

Real time automatic risk prediction in ICU patients treated with ECMO

Filipe Miguel de Oliveira Ribeiro

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisor(s): Prof. João Miguel Raposo Sanches
Prof. Ana Catarina Fidalgo Barata
Dr. João Miguel Ferreira Ribeiro

Examination Committee

Chairperson: Prof. José Eduardo Charters Ribeiro da Cunha Sanguino

Supervisor: Prof. Ana Catarina Fidalgo Barata

Member of Committee: Maria Margarida Campos da Silveira

November 2024

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I would like to express my heartfelt gratitude to all those who contributed to the success of this project.

First and foremost, I extend my sincere thanks to Professor João Sanches and Catarina Barata for their decision to take part in this challenging endeavor. Your unwavering support, insightful guidance, and invaluable feedback throughout this journey have been instrumental in my personal and professional growth. The lessons I've learned under your mentorship have pushed me to test my limits and will undoubtedly continue to shape my future, no matter what path I take.

Secondly, I want to express my gratefulness to the Intensive Care Department of Hospital de Santa Maria for their willingness to collaborate and provide access to essential resources, which were crucial for carrying out this work. I owe a special thanks to the dedicated team in this department, whose openness and collaboration were inspiring and ultimately decisive in making this project a reality.

A particular acknowledgment goes to Doctor João Ribeiro, the Director of the Intensive Care Department at Hospital de Santa Maria and Coordinator of the ECMO Reference Center at ULS Santa Maria, and my co-supervisor, whose continuous support and drive, thoughtful guidance, and dedication were essential to the success of this project.

Finally, I would like to express my profound appreciation to Inês Mendes, my family, and friends for their unwavering encouragement and support throughout this journey. Your belief in me through the highs and lows of this journey was a constant source of motivation and inspiration, for which I am deeply grateful.

Abstract

Extracorporeal Membrane Oxygenation (ECMO) is a therapeutic intervention employed in intensive care medicine that provides life support to critically ill patients whose lungs and heart function are severely compromised and proved decisive during the COVID-19 pandemic. ECMO support relies on a complex network of technologically advanced systems that monitor the patient's clinical condition throughout hospitalization, generating multidimensional and multidomain datasets. Assessing ECMO datasets is challenging, presenting a promising opportunity for applying Machine Learning (ML) techniques.

Using 81 labeled Multivariate Time Series (MTSs) from patients with COVID-19 pneumonia treated with ECMO support, a Support Vector Machine (SVM) with varying kernel functions and a Random Forest model was trained to distinguish between clinical deterioration and improvement throughout hospitalization. The Random Forest model achieved the best predictive performance and calibration. Ultimately, feature importance analysis provided insights into its predictions, enhancing interpretability and practical applicability.

The Random Forest model with 100 trees was then used to compute a risk score, which provided a real-time estimate of the risk of clinical deterioration throughout hospitalization under ECMO support for each patient. The score effectively anticipated periods of clinical decline and improvement, achieving an Area Under the Receiver Operating Characteristics (AUROC) curve of 0.9176, 0.8944, and 0.8556 for time intervals preceding these periods of 4, 8, and 12 hours, respectively.

This study demonstrated the pivotal role of ML systems in supporting physicians to assess complex patient cohorts, uncovering insights that might have otherwise remained undetectable.

Keywords: Intensive Care Unit, Extracorporeal Membrane Oxygenation, Machine Learning, Time Series, Risk Prediction, Risk Score

Resumo

Oxigenação por Membrana Extracorporeal (ECMO) é uma intervenção terapêutica usada em medicina intensiva que fornece suporte vital a doentes críticos com disfunção cardíaca e pulmonar, tendo sido decisiva durante a pandemia COVID-19. O suporte de ECMO depende de uma rede complexa de sistemas tecnologicamente avançados que monitorizam a condição clínica dos doentes ao longo do internamento, produzindo dados multidimensionais e multidisciplinares. Analisar estes dados representa um desafio, mas também uma oportunidade promissora para aplicação de técnicas de Machine Learning (ML).

81 séries temporais multivariadas etiquetadas relativas a doentes com pneumonia COVID-19 tratados com ECMO foram usadas para treinar modelos Support Vector Machine (SVM), usando diferentes kernels, e Random Forest para discernir entre deterioração e melhoria clínica. O modelo Random Forest obteve o melhor desempenho preditivo e calibração. A análise da relevância das variáveis ofereceu ainda informações valiosas relativas às previsões deste modelo, aumentando a sua interpretabilidade e aplicabilidade prática.

O modelo Random Forest com 100 árvores foi utilizado para calcular um score de risco, que fornece uma estimativa em tempo real do risco de deterioração clínica durante o internamento sob aplicação de ECMO. O score antecipou eficazmente períodos de deterioração e melhoria clínica, alcançando áreas sob a curva ROC de 0,9176, 0,8944 e 0,8556 para intervalos de tempo precedentes destes períodos de 4, 8 e 12 horas, respetivamente.

Este estudo demonstra o suporte crucial que sistemas de ML podem oferecer no que respeita à avaliação médica de doentes complexos, revelando informações que de outra forma poderiam permanecer ocultas.

Palavras-chave: Unidade de Cuidados Intensivos, Oxigenação por Membrana Extracorporeal, Aprendizagem Automática, Séries Temporais, Previsão de Risco, Score de Risco

Contents

1	Introduction	1
1.1	Background and Motivation	1
1.2	Research Objectives	3
1.3	Project Contributions	4
2	Background	5
2.1	Intensive Care Medicine	5
2.1.1	Historical Development and Emerging Trends in Intensive Care Medicine	5
2.1.2	Intelligent Intensive Care Units: Addressing the Challenges and Limitation of Con- ventional Intensive Care Medicine	6
2.2	Extracorporeal Membrane Oxygenation	10
2.2.1	Overcoming the Limitations of Mechanical Ventilation: The Rise of Extracorporeal Membrane Oxygenation	11
2.2.2	Extracorporeal Membrane Oxygenation: Historical Context and General Perspective	11
2.2.3	Foundations of Extracorporeal Membrane Oxygenation: Lung Physiology and Oxy- gen Metabolism	12
2.2.4	Physiology of Extracorporeal Membrane Oxygenation	15
2.3	Chapter Conclusions	17
3	Related Work	19
3.1	Machine Learning Risk Prediction in Healthcare	19
3.1.1	Introduction to Machine Learning Risk Prediction in Healthcare	20
3.1.2	Machine Learning Risk Prediction in Healthcare: Problem Framing	21
3.1.3	Machine Learning Risk Prediction in Healthcare: Study Review and Methodologi- cal Insights	23
3.2	Machine Learning Risk Prediction in Extracorporeal Membrane Oxygenation	25
3.3	Chapter Conclusions	28
4	Data	29
4.1	Criteria-Based Patient Selection	29
4.2	Data Acquisition	31

4.3	Data Preparation	33
4.3.1	Candidate Predictor Variables	33
4.3.2	Outcome	35
4.4	Exploratory Data Analysis	36
4.4.1	Patient Cohort: Descriptive Statistics and General Characteristics	37
4.4.2	Data Quality Assessment	38
5	Research Methodology	40
5.1	Phase 1: Development of a Machine Learning Model for Classification of Clinical Deterioration and Improvement	40
5.1.1	Data Preprocessing	41
5.1.2	Feature Engineering	45
5.1.3	Model Training and Hyperparameter Optimization	49
5.1.4	Model Evaluation and Selection	50
5.1.5	Model Interpretability	52
5.2	Phase 2: Development of a Machine Learning Risk Score for Real-Time Assessment of the Patient's Clinical Condition	53
5.2.1	Development and Optimization	53
5.2.2	Evaluation and Selection	55
6	Results and Discussion	58
6.1	Phase 1: Development of a Machine Learning Model for Classification of Clinical Deterioration and Improvement	58
6.1.1	Data Preprocessing	58
6.1.2	Feature Engineering	62
6.1.3	Model Training and Hyperparameter Optimization	64
6.1.4	Model Evaluation and Selection	65
6.1.5	Model Interpretability	68
6.2	Phase 2: Development of a Machine Learning Risk Score for Real-Time Assessment of the Patient's Clinical Condition	71
6.2.1	Development and Optimization	72
6.2.2	Evaluation and Selection	73
7	Conclusions and Future Work	77
	Bibliography	80

List of Tables

2.1	Functions inherent to the systems within intelligent ICUs as described by Mao et al. [12].	8
2.2	Adverse effects frequently observed in patients undergoing mechanical ventilation [14].	11
3.1	Summary and key findings from several studies on the development of ML risk prediction models in healthcare.	24
3.2	Summary and critical analysis of two studies on ML-based risk prediction models for ECMO support.	27
4.1	High-level clinical categories and corresponding candidate predictor variables. *Categorical variables. **Binary presence variables (0 and 1 indicating absence and presence, respectively).	34
4.2	Temporal categories and corresponding candidate predictor variables.	34
4.3	Description of each possible value of the established ternary outcome variable.	35
4.4	Results obtained by applying the data labeling strategy to the multiple patients' MTSs.	36
4.5	Patient cohort descriptive statistics. *Excludes Sars-CoV-2 infection. SD: standard deviation.	37
5.1	Minimum and maximum accepted value for each of the continuous numerical variables (static and dynamic). *beats per minute; **rotations per minute; *** cycles per minute.	42
5.2	Available categories for selected categorical variables (excluding binary presence variables). *The categories for variable Obesity Degree were established based on guidelines and directives provided by the U.S. Centers For Disease Control and Prevention [42].	42
6.1	Results obtained after transforming each patient's MTS into multiple labeled windows, each comprising three consecutive time points	62
6.2	Performance metrics for the ML models. * Polynomial; ** Random Forest.	66
6.3	Collaborative evaluation of the 7 most relevant features determining model predictions for clinical deterioration (outcome class -1), based on physicians' feedback. The table highlights clinical insights and contradictions revealed through SHAP analysis.	70

List of Figures

2.1	Overview of the comprehensive nature of data typically available for each patient hospitalized within the ICU.	7
2.2	Architecture of an optimal intelligent ICU.	9
2.3	Illustration of the ECMO circuit applied to a patient (adapted from [15]).	13
3.1	Implementation of the sliding window framing structure, with "ONSET" reflecting the time of occurrence of singular clinical event in time.	22
4.1	Flowchart illustrating the criteria-based patient selection process and indicating the number of patients included and excluded after each step.	30
4.2	Data table containing information for a unique patient referring to a single day of hospitalization under ECMO support. These tables frequently contained outliers, as observed, for instance, in the measurements of FiO_2 at 00h and 08h (30.0 would be the valid measurement instead of 0.30).	32
4.3	Data stored within the <i>Picis</i> system for a single patient undergoing ECMO support.	32
4.4	Visualization depicting the data labeling process: the blue region denotes both past and present information available to physicians for analysis, guiding outcome assignment decisions for the current instant; meanwhile, the gray region reflects forthcoming data lying beyond current knowledge or access, thus inaccessible to physicians at the current instant. *For patients with a length of stay under ECMO support of less than 12 days, data labeling was conducted for the entire hospitalization period under these conditions.	36
4.5	Count of patients per variable for which no valid measurements are available.	39
4.6	Count of patients per variable for which invalid measurements (either missing or incorrect) surpass 50%.	39
5.1	Flowchart illustrating the data preprocessing pipeline. * Represented as a single block since the multiple patient-specific correlation matrices are aggregated and assessed jointly (this process is described in detail below). MTS: Multivariate Time Series.	41

5.2	Flowchart illustrating the feature engineering pipeline. *Represented as a single block since windows selected from different patients are combined and then assigned to the training and validation subsets according to predefined criteria (further details discussed below). **Represented as a single block since the feature extraction pipeline is initially fitted onto the training subset and subsequently used to transform (i.e., extract features from) both the training and validation subsets (further details discussed below). MTS: Multivariate Time Series.	45
5.3	Modified sliding window method applied to convert the multiple patient Multivariate Time Series (MTSs) into sets of labeled instances (8-hour windows). *For patients with a length of stay under ECMO support of less than 12 days, the sliding window spanned the entire period of hospitalization under ECMO support.	46
5.4	Structure of the confusion matrix generated to evaluate the ML models' performance in distinguishing between clinical deterioration (outcome class -1) and improvement (outcome class +1).	50
6.1	Incidence of outliers and missing values across all variables within the multiple patient MTSs.	60
6.2	Correlation matrix obtained after aggregating the multiple patient-specific correlation matrices computed considering the Spearman's rank correlation coefficient.	61
6.3	Confusion matrix for the Random Forest model with 100 trees, computed using the validation dataset.	67
6.4	Reliability diagrams (or calibration curves) for the uncalibrated (blue curve) and calibrated (green curve) versions of the Random Forest model with 100 trees.	68
6.5	Display of the top 10 features ranked by their mean absolute SHAP values, with features listed in descending order of importance for predictions. Each feature plot illustrates SHAP values across observations (x-axis), color-coded to represent low (blue) to high (red) feature values. The absolute value indicates the contribution extent, with the sign denoting positivity or negativity.	69
6.6	Risk score for patient 44 of the study cohort.	74
6.7	Risk score for patient 76 of the study cohort.	74
6.8	Risk score for patient 53 of the study cohort.	75
6.9	Risk score for patient 57 of the study cohort.	75
6.10	ROC curves obtained for the risk score considering time windows of 4, 8, and 12 hours preceding periods of clinical deterioration and improvement.	76

Acronyms

AI Artificial Intelligence

AKI Acute Kidney Injury

ANN Artificial Neural Network

ARDS Acute Respiratory Distress Syndrome

AUROC Area Under the Receiver Operating Characteristics

DL Deep Learning

ECMO Extracorporeal Membrane Oxygenation

EHR Electronic Health Record

EWS Early Warning Score

FN False Negative

FP False Positive

FPR False Positive Rate

ICU Intensive Care Unit

IoT Internet of Things

LightGBM Light Gradient Boosting

MEWS Modified Early Warning Score

ML Machine Learning

mNEWS2 modified National Early Warning Score

MTS Multivariate Time Series

NEWS National Early Warning Score

PEEP Positive End-Expiratory Pressure

RBF Radial Basis Function

ROC Receiver Operating Characteristics

SHAP SHapley Additive exPlanations

SVM Support Vector Machine

TN True Negative

TP True Positive

TPR True Positive Rate

VA-ECMO Venoarterial Extracorporeal Membrane Oxygenation

VV-ECMO Venovenous Extracorporeal Membrane Oxygenation

XGBoost Extreme Gradient Boosting

Chapter 1

Introduction

This chapter introduces the dissertation "Real time automatic risk prediction in ICU patients treated with ECMO", providing a comprehensive overview of the study's background and inherent motivation (section 1.1), outlining the established research objectives (section 1.2), and highlighting its contributions to the scientific community (section 1.3).

1.1 Background and Motivation

According to the American College of Physicians, intensive care medicine constitutes a highly comprehensive medical science field encompassing the diagnosis and treatment of patients with diverse clinical conditions that represent the extreme of human disease severity, posing an imminent or foreseeable threat to their survival [1].

The technological progress that has characterized recent years has markedly contributed to the advancement of intensive care medicine, fostering the development of advanced medical procedures and transformative therapeutic interventions, thereby enhancing the ability to manage complex clinical scenarios. The growing integration of advanced technologies within the Intensive Care Unit (ICU) has enabled continuous monitoring of various clinical variables and physiological parameters, providing physicians with extensive information and promoting a more in-depth assessment of the patient's clinical condition. Consequently, the volume of data generated in modern Intensive Care Units (ICUs) is exceptionally high, offering critical insights that enhance medical evaluation and support informed decision-making. Additionally, the availability of this extensive data serves as a foundation for establishing collaboration protocols between multidisciplinary research institutions, laying the groundwork for comprehensive studies and research projects.

Most recently, the COVID-19 pandemic further underscored the pivotal role of intensive care medicine in hospital settings. The unprecedented global health crisis imposed immense strain on healthcare services, with successive waves of infection leading to numerous patients experiencing critical respiratory failure and other severe complications necessitating ICU treatment. These patients often presented with near-fatal clinical conditions that severely compromised vital functions, posing an imminent threat

to their survival. The highly challenging and uncertain clinical scenarios observed during the pandemic necessitated the employment of advanced medical techniques, including life support interventions (e.g., respiratory, cardiac, liver, or kidney function support) to stabilize patients and create the conditions for partial or complete recovery.

Mechanical ventilation, which exclusively performs air exchange, was the primary intervention for critically ill COVID-19 patients. However, in cases where it failed to ensure adequate oxygen delivery to all tissues and organs, Extracorporeal Membrane Oxygenation (ECMO) emerged as pivotal for providing critical support in near-fatal conditions, ultimately improving patient survival rates.

Moreover, ECMO has proven to be a crucial intervention not only for critically ill COVID-19 patients but also for patients suffering from a broad spectrum of life-threatening clinical conditions (e.g., Sepsis) requiring intensive support [2]. Due to its complexity, invasive nature, inherent risks, and limited availability, ECMO is typically employed as a last resort when other therapeutic interventions have failed to improve the patient's clinical condition. ECMO facilitates gas exchange directly within the blood by artificially promoting the oxygenation and removal of carbon dioxide (CO₂) within the blood drained from the venous systemic circulation and ensures appropriate blood circulation, effectively simulating lungs and heart functions. By addressing the limitations of more conventional therapeutic interventions (e.g., mechanical ventilation), ECMO plays a crucial role in modern intensive care medicine, offering a life-saving option for patients with severe cardiac or respiratory failure for whom alternative pathways have no positive effect.

The clinical severity and volatility of patients under ECMO support, regardless of underlying health conditions, make this patient cohort one of the most complex and challenging to manage, requiring continuous evaluation by expert physicians to support high-value clinical decisions. Since ECMO support frequently involves the simultaneous application of supplementary interventions that support patient treatment, physicians must be adept at processing and analyzing the large volumes of data generated through an extensive network of monitoring systems. This comprehensive data encompasses a wide range of parameters (e.g., physiological and laboratory), providing physicians with real-time information that enhances medical evaluation, thus improving the detection of adverse events and enabling engagement in early, potentially preventive, therapeutic interventions that may be critical to patient safety.

This high-volume data flow is typically transferred to dedicated information systems through automated processes. However, this information is often recorded in an unstructured and decentralized manner, compromising accessibility and usability. Furthermore, the processing and analysis of this data remains excessively labor-intensive, time-consuming, and costly, demanding a high staff-to-patient ratio, which is often difficult to achieve in critical care settings.

These limitations, coupled with the scarcity of expert physicians capable of assessing patients under ECMO support, who are often required to operate in highly demanding environments and make critical decisions under conditions that push the limits of human cognition [3], necessitate the development of innovative mechanisms. These innovations should foster the combination of the increasing availability of medical knowledge with a digital, technological, and data-driven perspective. In this context, Artificial Intelligence (AI), particularly Machine Learning (ML), emerges as a promising avenue due to its ability

to process, analyze, and learn from large volumes of multidomain and multidimensional data.

Physicians frequently risk experiencing information overload when managing patients under ECMO support, resulting from the high-volume data flow generated through the intricate network of monitoring systems and supporting technology. This significantly compromises their ability to accurately assess these patients' clinical condition throughout hospitalization, resulting in substantial uncertainty and an inability to anticipate relevant clinical dynamics, such as deterioration and improvement. As described by Komorowski [4], ML techniques can compensate for these limitations, including the shortage of ECMO specialists. Systems that integrate this technology can process comprehensive datasets and extract informative and interpretable insights that could enable less experienced physicians to assess more complex patient cohorts, such as those referring to ECMO support. This optimization of resource availability ensures the possibility of allocating expert physicians to other crucial tasks. Additionally, these systems can help overcome one of the most pressing challenges in intensive care medicine, extreme individual heterogeneity, by providing more patient-centered care, thus improving patient outcomes.

Overall, the exceedingly complexity and critical nature of intensive care medicine, combined with the data-rich environment of ICUs, makes this one of the most conducive fields for developing and implementing ML systems. Integration of ML in critical care settings is increasingly prominent, especially in managing intricate interventions, such as ECMO, enabling more robust management of data produced continuously. As a result, future research should address the pressing need for integrating advanced systems into clinical practice to improve patient outcomes and optimize critical care resource utilization.

1.2 Research Objectives

This study emerged from the firm belief that the increasing digitization of healthcare and the ongoing development of technologically advanced equipment and data-driven systems can profoundly transform clinical practice. This trend holds particular promise for intensive care medicine, a data-intensive field focused on managing highly complex and critical patient cohorts (e.g., patients under ECMO support).

Addressing these challenges involved the establishment of a formal collaboration protocol (submitted and approved by the Academic Center Ethics Committee, **Code Reference 175/22**) with one of the most prestigious medical institutions in Portugal: the ECMO Referral Center, Intensive Care Department at the University Hospital Santa Maria, Local Health Unit Santa Maria, Lisbon. This partnership provided a solid foundation for conducting a comprehensive study to tackle the challenges outlined in section 1.1. Ultimately, the study focused on the development of a ML tool through a methodology including both conventional and innovative steps, designed based on the following research objectives:

1. **Develop a data acquisition and registration infrastructure** aimed at ensuring a greater degree of automation of data management processes. Ideally, this infrastructure should ensure increased centralization, ultimately facilitating data accessibility and enhancing storage security.
2. **Build a robust data processing framework** to transform unstructured datasets of patients under ECMO support. This framework should automatically convert raw data into a structured format

that accurately preserves its multidomain, multidimensional, and time-sensitive nature while circumventing inconsistencies typically observed in medical datasets.

3. **Design and implement an innovative data labeling strategy** that ensures the datasets' appropriate annotation, enabling the application of conventional supervised learning algorithms for predictive modeling and analysis.
4. **Implement a robust ML methodology** to create a predictive model capable of accurately distinguishing between periods of clinical deterioration and improvement throughout hospitalization under ECMO support.
5. **Develop a risk score**, which integrates the outputs of the ML model, that delivers accurate and interpretable real-time estimates of the risk of clinical deterioration throughout hospitalization under ECMO support for each patient. Ideally, the risk score should enhance the early detection of relevant clinical dynamics (e.g., deterioration and improvement) by providing physicians with valuable information (potentially previously inaccessible or undetectable) that supports medical evaluation and enables the adoption of a more personalized approach regarding this therapeutic intervention, ultimately increasing the likelihood of positive patient outcomes.

This study began with a comprehensive analysis of the study's background, providing technical details that facilitate an in-depth understanding of intensive care medicine and ECMO (chapter 2), followed by a literature review (chapter 3) that, despite scarce on the specific topic of ML applications in the context of ECMO, provided valuable methodological insights that informed several decisions across this work, essential to meet the outlined research objectives. Then, a robust data pipeline covering data selection, acquisition, and processing (including outcome definition and preliminary filtering) was established (chapter 4), aiming to achieve research objectives 1, 2, and 3. Chapter 5 describes the two-phase methodology employed, aiming to achieve research objectives 4 and 5, while chapter 6 presents and discusses the results obtained from implementing the methods described across chapters 4 and 5. Ultimately, chapter 7 summarizes the main findings within each chapter, laying the groundwork for future research in this critical area.

1.3 Project Contributions

This study was designed to make a meaningful and recognizable contribution to the scientific community by addressing the complex challenge of managing patients under ECMO support through a ML-based methodology. Its innovative approach and the promising results achieved were formally recognized through the acceptance and presentation of project papers at the **7ª Conferência Anual da RedeSAÚDE** [5] and the **46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society** [6].

This recognition further underscores the study's impact, with its findings to be published in the **IEEE Xplore digital library** (publication date pending).

Chapter 2

Background

This chapter provides an extensive overview of the study's background, offering a detailed review of its key elements. As outlined in chapter 1, this study emerged from a research initiative based on the collaboration protocol established with the ECMO Referral Center, Intensive Care Department at the University Hospital Santa Maria, Local Health Unit Santa Maria, Lisbon. This partnership granted access to high-value resources (both theoretical and practical) that supported the development of this chapter.

Building on the motivations and research objectives outlined in chapter 1, this chapter delves deeper into the study's context. Therefore, a comprehensive overview of intensive care medicine, highlighting current trends and evolving practices, is initially provided (section 2.1). Following this section, a thorough technical analysis of ECMO is presented, providing a comprehensive understanding of this complex yet essential therapeutic intervention (section 2.2).

2.1 Intensive Care Medicine

Introductory Note: some of the content provided in this section derived from discussions with the physicians that supported or participated in this study.

2.1.1 Historical Development and Emerging Trends in Intensive Care Medicine

The concept of intensive care (or critical care) medicine arose from the Copenhagen poliomyelitis epidemic of 1952 when numerous patients developed respiratory paralysis, requiring artificial ventilation [7]. The significant influx of critically ill patients overwhelmed ventilator facilities existing at the infectious disease hospital, introducing the need to adopt non-conventional solutions, including relying on medical students to provide manual positive pressure by repeatedly squeezing a rubber bag attached to a tracheostomy tube around the clock [8]. Given these patients' critical condition and the inherent need for continuous monitoring and surveillance, Bjorn Ibsen, an anesthetist who played a pivotal role in identifying initial treatment options such as positive pressure ventilation, proposed treating these patients in specially designed medical wards. This approach ensured the constant presence of at least one nurse

per patient. As a result, the specialty of intensive care was born, laying the groundwork for developing modern, technologically advanced ICUs [9].

Nowadays, the ICU represents a highly complex, intense, and challenging environment that relies on the convergence of multiple disciplines, cooperation between physicians with different specializations, and advanced technology for quality care provision. The ICU provides critical care to patients with severe, potentially life-threatening clinical conditions, often requiring temporary assistance and, in some cases, replacement of multiple organ systems [7]. The ICU's ability to provide this specialized care underscores its pivotal role within hospital settings.

The intricate nature of intensive care medicine, which combines technologically advanced procedures and highly complex therapeutic interventions, requires ICU physicians to develop a wide range of technical and practical competencies. However, their field of action is not exclusively restricted to these aspects, expanding to other domains such as ethics. Ethical issues are prevalent in the ICU because patients are often unconscious and legally incompetent, requiring ICU physicians to be proficient in addressing complex dilemmas, including ensuring effective communication, counseling patients' families, and making end-of-life decisions. As a result, ICU physicians bear significant responsibility, making it essential to develop systems and methods that streamline their practice and support them in managing the wide range of clinical demands and responsibilities they face.

Most recently, notable technological advancements have enabled the optimization of several processes within the ICU, creating favorable conditions for the increasing adoption of patient-centered practices, which take into account patient-specific characteristics. This innovative, increasingly frequent approach allows the optimization of therapeutic interventions, facilitating the minimization of underlying compromising side effects and ultimately improving patient outcomes.

Increasing technological innovation and integration within ICUs have driven significant improvements in the quality and effectiveness of care, enabling more precise responses to critical challenges in intensive care medicine, such as patient heterogeneity. These advancements are increasingly viewed as part of a broader transition toward intelligent ICUs, often presented as a promising solution to the most significant limitations currently facing ICUs. The following section delves into intelligent ICUs and their contribution to the general improvement of intensive care medicine.

2.1.2 Intelligent Intensive Care Units: Addressing the Challenges and Limitation of Conventional Intensive Care Medicine

The transition to intelligent ICUs, driven by continuous technological advancements, stems from the growing recognition of the challenges and limitations inherent in conventional intensive care medicine and is becoming increasingly urgent. ICUs face the immense responsibility of managing critically ill patients, a task made particularly difficult by the extreme volatility that characterizes these individuals. With rising life expectancy and an aging population, the demand for intensive care is escalating, intensifying challenges such as the ongoing shortage of ICU physicians. These professionals are consistently subjected to heavy workloads and exposed to high occupational pressure, resulting in frequent states

of exhaustion and reduced efficacy. This structural problem urgently requires strategies that enhance working conditions within ICUs, minimizing physician fatigue and preventing negative outcomes for both staff and patients.

Most recently, the COVID-19 pandemic revealed the extent of these challenges, causing an unprecedented influx of critically ill patients that stretched ICUs to their limits. This crisis underscored the necessity of innovative solutions, forcing ICUs to adopt unconventional strategies, which ultimately highlighted the adaptability and resilience of critical care [10]. This experience further demonstrated the urgent need for the transition to intelligent ICUs, which offer the potential to mitigate these structural problems, support overwhelmed medical staff, and enhance patient care.

Intelligent ICUs integrate advanced systems that enable the optimization, automation, and standardization of various processes, thereby improving their overall efficacy while simultaneously reducing susceptibility to errors caused by human factors (e.g., by mitigating the impact of subjectivity inherent in medical reasoning). These improvements, primarily driven by the precise integration of sophisticated equipment, facilitate the application of complex therapeutic interventions (e.g., mechanical ventilation and ECMO support). The enhanced monitoring capacity associated with this equipment leads to the production of increasingly larger volumes of data, adding to the already available critical information for each patient. Figure 2.1 provides an overview of the comprehensive nature of data typically available for each patient hospitalized within the ICU.

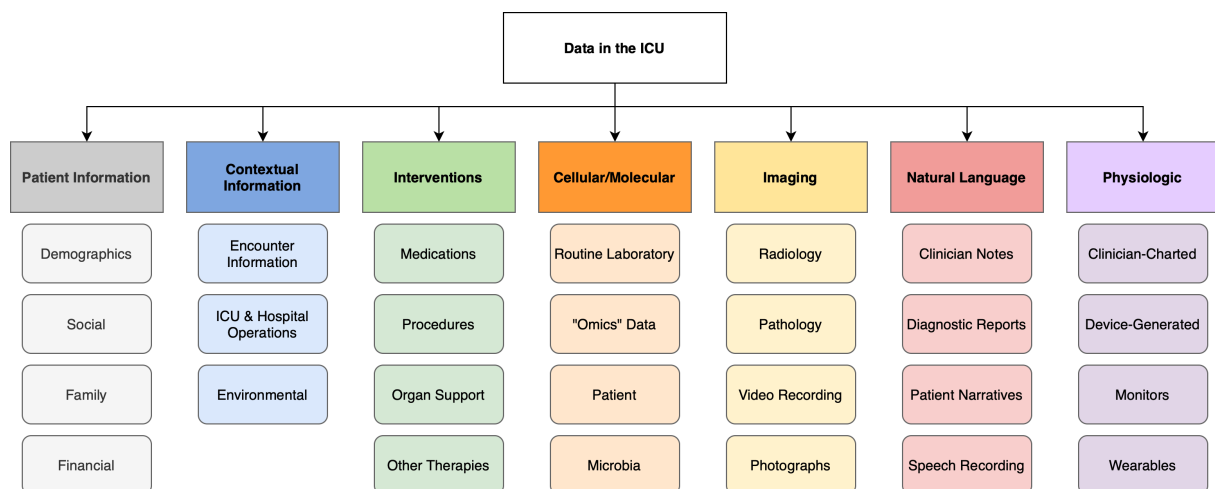


Figure 2.1: Overview of the comprehensive nature of data typically available for each patient hospitalized within the ICU.

