University of Modena and Reggio Emilia
XXXVII cycle of the international doctorate school in
Information and Communication Technologies

Doctor of Philosophy dissertation in
Computer Engineering and Science

Digital Transformation and Machine Learning Applied to Public Healthcare

Paolo Perliti Scorzoni

Advisor: Prof. Costantino Grana
Co-Advisor: Prof. Federico Bolelli
PhD School Director: Prof. Luigi Rovati

Modena, 2025

Abstract

Public healthcare is essential to societal welfare, yet healthcare systems worldwide are increasingly burdened by demographic shifts, financial constraints, and delays in adopting enabling technologies. This thesis investigates how Computer Vision (CV) and Machine Learning (ML) can address these challenges within the context of the Italian National Health System (Servizio Sanitario Nazionale, SSN), proposing solutions to improve operational efficiency, service delivery, and resource management.

The dissertation presents practical applications of Artificial Intelligence through specific case studies. The first demonstrates the use of CV for automated recognition of nasopharyngeal swab results in COVID-19 self-testing, aimed at reducing administrative workload and supporting pandemic management. Subsequently, the study presents ML predictive models developed to forecast hospital Length of Stay (LOS), designed to optimize bed allocation and streamline patient flow in general hospital wards and Emergency Departments. Finally, the thesis introduces a novel Cooperative Electronic Bed Management System (CEBMS), which leverages real-time data and predictive analytics to enhance coordination and improve resource utilization across hospital networks.

Through the implementation of AI-driven solutions, this work highlights the transformative potential of digital technologies in public healthcare, providing concrete proposals to modernize infrastructure, consolidate decision-making processes, and uphold the quality and sustainability of the Italian SSN. These findings contribute to advancing the dialogue on digital transformation as a strategic imperative for the evolution of modern healthcare systems.

Sommario

La sanità pubblica costituisce un elemento cardine per il benessere della società, nondimeno i sistemi sanitari, a livello globale, risultano gravati da mutamenti demografici, vincoli economici e ritardi nell'adozione di tecnologie abilitanti. La presente tesi analizza come la Computer Vision (CV) e il Machine Learning (ML) possano rispondere a tali sfide nel Servizio Sanitario Nazionale (SSN) italiano, proponendo soluzioni per migliorare l'efficienza operativa, la qualità dell'assistenza e la gestione delle risorse.

La ricerca presenta alcune applicazioni pratiche dell'Intelligenza Artificiale attraverso casi di studio specifici. Il primo illustra l'impiego della CV per il riconoscimento automatico degli esiti dei tamponi nasofaringei nell'autodiagnosi del COVID-19, puntando alla riduzione dei carichi amministrativi e ad una gestione più efficace della pandemia. Successivamente, vengono presentati modelli predittivi di ML sviluppati per la stima della durata della degenza ospedaliera (Length of Stay, LOS), volti all'ottimizzazione dell'allocazione dei posti letto e alla razionalizzazione del flusso dei pazienti all'interno dei reparti ospedalieri e nei dipartimenti di emergenza. Infine, la tesi introduce un innovativo Sistema Elettronico per la Gestione Cooperativa dei Posti Letto (CEBMS), il quale sfrutta dati in tempo reale e analisi predittive per potenziare il coordinamento e migliorare l'utilizzo delle risorse intra- e inter-ospedaliere.

Attraverso l'implementazione di soluzioni basate sull'intelligenza artificiale, il presente lavoro evidenzia il potenziale trasformativo delle tecnologie digitali nella sanità pubblica, formulando proposte concrete per l'ammodernamento delle infrastrutture, il consolidamento dei processi decisionali e la salvaguardia della qualità e della sostenibilità dell'SSN italiano. Tali risultati contribuiscono al dibattito sulla trasformazione digitale quale elemento strategico per il progresso dei sistemi sanitari moderni.

Review committee composed of:
Prof. Prati Andrea, Università degli Studi di Parma
Prof. Murino Vittorio, Università degli Studi di Verona

Never let a good crisis go to waste.
(Winston L. S. Churchill)

Contents

List of Figures

List of Tables

Listings

List of Abbreviations

ACC	Accuracy
ADASYN	Adaptive Synthetic sampling
ADT	Admission, Discharge, Transfer
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUPRC	Area Under the Precision-Recall Curve
AUROC	Area Under the Receiver Operating Characteristic Curve (AUROC)
AUSL	Azienda Sanitaria Locale (English: Local Health Authority)
CLI	Command-Line Interface
CNN	Convolutional Neural Network
CSV	Comma-Separated Value
CT	Computed Tomography
CV	Computer Vision
DL	Deep Learning
DNN	Deep Neural Network
DT	Digital Transformation
EBMS	Electronic Bed Management System
ED	Emergency Department
EHR	Electronic Health Record
EMR	Electronic Medical Record
FSE	Fascicolo Sanitario Elettronico (English: Electronic Health Record (EHR))
GB	Gradient Boosting Tree
HDR	Hospital Discharge Register

HIS	Hospital Information System
ICU	Intensive Care Unit
KNN	K-Nearest Neighbor
LGBM	Light Gradient-Boosting Machine (LightGBM)
LOS	Length of Stay
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
OECD	Organization for Economic Cooperation and Development
PCA	Principal Component Analysis
PLOS	Prolonged Length of Stay
PNRR	Piano Nazionale di Ripresa e Resilienza
PRC	Precision
REC	Recall
RER	Regione Emilia-Romagna (Italy)
RF	Random Forest
RMSE	Root Mean Squared Error
SDO	Scheda di Dimissione Ospedaliera (English: Hospital Discharge Register (HDR))
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
UI	User Interface
WHO	World Health Organization
XGB	Extreme Gradient Boosting (XGBoost)
YOLO	You Look Only Once

Acknowledgments

By now, my family knows: when I'm facing a crisis, it's not *if* things will go wrong, but *how* badly they'll go. An industrial PhD is not the most predictable outcome, but definitely an interesting one.

Let's begin with the acknowledgments. To my advisors, Costantino Grana and Federico Bolelli: with professionalism and boundless patience, they managed to guide me to the end of this journey. I'm grateful for their encouragement, but most of all, for their honesty. Being told, "This is a terrible idea, but do it anyway to see what happens," is exactly what you need to survive a PhD. A heartfelt thank you.

I'm also thankful to the Department of Engineering "Enzo Ferrari" and UNIMORE for welcoming me back. It was both pleasant and slightly embarrassing, especially with all those freshmen who assumed I was a teacher (I let them believe it, enjoying the perks of seniority).

My appreciation extends to Miliaris, the company that sponsored my PhD and where I've invested almost twenty years of my life. It's like a second home, except you have to pay for coffee. Thanks to all my colleagues and friends for their support, affection, and occasional sympathetic glances.

To my family, who endured *all* my cyclical phases of panic and exhaustion: I suspect that, at some point, they started putting something in my tea. Thank you for reminding me that there's life beyond the monitor.

Finally, a loving thought goes to my father. Before passing away, he told me to invest in education. You were right, Dad. I'm trying to instill this wisdom in my kids. The prospect of becoming a successful YouTuber is still prevailing, but I'm working on it.

For those just starting out: be curious, don't settle for easy answers, put in the effort, and believe in yourself. We need you.

Chapter 1

Introduction

Public healthcare constitutes a fundamental pillar for the welfare and development of numerous nations, including Italy, where the Constitution enshrines and guarantees universal and free access to primary care. Nevertheless, during the past twenty years, the Italian healthcare system, like many others worldwide, has confronted increasingly complex challenges that compromise its sustainability over time. Among the most significant issues are persistent reductions in public expenditure, imposed by the need to contain national debt; the progressive demographic shift towards an aging population and its concomitant rise in chronic disease burden and long-term care requirements; and a significant delay in the digital transformation of public infrastructures, exacerbated by a fragmented and inconsistent management across regions, often resulting in ineffective coordination. These factors collectively contribute to increasing the pressure on the National Health System (SSN), limiting its ability to respond adequately and effectively to the growing and diverse needs of its users.

This dissertation aims to explore the application of Computer Vision (CV) and Machine Learning (ML) within the context of Italian public healthcare to address global challenges. Following a brief literature review on digital transformation and its implications for public health systems, particularly in Italy, the study presents a series of case studies and practical implementations across diverse healthcare settings nationwide. These examples highlight the transformative potential of CV and ML in enhancing operational efficiency, augmenting user-perceived quality, and optimizing resource management within the Italian healthcare system.

The first case study investigates the application of CV in the automated recognition of nasopharyngeal swabs for COVID-19 self-testing. A pilot experiment conducted in Emilia-Romagna illustrates how the implementation of an automated outcome classification system, integrated with the Electronic Health Record (FSE), can contribute not only to tracking new infections and containing viral spread but also to alleviating the workload of administrative and healthcare personnel during critical phases of a pandemic.

The dissertation then examines ML techniques for predicting the Length of Stay (LOS) of patients in hospital wards. In collaboration with Ospedale di Sassuolo, a general hospital in northern Italy, the study demonstrated that ensemble models, in particular, exhibit robust accuracy in estimating the duration of hospitalizations, favouring the timely identification of potential "bed blockers," more precise scheduling of discharges and admissions, and the efficient allocation of hospital beds.

A further application of ML addresses the prediction of prolonged LOS in Emergency Departments (EDs), whose management involves significantly different operational dynamics compared to traditional wards. By analyzing real-time information collected during the triage phase, the predictive models facilitate more effective management of patient flow in emergency-urgency care settings, enabling the implementation of targeted and proactive interventions.

Finally, the thesis introduces and describes the first Cooperative Electronic Bed Management System (CEBMS), a novel variant of a computerized bed management system, designed for Bed Managers and Hospital Administrations operating within the same territorial network. This system collects and aggregates data from primary hospital patient flows, such as ED and ADT (Admission/Discharge/Transfer), processing them through predictive models to provide valuable insights that support operational decision-making. CEBMSs also allow real-time sharing of reports and statistics among nearby hospitals, offering a comprehensive overview of bed occupancy and availability, as well as specific alert situations (*e.g.*, COVID-19 outbreaks). Moreover, they enable coordinated management of patient admissions, thereby fostering more effective synergy among hospital structures at the provincial level and improving delivery responsiveness to healthcare needs.

Digital transformation represents not only an imperative objective but also a strategic opportunity to reform healthcare systems, modernize infrastructures, and strengthen resilience in addressing contemporary challenges —such as the 2020 pandemic— and future demands. Drawing from empirical evidence and implementation outcomes, this work offers a constructive contribution to the ongoing debate on how the adoption of Artificial Intelligence (AI) in Italian healthcare settings can drive meaningful advancements, while maintaining an unwavering focus on the quality of care provided to patients and the sustainability of a public healthcare system internationally recognized for its excellence.

The complete source code developed during this doctoral research is available on GitHub.

Activities Carried Out During the PhD

Advanced Training Programs

– *H2020 - School in AI:* The “School in AI: Deep Learning, Vision and Language for Industry” is an intensive two-week educational program (in English) held at UNIMORE. It consists of an initial week of “foundational” AI training followed by an advanced strengthening phase, culminating in a practical project development phase where participants collaborate with industry partners under expert tutorial guidance. The program is fully funded by the Regione Emilia-Romagna through the “Advanced Schools in Artificial Intelligence in Emilia-Romagna” initiative. The school is organized in partnership with AIACADEMY, AIMAGELAB, and AIRI, and is part of both the AI4CH project and the Horizon 2020 HumanE-AI-Net project.

– *H2020 - DeepHealth Winter School:* The “Deep-Learning and HPC to Boost Biomedical Applications for Health (DeepHealth)” project, funded by the European Commission, aims to develop a unified framework that leverages heterogeneous HPC and Big Data architectures to support new and more efficient ways of diagnosis, monitoring and treatment of diseases. Two libraries, the European Distributed Deep Learning Library (EDDLL) and the European Computer Vision Library (ECVL), have been developed and incorporated into the DeepHealth framework for manipulating and processing the images in a more efficient way, increasing the productivity of professionals working on biomedical images.

Conferences Attended

– International Conference on Intelligent Human Systems Integration - IHSI, Roma, Italy, 2025

– Artificial Intelligence for Healthcare Applications - AIHA, Kolkata, India, 2024

– International Conference on Human Interaction and Emerging Technologies - IHIET, Venezia, Italy, 2024

– International Conference on Image Analysis and Processing - ICIAP, Udine, Italy, 2023.

Seminars Attended

– "Understanding "Searching as Learning" in the AI Era" —Prof. Chei Sian Lee (Nanyang Technological University, Singapore)— IUAV University, Venice, Italy, August 27, 2024.

– "Establishing Sustainable Health Services for the Medication of Elderly Chronic Diseases" —Dr. Yuxin Sheng (Southeast University, Nanjing, Jiangsu, China)— IUAV University, Venice, Italy, August 27, 2024.

– "The Basic Principles of Project Management" —Prof. Massimo Bertolini (UNIMORE)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, May 7, 2024.

– "Towards Robot Consciousness" —Prof. Antonio Chella (Università degli Studi di Palermo)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, January 26, 2024.

– "SQL and Large Language Models: A Marriage Made in Heaven?" —Prof. Paolo Papotti (EURECOM, France)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, October 30, 2023.

