monitoring changes in DTD over time, ICU staff can gain insights into the patient's medical progress and identify any potential delays or complications that may require additional interventions. This information can highly improve the prediction models as it provides more information than LOS. DTD is closely related to the concept of length of stay (LOS) but, unlike this, DTD is not a constant parameter and it is not predicted on the patient condition in the first 24-48 hours after admission, but on the evolving condition of the patient.

There is a long tradition of LOS prediction with machine learning methods (Awad, Bader–El–Den, and McNicholas, 2017; Verburg et al., 2017; Peres et al., 2021). However, some studies have focused on patients suffering from a certain specific disease. For example, some works a related to ICU patients with cardiac surgery (Rowan et al., 2007; Hachesu et al., 2013; LaFaro et al., 2015; Mollaei et al., 2021), traumatic patients (Van Houdenhoven et al., 2007; Gholipour et al., 2015), cancer diseases (Muhlestein et al., 2019), or patients with post-surgical problems (Su et al., 2021). Other authors, in order to overcome the low predictive capacity in short stays, some works concentrate only on stays longer than five days, such as (Gusmão Vicente, 2004; Nassar and Caruso, 2016).

The problem of estimating the LOS for any patient admitted at an ICU, has been also approached using different kinds of statistical and machine learning methods. Table 2.1 shows the number of patients, number of parameters and the prediction methods used in the related works for LOS prediction.

The oldest works used some statistical techniques and different kinds of regression models (Moran and Solomon, 2012; Verburg et al., 2014; Li et al., 2019; Huang et al., 2013). Some other approaches use traditional machine learning methods, such

as Random Forest, Support Vector Machine or Neural Network (Caetano, Laureano, and Cortez, 2014; Abd-ElrazekaAhmed et al., 2021; Verburg et al., 2014; Gholipour et al., 2015; Chrusciel et al., 2021; Alghatani et al., 2021; Houthooft et al., 2015). Some more recent works, have also included advanced neural networks and deep learning (Ayyoubzadeh, 2020; Ma et al., 2020; Wu et al., 2021).

The prediction of DTD has awaken much less interest, possibly due to the added difficulties of collecting ICU data in a daily basis (Artis et al., 2019) or to the sophistication of the technologies for longitudinal data analysis (Caruana et al., 2015). Table 2.2 shows the same information from the previous mentioned articles for DTD prediction for the two existing works (to the best of our knowledge), (Temple, Lehmann, and Fabbri, 2015; Ruyssinck et al., 2016) These two works make a quite reduced study, with a small number of parameters, and focusing on some concrete values of DTD.

From this study of the related work, we can see that although there are numerous articles focused on the LOS in the ICU, there is still a significant unexplored territory in the field of DTD.

In order to build a good prediction model with Machine Learning methods, it is mandatory to have a large and representative dataset. In the next section, the two datasets used in this work are described and the data preparation is detailed.

2.2 Private database

2.2.1 Introduction

The first database used in this doctoral thesis was provided by the main reference public hospital in Tarragona (Catalonia, Spain). Thanks to the remarkable system for patient care through the efforts of its nurses in the Intensive Care Unit (ICU) and their commitment to monitoring patients and meticulously recording daily data, Hospital Joan XXIII has successfully created a comprehensive database that stores health information in a structured way.

The continuous monitoring performed by the nursing staff plays a pivotal role in ensuring the accuracy and reliability of the data within the database. By closely observing patients, noting vital signs, administering treatments, and documenting every pertinent detail, the nurses contribute significantly to the collection of reliable knowledge available for analysis and research.

The data analyzed in this study correspond to patients who were admitted to the Intensive Care Unit (ICU) between January 2016 and November 2019. In a preliminary research conducted before the development of this thesis, all patients were included in the dataset, regardless of their outcome upon leaving the ICU. It was in this initial phase of our research that we considered the possibility of patients leaving the ICU either alive or deceased. However, we later decided to exclude this possibility from our analysis due to the negative impact it could have on the patient's discharge prediction. We understood that a patient's death during their stay in the ICU would be considered as an interruption in terms of providing adequate medical treatment and healing, and we wanted to avoid such an interpretation.

Reference	Patients	Parameters	Methods	Results
(Moran and Solomon, 2012)	111,663	-	OLS, LMM, Treatment effects model, Skew-normal model, Skew-t model, GLMs, Extended estimating equations, Finite mixture model.	R2=0.17-0.22 and MAE=2.2-2.4
(Huang et al., 2013)	284	6	Linear regression analysis and ANN	RMSE = 5.20
(Caetano, Laureano, and Cortez, 2014)	15,253	14	Random Forest	R2=0.81%
(Verburg et al., 2014)	32,667	67	Ordinary Least Squares, Generalized Linear Model, Poisson regression, Negative binomial regression, Gamma regression with a logarithmic link function, Cox Proportional Hazards regression.	R2=0.09-0.20, RMSPE=7.28-8.74, MAPE=3.00-4.42
(Houthooft et al., 2015)	14,480	6	Support Vector Regression	MAE=1.79
(Nassar and Caruso, 2016)	2,955	19	-	Kappa=0.22
(Ayyoubzadeh, 2020)	526	16	Naïve Bayes, Generalized Linear Model, Logistic Regression, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees, Support Vector Machine and AdaBoost	Accuracy = 83.91%, Sensitivity=80%
(Abd-ElrazekaAhmed et al., 2021)	231	31	Neural Network, Classification, Regression Trees, Tree Bagger, Random Forest, Support Vector Machine, K-nearest Neighbor, Naive Baye, Fuzzy Rules	Accuracy(Fuzzy)= 92%
(Ma et al., 2020)	4,000	15	Extreme Learning	Accuracy=0.82, Specificity=1, Sensitivity=0.6150
(Alghatani et al., 2021)	44,626	12	Linear regression, Linear discriminant analysis, Random Forest, K-Nearest Neighbors, Support Vector Machine and Extreme Gradient Boosting	Best accuracy 89%
(Wu et al., 2021)	117,306 (eicu) and 42,932 (MIMIC-III)	26	Random Forest, Support Vector Machine, Deep Learning, and Gradient Boosting Decision Tree	GBDT: Brier score=0.164 AUROC(internal vald.)=0.742 AUROC(external vald.)=0.747
(Chrusciel et al., 2021)	5,006	9	Random Forest	Accuracy: Unstructured=76.3% Structured=75.0

TABLE 2.1: Description of the number of patients, parameters, methods and prediction models used in articles related to LOS prediction

Reference	Patients	Parameters	Methods	Results
(Ruyssinck et al., 2016)	4,693	26	Random Forest	For DTD =2 AUC=0.749 and for DTD =10 AUC=0.614
(Temple, Lehmann, and Fabbri, 2015)	4,023	19	Random Forest and baseline	Best: MAE = 1.4

TABLE 2.2: Description of the number of patients, parameters, methods and prediction models used in articles related to DTD prediction

The decision to exclude deceased patients from the analysis was not solely based on the concern that their inclusion might disrupt the discharge process and potentially affect the effectiveness of the predictive models. It was also driven by the belief that cases involving patient mortality should be treated separately, using specialized methods for mortality prediction.

By focusing on patients who survived their ICU stay, the study aims to develop predictive models that are specifically tailored to forecast outcomes for this particular group. This allows for a more targeted approach to improving patient care strategies and delivering accurate predictions within the context of ICU survival.

2.2.2 Data pretreatment

To establish a robust and dependable foundation, the data underwent a meticulous treatment process. The initial stage involved scrutinizing all the available parameters and eliminating any outliers present. Outliers refer to values that greatly deviate from the average and can be considered anomalous. By removing these outliers, the risk of distorting the results of subsequent analyzes was mitigated. This step ensured that the data used for further analysis remained reliable and free from any potentially misleading values.

In addition, consultations were held with medical professionals to identify any additional anomalous data that may have been overlooked during the outlier removal process. Recognizing the importance of these data points for the predictive models, it was crucial to ensure their proper inclusion in the data analysis. The excluded anomalous data can be characterized by the following traits:

- Given that a portion of the database consists of manually entered data by medical staff, it is important to address human errors that may inadvertently impact the predictive model. These errors primarily manifest as decimal omissions (e.g., recording a patient's body temperature as 374 ºC instead of 37.4 ºC), accidental addition of values (e.g., a saturation level of 199 % instead of a plausible range), or the inadvertent substitution of one value with another (e.g., recording a glucose level of 927 mg/dL instead of 127 mg/dL).
  To identify and mitigate these errors, a method involving the detection of outliers was implemented. By eliminating the 5 % of values located at the extremes, it became possible to pinpoint these extreme values, which often corresponded to the aforementioned human errors. Consequently, these extreme values were effectively removed from the database, ensuring a cleaner and more reliable dataset for further analysis and modeling.

- Another scenario in which outliers were found involved cases where the values, although within the accepted range of normality, were ultimately deemed invalid. This situation was particularly evident on the day when patients are discharged from the ICU. It was considered necessary to exclude such values, as it was expected that a patient would only be discharged when their medical records were considered valid and they no longer required constant monitoring by medical personnel. For instance, a patient with a blood glucose level exceeding 400 on the last day of their ICU stay would be considered an invalid value. To identify and remove these types of values, a meticulous examination of each parameter was conducted in collaboration with the medical staff, and a set of rules was established for each parameter.

- Among the parameters included in the dataset, there are binary variables that may contain anomalous data that cannot be easily identified through the traditional outlier detection approach. However, a similar process was employed to address this issue. With the assistance of the medical staff, a set of rules was established to identify and eliminate values that were incompatible with the patient's duration of stay. By applying these predefined rules to the binary parameters, any values that violated these rules were identified and subsequently removed from the dataset.

As part of the data preprocessing phase for developing the predictive models, a comprehensive examination was conducted to identify variables that contained missing values. This missing data can appear from various causes, including errors in extracting information from medical equipment into the database or inaccuracies in manually inputting data by healthcare professionals.

To handle these missing values, two primary strategies were employed: imputation and repetition of values:

- The first approach used to handle missing values involved replacing them with the average of the existing values within the same variable. This technique ensures that a representative value is obtained, minimizing the impact on the subsequent analysis. Given that the database comprises daily values pertaining to each patient, the average was calculated using the value from the previous day and the value from the following day. By considering these neighboring values, a more accurate estimation of the missing data was achieved.

- In situations where a missing value is encountered either on the first or last day of a patient's stay, the standard procedure of calculating the average cannot be employed. Instead, a technique known as value repetition is applied. This method involves substituting the missing value with the same value observed in one of the neighboring time points. Specifically, if the missing value is on the first day, the value from the following day is replicated, while if the missing value is on the last day, the value from the previous day is duplicated. This approach ensures that the dataset remains consistent and avoids introducing gaps or inconsistencies caused by missing values in the temporal sequence.

Lastly, another data treatment was conducted, similar to the previous outlier analysis, but specifically targeting the length of stay of patients. While the typical duration of ICU stays ranges from around 2 to 7 days, there exists a subset of

patients whose stays extend significantly longer. Although these prolonged stays are less frequent, they can have a detrimental impact on the accuracy of discharge prediction. To address this issue, all patients with stays surpassing 14 days were excluded from the analysis. This approach ensured that the predictive models would focus on more typical lengths of stay, enhancing their effectiveness in forecasting patient discharges.

2.2.3 Parameter types

The set of parameters comprising the database can be classified into five different types, depending on their nature: demographic, categorical, Boolean, dynamic and scale:

- Demographic data: They encompass various characteristics of a patient that are closely linked to their cultural and socioeconomic background. These variables, such as age and gender, play a significant role in determining the likelihood of patient recovery and their resilience against specific complications. Age, for instance, can influence the body's ability to heal and combat diseases, with younger patients often exhibiting higher resistance to certain health challenges. Similarly, gender may contribute to the prevalence of specific diseases, as biological and societal factors can impact susceptibility and resilience.

- Categorical data: These parameters refers to patient characteristics that can be categorized into a set of options based on their respective groups. Those groups can be the admission type of the patient (Emergency, Urgent, etc.) or treatment type (Surgical, Intensive, etc.) for example. These variables can be represented by text values or numeric variables that represent the specific group to which a patient belongs. It's important to note that categorical variables do not follow a numerical scale and all values within a category carry equal importance. Typically, categorical variables remain constant throughout a patient's stay and provide valuable insights into their specific group affiliations.

- Boolean data: The boolean data refers to parameters that can only have two possible options. While these variables can be represented in either numeric or text format, for the latter case, it was converted into a numeric value (0 or 1) to ensure compatibility with predictive models that exclusively accept numeric variables. Throughout a patient's entire stay, certain boolean variables may vary (e.g., the presence of a central venous catheter or CVC), whereas others remain fixed, such as gender, which is categorized as demographic information. These variables play a crucial role in assessing the likelihood of a patient's imminent discharge from the ICU. For instance, in cases involving mechanical ventilation, a patient cannot be discharged from the ICU while still relying on mechanical ventilation.

- Dynamic data: It refers to parameters that exhibit changes over the course of a patient's stay, providing valuable insights into their progression. These data points can be collected through devices connected to the patient during their hospitalization, as well as inputted into the database by the medical staff at the conclusion of their shifts. Dynamic data encompasses individual parameters, such as temperature or blood pressure, as well as score variables derived from calculations involving a combination of variables not explicitly stored

in the database. These score variables offer a representation of the patient's progression based on the overall outcome or result of a specific operation or assessment.

- Scale data: Scale variables are parameters that reflect the patient's health status using other variables. They provide insights into the severity or gravity of a patient's condition, either through a predefined set of values or a calculation involving other variables. An example of a scale variable is the Sequential Organ Failure Assessment (SOFA), which assesses the severity of the patient. By analyzing the patient's health status using indicators such as bilirubin or creatinine levels, a final value is derived. Another example is the APACHE-II, a classification system used to measure the severity or gravity of diseases.

2.2.4 Time treatment

The main objective behind creating these models is to forecast the day of patient discharge. That is why the data needs to be disaggregated on a daily basis. Each day of a patient's stay is considered as an independent clinical case from the rest of the days. The only variable that indicates the temporal progression is the number of days since admission (referred to as "Times Since Admission" or TSA). This enables the model to predict the day of discharge based on variables that describe the patient's clinical condition, rather than the specific day they are taken. This approach involves excluding variables that are not constant and are recorded every day.

Name	Sort	Mean	Stdev	Min	Max	Explanation
Age	DE	58.95	15.5	18	99	Age of the patient at ICU admission
PrefHospDays	N,CP	2.28	5.93	0	35	Days in hospital before ICU admis.
APACHEII	CP	21.45	8.82	0	51	APACHE-II score evaluated at ICU admis.
MAP	CP	71.87	15.49	20	137	Mean arterial pressure in the day
HR	CP	114.18	23	45	200	Heart rate mean value in the day
Tmp	CP	36.92	0.72	32.4	41.3	Body temperature mean value in the day
Glu(min)	CP	108.23	26.29	20	300	Glucose minimum value in the day
Glu(max)	CP	151.98	48.11	63	330	Glucose maximum value in the day
Glu(stdev)	CP	18.54	16.11	0	82.73	Glucose variation value in the day
STRATIFY	CP	2.67	1.12	0	5	Scale for identifying falll risk factors
SOFA-Cardio	CP	0.76	1.19	0	4	SOFA SCORE (cardiovascular system)
SOFA-CNS	CP	1.13	1.29	0	4	SOFA SCORE (nervous system)
SOFA-Cong	CP	0.33	0.76	0	4	SOFA SCORE (coagulation)
SOFA-Liver	CP	0.14	0.52	0	4	SOFA SCORE (Liver)
SOFA-Renal	CP	0.27	0.75	0	4	SOFA SCORE (Kidneys)
SOFA-Resp	CP	0.56	0.88	0	4	SOFA SCORE (respiratory system)
SOFA-Total	SC	3.21	2.82	0	20	Addition of all the SOFA Scores
NAS	SC	52.14	17.09	0	148	Nursing Activity Score (NAS)
EMINA	SC	8.51	2.82	0	15	Risk of developing pressure ulcers

TABLE 2.3: Description of the patient in the heterogeneity study for numerical values (N: numerical, DE: demographic, CP: clinical parameter, SC: scale.

2.2.5 Final description

All the information has been anonymized to respect patient privacy. There is no ID by which to track them, nor have their names or surnames been added, as well as data related to their country of origin or address. After pre-treatment of the data

Name	Type	Proportion	Explanation
AdmWard	C,AD	ER:42.57% Surgery:19.09% Scheduled:18.52% Others:19.81	%ICU admission type
APACHEAdm	C,AD	Postoperative:16.82% HF:16.55% Trauma:11.04% RF:10.12% Neurological:8.18% Other:37.7%	APACHE admission group
CHE	C,CP	5:21.66% 0:15.28% 3:14.12% 6:12.45% 2:11.49% Rest:25%	Charlson score
Gender	CB,DE	Male:67.33% Female:32.67%	Gender of the ICU patient
PatType	CB,DE	Surgical:54.85% Medical:45.15%	Type of ICU patient
AdmType	CB,DE	Urgent:64.15% Scheduled:35.85%	ICU admission type
AE	CB,CP	0:97.2% 1:2.8%	Adverse event
CA	CB,CP	N:62.89% Y:37.11%	Arterial catheter
SU	CB,CP	Y:80.92% N:19.08%	Urinary catheter
CVC	CB,CP	Y:51.87% N:48.13%	Central Venous catheter
Insuline	CB,CP	Y:53.29% N:46.71%	Insuline
VA	CB,CP	N:80.25% Y:19.15%	Vasoactive drugs
SA	CB,CP	N:58.75% Y:41.25%	Sedative analgesics
ATF	CB,CP	N:94.08% Y:5.92%	Antifungicals
VMI	CB,CP	N:58.46% Y:41.51%	Invasive mechanical vent.
VMNI	CB,CP	N:95.71% Y:4.29%	Non-Invasive mechanical vent.
Isol	CB,AD	N:82.19% Y:17.81%	Isolation
LTSV	CB,AD	N:97.84% Y:2.12%	Life support limitation

TABLE 2.4: Description of the patients in the heterogeneity study for categorical and binary values (C: categorical, CB: categorical binary, DE: demographic, CP: clinical parameter, AD: administrative.

and discarding the patients whose parameters did not meet the need to have the data broken down by days, a database of 3973 patients discharged alive between 2016-19 at the Joan XXIII hospital is available.

The database is composed of 5 categorical variables (Type of patient, Type of admission, Admission room, APACHE admission group and Main diagnosis), 11 Boolean variables (Arterial catheter, Urinary catheter, Central Venous catheter, Insulin, Vasoactive drugs , Sedoanalgesia, Antifungals, Antibiotics, Invasive mechanical ventilation, Non-invasive mechanical ventilation, Isolated) 4 demographic variables (Age, gender, Year and previous days of admission), 12 dynamic variables (Blood pressure, Heart rate, Temperature, Minimum glucose, Glucose maximum, Glucose standard deviation, Days from admission and Days until discharge) and 13 scale variables (APACHE-II, Comorbidity, Adverse elements, Stratify, NAS, EMINA, SOFA Cardio, SOFA CNS, SOFA Coagulant, SOFA Liver, SOFA Renal , SOFA Respiratory and SOFA Total). Table 2.3 shows a description of the numerical and Table 2.4 shows a description of the categorical and binary variables of the database.