According to Zelechower et al. [11], effectively leveraging the increasing volume of available data relies on several critical factors, including internal information exchange through efficient communication systems and the maintenance of accessible and well-structured databases. Intelligent ICUs distinguish themselves by incorporating robust data management frameworks that optimize data acquisition (e.g., through real-time monitoring equipment), transmission (e.g., through seamless communication channels for data exchange between distinct elements of the ICU), registration, and analysis. In fact, Mao et al. [12] suggests that intelligent ICUs should encompass the following systems: **monitoring system**,

communication system, **analytical system**, and **alarm system**, whose primary functions are detailed in table 2.1.

Table 2.1: Functions inherent to the systems within intelligent ICUs as described by Mao et al. [12].

System	Functions
Monitoring System	Encompasses equipment that continuously collects patient-specific data (e.g., physiological parameters and clinical variables), gathers medical equipment operational status indicators, and more. This uninterrupted stream of data enables real-time monitoring of critical ICU components, offering a comprehensive overview of the ICU.
Communication System	Facilitates the transmission and integration of data collected through the monitoring system across distinct ICU components (e.g., between medical equipment and information systems), enables remote access to real-time data for physicians, and enhances interoperability and coordination within the ICU.
Analytical System	Employs advanced methods (e.g., built-in parameters ranges, algorithms) to process and analyze data transferred to the ICU's information system, enabling the extraction of relevant and actionable clinical insights, ultimately supporting medical evaluation and decision-making.
Alarm System	Triggers alarms based on outputs from the analytical system, promptly alerting physicians to abnormal changes in the patient's clinical condition. This improves diagnostic accuracy and enhances early detection of relevant clinical dynamics, ultimately facilitating timely intervention.

Current research primarily focuses on developing increasingly precise monitoring systems that generate large volumes of data, which is essential as it forms the foundation for the functioning of all other systems. However, emerging technologies such as **5G Communication**, **Internet of Things (IoT)**, **AI**, and **Robotics** hold significant potential for further enhancing the remaining systems within intelligent ICUs.

Figure 2.2 provides a detailed overview of the architecture of an optimal intelligent ICU, showcasing the transformative improvements driven by increasingly robust data management frameworks, pivotal for optimizing clinical workflows and improving the quality and efficacy of care.

Despite the significant benefits associated with integrating emerging technologies such as AI, IoT, and Robotics, current ICU infrastructures demonstrate an inability to fully harness these advances, which is primarily related to issues referring to data acquisition, transmission, and integration (e.g., incompatibilities between monitoring devices, proprietary limitations imposed by manufacturers, and the absence of standardized data management processes). Ultimately, these obstacles hinder the seamless flow of information, making it difficult to achieve the full benefits of intelligent ICU systems [13].

To overcome these challenges, advancements in clinical informatics and information technology are essential. In particular, improvements regarding the data acquisition, transmission, integration, and analysis frameworks are required to handle the increasing volume of data generated within modern

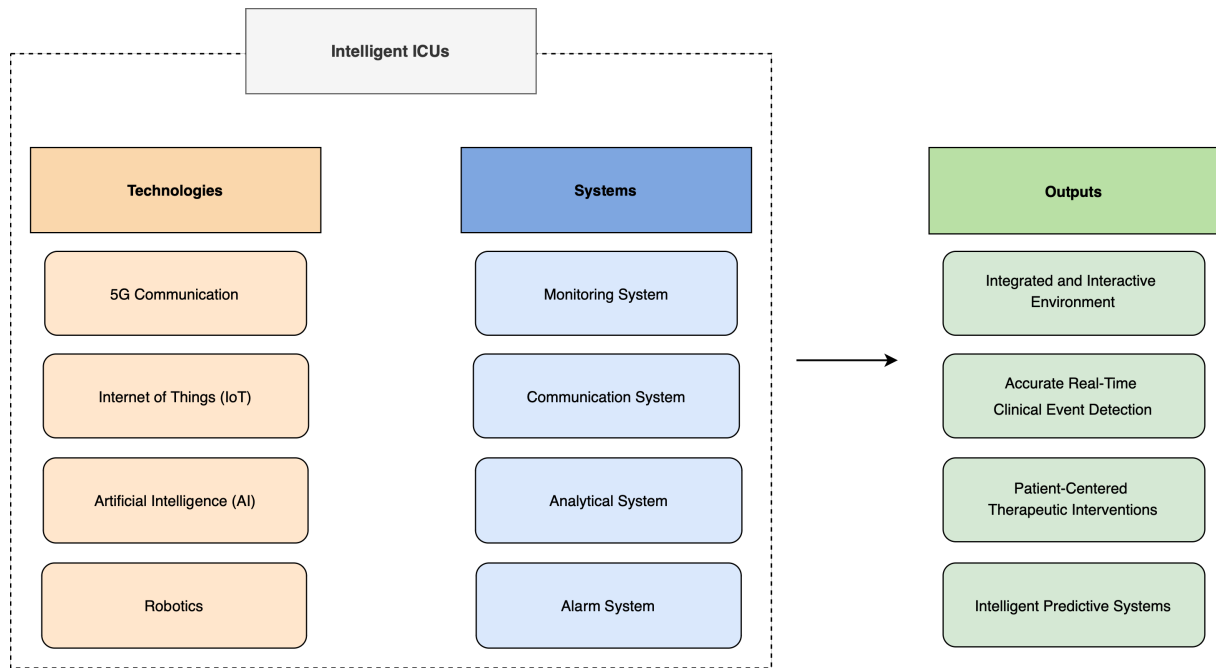


Figure 2.2: Architecture of an optimal intelligent ICU.

ICUs. Indeed, the Intensive Care Department of Hospital de Santa Maria showcased these limitations, as it relied on rudimentary methods (e.g., manual recording of clinical variables into spreadsheets) to manage highly complex data.

Ultimately, as shown in figure 2.2, the efficient integration of emerging technologies and advanced systems can lead to several transformative improvements, including the following:

- **Integrated and Interactive Environment:** efficient communication systems, combined with enhanced interoperability among devices (i.e., ability to communicate and transfer data between one another), enable a more synchronized and accurate response to clinical situations. In addition to facilitating real-time information sharing, these environments allow physicians to remotely adjust device parameters, drug administration settings, and monitoring configurations, ultimately enhancing overall ICU coordination and responsiveness.
- **Accurate Real-Time Clinical Event Detection:** the ability to process and analyze large volumes of data in real time improves clinical event detection. This capability allows for earlier and more accurate diagnosis, increasing the likelihood of timely interventions that can prevent deterioration and improve patient outcomes.
- **Patient-Centered Therapeutic Interventions:** advanced monitoring systems capable of monitoring and assessing patient-specific characteristics more accurately, enabling physicians to design more targeted (ultimately precise) therapeutic interventions. This patient-centered approach optimizes care (e.g., by reducing unnecessary interventions and minimizing risks), enhancing recovery and improving patient outcomes.
- **Intelligent Predictive Systems:** AI-powered systems that analyze patient-specific information in

real-time, predicting relevant clinical dynamics, including before they manifest. These predictive capabilities allow physicians to act early and potentially preventively, mitigating risks and ultimately improving prognostic outcomes. Recent studies suggest that predictive models can be instrumental in preventing critical events from unfolding by providing early warning of potential clinical complications.

In conclusion, the shift to intelligent ICUs, powered by the continuous integration of increasingly sophisticated data management frameworks and advanced technologies, represents an irreversible, potentially revolutionary trend, offering a solution to many of the structural challenges faced by conventional intensive care medicine, and ultimately benefiting both the medical staff and patients.

2.2 Extracorporeal Membrane Oxygenation

Following the previous section, ECMO support represents a critical therapeutic intervention that can significantly benefit from continuous technological advancement and integration within ICUs and the shift towards more intelligent architectures. ECMO, which provides life support to patients with severe cardiac and pulmonary dysfunction, relies on a complex network of systems that monitor a wide range of physiological parameters and clinical variables, generating multidimensional and multidomain datasets. Despite their informative value, these are often unstructured, which poses a challenge to their accurate analysis.

Moreover, the technical and functional specifications of ECMO devices often complicate data transmission, limiting accessibility and hindering the ability to conduct high-value analyses (e.g., identification of clinical trends, such as deterioration and improvement, to enable timely medical intervention). In fact, the Intensive Care Department at Hospital de Santa Maria relies on manual recording of ECMO-generated data, a process highly susceptible to errors.

The complexity of the datasets produced for patients under ECMO support adds to the challenges of analyzing them, even though sufficiently experienced physicians can still extract valuable clinical insights. **This challenge is further amplified by the absence of integrated analytical systems within ICU information systems, as observed in the Intensive Care Department of Hospital de Santa Maria.**

Addressing this issue was at the core of this study. The limited body of literature on the analysis and interpretation of ECMO-generated datasets underscores the pressing need for advancements in this area.

Ultimately, this section explores relevant technical and functional aspects of ECMO support, focusing in great detail on the physiological principles underlying this therapeutic intervention and explaining medical concepts essential for understanding its role in managing severe respiratory failure. By delving into ongoing physiological mechanisms, this section aims to connect closer with the intricacies of intensive care medicine and the broader medical field. It should be noted that some of the information provided in this section derived from the ECMO protocol specific to Hospital de Santa Maria, as well as valuable discussions with the physicians who supported and participated in this study.

2.2.1 Overcoming the Limitations of Mechanical Ventilation: The Rise of Extracorporeal Membrane Oxygenation

Modern extracorporeal life support systems represent relatively recent advancements, enabling the temporary support of organ function. As outlined in section 2.1.1, the emergence of intensive care medicine was driven by the necessity to support patients with failing lungs, a challenge initially addressed through the development of **mechanical ventilators**. These ventilators have since played a pivotal role in providing respiratory support and oxygenation to millions of patients worldwide.

Despite their critical function of assuring appropriate respiratory support and oxygen delivery to patients with severely compromised lung function (e.g., resulting from infection or inflammation), mechanical ventilators are not exempt from potentially adverse side effects. Pham et al. [14] delves into several commonly observed adverse effects and their respective causes, which are summarized in table 2.2.

Table 2.2: Adverse effects frequently observed in patients undergoing mechanical ventilation [14].

Adverse Effect	Cause
Barotrauma	Unregulated or maladjusted pressure application.
Volutrauma	Excessive or maladjusted air volume administration.
Heart Failure	Excessive constriction of the heart.
Oxygen Toxicity	Abnormally high oxygen concentration.
Muscle Atrophy	Inadequate sedation management.

These limitations highlight the need for alternative techniques that can address respiratory failure without inducing the harmful effects associated with mechanical ventilation. Extracorporeal life support systems, such as ECMO, offer a promising solution by supporting lung function while avoiding many of the adverse effects linked to ventilators. As a result, ECMO is increasingly seen as a vital and forward-looking intervention in intensive care medicine [15].

2.2.2 Extracorporeal Membrane Oxygenation: Historical Context and General Perspective

ECMO was first developed over 50 years ago at the University of Michigan to support respiratory function in patients with severe thoracic blunt trauma [16]. This technique involved blood extraction, external oxygenation, and return to the patient's body. While this early application was successful in saving a patient, the ECMO system at the time proved to be highly complex and resulted in severe systemic complications (e.g., hemorrhagic episodes, multisystemic organ failure), ultimately leading to high mortality rates in subsequent studies. In 1979, Zapol et al. [17] conducted a study with 90 patients suffering from acute respiratory failure, where 48 were treated through conventional mechanical ventilation, and 42 were treated using ECMO support in complementarity. The outcomes were relatively discouraging, with only eight patients surviving (evenly split between both groups). Consequently, ECMO remained

largely experimental and primarily confined to research centers for the following three decades.

Significant advancements regarding extracorporeal life support systems occurred with the development of miniaturized extracorporeal circuits, which reduced the priming volume (i.e., the amount of fluid required to fill the tubing and components of the extracorporeal circuit), thereby decreasing the risks of hemorrhagic episodes, thrombosis, and systemic inflammation. These improvements were critical during the 2009 H1N1 influenza pandemic, which started in Australia [18], where ECMO support was successfully implemented and proved decisive in treating critically ill patients. In Portugal, the H1N1 pandemic peaked in 2010, promoting the establishment of national ECMO programs at Hospital de Santa Maria in Lisbon and Hospital de São João in Porto, which have since become the country's primary ECMO referral centers.

In its most conventional configuration, extracorporeal life support involves the insertion of an extracorporeal circuit between two large central veins of the patient. This circuit consists of large cannulas and an artificial membrane designed to simulate the function of the native lung. As illustrated in figure 2.3, the circuit operates with a low priming volume (less than 500 mL) and is designed to accommodate a low-resistance flow, allowing blood flow rates between 0.5 and 7 liters per minute (L/min). Achieving high flow rates is critical to meet the patient's cardiac output (typically constrained between 4.5 and 6 L/min) demands. The relevance of ensuring a balanced relationship between extracorporeal flow and cardiac output will be discussed later in this section.

Achieving these flow rates requires the ECMO circuit to be connected to a central vein (typically the inferior vena cava) to drain the patient's deoxygenated blood. An electromagnetic pump generates negative pressure (ranging from -100 to -200 mmHg), drawing blood from the vein into the circuit. This pump, positioned between the drainage cannula and the oxygenation membrane, minimizes mechanical stress on blood cells, thus reducing the risk of hemolysis, a phenomenon typically observed in older roller pump systems. After the pump, the pressure increases as the blood flows toward the oxygenation membrane, often exceeding 300 mmHg. Once the blood passes through the membrane, a pressure drop (known as transmembrane pressure drop) of around 20 to 35 mmHg occurs (the exact value depends on the flow rate) before returning the blood to the patient through a separate cannula inserted in another central vein (typically the superior vena cava). This configuration ensures that the right heart receives a blood volume equivalent to that seen under normal physiological conditions.

This is the most frequently adopted configuration to treat cases of severe respiratory failure, known as Venovenous Extracorporeal Membrane Oxygenation (VV-ECMO). The respiratory support provided by ECMO relies on the oxygenation membrane's physiology, discussed in the following section on lung physiology.

2.2.3 Foundations of Extracorporeal Membrane Oxygenation: Lung Physiology and Oxygen Metabolism

The primary respiratory function of the native lung is to oxygenate venous blood and remove CO₂ produced by cellular metabolism. Under normal conditions, oxygen from the atmosphere is transferred

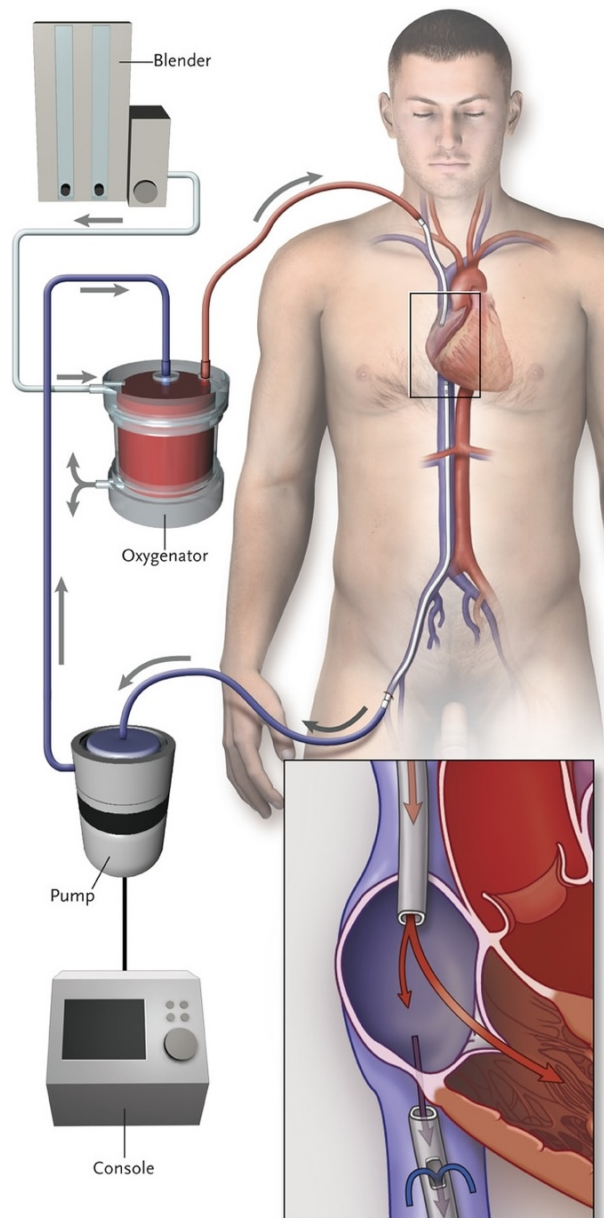


Figure 2.3: Illustration of the ECMO circuit applied to a patient (adapted from [15]).

from the pulmonary alveolus into the bloodstream, primarily binding to hemoglobin within erythrocytes (also known as red blood cells) through a process named hematoses. Hemoglobin, which constitutes around 98% of the erythrocyte cytoplasmic content, is the key molecule responsible for oxygen transportation across the human body. Each erythrocyte contains approximately 300 million hemoglobin molecules, where a healthy hemoglobin concentration corresponds to around 150 grams per liter (g/L), while values below 120 g/L indicate that the patient has anemia.

Hemoglobin's oxygen-binding capacity (essential to ensure it can effectively transport oxygen across the human body) depends on its heme groups, with each hemoglobin molecule containing four such groups. Each heme group consists of a porphyrin ring linked to an iron atom in its ferrous state (Fe^{2+}), which binds oxygen to form oxyhemoglobin, responsible for giving blood its characteristic bright red color, contrasting with the bluish color of deoxygenated (venous) blood.

Physiologically, when exposed to sufficient oxygen levels, all hemoglobin molecules become fully saturated with oxygen, a state known as 100% oxygen saturation. Under these conditions, each gram of hemoglobin can carry approximately 1.34 mL of oxygen (70 times more elevated than the amount dissolved in plasma).

This knowledge is critical for calculating the oxygen content in arterial blood, expressed in milliliters per liter (mL/L), achievable through equation 2.1, where $[Hg]$ (measured in g/L) represents the concentration of hemoglobin, $SatO_2$ depicts the hemoglobin oxygen saturation, and PaO_2 reflects the oxygen partial pressure in blood.

$$CaO_2 \text{ [mL/L]} = 1.34 \cdot [Hg] \text{ [g/L]} \cdot SatO_2 \text{ [%]} + 0.03 \cdot PaO_2 \text{ [mmHg]} \quad (2.1)$$

Monitoring oxygen metabolism is critical for understanding the patient's physiological state and assessing the effectiveness of therapeutic interventions in critical care settings. The total amount of oxygen delivered to the tissues and organs per minute, known as oxygen delivery (DO_2), can be easily computed using information regarding the oxygen content in arterial blood and the cardiac output (measured in L/min), which represents the volume of blood pumped by the heart and whose assessment relies on one of several available bedside techniques, ranging from Doppler echocardiography to minimally invasive hemodynamic monitoring. Oxygen delivery is then calculated using equation 2.2, where CO represents the cardiac output and CaO_2 depicts the oxygen content in arterial blood (computed through equation 2.1).

$$DO_2 \text{ [mL/min]} = CO \text{ [L/min]} \cdot CaO_2 \text{ [mL/L]} \quad (2.2)$$

In normal conditions, the human body fully saturates hemoglobin in the lungs ($SatO_2$ while breathing atmospheric air), resulting in an oxygen delivery of around 800 to 1000 mL/min, assuming a healthy cardiac output of 4.5 to 6 L/min. The body typically consumes 180 to 250 mL/min of oxygen, corresponding to an oxygen extraction ratio of approximately 20-25%. When the lung function deteriorates, blood oxygenation decreases while the extraction ratio rises. These changes in oxygen metabolism can be assessed by calculating oxygen consumption (VO_2), which correlates with the difference between oxygen content in arterial blood (CaO_2) and venous blood (CvO_2), as shown in equation 2.3, where CO refers to the cardiac output.

$$VO_2 \text{ [mL/min]} = CO \text{ [L/min]} \cdot (CaO_2 \text{ [mL/L]} - CvO_2 \text{ [mL/L]}) \quad (2.3)$$

In parallel, the body continuously produces CO_2 , a volatile acid that requires systematic removal from the human body to maintain pH balance (i.e., with pH values ranging from 7.35 to 7.45). CO_2 is highly soluble and diffuses rapidly from the bloodstream into the pulmonary alveolus, where its removal relies on a process known as alveolar ventilation (most commonly referred to as regular breathing). Normal arterial CO_2 partial pressure ($PaCO_2$) is typically constrained within the interval [38, 42] mmHg and can be easily measured through appropriate blood gas analysis. When $PaCO_2$ exceeds 42 mmHg, CO_2 sig-

nificantly accumulates, forming carbonic acid (H_2CO_3), which releases H^+ ions into the bloodstream, ultimately lowering blood pH. Conversely, increased alveolar ventilation (as seen during systemic inflammation) can cause PaCO_2 to fall below the 38 mmHg minimum threshold, resulting in an elevation of blood pH. Notably, CO_2 diffuses approximately 40 times faster than oxygen, meaning respiratory failure due to primary lung disease typically leads to hypoxemia (low blood oxygen) with hypercapnia (elevated levels of CO_2) being less common.

2.2.4 Physiology of Extracorporeal Membrane Oxygenation

Hypoxemia frequently occurs in patients with respiratory failure due to lung disease. The initial medical response to this condition typically involves increasing the oxygen concentration of the inspired air, thus raising the oxygen levels at the pulmonary alveolus. In more critical cases, invasive mechanical ventilation may be necessary to reduce the patient's breathing effort, restore adequate alveolar ventilation (i.e., ensure effective elimination of CO_2), and administer oxygen through controlled pressure and concentration (FiO_2).

For the most severe cases of respiratory failure, the hypothesis concerning the implementation of extracorporeal respiratory support (e.g., ECMO) should be raised. This is usually required in the presence of refractory hypoxemia (when pure oxygen administration fails to normalize oxygen levels, increasing the risk of oxygen toxicity from free radicals), elevated ventilatory pressures (expressed through high plateau pressures, which indicate the lung's pressure at the end of inspiration), or difficulties in CO_2 removal, which results in increased levels of PaCO_2 .

As previously outlined, ECMO involves integrating a controlled extracorporeal circuit into the patient's circulatory system. This circuit encompasses a specially designed membrane through which blood passes once drained from the patient's body, mimicking lung function by enabling effective gas exchange. This polypropylene-made membrane, which resembles alveolar capillaries, consists of numerous fibrous capillaries with a total surface area of 1.8 m^2 through which blood is pumped, with a physician-determined flow rate (usually between 0.5 and 7 L/min). Respiratory support is provided by ventilating the membrane with a sweep gas (with a flow rate between 0.1 and 12 L/min), which enables CO_2 removal from the blood drained from the patient. Ultimately, this process lowers PaCO_2 in the post-filter blood, leading to a reduction in the levels of CO_2 within the arterial blood, which will be reinfused into the patient. Although ECMO is often seen primarily as an oxygenation support tool, it is equally effective in removing CO_2 from the patient's body. Nonetheless, the primary reason for implementing ECMO support remains the need to improve oxygenation.

To fully understand ECMO physiology, it is imperative to revisit the principles of lung physiology and oxygen metabolism, detailed in section 2.2.3. The membrane within the ECMO circuit mimics healthy alveolar function while sweeping gas ventilation simulates healthy air exchange. The sweep gas flow (measured in L/min), with a predefined (through the blender) oxygen concentration, is typically synchronized with blood flow through the membrane (figure 2.3). In cases dominated by hypoxemia (particularly in the early stages of disease), the sweep gas often consists of pure oxygen ($\text{FiO}_2 = 1.0$), thus ensuring

full hemoglobin saturation in the blood passing across the membrane (appropriate post-filter blood gas analysis can confirm full hemoglobin saturation). Ultimately, the amount of oxygen transferred from the sweep gas to the blood passing through the membrane depends on the respective hemoglobin concentration (as hemoglobin molecules are the primary oxygen carriers).

Appropriate pre-filter blood gas analysis enables a comprehensive understanding of the composition of venous blood (i.e., blood drained from the patient), including its oxygen content, computed through equation 2.4, where $[Hg]$ represents the hemoglobin concentration, $SatO_2$ depicts the hemoglobin oxygen saturation, and PaO_2 reflects the oxygen partial pressure, all measured in venous blood.

$$C_{pre-filter}O_2 \text{ [mL/L]} = [1.34 \cdot [Hg] \text{ [g/L]} \cdot SatO_2 \text{ [%]} + 0.03 \cdot PaO_2 \text{ [mmHg]}]_{pre-filter} \quad (2.4)$$

Conversely, appropriate post-filter blood gas analysis enables a comprehensive understanding of the composition of arterial blood (i.e., filtered blood that will be reinfused into the patient's circulatory system), including its oxygen content, computed through equation 2.5, where $[Hg]$ represents the hemoglobin concentration, $SatO_2$ depicts the hemoglobin oxygen saturation, and PaO_2 reflects the oxygen partial pressure, all measured in arterial blood.

$$C_{post-filter}O_2 \text{ [mL/L]} = [1.34 \cdot [Hg] \text{ [g/L]} \cdot SatO_2 \text{ [%]} + 0.03 \cdot PaO_2 \text{ [mmHg]}]_{post-filter} \quad (2.5)$$

A patient undergoing ECMO support requires continuous monitoring of the ECMO pump flow, a key determinant of oxygen transfer across the extracorporeal membrane. Similar to systemic oxygen delivery (which depends on cardiac output), represented in equation 2.2, the amount of oxygen transferred through the ECMO circuit per minute (ECMO oxygen transfer) is computed through equation 2.6, where $ECMO \text{ Pump Flow}$ represents the flow rate of the ECMO circuit pump, while $C_{pre-filter}O_2$ and $C_{post-filter}O_2$ denote pre-filter and post-filter oxygen content (computed through equations 2.4 and 2.5), respectively.

$$ECMO \text{ } O_2 \text{ Transfer [mL/min]} = ECMO \text{ Pump Flow [L/min]} \cdot (C_{post-filter}O_2 - C_{pre-filter}O_2) \text{ [mL/L]} \quad (2.6)$$

Under typical ECMO conditions, patients with hemoglobin levels ranging from 80 to 100 g/L and an ECMO pump flow between 4 and 6 L/min receive between 186 and 350 mL of oxygen per minute, which generally satisfies oxygen demands. However, hypercatabolic patients (i.e., patients presenting with excessive metabolic breakdown of complex substances) or those presenting with a higher body mass index may require more significant oxygen transfer. These scenarios may require increasing hemoglobin levels through erythrocyte transfusion to meet elevated consumption needs.

When ECMO support starts, native lung function typically diminishes significantly, resulting in minimal or even absent oxygenation and CO_2 removal by the patient's lungs. Although the physiological mechanisms behind this suppression are beyond the scope of this discussion, a key finding is that calculations of native lung oxygen transfer may yield negative values, which occur due to the lung's tissue

oxygen consumption and its near-zero contribution to oxygen enrichment. Additionally, a minor degree of blood recirculation in the ECMO circuit (usually below 10%) can also affect oxygen levels, though its impact is generally negligible.

Both native lung oxygen transfer and ECMO oxygen transfer hold relevant information regarding the prospects of patient recovery, supporting the prediction of lung function restoration. As the patient's clinical condition improves, minor increases in native lung oxygen transfer should be detected, followed by a gradual decline in ECMO oxygen transfer as the lungs resume functioning to meet the body's oxygen demands.

To analyze native lung oxygen transfer, it is essential to initially estimate total body oxygen consumption and determine the ECMO system's relative contribution. When native lung function is practically absent, ECMO oxygen transfer should equal total body oxygen consumption, as significant oxygen losses are nonexistent. As native lung function starts recovering, the ECMO system's contribution diminishes accordingly.

Total body oxygen consumption (VO_2) can be estimated by comparing arterial blood oxygen content (CaO_2) and venous blood (i.e., pre-filter) oxygen content ($C_{pre-filter}O_2$), as shown in equation 2.7, where CO represents cardiac output, and CaO_2 and $C_{pre-filter}O_2$ are computed through equations 2.1 and 2.4, respectively.

$$VO_2 [\text{mL/min}] = CO [\text{L/min}] \cdot (CaO_2 - C_{pre-filter}O_2) [\text{mL/L}] \quad (2.7)$$

Through this estimate, native lung oxygen transfer can then be determined by simply computing the difference between total body oxygen consumption and ECMO oxygen transfer, as shown in equation 2.8, where $ECMO O_2 Transfer$ (ECMO oxygen transfer) and VO_2 (total body oxygen consumption) are computed through equations 2.6 and 2.7, respectively.

$$Native Lung O_2 Transfer [\text{mL/min}] = VO_2 [\text{mL/min}] - ECMO O_2 Transfer [\text{mL/min}] \quad (2.8)$$

The physiological variables explored across this section serve as robust indicators of the patient's clinical condition throughout hospitalization under ECMO support, providing critical insights for physicians to assess the effectiveness of ongoing treatments and determine the need for alternative strategies or supplementary therapeutic interventions. Moreover, additional physiological variables (e.g., related to the functioning of other organ systems) are typically monitored in patients undergoing ECMO support, highlighting the complexity of balancing multiple systems to ensure the efficacy of this therapeutic intervention.

2.3 Chapter Conclusions

The growing complexity of critical care environments, combined with recent technological advancements, creates a pathway for developing increasingly intelligent ICUs, which incorporate multiple sys-

tems that augment conventional medical practices through advanced monitoring, efficient data transmission, and analytical frameworks. Within the context of ECMO, such systems play a pivotal role in managing the multidimensional and multidomain data generated throughout extracorporeal support, facilitating the accurate and timely production of clinical insights. This ability is particularly relevant for ECMO patients, whose conditions are often volatile and challenging to interpret.

The intricacies and complexity of ECMO support, demonstrated across this chapter, require significant expertise in operating the device and interpreting the extensive datasets produced. Expert physicians who supported or participated in this study emphasized that these patients' critical condition and volatility often compromise their ability to fully understand clinical dynamics, leading to a frequent state of uncertainty. Sudden and unexpected changes in the patient's clinical condition frequently arise, further complicating the physicians' ability to interpret the patient's condition through the exclusive use of medical knowledge and conventional methods.