- "Federated Learning in Healthcare: Challenges and Research Directions" —Dr. Mirko Polato (Università di Torino, Italy)— Università di Udine, Italy, September 11, 2023.
- "Morphing-Attacks against Binary Fingervein Templates" —Dr. Tobias Mitterreiter (University of Salzburg, Austria)— Università di Udine, Italy, September 11, 2023.
- "Algebraic Vision" —Prof. Tomas Pajdla (Czech Technical University, Prague)— Università di Udine, Italy, September 12, 2023.
- "AI Hardware and Real-World AI" —Prof. Andrew Fitzgibbon (Microsoft)— Università di Udine, Italy, September 13, 2023.
- "Data-driven Learning-based Surface Anomaly Detection" —Prof. Danijel Skočaj (University of Ljubljana, Slovenia)— Università di Udine, Italy, September 14, 2023.
- "Academic English Workshop II" —Prof. Silvia Cavalieri (UNIMORE)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, June 13, 2023.
- "Academic English Workshop I" —Prof. Silvia Cavalieri (UNIMORE)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, May 2, 2023.
- "IP4ENGINEERS" —Prof. Isabella Ferrari (UNIMORE)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, March 14, 2023.
- "Trends and Challenges in Machine Learning-based Malware Detection" —Prof. Fabio Pierazzi (King's College, London)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, October 19, 2022.
- "How to Implement Your Own Deep Learning Framework" —Prof. Roberto Paredes (Universitat Politècnica de València, Spain)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, September 14, 2022.

- "Programming with Python" —Prof. Julien Bloino (Scuola Normale Superiore di Pisa, Italy)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, July 18, 2022.

- "Machine Learning Applications in Trading and Portfolio Management" —Prof. Petter Kolm (New York University, USA)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, June 7, 2022.

- "The Evolution of the Methods of Software Engineering" —Dr. Fabio Mora (Italian Agile Movement)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, May 19, 2022.

- "Modern Analytical Science Needs Modern Data Analysis Approach. Artificial Intelligence in Scientific Data Analysis" —Prof. José Amigo (University of the Basque Country, Spain)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, April 11, 2022.

- "Kernel-based Methods in Classification and Regression" —Prof. Federico Marini (Università di Roma "La Sapienza", Italy)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, April 7, 2022.

- "Research Challenges in Leonardo Labs: Applied Deep Learning in the Industry" —Dr. Alessandro Nicolosi (Leonardo Labs R&D, Italy)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, April 6, 2022.

- "High-Performance Computing and Large-Scale Models" —Prof. Giuseppe Fiameni (UNIMORE)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, February 24, 2022.

- "Explainable Artificial Intelligence (XAI)" —Prof. Natalia Díaz-Rodríguez (University of Granada, Spain)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, February 22, 2022.

- "Continual Learning for AI" —Dr. Vincenzo Lomonaco (Università di Pisa, Italy)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, February 22, 2022.

- “Artificial Intelligence for Bioinformatics” —Prof. Elisa Ficarra (UNIMORE)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, February 22, 2022.
- “Adventures in Geometric Deep Learning” —Prof. Emanuele Rodolà (Università di Roma “La Sapienza”, Italy)— Università degli Studi di Modena e Reggio Emilia, Modena, Italy, February 21, 2022.

Online Courses and Certifications

- “Regolamento AI ACT: A cosa devono prestare attenzione le aziende?” —Avv. Sofia Telese— Consulenti Privacy, November, 2024.
- “PyTorch for Deep Learning with Python Bootcamp” —Prof. Jose Marcial Portilla— Udemy, May, 2023.
- “Python for Computer Vision with OpenCV and Deep Learning” —Prof. Jose Marcial Portilla— Udemy, February, 2023.
- “NLP - Natural Language Processing with Python” —Prof. Jose Marcial Portilla— Udemy, August, 2022.
- “Python for Machine Learning & Data Science Masterclass” —Prof. Jose Marcial Portilla— Udemy, July, 2022.
- “Python and Flask Bootcamp” —Prof. Jose Marcial Portilla— Udemy, April, 2021.
- “Python and Django Full Stack Web Developer Bootcamp” —Prof. Jose Marcial Portilla— Udemy, April, 2021.
- “Complete Python Bootcamp” —Prof. Jose Marcial Portilla— Udemy, March, 2021.
- “Trasformazione Digitale nella Pubblica Amministrazione” — Agenzia per l’Italia Digitale (AgID)— October 2019–July 2020.

Professional positions

- CTO at Miliaris: leading the development team of the company

Grants and Awards

– Best scientific communication for the project "MondrIAn: Smart Bed Management" at "47th National Congress of ANMDO (Italian National Association of Hospital Management)", Bologna, Italy, June 2022.

Chapter 2

Digital Transformation in Public Healthcare: a Brief Literature Survey

In this chapter, we briefly introduce the concept of Digital Transformation and its potential impact on public health systems, exploring both the opportunities and challenges it presents. Additionally, a short overview of the Italian context is provided.

2.1 A gentle introduction to Digital Transformation

The advent of COVID-19 in early 2020 [1] induced profound changes in business practices and economic structures worldwide, with ramifications expected to persist for years [2, 3, 4]. Among the most significantly impacted sectors, the healthcare industry has experienced unprecedented pressure on its institutional capacities [5, 6, 7], requiring structural reconfiguration and novel approaches [8]. A pivotal component in this adaptive process has been the expedited adoption of digital technologies, implemented at an extraordinary scale and pace [9, 10] throughout the healthcare landscape. This wide-ranging shift falls under the broader framework of

Digital Transformation (DT) [11].

As outlined in Regulation (EU) 2021/694 of the European Parliament and the Council of 29 April 2021, which established the Digital Europe Program, digital transformation refers to the integration of digital technologies into businesses and service processes, leading to significant organizational and operational changes [12]. Thus, digital transformation entails more than the mere adoption of technology, as it involves a fundamental redesign of how organizations function and deliver value. This process requires structural, cultural, and operational adjustments that extend beyond technological implementation alone. However, despite its increasing use in both academic and professional discourse, the term "digital transformation" still lacks a clear, universally accepted definition, reflecting the complexity and evolving nature of the phenomenon. This is particularly evident in the public health domain, where terms such as digitization, digitalization, and digital public health are frequently used interchangeably. Specifically:

- *Digital Public Health* refers to the incorporation of digital technologies into public health service delivery. The term gained prominence following Public Health England's release of its "digital-first" strategy in 2017 and has since been applied to a broad range of public health initiatives. Its growing emphasis is reflected in the increasing number of graduate programs, specialized journal issues and conferences, and intergovernmental forums focused on this subject.

- *Digitization* denotes the technical process of converting analog or physical information into digital formats, such as transforming paper-based patient records into electronic files that can be accessed and shared digitally.

- *Digitalization* involves using digital technologies to improve or automate existing processes. Examples in healthcare include digital patient check-ins, electronic prescription management, and the use of electronic health records (EHRs) to enhance administrative efficiency. While digitalization streamlines operations, it does not fundamentally alter the nature of healthcare services.

- *Digital Transformation*, by contrast, represents a deeper and more disruptive change. It not only improves but completely reconfigures

> healthcare systems through the integration of a wide range of technologies. This transformation is evident in innovations such as AI-powered diagnostics, predictive analytics using big data, and patient-centered care models supported by real-time data and IoT technologies.

Unlike digitization and digitalization, which represent different levels of complexity and integration of digital technologies in public health practices [13], digital transformation reshapes how healthcare services are delivered, altering organizational structures and the way stakeholders — including patients, clinicians, and administrators— interact within the system. This comprehensive process fosters innovation in healthcare management, research, and public health policy, with a focus on meeting evolving needs through a patient-centered approach. Technologies driving this transformation include telemedicine [14], wearable devices, 3D printing, AI-based diagnostics, big data analytics [15], and blockchain [16].

2.2 The Role of DT in Public Healthcare

The rising influence of digital technologies in healthcare and public health is reflected in the expanding body of scientific literature. Although research on Information and Communication Technology (ICT) and digitalization began nearly fourty years ago, the past decade has witnessed a surge in publications, particularly those examining the combination of ICT and DT, with the highest volume of articles published between 2019 and 2021 [17].

According to the World Health Organization's Health Innovation Group (WHIG), innovation in healthcare addresses unmet health needs through new processes, tools, therapies, medical procedures, and innovative approaches to education, training, management, and procurement. As demonstrated by the COVID-19 pandemic, digital transformation represents a key driver of healthcare innovation. The crisis accelerated the adoption of telehealth services and digital tools for managing patient care and tracking the spread of the virus, demonstrating the effectiveness of digital solutions in responding to public health emergencies. Since then, public health organizations have increasingly embraced digital technologies to strengthen disease surveillance, patient management, and health promotion initiatives.

Digital transformation also expedites the acquisition and analysis of large-scale clinical data. Through big data analytics, healthcare providers can identify emerging trends, bolster disease prevention strategies, and develop targeted interventions for specific populations [18, 19], ultimately leading to better outcomes. Technologies such as digital imaging support patient engagement and promote shared decision-making. Furthermore, digital platforms that integrate data across different healthcare services and regions can improve coordination among providers, minimize redundancies, reduce costs, and enable more efficient delivery of health services.

The digital transition is also revolutionizing managerial practices. As Bernardi and Exworthy [20] highlighted, "clinical leaders" will play a crucial role in driving innovation in information technology within healthcare settings. Clinical leaders, typically healthcare professionals with both medical expertise and management responsibilities, are uniquely positioned to bridge the gap between technological capabilities and clinical needs. Their dual understanding allows for more effective implementation of digital solutions that align with healthcare objectives and operational realities.

Collectively, these advancements in clinical practice, healthcare administration, patient engagement, and medical research are contributing to the development of more responsive and resilient public healthcare systems.

2.3 Medical Applications

The Fourth Industrial Revolution [21], characterized by the convergence of physical and digital technologies, has profoundly impacted the global healthcare sector over the last few years. While ICT has been part of healthcare since the 1980s [22], the concurrent adoption of cloud computing, augmented reality, and cognitive technologies has accelerated this transition, facilitating advancements in healthcare strategies, services, and management [23].

Data-intensive healthcare environments have made artificial intelligence an essential tool. Machine learning, specifically, enables the analysis of large datasets to identify complex patterns, reducing the need for manual investigation. As a result, AI is recognized as a powerful means of enriching decision-support systems, reducing clinical risks, and lowering costs, thus creating value for both providers and patients [24]. AI-driven applications are also transforming fields such as drug discovery (*e.g.*, expediting the process of identifying and developing new pharmaceuticals), image diagnostics (*e.g.*, detecting correlations between symptoms and diseases), medical device therapy (*e.g.*, artificial body parts that translate brain signals), and medical education through simulators [25].

The Internet of Things (IoT) is another emerging area of medical applications, involving the integration of physical objects into healthcare information networks. IoT enables smart systems capable of monitoring, tracking, and storing patient records for continuous care and analysis, thereby supporting remote observation, contact tracing [26], disease risk assessment, and predictive systems. Following the development of the internet, IoT is anticipated to drive the next major digital revolution, particularly through devices like health trackers, skin sensors, cardiac monitors, glucometers, and ingestible smart pills. These innovations empower individuals to monitor their health metrics and receive timely alerts. IoT has been widely applied in delivering effective healthcare services to diverse patient populations, including the elderly and those with chronic illnesses [27], demonstrating its potential in healthcare systems. When combined with personal mobile devices, IoT enables even greater innovations, such as patient support systems through digital apps that enhance compliance, accuracy, and monitoring, encouraging the broader adoption of these technologies [28].

The widespread adoption of AI and IoT raises important concerns regarding data privacy and integrity, as these technologies necessitate secure systems to manage device identities and data transfer. Blockchain technology has emerged as a promising solution, operating as a decentralized ledger that enables robust, transparent transactions while preserving data trustworthiness. In the context of healthcare, blockchain has the potential to redesign information exchange between users, IoT devices, and servers, contributing to greater trust and reliability: by enabling tamper-proof record-keeping, blockchain allows both patients and healthcare providers to share sensitive information with greater confidence. Additionally, blockchain's decentralized nature reduces the risk of data breaches by eliminating single points of failure, a common vulnerability in traditional centralized systems. This makes blockchain particularly valuable for applications such as patient records management [29], consent tracking [30], and the secure sharing of data across different healthcare institutions and platforms.

Cloud computing is another key component in healthcare [31], offering dynamic, transparent, and cost-effective platforms for sharing information and hosting critical services. Nowadays, many healthcare services are delivered through Software as a Service (SaaS) models, ranging from patient scheduling and billing to advanced diagnostic tools. A major application is the provisioning of Electronic Health Records on regional data centers, which allows for secure and efficient management, storage, and retrieval of patient data across multiple healthcare facilities, ensuring the availability of medical information and supporting coordination of care. Another example is the integration and optimization of hospital supply chains [32, 33], which streamline inventory management and operational processes. Cloud-based solutions reduce the need for expensive on-site hardware and maintenance while guaranteeing high service-level agreements (SLAs), regulatory compliance (*e.g.*, HIPAA), and the flexibility and scalability required to meet varying healthcare demands. Furthermore, this approach promotes sustainability goals by reducing the environmental impact associated with traditional IT infrastructure.

Finally, telemedicine stands out as one of the most transformative innovations in healthcare, with far-reaching technological, cultural, and social implications. This novel approach enhances the accessibility and efficiency of healthcare services, addressing challenges arising from demographic and socioeconomic shifts, such as aging populations, increased mobility, and the growing need to manage large volumes of health data within constrained budgets. Telemedicine is particularly beneficial for underserved communities, overcoming geographical barriers between healthcare providers and patients, and enabling remote clinical assessments in rural areas with limited access to healthcare facilities [34, 35, 36]. While obstacles to standardization and widespread adoption exist —such as high equipment costs, connectivity issues, the need for technical training [37], and patient skepticism— major health organizations, including the World Health Organization (WHO), acknowledge telemedicine's potential to improve healthcare access, quality, and cost-effectiveness. Its implementation can lead to optimal resource allocation, potentially reducing the strain on physical healthcare facilities and allowing for precise in-person care where necessary. In addition, telemedicine facilitates continuous patient monitoring and timely interventions, which are essential for managing chronic conditions and preventing acute health crises.

2.4 Challenges and Opportunities

The potential for digital transformation in public healthcare is considerable, yet several challenges remain. Privacy concerns, data security, the digital divide, and the lack of a unified, up-to-date regulatory framework pose substantial barriers to the widespread adoption of digital technologies.