2.3 Public database

2.3.1 Introduction

The database provided by the Hospital Joan XXIII was used in the first study for training the ICU patient discharge prediction algorithm. However, due to the limited number of patients in the database, there was a lack of progress in improving the

predictions. As a result, the potential use of a larger database was considered. There are many public database that provide access to a vast amount of real world ICU patient data. Some of the most used databases that are widely used in this domain are:

- MIMIC-IV: The Medical Information Mart for Intensive Care (MIMIC-IV) is a widely-used database that provides access to anonymous health records of over 40,000 patients admitted to the intensive care unit between 2008 and 2019. It encompasses a diverse range of information, including vital signs, laboratory measurements, medications, demographics, and clinical outcomes.
- MIS Database: The Medical Information System (MIS) Database is a comprehensive ICU database that contains clinical data from patients admitted to the ICU. It includes information such as vital signs, laboratory results, medications, and outcomes. The database is designed for research purposes and can be accessed by researchers to study various aspects of critical care.
- eICU: The eICU Collaborative Research Database is a comprehensive database focused on intensive care clinical information. It is employed for research purposes and enhancing medical care in intensive care units (ICUs). It contains extensive details about patients admitted to ICUs, encompassing demographic data, diagnoses, treatments, and clinical outcomes. The eICU database was developed as part of the Philips eICU project, which is a remote monitoring system implemented in ICUs across hospitals worldwide. It comprised information on over 200,000 patients admitted to ICUs in hospitals across the United States between 2014 and 2015.

The eICU database offers extensive ICU patient data and an ease of data processing in a daily format that aligns with the research objectives of this thesis. The database obtains data from electronic medical record systems employed in the ICU to capture patient information throughout their stay. This comprehensive database is utilized for clinical research purposes, including studying critical illness epidemiology, forecasting patient outcomes, and enhancing the quality of ICU treatment.

However, it is important to note that access to the database is subject to regulation and limitations due to ethical considerations and privacy concerns. Stringent restrictions and ethical regulations have been implemented to ensure the protection of patients' privacy and the confidentiality of the information stored in the database. We followed all the required conditions to get access to this database for the purpose of this research.

2.3.2 Data pretreatment

Data extraction follows the same scheme as the previous database, that is, transforming the data to obtain a single daily line for each day the patient is in the ICU. To this end, a set of treatments has been done as described below:

- Only variables that had daily data available will be used. This implies that the laboratory samples were discarded because they were specific data. Variables

with gaps of more than two days were also discarded since they did not allow the same treatment to be carried out as with the previous base of interpolation or repetition of values depending on the day of stay in which the missing value is found.

- Another new feature of the private database is the ability to determine the patient's destination on the day of discharge. This information allows us to determine whether the patient is transferred to a different ward or, alternatively, if they are transferred to the operating room or another location that does not necessarily indicate their recovery. In such cases, only the duration of the patient's stay in the recovery ward is considered, as being transferred to the operating room is viewed as an interruption of the stay and could potentially introduce confusion to the algorithm.

- Similar to the transfer of a patient to an operating ward, mortality is considered an interruption in the patient's progress towards recovery. As a result, all patients who have passed away have been excluded from the analysis.

2.3.3 Parameter types

The variables were categorized into five distinct types as done in the public database (Section 2.2):

- Demographic data: The variables that describe the cultural and socioeconomic profile of a patient are much more extensive in this database. In addition to age or gender, others such as ethnicity or occupation are also included.

- Categorical data: The variables used to classify patients into specific groups based on a predefined set of options have been expanded. Alongside variables such as admission type or prior treatment, we now also consider the patient's destination, enabling us to differentiate between those who were transferred to the ward and those whose treatment was interrupted due to surgery or a similar procedure.

- Boolean data: A boolean variable refers to a data that can have two possible values: "true" or "false." It represents a binary condition or state related to a specific aspect of the patient's condition or treatment.

- Dynamic data: Dynamical data refers to patient information that changes over the course of their hospital stay.

- Scale data: Scale data reflects patient's health based on a numerical value.

2.3.4 Time treatment

To ensure accurate prediction for patients upon their discharge from the ICU, all patient data is analyzed on a daily basis. To have more information, in addition to the average daily values, two new variables have been introduced: the maximum and minimum values for both boolean and numerical variables.

Initially, the numerical value for each patient in the private database was an average of the daily values throughout their stay. In this database the maximum and minimum values attained on each day are also included. It should be noted that these values are obtained after removing outliers, which improves the accuracy of the data.

Regarding boolean variables, the maximum and minimum values represent the highest and lowest values, respectively, recorded for each day of a patient's stay. This information allows us to determine if there has been a change in the patient's condition or if it has remained stable.

As part of the daily treatment process, certain variables that did not meet the requirement of evolving throughout the patient's stay, such as laboratory variables (test results that do not change), were excluded.

Upon analyzing variables broken down by the hour, it was discovered that the last day of a patient's stay in the ICU typically corresponds to the final hours rather than the entire day. Consequently, due to the lack of medical input during these hours, the patient's last day of stay was eliminated.

2.3.5 Final description

All the information is anonymized and it is impossible to identify the patient. The internal IDs corresponding to the different tables that compose the database incorporate IDs to link the tables between them but do not allow further investigation.

After pretreatment of the data and discarding the patients whose parameters did not meet the need for data broken down by days, a database of 16,585 discharged-alive patients is available.

The database is composed of 2 categorical variables (APACHE admission group and Unit Type), 6 Boolean variables (Mean, Minimum and Maximum non-invasive mechanical ventilation and Mean, Minimum and Maximum invasive mechanical ventilation), 3 demographic variables (Age, Gender and Ethnicity), 12 dynamic variables (Mean, Minimum and Maximum Temperature, Mean, Minimum and Maximum SpO2, Mean, Minimum and Maximum Blood Pressure, Mean, Minimum and Maximum Heart Rate) and 6 scales (Mean, Minimum Pain Scale and Maximum, Glasgow Comma Score Mean, Minimum and Maximum).

Table 2.5 shows a description of the variables variables of which the database is composed.

	HR	Tmp	Glu(min)	Glu(max)	Glu(std)	NAS	EMINA	TSA	STRATIFY	SOFA Cardio	SOFA CNS	SOFA Coag	SOFA Liver	SOFA Renal	SOFA Resp	SOFA Total
Count	34026	34026	34026	34026	34026	34026	34026	34026	34026	34026	34026	34026	34026	34026	34026	34026
Mean	114.61	37.01	108.74	153.68	18.91	53.52	8.94	13.22	2.78	0.88	1.29	0.33	0.16	0.28	0.67	3.61
std	23.04	0.77	26.40	47.93	16.09	17.57	2.86	17.37	1.12	1.27	1.36	0.76	0.55	0.73	0.93	2.99
min	45.00	32.40	20.00	63.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	99.00	36.50	92.00	120.00	8.00	41.00	7.00	3.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50%	112.00	36.90	105.00	142.00	14.57	49.00	9.00	6.00	3.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
75%	128.00	37.40	122.00	175.00	25.15	62.00	11.00	17.00	4.00	1.00	2.00	0.00	0.00	0.00	0.00	1.00
Max	200.00	42.00	200.00	330.00	82.73	155.00	15.00	159.00	5.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00

TABLE 2.5: Statistical description of the dynamic numeric variables in the dataset: heart rate (HR); temperature (Tmp); glucose (minimal, maximal, and standard deviation within the day); nursing workload score (NAS); pain scale (EMINA); time since ICU admission (TSA); risk of falling (STRATIFY); SOFA scales (cardio, central nervous system, coagulation, liver, renal, respiratory, and total).

2.4 Preliminary analysis of DTD

An initial study was conducted using the database provided by the Hospital Joan XXIII, presented in Section 2.2. This database contains information on patients who have been admitted in the ICU. From the time a patient is admitted, the team of healthcare professionals provides continuous care 24 hours and 7 days per week.

The Hospital Joan XXIII database contains a total of 3,973 patients who were admitted between January 2016 and November 2019. The database has been provided with a collection of initial treatments, which are outlined and described as follows:

- All patients with a stay of more than 14 days have been discarded. This is because in some cases the length of stay can take months, so we avoid similar cases by limiting the stay to two weeks. This allows training the model with cases in which a faster evolution of the patient is shown.
- The initial database has multiple variables to identify the evolution of the patient. However, not all variables are important and this can negatively impact the result of the predictions. That is why, using Feature Importance technique, we can identify those parameters that are most related to patient discharge. The four techniques used to define the importance of variables are Information Gain, Correlation, Gini Index and SVM.
- Another important factor to address is how to handle scale variables. These variables consist of scores derived from formulas that incorporate several parameters. As these variables already capture the overall value and were already included in the initial database provided by the hospital, they are removed from the database to prevent duplicated values. This case can be clearly seen in the SOFA scale where data such as creatinine or bilirubin intervene.

In total, there are 49 important variables that are clinically relevant. These variables include demographic data, laboratory results (such as creatinine, platelet count, and blood sugar level), clinical information (like oxygen saturation and primary and secondary diagnoses), and pharmaceutical data (such as the use of sedatives, vasopressors, and insulin). Additionally, the daily values of 15 scales that encompass other patient parameters like APACHE-II, CaM-ICU, SOFA, EMINA, NAS, CHE, and NAS have also been incorporated into the database.

2.4.1 Machine Learning for a basic DTD prediction model

The task of predicting the discharge day for critically ill patients in the ICU is a significant challenge. To face this problem, Machine Learning tools, specifically the Random Forest model, have been employed.

Random Forest enables the construction of multiple decision trees using the provided training data and some parameters. Each decision tree is trained independently and generates a prediction for the desired outcome. During the training process, the bootstrap technique is employed, which involves selecting random samples from the training data with replacement. This means that some data points may appear multiple times, while others may not appear at all. Additionally, only a random subset of features is considered at each split of every decision tree. Ultimately, the

Random Forest combines the predictions from all the decision trees and produces an averaged output as the final prediction. This approach reduces variance and enhances accuracy compared to using a single decision tree.

To properly train the Random Forest model, a 10-fold cross validation was utilized. This technique assesses the model's predictive capability by dividing the dataset into ten equal parts. The model is trained using nine of these parts and validated using the remaining one. This method helps prevent overfitting, which occurs when the model becomes overly specific to the training data and fails to generalize to new data.

Once the model was trained, its performance was evaluated using three metrics: mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R2). These metrics offer a comprehensive and accurate assessment of the model's effectiveness. The MAE measures the average absolute difference between the predictions and the actual values, indicating the average deviation of the predictions from the true values. On the other hand, the RMSE calculates the square root of the difference between the predictions and the actual values, assigning higher weight to larger deviations. Lastly, the R2 is a statistical measure that indicates the proportion of data variability that can be explained by the model.

2.4.2 Analysis of results

Out of the total 3973 patients included in the database, 67.33 % of them are male. The average age of the patients is 58.95 years, with a standard deviation of 15.5 years. The mortality rate is extremely low, and 85 % of the patients were hospitalized for a duration ranging from 2 to 7 days. Table 2.3 and table 2.4 shows the results obtained in detail. When examining the distribution of DTD for each patients we notice that some parameters demonstrate similarities, as depicted in figure 2.1.

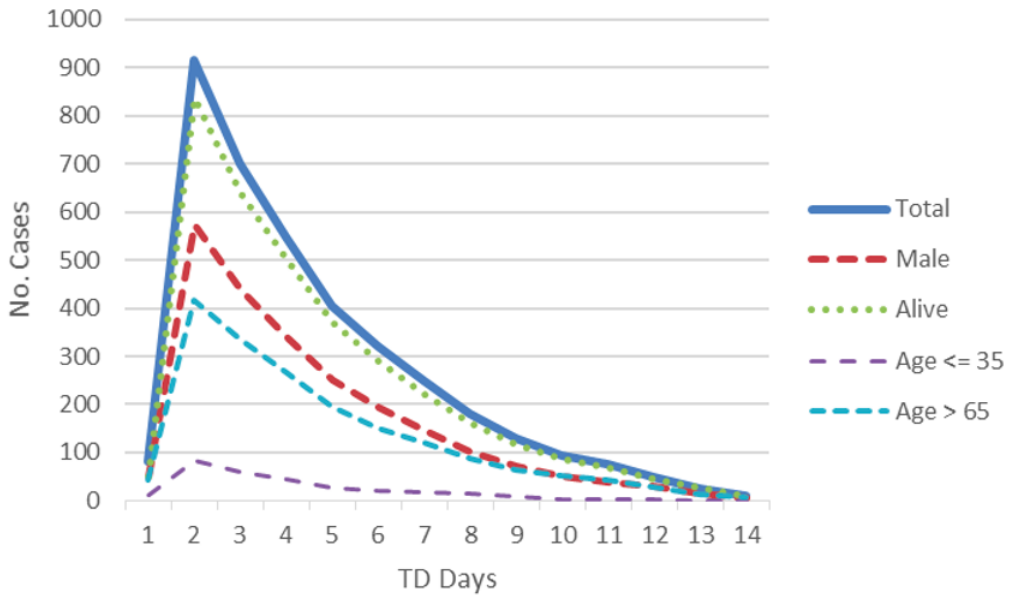


FIGURE 2.1: Distribution of LOS for different parameters

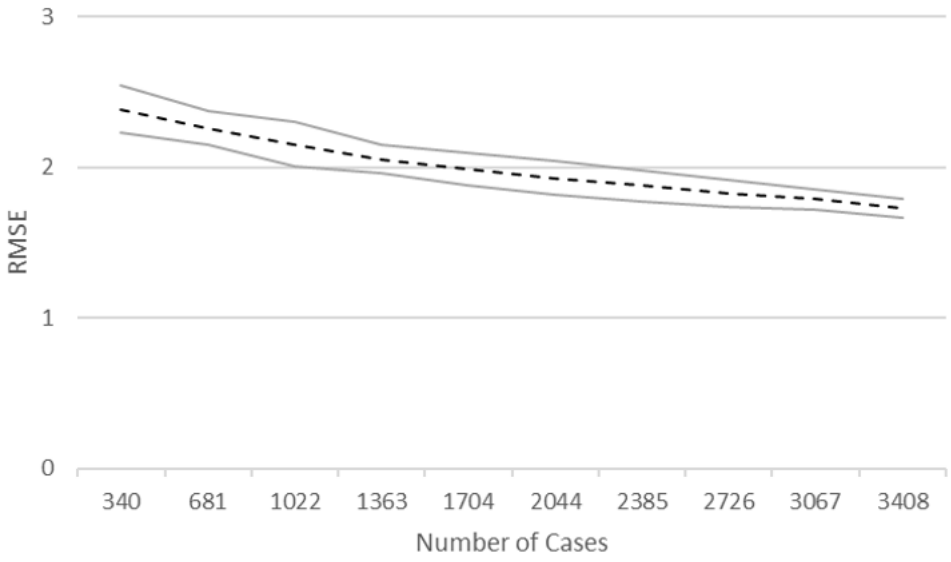


FIGURE 2.2: Progression of RMSE mean (dotted line) and RMSE mean ± st.dev (straight lines) for different cases.

In the process of applying the 10-fold cross-validation, we obtained certain values for Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R2) which were 1.34, 1.73, and 0.61, respectively. These results show an improvement compared to previous studies. Specifically, if we focus on the progression of RMSE and examine figure 2.2, we can observe that as we include more cases, the results continue to improve. Overall, the root mean square error RMSE values suggest predictions of less than a day for stays shorter than 10-11 days and less than two days for stays ranging from 11 to 13 days.

In this previous work, published in (Cuadrado et al., 2019), we did not break down our patient data by race or gender, nor did we create initial patient groups. This was primarily due to the limited amount of data that was available to us. As a result, we could not analyze at the beginning the impact that race or gender might have on patient outcomes or develop targeted interventions that could benefit specific patient groups.

Although we were unable to consider these factors in our analysis, our research still provided valuable insights into the care process and outcomes of ICU patients. Our initial approach helped us to focus on the key factors that impact the length of stay and days to discharge, which allowed us to identify areas for improvement in the care process, which motivated the work presented in the rest of chapters of this doctoral thesis.

Chapter 3

Methods and measures to quantify ICU patient heterogeneity

3.1 Introduction

In ICUs, it is essential to find comprehensive patterns or to develop predictive models to differentiate among patients with different DTDs. But, before we can identify those patterns or make accurate DTD predictors, it is important to know more about ICU patient heterogeneity and gain insight on the clinical parameters that could help identifying patients with different DTDs.

Focused on patients discharged alive, in this work, we developed four alternative methods and their corresponding measures to quantitatively ascertain the heterogeneity of ICU patients with regard to their DTD values.

The first method (clinical parameter analysis) assumes that, as time passes, the values observed for some (or all) clinical parameters converge to "normality"[1] values, which justify patient's discharge from the ICU. On the other hand, the variability of values of these clinical parameters decreases as the patients approach to their discharge day.

The second method (severity scales analysis) assumes that some clinical scales measuring patient's condition severity or care needs, such as SOFA (Moreno et al., 1999; Lambden et al., 2019), NAS (Miranda et al., 2003), or EMINA (Roca-Biosca et al., 2015), tend to normality values as the ICU patient stabilizes and improves along the consecutive days in the ICU.

The third method (confusion analysis) uses a similarity function between patients to calculate how many ICU patients discharged in x days are similar to ICU patients discharged in y days. If these measures are expressed as percentages, we obtain that if $x = y$, 100% of the patients are similar, if $x < y$, the percentage describes the risk of premature discharges, if $x > y$, the risk of overdue discharges, and if $x \neq y$, the risk of wrong patient discharges in x days. Here, patient heterogeneity is seen as the proportion of cases not discharged in x exact days, that are indistinguishable from cases that are discharged exactly in x days.

The fourth method (cluster analysis) borrows internal evaluation methods from the theory of cluster analysis to calculate the quality of DTD groups. Three of the most common indices to calculate grouping quality are Davies-Bouldin (Davies and

[1] Here, normality must be understood in the context of an ICU.

Bouldin, 1979), Dunn (Dunn, 1974), and silhouette (Rousseeuw, 1987). These measures can be modified to describe the level of heterogeneity among cases which are clustered according to their DTD values.

The formalization of these four methods and metrics is done in section 3.2.2. An study about the heterogeneity of patients at Joan XXIII Hospital is presented. The results are exposed and discussed in section 3.3. All the studies confirmed the large heterogeneity of ICU patients in this reference hospital.

3.2 Methods

Four methods to measure ICU patient heterogeneity with regard to their respective days to discharge are formalized in this section.

3.2.1 Formal definition of the base concepts

Our analysis of heterogeneity is based on the study of N ICU patients discharged alive, with P_i ($i = 1, ..., N$) being a sequence of daily descriptions of the patient in terms of m clinical parameters; i.e., $P_i = (d_{i1}, ..., d_{i\ell_i})$, with ℓ_i the length of stay (total days of stay) of the i-th patient, $d_{ij} = (\mathrm{v}_{ij1}, ..., \mathrm{v}_{ijm})$ the description of the i-th patient in her j-th day before ICU discharge (DTD), and v_{ijk} the value of the k-th parameter of patient P_i in her j-th DTD. We assume d_{ij} structures do not contain missing v_{ijk} values. This is a reasonable assumption when dealing with ICU patients, whose clinical parameters are either automatically taken by ICU devices or systematically recorded by ICU staff. Parameters which are measured several times along the day can be replaced by one or more alternative parameters representing aggregated values such as the minimum, maximum, average, or the standard deviation in the day.