This inherent unpredictability and the frequent inability to draw comprehensive conclusions from the data validate the core premise of this study: automatic systems powered by advanced data processing and analytical technologies such as ML can significantly complement clinical practice. These systems can uncover insights and patterns previously undetectable (or even inaccessible through conventional analysis) in datasets referring to patients undergoing ECMO support, enabling physicians to access high-value, real-time information that can enhance decision-making and improve patient outcomes.

Chapter 3

Related Work

This chapter provides technical insights into the development of ML risk prediction models in healthcare. A comprehensive overview is initially delivered (section 3.1), introducing ML as a promising tool for addressing the challenges inherent in risk assessment within healthcare, presenting technical details into how risk prediction problems can be framed to ensure compatibility with ML procedures, and reviewing several studies that employed distinct problem-framing structures and ML models for predicting the risk of adverse clinical events in critical care settings, ultimately providing methodological insights that inspired this study. Section 3.2 explores ML risk prediction in the context of ECMO, highlighting current limitations and constraints. Ultimately, section 3.3 concludes this chapter by identifying future research opportunities based on gaps identified throughout the literature.

3.1 Machine Learning Risk Prediction in Healthcare

Progressive technological advancements have led to the deployment of increasingly precise medical equipment in hospital settings, enabling continuous monitoring of patient-specific information and leading to the generation of high-dimensional datasets (typically stored in dedicated systems). However, this data is often underutilized due to its complexity and the labor-intensive, time-consuming nature inherent to analyzing it. This issue is further aggravated by low staff-to-patient ratios commonly observed within ICUs. According to Almenyan et al. [19], there is an evident relationship between nursing workload and patient safety, with inadequate or insufficient nursing care contributing to an elevated risk of severe complications (e.g., pneumonia and in-hospital infections), more prolonged hospitalization periods, and higher mortality rates. Conversely, Kahn et al. [20] found that ICU mortality rates seem more resistant to physicians' overload, noting nonetheless that these findings may not be generalizable to ICUs organized differently or with lower resource availability compared to those in the study.

Emerging technologies such as ML present promising avenues for addressing these challenges, with ML-based risk prediction models (e.g., for predicting clinical deterioration, onset of critical events, and mortality risk) enhancing improved resource allocation and enabling more efficient clinical assessments and decision-making for critically ill patients.

This section introduces the concept of ML-based risk prediction models in healthcare, providing a theoretical overview followed by a technical review of the intricacies of problem framing.

3.1.1 Introduction to Machine Learning Risk Prediction in Healthcare

Previous research has demonstrated that predictive models for clinical deterioration can detect adverse clinical events with notable accuracy, thus aiding in early intervention. As noted by Blythe et al. [21], a framework widely adopted for developing these models is the Early Warning Score (EWS), which encompasses both traditional applications such as the Modified Early Warning Score (MEWS) and the National Early Warning Score (NEWS) and more advanced configurations that combine knowledge-driven information with data-driven techniques such as statistical analysis and ML, which present a notable capacity to estimate relationships between observed variables (e.g., vital signs, laboratory data, sociodemographic information) and adverse events.

Despite their potential, several studies focusing on assessing the impact of integrating EWSs into clinical practice have reported mixed results, with a significant portion not finding detectable improvements in clinical outcomes (e.g., in-hospital cardiac arrest and mortality) compared to clinical judgment alone. Blythe et al. [21] recognizes the wide range of reasons that can lead to these observations but points out the following: **inefficient integration with clinical workflows**, primarily reflected in the lack of sufficiently interpretable information made available, ultimately hindering the physicians' ability to conduct more precise and rapid assessments of the patient's clinical condition and intervene early to prevent additional deterioration. Baker and Gerdin [22] demonstrated how accurately addressing these issues can improve patient outcomes.

Highly skilled and experienced nurses and physicians can often identify markers of clinical deterioration over equivalent periods as those of deterioration models [23], thus demonstrating that **these models should aim to complement and support clinical evaluation and decision-making rather than replace medical expertise**. Guaranteeing that these models play a supporting role in clinical practice requires ensuring that their outputs are easily accessible and interpretable. These conditions enable physicians to spend less time assessing the reliability of model predictions and focus on providing the necessary care to the deteriorating patient. Following these observations, Baker and Gerdin [22] note that future research should focus on refining models regarding their **performance** (e.g., in terms of discrimination and calibration) and **practical utility**, which can be achieved by ensuring that the models' outputs are interpretable and appropriately validated through extensive real-world testing across diverse hospital settings.

As Jahandideh et al. [24] and Muralitharan et al. [25] demonstrate through extensive reviews, ML-based models, particularly Support Vector Machine (SVM), Random Forest, and Artificial Neural Network (ANN), revealed enhanced accuracy in predicting clinical deterioration for hospitalized patients, often outperforming traditional EWS approaches (e.g., MEWS and NEWS). These models leverage their ability to process complex relationships within highly comprehensive datasets, enabling more precise and timely risk assessments.

Maximizing these models' effectiveness and practical utility requires increasingly embedding them within hospitals' digital infrastructures such as Electronic Health Records (EHRs). This integration promotes enhanced accessibility to physicians, enabling them to access patient-specific information (e.g., estimates of the risk of clinical deterioration) in real time, ultimately improving the quality and precision of care. However, achieving these results requires digital infrastructures to automatically collect and process patient data, which is crucial to ensure models use comprehensive, high-quality information to generate risk predictions. **Additionally, the physicians' response to alerts significantly impacts patient outcomes, with protocolized and well-structured response workflows proving more effective in ensuring timely and precise interventions [21]. As a result, future research should also focus on refining alarm-response workflows to improve patient outcomes.**

The following section explores different framing strategies that can enhance the applicability of ML frameworks, building on the promising results of ML-based models in assessing the risk and predicting adverse clinical events.

3.1.2 Machine Learning Risk Prediction in Healthcare: Problem Framing

The performance, applicability, and practical utility of ML-based risk prediction models depend significantly on appropriate model selection (including the corresponding hyperparameters), evaluation frameworks, and interpretability methods. However, as emphasized by Muralitharan et al. [25], appropriate data preparation, including setting predictor variables and defining the outcome variable, data preprocessing, and feature engineering, are equally important. Ultimately, the framing strategy adopted significantly depends on the data preparation strategies selected.

Lauritsen et al. [26] provides a comprehensive overview of the basic concepts of problem framing for developing ML-based risk prediction models, demonstrating the application of four different framing structures to the same generic dataset (referring to hospitalized patients) and assessing their impact on the performance (in terms of discrimination and calibration concerning sepsis prediction) and clinical relevance (reflected by the applicability and practical utility in real-life healthcare settings) of five ML models, including Random Forests and Extreme Gradient Boosting (XGBoost). The four framing structures considered were the following: **fixed time to onset**, **sliding window**, **sliding window with dynamic inclusion**, and the **on clinical demand**, each characterized by unique time intervals for data acquisition and outcome prediction.

Among the four framing structures outlined, **sliding window** stands out as the most clinically applicable and useful one due to its ability to provide continuous real-time assessments of the patient-specific clinical risk throughout hospitalization, as well as converting a sequential supervised learning problem into a standard supervised learning format, particularly under conditions of limited data availability. Figure 3.1 illustrates the sliding window implementation, which segments sequential data into overlapping windows, generating numerous samples that can be assessed through conventional ML algorithms. These samples preserve critical temporal dependencies by integrating the following components:

- **Observation window:** retrospective time window measured from prediction time, containing the

measurements of the predictor variables, evaluated to generate predictions. The observation window's length can be fixed across the entire hospitalization period or vary based on data availability (e.g., for the initial period of hospitalization, there may be no data available to generate observation windows with the established length, resulting in incrementally larger windows until sufficient data is available).

- **Prediction time:** time point for which an informed prediction (based on the analysis of the observation window) is generated.
- **Prediction window:** time window considered to extrapolate the value of the outcome variable for prediction time (i.e., the values of the outcome variable observed throughout this window are combined to determine the outcome for prediction time). Setting an appropriate prediction window is crucial to enhance model performance and ensure its clinical relevance, as shorter windows may provide higher accuracy but limit response time, while longer windows may enable timely interventions but present with reduced predictive performance.
- **Window shift:** time interval between consecutive samples, depicting the shift between successive observation windows.

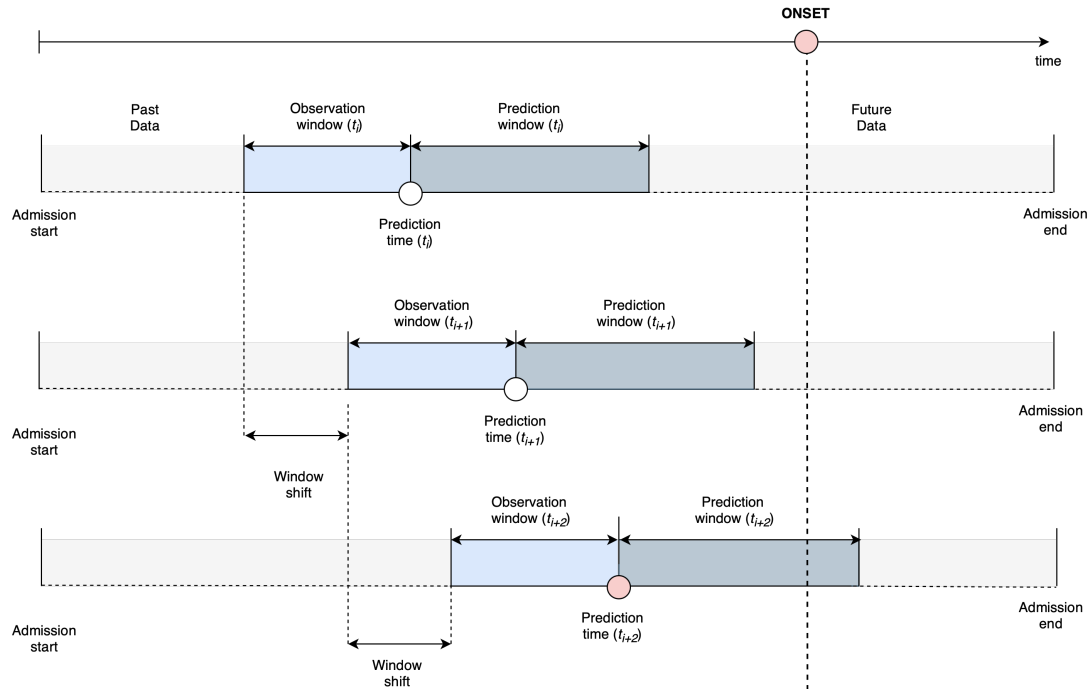


Figure 3.1: Implementation of the sliding window framing structure, with "ONSET" reflecting the time of occurrence of singular clinical event in time.

The onset depicted in figure 3.1 refers to a singular clinical event in time (e.g., the onset of sepsis), occurring only once throughout the observation period (from admission to discharge). If the onset is positioned to the right of the prediction window, it indicates that the onset has not yet occurred and will not occur within the window of interest, as shown in instances t_i and t_{i+1} . Conversely, if the onset

coincides with or falls to the left of the prediction window, it indicates that the onset will occur within the window of interest (instance t_{i+2}) or has already occurred.

Lauritsen et al. [26] demonstrated that the best-performing (in terms of discrimination and calibration) ML-based risk prediction model was framed as a fixed time to onset. **However, it proved clinically not applicable as the inherent framing structure depends on specific future information that may not be available in real-time.**

In conclusion, while ensuring that the ML models for risk prediction perform well (in terms of discrimination and calibration) is crucial to ensure their clinical applicability, it is equally relevant to ensure that a framing structure that enables the models' alignment with clinical workflows is selected. Additionally, choosing an appropriate framing structure can ensure that the models limit the number of False Positives (FPs) generated, thus reducing alarm fatigue (typically observed in high-intensity clinical settings such as the ICU). Ultimately, this choice can also contribute to ensuring that the models provide interpretable insights that align with the physicians' needs for better understanding and evaluating the patients, thus enabling them to improve the quality of care and patient outcomes.

3.1.3 Machine Learning Risk Prediction in Healthcare: Study Review and Methodological Insights

Several studies whose data preparation frameworks relied on the framing structures outlined in section 3.1.2 were selected for further assessment, forming the basis for designing a research methodology that aligns with this study's settings and objectives.

The selected studies, summarized in table 3.1, focused on developing ML-based risk prediction models applicable to intensive care medicine, a critical, data-rich medical field that can substantially benefit from integrating automated systems capable of analyzing large volumes of data and generating actionable insights that support clinical evaluation and decision-making. Ultimately, these studies demonstrate the promising impact of ML models for assessing the risk and predicting adverse clinical events within ICUs, reflected in the physicians' enhanced ability to make better-informed decisions and thus ensure early intervention amidst critical clinical scenarios, possibly preventing irreversible and fatal deterioration.

The studies summarized in table 3.1 leverage the extensive, highly complex, and comprehensive datasets available within EHRs to develop ML models for predicting adverse clinical events. Implementing appropriate data preparation strategies was essential to the success observed across these studies, enabling the models to predict adverse outcomes in advance. Each study first established a specific outcome and then determined the most suitable framing structure. In studies focused on early event detection, the sliding window structure, outlined in section 3.1.2, stood out, as seen in the studies conducted by Tomašev et al. [27] and Noy et al. [28], who used the sliding window structure considering prediction windows of 48 hours and 21 hours (ranging from 7 to 30 hours after prediction time), respectively.

Table 3.1: Summary and key findings from several studies on the development of ML risk prediction models in healthcare.

Reference	Summary and Key Findings
Tomašev et al. [27]	The authors present a Deep Learning (DL) model for predicting the risk of Acute Kidney Injury (AKI) up to 48 hours in advance , aiming to enhance early prediction and facilitate timely intervention amidst this adverse clinical event (commonly observed within the ICU). By continuously updating risk predictions through real-time sequential data assessment, this model enables uninterrupted monitoring of the patient's risk for AKI throughout hospitalization.
Noy et al. [28]	The authors present a ML model for predicting the clinical deterioration of COVID-19 inpatients, defined according to the value of the modified National Early Warning Score (mNEWS2) adapted for COVID-19. The model focused specifically on identifying patients at higher risk of requiring increased levels of care within the next 7 to 30 hours . The model achieved notable predictive performance, enabling physicians to adopt preventive measures for attenuating or preventing the effects of clinical deterioration.
Liu et al. [29]	The authors developed a ML model for predicting septic shock in ICU patients before its onset. This study introduced an innovative approach by assuming the existence of a novel "pre-shock" clinical state that precedes septic shock. ML techniques were applied to characterize this state through extensive analysis of EHRs from the MIMIC-III database. The best-performing model exhibited a notable ability to predict the patient's transition to the "pre-shock" state, effectively enabling the early identification of septic shock by achieving a median early warning time of 7 hours. Ultimately, the authors aimed to provide physicians with a time window during which they could intervene to reverse or prevent clinical deterioration and mortality.
Garcia-Gutiérrez et al. [30]	The authors developed a ML-based predictive model for identifying patients at risk of clinical deterioration to facilitate early detection of individuals who may require progressively more intensive care (e.g., intensive ventilatory support). This model supports optimized resource allocation throughout hospitalization by ensuring patients receive appropriate care and the resources needed to address their clinical conditions. Despite being tailored for COVID-19 patients, the methodology implemented in this study is generalizable for other patient cohorts and healthcare settings.
Hyland et al. [31]	The authors developed a ML-based early warning system for predicting circulatory failure in ICU patients before its onset . This model enabled physicians to anticipate and respond to circulatory failure events up to 8 hours in advance, thus facilitating timely intervention. A key strength of this study was the model's low false-alarm rate, which enhanced its applicability and practical utility by reducing alarm fatigue in an environment already characterized by intense alarm activity.

Furthermore, these studies highlighted the high predictive performance that ML and DL models can achieve when sufficient data is available. Given the critical clinical context in which these studies were conducted, feature importance was consistently evaluated, with the SHapley Additive exPlanations (SHAP) method being the predominating technique. This approach to interpretability, as used by Noy et al. [28], Garcia-Gutiérrez et al. [30], and Hyland et al. [31], allows clinicians to better understand the most influential factors driving model predictions, supporting better-informed decision-making. Ultimately, assessing feature importance was also crucial to enhance the models' applicability and practical utility.

3.2 Machine Learning Risk Prediction in Extracorporeal Membrane Oxygenation

As noted in section 3.1, inadequate staff-to-patient ratios can significantly impact the quality of care and patient outcomes, particularly considering complex therapeutic interventions requiring constant monitoring such as ECMO support. The research by Lucchini et al. [32] suggests that a 1:1 nurse-to-patient ratio is critical for patients under ECMO support. However, an aging population and the rising demand for this therapeutic intervention make it increasingly challenging to ensure the ideal conditions for patients under ECMO support. As a result, there is a pressing need to develop and integrate technologically advanced, automated data analysis and risk prediction systems to support the management of this patient cohort.

Several risk scores have been developed to assess the prognosis of patients hospitalized under ECMO support. The following well-established, widely referenced scores apply straightforward mathematical formulations incorporating specific patient characteristics (e.g., demographic information, comorbidities, physiological parameters, and clinical variables) to produce individualized risk assessments. The following scores stand out as some of the most commonly applied in the context of ECMO support:

- **PRESET Score:** specifically designed for patients under ECMO support due to severe respiratory failure, the PRESET Score is typically computed at the time of ECMO initiation, providing an individualized static estimate of survival using patient-specific demographic information (e.g., age) and clinical variables (e.g., platelet count and bilirubin levels). The PRESET Score is not updated throughout hospitalization, thus not reflecting changes in the patient's clinical condition [33].
- **RESP Score:** similar to the PRESET Score, the RESP Score, specifically designed for patients under ECMO support due to acute respiratory failure, is usually computed at the time of ECMO initiation, providing an individualized static estimate of survival using patient-specific clinical indicators, including cardiac, respiratory, and central nervous system dysfunction, and immunocompromised state [34].
- **Murray Score:** contrary to the PRESET Score and RESP Score, the Murray Score is designed primarily to assess the severity of lung injury rather than predict survival outcomes. The Murray Score

is commonly computed at the time of ECMO initiation, providing an individualized assessment of the severity of lung injury, comprised between 0 and 4, based on the evaluation of four parameters: chest X-ray findings, hypoxemia (measured as the ratio between PaO_2 and FiO_2), Positive End-Expiratory Pressure (PEEP), and lung compliance. Patients exhibiting a Murray Score above 3 are often considered possible candidates for ECMO support, particularly if conventional ventilation methods prove insufficient to provide the patient with the necessary conditions for recovery [35].

- **SOFA Score:** the SOFA Score assesses multi-organ dysfunction by evaluating parameters across multiple systems (e.g., respiratory, cardiovascular, and renal). The SOFA Score is frequently employed to monitor the patient's clinical condition and overall stability throughout ECMO support, thus providing a dynamic measure. **Despite the dynamic nature inherent to the SOFA Score, it does not constitute a predictive score** [36].

The Murray Score provides physicians with information that can help them assess the need for a patient with severe respiratory failure to undergo ECMO support. The PRESET and RESP Scores, typically computed at ECMO initiation, provide physicians with information that facilitates the determination of prognosis for patients initiating ECMO support, thus supporting decision-making regarding which strategies are most suitable to ensure the patient's clinical demands. Finally, the SOFA Score provides insights that enable the physicians to better evaluate the patient's clinical condition and overall stability throughout hospitalization, guiding them to consider alternative strategies if the score indicates clinical deterioration.

While conventional scoring systems constitute valuable tools for supporting physicians in assessing critically ill patients, they are inherently limited, exclusively providing static or periodic updates on the patient's clinical condition and overall stability. Additionally, these systems rely on straightforward mathematical formulations, thereby not leveraging the extensive, highly comprehensive datasets available for each patient within the ICU and ultimately overlooking more complex, potentially nonlinear relationships between different variables. Conversely, ML models emerge as a promising solution, providing physicians with continuously updated risk estimates and data-driven predictions that capture complex underlying patterns and trends in the data. By enabling a more dynamic and accurate assessment of the patient's clinical condition and overall stability, ML can significantly enhance the management of patients under ECMO support and improve their outcomes.

Despite the increasing application of ECMO support for treating a wide range of critical clinical complications, research on ML-based solutions for this patient cohort remains relatively limited. The scarcity of studies within this area is associated with the complexity and critical nature inherent in the data of patients under ECMO support and the existing challenges concerning the accessibility to standardized, well-structured datasets. Existing ML applications primarily focus on fixed clinical outcomes (i.e., singular events in time) such as mortality, thereby neglecting the volatile clinical nature of patients throughout hospitalization under ECMO support. To illustrate the typical approach adopted in developing ML solutions for ECMO support, two studies, summarized in table 3.2, have been selected from the literature, each focusing on the development of ML-based risk prediction models for mortality considering the main

ECMO configurations: Venoarterial Extracorporeal Membrane Oxygenation (VA-ECMO) and VV-ECMO.

Table 3.2: Summary and critical analysis of two studies on ML-based risk prediction models for ECMO support.

Reference	Summary and Critical Analysis
Ayers et al. [37]	The authors developed a deep neural network model to predict survival to discharge for patients under VA-ECMO , exclusively relying on laboratory data from the first 48 hours following ECMO implementation . This relatively restricted time window enhances the model's generalizability, making it adaptable to a wide range of clinical settings and more easily integrated. Additionally, by achieving notable predictive performance, the model can provide critical assistance to physicians in assessing patient prognosis at an early stage of ECMO support, facilitating timely adjustments to treatment strategies and therapeutic interventions, thereby supporting improved quality care.
Lee et al. [38]	The authors developed ML models using XGBoost and Light Gradient Boosting (LightGBM) to predict 90-day mortality for patients under VV-ECMO , exclusively relying on EHRs available at ECMO initiation . Unlike the study by Ayers et al. [37], this study considered a more comprehensive dataset, encompassing 51 variables. The models revealed notable predictive performance, surpassing conventional scores such as PRESERVE and RESP. Ultimately, these models provide physicians with an effective tool to determine which patients present with higher success probabilities concerning the outcome of VV-ECMO. The data in this study enabled the models to capture complex dynamics, patterns, and trends in the data, enhancing predictive accuracy. However, the models may become less generalizable when considering a more comprehensive dataset, as ensuring access to all necessary variables could be challenging in distinct clinical settings.

These studies demonstrated the valuable contribution that ML risk prediction models can have in assessing the prognosis of patients under ECMO support. By generating data-driven insights, these models support physicians in clinical evaluation and decision-making, enabling timely adjustments of treatment strategies and therapeutic interventions. Additionally, by identifying the key features driving their predictions, these models provide physicians with information that enhances their understanding of complex dynamics that characterize patients under ECMO support.

However, these models also present with significant limitations, as they focus exclusively on predicting patient mortality using data from the period around ECMO initiation, thus neglecting how clinical decisions made throughout hospitalization under ECMO support influence the patients' clinical trajectories and outcomes. This gap is particularly significant given the high volatility that characterizes patients under ECMO support, which increases the difficulty in assessing the patient's clinical condition in real-time. Consequently, future research should aim to establish alternative outcomes reflecting the patient's clinical dynamics throughout hospitalization under ECMO support and model them, offering physicians continuous support in navigating this complex and volatile patient cohort.

3.3 Chapter Conclusions

This chapter underscored the critical influence of selecting appropriate problem-framing structures on the performance of ML risk prediction models, their clinical utility, and their overall applicability. Among the various structures highlighted in section 3.1.2, the **sliding window method** stood out due to its inherent value in healthcare applications. This structure enables the real-time assessment of patients' clinical trajectories, thus offering a promising pathway for supporting clinical practice and improving decision-making in real-time.

While several studies have leveraged this approach to improve risk assessment for several adverse clinical events, as outlined in section 3.1.3, work focused on extending these efforts to ECMO support remains limited, with most studies considering alternative framing structures (e.g., fixed time to onset). However, the critical nature, complexity, and volatility inherent to patients under ECMO support make the sliding window method particularly promising to positively impact current clinical practices and thus improve the quality of care (primarily through enabling accurate real-time risk assessments based on extensive data analysis).

Ultimately, this gap set the stage for conducting this study, which aims to demonstrate the promise of developing automatic systems capable of processing large volumes of complex data and generating real-time risk assessments for supporting physicians in managing patients under ECMO support.

Chapter 4

Data

The processes outlined in this chapter were conducted in strict collaboration with physicians who supported and participated in this study, which was essential to ensure the alignment of the implementation choices made with the intricacies of the data management frameworks within the ICU at Hospital de Santa Maria.

At first, the patient cohort was selected using a criteria-based approach (section 4.1), followed by the acquisition of the corresponding data (section 4.2). The unorganized and unstructured nature of the data introduced the need to implement multiple preparatory steps, including identifying and categorizing relevant variables and establishing the study's outcome (section 4.3). Lastly, after generating the different patient datasets, these were subjected to preliminary analysis and processing, focusing on gaining a deeper understanding of the data and filtering inconsistencies that could compromise the study's results (section 4.4).

4.1 Criteria-Based Patient Selection

As previously mentioned, the COVID-19 pandemic has significantly underscored the relevance of intensive care medicine. Considering the study's time frame, only patients admitted to the ICU with a primary diagnosis of COVID-19 pneumonia were assessed for eligibility, totaling 545. This preliminary filtering ensured the selection of a more uniform and homogeneous population, including patients with similar clinical profiles, facilitating subsequent processes and analyses.

Despite the narrowing of the number of patients likely to be included in the study, the established inclusion and exclusion criteria aimed to select a study population broad enough to include patients with distinct physiological characteristics and clinical manifestations, reflecting more accurately the real-world diversity observed in the ICU.

The study's context and objectives guided the establishment of inclusion criteria, resulting in the exclusive inclusion of patients submitted to ECMO support and concurrent invasive mechanical ventilation. Due to the rarity and unique nature of pediatric cases, the study focused exclusively on adult patients (age ≥ 18 years old). Applying these criteria to the initially selected population resulted in a substantial

reduction, with 452 exclusions, leaving 93 eligible patients.

The need to refine the study population led to the establishment of exclusion criteria, allowing the removal of patients whose characteristics significantly deviate from standard observations. Initially, patients subjected to the awake-ECMO modality, an uncommon therapeutic approach, were excluded, ensuring equivalent hospitalization conditions. A standard minimum duration of 7 days for ECMO support is typical, with shorter hospitalizations denoting rare occurrences marked by unexpected and particular developments, such as sudden clinical deterioration. To capture these uncommon events and enhance patient diversity within the study, a threshold of 3 days was established as the minimum accepted duration for ECMO support. After setting the conditions for patient inclusion in the study, data quality was assessed to ensure compliance with minimum quality standards, preventing issues that could negatively impact the study's results. Therefore, patients whose corresponding datasets had a ratio of missing values and outliers greater than 20% were excluded. Lastly, as some variables required manual extraction from patients' medical records stored within the ICU database, extreme incompleteness or unavailability of these files automatically resulted in patient exclusion. Applying these criteria led to the exclusion of 12 individuals, **resulting in a final study population comprising 81 patients.**

Figure 4.1 illustrates the criteria-based patient selection process considering the inclusion and exclusion criteria above defined.

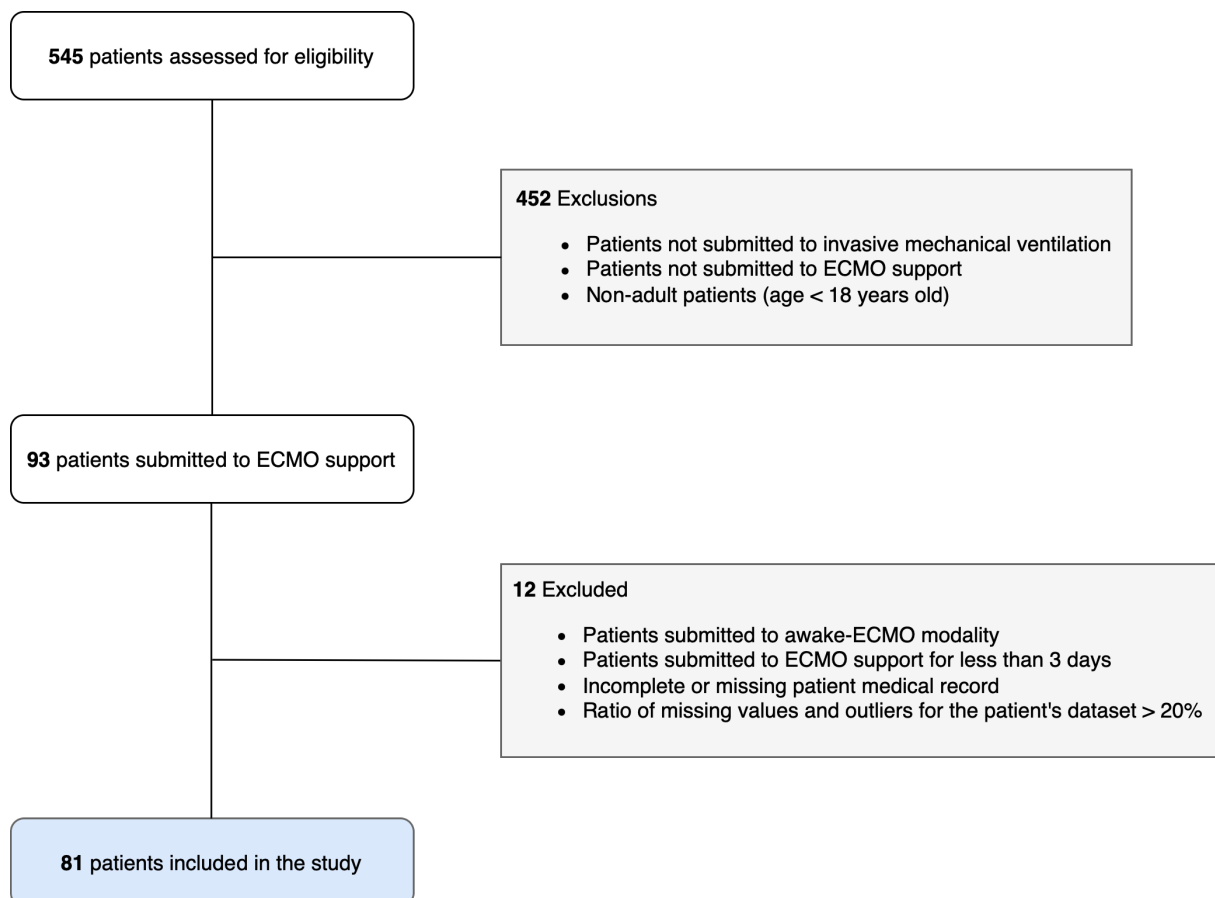


Figure 4.1: Flowchart illustrating the criteria-based patient selection process and indicating the number of patients included and excluded after each step.