Healthcare industry is inherently complex and involves diverse stakeholders —governments, public institutions, private organizations, startups, patients, healthcare professionals, researchers, and academics— resulting in information silos. This fragmentation is further compounded by intricate regulatory mechanisms [38] that, while necessary for ensuring safety and quality control, can act as a hindrance to development. Strict procedures may slow innovation, particularly for emerging technologies whose disruptive impacts are difficult to anticipate [39, 40].

Financial constraints and functional concerns also complicate the process of DT [41]. Many healthcare organizations either have a partial understanding of the potential benefits that digital transition can offer or overestimate its costs. This can lead to reluctance to abandon long-established business practices in favor of new, digitally driven processes [42, 43]. Importantly, DT should not be viewed solely as a technological investment but rather as a holistic transformation requiring the engagement of the entire organization and its key stakeholders [43, 44], clear communication of innovation goals, and alignment toward shared objectives. Collaborative approaches —such as co-design, co-creation, and open innovation— are essential for driving this transformation [45, 46].

From the patient's perspective, resistance to digital technology adoption largely stems from privacy concerns, particularly the reluctance to share sensitive personal information [43] in light of escalating global cybersecurity threats. On one hand, the increasing frequency and sophistication of phishing attempts heighten uncertainty and suspicion among individuals; on the other hand, recent cyberattacks targeting hospital information systems and major data breaches raise serious questions about patient data security and integrity, underscoring the vulnerabilities of traditional centralized EHR systems [47, 48]. In addition, healthcare organizations frequently rely on legacy practices that may be outdated or incompatible with newer technologies, contributing to generalized distrust in how both public and private entities manage personal information [49]. Apprehensions are especially pronounced in areas like clinical genomics, where the handling of highly sensitive genetic data raises ethical issues related to informed consent and result disclosure. Although existing regulations provide some guidance, best practices in this field remain uncertain [50].

Subtle yet significant considerations also surround data sovereignty (*e.g.*, determining ownership of clinical data) and the roles of public and private actors in digital public health infrastructure. The proliferation of user-generated health data from non-conventional sources adds further pressure on policymakers to integrate these streams into a highly coordinated digital health ecosystem, reflecting users' increasing demand to leverage supplementary health information.

Patient engagement is another critical factor in achieving digital transition in healthcare. DT places a strong emphasis on improving patient experience, with a focus on making healthcare services more user-centered [51]. Since patients act as both co-producers and beneficiaries of services [52], the effectiveness of treatments depends significantly on patient involvement in managing and monitoring their own health [53].

Finally, the successful implementation of DT in healthcare also hinges on developing digital competencies among healthcare professionals. These skills are essential for fostering cross-functional collaborations, moving away from the siloed operations that have historically characterized healthcare systems. Poor coordination among digital initiatives has been a long-standing issue, hampering the development of a cohesive digital public health strategy. To overcome this, both healthcare organizations and policymakers must prioritize the provision of training and resources to support the digital transition [13].

Despite these challenges, DT presents great opportunities within healthcare. Beyond improving patient experiences, digital technologies can enhance organizational efficiency by automating administrative tasks and streamlining processes. This can result in cost savings, better resource allocation, and increased employee productivity [54]. By automating routine tasks, healthcare workers can focus on higher-value activities, while digital tools such as collaboration platforms and video conferencing endorse smart working and knowledge sharing [55, 56].

Moreover, the introduction of digital technologies into public healthcare creates opportunities for new players to enter this heavily regulated market. Greater participation from third-party actors is both welcome and beneficial, as it can accelerate innovation and promote collaboration within the healthcare ecosystem. Smart hospitals exemplify this frontier of innovation, integrating advanced technologies into healthcare infrastructure. Nevertheless, public authorities require the necessary knowledge and expertise to design, construct, and operate these facilities effectively.

In conclusion, the digital transformation of healthcare represents a complex paradigm shift, fundamentally redefining how organizations operate and deliver value to communities [57, 51]. It creates opportunities to innovate patient care, streamline operational processes, and establish new business models. Recognizing this potential, the European Union has prioritized digital health in its European Strategic Plan 2019–2024 (European Commission).

2.5 A Focus on the Italian Context

The digital transformation of healthcare in Italy is an integral part of the national strategy aimed at modernizing services, reducing inequalities, and enhancing efficiency to meet the evolving needs of the population. Central to this transformation is the National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) [58], part of the Next Generation EU (NGEU) programme, which allocates substantial funding —totaling €248 billion— across three strategic pillars aligned with European priorities: digitization and innovation, ecological transition, and social inclusion.

The plan is structured around six missions, two of which are particularly relevant to healthcare transformation. Mission 1 ("Digitization, Innovation, Competitiveness, Culture") allocates €49.2 billion to promote the country's digital transformation and support innovation in the production system. Mission 6 ("Health") designates €18.5 billion to strengthen local prevention and health services, modernize and digitize the healthcare system, and ensure equal access to care.

Within Mission 6, several key areas have been identified as focal points for digital transformation:

– *Fascicolo Sanitario Elettronico (FSE).* The Fascicolo Sanitario Elettronico, or Electronic Health Record (EHR), is a centralized, interoperable platform enabling comprehensive access to medical records for both patients and healthcare providers. It serves as a digital repository for healthcare and social-healthcare data and documents generated during clinical events, including records from both the National Health Service and private providers. The system aims to make healthcare more inclusive by ensuring consistent access to health data and essential services, regardless of geographic location. While the Italian government targets national adoption of FSE by 2026, current implementation reveals significant regional disparities, creating challenges in data interoperability and exchange.

– *Telemedicine.* Telemedicine has emerged as a critical element of Italy's healthcare strategy, particularly in response to the COVID-19 pandemic. It enables remote consultations, monitoring, and care, reducing the need for in-person visits and facilitating access to healthcare services, especially for underserved or remote populations. The PNRR sets ambitious targets for telemedicine, aiming to serve over 200,000 patients by 2025. A key aspect of this strategy is the development of a national telemedicine platform, which is essential for supporting chronic disease management and improving health outcomes.

– *Big Data and Artificial Intelligence.* The integration of big data and AI into Italy's healthcare system is advancing, particularly in the areas of predictive and preventive medicine. These technologies can process vast amounts of data, far exceeding the capacity of individual healthcare providers, allowing for early identification of health risks and personalized care plans. This shift from traditional evidence-based medicine to data-driven predictive models has the potential to revolutionize clinical decision-making and chronic disease management. However, addressing concerns surrounding data privacy, security, and the ethical use of AI remains central to public debate and policy development.

Despite the considerable progress achieved over the last five years, several challenges continue to hinder the comprehensive DT of Italian healthcare. These obstacles include digital illiteracy, a lack of standardized digital competencies among healthcare professionals, and pronounced regional inconsistencies in digital infrastructure. Furthermore, the underutilization of available digital services by both healthcare providers and citizens — evidenced by a relatively low percentage of the population regularly using the FSE and telemedicine platforms— further exacerbates inefficiencies in healthcare delivery.

The PNRR is instrumental in addressing these challenges by allocating significant financial resources towards healthcare modernization. The plan emphasizes the development of digital skills among healthcare workers and patients alike, aiming to bridge the digital divide between different regions and demographic groups. Investments in telemedicine, AI, big data, and healthcare infrastructure are deemed critical to overcoming the existing barriers to digital transformation and achieving a fully integrated digital healthcare system by 2026.

Chapter 3

Automated Image Classification of Self-Administered SARS-CoV-2 Nasal Swabs

This chapter introduces a novel application of Computer Vision (CV) techniques for the automated classification of self-administered SARS-CoV-2 diagnostic test images. The study was motivated by the urgent need for efficient and reliable assessment of test outcomes submitted to the regional Electronic Health Record system in Emilia-Romagna, Italy, during the COVID-19 pandemic. An initial corpus of 1,505 annotated images was curated and utilized to train a YOLOv5-based classifier. The classifier achieved a remarkable accuracy of 97.6% on the holdout set and 96.9% accuracy on a larger validation set of 144,844 images, demonstrating its robustness in categorizing outcomes as either positive or negative, while also detecting potential anomalies such as improper image submissions or fraudulent attempts. The integration of this model into public health workflows holds the potential to significantly alleviate the burden on human personnel during emergencies, improve the accuracy of epidemiological data monitoring, and enhance the overall efficiency of pandemic response efforts.

3.1 Background and Motivation

The global COVID-19 pandemic posed unprecedented challenges to healthcare systems worldwide, necessitating novel approaches to manage and contain the spread of the virus. In Emilia-Romagna, Italy, one such strategy involved the introduction of self-administered SARS-CoV-2 tests, a measure designed to facilitate rapid and extensive screening among the population. Through Regional Government Decree No. 33 of January 17, 2022 (Approval of the Document "Procedure for Self-Testing for the Detection of Coronavirus and Transmission of Results" [59]), the Regione Emilia-Romagna (RER) approved a pilot project aimed at simplifying the procedures for tracking and managing citizens infected with COVID-19.

As of January 19, 2022, citizens residing in Emilia-Romagna who had received their third dose of the anti-COVID vaccine are eligible to conduct self-administered tests and subsequently upload photographic evidence of the test results to their Fascicolo Sanitario Elettronico (FSE), the regional Electronic Health Record (EHR) system.

The procedure is straightforward [60]: asymptomatic individuals requiring SARS-CoV-2 screening may perform a home test using any of the numerous rapid antigen kits certified for self-administration[1] – readily available in pharmacies and supermarkets. In the event of positivity, the individual can report the outcome within their FSE. The process entails declaring the test result, along with the date and time of test execution, the commercial name of the kit utilized, its lot number, and the expiration date. Additionally, visual documentation must be provided, consisting of a photograph clearly showing both the commercial name and barcode of the diagnostic kit, together with the outcome (a single image not exceeding 4MB).

Upon submission of a positive test result, the individual receives a domiciliary fiduciary isolation order from their local health authority (AUSL) within 24 hours, formally initiating the isolation period. Once the isolation period is completed, the citizen may repeat the home test and submit any negative result via their FSE[2]. Should the test remain positive, additional

[1] The packaging must bear the CE mark followed by a 4-digit code certifying that the test is valid for home use, without assistance from a healthcare professional. A continuously updated list of recognized valid tests is available at https://salute.regione.emilia-romagna.it/tamponi-autotesting.

[2] The citizen must also provide a self-declaration attesting the absence of symptoms

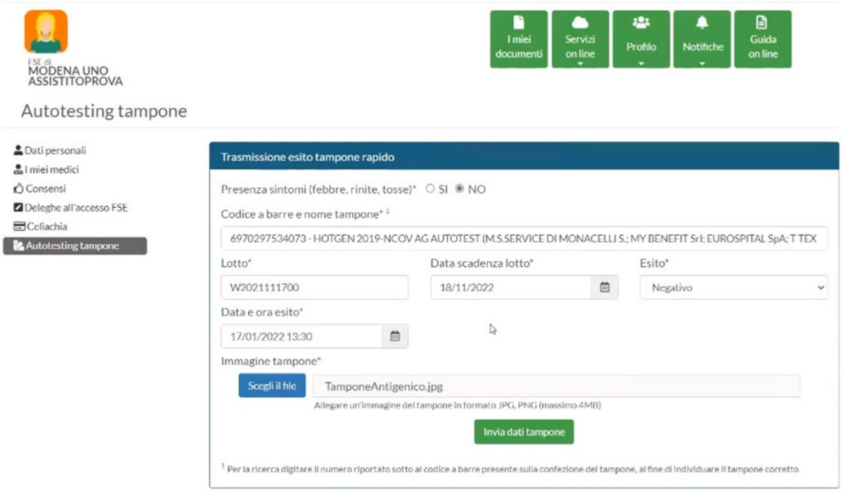


Figure 3.1: Reporting Procedure for Self-Administered Test Results to FSE.

Figure 3.2: Example of an Image Properly Uploaded to FSE.

self-administered tests can be performed in subsequent days, with results transmitted as soon as a negative outcome is achieved.

The adoption of this methodology proved to be crucial during a period when healthcare systems were under immense pressure, necessitating rapid and efficient monitoring of large numbers of individuals, particularly those presenting as asymptomatic or mildly symptomatic.

Approximately two million people were included in this pilot project.

However, while self-administration of tests provided logistical benefits, it also introduced issues related to the reliability of outcomes submitted by citizens. Considering the critical role of accurate results in informing public health decisions, the potential for incorrect or fraudulent reporting was a significant concern. Furthermore, manual verification of such a large volume of images rapidly evolved into a time-consuming task for healthcare workers, underscoring the necessity for automated solutions that could assist in the rapid and accurate assessment of test results.

In response to these challenges, this research explored the application of computer vision techniques, specifically through the use of the YOLOv5 model, to automate the classification of nasopharyngeal swab images. The primary objective was to develop a system capable of not only classifying test results as positive or negative but also detecting anomalies or inconsistencies that might indicate improper submissions or potential tampering of the images. By integrating the system into the regional EHR framework, the proposed methodology aimed to improve the efficiency of public health monitoring during pandemics, reduce the reliance on manual processes, and enhance the accuracy of data used in public health decision-making.

3.2 Related Works

In the field of Computer Vision (CV) and Machine Learning (ML), automated classification of medical images has gained significant attention, particularly during the COVID-19 pandemic. Numerous studies have investigated the application of deep learning (DL) models to assist in the diagnosis and monitoring of the virus, focusing on various imaging modalities, including radiological images (such as X-rays and CT scans), ultrasound images, and magnetic resonance imaging (MRI).

associated with the acute phase of the disease (fever, cough, rhinitis, cold) for at least three days prior to taking the test.

One of the earliest and most notable applications involved the use of convolutional neural networks (CNNs) for the detection of COVID-19 in chest X-ray images. For instance, Wang and Wong [61] introduced COVID-Net, a deep CNN specifically designed for identifying COVID-19 from chest radiographs. The model demonstrated high sensitivity and specificity, establishing the potential of DL for rapid and accurate diagnosis during the pandemic.