All m clinical parameters have a minimum value and a maximum value (see Eq.3.1 and Eq.3.2) that are used to normalize the v_{ijk} values in an interval $[0, 1]$ with the min-max normalization Eq.3.3.

$$min_k = \min_{\substack{i=1,...,N \\ j=1,...,\ell_i}} \{\mathrm{v}_{ijk}\} \tag{3.1}$$

$$max_k = \max_{\substack{i=1,...,N \\ j=1,...,\ell_i}} \{\mathrm{v}_{ijk}\} \tag{3.2}$$

$$v_{ijk} = \frac{\mathrm{v}_{ijk} - min_k}{max_k - min_k} \tag{3.3}$$

Patient descriptions d_{ij} are used to define DTD groups. A group DTD_j is defined as the set of all the patient descriptions corresponding to the j-th day before discharge; see Eq.3.4. The number of patient descriptions in a DTD_j group is $n_j = \#\text{DTD}_j$. Notice that, if $j < j'$, then $n_j \geq n_{j'}$ because some patients could have a length of stay between j and j' days.

$$\text{DTD}_j = \bigcup_{i=1,...,N} \{d_{ij}\} \tag{3.4}$$

Clinical parameters can be numerical or categorical. The similarity between two normalized values v and v' of a given numerical parameter is $sim_j(v, v') = 1 - (v - v')^2$. If the parameter is categorical, the similarity between two categories v and v' is $sim_j(v, v') = 1$, if $v = v'$, or 0, otherwise.

Using these parameter similarity functions, we define the similarity between any two patient descriptors d_{ij} and $d_{i'j'}$ as their root mean square resemblance (Eq.3.5).

$$sim(d_{ij}, d_{i'j'}) = \sqrt{\frac{1}{m} \cdot \sum_{k=1}^{m} sim_k(v_{ijk}, v_{i'j'k})} \tag{3.5}$$

Based on the previous concepts, we can formally describe the four methods and related metrics that we propose to calculate DTD-based heterogeneity of patients in an ICU. For the sake of simplicity, we consider DTD groups as $\text{DTD}_i = \{d_{i1}, ..., d_{in_i}\}$ $(i = 1, ..., \ell)$, with ℓ the largest observed LOS, and the patient descriptors in the DTD_i group as $d_{ij} = (v_{ij1}, ..., v_{ijm})$, with v_{ijk} the normalized value of the k-th clinical parameter for the patient description d_{ij} of a patient that is in her i-th day before discharge.

3.2.2 Method 1: clinical parameters analysis

The clinical parameters should converge to normality values as the patients' day of discharge alive approaches. Simultaneously, the variability of the observed values should decrease as the discharge day is closer. Heterogeneity here corresponds to how much the average value of the parameters in a DTD_i group deviates from the normality values or normality range and, alternatively, how much the variability of these values for patients within a DTD_i group decreases as i approaches to one.

In order to measure these values, we use the mean and standard deviation Eq.3.6 and Eq.3.7 of the clinical parameter k for the patients in DTD_i, and represent the functions $f_k(x) = mean(k, x)$ and $g_k(x) = stdev(x, k)$ $(k = 1, ..., m)$, for each one of the m clinical parameters.

$$mean(k, i) = \frac{1}{n_i} \cdot \sum_{j=1}^{n_k} v_{ijk} \tag{3.6}$$

$$stdev(k,i) = \sqrt{\frac{1}{n_i} \cdot \sum_{j=1}^{n_k} (v_{ijk} - mean(k,i))^2} \tag{3.7}$$

The graphical representations of $f_k(x)$ and $g_k(x)$ show the evolution of the average value and variability of each clinical parameter as the discharge day is nearer. This allows visual analysis of the patients' heterogeneity with regard to their DTD.

3.2.3 Method 2: severity scales analysis

Several clinical scales or scores, such as SOFA, NAS, or EMINA, are frequently used in ICU to measure different dimensions of the patient complexity, condition severity, or care requirements. These scales simplify medical interpretation of the patient's organ failure (SOFA), the percentage of nursing activity required (NAS), or the risk of developing pressure ulcers as a combined assessment of mental state, mobility, incontinence, nutrition and activity (EMINA).

When treated as these medical parameters, the graphical representation of means and standard deviations of scales in a DTD_i group, as i evolves, provides information to analyze patient heterogeneity in terms of the evolution of the patient complexity. Intuitively, as a patient approaches to the discharge day, alive, scales should get closer to their respective normality values and the fluctuation of scale values among people in the same DTD group should reduce. Any other behaviour implies a greater heterogeneity between ICU patients.

3.2.4 Method 3: confusion analysis

The patient similarity function in Eq.3.5 can be used to identify which patients in DTD_i are similar to which patients in DTD_j ($i, j = 1, ..., \ell$).

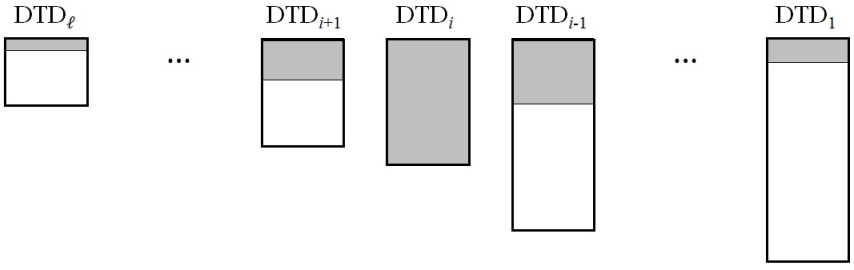
We use the similarity function defined in Eq.3.5 in combination with a threshold parameter $\delta \in [0,1]$, such that $S_\delta(i,j)$ in Eq.3.8 represents the set of all the patients in DTD_i which have a degree of similarity δ or higher to some patient in DTD_j. As δ approaches to one, similarity is more demanding and less patient descriptions d' in DTD_i are similar to the pivoting descriptions d in DTD_j.

$$S_\delta(i,j) = \bigcup_{d \in DTD_j} \{d' \in DTD_i : sim(d,d') \geq \delta\} \tag{3.8}$$

The number of patient descriptions in $S_\delta(i,j)$ is $n_\delta(i,j)$ and the total number of patient descriptors similar to patients in DTD_i is calculated as $s\delta(i) = \sum_{j=1}^{\ell} n_\delta(j,i)$ in Eq.3.8.

In Figure 3.1, each block represents one of the ℓ DTD groups DTD_ℓ, ..., DTD_1, with DTD_i the pivoting group. The dark areas are the patients in each group that are similar to some patients in the pivoting group. The table at the bottom shows some properties of the DTD groups: the number of patients in the group (n_j) and the

subset ($S_\delta(j,i)$), the number ($n_\delta(j,i)$), and the proportion ($\frac{n_\delta(j,i)}{n_j}$) of patients in the group who are similar to some of the DTD$_i$ cases.



#DTD$_i$	$1 \le n_\ell$	$\le$	...	$\le$	n_{i+1}	$\le$	n_i	$\le$	n_{i+1}	$\le$	...	$\le$	n_1
Similar to DTD$_i$ (cases or dark areas)	$S_\delta(\ell, i)$		...		$S_\delta(i+1, i)$		$S_\delta(i, i)$		$S_\delta(i-1, i)$		...		$S_\delta(1, i)$
Similar to DTD$_i$ (number of cases)	$n_\delta(\ell, i)$		...		$n_\delta(i+1, i)$		$n_\delta(i, i)$		$n_\delta(i-1, i)$		...		$n_\delta(1, i)$
Similar to DTD$_i$ (proportions)	$\frac{n_\delta(\ell, i)}{n_\ell}$		...		$\frac{n_\delta(i+1, i)}{n_{i+1}}$		$\frac{n_\delta(i, i)}{n_{i+1}}$		$\frac{n_\delta(i-1, i)}{n_{i+1}}$		...		$\frac{n_\delta(1, i)}{n_1}$

FIGURE 3.1: Confusion-based analysis of patient heterogeneity.

With Eq. 3.9 and Eq. 3.10, we can also calculate the average number of patients which are similar to other patients who were discharged n days before (or after).

$$b_\delta(n) = \frac{1}{\ell - n} \cdot \sum_{i=n+1}^{\ell} \frac{n_\delta(i, i-n)}{n_{i-n}} \tag{3.9}$$

$$a_\delta(x) = \frac{1}{\ell - n} \cdot \sum_{i=1}^{\ell-n} \frac{n_\delta(i, i+n)}{n_{i+n}} \tag{3.10}$$

The next values are used to define the following indicators and four confusion ratios to measure heterogeneity as a degree of confusion, premature discharge, overdue discharge, and feasible discharge error, respectively:

- The proportion of cases in DTD$_i$ which can be confused with patients in DTD$_j$ (i.e., patients that are discharged in i days who are similar to patients discharged in j days) (Eq.3.11).

$$c_\delta(i, j) = \frac{n_\delta(i, j)}{n_i} \tag{3.11}$$

- The proportion of cases similar to DTD$_i$ patients which are discharged in less than i days (i.e., premature discharges) (Eq.3.12)

$$p_\delta(i) = \frac{1}{s(i)} \cdot \sum_{j<i} n_\delta(i, j) \tag{3.12}$$

- The proportion of cases similar to DTD_i patients which are discharged in more than i days (i.e., overdue discharges) (Eq.3.13)

$$o_\delta(i) = \frac{1}{s(i)} \cdot \sum_{j>i} n_\delta(i,j) \tag{3.13}$$

- The proportion of cases similar to DTD_i patients which are discharged in a number of days different from i (feasible discharge errors) (Eq.3.14)

$$e_\delta(i) = \frac{1}{s(i)} \cdot \sum_{j\neq i} n_\delta(i,j) \tag{3.14}$$

3.2.5 Method 4: cluster analysis

In statistical classification, cluster analysis is based on the premise that elements in the same cluster are similar and elements in different clusters are dissimilar. Combining these intra-cluster similarity and inter-cluster dissimilarity concepts, there are several indices to assess the quality of a given clustering. Among them, Davies-Bouldin (Davies and Bouldin, 1979), Dunn (Dunn, 1974), and average silhouette (Rousseeuw, 1987) are some of the most used. The definition of these distance-based indices on the basis of ICU patient similarity indices is done with Eqs. (3.15), (3.16), and (3.17).

$$DB = \frac{1}{\ell} \sum_{i=1}^{\ell} \max_{1\leq j\neq i\leq \ell} \frac{2 - (m_i(C_i) + m_j(C_j))}{1 - sim(C_i, C_j)} \tag{3.15}$$

$$D = \frac{1 - \max_{1\leq i<j\leq \ell}\{sim(C_i, C_j)\}}{1 - \min_{1\leq i\leq \ell}\{m_i\}} \tag{3.16}$$

$$silhouette = \frac{1}{\sum_{i=1}^{\ell} n_i} \cdot \sum_{i=1}^{\ell} \sum_{d\in DTD_i} \frac{m_i(d) - m(d)}{1 - \min\{m(d), m_i(d)\}} \tag{3.17}$$

These are based on the identification of one representative patient description C_i for each DTD_i group, with Eq.3.18. This representative is the patient in the DTD_i group with the greatest average similarity to the rest of cases in DTD_i.

$$C_i = \arg \max_{d\in \text{DTD}_i} m_i(d) \tag{3.18}$$

The average similarity of any patient description d to a DTD_i group is calculated with Eq.3.19, after removing d from DTD_i, if $d \in DTD_i$. The average similarity within a DTD_i group is defined by Eq.3.20, and $m(d) = \max_{j \neq i} m_j(d)$ is the greatest average similarity of any patient description d in DTD_i to any other DTD group.

$$m_i(d) = \frac{1}{n_i} \cdot \sum_{d' \in DTD_i} sim(d, d') \tag{3.19}$$

$$m_i = \frac{1}{n_i} \cdot \sum_{d \in DTD_i} m_i(d) \tag{3.20}$$

The Davies-Bouldin, Dunn, and silhouette indices can be used to summarize the quality of the DTD groups according to the similarity of the patient descriptions that they contain, and therefore, provide information on the degree of heterogeneity of the patients in the DTD groups.

3.2.6 UCI patient heterogeneity analysis

All the adult patients admitted in the ICU of the University Hospital Joan XXIII, Spain, in the years 2016-2019 were taken to quantify the patient heterogeneity within that service. Only patients discharged alive were considered. The daily information of these patients in the 21 days previous to discharge was used in the analysis. Table 2.3 from summarizes the mean, standard deviation, min, and max values of the 19 numerical clinical parameters, and table 2.4 summarizes the proportions of the categories of the 18 categorical parameters. A distinction is made between numerical (N), categorical (C), and binary (CB) parameters, and also between demographic (DE), clinical (CP), scales (SC)[2], and administrative (AD) parameters. Both tables can be found in section 2.2.

A total number of 3,973 patients were involved in the study, with a mean LOS of 8.56 days per patient (6.95 after trimming the data to 21 days). The total number of daily patient descriptions was 27,611.

3.3 Results and discussion

The four methods proposed in this thesis chapter were applied to the data about the ICU patients described in section 3.2.6, to measure patient heterogeneity among these patients with regard to their days to discharge.

3.3.1 Patients heterogeneity based on clinical parameters analysis

Figure 3.2 describe the evolution of the mean and standard deviation of all the clinical parameters in the different DTD groups (f(x) and g(x) normalized to 0-1). The horizontal axes represent the number of days to discharge. In the first figure, most

[2]Scales different from SOFA, NAS, and EMINA were considered as clinical parameters in the analysis.

of the parameters remain stable as patients move from DTD_{21} to the day before discharge (DTD_1). Some others, such as SOFA-Cardio, SOFA-CNS, and SOFA-Resp, show a clear decrease from (denormalized) mean values 1, 2, and 1, down to normality value 0. Minimum and maximum mean daily glucose values also decrease from 159-111 to 143-104 mg/dl, and mean heart rate from 118 to 111 bpm. On the contrary, the mean arterial pressure (MAP) increases from 66 to 77 mmHg.

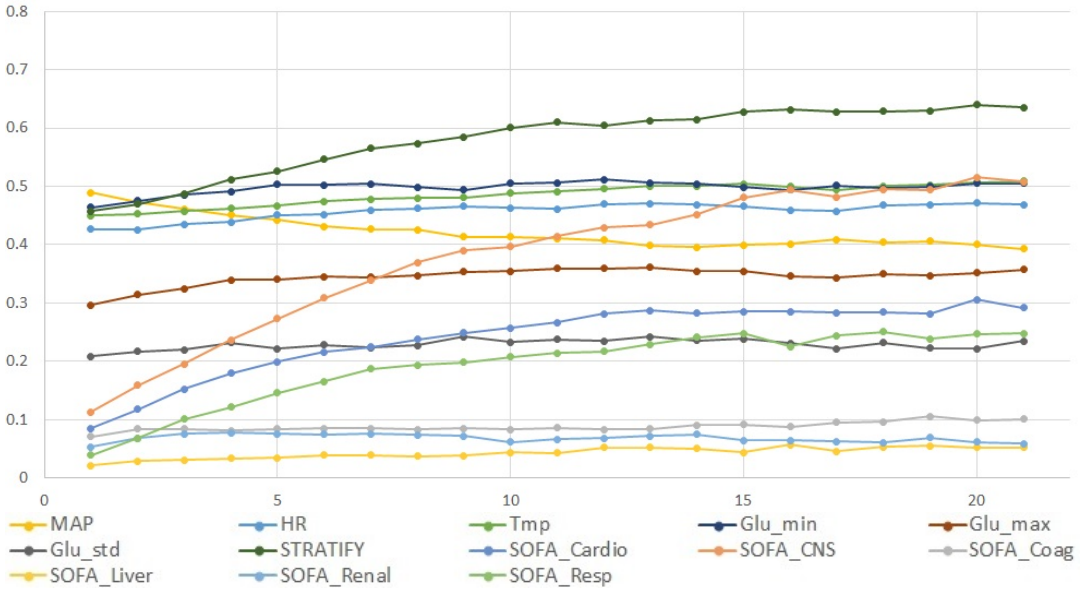


FIGURE 3.2: Clinical parameter of ICU patient heterogeneity for mean values.

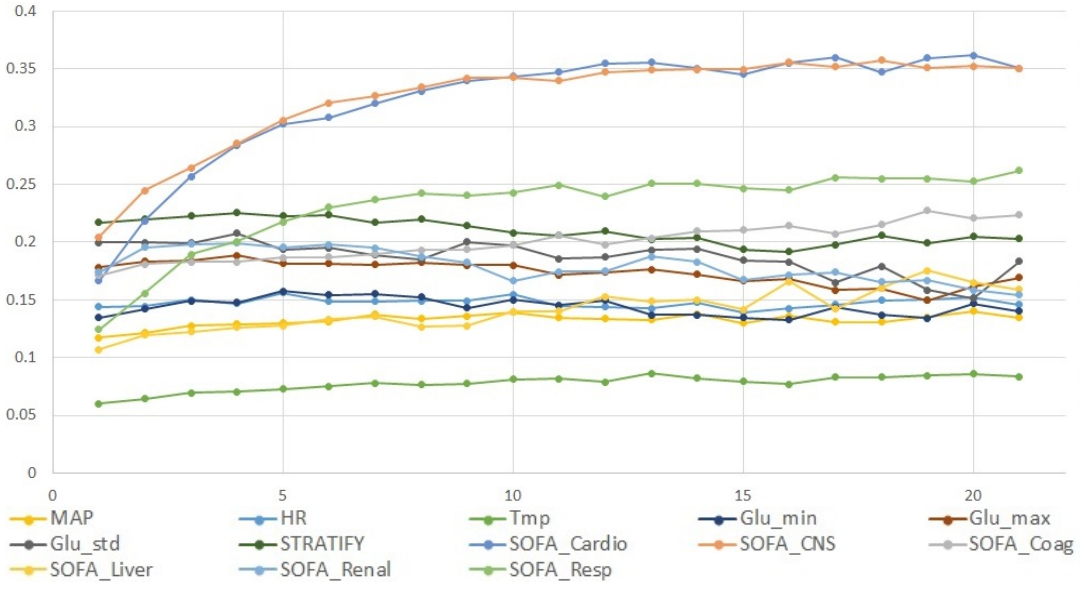


FIGURE 3.3: Clinical parameter of ICU patient heterogeneity for standard deviation.

Figure 3.3 describes variability of the clinical parameter values as the discharge day approaches. Decrease of variability is evident for several parameters, particularly in the last 10 days before discharge. For previous days, most of the parameters

show continuous oscillations in variational peaks and instability. The largest variability reductions are observed for SOFA-Cardio, SODA-CNS, and SOFA-Resp. Except for patient temperature, whose variability is always at a reasonable level, and SOFA-Cardio and SOFA-CNS whose variability is above 25% until the day before discharge, the rest of clinical parameters show a variability in the last 10 days between 10% and 25%. Among them, SOFA-Resp drastically reduces variability in the last days, but the rest keep a smooth oscillation that concludes with STRATIFY and Glu-std above 20%, Glu-max SOFA-Renal, SOFA-Coag, and SOFA-Cardio with 17%, and HR, Glu-min, SOFA-Resp, MAP, SOFA-Liver between 11% and 14%, the day before discharge.