4.2 Data Acquisition

As mentioned in section 1.2, this retrospective and non-interventional study was carried out in strict collaboration with the Intensive Care Department at Hospital de Santa Maria. Due to its nature and context, a detailed protocol was initially submitted to the Academic Center Ethics Committee, addressing several relevant procedural intricacies, mainly focusing on data acquisition and utilization. Submission review and further discussions culminated in the study's protocol approval: **Code Reference 175/22**.

The ICU incorporates multiple cutting-edge devices, such as ECMO, which provide physicians with critical data and insights into the patient's clinical condition and progression. These instruments monitor multiple variables, generating extensive and comprehensive datasets typically stored within the ICU's database.

Data acquisition entailed direct access to the database of the ICU at Hospital de Santa Maria. Multiple anonymization protocols were rigorously applied to eliminate protected personal and sensitive information, ensuring compliance with relevant guidelines and regulations.

Within the ICU at Hospital de Santa Maria, data from patients undergoing ECMO support is automatically transferred to the unit's database daily, adhering to established protocols. The frequency of these transfers varies based on the monitored variables, occurring at intervals of 4, 8, 12, and 24 hours.

For the selected patient cohort, multiple patient-specific data tables (illustrative example available in figure 4.2) were generated daily, providing critical information for physicians to conduct thorough retrospective assessments of each patient's clinical condition. Consequently, several data files spanning the entire hospitalization period were available for each patient. Since these files were easily extractable, physicians made them available without requiring accessing the unit's database.

Most variables considered in subsequent analyses were directly sourced from these files. However, specific parameters, such as patient comorbidities, required manual extraction from the ICU's information system, known as *Picis* [39], represented in figure 4.3. Due to the intricate nature of this system, prior training was received to gain familiarity, culminating in the acquisition of access credentials, crucial for assessing required information.

Based on the experience of physicians regarding the management and treatment of these patients, the data collected for each patient was confined to the initial **12 days** of hospitalization under ECMO support, a period deemed crucial in outcome assessment, based on the consensus that the patient's clinical trajectory during this period serves as a robust prognostic determinant. Specifically, a positive clinical trajectory within this period often signals a favorable prognosis, whereas a negative trajectory suggests a heightened likelihood of future clinical deterioration. For patients whose length of stay under ECMO support was less than 12 days, the corresponding dataset encompassed the entire hospitalization period under these conditions.

Voltar Menu inicial	Data	7/12/2020					
	Dia	2					
	Hora	0h	4h	8h	12h	16h	20h
ECMO/PECLA	Fio2	100.00	100.00	100.00	100.00	100.00	100.00
	RPM	2700.00	2700.00	2700.00	2700.00	2500.00	2500.00
	LPM	3.56	3.58	3.59	3.50	3.10	3.17
	LO2/min	5.00	5.00	5.00	5.00	5.00	5.00
	Tsistema	36.50	36.50	36.50	36.50	36.50	36.50
Dados Doente	Tdoente	36.10	36.30	36.40	36.00	36.00	36.20
	D. Cardiaco	5.80	5.60	5.50	6.10	5.80	5.90
Pressões Filtro	P. Pré Filtro			129.00			115.00
	P. Pós Filtro			110.00			99.00
	PTM			19.00			16.00
G. Membrana	HBG (gases membrana)			9.40			9.20
	Po2 Pré Filtro			38.90			38.20
	Po2 Pós Filtro			461.00			451.00
	Sato2 Pré Filtro			76.60			74.70
	Sato2 Pós Filtro			98.10			98.10
Análises	Hgb			9.20			9.00
	plaquetas			204.00			201.00
	dímeros			1.27			
	fibriogenio			556.00			
	aPTT	45.20	41.90	41.40	44.50	42.40	45.40
G.Arterial	pH	7.52	7.53	7.52	7.47	7.46	7.47
	Paco2	47.30	45.60	46.60	48.70	50.70	49.20
	Pao2	87.40	93.90	103.00	96.30	93.80	84.70
	HCO3	37.60	37.30	37.00	34.20	34.00	34.00
	SO2	97.20	97.50	97.90	97.10	96.90	96.40
	K+	4.10	4.00	3.70	3.70	3.80	3.80
	Na+	148.00	148.00	148.00	148.00	148.00	147.00
	Ca2+	1.18	1.20	1.17	1.18	1.21	1.15
	Glicemia	169.00	149.00	141.00	175.00	158.00	117.00
	Lactatos	12.00	12.00	11.00	10.00	12.00	8.00
Oximetria	Spo2	1.00	100.00	97.00	96.00	97.00	96.00
Ventilador	M. Ventilatorio	PC (12)	PC (12)	PC (12)	PC (12)	PC (12)	PC (12)
	Fio2	0.30	30.00	0.30	30.00	30.00	30.00
	Peep	5.00	5.00	5.00	5.00	6.00	6.00
	FR	12.00	12.00	12.00	12.00	18.00	18.00
	Vt	177.00	151.00	156.00	135.00	160.00	162.00
	Pplat	16.00	16.00	16.00	19.00	17.00	17.00
	Compliance	15.40	15.40	14.00	18.00	14.70	15.50

Figure 4.2: Data table containing information for a unique patient referring to a single day of hospitalization under ECMO support. These tables frequently contained outliers, as observed, for instance, in the measurements of FiO₂ at 00h and 08h (30.0 would be the valid measurement instead of 0.30).

Parâmetros fisiológicos		Parâmetros em Tempo Real																											
Rectal Temperature	37						37																						37
FIO2 (ECMO)		100																											
FIO2	30	30	30	30	30	30	30	30	30	30	100	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30
PTM (ECMO)		27																											
Temperatura do doente (ECMO)		36.5																											
Temperatura (Reg. Manual)		37.1					37.0		36.9			37.0		36.0	36.2														37.1
Glicemia Capilar (Reg. Manual)		115									105																		
Peso (Reg. Manual)											114.0																		
Pulso	97	94	85	89	82	88	63	80	74	101	132	79	77	77	78	75	88	90	102	74	80								
Frequência Cardíaca	96	96	83	81	82	89	57	80	74	101	131	73	84	70	80	75	90	89	74	72	82								
Pre. Art. Sistólica - Inv	138	137	103	145	103	151	164	104	112	300	177	131	138	124	145	147	136	108	106	110	172								
Pre. Art. Diastólica - Inv	63	60	50	69	45	76	83	55	60	159	102	76	79	63	79	80	70	53	54	56	82								
Pre. Art. Média - Inv	83	80	65	88	60	99	104	68	75	210	128	93	97	86	100	100	87	68	67	68	107								
Modo Ventilatório	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC								
SpO2	88	86	91	90	93	94	97	94	94	72	84	96	93	94	98	95	94	97	100	100	88								
Rotações/min (ECMO)		3750																											
Litros/min (ECMO)		5.18																											
Litros O2/min (fluxo gás ECMO)		8																											
Temperatura do sistema (ECMO)		36.5																											
Frequência Respiratória	24	27	17	21	18	22	17	14	14	14	15	12	12	12	12	12	12	12	12	12	12								
Vol Corrente Inspiratório	141	97	117	90	126	133	99	95	75	44	65	54	50	48	44	40	41	40	41	46	57								
Volume Mínimo	2	2	2	2	2	2	1	1	1	1	1	1	1	1	0	1	0	0	0	0	1								
Pressão Pausa	23	28		26				200		24				23	23														
Pressão Controlada	12	12	12	12	12	12	12	12	12	14	14	12	12	12	12	12	12	12	12	12	12								
FIO2	30	30	30	30	30	30	30	30	30	100	30	30	30	30	30	30	30	30	30	30	30								
Relação I/E	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0	1:2.0								
Pressão do Cuff		30						30																					
Nível do tubo traqueal		22						22																					
PEEP	12	11	11	10	11	11	11	11	12	12	10	12	11	11	12	12	12	12	12	12	12								
Vol Corrente Expiratório	104	77	104	83	87	100	67	90	72	38	66	52	48	44	40	39	39	38	39	37	46								
ECOD2								0																					
Vent Resp Rate/min Settin	14	14	14	14	14	14	14	14	14	14	12	12	12	12	12	12	12	12	12	12	12								
PEPset	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12								
STI	-0	-0	0	-0	-0	-0	-0	0	0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	0	0								
Exp. airway pressure	11	11	11	12	12	10	27	12	12	12	12	14	12	12	16	12	24	25	12	21	24								
Inspiratory Minute Volume	3	3	2	2	2	3	2	1	1	1	1	1	1	1	1	1	1	0	0	1	1								
PIP	26	26	26	27	26	25	25	26	24	26	26	24	24	24	24	24	24	24	24	24	24								
Barometric Pressure	1011	1011	1011	1011	1011	1011	1011	1011	1011	1011	1011	1010	1010	1010	1010	1010	1011	1011	1011	1012	1012								
Leakage Volume	18	16	10	5	24	20	7	8	5	5	3	4	6	8	5	8	7	6	4	11	8								
E6 min								0																					
Tempo Subida Inspiratória	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5								
Inspiratory Time, sec	1.4299	1.4299	1.4299	1.4299	1.4299	1.4299	1.4299	1.4299	1.4299	1.4299	1.6699	1.6699	1.6699	1.6699	1.6699	1.6699	1.6699	1.6699	1.6699	1.6699	1.6699								
sec Pulse Rate BP Tracing	96	96	83	82	81	88	58	79	74	101	130	68	78	92	81	76	90	89	74	69	79								
Índice Bispectral		37						41																					

Figure 4.3: Data stored within the *Picis* system for a single patient undergoing ECMO support.

4.3 Data Preparation

The data collected from multiple patients in this study underwent additional preparation to meet the study's objectives and ensure structural compatibility with subsequent analyses. As discussed in section 4.3.1, the complexity and comprehensiveness of the patient datasets introduced the need to filter the data, excluding multiple variables, leaving the remaining ones as candidate predictor variables. As the data was intended for training a supervised learning algorithm, the necessity to establish an outcome variable to label the patient datasets arose. Section 4.3.2 offers a detailed overview of the methodology applied to build this variable, elucidating the resulting outcomes.

4.3.1 Candidate Predictor Variables

After collecting information and gathering data from the multiple patient data files, the preliminary patient datasets were transformed into MTS, ensuring structural compatibility with subsequent analyses and modeling techniques. At the same time, the datasets' temporal properties were preserved, complying with the assumption that this was critical to optimize overall performance and enhance results throughout the study.

The extreme comprehensiveness of the MTSs obtained led to the exclusion of multiple variables associated with low anticipated clinical value, defined as variables that do not contribute to increasing the physicians' ability to accurately assess the patient's clinical condition nor present with any correlation with the clinical situation of the patients within the study cohort. Additionally, the exclusion also encompassed variables that could introduce bias in the study, including those dependent on medical judgment and intervention (i.e., indicators of noradrenaline, antibiotics, sedation, and muscle relaxants administration). This process was conducted in cooperation with physicians, aiming to simplify the patients' MTSs and, consequently, subsequent analyses.

The resulting patient-specific MTSs comprised a total of **41** variables, corresponding to the study's candidate predictor variables, mapped into the following high-level clinical categories:

- **Demographics and Comorbidities:** encompasses demographic information and indicators of the presence or absence of specific underlying health conditions (comorbidities).
- **Immediate Pre-ECMO Condition:** provides insights into the patient's overall clinical condition immediately before ECMO implementation.
- **Clinical Data:** includes a broad spectrum of patient vital signs and physiological measurements.
- **ECMO Physiology:** consists of parameters directly retrieved from the ECMO device, providing information regarding its configuration and functional status.
- **Integrated Patient Physiology:** offers a comprehensive overview of the patient's circulatory and respiratory status.
- **Native Lung Physiology:** provides insights into the physiological and functional condition of the patient's native lung.

- **Laboratory Data:** includes the results of laboratory tests conducted throughout hospitalization.

Table 4.1 presents the candidate predictor variables that fall into each high-level clinical category.

Table 4.1: High-level clinical categories and corresponding candidate predictor variables. *Categorical variables. **Binary presence variables (0 and 1 indicating absence and presence, respectively).

High-Level Clinical Category	Candidate Predictor Variables
Demographics and Comorbidities	Age, Sex*, Diabetes**, Hypertension**, Obesity Degree*.
Immediate Pre-ECMO Condition	Infection**, SAPS II, Murray Score.
Clinical Data	Body Temperature, Heart Rate, Non-VAP Infection**, VAP**.
ECMO Physiology	Air Flow, O ₂ Concentration, Rotations, ECMO Pump Flow, Delta Pressure, PaO ₂ Pre-Filter, PaO ₂ Post-Filter, SatO ₂ Pre-Filter, SatO ₂ Post-Filter.
Integrated Patient Physiology	Cardiac Output, Lactate, pH, Arterial PaO ₂ , Arterial PaCO ₂ , Arterial HCO ₃ ⁻ , Arterial SatO ₂ , Plateau Pressure, Lung Compliance, Renal SOFA, Respiratory SOFA, SOFA.
Native Lung Physiology	Ventilatory Mode*, FiO ₂ , PEEP, Respiratory Rate, Tidal Volume.
Laboratory Data	Hemoglobin, Platelets, Sodium (Na ⁺).

The candidate predictor variables within the MTSs present with distinct temporal behaviors. Thus, to facilitate subsequent data processing, these were divided into the following temporal categories:

- **Static (Time-Independent):** the value of these variables remains constant throughout hospitalization, and they were expanded to ensure correct integration within the MTSs.
- **Dynamic (Temporal):** the value of these variables changes throughout hospitalization.

Table 4.2 presents the candidate predictor variables that fall into each temporal category.

Table 4.2: Temporal categories and corresponding candidate predictor variables.

Temporal Category	Candidate Predictor Variables
Static (Time-Independent)	Age, Sex, Obesity Degree, Diabetes, Hypertension, Infection, SAPS II, Murray Score.
Dynamic (Temporal)	Body Temperature, Heart Rate, VAP, Non-VAP Infection, Air Flow, O ₂ Concentration, Rotations, ECMO Pump Flow, Delta Pressure, PaO ₂ Pre-Filter, PaO ₂ Post-Filter, SatO ₂ Pre-Filter, SatO ₂ Post-Filter, Cardiac Output, Lactate, pH, Arterial PaO ₂ , Arterial PaCO ₂ , Arterial HCO ₃ ⁻ , Arterial SatO ₂ , Plateau Pressure, Lung Compliance, Renal SOFA, Respiratory SOFA, SOFA, Ventilatory Mode, FiO ₂ , PEEP, Respiratory Rate, Tidal Volume, Hemoglobin, Platelets, Sodium (Na ⁺).

The first phase of data preparation has resulted in the generation of 81 patient-specific unlabeled MTSs spanning the first 12 days of hospitalization under ECMO support or the complete hospitalization period in case the patient recovered or died before reaching the 12th day of hospitalization. As mentioned earlier, since the goal was to train a supervised learning algorithm, labeling the MTSs became imperative. The following section provides a detailed overview of the characteristics of the established outcome variable and the corresponding data labeling strategy.

4.3.2 Outcome

The outcome variable and corresponding data labeling strategy markedly differ from those outlined in the literature, constituting an innovative element of this study.

Section 3.1.3 demonstrates that prior research referring to the application of ML within critical care predominantly focused on assessing well-defined outcomes, typically characterized by straightforward defining criteria. For instance, Liu et al. [29] and Hyland et al. [31] consider the onset of sepsis, diagnosed using the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [40], and the onset of circulatory failure, diagnosed based on the assessment of 3 types of variables: lactate (arterial and venous), mean arterial pressure, and presence of vasoactive drugs, as the primary outcomes of their studies, respectively.

Contrary to this trend, the outcome variable of the current study aims to capture the intricate real-time patient-specific clinical dynamics spanning the complete hospitalization period considered. To achieve this goal, a binary variable was initially defined. However, the complexity and uncertainty associated with accurately evaluating the patients' clinical condition in different instances introduced the need for an alternative strategy. As a result, a ternary variable had to be established, allowing the inclusion of periods posing heightened challenges to physicians in making precise clinical inferences. The three scenarios represented by this ternary variable, provided below, are derived from the physicians' assessment and interpretation of the patient's clinical dynamics at specific time instants of the corresponding MTSs.

Table 4.3: Description of each possible value of the established ternary outcome variable.

Outcome	Description
-1	Based on past and current data analysis, physicians infer that the patient's clinical condition is deteriorating.
0	Based on past and current data analysis, physicians are unable to confidently infer the patient's clinical trajectory, although it remains plausible that clinical developments may be unfolding.
+1	Based on past and current data analysis, physicians infer that the patient's clinical condition is improving.

The employed data labeling strategy aimed to replicate real-world conditions, where access to past and present information is exclusive, and knowledge of the patient's clinical evolution and outcome is restricted. Consequently, the multiple patient-specific MTSs were individually presented to physicians for blinded analysis, culminating in the assignment of an outcome value to every time instant within the

respective MTS. The visual depiction of this process is elucidated in figure 4.4, which simulates the comprehensive assessment conducted by physicians on a single patient's MTS.

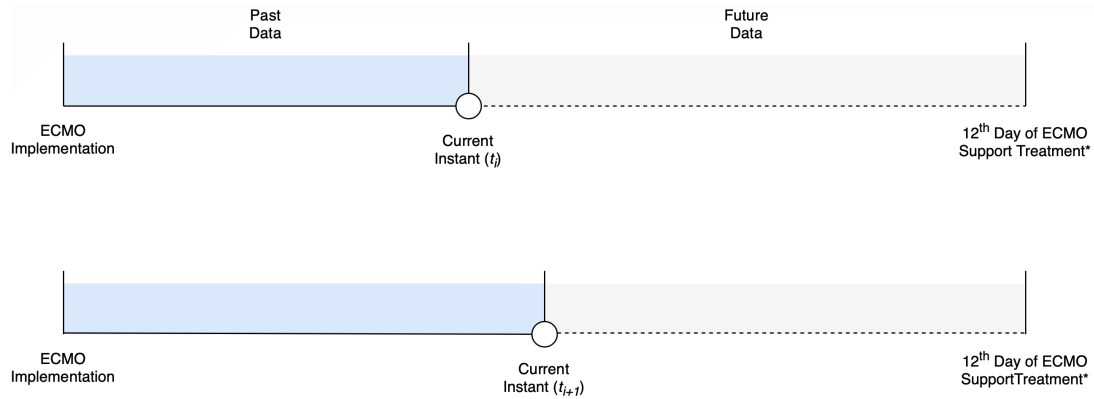


Figure 4.4: Visualization depicting the data labeling process: the blue region denotes both past and present information available to physicians for analysis, guiding outcome assignment decisions for the current instant; meanwhile, the gray region reflects forthcoming data lying beyond current knowledge or access, thus inaccessible to physicians at the current instant. *For patients with a length of stay under ECMO support of less than 12 days, data labeling was conducted for the entire hospitalization period under these conditions.

After applying the strategy depicted in figure 4.4 to label the 81 patient-specific MTSs considering the established ternary outcome variable, the results illustrated in table 4.4 were achieved.

Table 4.4: Results obtained by applying the data labeling strategy to the multiple patients' MTSs.

Outcome	Total instances	Patients with instances exclusively labeled with outcome class
-1	641 (12.11%)	2 (2.47%)
0	3583 (67.72%)	14 (17.28%)
+1	1067 (20.17%)	0 (0.0%)

The results outlined in table 4.4 underscore a prevalent trend: most instances across the multiple patient MTSs were labeled with outcome class 0. These findings highlight the complexity and uncertainty inherent in evaluating these patients' clinical dynamics, emphasizing the need for integrating more advanced, automated systems in the ICU (especially in the context of ECMO support). By leveraging this integration, physicians gain invaluable assistance in assessing the intricacies of patient care, enhancing their ability to make informed decisions, and ultimately impacting the quality and efficacy of care.

4.4 Exploratory Data Analysis

Even though the study population's size was relatively limited, its complexity was significant, necessitating the implementation of a robust exploratory data analysis framework. The first step focused on generating informative insights regarding the patient cohort, providing a more comprehensive picture of patient characteristics (section 4.4.1). The second step involved assessing data quality to avoid error propagation (e.g., derived from inconsistencies associated with data acquisition and registration) and

thus contamination of subsequent analyses, ultimately culminating in the removal of variables that were not compliant with minimum quality standards (section 4.4.2).

4.4.1 Patient Cohort: Descriptive Statistics and General Characteristics

Despite the small number of patients in this study, the differences between individuals ensured that the population remained highly comprehensive and representative. As described in section 4.1, this results from the implemented criteria-based patient selection process, which aimed to produce a relatively uniform and homogeneous population while ensuring that patients exhibited distinct physiological characteristics and clinical manifestations, reflecting real-world diversity.

To effectively describe the patient cohort, multiple key variables were assessed, including **Age**, **Sex**, **Comorbidities** (Diabetes, Hypertension, and Infection), **ECMO support duration**, and **Mortality**. These were selected since they serve as accurate descriptors that enable the characterization of the study population. Table 4.5 provides descriptive statistics computed using the selected variables, including a brief description to support the analysis.

Table 4.5: Patient cohort descriptive statistics. *Excludes Sars-CoV-2 infection. SD: standard deviation.

Variable	Description	Statistics
Age	Patient's age	Mean: 49.4, Median: 52.0, SD: 12.0, Range: 21-74
Sex	Patient's sex	Male: 56 (69%), Female: 25 (31%)
Diabetes	Presence of diabetes	Yes: 11 (14%), No: 70 (86%)
Hypertension	Presence of hypertension	Yes: 28 (35%), No: 53 (65%)
Infection	Presence of infection*	Yes: 16 (20%), No: 65 (80%)
ECMO Duration	Length of stay subject to ECMO	Mean: 30.5, Median: 26.0, SD: 25.3, Range: 5-193
Mortality	Patient's survival outcome	Survival: 65 (80%), Death: 16 (20%)

Table 4.5 presents valuable information regarding the study's patient cohort. The mean age was 49.4 years old, contradicting the widespread perception that patients with COVID-19 pneumonia requiring ECMO support were predominantly elderly. Most patients were male (69%), and a relatively high incidence of comorbidities was observed, with diabetes, hypertension, and infection affecting 14%, 35%, and 20% of individuals, respectively. The duration of ECMO support varied significantly, with a mean of 30.5 days, reflecting the uncertainty and volatility inherent in this therapeutic intervention. Additionally, the mortality rate reached 20%, which, despite not seeming excessively high, is substantial considering the cohort's mean age, thus underscoring the critical condition that characterizes patients undergoing ECMO support, irrespective of underlying health issues, age, or other factors.

4.4.2 Data Quality Assessment

As outlined in section 4.2, the data from patients undergoing ECMO support is automatically transferred and recorded in the ICU database. However, this process is occasionally inconsistent, requiring direct human intervention, inevitably increasing susceptibility to errors. Consequently, a robust methodology for assessing data quality was implemented, focusing on identifying and filtering out potentially compromising factors.

Given the patient cohort's limited dimension, the strategic priority was to minimize patient exclusions. Thus, the methodology was designed based on the assumption that the negative effect associated with eliminating variables from the patients' MTSs would be minimal compared to that observed in the case of removing patients. As a result, the strategy applied focused on individually assessing the quality of the different variables across the multiple patient-specific MTSs, seeking to identify the most compromising, therefore limiting the potential of reverse causation affecting model predictions in a subsequent phase.

The methodology's first step involved identifying variables associated with a percentage of missing values and outliers exceeding 30% for at least one patient. This threshold was determined based on combining approaches observed across reviewed studies outlined in section 3.1.3 (Noy et al. [28] and Garcia-Gutiérrez et al. [30] consider a threshold of 40% and 25% for variable removal, respectively) and input provided by physicians. The result was the identification of **13** potentially compromising variables: Non-VAP Infection, Air Flow, Rotations, Delta Pressure, PaO₂ Post-Filter, Plateau Pressure, Lung Compliance, Renal SOFA, Respiratory SOFA, SOFA, FiO₂, PEEP, and Tidal Volume.

These variables underwent additional scrutiny through a straightforward methodology comprising two sequential steps. Initially, attention was directed towards identifying and removing particularly anomalous variables. Subsequently, critical variables failing to meet the established minimum quality standards were also identified and removed. Additional details are listed below:

1. Variables that were entirely missing or incorrect for at least one patient and deemed irrecoverable were selected for removal. Figure 4.5 illustrates the frequency with which this scenario occurs for each of the 13 variables initially identified. Following assessment, **4** variables were removed from the multiple patient-specific MTSs: **Delta Pressure, Renal SOFA, Respiratory SOFA, and SOFA.**
2. The remaining 9 variables from the initial group of 13 were then evaluated for overall quality, with the following **2** variables being removed from the multiple patient-specific MTSs due to the abnormally high frequency with which the corresponding ratio of missing values and outliers exceeds 50%, as illustrated in figure 4.6: **Plateau Pressure and Lung Compliance.**

In conclusion, the data quality assessment methodology resulted in the removal of **6** variables from the multiple patient-specific MTSs, including Delta Pressure, Renal SOFA, Respiratory SOFA, SOFA, Plateau Pressure, and Lung Compliance. **Consequently, a total of 81 patient MTSs comprising 35 variables were retained for subsequent analyses.**

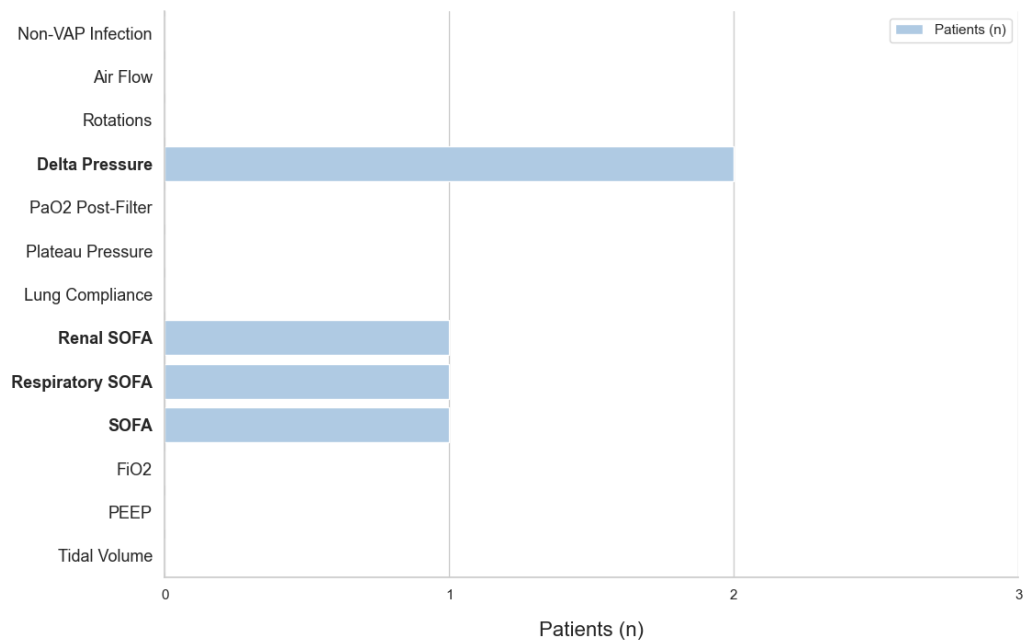


Figure 4.5: Count of patients per variable for which no valid measurements are available.

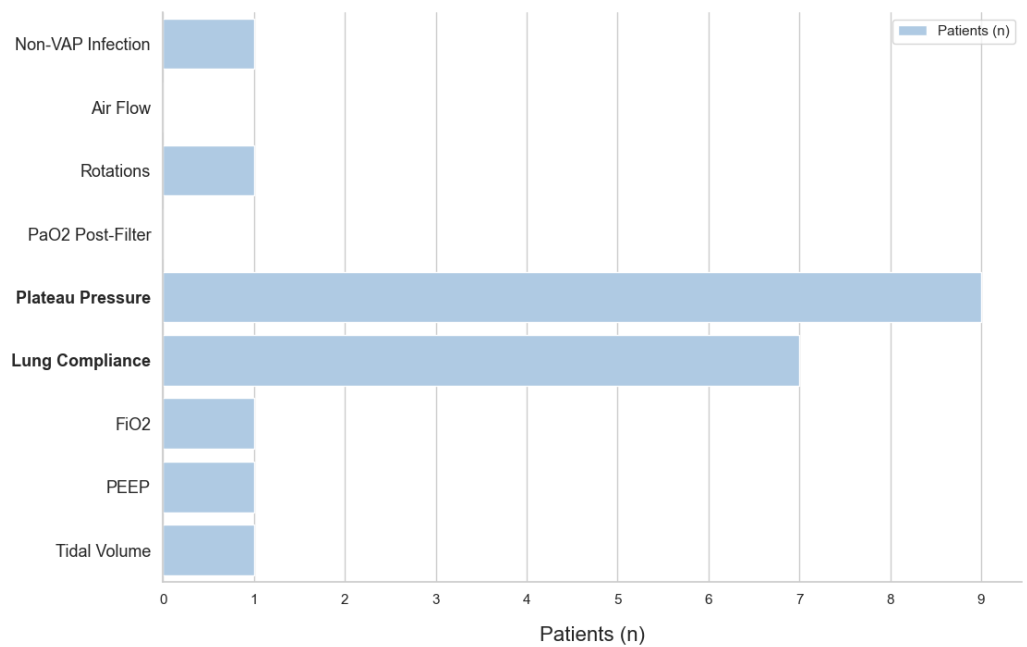


Figure 4.6: Count of patients per variable for which invalid measurements (either missing or incorrect) surpass 50%.

Chapter 5

Research Methodology

The research methodology implemented in this study involved processing and modeling multidimensional and multidomain datasets obtained through the protocolized monitoring system applied to all patients undergoing ECMO support in the ICU at Hospital de Santa Maria. Considering the study's objectives, the following two-phase design was established:

- **Phase 1:** involved processing multiple patient datasets appropriately to enable the development of a well-performing, calibrated, and interpretable ML model capable of accurately distinguishing between clinical deterioration and improvement (section 5.1).
- **Phase 2:** involved computing a risk score, using the ML model outputted through the preceding phase, that provides a real-time estimate of the risk of clinical deterioration throughout hospitalization under ECMO support for each patient, and its evaluation through qualitative and quantitative methods (section 5.2).