Similarly, Zhang *et al.* [62] applied ML techniques to analyze CT scans of infected patients, achieving remarkable accuracy in differentiating COVID-19 from other forms of pneumonia.

Sarki *et al.* [63] proposed transfer learning models, including VGG16, InceptionV3, and Xception, for both binary (Normal/COVID-19) and multi-class (Normal/COVID-19/Pneumonia) classification scenarios. Their results demonstrated the effectiveness of CNNs in COVID-19 diagnosis, achieving 100% accuracy for binary classification and 87.50% accuracy for multi-class classification.

Mehboob *et al.* [64] developed a transformer-based vision model to address the limitations of conventional CNNs that require extensive labeled datasets. Their approach incorporated a self-attention mechanism using CT slices. Compared to CNN and ensemble classifiers, the proposed method proved more effective in detecting COVID-19, with 98% accuracy in binary classification of the SARS-CoV-2 dataset.

Computer Vision has also been employed in pandemic-related ancillary tasks, such as monitoring social distancing[3] and face-mask detection[4]. In [65], Rezaei and Azarmi developed a hybrid CV and Deep Neural Network (DNN) model for automated people detection in crowded areas through CCTV security cameras. The model demonstrated high reliability in discerning human presence and evaluating adherence to prescribed interpersonal distance protocols.

In Chowdary *et al.* [66], the InceptionV3 deep learning model was used to automate the mask detection process. The Simulated Masked Face Dataset (SMFD) was used for training and testing the fine-tuned Incep-

[3]Social distancing is a recommended measure by the World Health Organization (WHO) to mitigate the transmission of COVID-19 in public spaces. Most governments and national health authorities have established a 1-meter physical distancing requirement in shopping centers, schools, and other indoor areas as a safety protocol.

[4]According to WHO guidelines for preventing the spread of coronavirus, the use of face masks is among the most efficient methods for reducing COVID-19 transmission. Wearing a face mask alone can diminish the risk of infection by over 70%.

tionV3 model. The model achieved an accuracy of 99.9% during training and 100% during testing.

The present study extends this line of research by applying CV to the classification of SARS-CoV-2 nasopharyngeal swab images. This approach builds upon the strengths of previous work, such as the use of CNNs for COVID-19 diagnosis, while adapting them to a new, highly relevant application in public health. The focus on self-administered tests and the integration of anomaly detection represents a novel contribution to the field, addressing both technical and ethical challenges posed by the reliance on self-reported data during a public health emergency. To the best of our knowledge, this is the first experiment of its kind.

3.3 Materials and Methods

The regulatory requirement for the Regione Emilia-Romagna to conduct random audits on a minimum of 10% of antigen swab images uploaded to the FSE, coupled with the high participation rate of citizens in the self-testing campaign [67], prompted a collaborative effort between RER and the University of Modena and Reggio Emilia to refine the regional pilot project. The proposal was structured across multiple levels, specifically:

1. Application of DL techniques to the images uploaded to the FSE for automatic detection of and classification of swab results (positive/negative).

2. Identification of potential inconsistencies between a citizen's declaration and the image content (*e.g.*, mismatched outcomes or diagnostic kit names) and detection of potential image manipulations, particularly related to the rapid antigen test result (fraud attempts).

3. Implementation of scalable RESTful APIs[5] for image validation (*e.g.*, ensuring the photograph contains at least one swab), correction (*e.g.*, conversion to a valid color space, perspective rectification, etc.), and outcome classification.

[5]The APIs (Application Programming Interface) can be directly invoked by the regional FSE system to perform preliminary online validation during the citizen's form submission process. Further details are available in Appendix B.

Steps 2 and 3 were contingent upon the evaluation of the results achieved during the first phase of the experiment, which is fully detailed in the following sections.

3.3.1 Dataset

The initial dataset, provided by RER, comprised 3,272 images of varying dimensions (all in JPEG format, not exceeding 750×750 pixels) depicting SARS-CoV-2 nasopharyngeal swabs uploaded by citizens to their FSE. Of these, 239 images were excluded due to duplication, while 22 exhibited anomalies in their color space that required adjustment.

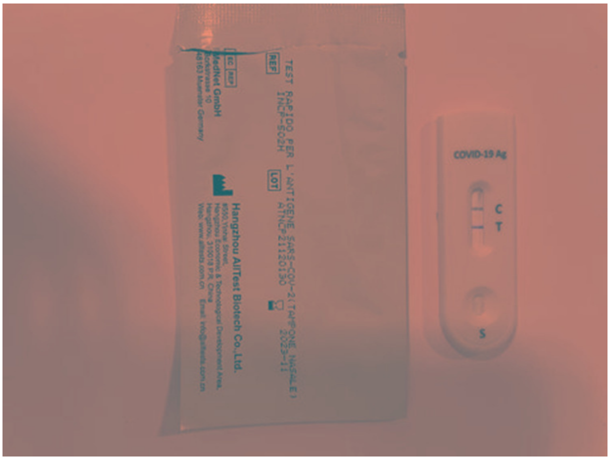


Figure 3.3: Example of an Image with Incorrect Color Space.

The first 1,505 images were annotated and used for model training, while the remaining 1,528 were reserved for the testing phase to evaluate performance. Additionally, RER provided a Comma-Separated Values (CSV) file consisting of two columns: a numerical identifier for each uploaded image (corresponding to the JPG file name) and the result declared by the citizen on the FSE platform (0 = Negative / 1 = Positive). No Personally Identifiable Information (PII) was supplied, ensuring compliance with current privacy and data protection regulations (Regulation EU 2016/679, European Parliament, 27 April 2016).

3.3.2 Detection of Rapid Antigen Tests

At the time of this investigation, no public datasets related to SARS-CoV-2 rapid antigen tests were available. Consequently, it was necessary to first annotate a substantial number of images before proceeding with neural network training.

OpenCV

An initial attempt was made using the OpenCV library (v3.4).

OpenCV [68], an acronym for Open-Source Computer Vision Library, is a widely used, open-source, cross-platform library designed for real-time CV and ML tasks. It offers a comprehensive suite of optimized algorithms for image and video analysis, including object detection, facial recognition, motion tracking, and image segmentation. Due to its efficiency in processing visual data, OpenCV has been extensively adopted in both academic research and various industrial applications, encompassing robotics, autonomous vehicles, and medical imaging.

The experimental setup involved the preparation of two distinct directories [69]: a "positive" folder containing 1,505 images of the "target" object to be recognized (positive samples), and a "negative" folder comprising generic images without swabs (negative samples)[6]. Files in the "positive" folder were annotated using the `opencv_annotation` tool, which generated a text file (*pos.txt*) with one row per annotated image.

```
positive/4494242_1.jpg 1 458 217 82 202
positive/4494245_1.jpg 1 1 161 748 138
positive/4494250_1.jpg 1 122 239 487 125
positive/4494251_1.jpg 1 420 220 66 331
positive/4494252_1.jpg 1 132 2 284 746
positive/4494255_1.jpg 1 47 399 677 130
positive/4494256_1.jpg 1 53 237 688 121
positive/4494260_1.jpg 1 132 15 488 98
positive/4494261_1.jpg 1 129 49 604 191
positive/4494262_1.jpg 1 97 22 436 91
```

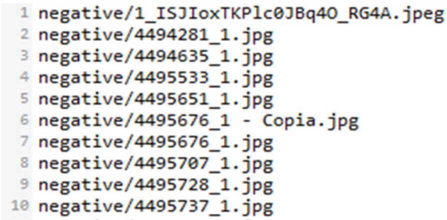
Figure 3.4: Excerpt from *pos.txt* File.

[6]The use of the terms "positive" and "negative", as applied to sample images, should not lead to misinterpretation: in the context of OpenCV, these terms refer to samples containing / not containing the object of interest for classification, respectively. Therefore, they are unrelated to the actual outcomes of antigen tests.

Each row adhered to the specific structure below, including image file information, bounding box coordinates, and dimensions:

```
data[0] = image file
data[1] = number of bounding boxes
data[2] = x1 (upper left corner x coord. of bounding box #1)
data[3] = y1 (upper left corner y coord. of bounding box #1)
data[4] = bbw1 (width of bounding box #1)
data[5] = bbh1 (height of bounding box #1)
data[6] = x2 (upper left corner x coord. of bounding box #2)
data[7] = y2 (upper left corner y coord. of bounding box #2)
data[8] = bbw2 (width of bounding box #2)
data[9] = bbh2 (height of bounding box #2)
...
```

For the images in the "negative" folder, a simple text file listing the paths (*neg.txt*) was created.



```
negative/1_ISJIoxTKPlc0JBq4O_RG4A.jpeg
negative/4494281_1.jpg
negative/4494635_1.jpg
negative/4495533_1.jpg
negative/4495651_1.jpg
negative/4495676_1 - Copia.jpg
negative/4495676_1.jpg
negative/4495707_1.jpg
negative/4495728_1.jpg
negative/4495737_1.jpg
```

Figure 3.5: Excerpt from *neg.txt* File.

Using *pos.txt* file as input, the `opencv_createsamples` tool was utilized to produce the *pos.vec* binary file, which contained a collection of generated positive samples in a format optimized for training object-detection classifiers:

```
opencv_traincascade -data cascade -vec pos.vec -bg neg.txt
    -w 48 -h 48 -numPos 100 -numNeg 250 -numStages 10
    -minHitRate 0.999 -maxFalseAlarmRate 0.5
    -precalcValBufSize 1024 -precalcIdxBufSize 1024
```

Subsequently, the *pos.vec* and *neg.txt* files were provided as input to the `opencv_traincascade` command, which executed model training [69]. This operation generated an XML file containing the specifications of the trained model.

```xml
<?xml version="1.0"?>
<opencv_storage>
<cascade>
  <stageType>BOOST</stageType>
  <featureType>HAAR</featureType>
  <height>48</height>
  <width>48</width>
  <stageParams>
    <boostType>GAB</boostType>
    <minHitRate>9.9900001287460327e-01</minHitRate>
    <maxFalseAlarm>5.0000000000000000e-01</maxFalseAlarm>
    <weightTrimRate>9.4999999999999996e-01</weightTrimRate>
    <maxDepth>1</maxDepth>
    <maxWeakCount>100</maxWeakCount></stageParams>
  <featureParams>
    <maxCatCount>0</maxCatCount>
    <featSize>1</featSize>
    <mode>BASIC</mode></featureParams>
  <stageNum>10</stageNum>
  <stages> (...)
          2.6334139704704285e-01 -7.7368384599685669e-01</leafValues></_></weakClassifiers></_></stages>
  <features> (...)
      <tilted>0</tilted></_></features></cascade>
</opencv_storage>
```

Figure 3.6: Excerpt from *cascade.xml* File Generated by the `opencv_traincascade` Command.

The resultant XML file was finally employed to evaluate the model's performance on the test set using a command-line interface (CLI) script written in Python.

```
python manage.py swabs_detection \training_set\positive
    --cascade \training_set\cascade\cascade.xml
```

However, the results were unsatisfactory (Figure 3.7), suggesting the need to explore alternative methodological approaches.

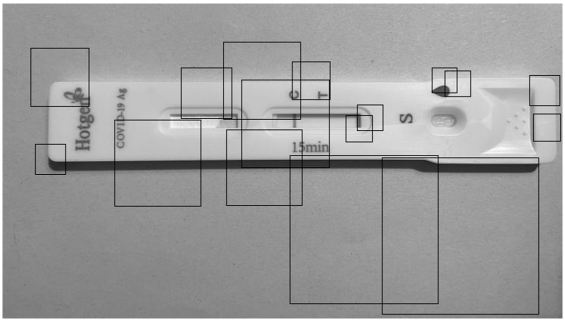


Figure 3.7: Recognition of a SARS-CoV-2 Rapid Antigen Test Using OpenCV.

YOLOv5

YOLOv5 [70] represents a state-of-the-art, real-time object detection system built upon the *You Only Look Once* (YOLO) architecture. It leverages a single CNN to simultaneously predict bounding boxes and class probabilities for multiple objects within an image.

YOLOv5 is characterized by higher accuracy, reduced computational requirements, and faster processing speed compared to its predecessors. The model utilizes anchor-based detection, feature pyramid networks for multi-scale predictions, and incorporates various optimizations —such as mosaic augmentation, adaptive image padding, adaptive anchor learning, and cross-GPU batch normalization— to achieve superior performance across a wide range of datasets and hardware configurations. The YOLO architecture has been extensively used in various medical image analysis tasks [71, 72, 73, 74, 75], supporting its adoption in the present study.

Alas, the annotation formats employed by YOLOv5 and OpenCV are incompatible. To avoid re-annotating the entire dataset, the DarkLabel tool [76] was initially used to generate example annotation files in YOLOv5 format, followed by the implementation of a Python script (Listing 3.1) to convert the OpenCV format (a single text file for all images) into the YOLOv5 format (one text file per image).

```python
img = cv2.imread(data[0])
ih, iw, _ = img.shape
bb_num_count = int(data[1])  # number of bounding boxes

out_line = ""
for i in range(bb_num_count):
  x = int(data[2 + (4 * i)])
  y = int(data[3 + (4 * i)])
  bbw = int(data[4 + (4 * i)])
  bbh = int(data[5 + (4 * i)])

  ncx = round((x + bbw / 2) / iw, 6)
  ncy = round((y + bbh / 2) / ih, 6)
  bbwp = round(bbw / iw, 6)
  bbhp = round(bbh / ih, 6)

  out_line += f"0 {ncx} {ncy} {bbwp} {bbhp}\n"

file_without_ext = pathlib.Path(data[0]).stem
out_file = os.path.join(destination_path, file_without_ext + ".txt")
with open(out_file, mode="w") as annotations_file_out:
  annotations_file_out.write(out_line)
```

Listing 3.1: OpenCV to YOLOv5 Annotation Format Conversion.