These high variabilities explain the great heterogeneity of patients with regard to their clinical parameters, even if we only consider their conditions the day before discharge.

3.3.2 Patients heterogeneity based on scales analysis

When we focus on the scale parameters of the clinical conditions of the patients, we can observe some interesting patterns. Figure 3.4 shows the reduction of mean values of SOFA-Total, NAS, and EMINA scores in the last days before patient discharge, with a trend towards normality values 2, 44, and 7, respectively.

Their variability in figure 3.5 stabilizes in a decreasing trend in the last five days, with SOFA-Total and NAS reaching the lowest mean variation close to 10%. Such variations are considered high, particularly for patients in their previous day to discharge, and they reflect patient heterogeneity, concerning severity scales.

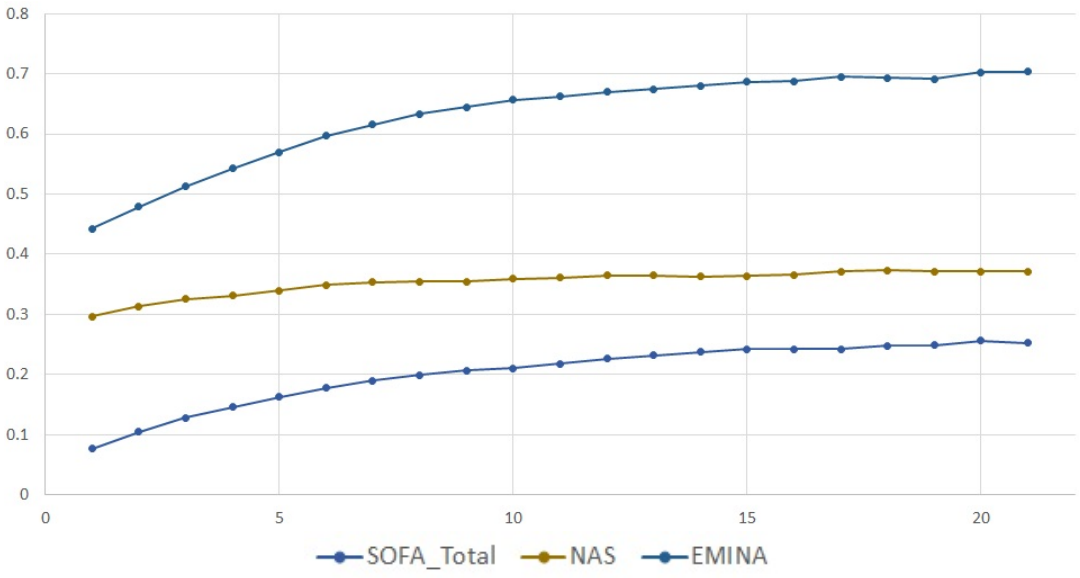


FIGURE 3.4: Scale analyses of ICU patient heterogeneity for mean values.

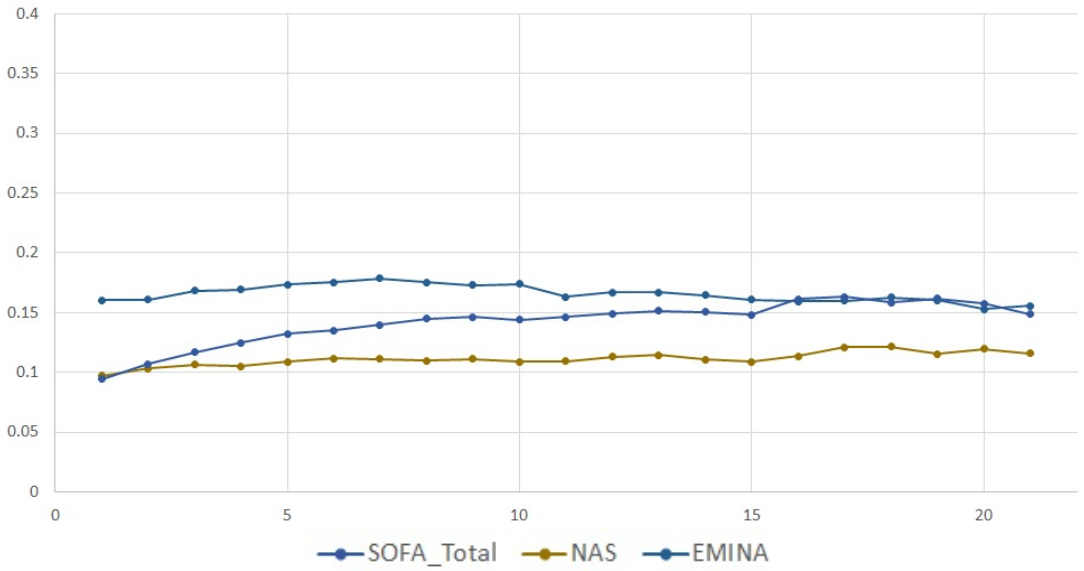


FIGURE 3.5: Scale analyses of ICU patient heterogeneity for standard deviation.

3.3.3 Patients heterogeneity based on confusion analysis

Table 3.1 shows the $n_\delta(i,j)$ matrix and the confusion measures for $\delta = 0.9$. This value of δ states that two patient descriptions are similar only if a similarity of 0.9 or higher is reached on a similarity scale from 0 to 1. The matrix shows that the largest values are in the diagonal $n_\delta(i,i)$, since obviously all the n_i patient descriptors in DTD_i are similar to some descriptor of a patient that will be discharged in i days. We also observe that the values close to the diagonal are also large and that they decrease as we move away from the diagonal. This means, that there is a degree of confusion between patients that are discharged in close days, and this confusion decreases as the difference between days to discharge gets larger. For example, position (4, 5) indicates that 715 patient descriptions, out of the 2488 descriptions of the patients 4 days before discharge, resemble 90% to some patients who are discharged in 5 days. Similarly, there are 416 patients to be discharged in 4 days that are similar to some patients that are discharged in one day (see position (4, 1)).

In table 3.2 we have the confusion measures defined in this chapter (for = δ 0.9). Columns p_δ and o_δ measure premature and overdue discharges. For example, in average, 19% of the patients who resemble some patients discharged in two days, are in fact discharged in one day, 29% of the cases similar to patients discharged in three days, are discharged in less than three days, and 32% of those similar to patients to be discharged in four days are discharged in less than four days. Similarly for column o_δ, 47% of the patients looking like patients to be discharged in one day, are in fact, discharged in two or more days, and 35% of patients similar to patients discharged in two days are discharged in more than two days. Adding the values in both columns we calculate the feasible average discharge error, which is represented in column e_δ. That is to say, the number of patients not discharged in i days who resemble some cases discharged in i days.

If we analyze these results in terms of patient heterogeneity, we can confirm that a large proportion of patients (37% on average) closely resemble the patients who

i\j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	3972	1491	748	416	254	155	108	73	62	39	40	28	18	17	8	7	7	7	8	2	3
2	1517	3799	1253	583	357	207	125	82	69	44	36	32	20	13	12	6	7	10	8	3	4
3	835	1316	3150	888	470	268	162	112	67	44	39	33	23	15	8	10	9	9	7	4	3
4	472	632	902	2488	709	358	200	122	88	61	45	38	27	14	13	10	7	9	4	4	5
5	281	383	478	715	2009	557	287	164	106	78	58	52	35	19	18	13	9	9	7	3	5
6	177	227	273	363	559	1635	414	219	151	99	73	60	33	26	21	17	10	11	6	8	8
7	124	143	175	208	289	417	1379	360	199	112	89	68	45	32	28	22	15	12	11	8	8
8	84	91	114	127	164	218	360	1169	291	168	114	75	53	43	30	24	19	20	12	6	4
9	69	74	68	92	106	150	195	291	1006	294	179	110	66	55	43	35	25	15	11	3	6
10	48	49	47	64	79	99	110	168	295	896	277	163	103	67	52	39	27	18	12	9	8
11	48	42	40	46	56	73	87	113	180	278	809	271	133	93	63	50	32	25	18	9	11
12	32	36	35	41	51	61	68	76	110	163	271	740	223	117	94	58	38	30	15	12	11
13	22	22	24	28	35	34	44	54	68	103	134	223	679	191	122	79	67	43	25	18	17
14	21	14	15	15	19	26	32	43	56	67	93	117	191	619	183	107	65	53	32	22	16
15	9	13	8	13	18	21	28	30	43	52	64	94	122	183	568	174	114	65	41	36	24
16	12	8	10	12	13	17	22	24	35	39	51	59	79	107	174	524	155	95	55	42	33
17	8	7	10	8	9	10	15	19	25	27	32	38	67	65	114	155	493	145	78	51	34
18	9	11	9	11	10	11	12	20	15	18	25	30	43	53	65	95	145	464	154	88	53
19	11	9	7	5	7	6	11	12	11	12	18	15	25	32	41	55	78	154	430	132	73
20	2	3	4	4	3	8	8	6	3	9	9	12	18	22	36	42	51	88	132	406	124
21	3	4	3	5	5	8	8	4	6	8	11	11	17	16	24	33	34	53	73	124	376

TABLE 3.1: n$\delta = 0.9$(i, j) matrix similarity for $\delta = 0.9$

were discharged earlier, and that a non inconsiderable number of patients (26% on average) are very similar to patients who were discharged later. In total, 63% of the patients discharged one day resemble patients who where discharged other days, on average. This represents a high heterogeneity in terms of DTD prediction since patients who share practically the same clinical description may have different discharge days.

Finally, columns b_δ and a_δ calculate equations 3.9 and 3.10, for $n = 1, ..., 10$. They represent the average proportion of patients that resemble other patients that are discharged n days before or after. For example, in average 28% of the patients are similar to patients who are discharged one day before the first ones, and 16% are similar to patients who are discharged two days before. Similarly, 32% of the patients are equivalent to the patients discharged one day later.

3.3.4 Patients heterogeneity based on cluster analysis

All patient descriptions in the dataset define a clustering which is determined by the days to discharge of each patient description. These clusters correspond to the DTD_i groups ($i = 1, ..., 21$). The elements in DTD_i describe patients who are i days before discharge, and they are expected to be mutually more similar than patient descriptions in different DTD groups.

Results in Table 3.3 shows that when calculating Davies-Bouldin, Dunn, and silhouette indices for DTD groups we obtain the values 11.43, 0.037, and -0.054. According to (Halkidi, Batistakis, and Vazirgiannis, 2001; Pakhira, Bandyopadhyay, and Maulik, 2004; Kaufman and Rousseeuw, 1990), these values suggest that the

i\j	ni	s(i)	pδ	oδ	eδ	bδ
1	3972	7463	0,00	0,47	0,47	0,28
2	3799	8187	0,19	0,35	0,54	0,16
3	3150	7472	0,29	0,29	0,58	0,10
4	2488	6208	0,32	0,28	0,60	0,06
5	2009	5286	0,35	0,27	0,62	0,05
6	1635	4390	0,36	0,26	0,63	0,03
7	1379	3744	0,36	0,27	0,63	0,02
8	1169	3186	0,36	0,27	0,63	0,02
9	1006	2893	0,36	0,29	0,65	0,01
10	896	2630	0,36	0,29	0,66	0,01
11	809	2477	0,39	0,28	0,67	
12	740	2282	0,41	0,26	0,68	
13	679	2032	0,39	0,28	0,67	
14	619	1806	0,39	0,26	0,66	
15	568	1720	0,41	0,26	0,67	
16	524	1566	0,42	0,24	0,67	
17	493	1410	0,43	0,22	0,65	
18	464	1341	0,43	0,22	0,65	
19	430	1144	0,44	0,18	0,62	
20	406	990	0,46	0,13	0,59	
21	376	826	0,54	0,00	0,54	
		mean:	0,37	0,26	0,62	

TABLE 3.2: Confusion measures of ICU patient heterogeneity: $DTD_i = l$ cardinalities n_i, and confusion indices $p_\delta(i)$, $o_\delta(i)$, $e_\delta(i)$, $b_\delta(n)$, $a_\delta(n)$ $(n = 1, ..., 10)$ for $\delta = 0.9$.

DTD clusters provide 'no substantial structure'. In other words, patients with a similar DTD value are not necessarily similar, and patients with different DTD values are not necessarily different, in clinical terms. This heterogeneity could be partially attributed to the inclusion of patients with a large number of stay at ICU in the study. For example, it would be natural for the patients in the DTD_{21} group to be very different from each other and thereby affect the final value of the indices. However, when we consider exclusively the patients during their last week in the ICU, the values obtained are 16.60, 0.03767 and -0.0403. And if only patients in the last three days are considered, the values are 1.806, 4.21 and -0.0133, still very far from those that describe a good classification. The Davies-Bouldin, Dunn and silhouette indices therefore confirm the heterogeneity of the patients, according to their DTDs.

Group	All periods	Last 7 days	Last 3 days
Davies-Bouldin	11.43	16.60	1.806
Dunn	0.037	0.03767	4.21%
Silohuette	-0.054	-0.0403	-0.0133

TABLE 3.3: Results for patient heterogeneity based on cluster analysis.

3.4 Conclusions

In this chapter, several measures of patient's heterogeneity have been defined, each measure takes a different perspectives of heterogeneity, using different indicators to establish a procedure for evaluating the differences between patients at ICU. This work contributes with a new set of methods and metrics to quantify patient heterogeneity in ICUs with a focus in the prediction of the days to discharge. These methods and metrics come to cover a gap on the current available tools for patient heterogeneity analysis.

The proposed metrics have been tested on the dataset of the ICU of the University Hospital Joan XXIII. Results confirm a high heterogeneity of conditions of the patients within the same DTD group and also confusion between patients with distant DTDs, which may contain very similar patients. This is a clear indicator of the complexity of making accurate DTD predictors, and it might be one of the reasons for such few published works on this relevant ICU issue.

Even the results concern an ICU of a single hospital, instead of evaluating a multi-center dataset, the inclusion of all the clinical cases seen in the ICU of the reference hospital in the Tarragona province in Spain, providing ICU services to a population of 750,000 inhabitants and including all the survival cases in four consecutive years, makes the study possibly representative of many other ICUs.

Due to this large coverage of this hospital, its ICU receives a great variety of patient types. Therefore, we can consider that the study done involves all the types of ICU patients, making the analysis of heterogeneity very relevant. Constraining our study to concrete patient types (e.g., surgical, scheduled, or emergency) would naturally decrease the heterogeneity values. However, this is not the purpose of the work on this doctoral thesis, which is develop a general method to predict the days to discharge for any ICU patient. In fact, other ICU studies such as (Vranas et al., 2017) use to involve all the ICU population in their works.

Other factors with a possible influence in the results are the internal organization of the ICU, which may have decided some patient discharges based on organizational reasons rather than pure clinical reasons. This information is not available, so it has not been possible to take it into account in this study.

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

Chapter 4

Proposing derived variables for Days-to-Discharge prediction

4.1 Introduction

In order to confront the long-term prediction limitation of static methods, dynamic tools propose a day by day prediction of the days to discharge (DTD). So, a patient with total LOS $= x$, in day $y < x$ is expected to have a DTD prediction of $x - y$ days. This dynamic approach presents some conceptual benefits: (1) predictive errors in day y can be fixed in subsequent days and clinical decisions readjusted, (2) predictions are current and not based on the patient condition at the admission time days ago, (3) as patients approach discharge, their clinical parameters tend to be normal[1], and it is expected that the accuracy of dynamic predictions may improve with respect to the static predictions in the admission day.

However, not many works have been published on the prediction of DTD for ICU patients. As demonstrated in Chapter 2, our preliminary study (Cuadrado et al., 2019), Random Forest was applied to construct a DTD predictor for general ICU patients which achieved an average root mean square error of 1.73 days, but patients has a maximum stay of 7 days at the ICU. If we want to extend the study to include more critical patients who stay at ICU for a longer time, we must take into account the problem of heterogeneity identified in Chapter 3. Patients heterogeneity may explain the reason why two clinically similar patients have very different discharge days, and also why clinically different patients may have the same discharge day. To try to reduce heterogeneity, we must study other kinds of indicators that may help in the estimation of DTD.

In this chapter we present a series of studies with Artificial Intelligence techniques that may be useful for the analysis and understanding of DTD in ICU. In particular, we have focused on two other dimensions: biomarkers and phenotype recognition.

- *Identification of biomarkers for DTD prediction*: In ICUs, multiple clinical signs from patients are continuously recorded. Some of these values are aggregated to provide a daily clinical summary of the patient's condition, when combined with other data that are captured once a day. These could be the subject of a

[1]This is only true for patients discharged alive for whom ICU discharge is due to stabilization of their vital signs.

feature selection process to determine which clinical parameters are more relevant to determine the DTD of ICU patients, thus acting as DTD biomarkers.

- *DTD phenotype recognition*: A consequence of ICU patient heterogeneity is that similar patients may have very different DTDs, and the opposite. This defines DTD prediction in ICUs as a complicated task that could be simplified by the identification of patient phenotypes representing subgroups of patients, all of them with identical DTD. Technically, these phenotypes should have DTD precisions close to one, and DTD recalls as high as possible.

In section 4.2, we describe the methods that we followed for the analysis of these new types of indicators. In section 4.3, we discuss about the clinical implications and benefits that these results can entail. The conclusions are exposed in section 4.4.

4.2 Methods and Results

In this section, we address DTD-related information whose analysis can contribute to a better understanding and management of patients in ICUs. For each one of these issues, we describe the methods and Artificial Intelligence techniques that we have used to deal with them, and the results obtained. All the studies were made on the public dataset described in Section 2.3.

4.2.1 The data

Important information about patients admitted in the University Hospital Joan XXIII in Spain was captured in a database providing daily description of all the ICU cases in terms of demographic, clinical, and treatment features which are essential in the professional decision of discharging patients from ICUs.

All the 3,973 patients discharged alive in 2016-19 were considered in this study. The average LOS was 8.6 days, with 1 and 159 days the minimum and maximum LOS. Only information of patients in their seven last days in the ICU was considered. The average age was 59.9, with all patients between 18 and 99 years, and 1,424 (35.8%) being women. There were 2,878 (72.4%) medical cases and 1,095 (27.6%) surgical.