5.1 Phase 1: Development of a Machine Learning Model for Classification of Clinical Deterioration and Improvement

This phase involved implementing a ML pipeline encompassing the following steps:

1. **Data Preprocessing:** ensure the quality and consistency of the multiple patient MTSs (e.g., through data imputation and outlier detection and removal) (section 5.1.1).
2. **Feature Engineering:** represent the several patient MTSs in a modeling-suitable format that preserves the datasets' temporal properties, thereby enhancing the ML models' ability to extract relevant insights (e.g., patterns and trends) from the data (section 5.1.2).
3. **Model Training and Hyperparameter Optimization:** involves the supervised training and hyperparameter optimization of ML models (SVM and Random Forest) to distinguish between clinical deterioration and improvement (section 5.1.3).

4. **Model Evaluation and Selection:** evaluate the performance of the different ML models through performance metrics' computation and calibration assessment (section 5.1.4).
5. **Model Interpretability:** produce valuable insights into the key features that influence the models' predictions, enabling interpretability and ultimately enhancing applicability (section 5.1.5).

5.1.1 Data Preprocessing

Each patient's MTS comprised manually registered information and data automatically transferred to the ICU's information system, as described in section 4.2. Direct human intervention in the data acquisition and registration process increases susceptibility to errors. Consequently, implementing a robust data preprocessing pipeline (figure 5.1) to prevent error propagation and contamination of subsequent processes was essential to ensure the reliability of further analyses.

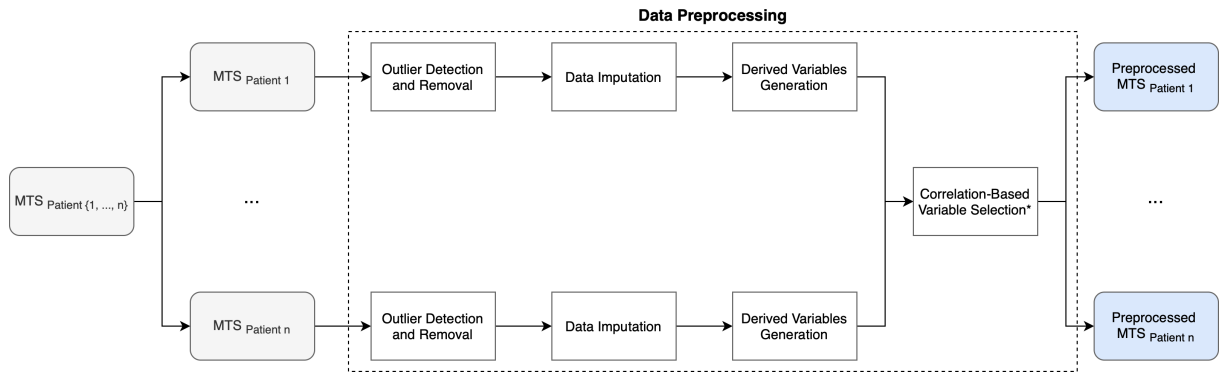


Figure 5.1: Flowchart illustrating the data preprocessing pipeline. * Represented as a single block since the multiple patient-specific correlation matrices are aggregated and assessed jointly (this process is described in detail below). MTS: Multivariate Time Series.

Outlier Detection and Removal

The volatility that characterizes the patient cohort prompted the decision to implement a rules-based approach for outlier detection rather than other commonly considered methods (i.e., statistical-based, distance-based, clustering-based, and density-based [41]).

Consequently, to screen for grossly incorrect measurements resulting from the data acquisition and registration process, a range of valid values per each **continuous numerical variable**, both static (time-independent) and dynamic (temporal), was established through a collaborative process involving physicians. This approach enabled the inclusion of extreme, pathological values that would typically be flagged as outliers but could represent natural occurrences within the patient cohort. At the same time, it also guaranteed the exclusion of erroneous data points.

Table 5.1 illustrates the set of rules established for the continuous numerical variables.

Categorical variables, both static (time-independent) and dynamic (temporal), underwent a straightforward verification process to ensure that each value corresponded to one of the predefined categories. Table 5.2 depicts the available categories for a specific subset of categorical variables.

Table 5.1: Minimum and maximum accepted value for each of the continuous numerical variables (static and dynamic). *beats per minute; **rotations per minute; *** cycles per minute.

Variable	Minimum	Maximum	Unit of Measurement
Age	18	-	-
SAPS II	0	163	-
Murray Score	0	4	-
Body Temperature	32.0	42.0	° C
Heart Rate	20	220	bpm*
Air Flow	0	15	LO ₂ /min
O ₂ Concentration	0.21	1.0	%
Rotations	500	5,000	rpm**
ECMO Pump Flow	0.4	7.0	L/min
PaO ₂ Pre-Filter	20	100	mmHg
PaO ₂ Post-Filter	50	500	mmHg
SatO ₂ Pre-Filter	0.3	1.0	%
SatO ₂ Post-Filter	0.4	1.0	%
Cardiac Output	1.0	15.0	L/min
Lactate	1.0	200.0	mg/dL
pH	6.5	8.0	-
Arterial PaO ₂	40.0	500.0	mmHg
Arterial PaCO ₂	25.0	150.0	mmHg
Arterial HCO ₃ ⁻	2.0	45	mmol/L
Arterial SatO ₂	0.21	1.0	%
FiO ₂	0.21	1.0	%
PEEP	0.0	30.0	cmH ₂ O
Respiratory Rate	5	40	cpm***
Tidal Volume	1.0	1,000.0	mL/cycle
Hemoglobin	4.0	17.0	g/dL
Platelets	1.0	1,000.0	cel/ μ L $\times 10^3$
Sodium (Na ⁺)	110.0	180.0	mmol/L

Table 5.2: Available categories for selected categorical variables (excluding binary presence variables). *The categories for variable Obesity Degree were established based on guidelines and directives provided by the U.S. Centers For Disease Control and Prevention [42].

Variable	Categories
Sex	M (Male), F (Female)
Obesity Degree*	0 (Non-Obese), 1 (Class 1 Obesity), 2 (Class 2 Obesity), 3 (Class 3 Obesity)
Ventilatory Mode	ESP (Spontaneous), PS (Pressure Support), PC (Pressure Control), VC (Volume Control)

The remaining categorical variables (not depicted in table 5.2) were binary indicators reflecting the presence or absence of specific clinical conditions (e.g., Diabetes, Hypertension, Infection, VAP, Non-VAP Infection). Outlier detection for these variables was equally straightforward, verifying that values confirmed the valid categories: 0 (absence) and 1 (presence).

Data Imputation

Similarly to outlier detection, the data imputation strategies employed were defined based on the specific characteristics of each variable, as outlined below:

- **Static (time-independent) variables:** missing data for these variables was manually recovered from the ICU's information system. Binary presence variables Diabetes, Hypertension, and Infection, were included in this category as they represent conditions that remain constant throughout the patient's hospitalization under ECMO support.
- **Dynamic (temporal) variables:** imputation strategies for dynamic variables differed based on whether the variable was categorical or continuous. Binary indicators for the presence of VAP and Non-VAP infection were treated as categorical variables.
 - **Categorical variables:** in line with the approach employed by Tomašev et al. [27] for imputing missing values in the dataset's ground-truth labels, these variables underwent backward and forward filling within time windows of up to three days. Specifically, the last value available within a time window of less than three days was copied forward in time until a subsequent measurement was available, assuming it fell within the 3-day window. If the measurement corresponding to the first time instant was missing, the same rule applied but with backward filling, where the first available value was copied backward in time, under the condition that it belonged to the 3-day window. Beyond the 3-day mark, missing values were not imputed but instead classified as unknown, preventing outdated, potentially misleading information from being incorporated into the patient MTSSs.
 - **Continuous numerical variables:** similarly to the approach implemented by Noy et al. [28], these variables were imputed considering the multivariate Iterative Imputer algorithm, which draws inspiration from Multivariate Imputation Chain Equation [43]. This algorithm applies regression to predict missing values for each variable by considering other variables iteratively. In each iteration, missing values for all variables are estimated sequentially, generating a dataset at each step. This process continues until convergence, with the final iteration yielding the imputed dataset.

Derived Variables Computation

The computation of additional variables (termed **derived variables**), established cooperatively with physicians, aspired to augment and enrich the patient MTSSs by enabling the inclusion of high-value data (i.e., data that captures key physiological indicators critical for assessing the clinical condition of patients undergoing ECMO support). These variables are relatively complex and scarcely represented in the existing literature (as they are exclusive to ECMO support), supporting the idea that their inclusion might allow the ML models to access information that reinforces their ability to capture relevant clinical dynamics.

The mathematical formulations used to compute the derived variables represented via equations 5.1 to 5.9 (where $\Delta Filter\ O_2\ Content(t) = Post-Filter\ O_2\ Content(t) - Pre-Filter\ O_2\ Content(t)$), are closely aligned with the physiological principles discussed within section 2.2, in which the meaning and relevance of each variable in these equations is clearly defined.

$$Pre-Filter\ O_2\ Content(t) = 1.34 \cdot Hemoglobin(t) \cdot Pre-Filter\ SatO_2(t) + 0.03 \cdot Pre-Filter\ PaO_2(t) \quad (5.1)$$

$$Post-Filter\ O_2\ Content(t) = 1.34 \cdot Hemoglobin(t) \cdot Post-Filter\ SatO_2(t) + 0.03 \cdot Post-Filter\ PaO_2(t) \quad (5.2)$$

$$ECMO\ O_2\ Transfer(t) = ECMO\ Pump\ Flow(t) \cdot \Delta Filter\ O_2\ Content(t) \quad (5.3)$$

$$Arterial\ O_2\ Content(t) = 1.34 \cdot Hemoglobin(t) \cdot Arterial\ SatO_2(t) + 0.03 \cdot Arterial\ PaO_2(t) \quad (5.4)$$

$$Arterial\ O_2\ Delivery(t) = Cardiac\ Output(t) \cdot Arterial\ O_2\ Content(t) \quad (5.5)$$

$$Venous\ O_2\ Return(t) = Cardiac\ Output(t) \cdot Pre-Filter\ O_2\ Content(t) \quad (5.6)$$

$$Body\ O_2\ Consumption(t) = Arterial\ O_2\ Delivery(t) - Venous\ O_2\ Return(t) \quad (5.7)$$

$$Native\ Lung\ O_2\ Transfer(t) = Body\ O_2\ Consumption(t) - ECMO\ O_2\ Transfer(t) \quad (5.8)$$

$$\Delta Tidal\ Volume(t) = Tidal\ Volume(t) - Tidal\ Volume(0) \quad (5.9)$$

Correlation-Based Variable Selection

The datasets' multidimensional and multidomain nature, which enhances the likelihood of highly related variables with non-linear associations, emphasizes the need to conduct a robust examination of underlying relationships and dependencies within the data.

As a result, the correlation between all the datasets' eligible variables was assessed. As described by Papana [44], several correlation measures are available, including the Pearson correlation coefficient (accounts exclusively for the variables' linear relationships and is sensitive to data distributions) and

Spearman's rank correlation coefficient (alternative non-parametric correlation characterized by a notable ability to capture monotonic relationships within the data). Based on the datasets' characteristics, **Spearman's rank correlation coefficient was selected as the primary correlation measure.**

Spearman's rank correlation was applied to compute correlation matrices for the multiple patient MTSs. As the main objective of this analysis focused on assessing dynamic changes and patterns over time, static (time-independent) variables, categorical (including binary presence) variables, and variables with zero variance for at least one of the several patient MTSs were excluded. The resulting matrices were then aggregated into a single matrix by averaging their values, providing a proxy for the correlations identified across the multiple patient MTSs. The final matrix was ultimately considered to filter out highly correlated variables.

5.1.2 Feature Engineering

After completing data preprocessing, the multiple patient MTSs were subject to the application of a robust feature engineering pipeline, illustrated in figure 5.2. This pipeline was designed based on the premise that preserving the datasets' temporal properties would enhance the ability to derive relevant insights through subsequent analyses.

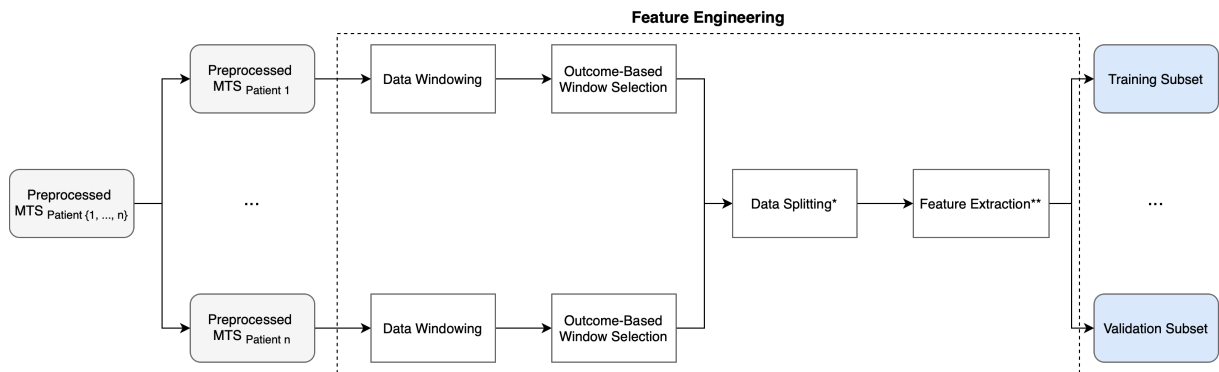


Figure 5.2: Flowchart illustrating the feature engineering pipeline. *Represented as a single block since windows selected from different patients are combined and then assigned to the training and validation subsets according to predefined criteria (further details discussed below). **Represented as a single block since the feature extraction pipeline is initially fitted onto the training subset and subsequently used to transform (i.e., extract features from) both the training and validation subsets (further details discussed below). MTS: Multivariate Time Series.

Data Windowing

The multiple preprocessed patient MTSs were initially transformed to reframe the sequential supervised learning problem into a conventional supervised learning format.

Unlike the scenario discussed in section 3.1.2, the outcome variable in this study was dynamic, exhibiting frequent fluctuations over short time intervals. As a result, different values of the outcome variable could occur within the same prediction window, where oscillations between states of clinical

deterioration (outcome class -1), uncertainty (outcome class 0), and improvement (outcome class +1) were frequent.

As previously mentioned, the patient cohort's volatility introduced additional difficulties (regardless of physicians' experience) in assessing the patient's clinical condition throughout hospitalization in real-time, making predicting future clinical trajectories even more challenging. As a result, to ensure greater alignment with the outcome variable's dynamic nature, slight modifications were introduced to the structure depicted in figure 3.1. The modified sliding window method, illustrated in figure 5.3, presented the characteristics listed below:

- The **observation window** has a fixed length, spanning three consecutive time points, equivalent to an 8-hour period (with each timestep having a duration of 4 hours).
- The **prediction window** is omitted, meaning that prediction time coincides with the last time point of the observation window. Consequently, the focus shifted to answering the question: "**Based on the assessment of the observation window, what is the patient's current clinical condition?**", rather than "How will the patient's clinical condition evolve within the next x hours?", where x would otherwise depend on the prediction window's length.
- The **prediction time** corresponds to the last time point of the observation window, with the outcome variable at this time point serving as the predicted value.
- The **window shift** is equivalent to a timestep, meaning that consecutive instances (in practice, observation windows) are separated by a 4-hour interval.

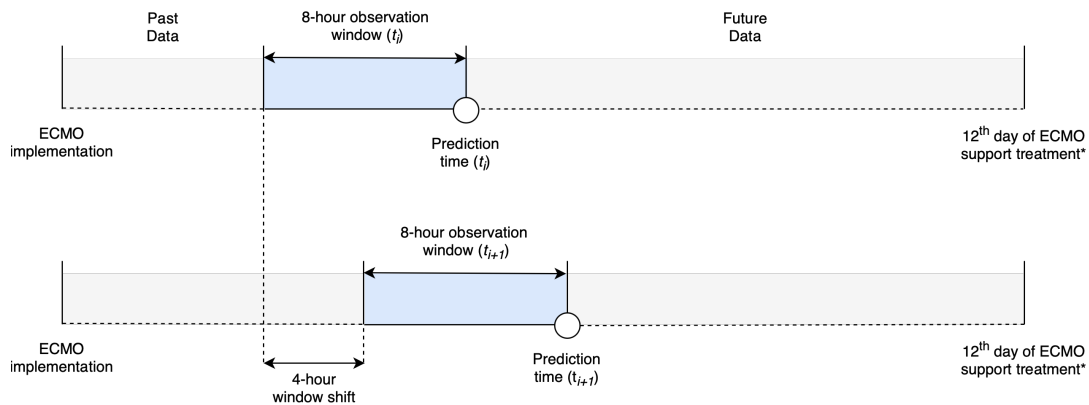


Figure 5.3: Modified sliding window method applied to convert the multiple patient MTSs into sets of labeled instances (8-hour windows). *For patients with a length of stay under ECMO support of less than 12 days, the sliding window spanned the entire period of hospitalization under ECMO support.

The application of the modified sliding window method resulted in the generation of several patient-specific datasets, each containing numerous independent windows. The data within each window was vectorized and condensed into a single row in the resulting dataset. Since each window encompassed, for each variable, measurements obtained at three consecutive time points, a suffix s was appended to the name of each dynamic (temporal) variable, where $s \in \{1, 2, 3\}$ denotes the first, middle, and

last (most recent) measurement (e.g., Heart Rate_1, Heart Rate_2, Heart Rate_3), respectively. Additionally, no suffix has been added to the name of static variables, as their value remains unchanged throughout hospitalization under ECMO support. Ultimately, the dimension of the vector computed from each window was given by equation 5.10, where n_{vector} depicts the dimension (i.e., number of elements) of the vector, $n_{dynamic}$ represents the number of dynamic (temporal) variables, and n_{static} corresponds to the number of static variables.

$$n_{vector} = 3 \cdot n_{dynamic} + n_{static} \quad (5.10)$$

Outcome-Based Window Selection

Following the execution of the previous step, multiple patient-specific datasets, each comprising numerous instances (8-hour windows), were generated. Extrapolating information available in table 4.3, one can conclude that windows labeled as 0 reflect periods where physicians could not confidently infer the patient's clinical condition (although remaining possible that clinical developments may be occurring). Therefore, to prevent the model from using data associated with medical ambivalence and uncertainty during training, these windows were excluded, resulting in the exclusive utilization of **windows labeled as -1 and +1** for subsequent steps.

Data Splitting

The strategy implemented for partitioning the numerous patient-specific datasets into training and validation subsets was designed based on the assumption that the model could accurately classify the patient's clinical condition (i.e., assign a label of -1 or +1 to indicate clinical deterioration or improvement, respectively) at a specific time point (prediction time) through exclusively assessing the corresponding observation window, thereby not requiring access to preceding data. As a result, the implemented strategy performed random assignment of patient-specific windows to the training and validation subsets, following a process based on the criteria listed below:

- The windows from the training subset for the same patient should precede those assigned to the validation subset. This precaution prevents information leakage, ensuring the training subset does not contain future data relative to the validation subset.
- The allocation of windows between the training and validation subsets should maintain the balance of outcome classes.

Garcia-Gutiérrez et al. [30] proposes a data splitting strategy that assigns all data from each patient-specific dataset to either the training or validation subset. However, this approach risks introducing imbalances across the two subsets, potentially failing to capture the diversity of the patient cohort. As a result, model learning and performance could be compromised, particularly if the model does not encounter specific patient subgroups during training. Therefore, to mitigate these risks, the windows from each patient-specific dataset were randomly assigned to the training and validation subsets, ensuring

compliance with the aforementioned criteria. This approach ensures that both subsets adequately represent the patient cohort's characteristics, allowing the model to learn patterns that reflect the cohort's inherent heterogeneity.

Ultimately, a **4:1 ratio was employed to allocate patient-specific windows labeled as -1 and +1 to the training and validation subsets.**

Feature Extraction

After splitting the data into the training and validation subsets, the next step involved transforming the data to ensure compatibility with modeling techniques. As a result, a feature extraction pipeline comprising the following two steps was designed: **scale numerical variables** and **encode categorical variables**.

Numerical variables were scaled using the `StandardScaler` method, implemented via the `scikit-learn` library [45], which standardizes the data by centering it (subtracting the mean) and scaling it to unit variance. The corresponding mathematical formulation is presented in equation 5.11, where x denotes the original value, μ represents the mean, and σ depicts the standard deviation.

$$z = \frac{x - \mu}{\sigma} \quad (5.11)$$

This method standardizes each variable to ensure a mean of 0 and a standard deviation of 1. These transformations are particularly relevant when considering ML models sensitive to feature magnitudes, preventing variables whose values are concentrated within a substantially higher range relative to the others from potentially dominating model training.

To ensure compatibility with ML models requiring numerical inputs, categorical variables within the datasets were encoded using the `OneHotEncoder` method, implemented via the `scikit-learn` library [45]. This method's underlying mathematical formulation converts a categorical variable with k unique values (categories) into k binary features (i.e., it transforms each categorical variable into a binary vector where each category is represented by a separate column containing values of 0 or 1). As an illustrative example, the result obtained from encoding the categorical variable Obesity Degree depicted in table 5.2 would be the following:

$$\bullet \{ \mathbf{0} = [1, 0, 0, 0], \mathbf{1} = [0, 1, 0, 0], \mathbf{2} = [0, 0, 1, 0], \mathbf{3} = [0, 0, 0, 1] \}.$$

The feature extraction pipeline was initially fitted to the training subset and ultimately applied to both the training and validation subsets, transforming them into sets of features suitable for modeling. Fitting the pipeline to the training subset ensured that the scaling and encoding transformations were exclusively based on the characteristics of the training subset, preventing data leakage from the validation subset (i.e., ensuring that no information from the validation subset influenced model training).

5.1.3 Model Training and Hyperparameter Optimization

Following the transformation of the training and validation subsets into sets of features, two conventional ML models were employed: **SVM**, introduced by Cortes and Vapnik [46], and **Random Forest**, developed by Breiman [47]. The selection of these models, rather than more advanced ones (e.g., neural networks), was motivated by their ability to effectively handle data of small to moderate size (neural networks typically require large datasets to perform well), which is an essential attribute given the limited available data in this study. The theoretical insights regarding both models presented below resulted from the analysis of both foundational works:

- **SVM**: it performs well in high-dimensional spaces while presenting a notable ability to find the optimal decision boundary that maximizes the margin between distinct classes, making it especially effective in handling smaller datasets. This effectiveness is enhanced by the SVM's ability to generalize well with fewer data points, namely when paired with an appropriate kernel function.
- **Random Forest**: it constructs an ensemble of decision trees, averaging their outputs, which enables the achievement of remarkable predictive performance while simultaneously preventing overfitting, which is particularly challenging when dealing with smaller datasets.

The SVM model, implemented via the scikit-learn library [45], was trained considering distinct kernel functions: linear, polynomial of degree 3, and Radial Basis Function (RBF). The key hyperparameters of the SVM model, C , which controls the strength of regularization, and γ , which depicts the kernel coefficient for the polynomial and RBF kernel functions, were assigned the default values as established in the corresponding scikit-learn implementation: $C = 1$; $\gamma = \text{scale}$. Based on this equality, the numerical value of the hyperparameter γ is determined through equation 5.12, where $n_features$ denotes the total number of features and $X.var()$ represents the variance of the input features across the dataset.

$$\gamma = \frac{1}{n_features \times X.var()} \quad (5.12)$$

The Random Forest model was also implemented using the scikit-learn library [45]. The training process involved the optimization of the hyperparameter `n_estimators`, which depicts the number of trees within the ensemble. The optimization process encompassed an **iterative search approach** that included the following steps:

1. Perform a random search across an established interval (for the first iteration, the interval is defined by randomly selecting lower and upper boundaries, e.g., $[1, 1000]$).
2. Assess model performance (details on the methods employed available in the following section) at various equally distanced points across the interval.
3. Establish a narrower search interval, allowing for a more targeted search in regions corresponding to higher model performance.

The optimization process described above was executed repeatedly until a sufficiently narrow interval was reached. At that point, model performance was assessed for multiple equally distanced points,

culminating in selecting the optimal value (i.e., the value that maximizes model performance) for hyperparameter `n_estimators`.

Ultimately, since the models will be used to construct a risk score that provides a real-time estimate of the risk of clinical deterioration throughout hospitalization under ECMO support for each patient, they require calibration, which enables the correct quantification of the level of confidence (or uncertainty) associated with instance-wise predictions, which can be interpreted as reliable probability estimates. For the SVM model, the boolean parameter `probability` (designation used in the corresponding scikit-learn implementation) was set to `True`, activating the **Platt Scaling** calibration method, which applies a logistic regression model to the SVM's decision function outputs, converting them into calibrated probabilities [48]. The transformation follows equation 5.13, where $p_{calibrated}$ represents the calibrated probability estimate, $f(x)$ depicts the model's decision function output, and A and B are scalar parameters learned during the calibration process.

$$p_{calibrated} = \frac{1}{1 + e^{(Af(x)+B)}} \quad (5.13)$$

In contrast, the Random Forest did not contain any internal parameter for activating calibration, thereby requiring an external method to be employed: `CalibratedClassifierCV` using the `sigmoid` method (implemented via the scikit-learn library [45]), which functions equivalently to Platt Scaling.

5.1.4 Model Evaluation and Selection

After fitting the ML models into the training subset, their performance was assessed using the validation subset. Considering that the problem at hand consists of a binary classification task (with outcome classes -1 and +1 representing clinical deterioration and improvement, respectively), the first step within the evaluation pipeline involved computing a confusion matrix, whose structure is depicted in figure 5.4.

	Predicted Positive (Outcome Class -1)	Predicted Negative (Outcome Class +1)
Actual Positive (Outcome Class -1)	True Positive (TP)	False Negative (FN)
Actual Negative (Outcome Class +1)	False Positive (FP)	True Negative (TN)

Figure 5.4: Structure of the confusion matrix generated to evaluate the ML models' performance in distinguishing between clinical deterioration (outcome class -1) and improvement (outcome class +1).

As illustrated in figure 5.4, the confusion matrix provides four potential outcomes: **True Positive (TP)**, which occurs when the model correctly predicts clinical deterioration (outcome class -1) for an instance (8-hour window) labeled as -1; **FP**, where the model incorrectly predicts clinical deterioration (outcome class -1) for an instance (8-hour window) that in reality is labeled as +1; **False Negative (FN)**, where the

model incorrectly predicts clinical improvement (outcome class +1) for an instance (8-hour window) that in reality is labeled as -1; **True Negative (TN)**, which occurs when the model correctly predicts clinical improvement (outcome class +1) for an instance (8-hour window) labeled as +1.

The next step in the evaluation pipeline involved computing key performance metrics derived from the confusion matrix, commonly assessed in binary classification tasks:

- **Accuracy**: represents the proportion of correctly predicted instances (8-hour windows), referring to both clinical deterioration (outcome class -1) and improvement (outcome class +1), relative to the total number of instances (equation 5.14).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.14)$$

- **Precision**: represents the proportion of instances (8-hour windows) where the model correctly predicted clinical deterioration (outcome class -1), relative to all instances classified by the model as -1 (equation 5.15).

$$Precision = \frac{TP}{TP + FP} \quad (5.15)$$

- **Recall** (also known as Sensitivity): represents the proportion of instances (8-hour windows) where the model correctly predicted clinical deterioration (outcome class -1), relative to all instances labeled as -1 (equation 5.16).

$$Recall = \frac{TP}{TP + FN} \quad (5.16)$$

- **F-Score**: harmonic mean of Precision and Recall, particularly useful for balancing the trade-off between both metrics, which is especially relevant for imbalanced datasets (i.e., it provides a more balanced measure of model performance by penalizing large discrepancies between Precision and Recall) (equation 5.17).

$$F-Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (5.17)$$

After computing performance metrics derived from the confusion matrix, the **Reliability Diagram** (also known as Calibration Curve) was generated for each model to ensure that the models' decision function outputs could be interpreted as reliable probability estimates. These diagrams were analyzed qualitatively (i.e., through visual assessment), with better calibration depicted as higher proximity between the curve corresponding to the model and the diagonal representing perfect behavior. The reliability diagrams' generation involved computing the probability of each instance (8-hour window) being classified as positive (outcome class -1), denoted as $p(y_i = -1|x_i)$.

These diagrams provide a visual tool to assess how well the models' predicted probabilities align with the observed outcomes (ground truth), illustrating the comparison between the average predicted probabilities and the actual frequency of positives (i.e., the ratio of instances labeled as -1).

For each model, the predicted probabilities were divided into ten equally spaced intervals between 0 and 1. Then, the following two key metrics were calculated for each interval:

1. The average predicted probability for instances within that interval

2. The relative frequency of actual positives, determined using the ground-truth labels

Finally, the reliability diagram was generated by plotting the relative frequency of actual positives (y-axis) against the average predicted probability (x-axis) for the multiple intervals.

5.1.5 Model Interpretability

Considering the established objective of producing interpretable results, the **SHAP** method was implemented [49]. SHAP, which is rooted in cooperative game theory and ensures high local accuracy and consistency in feature contributions assessment, generates detailed explanations for individual predictions by quantifying each feature's contribution to the model's output, making it a suitable technique for understanding complex model behavior [50]. In the context of this study, focused on providing interpretable and actionable insights on the patient's clinical condition throughout hospitalization under ECMO support, SHAP's capacity to provide consistent explanations across features is invaluable, which is particularly important given the high intensity, complexity, and volatility of the clinical settings considered. The guidelines available at the GitHub repository provided by Lundberg and Lee [49] were central to implementing the SHAP method.

As described in section 5.1.2, each patient's MTS was transformed into a set of labeled 8-hour windows, where each window comprised three consecutive measurements for each variable, distinguishable through suffix $s \in \{1, 2, 3\}$ appended to the variables' names, depicting the first, second, and third (most recent) measurements. The SHAP method's application to this particular dataset, exclusively considering the validation subset, followed the two-step methodology described below:

1. **SHAP values computation:** the SHAP values were computed for each variable at different instances (8-hour windows). Since each variable appears three times within the same instance, the single SHAP value referring to each variable was computed by averaging the SHAP values obtained for the three corresponding measurements. This approach ensured that the temporal relationships captured across the different instances were appropriately reflected in the overall feature importance scores. The result is a matrix of SHAP values with dimensions $n \times m$, where n depicts the number of instances (8-hour windows) within the validation subset, and m represents the number of variables in the dataset.
2. **SHAP values selection:** as discussed in section 5.2, the primary objective of this study is to develop a risk score that provides a real-time estimate of the risk of clinical deterioration throughout hospitalization under ECMO support for each patient. Therefore, the SHAP values corresponding to outcome class -1 were selected for further analysis, as these values provide critical insights into the model's decision-making process for identifying periods of clinical deterioration.