After cloning the original YOLOv5 project and appropriately configuring the paths for the training, validation, and test datasets, the primary objective was to accurately identify the swab within the image. This task was formulated as a single-class object detection problem, where the only target class was designated as "swab." To achieve this goal, the model was initialized using the weights of the "small" YOLOv5 model, which served as the baseline. A total of eight experiments were conducted, systematically varying key hyperparameters such as batch size, number of training epochs, and the resolution of input images. These experiments aimed to assess the impact of different configurations on model performance and to determine the optimal setup for accurate swab detection.

Table 3.1 presents a summary of the configurations used in two specific models: the baseline model, which exhibited the lowest performance among all tested configurations, and the best-performing model, which demonstrated the highest accuracy.

Table 3.1: Configurations for YOLOv5 Single-Class (SC) Experiments.

Experiment	**Img Size**	**Batch Size**	**Epochs**	**Weight**	**Err. (%)**
Exp1	640	8	3	Small	39.4
Exp7	**640**	**16**	**100**	**Small**	**0.20**

As anticipated, increasing both the batch size and the number of epochs led to superior results. This outcome aligns with established findings in deep learning research, where larger batch sizes contribute to more stable gradient updates, and longer training durations allow models to learn more complex representations. In Figure 3.8, the results of Exp1 SC (on the left) are compared with those of Exp7 SC (on the right). The detected target is enclosed within a red rectangle, labeled with the name of the identified class ("swab") and the model's confidence level.

A crucial aspect of object detection models is selecting an appropriate confidence threshold, which determines the probability level at which a detection is considered valid. Setting the threshold too low increases the risk of false positives, whereas an excessively high threshold may lead to missed detections. In Exp7 SC, a confidence level of 0.4 resulted in seven false positives in the test set of 1,528 images, while increasing the confidence threshold to 0.7 reduced the number of false positives to only one (Figure 3.9).

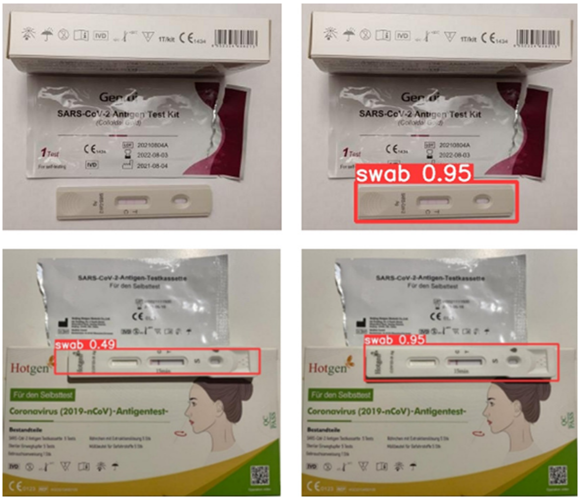


Figure 3.8: Detection Examples with Varying Epochs and Batch Sizes.

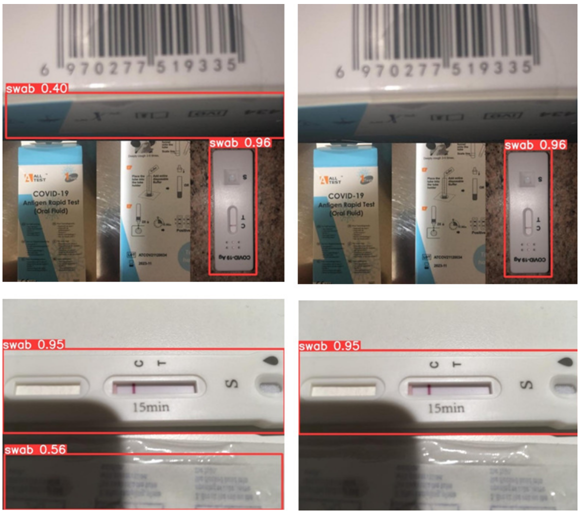


Figure 3.9: Comparative Examples of Detection at Distinct Confidence Levels.

3.3.3 Classification of Rapid Antigen Test Outcomes

Following the development of a model capable of identifying rapid antigen test kits within an image, the next challenge involved classifying the test outcome. Although interpretation of results may vary depending on the specific diagnostic kit utilized, the general principle is illustrated in Figure 3.10. For the purposes of this experiment, classification was restricted to "positive" (SARS-CoV-2 target detected) and "negative" (target not detected) outcomes. Including invalid results would have required an adequate number of images depicting null tests for network training, which is a relatively rare occurrence.

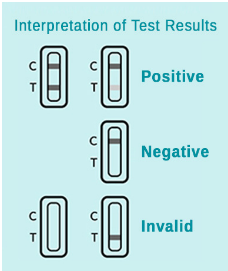


Figure 3.10: Interpretation of SARS-CoV-2 Rapid Antigen Test Results.

Template Matching with OpenCV

Initially, a template matching technique was applied to the photographs uploaded by users to the FSE platform. In template matching, the input image is systematically scanned to locate regions with a high degree of similarity to the portion used as a template.

Several images of antigen tests were manually selected, and the portion displaying the control window was isolated to serve as the template.

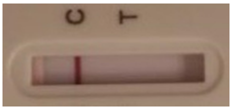


Figure 3.11: Control Window of a Rapid Antigen Kit Template for Negative Outcome.

OpenCV provides various methods for implementing template matching. Due to the confirmed presence of multiple test kits within certain images, the TM_COEFF_NORMED (Template Matching using Normalized Correlation Coefficient) method was chosen, with a similarity threshold set to 0.8. This method normalizes the correlation coefficient, making it more resilient to variations in lighting, scale, and contrast. Such normalization is particularly advantageous in images containing multiple objects, as it evaluates the similarity between the template and different regions of the image on a consistent scale, irrespective of absolute intensity differences. By assigning a normalized score between -1 and 1, TM_COEFF_NORMED can accurately identify regions that closely resemble the template, even in the presence of multiple potential matches within the same image. A similarity threshold of 0.8 ensures that only the most relevant matches are considered, reducing the occurrence of false positives.

Despite these theoretical benefits, empirical testing on a sample of images from the initial dataset revealed that template matching was not a viable approach for the context of this study, particularly considering the variety of diagnostic kits available to end-users[7] and the non-negligible challenges in interpreting test results[8].

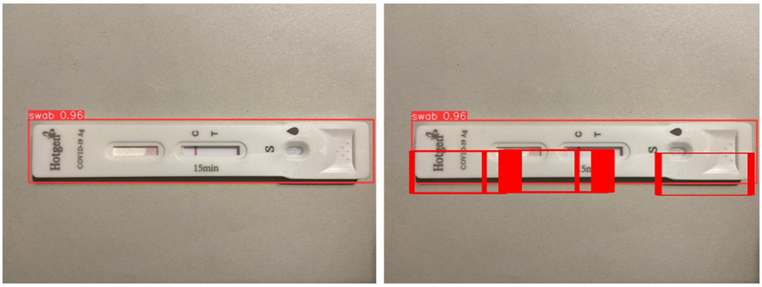


Figure 3.12: Template Matching on a Test Image.

As a result, it was decided to investigate the YOLOv5 neural network for both detection and multi-class classification.

[7]The arrangement of the C (Control) and T (Target) symbols varies across different test kits. In some cases, the symbols are reversed.

[8]The C/T marker lines are not always clearly defined.

Classification with YOLOv5

After modifying the YOLOv5 configuration file to define two classes ("swab_neg" for negative test results and "swab_pos" for positive outcomes), a Python script was implemented to convert and repurpose the previously used annotation files, rendering them suitable for multi-class classification. This process leveraged the CSV file provided by RER, which accompanied the initial dataset. The test result declared by the citizen at the time of submission to the FSE platform (0 = Negative / 1 = Positive) was used as a numerical indicator corresponding to the relevant class (0 = "swab_neg" and 1 = "swab_pos").

Additionally, a manual verification of all 1,505 images in the training set was conducted to ensure consistency between the class declared in the CSV file and the actual content of the corresponding image. Five evident discrepancies were identified, where outcomes had been declared as negative but were, upon visual inspection, positive. In these cases, the images were reclassified accordingly prior to initiating the training process.

The final training set included 465 images showing negative test results (approximately 31% of the total) and 1,040 showing positive results.

Extending the methodological framework employed in prior single-class detection assessment, we conducted a series of eight experiments to investigate the effects of varying batch size, the number of training epochs, and input sample dimensions. Table 3.2 presents a summary of the hyperparameter configurations employed for training the multi-class (MC) models in each evaluation.

Table 3.2: Configurations for YOLOv5 Multi-Class (MC) Experiments.

Experiment	**Img Size**	**Batch Size**	**Epochs**	**Check-points**
Exp1	640	16	100	Small
Exp2	640	16	300	Small
Exp3	640	32	300	Small
Exp4	750	32	300	Small
Exp5	750	64	300	Small
Exp6	750	64	300	Medium
Exp7	750	16	300	Medium
Exp8	750	32	300	Medium

3.4 Results

The test set comprised 1,528 images categorized as follows:

- 1,036 images contained one or more positive test results, totaling 1,045 positive swabs.
- 443 images contained one or more negative test results, amounting to 445 negative swabs.
- 49 images did not contain any test results.

Table 3.3 summarizes the overall performance metrics for each experiment, measured during the testing phase with a confidence threshold of 0.7. For positive swabs, the reported metrics include:

- *True Positive* (TP): the number of correctly classified positive swabs.
- *False Positive* (FP): the number of swabs incorrectly classified as positive (either actually negative or not present).
- *False Negative* (FN): the number of swabs erroneously classified as negative or not present (actually positive).

Similarly, for negative swabs:

- *True Positive*: the number of correctly classified negative swabs.
- *False Positive*: the number of swabs incorrectly classified as negative (either actually positive or not present).
- *False Negative*: the number of swabs erroneously classified as positive or not present (actually negative).

These metrics were subsequently aggregated and evaluated globally. Additionally, *True Negative* (TN) represents the number of images not containing any swabs and correctly classified as such. The model's performance was assessed using well-established metrics in image classification, including:

- *Precision* (PRC). Measures the proportion of correctly identified positive test results out of all predicted positives:

$$PRC = \frac{TP}{TP + FP}$$

In this context, it indicates how reliable the model is when it classifies a swab as positive. High precision is crucial because it ensures that when the model predicts a positive test result, the likelihood of being correct is high, thus reducing the risk of false positives, which could lead to unnecessary quarantine or anxiety.

- *Recall* (REC). Measures the proportion of true positives that were correctly identified by the model:

$$REC = \frac{TP}{TP + FN}$$

In this context, recall is critical because missing a positive test (false negative) could allow an infected person to go undetected, posing a public health risk.

- *Accuracy* (ACC). Measures the overall correctness of the model across all predictions:

$$ACC = \frac{TP + TN}{\text{Test set size}}$$

It provides a broad measure of how well the model performs across both positive and negative test results. High accuracy is important because it demonstrates the model's overall reliability, ensuring that the majority of images, whether positive or negative, are correctly classified.

Table 3.3: Precision, Recall, and Accuracy Metrics.

Experiment	**TP**	**FP**	**FN**	**TN**	**PRC ↑**	**REC ↑**	**ACC ↑**
Exp1	**1,443**	**28**	**36**	**49**	**0.981**	**0.976**	**0.976**
Exp2	1,418	53	61	49	0.964	0.959	0.960
Exp3	1,426	46	53	49	0.969	0.964	0.965
Exp4	1,442	30	37	49	0.980	0.975	0.976
Exp5	1,436	33	43	49	0.978	0.971	0.972
Exp6	1,433	39	46	49	0.974	0.969	0.970
Exp7	1,430	45	49	49	0.969	0.967	0.968
Exp8	1,439	35	40	49	0.976	0.973	0.974

The experiment yielding the best results, Exp1, reported 36 FN, of which 8 were due to failure in detecting a swab in the image, and 28 FP (negative swabs classified as positive and vice versa), resulting in a total error rate of 2.36% for the test set. The Exp1 model exhibited strong overall performance, with a high recall of 0.976, indicating the model's effectiveness in identifying true positives (particularly positive swabs), which is critical for reliable health monitoring during a pandemic.

Figures 3.13 through 3.16 illustrate the results. The detected target is outlined by a rectangle, with its color depending on the class (red for negative results, pink for positive outcomes), and labeled with the identified class name ("swab_neg" for negative, "swab_pos" for positive) along with the model's confidence level.

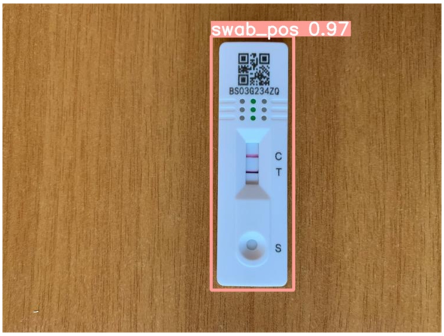


Figure 3.13: Correctly Classified Positive Antigen Test.

Figure 3.14: Correctly Classified Negative Antigen Test.

Notably, the model successfully classified images depicting swabs with slightly positive results (positive marker barely visible to the naked eye, Figure 3.15), and poor-quality images (Figure 3.16).

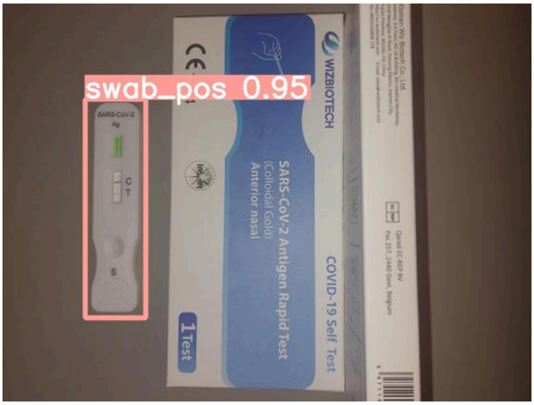


Figure 3.15: Correctly Classified Slightly Positive Antigen Test.