Forty-five parameters were considered: 5 categorical, 11 Boolean, and the rest numeric. Eleven were static in the sense that they remained constant during all the ICU stay of the patient. These were: "Age", "Gender", "DisYear", "PatType", "AdmType", "AdmWardGroup", "PrevHospDays", "APACHE_Adm_Group", "Pincipal_Diag_G", "APACHEII", and "CHE"[2]. Other parameters were taken at the rate of once a day, including signs (e.g., platelets, bilirubin, or creatinine), scales (e.g., NAS -

[2]"Age": age of the patient at admission time, "Gender": female or male, "DisYear": year of ICU diagnosis, "PatType": patient type as medical or surgical, "AdmType": either scheduled or urgency, "AdmWardGroup": source among ER, surgery, other hospital areas, or other hospital, "PrevHospDays": number of days in hospital previous to ICU admission, "APACHE_Adm_Group": post-operations, heart failure, respiratory failure, trauma, etc., "Pincipal_Diag_G": principal diagnosis group (e.g., respiratory system, infection, external injuries, etc.), "APACHEII": APACHE II score value at ICU admission, and "CHE": Charlson comorbidity index.

nursing work load, EMINA - pain scale, STRATIFY - risk of falling, or SOFA scores), and treatment actions (most of them Boolean; e.g. arterial catheter, urinary catheter, central venous catheter, insulin treatment, vasoactive drugs, analgesics, antibiotics, mechanical ventilation -either invasive or not, etc.). The remaining five parameters whose frequency of observation was below one day were aggregated per day and represented by their mean, standard deviation, min, or max daily values (e.g., heart rate, temperature, or glucose). As presented in Chapter 2, Table 2.5 describes the dynamic numeric variables in the data set used.

4.2.2 Identification of DTD Biomarkers

The complexity of DTD analysis may be not only related to the heterogeneity of ICU patients. Consequently, a different approach could be taken, based on the clinical parameters used to describe the ICU patients. To confirm this hypothesis, the relevance of these parameters in relation to the DTD was analyzed with several supervised feature selection methods. These methods were also used to identify low-relevant features.

Two approaches were followed: one which considered DTD as a n-ary variable and calculated the relevance of all parameters respect to this variable, and another one that binarized the DTD variable by applying dummy coding without comparison group. In this last one, n DTD binary new parameters DTD_x ($x = 1, ..., n$) were obtained, such that DTD_x was 1 for patients with DTD$= x$, and 0 for patients with DTD$\neq x$.

For the first method, filter-type feature selection was applied with five selection functions (Jović, Brkić, and Bogunović, 2015): information gain, information gain ratio, correlation, Chi square, and Gini index. The results were min-max normalized to allow cross comparison. The ten best parameters were kept as the most significant in the study of DTD. They were:

- SOFA-CNS
- EMINA
- sedative/analgesic
- SOFA-Resp
- arterial catheter
- SOFA-Cardio
- SOFA-Total
- vasoactive drugs
- central venous catheter
- NAS

For the second method, we obtained the list of the most significant parameters of every DTD value considered as an individual group . The normalized relevance of these features was used to select those which were among the 30% best ones to

TABLE 4.1: Significant biomarkers for each of last 7 DTD with a 30% threshold.

DTD group	Most significant parameters for biomarkers
DTD = 1	EMINA, SOFA-Total, SOFA-CNS, SOFA-Resp
DTD = 2	EMINA, SOFA-Total, SOFA-CNS, SOFA-Resp, STRATIFY, Tmp
DTD = 3	EMINA, SOFA-Total, SOFA-CNS, SOFA-Resp, STRATIFY, Tmp, TSA
DTD = 4	EMINA, SOFA-CNS, SOFA-Total, TSA, Tmp, SOFA-Resp, STRATIFY
DTD = 5	Tmp, TSA, SOFA-CNS, EMINA, SOFA-Total, SOFA-Resp, STRATIFY
DTD = 6	TSA, Tmp, Glucose min
DTD = 7	EMINA, Glucose min, HR, TSA, NAS, APACHE-II, Sofa-Resp

be considered for biomarkers. Table 4.2 shows the resulting features for each DTD group.

The level of overlapping of values of the top three DTD possible biomarkers (i.e., EMINA, SOFA_Total, and SOFA_CNS) is represented with the boxplots in Figure 4.1.

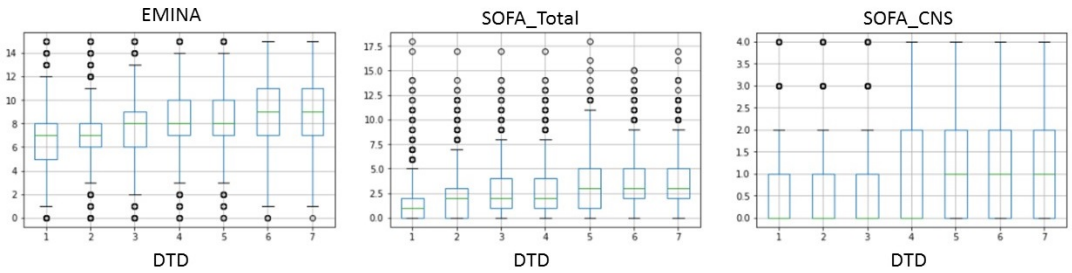


FIGURE 4.1: Overlapping of EMINA, SOFA_Total, and SOFA_CNS values among DTD groups.

4.2.3 Phenotype Extraction

The diversity of ICU patients in every DTD group limits the possibility of finding general DTD prediction models. In order to alleviate this problem of finding general descriptions for all patients in a DTD, we worked to identify patient phenotypes as descriptions of subgroups of patients with the same DTD who cannot be confused with patients with other DTDs.

To this end, we followed two methods. The first method applied subgroup discovery (Herrera, 2011; Helal, 2016) in order to find interesting associations among different variables with respect to the DTD, our property of interest. The second method combined feature selection (Chandrashekar, 2014) and unsupervised clustering, focused on numeric parameters, in order to first reduce the dimension of the problem, and then, based on the selected features, find out interesting subgroups of patients (i.e., phenotypes), with the use of k-means.

The first method, which used all the parameters of the dataset, discovered a subgroup of patients for each DTD, each one with a possibly different feasible phenotype. The second method selected the most relevant numerical parameters for every DTD group (i.e., group of patients with the same DTD). These are the ones contained

in Table 4.2.

TABLE 4.2: Feasible phenotype for patients in their last seven DTD

DTD group	Feasible phenotype
DTD = 1	SOFA-CNS = 0, SOFA-Resp = 0
DTD = 2	SOFA-CNS = 0, SOFA-Resp = 0
DTD = 3	SOFA-Resp = 0
DTD = 4	SOFA-Resp = 0
DTD = 5	SOFA-Liver = 0, TSA<2
DTD = 6	SOFA-Liver = 0, SOFA-Total = [2:4]
DTD = 7	SOFA-Total = [4:6]

During construction of phenotypes, the k-means algorithm was used to obtain three subclusters per DTD group. The selection of $k = 3$ for each group of patients with DTD=x attends to our intention to separate those patients who are closer to patients with DTD $< x$ (i.e., patients in a better health condition), and patients who are closer to patients with DTD $> x$ (i.e., patients in a worst health condition), from those which are dissimilar to these (i.e., patients with a health condition that justifies their discharge in x days), and therefore define a proper subgroup of patients for DTD $= x$. For these proper groups ($x = 1, ..., 7$) and for their respective selected features, a 95% confidence interval was calculated, in order to define one phenotype per DTD group.

As a result we obtained 21 feasible phenotypes, one for each one of the $k = 3$ subgroups of each one of the seven DTD groups. To evaluate the quality of the phenotypes, the dataset was split in two subsets: patients admitted from 2016 to 2018 were used to obtain the phenotypes as previously explained, and patients admitted in 2019 were used to make the testing.

We applied a Random Forest method to create a prediction model using the phenotypes. A a 10-fold cross validation was used, obtaining an accuracy of 75%, a sensitivity of 23%, and a F1-value of 24%. We can see that these are low quality scores for a prediction model, indicating that these phenotypes are not able to properly represent the knowledge required for DTD estimation. There results confirm the complexity of generating clinical decision models for the problem of the stay duration in ICU.

4.3 Discussion

Our experiences in the analysis of the days to discharge of patients in an ICU have shown that this is a complex area of work. We explored four different DTD-related issues with multiple methods and the help of a database on all the patients discharged alive in an interval of six years from an ICU.

The study of *patient heterogeneity* can confirm the complexity of the field, and having metrics to quantify this heterogeneity is not only good for benchmarking but also to gain insight on different interpretations of what heterogeneity in ICU-DTD means. When we gather all the patients in their x-th day before discharge in a DTD$_x$

group, the longitudinal analysis of the progression of the means (and standard deviations) of their clinical parameters contributes to identify whether there are some parameters that we have to look at in order to determine if a patient is close to discharge or not. In (Cuadrado et al., 2021), we found that SOFA-Cardio, SOFA-CNS, and SOFA-Resp are scores playing this role. Alternative scores such as SOFA-Total and NAS are also good at this purpose. An alternative interpretation of heterogeneity uses a distance function between ICU patients in order to calculate how confusing is to predict discharge for the patients in a DTD_x group by computing the quantity of patients in DTD_x who are similar to patients in other DTD groups. Our experiments showed that the degree of confusion is extremely high, which hinders the possibility of making good global DTD predictors, even for patients who are close to discharge.

The second issue analyzed concerns the identification of DTD *biomarkers* (Selleck, 2017). That is to say, measurable indicators concerning some biomedical condition of the ICU patients that could simplify the "diagnosis" of the patients' DTD group. Our application of several feature selection techniques identified SOFA-CNS, EMINA, and sedative/analgesic as best general biomarkers, but these may change if we focus on concrete DTD groups. For example, for patients in their 1, 2, 3, or 4 days to discharge EMINA, SOFA-Total, and SOFA-CNS seem to be the best biomarkers, but patient temperature (Tmp) becomes the most important for patients 5 or 6 days before discharge. According to (Selleck, 2017), for an indicator to become a good biomarker, it must meet (1) analytical validity (i.e., be accurate and reproducible), (2) clinical validity (i.e., be medically meaningful and useful differentiating between groups), (3) clinical utility (i.e., improve health care), and (4) other validities (e.g., cost-effectiveness, psychological implications, or ethical implications). Currently, the indices out of our study fail to satisfy clinical validity due to the high level of overlapping between DTD groups (see Fig. 4.1).

The third issue addressed in this work is the extraction of *phenotypes* from the data that could identify subgroups of patients with a positive day to discharge. Our first approach with subgroup discovery techniques obtained phenotypes based on SOFA_CNS, SOFA_Resp, SOFA_Liver, TSA, and SOFA_Total. In the same way as for biomarkers, the number of patients with different DTDs in the same phenotype is high. Our attempt to overcome this problem with the generation of biomarkers for each DTD_x group by means of a binary analysis of all patients with $DTD = x$, versus all patients with $DTD \neq x$, obtained low-sensitive DTD phenotypes that require futher improvement before they can be of practical use.

Our final issue concerned the application of supervised machine learning to construct *predictive models* of DTD for ICU patients. Our first approach with regular algorithms such as decision trees, naïve Bayes, or logistic regression were soon discarded in favor of ensemble methods such as Random Forest, which so far, is the best approximation that we have obtained for the analysis and prediction of the days to discharge from the ICU (Cuadrado et al., 2019).

4.4 Conclusion

Predicting DTDs of patients is essential to ICU management. Optimal prediction with models achieving an average error below one day is still far from being a reality. The high heterogeneity of ICU patients makes this a difficult objective. Here, we proposed two ways to analyze the DTD problem from different perspectives.

These methods will be applied and tested using real data from ICU patients in the next chapter. In addition, we discovered that before applying these methods, it might be helpful to filter out certain patients to reduce the overall diversity of cases, which could improve the accuracy of the predictions

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

Chapter 5

Predicting ICU patients discharge with a hybrid Machine Learning model

5.1 Introduction

We have seen in previous chapters that ICU patients may have a large heterogeneity, due to the diversity of their pathologies with an affectation of one or more threatened vital functions, which are potentially reversible. All this variability in admission extends throughout the patient's stay in the ICU, and is reflected in the great disparity in patient's evolution. Consequently, the days to discharge (DTD) prediction is a difficult task, although it is essential for the proper management of an ICU in terms of bed occupancy, pharmacological and non-pharmacological stock availability, staff provision, flow of patients to and from other hospital units, etc. (Bai et al., 2018).

We have also seen that ICU patients are assessed in terms of demographic parameters such as gender or age at the ICU admission, in addition to some clinical measures. During their ICU stage, some other clinical parameters such as temperature (T), heart rate (HR), mean arterial pressure (MAP), or peripheral oxygen saturation (SpO2) are systematically monitored, some of them continuously, some others at different discrete times during the day. These measurements are collected in the health information systems and are used for medical decision making (McKenzie et al., 2015). The hypothesis of this work is that these values can also be used to foresee the days to discharge (DTD) of the patient.

Most existing studies seen in Chapter 2 focused on the prediction of LOS at the admittance day. These LOS prediction methods reach root mean square errors (RMSE) of 0.47-8.74 days and mean absolute errors (MAE) of 0.22-4.42 days (Caetano, Laureano, and Cortez, 2014; Verburg et al., 2014; Li et al., 2019; Muhlestein et al., 2019). It is worth to note that the study that obtained the best results (Caetano, Laureano, and Cortez, 2014) applied the concept of tolerance, meaning that errors which were proportionally below the tolerance level were discarded in the calculation of the average errors (i.e., LOS errors below 0.4$\times$LOS did not count). The second study with best results (Li et al., 2019) (RMSE 0.88 and MAE 0.87) worked with a dataset with multiple covariates with 44-50% of missing values, whose management forcing a replacing value could have a high impact in LOS predictions.

In fact, some recent works (Kramer, 2017) question the capacity of computer-based predictive models based only on the condition of the patient in the first 24-48

hours after admission. Therefore, there is still a need to produce good, robust, and generic models to predict DTD in ICUs. A good DTD model must have a low average error and must be robust in front of ICU patient heterogeneity. In (Cuadrado et al., 2021), four methods to measure ICU patient heterogeneity respect to the DTD were described. Among them, the DTD confusion matrix determines the number of patients discharged in i days who are clinically indistinguishable from other patients who are discharged in j days ($j \neq i$). The results on 3,973 ICU patients with a mean stay of 8.56 days admitted to a tertiary hospital in Spain showed that, on average, 37% of the patients were clinically very similar to other patients who were discharged before, and 26% to patients who were discharged later (Cuadrado et al., 2021). Hence the difficulty of obtaining good DTD predictors and, perhaps, one of the reasons that DTD prediction has not received the same amount of attention as LOS prediction. Moreover, the only study we are aware of in this area obtained DTD predictive models with an average error above one day (Cuadrado and Riaño, 2021).

In this chapter we will use machine learning techniques with the public dataset described in Chapter 2, which has a population of highly heterogeneous ICU patients, with the aim to obtain a DTD predictive model with a mean absolute error below one day (i.e., almost perfect prediction).

5.2 Methods and Technologies

5.2.1 ICU Data Cohort

In this chapter, in order to consider the most difficult case of diversity in patients, we take the eICU Collaborative Research Database (Pollard, 2018), a public dataset that includes patients admitted in the ICUs of hospitals across the United States between years 2014 and 2015.

Only patients discharged alive were considered. Patients discharged on the same day of their admission were not considered. These inclusion criteria took into consideration other previous studies such as (Alghatani et al., 2021; Moitra et al., 2016; Hachesu et al., 2013). In eICU, this cohort encompasses 16,585 patients with a total of 84,032 rows corresponding to each of the days of treatment. From the data available, due to the need to have daily values, some covariates were discarded. The rest were selected based on previous studies (Cuadrado and Riaño, 2021).

Table 5.1 gives some descriptors of the data in the column All. In general, the average age is 63 years, with all patients between 18 and 90 years. A 45.61% are women. A 46.07% of patients are admitted in Medical-Surgical Intensive Care Unit (MSICU), 13.67% in Neurological Intensive Care Unit (NICU), 11.07% in Medical Intensive Care Unit (MICU) and 29.19% are under other ICU (e.g., CCU-CTICU, SICU, Cardiac ICU, etc.). Regarding duration, 26.30% remains one day, 11.08% two days,9.98% three days, 9.41% four days, 8.37% five days, 7.91% six days, 6.31% seven days and 20.65% remains eight days or more. The average length of stay is 5 days.

5.2.2 Patient Grouping

In order to achieve a good DTD prediction, we propose to divide the patients into three subgroups in order to reduce their heterogeneity. The subgroups proposed are based on the number of days staying in the ICU: short, medium and long stays.

Type	Freq.	name	All	Short stay	Medium stay	Long stay
N	E	Age	63.17,16.99,18.00,90.00	63.83,16.90,18.00,90.00	63.56,16.80,18.00,90.00	63.41,16.77,18.00,90.00
	D	Avg. T	36.93,0.52,32.58,40.01	36.85,0.48,32.64,39.89	36.9,0.50,32.64,39.89	36.92,0.51,32.64,40.01
		Min. T	36.48,0.66,25.00,39.89	36.44,0.62,25.00,39.30	36.46,0.64,25.00,39.30	36.48,0.66,25.00,39.39
		Max. T	37.37,0.67,32.60,43.61	37.25,0.62,33.00,42.00	37.32,0.65,33.00,43.61	37.36,0.67,33.00,43.61
		Avg. SpO2	96.73,2.13,80.39,100.00	96.61,2.09,80.04,100.00	96.66,2.12,80.14,100.00	96.70,2.13,81.49,100.00
		Min. SpO2	92.19,4.24,80.01,100.00	92.22,4.14,80.19,100.00	92.09,4.24,81.10,100.00	92.13,4.24,81.35,100.00
		Max. SpO2	99.15,1.43,81.13,100.00	99.01,1.53,80.22,100.00	99.10,1.46,80.92,100.00	99.14,1.44,81.94,100.00
		Avg. MAP	79.62,9.64,36.33,125.00	79.71,10.09,36.33,124.33	79.64,9.79,36.33,124.33	79.63,9.68,36.33,125.87
		Min. MAP	68.49,11.22,29.00,125.00	68.97,11.75,29.00,124.33	68.60,11.39,29.00,124.33	68.50,11.25,29.00,124.20
		Max. MAP	92.73,11.54,36.33,129.67	92.47,11.97,36.33,129.33	92.69,11.68,36.33,129.67	92.76,11.57,36.33,129.67
		Avg. HR	86.20,15.68,36.00,139.99	85.22,15.61,37.00,139.00	85.83,15.64,37.00,139.33	86.04,15.62,37.00,139.33
		Min. HR	73.44,15.34,36.00,137.40	73.00,15.40,36.00,138.13	73.21,15.38,36.00,139.12	73.33,15.32,36,141.83
		Max. HR	102.68,18.85,36.00,139.43	101.03,18.89,38.00,141.85	102.23,18.86,38.00,139.21	102.56,18.82,38.00,142.41
C	E	Gender (M/F) %	54.39/45.61	52.99/47.01	54.07/45.93	54.17/45.83
		UT (MSICU,NICU,MICU, Others) %	46.07 / 13.67 / 11.07 / 29.19	46.43 / 11.30 / 11.66 / 30.61	46.77 / 12.44 / 11.45 / 29.34	46.54 / 13.49 / 11.37 / 30.02
	D	MVI (0/1) %	91.75 / 8.25	94.01 / 5.99	92.42 / 7.58	91.97 / 8.03
		MVNI (0/1) %	98.15 / 1.84	97.75 / 2.24	97.93 / 2.07	98.06 / 1.94
S	D	Avg. GCS	12.52,3.31,3.00,15	13.39,2.78,3.00,15.00	12.85,3.11,3.00,15.00	12.61,3.25,3.00,15.00
		Min. GCS.	13.39,2.64,3.00,15.00	14.13,2.01,3.00,15.00	13.68,2.40,3.00,15.00	13.47,2.56,3.00,15.00
		Max. GCS.	12.82,2.95,3.00,15.00	13.65,2.36,3.00,15.00	13.14,2.73,3.00,15.00	12.91,2.88,3.00,15.00
		Avg. Pain	0.40,1.38,0.00,10.00	0.48,1.49,0.00,10.00	0.44,1.43,0.00,10.00	0.41,1.39,0.00,10.00
		Min. Pain	1.52,2.95,0.00,10.00	1.82,3.17,0.00,10.00	1.64,3.04,0.00,10.00	1.57,2.98,0.00,10.00
		Max. Pain	0.72,1.73,0.00,10.00	0.87,1.88,0.00,10.00	0.78,1.80,0.00,10.00	0.74,1.75,0.00,10.00

TABLE 5.1: Data set covariate description. For numerical parameters (N) we have Age, Temperature (T), SpO2, Heart rate (HR), and Mean Arterial Pressure (MAP) and each cell has its mean, stdev, minimum, and maximum values. For categorical data, each cell has the percentage values of Gender, Unit Type (UT), and mechanical ventilation invasive and non invasive (MVI/MVNI). For the two scale covariates, Glasgow Coma Score (GCS) and Pain Score, cells have its mean, minimum, and maximum values.