Finally, the SHAP values, denoting feature importance scores, were visualized using a **SHAP summary plot**. This easily interpretable, highly informative, and intuitive tool depicts the relationship between features and the selected outcome (in this case, clinical deterioration). In the SHAP summary plot, each row corresponds to a unique feature for which multiple dots are displayed, each depicting the SHAP

value corresponding to a specific instance (8-hour window) of the validation subset. The color of these dots expresses the feature value (with blue indicating lower values and red indicating higher values). On the other hand, their position along the x-axis reflects the feature's impact on the model's output for the selected outcome, where negative and positive values denote an inverse and direct relationship between feature and outcome, respectively.

5.2 Phase 2: Development of a Machine Learning Risk Score for Real-Time Assessment of the Patient's Clinical Condition

The second phase aimed to leverage the best-performing, calibrated, and interpretable ML model outputted through the preceding phase to develop a tool for assessing each patient's clinical condition throughout hospitalization under ECMO support in real time. Specifically, this tool consisted of a risk score that provided an estimate of the likelihood of clinical deterioration at each time point across the hospitalization period.

The development of the risk score involved extensive optimization of the underlying mathematical formulation (section 5.2.1) and performance evaluation through two complementary approaches: qualitative and quantitative, as detailed in section 5.2.2.

5.2.1 Development and Optimization

As outlined in section 5.1.2, model training exclusively involved instances (8-hour windows) labeled as -1 and +1, excluding the remaining, labeled with outcome class 0. The first step involved applying the ML model outputted through the first phase to the multiple patient MTSs, each encompassing the hospitalization period comprised between ECMO implementation and the 12th day of ECMO support (or the entire hospitalization period if the length of stay under ECMO support was less than 12 days). This process generated predictions and the corresponding decision function outputs (interpretable as reliable probability estimates after ensuring model calibration) for all instances of every patient, including those labeled as 0. The predictions and probability estimates corresponding to instances labeled as 0 were aggregated with the results obtained during model training and validation for instances labeled as -1 and +1, preserving the datasets' temporal order.

Initially, a curve connecting the probability estimates corresponding to outcome class -1 was generated for each patient, resulting in patient-specific scores illustrating the risk of clinical deterioration at different time instants during hospitalization. While providing instance-based probability information may be valuable, physicians were primarily interested in assessing trends observed over time.

To effectively incorporate these observations, the risk score's mathematical formulation was designed to balance the significance of individual probability estimates with the overall trends observed over time. The resulting mathematical formulation is depicted in equation 5.18, where $p_s(t)$ represents the **smoothed probability estimate** for clinical deterioration at time t , $\frac{dp_s(t)}{dt}$ is the **derivative (trend component)**, computed using $p_s(t)$ and depicting the rate of change of the smoothed probability esti-

mate, and α is an **optimization parameter** that controls the relative contribution of the trend component compared to the smoothed probability estimate.

$$s(t) = p_s(t) + \alpha \cdot \frac{dp_s(t)}{dt} \quad (5.18)$$

The probability estimates for each patient were smoothed using a rolling window with a moving average. The rightmost element of the window corresponds to the probability estimate at the current time instant ($p(t)$), ensuring that only past and present data were considered during smoothing, an essential requirement for real-world applicability. Letting n denote the rolling window length (number of instances within the rolling window), $p(t)$ represent the probability estimate for the current time instant t , and i depict the number of preceding instances (8-hour windows) relative to t , the smoothed probability estimate $p_s(t)$ was computed through equation 5.19.

$$p_s(t) = \begin{cases} \frac{1}{i+1} \sum_{k=0}^i p(t-k) & \text{for } i < n \\ \frac{1}{n} \sum_{k=0}^{n-1} p(t-k) & \text{for } i \geq n \end{cases} \quad (5.19)$$

Similarly to the process applied to smooth the probability estimates, the derivative (trend component) was computed using a rolling window. The rightmost element of the window depicts the smoothed probability estimate at the current time instant ($p_s(t)$), once again ensuring that only past and present data were considered during computation. Due to the patient cohort's volatility, sudden shifts in clinical dynamics are frequent (e.g., a patient showing signs of improvement may rapidly deteriorate). Therefore, to ensure sufficient sensitivity to fluctuations in $p_s(t)$ and preserve the ability to highlight relevant shifts in clinical dynamics, smoothing was not applied to the trend component, offsetting the risk of over-smoothing (especially considering that this signal results from already smoothed values). Letting m denote the rolling window length (number of instances within the rolling window), $p_s(t)$ represent the smoothed probability estimate for the current time instant t , and i depict the number of preceding instances (8-hour windows) relative to t , the derivative (trend component) $\frac{dp_s(t)}{dt}$ was computed through equation 5.20.

$$\frac{dp_s(t)}{dt} = \begin{cases} 0 & \text{for } i = 0 \\ \frac{p_s(t) - p_s(t-i)}{i} & \text{for } 1 \leq i < m \\ \frac{p_s(t) - p_s(t-m+1)}{m-1} & \text{for } i \geq m \end{cases} \quad (5.20)$$

The optimization of the risk score's mathematical formulation, depicted in equation 5.18, involved fine-tuning three essential components through a series of iterative trials: the length of the rolling window applied for smoothing the probability estimates (**smoothing window**); the length of the rolling window employed to compute the trend component (**trend window**); the parameter α , which controls the relative contribution of both components in the risk score.

For each iteration of the fine-tuning process, a unique combination of elements was evaluated and optimized based on specific performance criteria, detailed below, which were jointly assessed to ensure

the best possible outcome:

- **Smoothing window:** the smoothing window's length optimization required balancing the need to suppress short-term fluctuations, which may not signify relevant clinical events, with the requirement to avoid excessive distortion that could conceal meaningful trends in the patient's clinical condition.
- **Trend window:** the trend window's length optimization required balancing the need to simultaneously emphasize critical clinical dynamics (e.g., an upward trend within a period of elevated values for $p_s(t)$, and a downward trend within a period of lower values for $p_s(t)$), and soften periods of increased clinical uncertainty (e.g., an upward trend within a period of lower values for $p_s(t)$, and a downward trend within a period of elevated values for $p_s(t)$).
- α : the optimization of α , constrained within the interval $[0, 5]$, involved ensuring an appropriate weighting of both components of the risk score: $p_s(t)$ and $\frac{dp_s(t)}{dt}$.

The lengths of both rolling windows (i.e., the number of instances they encompass) were constrained within the interval $[2, 7]$, corresponding to a duration between 4 and 24 hours. Based on discussions with physicians, windows exceeding this range were excluded, as longer windows risked diminishing the risk score's ability to capture the evolving dynamics of the patient's clinical condition.

Finally, to ensure consistency and comparability across iterations, the risk score was scaled using the `MinMaxScaler`, implemented via the `scikit-learn` library [45], mapping it to the range $[0, 1]$.

The main objectives considered when evaluating the risk score's performance across iterations were the following:

- Ensure that the risk score demonstrates consistent behavior, aligning with the physicians' assessment of the patient's clinical condition throughout hospitalization, while simultaneously avoiding erratic behavior.
- Ensure that the risk score demonstrates the ability not only to accurately detect periods of clinical deterioration and improvement (expected assuming that the ML model performed well during training and validation) but also to predict them in advance.

Ultimately, assessing the achievement of these objectives necessitated both **qualitative** and **quantitative** approaches, as outlined in section 5.2.2.

5.2.2 Evaluation and Selection

The risk score's performance was evaluated through two complementary approaches: **qualitative** and **quantitative**.

Qualitative assessment of the risk score's performance encompassed plotting the risk score for the multiple patients in the study. Designing an interpretable and informative visualization for the risk score involved several steps, starting with establishing the plot's background, under which to represent the risk

score curve, which depicts the values computed through equation 5.18. The plot's background should be able to intuitively convey the ground truth, meaning the patient's clinical condition according to the physicians' assessment, thereby enabling the evaluation of the risk score's behavior and its ability to align with medical assessment. Accordingly, the color scheme described below was established for the plots' background:

- **Green:** indicates clinical improvement, representing instances (8-hour windows) labeled with outcome class +1 by physicians.
- **White:** indicates clinical uncertainty, representing instances (8-hour windows) labeled with outcome class 0 by physicians.
- **Red:** indicates clinical deterioration, representing instances (8-hour windows) labeled with outcome class -1 by physicians.

After generating the background for each patient, which required access to the ground-truth labels, patient-specific risk score curves were represented against the corresponding background. As detailed in section 5.1.2, each data point on the curve represents an 8-hour window, separated from adjacent data points (windows) by a 4-hour interval. Therefore, the risk score curve starts at the time point indicating 8 hours after ECMO implementation (this being the earliest point from which sufficient data for building 8-hour windows is available), meaning that estimates for the risk of clinical deterioration were not provided for the time points depicting ECMO implementation and the following 4 hours.

The risk score was designed to classify instances (8-hour windows) reflecting clinical deterioration and improvement in real-time (as evaluated in section 5.1.4) while simultaneously aiming to anticipate such events. Therefore, performance evaluation primarily focused on its behavior within predefined time windows preceding periods of clinical deterioration and improvement. The remaining plot sections do not yield such valuable insights: within white regions (depicting periods of clinical uncertainty), the risk score's behavior cannot be meaningfully evaluated as ground-truth assessments regarding clinical progression are not available for comparison (although erratic behavior is not anticipated); within red and green regions, for which physicians provided confident assessments of the patient's clinical state, the risk score's behavior is expected to align with the physicians' classification (assuming that the model performed well during training and validation).

As indicated by physicians, accurately anticipating these patients' clinical trajectories is highly challenging. Based on this input, **windows of 4, 8, and 12 hours preceding periods of clinical deterioration and improvement** were considered for subsequent evaluation.

Quantitative assessment of the risk score's performance involved computing **Receiver Operating Characteristics (ROC) curves** for the aforementioned time windows. These curves represent the **False Positive Rate (FPR)** on the x-axis (equation 5.21), against the **True Positive Rate (TPR)** on the y-axis (equation 5.22).

$$FPR = \frac{FP}{FP + TN} = 1 - Specificity \quad (5.21)$$

$$TPR = \frac{TP}{TP + FN} = \text{Sensitivity} \quad (5.22)$$

The computation of the ROC curves followed an iterative process, where progressively higher decision thresholds (values that define the boundary between predicting clinical deterioration and improvement) were applied within the range $[0, 1]$, with increments of 0.01. At each iteration, the risk score was evaluated over predefined time windows (4, 8, and 12 hours preceding periods of clinical deterioration and improvement). The threshold value determined the classification of the risk score's estimate for a given time window as one of the following: TP (the risk score correctly predicts clinical deterioration in advance), FP (the risk score incorrectly predicts clinical deterioration when in reality the patient's clinical condition is improving), FN (the risk score incorrectly predicts clinical improvement when in reality the patient's clinical condition is deteriorating), or TN (the risk score correctly predicts clinical improvement in advance). After performing these evaluations for each threshold on a patient-by-patient basis, the following key metrics were computed: **FPR**, which depicts the proportion of periods of clinical improvement misclassified as clinical deterioration, indicating the likelihood of a false alarm (incorrectly signaling clinical deterioration); **TPR**, which represents the proportion of periods of clinical deterioration successfully predicted by the risk score within the predefined time windows.

Finally, after computing the FPR and TPR for all decision thresholds and predefined time windows, the ROC curves were generated. The area under the curve was also calculated, serving as the decisive metric for selecting the optimal risk score (i.e., the smoothing and trend windows' length and the value of α , the three essential components considered for fine-tuning the risk score).

Chapter 6

Results and Discussion

The research methodology described in chapter 5 aimed to leverage ML techniques to process and model multidimensional and multidomain datasets referring to patients undergoing ECMO support. The ultimate objective consisted of creating a tool that supports physicians in high-quality medical judgment and decision-making by providing accurate and interpretable real-time estimates of the risk of clinical deterioration throughout hospitalization under ECMO support for each patient.

The evaluation of the outcomes of the proposed research methodology involved a comprehensive strategy combining technical analyses and context-based observations, as will be discussed throughout this chapter.

6.1 Phase 1: Development of a Machine Learning Model for Classification of Clinical Deterioration and Improvement

The first phase of the research methodology involved processing the multiple patient MTSs, transforming them into a format suitable for applying ML models to detect clinical deterioration and improvement. Despite limited data availability (the datasets from 81 patients, each constrained to the first 12 days of hospitalization under ECMO support, were considered), robust data preprocessing and feature engineering pipelines were employed. These steps facilitated the transformation of a sequential supervised learning problem into a standard supervised learning one, enabling conventional ML models to achieve substantial performance in distinguishing between clinical condition deterioration and improvement, underscoring the strength of methodological choices shaped by the study's inherent limitations.

The results obtained within each phase of the ML pipeline employed will be analyzed and discussed throughout this section.

6.1.1 Data Preprocessing

The first step in the data preprocessing pipeline involved performing outlier detection and removal. As outlined in section 5.1.1, selecting the most appropriate strategy required careful consideration of the

patient cohort's unique characteristics. Due to the critical clinical state and inherent volatility of patients hospitalized under ECMO support, certain variables exhibited erratic behaviors, resulting in extreme measurements. While such measurements would typically be flagged as outliers using conventional methods (e.g., statistical-based), they actually reflected natural clinical occurrences within this study. Consequently, using a **rules-based approach**, validated by physicians prior to implementation, proved to be highly effective at eliminating significant inaccuracies, mainly derived from manual data entry errors. The application of this strategy resulted in a relatively low outlier incidence, highlighting the robustness of the ICU's automated data collection and management systems while also underscoring the physicians' adherence to established protocols.

Following outlier detection and removal, missing values for each variable were imputed using strategies tailored to their specific characteristics. Missing values for static (time-independent) variables were manually recovered from the ICU's information system. For dynamic (time-dependent) variables, two distinct approaches were considered: for continuous numerical variables, the multivariate Iterative Imputer algorithm, a widely used method across similar studies, was applied; for categorical variables, missing values were imputed using backward and forward filling based on the first available measurement within a 3-day adjacent time window, with the missing value being classified as unknown if no valid measurement was found within this window. The latter approach minimized the risk of incorporating outdated, potentially misleading information into the patients MTSs, thus avoiding compromising the ML models' learning outcomes. Notably, this scenario did not arise, reaffirming the physicians' adherence to established protocols.

Figure 6.1 presents an overview of the incidence of outliers and missing values for all the variables across the multiple patient MTSs, with missing values generated through outlier removal not being considered.

Overall, both the outlier and missing value rates were relatively low, especially considering that an initial filtering process was employed to eliminate grossly anomalous variables that would have otherwise compromised data quality and, by extension, the validity of the results (further details available within section 4.4.2). It is relevant to note the absence of outliers and missing values for the static (time-independent) variables Age, Sex, Obesity Degree, Diabetes, Hypertension, and Infection. These variables were manually recorded in each patient's MTS by consulting the corresponding clinical file, thereby avoiding reliance on physician manual entry and automated data acquisition and registration systems. Ultimately, the impact of artificially introduced data was minimal.

After ensuring data quality and consistency, the subsequent step involved computing the derived variables represented via equations 5.1 to 5.9. These variables, established cooperatively with physicians, hold significant clinical relevance in the context of this study, further enhancing its innovative nature. Notably, an extensive literature review revealed that these variables have not been previously reported in similar studies.

The final step of the data preprocessing pipeline involved evaluating the correlation between specific variables across the multiple patient MTSs using Spearman's rank correlation coefficient. This analysis focused on identifying dynamic patterns and temporal relationships between variables, which ultimately

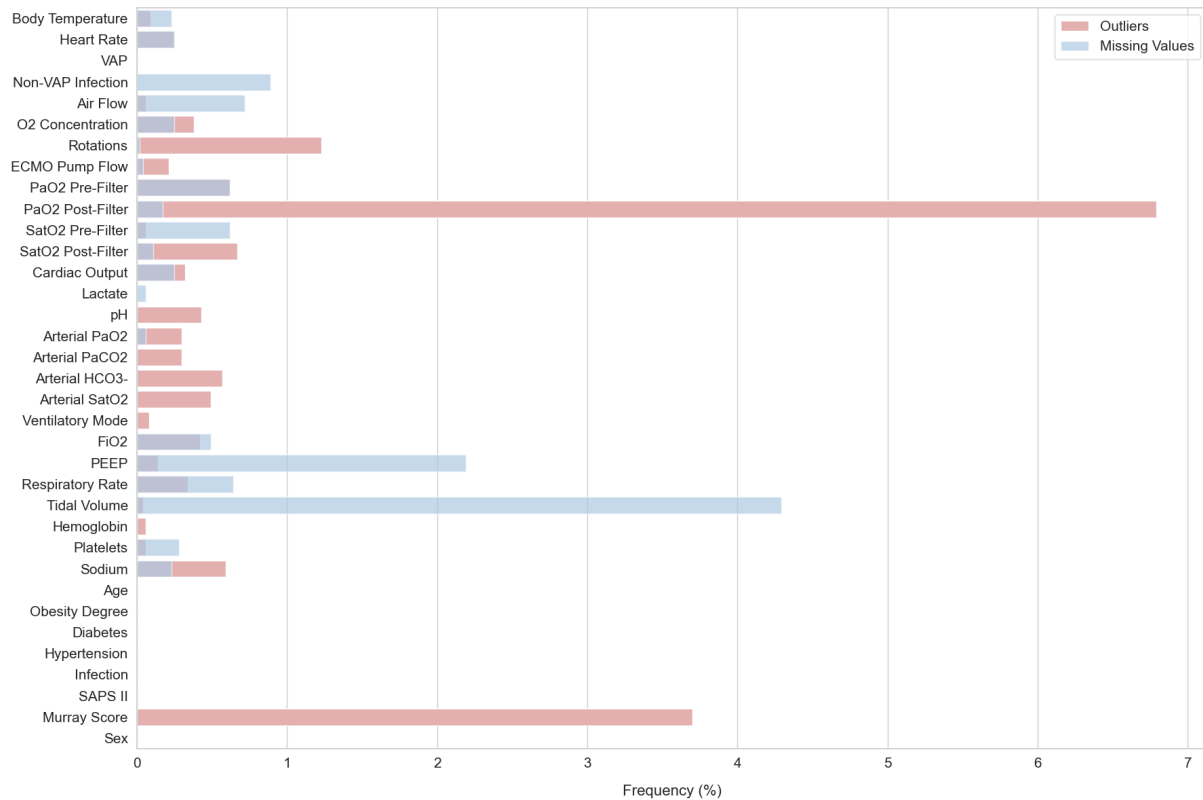


Figure 6.1: Incidence of outliers and missing values across all variables within the multiple patient MTSs.

led to the exclusion of several categories of variables:

- **Static (time-independent) variables:** these variables (Age, Sex, Obesity Degree, Diabetes, Hypertension, Infection, SAPS II, Murray Score) do not change within a patient's MTS and thus offer no meaningful insights in the context of temporal correlation assessment.
- **Nominal categorical variables:** these variables (VAP, Non-VAP Infection, Ventilatory Mode) were excluded since Spearman's rank correlation is unsuitable for unordered categories, as it ranks the data before computing correlations.
- **Variables with zero variance for at least one patient's MTS:** these variables (Rotations, SatO₂ Post-Filter, Respiratory Rate, FiO₂, PEEP, O₂ Concentration) were excluded since Spearman's rank correlation relies on rank-based differences, meaning variables that do not change over time provide no information for correlation analysis.

Derived variables computed in the preceding step were also omitted from this analysis, as their clinical relevance requires further consideration in subsequent stages.

Correlation matrices were initially computed for each patient. These matrices were then aggregated into a single, averaged matrix to provide a general overview of the correlations across the patient cohort. While averaging individual matrices does not offer a perfect reflection of the underlying correlations, the aggregated matrix, illustrated in figure 6.2, serves as a reliable proxy for the overall correlation structure.

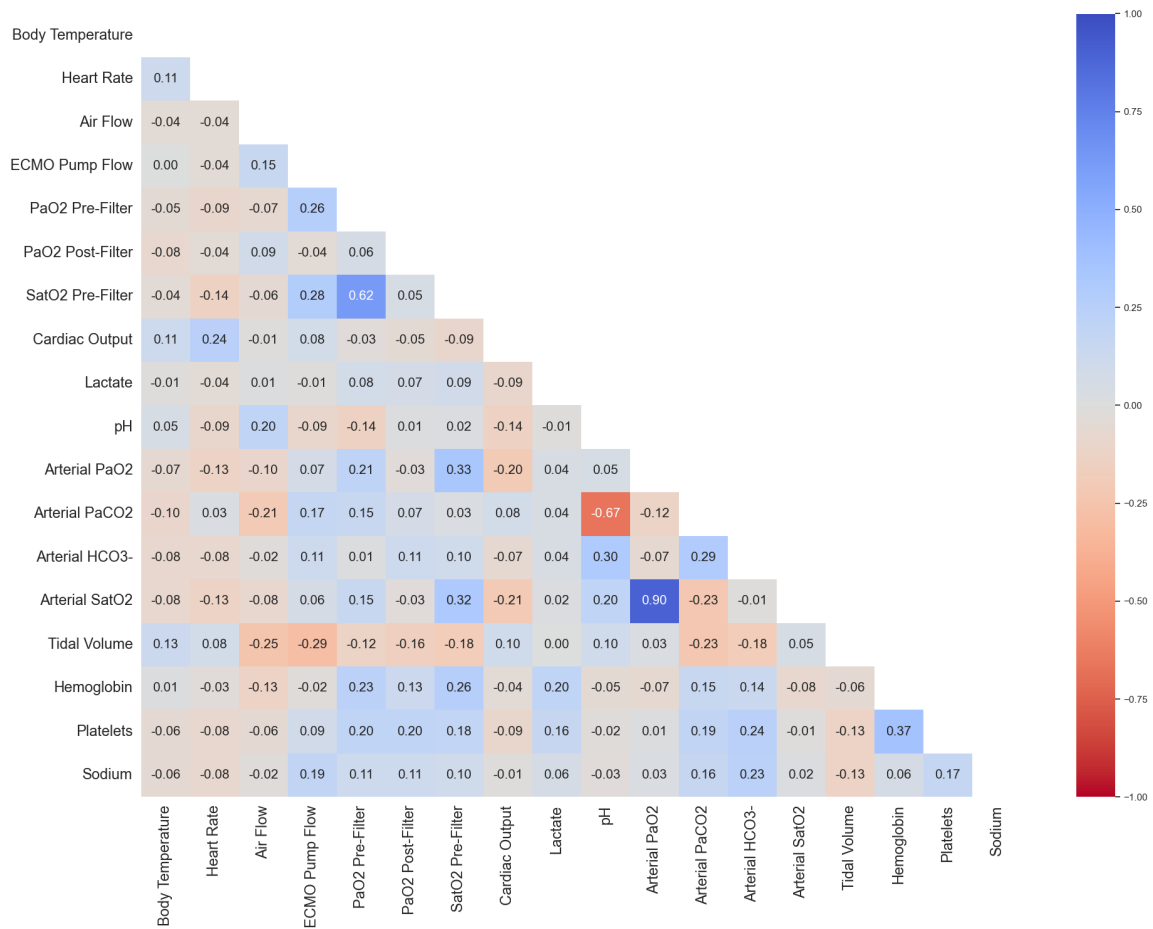


Figure 6.2: Correlation matrix obtained after aggregating the multiple patient-specific correlation matrices computed considering the Spearman's rank correlation coefficient.

The most notable correlations observed across the multiple patient MTSs encompassed the following variable pairs: (SatO₂ Pre-Filter, PaO₂ Pre-Filter), (Arterial PaCO₂, pH), and (Arterial SatO₂, Arterial PaO₂). To determine which variable to exclude from each pair, an initial analysis focused on evaluating model performance by iteratively considering different combinations of these six variables was conducted. However, model performance remained practically unchanged regardless of the combinations, leading to the abandonment of this approach. Furthermore, including all highly correlated variables did not lead to any performance increment, reinforcing the decision to proceed with variable exclusion.

Ultimately, an alternative approach focusing on the incidence of artificially added values was employed, leading to the exclusion of variables exhibiting the highest overall incidence of such values. Referring to figure 6.2, the variables with the higher incidence for each pair were: PaO₂ Pre-Filter (1.24% vs. 0.68%), pH (0.43% vs. 0.3%), and Arterial SatO₂ (0.49% vs. 0.36%). **Consequently, variables PaO₂ Pre-Filter, pH, and Arterial SatO₂ were excluded from the multiple patient MTSs.**

Excluding highly correlated variables that did not impact model performance reduced the datasets' dimensionality and overall complexity, preventing redundant information from being incorporated in subsequent analyses. Also, by considering fewer variables, the datasets' construction process was further simplified, enhancing applicability and ultimately facilitating the interpretation of results.

6.1.2 Feature Engineering

The first step within the feature engineering pipeline involved transforming each patient's MTS into multiple independent, labeled windows, each comprising three consecutive time points. This transformation was essential to preserve temporal dependencies within the data, a critical aspect regarding the analysis of clinical information where consecutive observations often reveal trends indicative of the patient's clinical condition progression throughout hospitalization under ECMO support. By incorporating consecutive time points within each window, the model could effectively capture patterns and temporal dynamics, which would not be possible if it relied exclusively on isolated observations. The results of this transformation are illustrated in table 6.1.

Table 6.1: Results obtained after transforming each patient's MTS into multiple labeled windows, each comprising three consecutive time points

Outcome	Total windows	Patients with windows exclusively labeled with outcome class
-1	632 (12.32%)	2 (2.47%)
0	3430 (66.87%)	14 (17.28%)
+1	1067 (20.8%)	0 (0.0%)

Table 6.1 shows that most resulting windows were labeled as 0. While ternary classification was initially considered, significant class imbalance and the specific characteristics of windows labeled as 0 (indicative of medical uncertainty) led to poor predictive performance (regardless of window length and overall training settings), with the model exhibiting a considerable difficulty in distinguishing between the three outcome classes. Based on these observations, windows labeled as 0 were excluded from the model development process, resulting in a reduction of 66.87% in the total number of windows, limiting data availability for model training. However, this approach enabled data quality to be maximized, leading to enhanced predictive performance and ultimately improved model applicability, as its development exclusively encompassed windows indicative of medical certainty. Consequently, subsequent analyses, including determining the optimal window length, were conducted exclusively considering windows labeled as -1 and +1 (binary classification task).

Extensive experimentation validated the decision to establish the optimal window length at three time points. Various window sizes, ranging from 1 (isolated observations) to 14 time points (representing a 48-hour period), were assessed to balance temporal dynamics and model performance. Although higher dimensional windows could theoretically capture more long-term trends, they also increased the complexity and dimensionality of the datasets, heightening the risk of model overfitting and leading to more substantial computational inefficiencies. Conversely, lower dimensional windows limited the model's ability to account for temporal dynamics, reflected in poorer predictive performance.

The results of these experiments revealed a significant performance decline when considering windows of length 1 (isolated observations) and 2 (4-hour period), with a 40% and 20% average reduction in performance metrics (section 5.1.4), respectively. Peak performance was achieved when considering windows comprising three consecutive time points, after which it stagnated up to six time points. Per-

formance gradually declined for windows with lengths constrained within the interval [7, 14], suggesting no benefit in considering higher dimensional windows to capture meaningful time-related information. This observation can be attributed to the volatility and critical clinical state inherent to the patient cohort, meaning that significant changes in their clinical condition typically occur suddenly (i.e., within shorter periods). Higher dimensional windows likely diluted this crucial information by blending time points corresponding to periods of clinical stability preceding deterioration or improvement, ultimately obscuring relevant trends and temporal dynamics.

Finally, vectorizing each window into a fixed-length feature vector which, by application of equation 5.10, resulted in a total of **87** elements, enabled conventional ML models (SVM and Random Forest) to process the temporal structure of the data without requiring more complex and computationally intensive architectures (e.g., neural networks). Ultimately, this approach maintained the balance between capturing essential temporal patterns and ensuring computational efficiency.

As shown in table 6.1, removing windows labeled as 0 resulted in the exclusion of 14 patients from the model development process. This exclusion had a notable impact, as it reduced the number of available patients by 17.28%, eliminating individuals with distinct characteristics, clinical trajectories, and treatment responses, ultimately limiting data availability for model development. Consequently, the ability of the remaining datasets (particularly the training subset) to capture the patient cohort's inherent complexity and diversity was significantly diminished, impacting the model's ability to learn patterns that effectively represent the heterogeneity of the patient population. This challenge underscored the importance of carefully designing and implementing an appropriate data splitting strategy to mitigate these limitations.

In contrast to the implemented methodology, a patient-based data splitting strategy (i.e., assigning all windows corresponding to a single patient to either the training or validation subset) was found to substantially decrease model performance, with a 15% average reduction in performance metrics. Furthermore, this approach inherently limits the model's exposure to different patient subgroups during training, likely contributing to the observed performance decline.

An alternative strategy, which involved randomly assigning patient-specific windows to both the training and validation subsets neglecting temporal relationships (i.e., without maintaining temporal order), was also considered. However, this approach not only facilitated data leakage, where the model trains on data that is influenced by preceding observations assigned to the validation subset) but also resulted in a non-negligible decline in model performance (average 8% reduction in performance metrics). This decline was likely due to the disruption of temporal relationships critical for modeling time-independent patient dynamics.

Ultimately, a total of 1334 windows were assigned to training subset and 365 to the validation subset. The findings above presented validated the chosen data splitting strategy, which ensured that patient-specific windows assigned to the validation subset were preceded by those used for training, thereby simulating real-world scenarios where the model must predict future outcomes based solely on past data. This approach not only improved the models' predictive performance but also enhanced its applicability by ensuring that the development process mirrored real clinical decision-making workflows.

In the feature extraction process, both the `StandardScaler` and `MinMaxScaler` methods, implemented via the `scikit-learn` library [45], were initially considered for scaling numerical variables. The selection of the most appropriate method was heavily influenced by the characteristics of the data, which consisted of multiple, highly volatile MTSs referring to patients hospitalized under ECMO support. Even after performing outlier detection and removal, substantial oscillations and significant, often sudden variations in the data were frequently observed, which, despite looking abnormal, reflected natural occurrences. As a result, certain variables had most values concentrated within a narrow range, with a few extreme values significantly deviating from the norm.