Figure 3.16: Correctly Classified Negative Test (Poor-Resolution Image).

The results obtained from the Exp1 model, however, did not account for the inherent imbalance in the initial dataset. Of the 1,505 images used for training, 1,040 were positive swab samples, while only 465 were negative swab specimens. Due to the unavailability of new samples, an additional 405 negative swab images from the test set were annotated, resulting in a training set of 1,910 images and a test set of 1,123 images.

The retrained model, referred to as Exp1_1910, yielded an accuracy of 98.8%. Nonetheless, when validated on a more extensive and heterogeneous dataset of 144,844 images, provided by RER as of June 12, 2022, the model's accuracy decreased to 96% (Table 3.4). This decline in performance was primarily attributed to the overrepresentation of diagnostic kits from a limited number of manufacturers (*e.g.*, Hotgen) in the original dataset, which limited the model's ability to generalize across a broader range of test kits.

Table 3.4: Precision, Recall, and Accuracy Metrics for Exp1_1910 on Extended Dataset.

Experiment	**TP**	**FP**	**FN**	**TN**	**PRC ↑**	**REC ↑**	**ACC ↑**
Exp1_1910*	137,370	4,979	5,803	1,671	0.965	0.960	0.960

* Extended test-set (144,844 images).

To overcome this limitation, an additional 170 images representing diverse diagnostic kits were annotated. These instances were selected from novel samples uploaded by users between June 15 and 19, 2022, focusing specifically on those that the Exp1_1910 model had incorrectly classified. The updated training set of 2,080 images, evenly distributed between 1,040 positive and 1,040 negative swabs, was then used to retrain a new model, namely Exp1_2080.

```
# TRAINING exp1_2080
python3 -u yolov5/train.py --img 640 --batch 16 --epochs 100
    --data yolov5/data/swabs_v2.yaml --weights yolov5/yolov5s.pt
    --project /checkpoints/SwabCV/ --name exp1_2080
```

Finally, the Exp1_2080 model was evaluated on the same set of 144,844 images employed to validate the Exp1_1910 experiment. This approach enabled a direct comparison between the performance of the two models under consistent hyperparameter configurations, while accounting for differences in their respective training sets.

```
# TEST exp1_2080
python3 manage.py swabs_detection '/datasets/SwabCV/testset_144844'
    --weights '/checkpoints/SwabCV/exp1_2080/best.pt' --confidence 0.7
    --metadata_in '/datasets/SwabCV/samples120922.csv'
    --metadata_out '/datasets/SwabCV/samples120922-OUT.csv'
    --use-db True
```

The Exp1_2080 model achieved an accuracy of 96.9% (Table 3.5), confirming its robustness, generalizability, and stability over time, even with the increased size and diversity of the new dataset. Moreover, it successfully classified images depicting less common diagnostic kit shapes (Figure 3.18).

Table 3.5: Precision, Recall, and Accuracy Metrics for Exp1_2080 on Extended Dataset.

Experiment	TP	FP	FN	TN	PRC ↑	REC ↑	ACC ↑
Exp1_2080*	138,608	3,984	4,517	1,719	0.972	0.968	0.969

* Extended test-set (144,844 images).

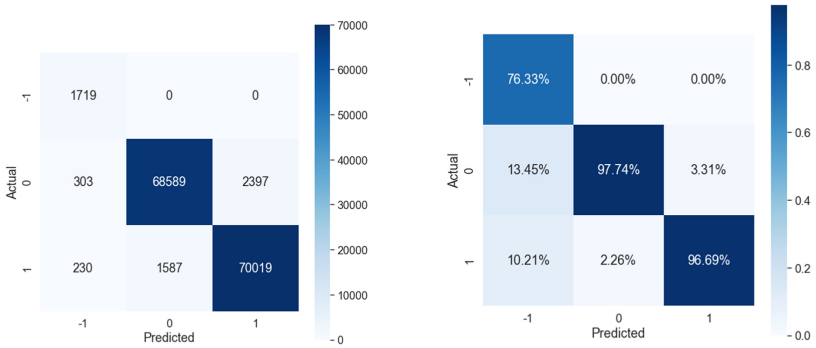


Figure 3.17: Confusion Matrix for Exp1_2080.

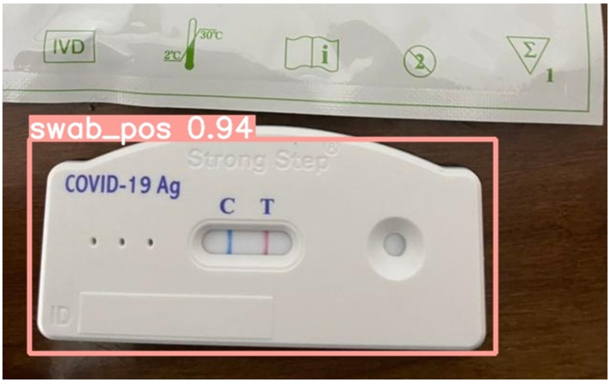


Figure 3.18: Correctly Classified Positive Antigen Test (Generalization).

The computation of metrics requires several methodological clarifications.

First, the model's ability to identify multiple target objects within individual images enabled the recording of both classification outcomes and target enumeration, distinguishing between cases of zero, single, and multiple occurrences. In instances where no swabs were recognized, the prediction result was assigned a value of -1.

Second, a manual inspection of the augmented validation set revealed the following:

– Among the 2,252 images wherein the model detected the absence of a swab, 1,719 were indeed devoid of swabs ("True Misses"), while 533 contained at least one swab ("False Misses"). For the 1,719 True Misses, the corresponding ground-truth labels were adjusted to -1 to prevent them from being mistakenly counted as model errors. This recalibration was necessary due to the mismatch between user-declared outcomes (restricted to binary values, as required by the FSE submission form) and the model's predicted outcomes, which is always -1 for True Misses. Interestingly, most False Misses occurred when swabs were positioned atop unfolded instruction leaflets.

– Of the 454 images in which the model determined the presence of multiple swabs, 299 actually contained more than one swab ("True

Multiple Detection"), while 155 contained at most one swab ("False Multiple Detection").

– In addition to the 4,517 images where the model's prediction diverged from the user-declared result, the True Misses and Multiple Detections instances must also be comprised in the total number of images requiring manual validation. True Misses need to be included because the FSE system inherently lacks a "-1" label, causing these instances to appear inconsistent with the user's declaration. Similarly, Multiple Detections must be considered, as users are expected to submit only a single result, necessitating manual validation by an operator for images containing multiple swabs. It is worth noting that False Misses are already included in the 4,517 misclassification count, as their declared outcome (0 or 1) differs from the predicted one (-1).

  Thus, the total number of images accounted for manual review is calculated as: 4,517 + 1,719 + 369[9] = 6,605 samples.

Compared to the Exp1_1910 model, which required 7,953 images to be verified by back-office operators, the Exp1_2080 model reduced the number of images needing review to 6,605, representing a decrease of 1,348 samples.

3.5 Discussion

This experimental study, the first of its kind in Italy, was made possible by several factors:

– The worsening of the SARS-CoV-2 pandemic in early 2022, particularly the rapid and significant rise in COVID-19 cases following the spread of the Omicron variant.

– A growing difficulty for Public Health Departments in promptly executing molecular confirmation tests for positive cases detected through rapid antigen tests performed in pharmacies.

[9]Calculated as the difference between the total instances of multiple swab detections within single images (n = 454) and the subset of these instances already incorporated within the 4,517 inferences where the predicted outcome diverged from the ground-truth label (n = 85), as identified through manual assessment. This computation ensures a non-duplicative count of discrepancies attributed to multiple swab detections.

– The Circular No. 705-08/01/2021 issued by the Italian Ministry of Health, which states: *"In a high-prevalence context, rapid antigen tests will have a high positive predictive value (PPV). Therefore, a positive rapid antigen test result is likely indicative of a true infection, not requiring confirmation with RT-PCR testing"*. In line with this directive, the Regione Emilia-Romagna sought to extend this approach to residents who had received their third vaccine dose within the preceding four months.

The developed model, based on the YOLOv5 architecture, exhibited high performance and scalability when evaluated on a dataset of 144,844 images of self-administered swabs, rendering it a valuable tool for verifying COVID-19 home test results. Further analysis of the validation dataset revealed 6,605 instances requiring human intervention.

By regulation Regione Emilia-Romagna's administrative personnel are required to manually validate at least 10% of the collected images. With 144,844 files uploaded over a six-month period, this would translate to a minimum of 14,484 *randomly* selected images requiring review. Assuming an optimistic validation rate of one image every 10 seconds[10], this process would take approximately 40 hours (or one full workweek). In contrast, employing the Exp1_2080 model would allow operators to limit their review to the 6,605 images with discrepancies between citizen declarations and the model predictions, amounting to 4.6% of the total dataset—a significant reduction compared to the mandated 10% manual review. Verification of these images would take approximately 18.3 hours, resulting in a 55% reduction in review time, which can be particularly valuable in the context of health crises characterized by constrained human resources[11]. As more data is collected, the manual review burden would typically increase proportionally. Conversely, by leveraging the proposed model, the need for human validation remains manageable, as human effort is directed only to-

[10]This assessment presupposes optimal working conditions characterized by uninterrupted workflow and streamlined error reporting functionality, where flagging problematic images requires only a single click rather than separate manual documentation. The estimation also assumes minimal occurrence of ambiguous cases requiring extended review. Given the repetitive and mechanical nature of the process, achieving a higher processing rate would be unlikely.

[11]Based on a more realistic validation rate of one image per minute, manual verification without automated support would require approximately 242 hours (or 1.51 working months). Implementation of the proposed model would reduce this duration to approximately 110 hours (equivalent to 0.69 months of standard working time).

ward ambiguous or edge cases, including True Misses (*i.e.*, images without swabs) and Multiple Detections (*i.e.*, images containing multiple swabs), ensuring scalability and allowing personnel to focus on more critical and high-level tasks.

Finally, as the model continues to learn from new data, its accuracy is expected to improve. Even in scenarios with a modest reduction in model performance (*e.g.*, precision of 0.94 and recall of 0.93), substantial time savings would still be achieved. As a supplementary test, a sample of 400 images was arbitrarily selected for review by a back-office operator. While the model flagged real anomalies in 192 instances (mostly involving very slightly positive swabs, likely indicating early stages of infection), the human operator identified 10.

3.6 Final Remarks

This study demonstrates the potential of AI-driven solutions in addressing critical challenges within public health, particularly in emergency scenarios where rapid and accurate data processing is crucial.

A YOLOv5-based model was developed and trained to detect and classify images of swabs submitted by citizens to their EHR (FSE) as evidence of self-administered COVID-19 nasopharyngeal tests.

Analysis of a six-month extended validation set, comprising 144,844 images, revealed 6,605 misclassifications (including 1,719 instances containing no swab and 369 containing multiple swabs in a single image) – 4.6% of the entire dataset – requiring validation. This proportion falls below the mandated minimum of 10% and primarily includes cases necessitating review, thus representing a significant improvement over random selection and verification, which could potentially overlook critical cases.

By automating a portion of the validation process, RER can ensure compliance with national regulations while expediting the processing of test results. The system also facilitates faster identification of positive cases —a key factor for timely virus containment interventions— while its proficiency in identifying potentially fraudulent or erroneous submissions provides an additional layer of security, which is essential in scenarios where public health resolutions partially rely on self-reported information from citizens.

Despite promising results, the study encountered several challenges. A primary issue was the variability in the quality of images submitted by residents. Factors such as lighting conditions, focus, and the presence of distracting elements in the background significantly affected performance. While the YOLOv5 model demonstrated robustness, misclassifications still occurred, particularly with severely compromised images.

Another limitation was the model's performance in detecting images that did not contain test kits. Although it correctly classified 76.3% of such images (i.e., recall), this rate is lower than its performance on positive and negative classifications. Improvements in preprocessing techniques, combined with the development of more sophisticated anomaly detection algorithms, could enhance the model's efficacy in handling these cases.

The integration of AI in healthcare, particularly in automated diagnostic systems, raises major ethical considerations. The potential for false positives or negatives, especially in a public health context, can lead to serious consequences, such as unnecessary anxiety for individuals or delayed interventions for positive cases. While automation can enhance efficiency, it is crucial to ensure that AI systems serve as tools to support, rather than replace, human decision-making. Future work should focus on developing guidelines and frameworks for the responsible and ethical use of AI in public health.

Several functionalities remain to be implemented, including:

– Extracting specific information from uploaded images (*e.g.*, the type of self-testing kit) and verifying consistency (*e.g.*, ensuring the use of approved kits only).

– Detecting improper submissions or fraud attempts (*e.g.*, images already submitted by other users[12]).

– Recognizing potential image manipulations.

– Identifying faces or other PII (*e.g.*, fiscal codes or driver's licenses) within images.

Incorporating these features into the submission process could facilitate proper image uploading to the FSE and alert users to potential anomalies, minimizing the necessity for subsequent manual verification.

[12] A rudimentary duplicate detection mechanism has been implemented within the system, using file hashing techniques for image comparison.

Among future enhancements, it is also worth considering the development of a task capable of automatically and regularly processing new images uploaded to the FSE system (*e.g.*, once or multiple times per day), providing back-office operators with a list of detected discrepancies or anomalies for review and correction.

As the field of AI in healthcare continues to evolve, computer vision is likely to play a pivotal role in developing advanced and reliable diagnostic tools, as well as in supporting the responsiveness and sustainability of healthcare systems across both clinical and administrative domains. This study suggests potential for extending the application of AI-driven solutions beyond COVID-19. For instance, similar models could be applied to other infectious diseases that rely on rapid diagnostic tests or used for screening purposes in routine medical examinations. The scalability of the YOLOv5 model, along with its ability to process large volumes of data in real-time, makes it a viable option for a wide range of healthcare applications.