Short stays encompasses patients with a length of stay up to seven days. Medium stays encompasses patients with a length of stay up to fourteen days (which includes all patients from short stays and also patients with a length of stay between eight and fourteen days). Long stay encompasses patients with a length of stay up to twenty one days (which includes all patients from short and medium stays and also patients with a length of stay between fifteen and twenty one days). Patients with a discharge on the same day as the admission (DTD=1) were excluded for not being clinically relevant.

For the eICU database, short stays encompasses 8,799 patients, medium stays 11,432 patients and long stays 11,981 patients. The average LOS for short stays is 4.21 days, for medium stays is 5.54 days and for long stays is 6.07 days.

5.2.3 Measuring patients heterogeneity

The whole cohort was subject to an heterogeneity analysis in order to contextualize the quality of the DTD predictive models obtained. Root mean square resemblance was used to measure the pairwise clinical similarity between ICU patient conditions in different days. Patient conditions with a resemblance value of 99% or above were considered clinically equivalent. Following the heterogeneity measures introduced in Chapter 3, we calculated the risk of discharging patients before (or after) time as the average percentage of patients discharged in i days ($i = 1, ..., 7$) who have an equivalent clinical condition to other patients that were discarded from the ICU in more (or less) than i days.

Additionally, Davies-Bouldin (DB), Dunn (D), and average silhouette (S) were also calculated for the each group using Eqs. (5.1), (5.2), and (5.3), in order to quantify the degree of heterogeneity of the clinical conditions of the patients that are discharged in the same number of days.

The heterogeneity measures are based on the number of remaining days, RD_i, with $i = 1..\ell$, being ℓ the maximum number of days of stay in the analysed group. Let us denote as C_i the representative patient of the set RD_i (i.e. the one with greatest similarity to the rest of patients with the same number of remaining days); and let us define $m_i(d)$ as the average similarity sim of a patient description d to any other patient in the group. As similarity function, we have used the root mean squared resemblance. Then, the indices can be defined as follows:

$$DB = \frac{1}{\ell} \sum_{i=1}^{\ell} \max_{1 \leq j \neq i \leq \ell} \frac{2 - (m_i(C_i) + m_j(C_j)}{1 - sim(C_i, C_j)} \tag{5.1}$$

$$D = \frac{1 - \max_{1 \leq i < j \leq \ell} \{sim(C_i, C_j)\}}{1 - \min_{1 \leq i \leq \ell} \{m_i\}} \tag{5.2}$$

$$S = \frac{1}{\sum_{i=1}^{\ell} n_i} \cdot \sum_{i=1}^{\ell} \sum_{d \in RD_i} \frac{m_i(d) - m(d)}{1 - \min\{m(d), m_i(d)\}} \tag{5.3}$$

Davies-Bouldin index provides a positive value which is higher as heterogeneity increases (Cuadrado et al., 2021). Dunn index also takes positive values. The higher the value, the most heterogeneous are the groups. Silhouette values are in the range [-1, +1], with values below 0.25 considered to reflect a high heterogeneity in the data. The higher the degree of heterogeneity in a group, the more difficult is to obtain a high predictive capacity of the data models constructed with machine learning.

5.2.4 Constructing DTD and LOS Prediction Models

In order to construct DTD and LOS prediction models with Artificial Intelligence, we used three of the most successful machine learning algorithms. These algorithms were Random Forest (Breiman, 2001), XGBoost (Chen and Guestrin, 2016), and lightGBM (Ke et al., 2017). The prediction methods were implemented using the libraries sklearn, lightgbm and xgboost. All models were constructed using 10-fold cross validation method for testing and training with the covariate described in Table 5.1.

For LOS predictions, the methods were trained with the data of the patients in their first 24 hours in the ICU for all patients together and also for the three subgroups separately. However, for DTD prediction, models have been trained not only on the patient conditions after 24-48h ICU admission, but also on the daily clinical condition of patient along their full ICU stay.

Each day of stay in the database is treated as an independent patient with a set of parameters, therefore temporal patterns have not been captured.

For validation, we calculated the mean absolute error (MAE) and the root mean square error (RMSE) of these three algorithms to predict DTD for the whole data set and also for the three subgroups separately.

5.3 Results

This section presents the results of the different steps of the analysis done to build a model of prediction of DTD. Firstly, we begin with the definition of the groups of patients and its parameters. Secondly, we study the patient heterogeneity within those groups. Thirdly, we build the DTD prediction model as well as a LOW prediction model and we compare them and integrate them into the final hybrid model proposed in this thesis as solution to the problem of predicting the duration of the stay at ICU.

5.3.1 Groups of Patients

The resulting groups of patients are described in table 5.2. Column "N" is the total number of days of data available for the patients in the group. The average lentgh of stay is given in the last column. As expected, the amount of data is larger for the

Long Stay group as patients spend more time at the ICU.

The parameters of the patients for the groups based on the number of days at ICU are summarized in table 5.1. Column "All" of this table shows the mean, standard deviation, min and max values of the 13 numerical variables (type N), the 6 scales variables (type S) and the percentages of the most frequent values of the 4 categorical (type C) variables of the 84,032 days of treatment of the whole set of all the patients. These values allow cohort comparison of the different groups of ICU patients (Short, Medium and Long stays) with the whole set of days of treatment in the ICU, from a population point of view.

Group	Interval	Patients	N	Avg. LOS
Short stay	$LOS > 1$ and $LOS \leq 7$	8,799	37,014	4.21
Medium stay	$LOS > 1$ and $LOS \leq 14$	11,432	63,298	5.54
Long stay	$LOS > 1$ and $LOS \leq 21$	11,981	72,761	6.07

TABLE 5.2: Subgroups based on LOS. N is the total number of days for the patients in the group. Last column shows the average LOS.

From the descriptive table of the groups (table 5.1), we obtain the following results in terms of characterization of the different groups. With respect to the numerical variables, in general, all the groups display similar values to the ones observed for the whole population of patients. Scales covariates shows the same similarities with some exceptions in the average values of GCS_avg, GCS_min and GCS_max (higher than the rest of subgroups and column All).

Categorical covariates shows more variations between column All and subgroups. The three subgroups present lower number of patients admitted in Medical-Surgical Intensive Care Unit (MICU) and a higher number admitted in Neurological Intensive Care Unit (Neuro ICU) and Medical Intensive Care Unit (MICU) respect to column All. Mechanical ventilation invasive (MVI) also shows difference between all columns. Short stays present higher values (3 points above column All) while Medium and Long stays present lower values (2 and 4 points respectively below column All). Non-invasive mechanical ventilation (MVNI) also shows lower values for medium and long stays (2 points below column All in both cases).

5.3.2 Patients Heterogeneity

Heterogeneity with respect to the DTD is measured for the whole data set and also within each one of the three subgroups defined in table 5.2. The heterogeneity values obtained are included in table 5.3.

We obtained a high heterogeneity of treatment days with respect to DTD: 9.10% the days of treatment, patients are 99% similar to other patients that were discharged later, and 7.40 % of the days, patients are similar to other patients discharged earlier. A Davies-Bouldin value of 63.61 and a silohuette value of -0.26 confirms the high heterogeneity in ICU patients.

High heterogeneity is also present in the three subgroups. For example, 17.36% of the days, patients are similar to other patients discharged later in Short stays,

Metric	All	Short stay	Medium stay	Long stay
premature discharge	9.10 %	17.36 %	13.01 %	9.84 %
overdue discharge	7.40 %	12.43 %	10.00 %	9.81 %
Davies-Bouldin	63.61	49.93	67.07	81.34
Dunn	0.0007	0.0012	0.0011	0.001
Silhouette	-0.26	-0.02	-0.04	-0.07

TABLE 5.3: Data and subgroup heterogeneity values.

and 12.43% to patients discarged before. However, the heterogeneity in terms of the Davies-boulding, Dunn and Silhouette is smaller in Short Stays than in the rest of groups. Silhouette index is the one that finds more cohesion when working with three subgroups in comparison of having the all dataset as a whole. As expected, we obtained the highest heterogeneity in the Long stay group, as it includes any patient with a DTD below 22 days, with scores of 81.34 in Davis-Boulding and Silohuette of -0.07. The more compactness of Short and Medium groups is encouraging for finding appropriate models of DTD for these groups.

5.3.3 DTD Prediction

DTD predictive models for patients in ICU were obtained using the patient descriptions of 84,032 days of treatment in the eICU dataset and also for the three subgroups (see table 5.2). An hyperparameter optimization was performed for all algorithms. The parameter used for the prediction models are described in table 5.4.

Models	Parameters
Random Forest	n.estimators=170, max.depth=80, max.features=5, min.samples split=5
LightGBM	boosting type='gbdt', num.leaves=131, max.depth=-1, learning rate=0.1, n.estimators=100, subsample for bin=2000, min.split gain=0.0, min.child weight=0.001, min.child samples=20, subsample=1.0, subsample freq.=0, colsample bytree=1.0, importance type='split'
XGBoost	base score=0.3, booster='gbtree', n.estimators=208, max.depth=5, learning rate=0.1, n.jobs=1, objective='reg: squarederror', verbosity=1

TABLE 5.4: Parameters optimized for the prediction models.

A 10-fold-cross validation method was used to obtain the corresponding mean predictive errors RMSE and MAE in their i-th day before discharge (i.e. remaining days from $i = 2, ..., 21$) for all patients and for every subgroup, which are gathered in table 5.5, 5.6 and 5.7. In bold, we highlighted the best predictions for each day of stay (i.e. each row).

R.D.	All		Short stay		Medium stay		Long stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	1.15	0.9	**0.56**	0.4	0.96	0.74	1.23	0.97
3	0.98	0.68	**0.41**	0.31	0.77	0.54	1.04	0.75
4	0.85	0.58	**0.48**	0.37	0.65	0.47	0.89	0.62
5	0.79	0.59	0.71	0.57	**0.66**	0.51	0.8	0.59
6	0.86	0.66	0.98	0.84	0.83	0.64	**0.82**	0.62
7	1.06	0.82	1.25	1.12	1.11	0.87	**0.99**	0.76
8	1.32	1.02			1.44	1.18	**1.26**	0.96
9	1.66	1.33			1.81	1.54	**1.6**	1.27
10	2.02	1.69			2.16	1.89	**1.95**	1.62
11	2.4	2.05			2.53	2.24	**2.33**	1.99
12	2.82	2.46			2.91	2.6	**2.76**	2.4
13	3.22	2.84			3.33	2.99	**3.15**	2.78
14	3.55	3.15			3.54	3.2	**3.49**	3.09
15	3.95	3.51					**3.88**	3.45
16	4.37	3.91					**4.3**	3.84
17	4.7	4.22					**4.64**	4.16
18	5.21	4.69					**5.15**	4.63
19	5.75	5.19					**5.64**	5.09
20	6.23	5.62					**6.02**	5.46
21	6.42	5.83					**6.22**	5.64

TABLE 5.5: RMSE and MAE of the Random forest model for DTD prediction for each number of remaining days (R.D.)

R.D.	All		Short stay		Medium stay		Long stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	1.4	1.13	**0.53**	0.37	1.12	0.89	1.49	1.22
3	1.17	0.86	**0.44**	0.32	0.89	0.66	1.22	0.92
4	1.08	0.82	**0.5**	0.37	0.82	0.63	1.08	0.82
5	1.14	0.9	**0.68**	0.49	0.92	0.72	1.09	0.85
6	1.32	1.05	**0.88**	0.67	1.13	0.88	1.22	0.96
7	1.61	1.27	**1.12**	0.89	1.44	1.14	1.48	1.17
8	1.92	1.53			**1.77**	1.42	1.78	1.44
9	2.3	1.86			**2.15**	1.76	2.17	1.81
10	2.68	2.22			**2.51**	2.1	2.54	2.15
11	3.09	2.62			**2.87**	2.49	2.96	2.51
12	3.59	3.08			**3.25**	2.88	3.44	2.96
13	3.96	3.45			**3.72**	3.3	3.8	3.32
14	4.3	3.77			**3.87**	3.48	4.15	3.62
15	4.75	4.18					**4.58**	4.02
16	5.18	4.6					**5.01**	4.43
17	5.45	4.88					**5.29**	4.73
18	6.02	5.43					**5.9**	5.31
19	6.79	6.15					**6.57**	5.96
20	7.11	6.46					**6.88**	6.28
21	7.39	6.64					**7.08**	6.36

TABLE 5.6: RMSE and MAE of the LightGBM model for DTD prediction for each number of remaining days (R.D.)

R.D.	All		Short stay		Medium stay		Long stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	2.09	1.75	**1.64**	1.59	1.76	1.46	2.24	1.91
3	1.78	1.34	**0.98**	0.87	1.39	1.07	1.88	1.45
4	1.59	1.21	**0.5**	0.4	1.21	0.96	1.63	1.24
5	1.57	1.28	**0.76**	0.62	1.32	1.09	1.54	1.24
6	1.78	1.49	**1.42**	1.28	1.68	1.39	1.67	1.39
7	2.22	1.85	2.13	1.97	2.21	1.87	**2.07**	1.71
8	2.73	2.29			2.79	2.45	**2.58**	2.16
9	3.36	2.92			3.44	3.12	**3.23**	2.8
10	4.02	3.59			4.07	3.78	**3.88**	3.46
11	4.72	4.31			4.72	4.43	**4.58**	4.18
12	5.5	5.09			5.39	5.11	**5.36**	4.97
13	6.23	5.84			**6.07**	5.77	6.1	5.73
14	6.79	6.38			**6.13**	5.79	6.68	6.27
15	7.53	7.1					**7.36**	6.95
16	8.26	7.83					**8.1**	7.67
17	8.74	8.32					**8.58**	8.16
18	9.6	9.16					**9.48**	9.05
19	10.5	10.03					**10.29**	9.84
20	11.02	10.51					**10.6**	10.09
21	10.96	10.14					**10.54**	9.77

TABLE 5.7: RMSE and MAE of the XGBoost model for DTD prediction for each number of remaining days (R.D.)

Random Forest is the algorithm that obtains the best DTD predictors in all subgroups, with the exception of short stays (LOS up to 7 days) where MAE and RMSE values in lightGBM model outperform for days 2, 5, 6 and 7 days. XGBoost is the worst DTD modeller in all subgroups, producing models with a MAE and RMSE above 1 day for short stays, above 6 days in medium stays and above 10 days for long stays.

For the RMSE values, the average difference between Random Forest and LightGBM is 0.06 for short stays (with LightGBM outperforming above 5 days), 0.22 for medium stays and 0.30 for long stays (with Random Forest outperforming every day in both groups).

For the MAE values, the average difference between Random Forest and LightGBM is 0.09 for short stays (with LightGBM outperforming above 5 days), 0.19 for medium stays and 0.27 for long stays (with Random Forest outperforming every day in both groups). The whole data set (group All) shows MAE and RMSE values below 1 day between days 3 and 6 but there is always a MAE and RMSE value in some of the other subgroups that outperforms the results obtained when using the whole dataset.

Broadly speaking, results show that Random Forest and lightGBM are good producing ICU patient DTD predictors for patients with a length of stay up to 7 days which are optimal in the sense that their root mean square error (RMSE) and their mean absolute predictive error (MAE) are always below one day (with the exception of the seventh day). Random Forest is also good producing DTD predictions for patients with a length of stay up to 14 and up to 21 days with RMSE and MAE values below one day in the last 7 days before discharge. This last week is the most crucial in the planning of at the ICUs, because gives the opportunity to know in advance, with quite small error, the date of discharge and therefore make a proper scheduling of beds, personnel and other resources, in addition to having the possibility of advancing the planning of the transfer of the patient to any other hospital unit.

Model	Measure	All		Short stay		Medium stay		Long stay	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Random Forest	avg.	3.0	2.6	0.7	0.6	1.7	1.5	2.9	2.5
	std.dev.	1.9	1.8	0.3	0.3	1.0	1.0	1.9	1.8
	avg. 2-6	0.9	0.7	0.6	0.5	0.8	0.6	1.0	0.7
	st.dev. 2-6	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2
LightGBM	avg.	3.6	3.1	0.7	0.5	2.0	1.7	3.5	3.0
	std.dev.	2.2	2.0	0.3	0.2	1.1	1.0	2.1	1.9
	avg. 2-6	1.2	1.0	0.6	0.4	1.0	0.8	1.2	1.0
	st.dev. 2-6	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.2
XGBoost	avg.	5.5	5.1	1.2	1.1	3.2	2.9	5.4	5.0
	std.dev.	3.4	3.4	0.6	0.6	1.8	1.8	3.3	3.3
	avg. 2-6	1.8	1.4	1.1	1.0	1.5	1.2	1.8	1.4
	st.dev. 2-6	0.2	0.2	0.5	0.5	0.2	0.2	0.3	0.3

TABLE 5.8: RMSE and MAE average and standard deviation for DTD prediction for each model

The average and deviation of the MAE for each of the 4 groups is given in table 5.8 taking into account all the days of each group (blue), and also considering only the remaining days below 7 in each group (green). We can see that Random Forest gives the lowest errors. The best average of the MAE is 2.5 with a deviation of 1.8 for the All and Long groups for the Random Forest model, but it is much smaller

for the Short stay (mean of 0.6, stdev of 0.3 and the Medium stay (mean of 1.5, stdev of 1.0. This indicates that using different prediction models for each case would lead to better results in general. Considering only the last week of stay at ICU, the predictions made are much better. The mean MAE obtained for Random Forest is between 0.5-0.7 with a maximum standard deviation of only 0.2.

5.3.4 DTD versus LOS Prediction

In order to have a global perspective of the predictive capacity of DTD models for ICU patients, we also constructed models for LOS prediction by training the Random Forest, the XGBoost, and the lightGBM algorithms using only the data of the patients in their first 24 hours in the ICU. The average errors (RMSE and MAE) after 10-fold cross validation are shown in tables 5.9, 5.10 and 5.11. Each row corresponds to the number of remaining days of stay at ICU.