The `MinMaxScaler` method scales numerical variables into a specific range (e.g., $[0, 1]$) based on their minimum and maximum values. However, this approach was unsuitable for the data, as the frequent occurrence of extreme values significantly impacted data distribution, causing the compression of most data points into a very narrow range for certain variables, with the extreme measurements occupying the remaining space. This uneven scaling compromised the ML models' ability to assess the differences among typical data points (as these were excessively compressed), thus compromising their ability to learn relationships and meaningful patterns.

Regarding encoding categorical variables, the `LabelEncoder` method, implemented via the `scikit-learn` library [45], which maps categories to integers, was also initially considered. However, in this study, the `LabelEncoder` method introduced misleading ordinal relationships into the data, leading the models to infer nonexistent hierarchical structures among categories. The data used within this study comprised several categorical variables, including Obesity Degree, represented by integer values ranging from 0 to 3, akin to the output of the `LabelEncoder` method. This representation introduced an artificial ordinal structure that lacked meaningful interpretation, hindering the models' learning process and, ultimately, their performance.

Ultimately, the feature extraction pipeline integrated the `StandardScaler` method for scaling numerical variables and the `OneHotEncoder` method for encoding categorical variables. These methods were carefully selected considering data characteristics and the specific strengths and limitations of the ML models employed (e.g., the SVM model's sensitivity to feature magnitudes, which necessitated the use of a scaling method like `StandardScaler`, which preserved the relative distributions of numerical features).

6.1.3 Model Training and Hyperparameter Optimization

Both the SVM and Random Forest models were trained and further evaluated for their ability to distinguish between clinical deterioration (windows labeled as -1) and clinical improvement (windows labeled as +1). The SVM model was trained using three distinct kernel functions: linear (employed as a baseline for comparison), polynomial with degree 3, and RBF, and did not undergo any additional hyperparameter optimization.

Conversely, the Random Forest model's training process involved optimizing the hyperparameter `n_estimators`, which depicts the number of trees in the ensemble, through an iterative search approach (further details available in section 5.1.3). The initial search window was set to $[0, 1000]$, followed by

a progressive narrowing of its length based on model performance, culminating in the identification of a sufficiently narrow search window: $[50, 150]$, associated with enhanced model performance. Within this search window, performance remained relatively stable, particularly across the range $[90, 110]$. Ultimately, the **optimal number of trees was set to 100**. The results derived from evaluating the performance of the SVM model with each kernel function and the Random Forest model with 100 trees are available in section 6.1.4.

Model calibration was another crucial step when training the models, as calibrated decision function outputs (i.e., probability estimates) were subsequently used to compute a risk score. Calibration performance was evaluated for the SVM model with three distinct kernel functions and the Random Forest model with 100 trees. Platt Scaling was selected as the optimal calibration method due to its reduced susceptibility to overfitting, a particularly relevant feature given the small size of the data.

Isotonic Regression, an alternative calibration method, was also considered. However, due to its significant proneness to overfitting in smaller datasets, this method was discarded. Isotonic Regression works by fitting a non-decreasing function to the data (lying as close to the observation as possible), making it highly flexible (particularly relevant for larger datasets) and capable of capturing intricate relationships. However, this can lead to overfitting when fewer data points are available, as sensitivity to small fluctuations in the data increases, ultimately impairing its ability to generalize effectively [48].

Ultimately, the selection of the SVM and Random Forest models was primarily motivated by their remarkable ability to handle datasets of small to moderate size, an essential feature given the limited data availability in this study. Furthermore, data scarcity posed a significant challenge, requiring careful selection of ML models capable of performing well under these conditions.

Additionally, these models also exhibit a notable ability to handle large datasets (particularly the Random Forest model), as they are capable of accurately learning complex, nonlinear relationships in the data. While the SVM can also capture nonlinear patterns, it requires a nonlinear kernel, such as polynomial (degree ≥ 2) and RBF. This feature offers a pathway for scalability, a key consideration in the model selection process, as future phases of this study should involve applying these models to increasingly larger datasets, making their ability to generalize to more complex data highly relevant.

6.1.4 Model Evaluation and Selection

Initially, a confusion matrix, structured as illustrated in figure 5.4, was computed for each model. From these, the following performance metrics were calculated: Accuracy (equation 5.14), Precision (equation 5.15), Recall (equation 5.16), and F-Score (equation 5.17), considered for evaluating the ML models' behavior and predictive performance. However, ultimately, **the primary metric used for selecting the best-performing model was the F-Score**.

Accuracy was excluded as a reliable performance metric due to its vulnerability to bias in the presence of class imbalance. As shown in table 4.4, there is a noticeable imbalance between the number of samples (and ultimately windows) corresponding to outcome class -1 and +1. This imbalance risked inflating accuracy, as a model biased toward predicting the majority outcome class (+1) could still achieve

a considerable accuracy score without genuinely reflecting predictive performance.

In contrast, Precision and Recall were established as critical performance metrics in this study, primarily due to their relevance in real-world clinical applications. A detailed discussion of their significance, with a contextual framework, is provided below:

- **Precision** measures the predicted positive rate, being particularly important for minimizing FPs, which occur when the model incorrectly predicts clinical deterioration for a window that, in reality, reflects clinical improvement. In this study, limiting FPs is critical since unnecessary alerts for clinical deterioration can lead to unreasonable, potentially compromising human (e.g., physicians and nurses) and material (e.g., medical equipment) resource relocation. Moreover, FPs could trigger unnecessary therapeutic interventions that might carry adverse side effects, ultimately impacting the patient's clinical trajectory. An excessively elevated rate of false alarms can also aggravate alarm fatigue among healthcare professionals, which is particularly undesirable in such a high-intensity and challenging environment like the ICU, already characterized by intense alarm activity (particularly considering the management of critically ill patients hospitalized under ECMO support).
- **Recall**: measures the TPR, being particularly important for minimizing FNs, which occur when the model incorrectly predicts clinical improvement for a window that, in reality, reflects clinical deterioration. In clinical practice, an automated system that frequently fails to detect clinical deterioration could lead to unintended relaxation in patient monitoring, facilitate premature cessation of essential therapeutic interventions, and cause resource misallocation. FNs pose a significant risk, as they may delay critical interventions, potentially leading to life-threatening situations that require urgent decision-making, ultimately increasing the likelihood of medical errors and jeopardizing the quality of care and patient outcomes.

The **F-Score** provides a balanced assessment of the models' performance, considering both FPs and FNs, by combining Precision and Accuracy into a single metric through their harmonic mean. This balanced evaluation is particularly relevant in this study, where it is imperative to minimize both types of errors to ensure an automated system based on the ML models developed does not compromise clinical decision-making, quality of care, or patient safety.

The performance metrics for the various ML models, computed using the validation subset, are presented in table 6.2.

Table 6.2: Performance metrics for the ML models. * Polynomial; ** Random Forest.

Metrics	Models			
	SVM (Linear)	SVM (Poly*)	SVM (RBF)	RF**
Precision	0.9567	0.9918	0.9767	0.9959
Recall	0.9644	0.9526	0.9526	0.9565
F-Score	0.9606	0.9718	0.9659	0.9758

Although the difference in performance between the models is minimal, the **Random Forest with 100 trees demonstrated the highest F-Score (0.9758)**, as highlighted in table 6.2. The corresponding confusion matrix is illustrated in figure 6.3.

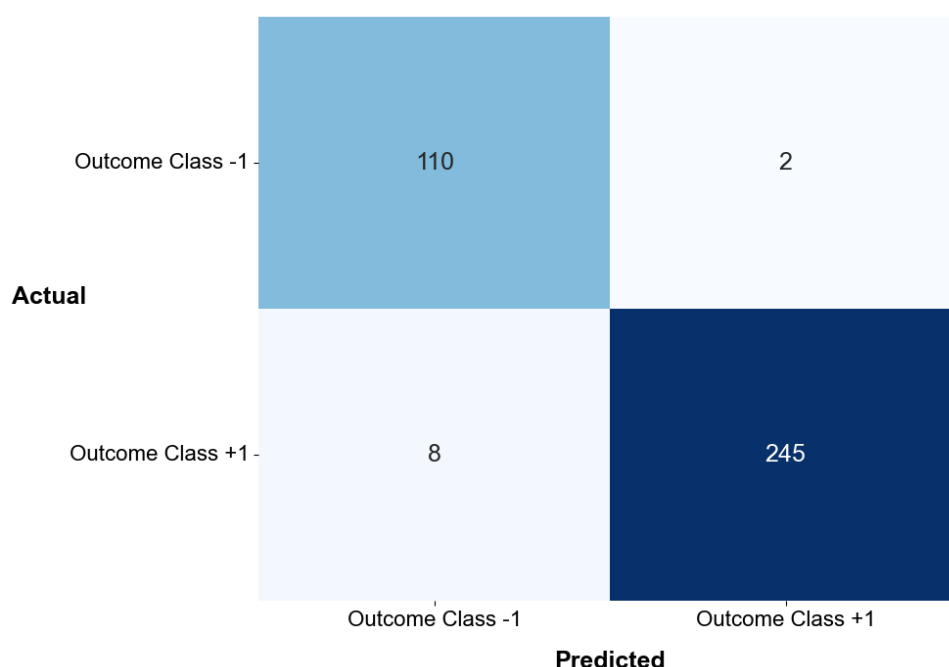


Figure 6.3: Confusion matrix for the Random Forest model with 100 trees, computed using the validation dataset.

The results presented in table 6.2 demonstrate the models' capability to accurately distinguish between clinical deterioration (windows labeled with outcome class -1) and clinical improvement (windows labeled with outcome class +1), an expected outcome given the clear separation between these outcome classes in the feature space. If the models had been unable to perform this task, which is easily achievable for experienced physicians, their added value (i.e., support medical evaluation and decision-making) would have been nullified.

Given that the selected model would be used to develop a risk score based on its decision function outputs, it was essential to assess the calibration of the models to ensure that these outputs could be interpreted as reliable probability estimates. As described in section 5.1.4, model calibration was evaluated qualitatively, relying exclusively on visual analysis of the reliability diagrams, also known as calibration curves. Consistent with the results observed for the performance metrics, the Random Forest model with 100 trees outperformed the remaining models in terms of calibration, this time with a more substantial margin. Figure 6.4 presents the reliability diagrams, computed using the validation subset, for both the uncalibrated and calibrated (via the Platt Scaling method) versions of the Random Forest model with 100 trees, enabling a direct comparison of their calibration performance.

The comparison between the reliability diagrams for the uncalibrated and calibrated versions of the Random Forest model with 100 trees demonstrated a substantial improvement in calibration following the application of the Platt Scaling method. The uncalibrated model (blue curve) exhibits significant deviations from the diagonal (perfect calibration), indicative of miscalibration, reflected in the tendency to

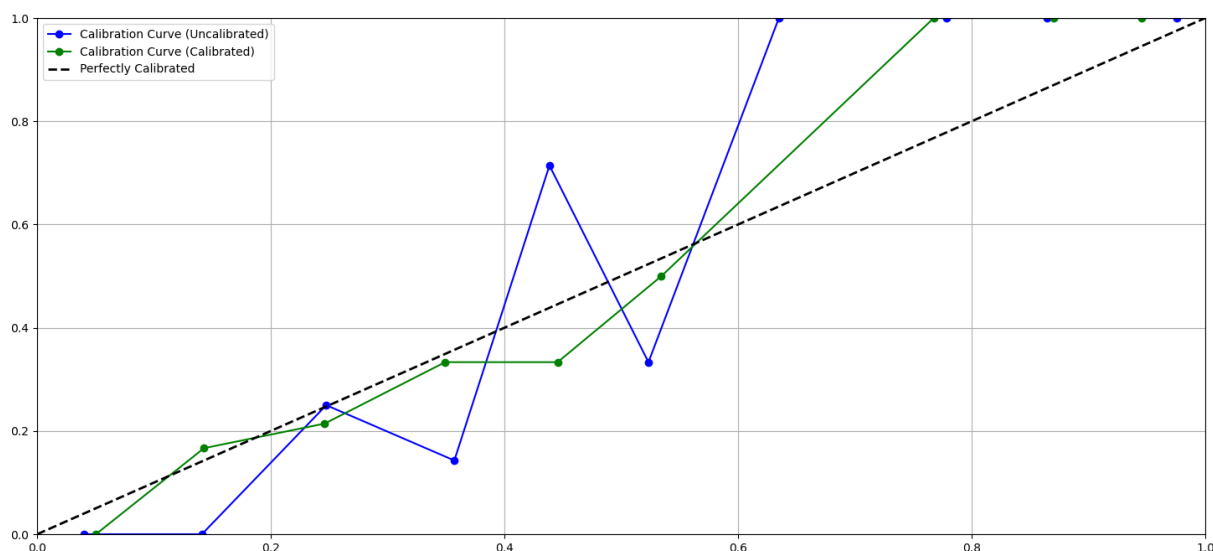


Figure 6.4: Reliability diagrams (or calibration curves) for the uncalibrated (blue curve) and calibrated (green curve) versions of the Random Forest model with 100 trees.

underestimate and overestimate the probabilities for outcome class -1. In contrast, the calibrated model (green curve) aligned more closely with the diagonal (particularly within the probability range $[0.2, 0.5]$), revealing that the calibration process successfully improved the precision of the model's probability estimates.

Although the calibrated model did not achieve perfect calibration, its performance was considerably better across all probability ranges, correcting the miscalibration (i.e., reducing the degree of underconfidence and overconfidence) observed in the uncalibrated model. However, the curves of both models displayed a more inconsistent behavior within the interval $[0.6, 1]$ on the x-axis, suggesting that the model struggled with higher probability predictions. This instability was attributed to the limited number of positive samples (windows labeled as -1), which totaled 112 instances, as can be inferred from the confusion matrix (figure 6.3). This small sample size likely increased statistical noise within this interval, making it challenging for the model to generate stable calibration results. Considering this limitation, probabilities within this range were cautiously interpreted.

Despite the fluctuations encountered when assessing the reliability diagrams, the calibrated Random Forest model with 100 trees revealed the ability to produce probability estimates that could be reasonably interpreted as reliable, fulfilling a critical requirement for developing an accurate, trustworthy risk score.

6.1.5 Model Interpretability

SHAP analysis was conducted to assess the contribution of each feature to the model's predictions regarding clinical deterioration (outcome class -1), aligning with the study's objective to develop an interpretable ML model for subsequent use in building a risk score. The SHAP method's implementation followed a methodology adapted to the characteristics of the data, as detailed in section 5.1.5.

Figure 6.5 presents the SHAP summary plot computed for the validation subset. To ensure the

model's generalization, a comparison with the SHAP summary plot referring to the training subset was performed. The analysis revealed that the top 10 features remained consistent across both subsets, with only a minor rank change involving variables Rotations and Native Lung O₂ Transfer. This consistency indicated that the patterns learned during training were effectively transferred to the validation subset, underscoring the model's ability to produce stable, reliable, and interpretable predictions.

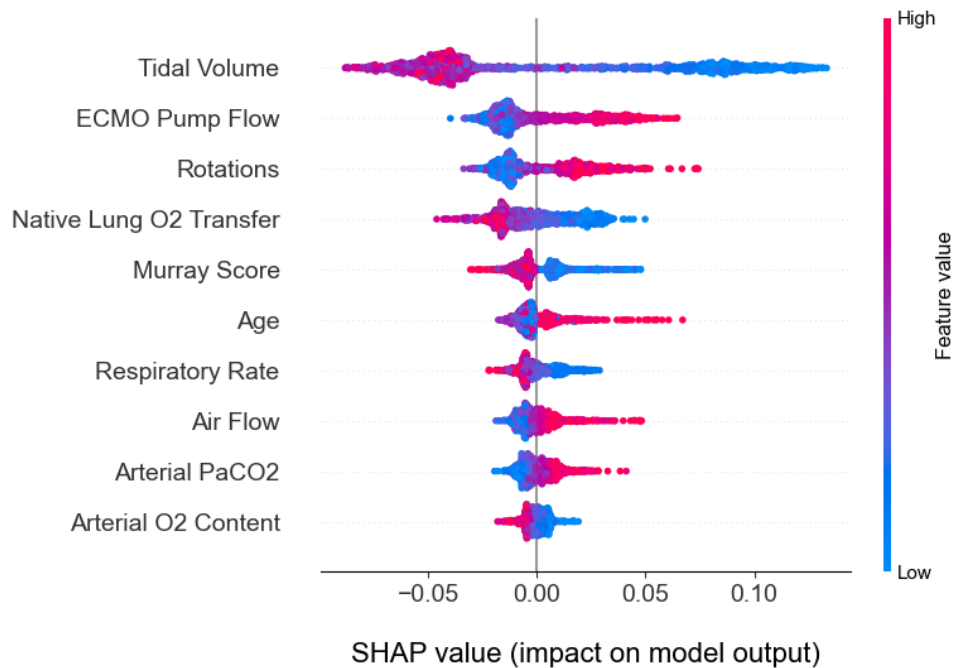


Figure 6.5: Display of the top 10 features ranked by their mean absolute SHAP values, with features listed in descending order of importance for predictions. Each feature plot illustrates SHAP values across observations (x-axis), color-coded to represent low (blue) to high (red) feature values. The absolute value indicates the contribution extent, with the sign denoting positivity or negativity.

The SHAP summary plot results were reviewed and discussed with physicians, who provided insights into the behavior of the multiple contemplated features, leading to a deeper understanding of the underlying clinical dynamics that characterize the patient cohort. Table 6.3 summarizes the outcomes of this collaborative analysis for the top 7 features, highlighting the positive impact that automated systems capable of processing and learning from multidomain and multidimensional datasets can have in medical practice.

Table 6.3: Collaborative evaluation of the 7 most relevant features determining model predictions for clinical deterioration (outcome class -1), based on physicians' feedback. The table highlights clinical insights and contradictions revealed through SHAP analysis.

Variable	Assessment
Tidal Volume	Represents the volume of air delivered to the lungs by the mechanical ventilator during each respiratory cycle. Lower tidal volumes (blue dots) suggest reduced mechanical compliance of the lung, signaling a worsening condition and impaired lung function.
ECMO Pump Flow	Represents the volume of blood circulating through the ECMO circuit each minute (measured in L/min). Higher pump flow (red dots) reflects greater reliance on extracorporeal support, thus indicating a deteriorating clinical condition. Conversely, reducing pump flow reflects improving clinical condition, ultimately signaling reduced dependence on ECMO support.
Rotations	Refers to the electromagnetic force that drives the blood through the ECMO circuit. Higher rotation speeds correspond to increased pump flow, which enables physicians to adjust ECMO support as required. The need for higher rotation speeds (red dots) indicates a worsening clinical condition.
Native Lung O ₂ Transfer	Represents the amount of oxygen transferred by the patient's native lungs from the pulmonary alveolus into the bloodstream, constituting a natural indicator of lung function. Lower native lung oxygen transfer reflects more significant impairment, leading to increased dependence on ECMO support.
Murray Score	Index employed to assess the severity of Acute Respiratory Distress Syndrome (ARDS) prior to ECMO implementation. A higher score (red dots) indicates more significant lung impairment, meaning a worsening clinical condition. Interestingly, a more elevated score is associated with a lower risk of clinical deterioration, possibly due to the already critical and extreme condition of these patients (nonetheless, conventional clinical reasoning would expect patients with higher Murray score values to exhibit worse clinical progression throughout ECMO support).
Age	Patients' age directly correlate with worse clinical outcomes, with older patients (red dots) facing a more elevated risk of clinical deterioration throughout hospitalization under ECMO support. This observation aligns with medical expectations, as older individuals often present with comorbidities and substantial clinical frailty.
Respiratory Rate	For patients undergoing ECMO support, a higher respiratory rate may indicate reduced sedation and gradual clinical improvement, contrasting with the characteristic physiological response, where worsening lung function and disease severity typically result in a more elevated respiratory rate, which serves as a compensatory mechanism for impaired gas exchange (i.e., the lungs' efforts increase to ensure sufficient oxygen delivery). However, for patients undergoing ECMO support, an increased respiratory rate is often associated with recovery, signaling greater autonomy and reduced dependence on ECMO support.

Ensuring that the results produced by these systems were interpretable to physicians fostered meaningful medical reflection, enabling physicians to access insights that align and contradict clinical reasoning (e.g., through highlighting typically undervalued or overlooked variables and identifying underlying dynamics that may be subtle or concealed from clinical observation). This disparity between physicians and the outputs of these systems presents an opportunity for enhanced analysis, facilitating the uncovering of hidden clinical insights and adoption of more appropriate (potentially effective) treatment strategies (i.e., strategies possibly overlooked in the absence of these results).

Analysis of the SHAP summary plot revealed that physicians found the feature importance rankings to be generally consistent with clinical reasoning, as no features were identified as nonsensical, relating to the outcome (clinical deterioration) in a way that aligns with their expectations (the only exception comes with variable **Murray Score**, whose behavior slightly contradicted conventional clinical reasoning, as outlined in table 6.3). Moreover, physicians noted some unexpected rankings for the following variables: **Native Lung O₂ Transfer** (expected to be ranked higher), and **Arterial PaCO₂** (presented with an unexpectedly high ranking).

Additionally, the SHAP summary plot highlighted the enriching impact of incorporating the innovative variables depicted in section 5.1.1, demonstrated by the fact that two of these variables (Native Lung O₂ Transfer and Arterial O₂ Content) were ranked among the top 10 most relevant features determining model predictions for clinical deterioration, thereby enhancing its predictive power.

Ultimately, this first phase outputted a well-performing, calibrated, and interpretable Random Forest model with 100 trees, which was used to build a risk score for assessing the patient's clinical condition throughout hospitalization under ECMO support.

6.2 Phase 2: Development of a Machine Learning Risk Score for Real-Time Assessment of the Patient's Clinical Condition

The demand for automatic systems (i.e., not requiring direct human intervention) capable of processing and analyzing large volumes of complex data has grown significantly due to expanding data availability, driven by continuous technological advancement (medical equipment is becoming increasingly sophisticated, resulting in enhanced monitoring capacity and the consequent generation of larger volumes of data). Its integration within specific medical settings can be positively transformative, particularly in high-intensity, volatile contexts (e.g., managing patients hospitalized in the ICU under ECMO support). The insights produced by these systems could enable physicians to assess previously inaccessible (or undetectable) information, reducing the risk of undervaluing or overlooking relevant dynamics (or trends), which could be the difference between being able or not to act early (or even preventively) in the face of complex clinical scenarios and avoid irreversible, potentially catastrophic outcomes.

The risk score developed aimed to provide a support tool to physicians, enabling them to access additional information that could stimulate more comprehensive reasoning and patient assessment, ultimately improving their ability to perform early detection of clinical dynamics and trends (e.g., deteriora-

tion or improvement of the patient's clinical condition).

6.2.1 Development and Optimization

The decision to assess the risk of clinical deterioration throughout hospitalization under ECMO support emerged from extensive discussions with physicians, who expressed a clear preference for systems that prioritize the detection of clinical decline over improvement. Due to the patient cohort's significant unpredictability and volatility, physicians find it particularly challenging to anticipate relevant clinical dynamics (e.g., deterioration and improvement). Ultimately, this uncertainty hinders the physicians' ability to track clinical progression across ECMO support effectively.

The early identification of patients at risk of clinical decline facilitates timely intervention, potentially preventing further deterioration and reducing the need for additional resources (human and material), as acknowledged by physicians and emphasized by Noy et al. [28]. Similarly, recognizing signs of clinical improvement early on may allow the cessation of therapeutic interventions carrying potentially harmful side effects, enhancing the availability of often scarce resources and ultimately enabling faster and more complete recovery.

These challenges and opportunities formed the basis for developing a risk score designed to identify and ideally anticipate periods of clinical deterioration and improvement throughout ECMO support. The risk score should enable physicians to intervene early and adopt treatment strategies to prevent severe, potentially irreversible (or even fatal) clinical deterioration and promote more favorable recovery.

The development and optimization of the risk score followed a comprehensive methodology, described in section 5.2.1, driven by the necessity to ensure more effective identification of clinical trends during ECMO support, unattainable through simply considering instance-based probability information. Physicians highlighted the importance of understanding patient dynamics over time, as isolated probability estimates can be misleading, particularly considering the patient cohort's critical clinical state and significant volatility.

The following two scenarios, which illustrated these issues, were observed for several patients in the cohort:

- **Scenario 1:** during certain intervals of hospitalization, the ML model consistently outputs elevated probability estimates, suggesting a high risk of clinical deterioration. However, since patients under ECMO support typically present a high risk of deterioration, this information alone is insufficient, as what truly matters to physicians is identifying and ideally anticipating clinical trends within these intervals. A decreasing trend in the risk score, despite elevated probability estimates, may signal that the patient's clinical condition is improving, suggesting that ECMO support is having a positive impact (this is a particularly relevant input since the invasive nature of ECMO support makes it prone to negatively affect the patient's clinical condition in subtle ways that may often go undetectable). Therefore, assessing the trajectory (rather than exclusively focusing on sporadic observations) of these clinical indicators can offer more valuable insights into the patient's clinical progression and response to treatment.

- **Scenario 2:** during certain intervals of hospitalization, the ML model consistently outputs low probability estimates, suggesting a lower risk of clinical deterioration. However, this does not necessarily reflect the success of ECMO support, as the underlying trend may still reveal that the patient's clinical condition is worsening, indicating that ECMO support and supplementary therapeutic interventions may require some adjustment. Conversely, a downward trend in the risk score would suggest ECMO support is succeeding in stabilizing the patient's clinical condition. Ultimately, this highlights how exclusively considering the model's probability estimates for each instance individually without assessing underlying trends may be misleading and lead physicians to draw inaccurate conclusions.

Given these insights, relying solely on individual (or static) probability estimates was deemed insufficient (i.e., it did not enable physicians to become better equipped to make better-informed decisions). Instead, a more comprehensive expression (equation 5.18) was established for the risk score, integrating two critical components: **smoothed probability estimates** and **derivative (trend)**, aimed at reflecting the immediate risk and the underlying clinical dynamics, respectively. Achieving an adequate balance between these components required extensive optimization, which focused on determining the optimal values for the following three parameters (optimization parameters): **smoothing window** (controls the extent of smoothing), **trend window** (defines the time frame for trend analysis), and α (weighting factor to balance the contribution of both components). Optimizing these parameters involved a comprehensive process, detailed in section 5.2.1, focused on maximizing the risk score's predictive utility (i.e., ensuring the risk score aligns with medical assessment during periods of clinical deterioration and improvement, which was expected considering the model's performance in distinguishing between clinical deterioration and improvement, discussed in section 6.1.4, and that it is capable of anticipating such periods). The results obtained are presented in section 6.2.2.

6.2.2 Evaluation and Selection

As outlined in section 5.2.2, assessing the performance of the risk score across the patient cohort involved employing qualitative and quantitative methods. These evaluations were conducted iteratively, with each iteration referring to the application of a specific combination of values for the optimization parameters (smoothing window, trend window, and α). The optimal risk score was ultimately represented via equation 6.1.

$$s(t) = p_s(t) + 1.5 \cdot \frac{dp_s(t)}{dt} \quad (6.1)$$

In this optimal configuration, the smoothing window was set to five consecutive instances (i.e., 8-hour windows), corresponding to a period of 16 hours, the trend window to three successive instances, equivalent to a period of 8 hours, and α was assigned the value of 1.5, ensuring optimal balance between both components of the risk score's mathematical formulation.

This configuration ensured compliance with the predefined optimization criteria, detailed in section 5.2.2. The 16-hour smoothing window effectively reduced short-term fluctuations (deemed uninforma-

tive by physicians) in the model's probability estimates throughout ECMO support. At the same time, it avoided excessive signal distortion, thus ensuring the preservation of relevant clinical trends. In contrast, the shorter 8-hour trend window enabled the risk score to capture critical shifts in the patient's clinical condition, a dynamic frequently observed due to the cohort's inherent unpredictability and volatility. Qualitative assessment of the risk score across the multiple patients in the cohort revealed that shorter trend window durations were insufficient for capturing meaningful clinical trends and patterns. Conversely, longer windows tended to overlook brief, often sudden shifts in the patient's clinical condition requiring immediate attention, thus compromising the physicians' ability to act early and prevent negative consequences.

The sensitivity of the risk score to sudden clinical shifts is demonstrated in figures 6.6 and 6.7, which present the risk score for patients 52 and 76 of the cohort, respectively, computed using the optimal configuration (equation 6.1).

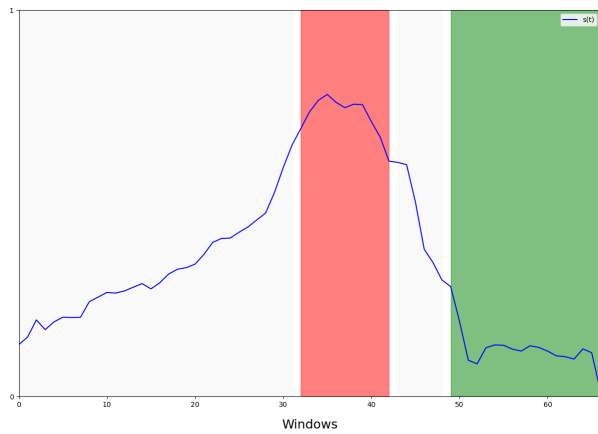


Figure 6.6: Risk score for patient 44 of the study cohort.

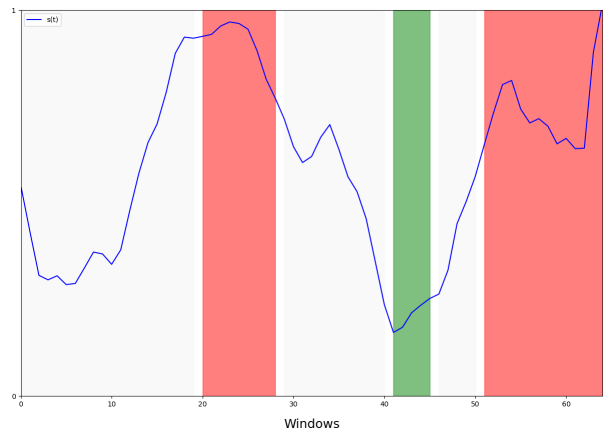


Figure 6.7: Risk score for patient 76 of the study cohort.

The ability of the risk score to exhibit upward and downward trends within time windows preceding clinical deterioration (red background) and improvement (green background), respectively, constituted a key performance indicator, as discussed in section 5.2.2. This behavior, consistently observed across the multiple patients in the cohort, enhanced the risk score's value in supporting early detection of critical clinical trends (both negative and positive), enabling physicians to make better-informed clinical decisions, intervene early, and take preventive measures whenever possible (e.g., interrupting a therapeutic intervention that may be harming the patient).