Chapter 4

Optimizing Resource Allocation in Hospitals: A Machine Learning Approach for Length-of-Stay Prediction

Effective hospital resource management hinges on established metrics such as Length of Stay (LOS) and Prolonged Length of Stay (pLOS). Reducing pLOS is associated with better patient outcomes and optimized resource utilization (*e.g.*, bed allocation). This chapter investigates several Machine Learning (ML) models for both LOS and pLOS prediction. The integration of ML, particularly ensemble models, has the potential to significantly improve LOS prediction and identify patients at high risk of pLOS. Such insights empower clinicians and bed managers to enhance patient flow, contributing to both care quality and operational efficiency in healthcare delivery systems.

4.1 Introduction

The Italian public healthcare system faces a complex challenge in managing bed availability. The past few decades have witnessed a 30% reduction in hospital bed capacity [77], along with a rise in bed occupancy rates, leading to congestion and extended patient stays. Italy's current provision of 11.6 beds per 100,000 inhabitants falls below the OECD (Organization for Economic Cooperation and Development) average of 16.9 [78]. This situation is further exacerbated by a significant increase in national health expenditure, which rose from approximately 80 billion euros in 2002 to 129 billion euros in 2022 [79, 80], with 20% allocated to inpatient care.

Managing patient flow has become increasingly problematic due to factors such as rising patient volumes, an aging population with higher comorbidity rates, delayed discharges to non-acute care settings, and evolving working practices (as emphasized by the COVID-19 pandemic). Current bed modeling techniques, often relying on midnight census data, lack the granularity needed for optimal space management. A more comprehensive understanding of patient flow dynamics and peak occupancy patterns is essential. In this context, the emergence of Artificial Intelligence (AI) and Machine Learning (ML) offers promising tools to assist bed managers in their daily operations.

4.2 Background

Length of Stay (LOS), defined as the duration between hospital admission and discharge (*i.e.*, total bed-days occupied by a patient), plays a fundamental role in assessing healthcare service quality. Previous research has demonstrated correlations between LOS and disease severity, readmission rates, and mortality [81, 82]. The reduction of LOS in public healthcare systems benefits both patients and hospitals: early discharge and faster turnover improve inpatient outcomes by preventing complications and reducing the risk of adverse events (such as falls, thrombosis, drug reactions, and hospital-acquired infections) [83, 84], while promoting patient autonomy [85]. On the other hand, hospitals gain advantages from optimized treatment strategies, improved resource utilization (*e.g.*, bed allocation), and better control over waiting lists [86]. Furthermore, LOS holds the advantage of uniform measurability, making it comparable even across different healthcare facilities on a global scale [87].

Conversely, Prolonged Length of Stay (pLOS) is associated with functional limitations, cognitive impairment, and a higher burden of comorbidities among patients [88]. Moreover, pLOS often results in cancellations of elective surgeries, increased resource utilization (including raising medical costs, especially in Intensive Care Units, ICUs), and potential delays in admitting critically ill patients. Notably, a small percentage of patients with pLOS can consume up to 50% of available resources [89, 90].

This holds particular relevance in Italy, where demographic shifts like increasing life expectancy (80.6 years for men and 84.8 years for women in 2022 [91]) contribute to an escalation in chronic and degenerative diseases. Timely identification of inpatients with extended stays (often referred to as "bed-blockers") is essential for formulating effective treatment plans. Thus, pLOS serves as a key metric, directly influencing healthcare expenditures and available capacity.

The present study aimed to develop ML-based models for predicting both LOS and pLOS in general patient populations. Several techniques, including regression, support vector machine, KNN, random forest, gradient boosting trees, neural networks, and ensembles, were compared to identify the most effective models. Additionally, an analysis of the most relevant features for accurate prediction was conducted.

4.3 Related Works

Over the past two decades, researchers have employed various statistical techniques to investigate LOS and the influence of covariates such as age, gender, illness severity, diagnosis, and hospital characteristics. More recently, Machine Learning and Deep Learning (DL) have emerged as promising alternatives to these established methodologies in healthcare research [92, 93]. Studies exploring LOS patterns exhibit considerable heterogeneity [94, 95], often focusing on broad patient cohorts [96], specific age ranges [97, 98, 99], explicit discipline areas [100, 101] and medical specialties [102, 103], surgical procedures [104, 105] and oncological surgeries [106, 107], and individual hospital departments. However, only a limited portion of the existing literature addresses the specific context of the Italian public healthcare system.

Trunfio *et al.* [108] analyzed 2,515 patients undergoing hip-replacement surgery at the University Hospital of Salerno, Italy. Their analysis revealed that Multiple Linear Regression yielded the highest performance in predicting LOS ($R^2 = 0.616$), whereas Random Forest and Gradient-Boosted Tree models achieved an accuracy of 71.76% in predicting LOS as a discrete target (less than 7 days, 7-12 days, over 12 days).

Olivato *et al.* [109] developed an ML-based system to predict pLOS in COVID-19 patients admitted to the "Spedali Civili" hospital in Brescia. Their model, trained on demographic information and laboratory test results from over 6,000 admissions, attained a ROC-AUC score of 0.76.

Zeleke *et al.* [110] investigated 12,858 inpatients admitted through the emergency department of an Italian hospital in Bologna. Their Gradient Boosting classifier achieved an accuracy of 75% in predicting pLOS (defined as any stay exceeding 6 days), while Ridge and XGBoost regressors were most effective in forecasting LOS as a continuous outcome, with a prediction error ranging between 6 and 7 days.

In another Italian study by D'Onofrio *et al.* [111], the application of Random Forest achieved a 77.79% accuracy in predicting LOS for 989 patients undergoing mastectomy surgery at A.O.R.N. "Antonio Cardarelli" in Naples.

Di Matteo *et al.* [112] implemented an ML-driven system to forecast prolonged LOS (defined as LOS > 7 days) for hip-/knee-arthroplasty patients at "Humanitas Research" Hospital in Milan. Leveraging combined clinical and textual data from 1,517 patients, their model achieved an AUC of 0.789.

While many of these studies focus on specific departments or rely on data partially unavailable at admission (such as laboratory results), our research addresses this gap by employing various supervised ML algorithms to predict LOS for general inpatients using readily available data extracted from a medico-administrative platform. We analyzed LOS as a continuous, multi-class, and dichotomous variable, encompassing all medical-surgical departments with the purpose of developing robust and adaptable models for effective generalization. This approach mirrors real-world scenarios where patients may be relocated to alternate wards (often regardless of their primary service) when a department reaches full capacity. Evaluating all medical units collectively provides greater consistency. Additionally, focusing on admission data ensures immediate implementation across diverse hospital settings.

4.4 Materials and Methods

4.4.1 Data Selection and Inclusion Criteria

This study was conducted at the Ospedale di Sassuolo S.p.A., a general hospital located in Emilia-Romagna, Italy. The facility is organized by intensity of care and comprises 19 clinical units. We analyzed a dataset of 12,471 hospitalizations from 10,145 unique patients discharged between February 2022 and November 2023. All patients had a minimum length of stay of 24 hours. Data were extracted from the hospital's EBMS (Electronic Bed Management System), which included information on patient demographics, admission type, clinical features, and hospitalization details. A summary of patient characteristics is provided in Table 4.1.

To minimize potential biases in model performance, patients undergoing Day Surgery or Day Hospital procedures were excluded due to their predetermined LOS of one day. Additionally, to ensure data integrity, we excluded patients deceased during hospitalization, inpatients with stays exceeding the 99.95th percentile of the LOS distribution (outliers), and maternity/infancy wards due to their distinct clinical characteristics and potential data collection biases.

We further expanded the initial dataset by integrating historical information regarding each patient's prior hospitalizations, including the number of previous admissions in the last 12 months (particularly those requiring ICUs or high-intensity care levels), the average and total lengths of stay during previous hospitalizations, and the average LOS for all patients admitted within the same service as the current hospitalization in the preceding 30 days (aiming to capture potential department-specific trends).

4.4.2 Descriptive Statistics

An in-depth analysis identified significant associations between the length of stay and other covariates, including:

- *Age and LOS.* Length of stay increased with age, particularly from 75 to 95 years, aligning with established literature. This finding corroborates the notion that age plays a more pronounced role in determining LOS among older individuals, who often present with more complex clinical conditions (*e.g.*, frailty and cognitive decline)

Table 4.1: Patient Characteristics for LOS Prediction.

#	Feature	Total	Type
Patient demographics			
1	Seniority (age divided into 10-year groups)	12,471	Categorical
2	Gender (M/F)	12,471	Categorical
3	From outer province? (Y/N)	12,471	Boolean
4	From outer administrative district? (Y/N)	12,471	Boolean
Information from current admission			
5	Admission month (1-12)	12,471	Categorical
6	Admission day of week (1-7)	12,471	Categorical
7	Admission on weekend (Y/N)	12,471	Boolean
8	Admission on working day (Y/N)	12,471	Boolean
9	Admission hour of day (0-23)	12,471	Categorical
10	Admission during night-time? (Y/N)	12,471	Boolean
11	Admission from outer facility? (Y/N)	12,471	Boolean
12	Intensity care (L/M/H)	12,471	Categorical
13	Admission from ER? (Y/N)	12,471	Boolean
14	Short-Stay Observation (SSO)? (Y/N)	12,471	Boolean
15	Single room? (Y/N)	12,471	Boolean
16	Bed type	12,471	Categorical
Clinical information			
17	Terminal patient (End-of-Life)? (Y/N)	12,471	Boolean
18	Bedridden patient? (Y/N)	12,471	Boolean
19	Multidimensional geriatric assessment requested? (Y/N)	12,471	Boolean
20	Integrated Home Care requested? (Y/N)	12,471	Boolean
21	Isolation required? (Y/N)	12,471	Boolean
22	COVID-19 isolation? (Y/N)	12,471	Boolean
23	Contact isolation? (Y/N)	12,471	Boolean
24	Structural isolation? (Y/N)	12,471	Boolean
25	Other type of isolation? (Y/N)	12,471	Boolean
26	Diagnosis (text)	12,471	Text
Information from current hospitalization			
27	Hospitalization area	12,471	Categorical
28	Hospitalization area type	12,471	Categorical
29	Specialty (service)	12,471	Categorical
30	Recent transfers count	12,471	Numeric
31	Past transfers count	12,471	Numeric
32	Movements count	12,471	Numeric
33	ICU movements count	12,471	Numeric
Information from previous hospitalizations			
34	Patient prev. hospitalizations count (prior 12 mo.)	12,471	Numeric
35	Patient prev. ICU hospitalizations count (prior 12 mo.)	12,471	Numeric
36	Patient prev. hosp. with high-intensity care count (prior 12 mo.)	12,471	Numeric
37	Patient prev. hospitalizations average LOS (prior 12 mo.)	12,471	Numeric
38	Patient prev. hospitalizations total LOS (prior 12 mo.)	12,471	Numeric
39	Specialty prev. hospitalizations average LOS (prior 1 mo.)	12,471	Numeric

and social issues (*e.g.*, limited social support networks and lack of caregivers), contributing to extended hospital stays.

- *Hospitalization Area and LOS.* As expected, patients admitted to the long-term care and rehabilitation units exhibited the highest LOS. Medical departments showed comparable LOS, while ICU experienced shorter stays, confirming its role in stabilizing critically ill patients before transfer to other units. The surgical area, dedicated to short-stay routine interventions, demonstrated a shorter LOS compared to other surgical areas, such as those focused on urology, general surgery, orthopedics, and cardiology.
- *Intensity of Care and LOS.* Patients receiving lower-intensity care tended to have longer LOS than those receiving medium- or high-intensity care. Although this may seem counterintuitive, it likely reflects the association of high-intensity care with critical units where patients remain only as long as necessary to stabilize their condition before being transferred to medium- or low-critical departments. This also highlights the remarkable efficiency of high-intensity care units in expediting patient recovery and facilitating discharge.
- *Specialty and LOS.* Similar patterns were observed for the "Specialty" variable compared to "Hospitalization Area". However, the medical service offers a more reliable indicator of LOS than the area, as the latter may be influenced by temporary patient placements due to limited bed availability.
- *Discharge Type and LOS.* Patients with an ordinary discharge tended to have shorter LOS compared to those discharged to other facilities or requiring a protected discharge. Patients with advanced tumor stages (near end-of-life) and those with significant mobility limitations (bedridden) exhibited longer LOS.

These findings highlight the complex interplay between LOS and both patient-specific factors and hospital characteristics, providing insights for patient care optimization and targeted interventions.

4.4.3 Models Development

The research employed two variants of the available dataset. Dataset A contained only structured data, including demographics, clinical information, and admission details. Conversely, Dataset B included additional features derived from unstructured free-text diagnoses documented by healthcare practitioners.

Initially, fourteen regression algorithms were implemented to predict LOS as a continuous variable. Performance evaluation metrics included mean absolute error (MAE), root mean squared error (RMSE), R-squared (R^2), and adjusted R-squared scores. Additionally, ten classification methods were employed to predict LOS as a multi-class target (1-3 days, 4-10 days, >10 days). Evaluation metrics for classification encompassed accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUROC). The same ten classifiers were also used to predict pLOS, defined as any hospitalization exceeding eight days, corresponding to the 75th percentile. A comprehensive discussion of the models employed in this study is provided in Appendix C.

To extract features from diagnoses, a text-cleaning process was implemented, involving the removal of stop words and irrelevant or non-domain-specific terms. A pre-trained BERT foundation model[1] was then employed to tokenize the text and generate embeddings. Subsequently, a data preprocessing pipeline was applied to normalize all numerical variables and to one-hot encode categorical features. This step ensured that all features were on a comparable scale and that categorical features were numerically represented. Furthermore, principal component analysis (PCA) was applied to embeddings from Dataset B to reduce the dimensionality of the data to 100 components, thereby mitigating computational complexity.