R.D.	All		Short stay		Medium stay		Long stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	1.2	0.9	**0.95**	0.84	1.45	1.26	1.63	1.4
3	1.16	0.8	**0.59**	0.49	1.18	0.98	1.43	1.16
4	0.99	0.69	**0.29**	0.21	0.87	0.65	1.14	0.84
5	0.99	0.72	**0.41**	0.31	0.67	0.47	1.01	0.69
6	0.95	0.73	0.79	0.68	**0.54**	0.42	0.8	0.57
7	1.13	0.87	1.25	1.12	**0.78**	0.6	0.8	0.62
8	1.31	0.98			1.06	0.83	**0.98**	0.75
9	1.63	1.25			1.44	1.2	**1.27**	0.98
10	1.95	1.57			1.84	1.6	**1.59**	1.29
11	2.27	1.88			2.25	1.99	**1.95**	1.63
12	2.77	2.37			2.71	2.43	**2.44**	2.11
13	3.27	2.84			3.25	2.93	**2.94**	2.58
14	3.46	3.04			3.53	3.19	**3.16**	2.79
15	4.04	3.58					**3.73**	3.33
16	4.36	3.9					**4.09**	3.67
17	4.7	4.22					**4.41**	3.99
18	5.03	4.52					**4.83**	4.36
19	5.77	5.22					**5.52**	5
20	6.25	5.68					**5.89**	5.35
21	6.47	5.85					**6.19**	5.62

TABLE 5.9: RMSE and MAE of the Random forest model for LOS prediction for each number of remaining days (R.D.)

The best results, which are marked in bold, are for the Random Forest algorithm. We can see that Random Forest method generally outperforms the other methods in all the days. Short stay subgroup results shows that LightGBM slightly outperforms for the last day of stay in ICU while Random Forest shows better results for every day in the other subgroups.

In general, results for DTD prediction models are better than LOS prediction models when applied in the last days of stay.

R.D.	All		Short stay		Medium stay		Long stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	1.45	1.14	**0.94**	0.8	1.62	1.44	1.87	1.66
3	1.36	0.99	**0.62**	0.46	1.27	1.03	1.53	1.24
4	1.24	0.94	**0.39**	0.29	0.96	0.71	1.23	0.91
5	1.3	1.02	**0.5**	0.37	0.8	0.59	1.09	0.78
6	1.38	1.1	0.81	0.64	**0.77**	0.6	0.99	0.75
7	1.67	1.32	1.12	0.89	**1.06**	0.83	1.14	0.9
8	1.94	1.5			**1.36**	1.06	1.36	1.07
9	2.28	1.78			1.72	1.38	**1.68**	1.32
10	2.65	2.13			2.11	1.76	**2.02**	1.6
11	2.99	2.45			2.49	2.14	**2.36**	1.92
12	3.63	3.08			3.02	2.64	**2.92**	2.44
13	4.28	3.69			3.69	3.28	**3.54**	3.02
14	4.29	3.69			3.83	3.39	**3.5**	2.95
15	5.09	4.48					**4.25**	3.67
16	5.37	4.77					**4.55**	3.94
17	5.57	5.01					**4.75**	4.16
18	6	5.41					**5.24**	4.61
19	7.11	6.51					**6.15**	5.5
20	7.59	7					**6.48**	5.82
21	7.88	7.1					**6.88**	6.12

TABLE 5.10: RMSE and MAE of the LightGBM model for LOS prediction for each number of remaining days (R.D.)

Differences between DTD and LOS prediction models become more evident when they are used to predict the discharge times of patients in their i-th day before discharge ($i = 2, ..., 21$), separately.

Figures 5.1, 5.2 and 5.3 show that DTD models have better performance (with lower values in both RMSE and MAE) when the stay arrives at its end. In particular, DTD models should be used in the last 1-3.5 days for short stays, 1-5 days for medium stays and 1-6 days for long stays. We observed that these values correspond to the average length of stay for every subgroup (i.e., 4.21 for shot stays, 5.54 for medium stays and 6.07 for long stays).

Since DTD models have been trained not only on the patient conditions after 24-48h ICU admission, but also on the daily clinical condition of patient along their full ICU stay, the amount of data to train the model is larger than in LOS models, resulting in better results for DTD models at the end of stay.

R.D.	All		Short stay		Medium stay		Long stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	2.04	1.69	**1.64**	1.59	2.52	2.42	2.84	2.7
3	1.92	1.43	**0.98**	0.87	1.95	1.76	2.37	2.1
4	1.67	1.28	**0.5**	0.4	1.39	1.1	1.83	1.46
5	1.68	1.38	**0.76**	0.62	1.07	0.83	1.52	1.13
6	1.76	1.47	1.42	1.28	**1.02**	0.85	1.26	1.02
7	2.19	1.81	2.13	1.97	1.51	1.26	**1.49**	1.26
8	2.6	2.09			2.03	1.71	**1.9**	1.55
9	3.14	2.58			2.68	2.36	**2.45**	2.05
10	3.78	3.28			3.36	3.1	**3.1**	2.75
11	4.3	3.81			4	3.75	**3.68**	3.31
12	5.2	4.74			4.8	4.55	**4.54**	4.2
13	6.15	5.69			5.76	5.48	**5.47**	5.11
14	6.3	5.85			6.05	5.77	**5.71**	5.34
15	7.49	7.07					**6.77**	6.42
16	8.01	7.56					**7.35**	6.96
17	8.29	7.83					**7.68**	7.31
18	8.92	8.48					**8.44**	8.05
19	10.33	9.92					**9.66**	9.29
20	10.98	10.53					**10.17**	9.81
21	11.09	10.34					**10.12**	9.47

TABLE 5.11: RMSE and MAE of the XGBoost model for LOS prediction for each number of remaining days (R.D.)

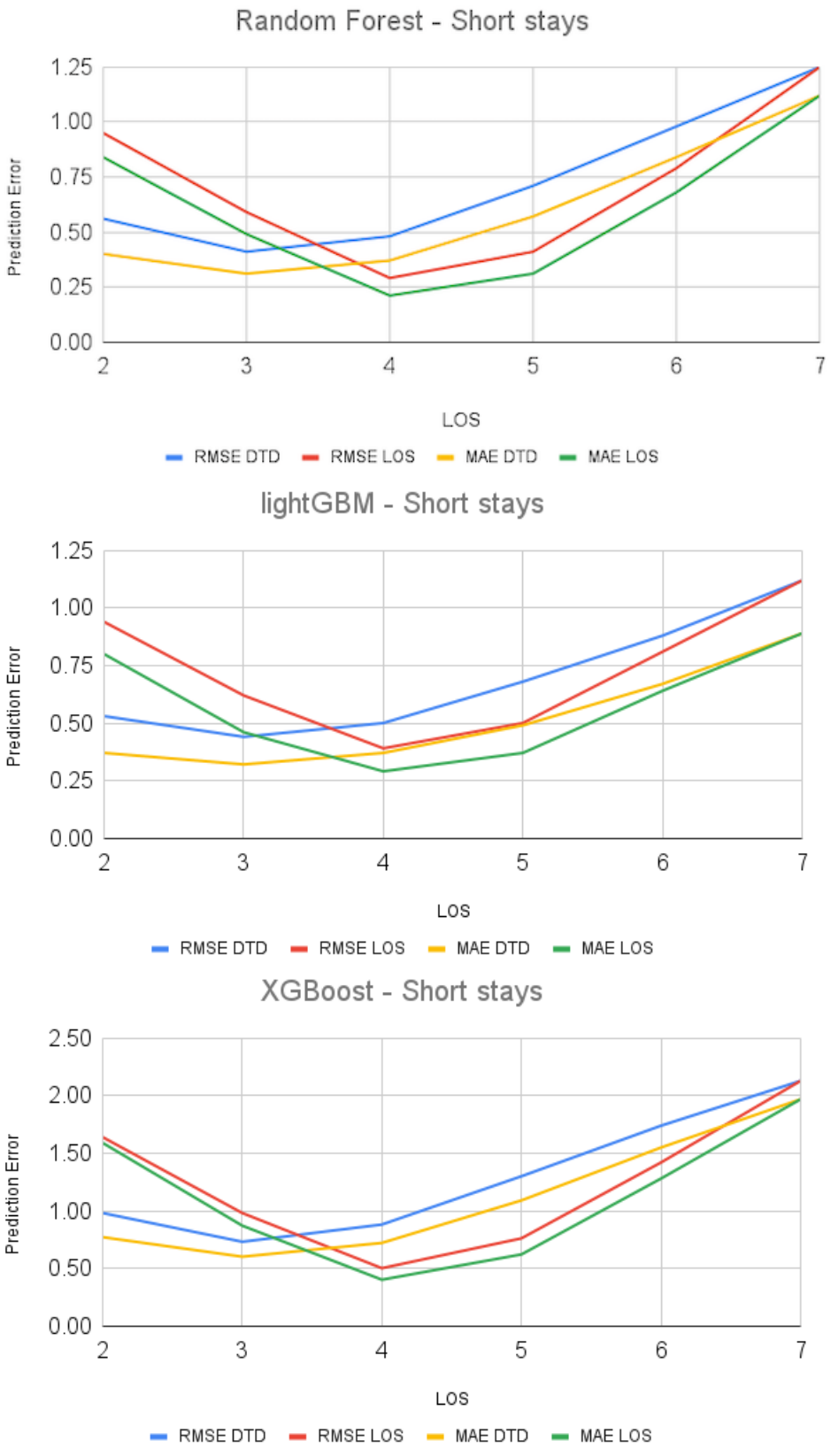


FIGURE 5.1: Prediction errors for Short Stays

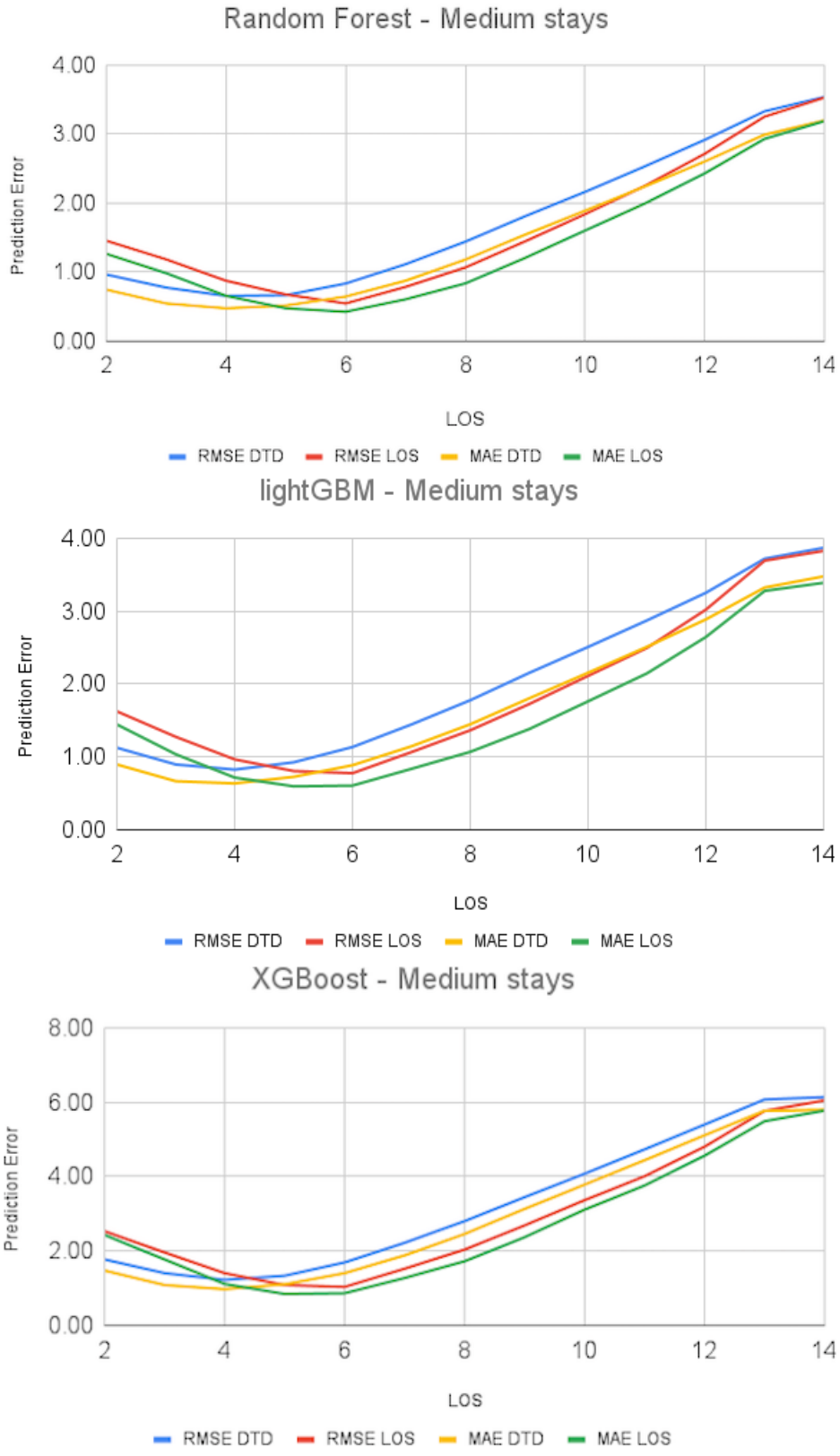


FIGURE 5.2: Prediction errors for Medium Stays

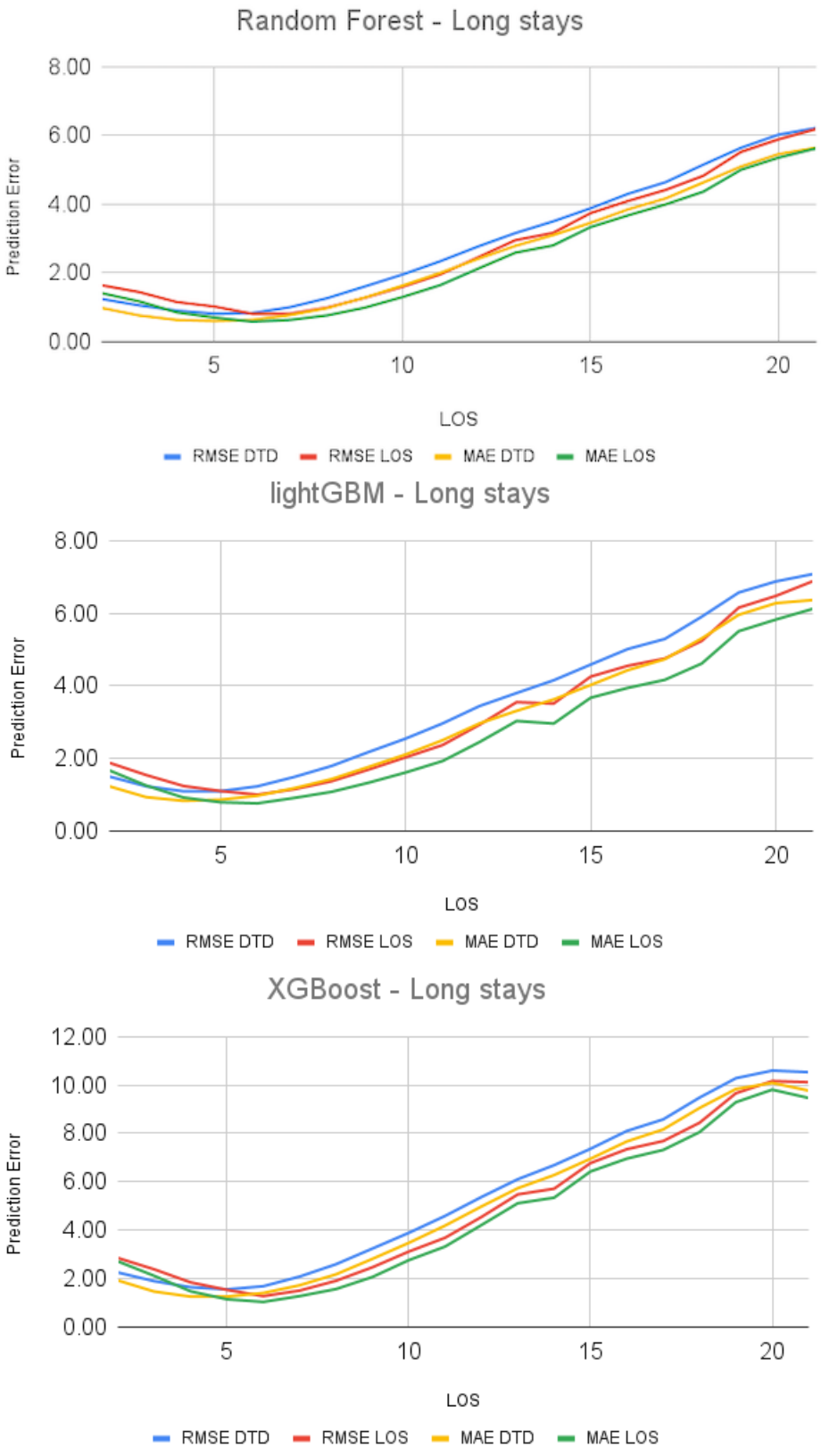


FIGURE 5.3: Prediction errors for Long Stays

5.4 Date of discharge prediction system

The most spread approach of predicting patient stays in ICU with the use of Artificial Intelligence models that consider only the data captured in the first 24-48h after admission does not give enough accuracy. Without undermining the importance of this type of prediction, it does not seem justified that, as the clinical condition of the patients evolve, the new information about the patient's state is not taken into account to dynamically predict the DTD of the corresponding patients.

In this chapter, we have empirically proven that the dynamic prediction models of DTD improve the quality of static prediction of LOS for the last 4-6 days of stay at ICU. We have also seen that building different prediction models for 3 different subgroups of patients improves the RMSE and MAE. The creation of three subgroups based on the number of days of stay of an ICU patient (short, medium and long) allow us to determinate when the DTD model prediction obtains better results than LOS model prediction since DTD model is trained with daily clinical condition of patient along their full ICU stay.

Therefore, a hybrid model could drastically improve the results in prediction models for ICU patients using LOS models at the beginning of stay and DTD models at the end of the stay.

Consequently, considering that we have built the LOW and DTD models with Random Forest, the proposal we make consists of the following steps:

- When a patient is addmitted at ICU, calculate LOS with the All model.
- if $LOS \leq 7$ days use the LOS-Short prediction model the first days, and the DTD-Short model in the last 4 days until discharge.
- if $LOS > 7 and \leq 14$ days use the LOS-Medium prediction model the first days, and the DTD-Medium model in the last 5 days until discharge.
- if $LOS > 14$ days (Long stay) use the LOS-Long prediction model the first days, and the DTD-Long model in the last 6 days until discharge.

This combination of models should give the date of discharge to the ICU personnel with enough advance to make a proper planning of the end of the patient's stay at this critical hospital unit, since the Mean Absolute Error is below 1 day in most of the predictions, and below 0.5 in the last days of stay.

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

Chapter 6

Conclusions and future work

6.1 Conclusions

This doctoral thesis has addressed the complex problem of calculating the duration of the stay of any patient admitted at an Intensive Care Unit.