The scenarios outlined in section 6.2.1 reflect some of the most challenging clinical dynamics to assess in this patient cohort. Despite this complexity, the risk score was able to effectively identify critical trends, as demonstrated in the following cases:

- In figure 6.8, the risk score depicts a clear downward trend during a period identified by physicians as clinical deterioration (between windows 21 and 27). This phase eventually transitioned to clinical uncertainty and later improvement, suggesting that ECMO support and supplementary therapeutic interventions may have had a positive effect. The risk score's behavior, signaling improvement despite elevated risk estimates and the physicians' assessment, highlights its ability to capture

subtle positive clinical changes, some undetectable or even inaccessible to physicians.

- In figure 6.9, the risk score shows the opposite behavior, exhibiting an upward trend during a period marked by physicians as clinical progression (between windows 15 and 22). This phase eventually transitioned to clinical uncertainty and later deterioration, suggesting that ECMO support and supplementary therapeutic interventions may have been insufficient. This upward trend, despite the low-risk estimates initially observed, highlights the risk score's sensitivity to emerging clinical complications (often undetectable and even inaccessible to physicians), ultimately prompting physicians to intervene earlier and reassess the patient's requirements and clinical condition.

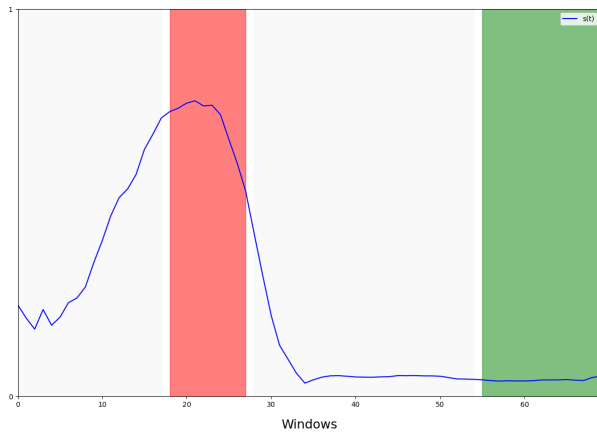


Figure 6.8: Risk score for patient 53 of the study cohort.

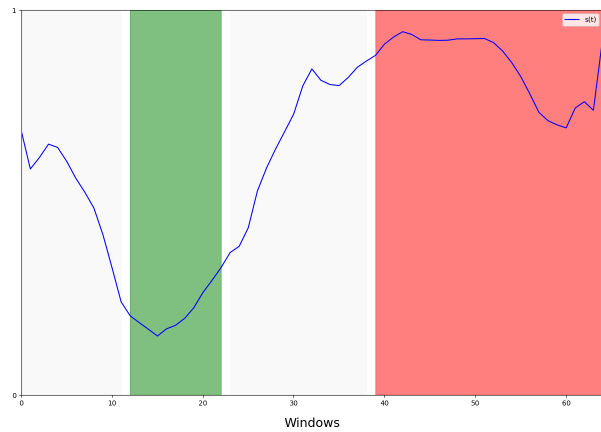


Figure 6.9: Risk score for patient 57 of the study cohort.

As demonstrated, evaluating the performance of the risk score primarily relied on qualitative analysis (i.e., assessing the score's behavior across the patients within the cohort). Quantitative assessment essentially served as a foundation for future advancements, such as the potential development of an alarm-generating system that provides a real-time estimate of the patient's clinical condition throughout ECMO support, signaling clinical deterioration and improvement, ultimately supporting medical evaluation and decision-making.

ROC curves, computed through a process described in section 5.2.2, aimed to quantify the risk score's ability to anticipate periods of clinical deterioration and improvement. The ROC curve's axes (FPR and TPR) were calculated by iteratively applying varying decision thresholds over time windows of 4, 8, and 12 hours preceding periods of clinical deterioration and improvement. The Area Under the Receiver Operating Characteristics (AUROC) curve provided an objective metric that guided the selection of the optimal values for the optimization parameters of the risk score's mathematical formulation. Ultimately, this quantitative assessment of the risk score's performance provided relevant insights into the model's predictive capabilities. **The results obtained for the optimal risk score configuration (equation 6.1) are illustrated in figure 6.10, with the risk score achieving an AUROC of 0.9176, 0.8944, and 0.8556 for time windows of 4, 8, and 12 hours preceding clinical deterioration and improvement, respectively.**

Figure 6.10 demonstrates that the model's predictive accuracy improves as the prediction time nears the onset of clinical deterioration and improvement. This outcome aligns with medical

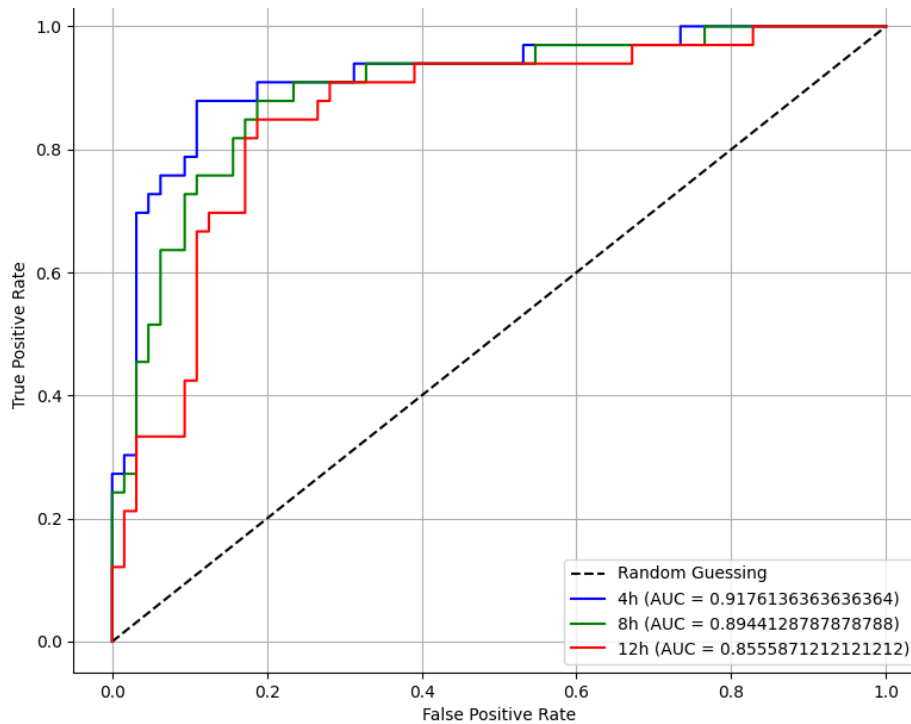


Figure 6.10: ROC curves obtained for the risk score considering time windows of 4, 8, and 12 hours preceding periods of clinical deterioration and improvement.

expectations as the patient's clinical features become increasingly pronounced, thus reflecting more evidently the shift in their clinical dynamics.

This analysis highlighted the relevance of ensuring a balance between sensitivity and specificity. For instance, a risk score designed to maximize sensitivity at the expense of compromising specificity would be highly adept at predicting impending clinical deterioration but ineffective at suppressing alarms when deterioration is absent, thus potentially resulting in an excessive number of false alarms, which would exacerbate alarm fatigue in an environment already overwhelmed by alarm activity. Such an imbalance could also lead to resource misallocation, possibly aggravating the problem related to the low staff-to-patient ratio typically observed under these conditions, ultimately compromising the physicians' ability to manage this patient cohort. Conversely, a risk score designed to maximize specificity at the expense of compromising sensitivity would reduce false alarm generation, however, it would risk missing critical clinical dynamics (indicative of deterioration), which could lead to misleading medical evaluations, delayed interventions, and ultimately worse patient outcomes. **Based on these considerations, the decision was made to implement the risk score as a real-time monitoring tool rather than developing a threshold-based alarm-generating system.**

Alternative implementations (e.g., threshold-based alarm-generating system) would require expanding the study to include a larger dataset, which could be achieved by considering extended hospitalization periods and incorporating data from additional medical institutions (rather than solely relying on a single one). This expansion is essential for conducting the necessary validation and testing, particularly relevant given the critical and highly volatile nature of critical care medicine (namely ECMO support). The following chapter, which concludes this dissertation, outlines promising avenues for future research.

Chapter 7

Conclusions and Future Work

This dissertation has demonstrated the potential of establishing partnerships between multidisciplinary institutions, granting access to exclusive resources crucial for conducting innovative studies with a valuable contribution to the scientific community. In collaboration with the Intensive Care Department at Hospital de Santa Maria, this study focused on a particularly complex scenario: managing critically ill patients under ECMO support. These patients, characterized by significant clinical volatility and uncertainty, are constantly subject to the application of advanced monitoring equipment, resulting in an extended and uninterrupted data stream. Analyzing this data is inherently complex, requiring a degree of expertise that few physicians possess.

Chapter 2 introduced the concept of intelligent ICUs, highlighting the potential of integrating advanced systems (monitoring, communication, analytics, and alarming) complemented by emerging technologies such as AI. This integration facilitates more efficient analysis of the continuous flow of data generated within ICUs. Specifically, patients under ECMO support produced large volumes of data reflecting the intricate nature of this therapeutic intervention. However, physicians considered this data insufficient to capture underlying biological processes triggered throughout ECMO support and the physiological intricacies inherent in this technique. As a result, an additional set of variables was defined in collaboration with physicians (equations 2.1 to 2.8), forming the basis for computing derived variables that were subsequently integrated into the dataset, thus augmenting and enriching it.

The critical nature and complexity of ECMO support, typically applied when the patient exhibits severe cardiac and respiratory dysfunction, results in a notable lack of publicly available datasets referring to this intricate therapeutic intervention. Thus, chapter 3 primarily focused on reviewing methodological approaches employed in studies of a similar nature, providing the foundation for the strategies employed across this work.

Given the limitations imposed by the unavailability of publicly accessible datasets referring to ECMO patients, the first step in this study involved designing a robust data preparation pipeline, detailed in chapter 4. This process, involving physicians who supported and participated in this study, comprised the selection of data from 81 patients with COVID-19 pneumonia who underwent ECMO support, a relatively small cohort given the labor-intensive and time-consuming nature of manually assembling

multidomain and multidimensional datasets. Another essential step in the datasets' assembly process was the definition of a dynamic ternary outcome variable (with possible values including -1, 0, and +1), designed to track the clinical evolution of patients throughout ECMO support. The approach employed to label each patient's dataset remains open to future refinement to overcome some of the inherent limitations in this strategy.

The research methodology, detailed in chapter 5, combined both conventional (i.e., typically observed in the literature) and innovative approaches, aligning with the study's objectives. The first step involved transforming the sequential supervised learning problem into a conventional supervised learning format, a particularly relevant step given the study's inherent limitations (e.g., low data volume). This transformation enabled the application of standard ML models (SVM and Random Forest) to develop a risk score capable of generating accurate real-time estimates of the risk of clinical deterioration throughout ECMO support for each patient.

Despite the study's constraints, the results obtained were promising. The ML models accurately distinguished between periods of clinical deterioration and improvement, with the SHAP methodology proving effective at making model predictions interpretable (a particularly relevant feature for physicians considering the critical nature and complexity of ECMO support). Ultimately, the computed risk score demonstrated a remarkable ability to anticipate underlying clinical dynamics (deterioration and improvement), thus reinforcing the potential of automatic systems capable of processing, analyzing, and learning from complex datasets to support medical evaluation and decision-making in intensive care.

Despite the study's promising results, several limitations were encountered across different stages, suggesting opportunities for future development. Key areas where further investigation could effectively address these limitations, thereby improving the quality, reliability, and ultimately generalizability of results, are summarized below.

The first objective out of the research objectives outlined in section 1.2 (developing an alternative data acquisition and registration infrastructure) was partially achieved, as specific segments of code developed in this study could be integrated into the ICU information system to enable the automatic transformation of raw, unstructured data into a format suitable for analysis, thus enhancing interpretability and usability. However, fully achieving this objective would require a broader knowledge base and additional resources, including direct collaboration with medical equipment manufacturers to ensure compatibility between devices and information systems (i.e., enable data transmission between source and receiver), as an in-depth understanding of each device's settings and configurations are often required.

Overcoming data acquisition and registration challenges is essential to address a significant limitation of this study: limited data availability. The criteria for establishing the patient population yielded a relatively small cohort, reflecting the labor-intensive, time-consuming nature of the data acquisition process. Ultimately, this limitation influenced several multiple methodological decisions across this study. Despite the results' robustness, a large, more diverse, and representative dataset would enhance the study's reliability and generalizability. The following two approaches are easily implementable and can help to mitigate or even overcome the outlined limitations:

- Instead of exclusively considering the first 12 days of hospitalization under ECMO support (or the

complete period of hospitalization in case the patient recovered or died before the established period of 12 days), data collection should contemplate the complete period of hospitalization for each patient (i.e., from ECMO implementation to ECMO discharge).

- The patient selection criteria should be reassessed to include patients with varying clinical situations undergoing ECMO support, allowing the establishment of a larger and more diverse patient cohort. This expansion can easily be achieved since the datasets considered, despite referring to patients with a primary diagnosis of COVID-19 pneumonia undergoing ECMO support, are exclusive to all ECMO patients, incorporating data and information available for these patients, regardless of the inherent clinical situation.

Despite its innovative nature, the data labeling strategy employed in this study was notably labor-intensive and time-consuming, requiring expert physicians to dedicate significant time to this process, which represents a challenging commitment considering their often work overload. These challenges foster additional consideration of alternative strategies. A straightforward approach would involve recruiting additional physicians to distribute the labeling workload more efficiently. Additionally, data augmentation techniques could synthetically expand the datasets, presenting an alternative pathway. Ultimately, more advanced solutions could include methods aimed at automating the labeling process, such as Active Learning, which selects the most informative samples for expert labeling, thereby reducing the overall efforts associated with this process.

With increased data availability, more advanced methodologies (e.g., the sliding window method configuration described in section 3.1.2) and algorithms (e.g., neural networks) could be applied to harness greater computing power, meaning the enhanced ability for data analysis, processing, and learning.

The core focus of this study was the development of a risk score aimed at providing real-time estimates of the risk of clinical deterioration for patients under ECMO support. As detailed in section 6.2.2, the risk score demonstrated notable performance, accurately identifying and anticipating periods of clinical deterioration and improvement. However, considering the critical nature of the environment for which the risk score was conceived, further refinement is crucial to enhance its interpretability, reliability, and applicability. The following approaches present promising pathways to achieve these objectives:

- Integrate the SHAP method into the risk score's real-time predictions, enabling dynamic visualizations of feature importance. This integration could enable physicians to conduct a more precise analysis of the risk score's output by providing access to critical information for understanding relevant underlying feature trends and dynamics deemed critical by the score.
- Consider more objective outcomes to assess the risk score's performance, such as whether the patient recovered from ECMO support (i.e., successful removal of ECMO support) or died. A well-performing risk score should demonstrate the ability to generate risk estimates that align with these outcomes.

Ultimately, integrating the risk score into the ICU provides physicians with a supplementary tool for

supporting clinical evaluation and decision-making. As a result, the risk score can also play a decisive role in enhancing ICU management through optimized resource allocation, as the data-driven insights it produces could provide less experienced physicians with critical information for supporting their ability to assess patients under ECMO support (simultaneously contributing to boosting their confidence in recognizing early warning signs requiring specialized attention), ultimately freeing up more experienced physicians and enabling them to allocate their time more efficiently.

The study's preliminary nature opens numerous avenues for future research and development, fitting into a future where growing collaboration between physicians and advanced, automatic systems is expected, ultimately improving the quality of care and patient outcomes.

Bibliography

- [1] American College of Physicians. *Critical Care Medicine*. Accessed on the 6th of June 2024. Available from: <https://www.acponline.org/about-acp/about-internal-medicine/subspecialties-of-internal-medicine/critical-care-medicine>.
- [2] Mayo Clinic. *Extracorporeal membrane oxygenation (ECMO)*. Accessed on the 8th of June 2024. Available from: <https://www.mayoclinic.org/tests-procedures/ecmo/about/pac-20484615>.
- [3] Morris, A. H. (2018). Human Cognitive Limitations. Broad, Consistent, Clinical Application of Physiological Principles Will Require Decision Support. *Annals of the American Thoracic Society*, 15(Suppl 1), S53-S56. URL: <https://doi.org/10.1513/AnnalsATS.201706-449KV>.
- [4] Komorowski, M. (2019). Artificial intelligence in intensive care: are we there yet? *Intensive Care Medicine*, 45(9), 1298-1300. URL: <https://doi.org/10.1007/s00134-019-05662-6>.
- [5] Universidade de Lisboa. 7.^a Conferência Anual da RedeSAÚDE. Accessed on the 25th of October 2024. Available from: <https://www.ulisboa.pt/evento/7a-conferencia-anual-da-redesaude>.
- [6] IEEE Engineering in Medicine and Biology Society. 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Accessed on the 25th of October 2024. Available from: <https://embc.embs.org/2024/>.
- [7] Kelly, F. E., Fong, K., Hirsch, N. and Nolan, J. P. (2014). Intensive care medicine is 60 years old: the history and future of the intensive care unit. *Clinical medicine (London, England)*, 14(4), 376-379. URL: <https://doi.org/10.7861/clinmedicine.14-4-376>.
- [8] West, J. B. (2005). The physiological challenges of the 1952 Copenhagen poliomyelitis epidemic and a renaissance in clinical respiratory physiology. *Journal of applied physiology (Bethesda, Md.: 1985)*, 99(2), 424-432. URL: <https://doi.org/10.1152/japplphysiol.00184.2005>.
- [9] Reisner-Sénélar, L. (2011). The birth of intensive care medicine: Björn Ibsen's records. *Intensive care medicine*, 37(7), 1084-1086. URL: <https://doi.org/10.1007/s00134-011-2235-z>.
- [10] Hermann, B., Benghanem, S., Jouan, Y., Lafarge, A., Beurton, A. and ICU French FOXES (Federation Of eXtremely Enthusiastic Scientists) Study Group (2023). The positive impact of COVID-19 on critical care: from unprecedented challenges to transformative changes, from the perspec-

- tive of young intensivists. *Annals of Intensive Care*, 13(1), 28. URL: <https://doi.org/10.1186/s13613-023-01118-9>.
- [11] Zelechower, J., Astudillo, J., Traversaro, F., Redelico, F., Luna, D., Quiros, F., San Roman, E. and Risk, M. (2017). Infrastructure for Big Data in the Intensive Care Unit. *Studies in Health Technology and Informatics*, 245, 1346. URL: <https://doi.org/10.3233/978-1-61499-830-3-1346>.
- [12] Mao, Z., Liu, C., Li, Q., Cui, Y. and Zhou, F. (2023). Intelligent Intensive Care Unit: Current and Future Trends. *Intensive Care Research*, 1-7. URL: <https://doi.org/10.1007/s44231-023-00036-5>.
- [13] De Georgia, M. A., Kaffashi, F., Jacono, F. J. and Loparo, K. A. (2015). Information technology in critical care: review of monitoring and data acquisition systems for patient care and research. *The Scientific World Journal*, 2015, 727694. URL: <https://doi.org/10.1155/2015/727694>.
- [14] Pham, T., Brochard, L. J. and Slutsky, A. S. (2017). Mechanical Ventilation: State of the Art. *Mayo Clinic Proceedings*, 92(9), 1382-1400. URL: <https://doi.org/10.1016/j.mayocp.2017.05.004>.
- [15] Brodie, D. and Bacchetta, M. (2011). Extracorporeal membrane oxygenation for ARDS in adults. *The New England Journal of Medicine*, 365(20), 1905-1914. URL: <https://doi.org/10.1056/NEJMct1103720>.
- [16] Hill, J. D., O'Brien, T. G., Murray, J. J., Dontigny, L., Bramson, M. L., Osborn, J. J. and Gerbode, F. (1972). Prolonged extracorporeal oxygenation for acute post-traumatic respiratory failure (shock-lung syndrome). Use of the Bramson membrane lung. *The New England Journal of Medicine*, 286(12), 629-634. URL: <https://doi.org/10.1056/NEJM197203232861204>.
- [17] Zapol, W. M., Snider, M. T., Hill, J. D., Fallat, R. J., Bartlett, R. H., Edmunds, L. H., Morris, A. H., Peirce, E. C., 2nd, Thomas, A. N., Proctor, H. J., Drinker, P. A., Pratt, P. C., Bagniewski, A. and Miller, R. G., Jr (1979). Extracorporeal membrane oxygenation in severe acute respiratory failure. A randomized prospective study. *JAMA*, 242(20), 2193-2196. URL: <https://doi.org/10.1001/jama.242.20.2193>.
- [18] Australia and New Zealand Extracorporeal Membrane Oxygenation (ANZ ECMO) Influenza Investigators, Davies, A., Jones, D., Bailey, M., Beca, J., Bellomo, R., Blackwell, N., Forrest, P., Gattas, D., Granger, E., Herkes, R., Jackson, A., McGuinness, S., Nair, P., Pellegrino, V., Pettilä, V., Plunkett, B., Pye, R., Torzillo, P., Webb, S., Wilson, M. and Ziegenfuss, M. (2009). Extracorporeal Membrane Oxygenation for 2009 Influenza A(H1N1) Acute Respiratory Distress Syndrome. *JAMA*, 302(17), 1888-1895. URL: <https://doi.org/10.1001/jama.2009.1535>.
- [19] Almenyan, A. A., Albuduh, A. and Al-Abbas, F. (2021). Effect of Nursing Workload in Intensive Care Units. *Cureus*, 13(1), e12674. URL: <https://doi.org/10.7759/cureus.12674>.
- [20] Kahn, J. M., Yabes, J. G., Bukowski, L. A. and Davis, B. S. (2023). Intensivist physician-to-patient ratios and mortality in the intensive care unit. *Intensive Care Medicine*, 49(5), 545-553. URL: <https://doi.org/10.1007/s00134-023-07066-z>.

- [21] Blythe, R., Parsons, R., White, N. M., Cook, D. and McPhail, S. (2022). A scoping review of real-time automated clinical deterioration alerts and evidence of impacts on hospitalised patient outcomes. *BMJ Quality & Safety*, 31(10), 725-734. URL: <https://doi.org/10.1136/bmjqs-2021-014527>.
- [22] Baker, T. and Gerdin, M. (2017). The clinical usefulness of prognostic prediction models in critical illness. *European Journal of Internal Medicine*, 45, 37-40. URL: <https://doi.org/10.1016/j.ejim.2017.09.012>.
- [23] Romero-Brufau, S., Gaines, K., Nicolas, C. T., Johnson, M. G., Hickman, J. and Huddleston, J. M. (2019). The fifth vital sign? Nurse worry predicts inpatient deterioration within 24 hours. *JAMIA Open*, 2(4), 465-470. URL: <https://doi.org/10.1093/jamiaopen/ooz033>.
- [24] Jahandideh, S., Ozavci, G., Sahle, B. W., Kouzani, A. Z., Magrabi, F. and Bucknall, T. (2023). Evaluation of machine learning-based models for prediction of clinical deterioration: A systematic literature review. *International Journal of Medical Informatics*, 175, 105084. URL: <https://doi.org/10.1016/j.ijmedinf.2023.105084>.
- [25] Muralitharan, S., Nelson, W., Di, S., McGillion, M., Devereaux, P. J., Barr, N. G. and Petch, J. (2021). Machine Learning-Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review. *Journal of Medical Internet Research*, 23(2), e25187. URL: <https://doi.org/10.2196/25187>.
- [26] Lauritsen, S. M., Thiesson, B., Jørgensen, M. J., Riis, A. H., Espelund, U. S., Weile, J. B. and Lange, J. (2021). The Framing of machine learning risk prediction models illustrated by evaluation of sepsis in general wards. *NPJ Digital Medicine*, 4(1), 158. URL: <https://doi.org/10.1038/s41746-021-00529-x>.
- [27] Tomašev, N., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., Mottram, A., Meyer, C., Ravuri, S., Protsyuk, I., Connell, A., Hughes, C. O., Karthikesalingam, A., Cornebise, J., Montgomery, H., Rees, G., Laing, C., Baker, C. R., Peterson, K., Reeves, R., Hassabis, D., King, D., Suleyman, M., Back, T., Nielson, C., Ledsam, J. and Mohamed, S. (2019). A Clinically Applicable Approach to Continuous Prediction of Future Acute Kidney injury. *Nature*, 572(7767), 116-119. URL: <https://doi.org/10.1038/s41586-019-1390-1>.
- [28] Noy, O., Coster, D., Metzger, M., Atar, I., Shenhar-Tsarfaty, S., Berliner, S., Rahav, G., Rogowski, O. and Shamir, R. (2022). A machine learning model for predicting deterioration of COVID-19 inpatients. *Scientific Reports*, 12(1), 2630. URL: <https://doi.org/10.1038/s41598-022-05822-7>.
- [29] Liu, R., Greenstein, J. L., Granite, S. J., Fackler, J. C., Bembea, M. M., Sarma, S. V. and Winslow, R. L. (2019). Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. *Scientific Reports*, 9(1), 6145. URL: <https://doi.org/10.1038/s41598-019-42637-5>.
- [30] Garcia-Gutiérrez, S., Esteban-Aizpiri, C., Lafuente, I., Barrio, I., Quiros, R., Quintana, J. M., Uranga, A. and COVID-REDISSEC Working Group (2022). Machine learning-based model for prediction of

- clinical deterioration in hospitalized patients by COVID 19. *Scientific Reports*, 12(1), 7097. URL: <https://doi.org/10.1038/s41598-022-09771-z>.
- [31] Hyland, S. L., Faltys, M., Hüser, M., Lyu, X., Gumbsch, T., Esteban, C., Bock, C., Horn, M., Moor, M., Rieck, B., Zimmermann, M., Bodenham, D., Borgwardt, K., Rätsch, G. and Merz, T. M. (2020). Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature Medicine*, 26(3), 364-373. URL: <https://doi.org/10.1038/s41591-020-0789-4>.
- [32] Lucchini, A., Elli, S., De Felippis, C., Greco, C., Mulas, A., Ricucci, P., Fumagalli, R. and Foti, G. (2019). The evaluation of nursing workload within an Italian ECMO Centre: A retrospective observational study. *Intensive and Critical Care Nursing*, 55, 102749. URL: <https://doi.org/10.1016/j.iccn.2019.07.008>.
- [33] Montero, S., Slutsky, A. S. and Schmidt, M. (2018). The PRESET-Score: the extrapulmonary predictive survival model for extracorporeal membrane oxygenation in severe acute respiratory distress syndrome. *Journal of Thoracic Disease*, 10(Suppl 17), S2040-S2044. URL: <https://doi.org/10.21037/jtd.2018.05.184>.
- [34] Joshi, H., Flanagan, M., Subramanian, R. and Drouin, M. (2022). Respiratory ECMO Survival Prediction (RESP) Score for COVID-19 Patients Treated with ECMO. *ASAIO Journal (American Society for Artificial Internal Organs: 1992)*, 68(4), 486-491. URL: <https://doi.org/10.1097/MAT.0000000000001640>.
- [35] Raghavendran, K. and Napolitano, L. M. (2011). ALI and ARDS: Challenges and Advances. *Critical Care Clinics*, 27(3), 429-437. URL: <https://doi.org/10.1016/j.ccc.2011.05.006>.
- [36] Moreno, R., Rhodes, A., Piquilloud, L., Hernandez, G., Takala, J., Gershengorn, H. B., Tavares, M., Coopersmith, C. M., Myatra, S. N., Singer, M., Rezende, E., Prescott, H. C., Soares, M., Timsit, J. F., de Lange, D. W., Jung, C., De Waele, J. J., Martin, G. S., Summers, C., Azoulay, E., Fujii, T., McLean, A. S. and Vincent, J. L. (2023). The Sequential Organ Failure Assessment (SOFA) Score: has the time come for an update? *Critical Care (London, England)*, 27(1), 15. URL: <https://doi.org/10.1186/s13054-022-04290-9>.
- [37] Ayers, B., Wood, K., Gosev, I. and Prasad, S. (2020). Predicting Survival After Extracorporeal Membrane Oxygenation by Using Machine Learning. *The Annals of Thoracic Surgery*, 110(4), 1193-1200. URL: <https://doi.org/10.1016/j.athoracsur.2020.03.128>.
- [38] Lee, H., Song, M. J., Cho, Y. J., Kim, D. J., Hong, S. B., Jung, S. Y. and Lim, S. Y. (2023). Supervised machine learning model to predict mortality in patients undergoing venovenous extracorporeal membrane oxygenation from a nationwide multicentre registry. *BMJ Open Respiratory Research*, 10(1), e002025. URL: <https://doi.org/10.1136/bmjresp-2023-002025>.
- [39] Picis. Accessed on the 25th of May 2024. Available from: <https://www.picis.com>.

- [40] Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J. D., Coopersmith, C. M., Hotchkiss, R. S., Levy, M. M., Marshall, J. C., Martin, G. S., Opal, S. M., Rubenfeld, G. D., van der Poll, T., Vincent, J. L. and Angus, D. C. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 801-810. URL: <https://doi.org/10.1001/jama.2016.0287>.
- [41] Smiti, A. (2020). A critical overview of outlier detection methods. *Computer Science Review*, 38, 100306. URL: <https://doi.org/10.1016/j.cosrev.2020.100306>.
- [42] Centers for Disease Control and Prevention. Accessed on the 27th of May 2024. Available from: <https://www.cdc.gov>.
- [43] van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. URL: <https://doi.org/10.18637/jss.v045.i03>.
- [44] Papana, A. (2021). Connectivity Analysis for Multivariate Time Series: Correlation vs. Causality. *Entropy (Basel, Switzerland)*, 23(12), 1570. URL: <https://doi.org/10.3390/e23121570>.
- [45] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É. (2018). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. URL: <https://doi.org/10.48550/arXiv.1201.0490>.
- [46] Cortes, C. and Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297. URL: <http://dx.doi.org/10.1007/BF00994018>.
- [47] Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32. URL: <https://doi.org/10.1023/A:1010933404324>.
- [48] Silva Filho, T., Song, H., Perello-Nieto, M., Santos-Rodriguez, R., Kull, M. and Flach, P. (2023). Classifier calibration: a survey on how to assess and improve predicted class probabilities. *Machine Learning*, 112, 3211-3260. URL: <https://doi.org/10.1007/s10994-023-06336-7>.
- [49] Lundberg, S. and Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. *Neural Information Processing Systems*. URL: <https://doi.org/10.48550/arXiv.1705.07874>.
- [50] Rodríguez-Pérez, R. and Bajorath, J. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34(10), 1013-1026. URL: [1026.https://doi.org/10.1007/s10822-020-00314-0](https://doi.org/10.1007/s10822-020-00314-0).