While there have been many studies conducted on estimating the length of stay, there has been a lack of articles focusing on days to discharge in a dynamic way. This lack of references was an additional handicap of this thesis, since we had no security about the potential of dynamic DTD calculation to predict ICU patient discharge. Addressing this gap in the literature, makes this work even more interesting for the medical community as well as for the point of view of an Artificial Intelligence doctorate.

With the thesis aim to obtain a prediction for ICU patient discharge below 1 day, this manuscript has described the work done during these years of doctorate study.

Chapter 2 made a review of related works and then focused on analyzing the various parameters of patients in the intensive care unit. This analysis serves as a foundation for evaluating the potential accuracy of predicting when patients will be discharged. Additionally, the conclusive outcomes of our research make it possible to apply our findings to both studies on length of stay and days to discharge. The chapter ended presenting a preliminary analysis to predict the days to discharge in a small hospital. Although the study had limitations because of the small number of patients, the initial findings emphasized the significant value of the investigation. The difficulty in developing predictors for days to discharge is the reason why there is a lack of articles addressing this topic.

Chapter 3 highlights the complexity of predicting DTD for heterogeneous patients in ICUs. To improve DTD prediction accuracy, four methods and their corresponding measures were proposed for analyzing patient heterogeneity and tested on patients admitted in Hospital Joan XXIII. Results confirmed high heterogeneity of conditions of the patients within the same DTD group and also confusion between patients with distant DTDs. The study involved all the types of ICU patients making the analysis of heterogeneity very relevant and improves the understanding of the heterogeneity in the ICU and the development of a more accurate DTD predictor.

Chapter 4 addressed the issue of heterogeneity's complexity by studying the formation of patient groups based on some derived variables, as second contribution. By categorizing patients according to their DTD biomarkers and identifying potential phenotypes, results shows that the complexity remains high, although it seems

that the creation of distinct groups of patients may help to build more specific prediction models.

Chapter 5 brought together the previously discussed ideas and implemented them in a different hospital that had a larger number of patients. The main contribution of the thesis is give in this chapter. The first step was to analyze the differences among patients in terms of how long they stayed in the intensive care unit, which led to the identification of three separate groups based on their duration of stay. Afterward, predictive models were used to forecast the patient outcomes. The findings indicated that models focusing on length of stay produced superior results for patients with shorter stays, whereas models concentrating on days to discharge were more effective for patients with longer stays. Therefore, a hybrid model combining days to discharge and length of stay is proposed in order to improve the results in prediction of the date of discharge.

6.2 Future work

Given the lack of articles focusing on days to discharge in the literature, it is crucial to continue exploring its potential contributions to predicting ICU patient discharge. This can involve conducting more extensive studies with larger sample sizes to validate and refine the initial findings. As mentioned in Chapter 4, the study faced limitations due to the small number of patients in the initial analysis for the private database. Conducting further research with a larger sample size would enhance the reliability and statistical power of the predictive models. It would also enable a more comprehensive examination of patient subgroups, leading to more accurate predictions for days to discharge. The study carried out with the public database has made it possible to observe that a greater number of patients improve the prediction results, so the expansion of the private database will favor the prediction results in hospitals.

The thesis also highlighted the challenges encountered in creating predictors for days to discharge, which might explain the lack of research in this domain. To address this, future efforts can concentrate on enhancing the accuracy of days to discharge prediction models by integrating a wider range of patient data, exploring innovative variables or biomarkers, and harnessing advanced machine learning methods. With the constant evolution of healthcare, new biomarkers and medical data are likely to emerge. By incorporating these novel factors into the predictive models, their precision and predictive abilities could be significantly improved. Consequently, ongoing research should focus on incorporating up-to-date and relevant biomarkers into the analysis to enhance the predictive capacity of days to discharge models.

Future work should focus on translating the research findings into practical applications in clinical settings, aiming for real-world implementation and impact. Collaborating with healthcare professionals is essential to ensure that the predictive models are effectively integrated into the existing healthcare workflow. By involving them, the research can be focused to address specific clinical needs and challenges. The development of decision support tools or guidelines based on the

predictive models could empower healthcare providers to make well-informed decisions regarding patient discharge from the ICU. Such tools could consider various factors, including patient vitals, medical history, and response to treatment, to generate timely and accurate predictions. Ultimately, the successful deployment of these tools would not only enhance patient care by facilitating optimized discharge decisions but also improve resource allocation within the ICU, ensuring that critical resources are efficiently used.

Finally, the future work of this thesis should focus on addressing the consideration of temporal treatment in the database to significantly improve the predictive models. Predicting the discharge of a patient from the ICU implies a deep understanding of the evolution of their health status over time. Therefore, it is crucial to adequately capture and analyze the temporal sequences of medical events and treatments that the patient underwent during their stay in the ICU. Incorporating this temporal dimension into the data will allow the models to detect more precise patterns and trends, which will translate into more accurate predictions about the appropriate time to discharge a patient. The focus on time sequence analysis could also help identify which specific treatments have had a significant impact on patient recovery, which would be of great use to medical staff in making informed decisions about hospital discharge. Consequently, this focus on temporal treatment in the database represents a promising opportunity to improve the efficacy and relevance of predictive models in the ICU setting and, ultimately, to improve the quality of care and outcomes for critical patients.

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

Bibliography

Abd-ElrazekaAhmed, M. A. et al. (2021). "Predicting length of stay in hospitals intensive care unit using general admission features". In: *Ain Shams Engineering Journal* 12.4, pp. 3691–3702. DOI: 10.1016/j.asej.2021.02.018.

Alghatani, Khalid et al. (2021). "Predicting Intensive Care Unit Length of Stay and Mortality Using Patient Vital Signs: Machine Learning Model Development and Validation". In: *JMIR Medical Informatics* 9.5, e21347. DOI: 10.2196/21347.

Artis, Kathryn A. et al. (2019). "Data Omission by Physician Trainees on ICU Rounds". In: *Critical Care Medicine* 47.3, pp. 403–409.

Awad, Aya, Mohamed Bader–El–Den, and James McNicholas (2017). "Patient length of stay and mortality prediction: A survey". In: *Health Services Management Research* 30.2, pp. 105–120.

Ayyoubzadeh, Seyed Mohammad et al.Gusmão (2020). "A study of factors related to patients' length of stay using data mining techniques in a general hospital in southern Iran". In: 8, p. 9. DOI: 10.1007/s13755-020-0099-8.

Bai, J. et al. (2018). "Operations research in intensive care unit management: a literature review". In: *Health Care Manag. Sci.* 21.1, pp. 1–24.

Bing-Hua, Y. U. (2014). "Delayed admission to intensive care unit for critically surgical patients is associated with increased mortality". In: *The American Journal of Surgery* 208.2, pp. 268–274.

Breiman, L (2001). "Random Forests". In: *Machine Learning* 45, 5–32. DOI: 10.1023/A:1010933404324.

Caetano, Nuno, Raul M. S. Laureano, and Paulo Cortez (2014). "A Data-driven Approach to Predict Hospital Length of Stay - A Portuguese Case Study". In: *Proceedings of the 16th International Conference on Enterprise Information Systems* 3, pp. 407–414. DOI: 10.5220/0004892204070414.

Caruana, Edward Joseph et al. (2015). "Longitudinal studies". In: *J Thorac Dis.* 7.11, E537–E540.

Chandrashekar G., Sahin F. (2014). "A survey on feature selection methods". In: *Computers & Electrical Engineering* 40, pp. 16–28.

Chen, T. and C. Guestrin (2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. DOI: 10.1145/2939672.2939785.

Chrusciel, J. et al. (2021). "The prediction of hospital length of stay using unstructured data". In: *BMC Med Inform Decis Mak* 21.1, p. 351. DOI: 10.1186/s12911-021-01722-4.

Cuadrado, D. and D. Riaño (2021). "ICU Days-to-Discharge Analysis with Machine Learning Technology". In: *Artificial Intelligence in Medicine*. Springer LNAI 12721, pp. 103–113. DOI: 10.1007/978-3-030-77211-6_11.

Cuadrado, David et al. (2019). "Pursuing Optimal Prediction of Discharge Time in ICUs with Machine Learning Methods". In: 11526, pp. 150–154.

Cuadrado, David et al. (2021). "Methods and measures to quantify ICU patient heterogeneity". In: *Journal of Biomedical Informatics* 117, p. 103768. DOI: 10.1016/j.jbi.2021.103768.

Davies, David L. and Donald W. Bouldin (1979). "A Cluster Separation Measure". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1.2, 224–227.

Droogh, Joep M et al. (2015). "Transferring the critically ill patient: are we there yet?" In: *Critical Care* 19.1, p. 62.

Dunn, J. C. (1974). "Well-Separated Clusters and Optimal Fuzzy Partitions". In: *Journal of Cybernetics* 4.1, 95–104.

Gholipour, C. et al. (2015). "Using an Artificial Neural Networks (ANNs) Model for Prediction of Intensive Care Unit (ICU) Outcome and Length of Stay at Hospital in Traumatic Patients". In: *J Clin of Diagn* 9.4, pp. 19–23. DOI: 10.7860/JCDR/2015/.

Gusmão Vicente F., Polito Lomar F. et al. (2004). "Can the experienced ICU physician predict ICU length of stay and outcome better than less experienced colleagues?" In: *Intensive Care Med* 30, pp. 655–659. DOI: 10.1007/s00134-003-2139-7.

Hachesu, P. R. et al. (2013). "Use of data mining techniques to determine and predict length of stay of cardiac patients". In: *Healthc Inform Res.* 19.2, pp. 121–129. DOI: 10.4258/hir.2013.19.2.121.

Halkidi, Maria, Yannis Batistakis, and Michalis Vazirgiannis (2001). "On Clustering Validation Techniques". In: *Journal of Intelligent Information Systems* 17, pp. 107–145.

Helal, S. (2016). "Subgroup Discovery Algorithms: A Survey and Empirical Evaluation". In: *J Comput Sci Technol* 31, pp. 561–576. DOI: 10.1007/s11390-016-1647-1.

Herrera F., Carmona C.J. González P. et al. (2011). "An overview on subgroup discovery: foundations and applications". In: *Knowl Inf Syst* 29, pp. 495–525. DOI: 10.1007/s10115-010-0356-2.

Houthooft, R. et al. (2015). "Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores". In: *Artificial Intelligence in Medicine* 63.3, pp. 191–207. DOI: 10.1016/j.artmed.2014.12.009.

Huang, Zhengxing et al. (2013). "Length of stay prediction for clinical treatment process using temporal similarit". In: *Expert Systems with Applications* 40.16, 6330 – 6339. DOI: 10.1016/j.eswa.2013.05.066.

J. Dasta T. Mclaughlin, S. Mody C. Piech (2005). "Daily cost of an intensive care unit day: The contribution of mechanical ventilation". In: *Crit Care Med* 33, pp. 1266–1271.

Jacobs, K. et al. (2022). "Variability drivers of treatment costs in hospitals: A systematic review". In: *Health Policy* 126.2, pp. 75–86. DOI: 10.1016/j.healthpol.2021.12.004.

Jović, A., K. Brkić, and N. Bogunović (2015). "A review of feature selection methods with applications". In: pp. 1200–1205. DOI: 10.1109/MIPRO.2015.7160458.

Kaufman, Leonard and Peter J. Rousseeuw (1990). *Finding groups in data. An introduction to cluster analysis*. John Wiley and Sons.

Ke, G. et al. (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: vol. 30, 3149–3157. DOI: 10.5555/3294996.3295074.

Kramer, A. A. (2017). "Are ICU Length of Stay Predictions Worthwhile?" In: *Critical Care Medicine* 45.2, pp. 379–380. DOI: 10.1097/CCM.0000000000002111.

Kramer A.A., Zimmerman J.E. (2010). "A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay". In: *BMC Medical Informatics and Decision Making* 10, p. 27. DOI: 10.1186/1472-6947-10-27.

Kross, E. K. et al. (2014). "Differences in End-of-Life Care in the ICU Across Patients Cared for by Medicine, Surgery, Neurology, and Neurosurgery Physicians". In: *Chest* 145.2, pp. 313–321. DOI: 10.1378/chest.13-1351.

LaFaro, R. J. et al. (2015). "Neural Network Prediction of ICU Length of Stay Following Cardiac Surgery Based on Pre-Incision Variables". In: *PLoS ONE* 10.12. DOI: 10.1371/journal.pone.0145395.

Lambden, S et al. (2019). "The SOFA score-development, utility and challenges of accurate assessment in clinical trials". In: *Crit Care* 23.1.

Li, C. et al. (2019). "Prediction of Length of Stay on the Intensive Care Unit Based on Least Absolute Shrinkage and Selection Operator". In: *IEEE Access* 7, pp. 110710–110721. DOI: 10.1109/ACCESS.2019.2934166.

Livieris I.E., Kotsilieris T. Dimopoulos I. Pintelas P. (2018). "Decision Support Software for Forecasting Patient's Length of Stay". In: *Algorithms* 11, p. 199. DOI: 10.1186/1472-6947-10-27.

Ma, X. et al. (2020). "Length of stay prediction for ICU patients using individualized single classification algorithm". In: *Comput Methods Programs Biomed.* 186:105224. DOI: 10.1016/j.cmpb.2019.105224.

Marshall, John C. et al. (2017). "What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine". In: *Journal of Critical Care* 37, pp. 270–276. ISSN: 0883-9441. DOI: https://doi.org/10.1016/j.jcrc.2016.07.015. URL: https://www.sciencedirect.com/science/article/pii/S0883944116302404.

McKenzie, M. S. et al. (2015). "An Observational Study of Decision Making by Medical Intensivists". In: *Critical care medicine* 43.8, 1660–1668. DOI: 10.1097/CCM.0000000000001084.

Miranda, D R et al. (2003). "Therapeutic Intervention Scoring System. Nursing Activities Score". In: *Crit Care Med* 331, pp. 374–382.

Moitra, V. K. et al. (2016). "Relationship Between ICU Length of Stay and Long-Term Mortality for Elderly ICU Survivors". In: *Critical care medicine* 44.4, 655–662. DOI: 10.1097/CCM.0000000000001480.

Mollaei, N. et al. (2021). "Length of Stay Prediction in Acute Intensive Care Unit in Cardiothoracic Surgery Patients". In: *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, pp. 1–5. DOI: 10.1109/ICBSII51839.2021.9445145.

Moran, J. and P. Solomon (2012). "A review of statistical estimators for risk-adjusted length of stay: analysis of the Australian and new Zealand intensive care adult patient data-base, 2008–2009". In: *BMC Medical Research Methodology* 12.1, p. 68. DOI: 10.1186/1471-2288-12-68.

Moreno, R et al. (1999). "The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. Working Group on Sepsis related Problems of the ESICM". In: *Intensive Care Med.* 25.7, pp. 686–96.

Muhlestein, W. et al. (2019). "Predicting Inpatient Length of Stay After Brain Tumor Surgery: Developing Machine Learning Ensembles to Improve Predictive Performance". In: *Neurosurgery* 85.3, pp. 384–393. DOI: 10.1093/neuros/nyy343.

Nassar, Antonio Paulo Jr and Pedro Caruso (2016). "ICU physicians are unable to accurately predict length of stay at admission: a prospective study". In: *Int J Qual Health Care* 28, pp. 99–103. DOI: 10.1093/intqhc/mzv112.

Pakhira, Malay K., Sanghamitra Bandyopadhyay, and Ujjwal Maulik (2004). "Validity index for crisp and fuzzy clusters". In: *Pattern Recognition* 37.3, pp. 487–501.

Peres, I. T. et al. (2021). "Predição do tempo de permanência em unidades de terapia intensiva: uma revisão concisa". In: *Rev Bras Ter Intensiva* 33.2, pp. 183–187. DOI: 10.5935/0103-507X.20210025.

Pollard T., Johnson A. Raffa J. et al (2018). "The eICU Collaborative Research Database, a freely available multi-center database for critical care research". In: *Scientific Data* 5 (1). DOI: 10.1038/sdata.2018.178.

Roca-Biosca, A. et al. (2015). "Validation of EMINA and EVARUCI scales for assessing the risk of developing pressure ulcers in critical patients (in Spanish)". In: *Enfermería Intensiva* 26, pp. 15–23.

Rossi, C. et al. (2006). "Variable costs of ICU patients: a multicenter prospective study". In: *Intensive Care Med* 32, 545–552. DOI: 10.1007/s00134-006-0080-2.

Rousseeuw, Peter J. (1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65.

Rowan, M. et al. (2007). "The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors". In: *Artificial Intelligence in Medicine* 40.3, pp. 211–221. DOI: 10.1016/j.artmed.2007.04.005.

Ruyssinck, Joeri et al. (2016). "Random Survival Forests for Predicting the Bed Occupancy in the Intensive Care Unit". In: *Computational and Mathematical Methods in Medicine*, pp. 1–7.

Selleck MJ, Senthil M Wall NR. (2017). "Making Meaningful Clinical Use of Biomarkers". In: *Biomark Insights* 23. DOI: 10.1177/1177271917715236.

Su, L. et al. (2021). "Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis 3.0 by Machine Learning Models". In: *Frontiers in Medicine* 8. DOI: 10.3389/fmed.2021.664966.

Tan S.S., Bakker J. et al. (2012). "Direct Cost Analysis of Intensive Care Unit Stay in Four European Countries: Applying a Standardized Costing Methodology". In: *Value in Health* 15, pp. 81–86. DOI: 10.1016/j.jval.2011.09.007.

Temple, Michael W., Christoph U. Lehmann, and Daniel Fabbri (2015). "Predicting Discharge Dates From the NICU Using Progress Note Data". In: *Pediatrics* 136.2, e395–e405.

Valentin, A., P. Ferdinande, and ESICM Working Group on Quality Improvement (2011). "Recommendations on basic requirements for intensive care units: structural and organizational aspects". In: *Intensive Care Med* 37, pp. 1575–1587. DOI: 10.1007/s00134-011-2300-7.

Van Houdenhoven, M. et al. (2007). "Optimizing intensive care capacity using individual length-of-stay prediction models". In: *Critical Care* 11.2, p. 42. DOI: 10.7860/JCDR/2015/9467.5828.

Verburg, I. W. M. et al. (2014). "Comparison of Regression Methods for Modeling Intensive Care Length of Stay". In: *PLoS ONE* 9.10. DOI: 10.1371/journal.pone.0109684.

Verburg, Ilona Willempje Maria et al. (2017). "Which Models Can I Use to Predict Adult ICU Length of Stay? A Systematic Review". In: *Critical Care Medicine* 45.2, e222–e231.

Vranas, KC et al. (2017). "Identifying Distinct Subgroups of ICU Patients: A Machine Learning Approach". In: *Crit Care Med.* 45.10, pp. 1607–1615.

Wu, J. et al. (2021). "Predicting Prolonged Length of ICU Stay through Machine Learning". In: *Diagnostics* 11.12, p. 2242. DOI: 10.3390/diagnostics11122242.

UNIVERSITAT ROVIRA I VIRGILI
MACHINE LEARNING METHODS FOR PREDICTING DAYS TO DISCHARGE IN INTENSIVE CARE UNITS PATIENTS
David Cuadrado Gómez

UNIVERSITAT
ROVIRA i VIRGILI