Each dataset was randomly partitioned into a training set comprising 80% of the admissions (9,976) and a holdout/validation set encompassing the remaining 20% (2,495 admissions), using stratified sampling. A five-fold cross-validation approach was employed for each task to compare algorithms and identify the top-performing models. Hyperparameter tuning was then conducted for each selected model. Final performance was assessed on the independent test set.

Moreover, Voting and Stacking ensemble methods were added in the final evaluation. These methods aggregate predictions from multiple base models (using voting and stacking aggregation techniques, respectively) to

[1]Specifically, the "bert-base-italian-xxl-uncased" was used, with an 81GB training corpus size and 13,138,379,147 tokens. The choice to employ a generic BERT model was necessitated by the lack of publicly available language models specifically trained for the Italian medical domain.

improve overall accuracy and reduce model bias.

Classification Models.

According to the Italian Ministry of Health [113], the national average LOS for acute care in 2020 was 7.5 days. In Emilia-Romagna, the region where the Ospedale di Sassuolo S.p.A. is located, the average LOS for acute care in 2020 was 7.6 days, closely aligned with national data. Guided by these benchmarks and the requirements of the hospital under investigation, it was decided to split LOS into three distinct groups:

- Group 1: LOS $\leq$ 3 days (5,830 hospitalizations)
- Group 2: 4 $\leq$ LOS $\leq$ 10 days (4,445 hospitalizations)
- Group 3: LOS $>$ 10 days (2,196 hospitalizations)

The chosen thresholds ensured a roughly even distribution of observations across groups. As a result, no data-balancing techniques were applied.

Binary Classification Models.

Consistent with national trends, this study adopted an 8-day threshold to classify hospitalizations as either "short" or "prolonged". Hospitalizations with LOS ranging from 1 to 8 days accounted for 9,578 cases, while prolonged LOS (>8 days) represented 2,893 admissions (23.2% of all stays). Elderly age groups (70-79 years, 80-89 years, and 90-99 years) experienced longer LOS compared to younger cohorts, corresponding to 28%, 37%, and 12% of all prolonged stays, respectively. The majority of cases were observed in general medicine (39%) and long-term care wards (16%). Additionally, 67.3% of bed-blockers required medium-intensity care, while 26.3% and 6.4% required low- and high-intensity care, respectively.

The prevalence of nearly three times more instances of class 0 ("short" LOS) compared to class 1 ("prolonged" LOS) posed a potential challenge, as the model might overfit to the majority class (class 0) and struggle to accurately identify the minority class (class 1). To address this imbalance, several techniques were initially employed, including Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic sampling (ADASYN). These methods aim to balance the class distribution by generating synthetic instances of the minority class, thereby reducing the impact of class imbalance on model performance. However, implementing these techniques did not yield significant performance improvements for our specific dataset. This suggests that incorporating ensemble methods, which combine predictions from multiple learners, might be a more effective strategy for handling class imbalance, particularly when considering metrics beyond accuracy. To maintain methodological coherence, the pLOS prediction task leveraged the same battery of classifiers and evaluation metrics employed for the multi-class LOS task.

4.5 Results

4.5.1 Regression Models

Among the sixteen regressors evaluated for predicting LOS as a continuous variable on Dataset A (Table 4.2), the StackingRegressor ensemble achieved the strongest performance (MAE 2.81, R^2 0.635), followed by VotingRegressor and XGBRegressor. When considering Dataset B, which incorporated unstructured data, CatBoostRegressor emerged as the superior model (MAE 2.73, R^2 0.649), followed by VotingRegressor and XGBRegressor.

Table 4.2: Results of Tuned Models on Datasets A and B (Regression Task).

	Dataset A				Dataset B			
Model	**MAE ↓**	**RMSE ↓**	R^2 ↑	**Adj.**R^2 ↑	**MAE ↓**	**RMSE ↓**	R^2 ↑	**Adj.**R^2 ↑
Stacking Regressor	**2.806**	**4.614**	**0.635**	**0.633**	2.705	4.622	0.633	0.632
Voting Regressor	2.824	4.617	0.634	0.633	2.722	4.537	0.647	0.645
XGB Regressor	2.844	4.634	0.632	0.630	2.776	4.585	0.639	0.638
CatBoost Regressor	2.831	4.639	0.631	0.629	**2.726**	**4.520**	**0.649**	**0.648**
Linear Regression	2.976	4.692	0.622	0.621	2.911	4.622	0.633	0.632
Ridge	2.963	4.706	0.620	0.618	2.862	4.620	0.634	0.632
GB Regressor	2.946	4.723	0.617	0.616	2.923	4.739	0.615	0.613
LGBM Regressor	2.903	4.770	0.609	0.608	2.756	4.616	0.634	0.633
Elastic-net	2.995	4.776	0.609	0.607	2.887	4.679	0.624	0.623
SVR	2.883	4.784	0.607	0.606	2.792	4.694	0.622	0.620
Lasso	3.051	4.867	0.594	0.592	3.004	4.811	0.603	0.601
RF Regressor	2.968	4.903	0.588	0.586	2.843	4.787	0.607	0.605
KNN Regressor	2.989	5.012	0.569	0.567	2.868	4.957	0.578	0.577
AdaBoost Regressor	3.601	5.275	0.522	0.521	3.576	5.285	0.521	0.519
MLP Regressor	3.298	5.587	0.464	0.462	3.090	5.350	0.509	0.507
DT Regressor	3.882	6.617	0.249	0.246	3.757	6.604	0.252	0.249

Legend: XGB (XGBoost), GB (Gradient Boost), LGBM (LightGBM), SVR (Support Vector Regressor), RF (Random Forest), KNN (K-Nearest Neighbors), MLP (Multi-Layer Perceptron), DT (Decision Tree).

The integration of embedded representations derived from free-text diagnoses resulted in a measurable, albeit slight, performance enhancement across all models. This improvement can be partially ascribed to the ability of embeddings to encapsulate the semantic meaning of diagnoses, a task that is challenging to accomplish solely through conventional structured features. In addition, embeddings offer the advantage of modeling the relationships among diverse diagnoses, which becomes particularly valuable in the presence of comorbidities. Importantly, the inclusion of admitting diagnosis does not introduce bias or confound the study endpoint (*e.g.*, data leakage), as this information is inherently available at the time of hospitalization.

4.5.2 Classification Models

A comparative analysis of classifier performance for forecasting LOS on Dataset A revealed the superior performance of VotingClassifierSoft (accuracy 73.55%, F1-score 73.25%, AUROC 87.94%). The StackingClassifier and CatBoostClassifier trailed closely behind. For Dataset B, VotingClassifierSoft again emerged as the top performer (accuracy 76.27%, F1-score 75.96%, AUROC 89.60%), followed by CatBoostClassifier and StackingClassifier. Consistent with the findings from the regression analysis, the employment of embeddings derived from diagnoses yielded a modest improvement in performance (Table 4.3).

Table 4.3: Results of Tuned Models on Datasets A and B (Multi-Class Classification Task).

	Dataset A				**Dataset B**			
Model	**Acc. ↑**	**F1 ↑**	**A.ROC ↑**	**A.PRC ↑**	**Acc. ↑**	**F1 ↑**	**A.ROC ↑**	**A.PRC ↑**
Voting Soft	**0.735**	**0.732**	**0.879**	**0.777**	**0.763**	**0.760**	**0.896**	**0.794**
Stacking	0.732	0.730	0.879	0.774	0.763	0.761	0.893	0.789
CatBoost	0.733	0.729	0.878	0.776	0.758	0.754	0.894	0.788
XGB	0.728	0.724	0.875	0.775	0.762	0.760	0.892	0.789
RF	0.730	0.728	0.873	0.767	0.750	0.745	0.886	0.778
GB	0.729	0.726	0.872	0.764	0.760	0.757	0.890	0.783
LGBM	0.724	0.722	0.870	0.760	0.748	0.745	0.887	0.778
Log. Regression	0.730	0.727	0.864	0.750	0.740	0.737	0.883	0.773
KNN	0.702	0.697	0.847	0.726	0.718	0.710	0.862	0.733
MLP	0.678	0.679	0.827	0.707	0.719	0.717	0.852	0.729
AdaBoost	0.710	0.707	0.787	0.651	0.738	0.737	0.817	0.677
DT	0.654	0.652	0.724	0.519	0.655	0.655	0.720	0.513

The following figures show the AUROC and AUPRC curves, along with the classification report and confusion matrix, for one of the models with the highest predictive accuracy, CatBoost, assessed on both Dataset A and Dataset B.

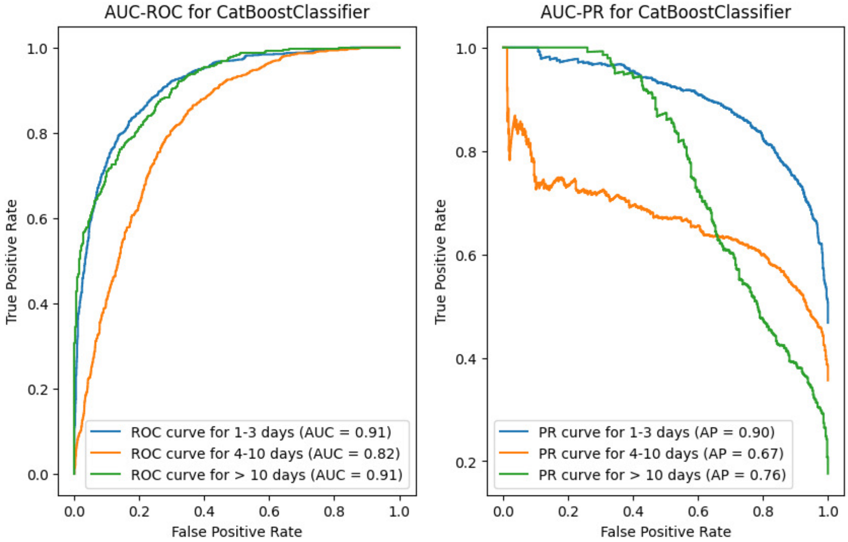
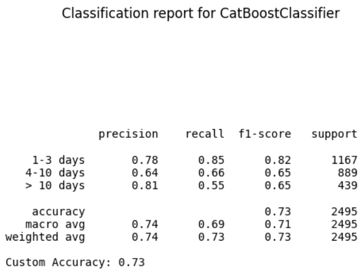


Figure 4.1: AUROC and AUPRC Curves for CatBoost Classifier on Dataset A (Multi-Class Classification Task).


Classification report for CatBoostClassifier

	precision	recall	f1-score	support
1-3 days	0.78	0.85	0.82	1167
4-10 days	0.64	0.66	0.65	889
> 10 days	0.81	0.55	0.65	439
accuracy			0.73	2495
macro avg	0.74	0.69	0.71	2495
weighted avg	0.74	0.73	0.73	2495

Custom Accuracy: 0.73

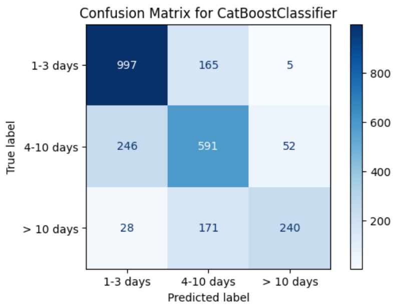


Figure 4.2: Classification Report and Confusion Matrix for CatBoost Classifier on Dataset A (Multi-Class Classification Task).

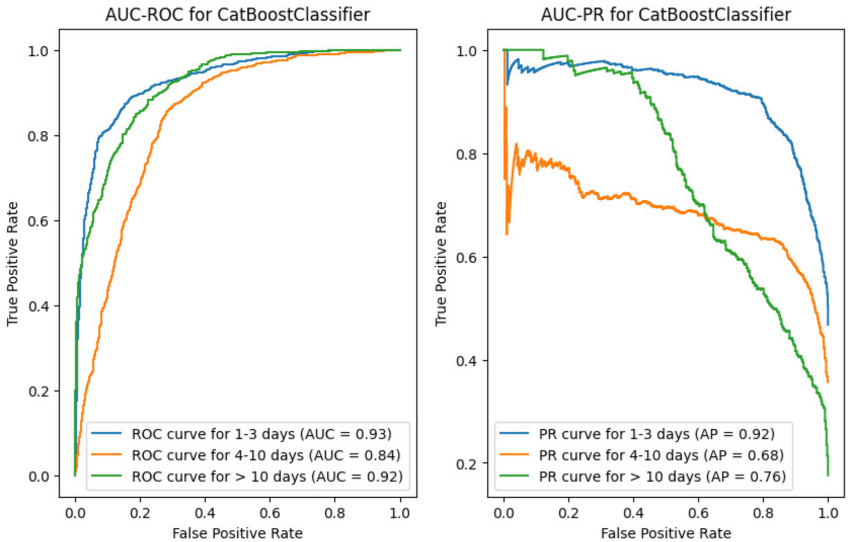
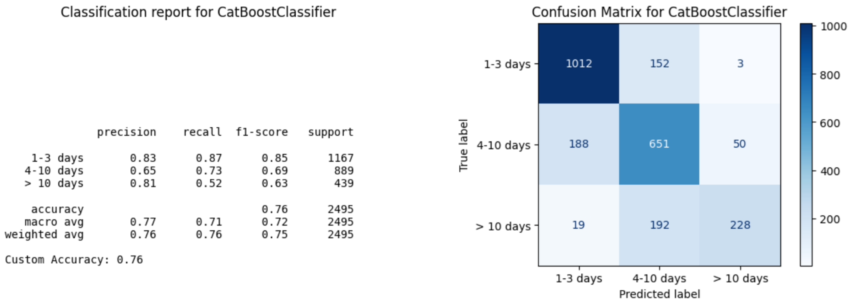


Figure 4.3: AUROC and AUPRC Curves for CatBoost Classifier on Dataset B (Multi-Class Classification Task).


Classification report for CatBoostClassifier

	precision	recall	f1-score	support
1-3 days	0.83	0.87	0.85	1167
4-10 days	0.65	0.73	0.69	889
> 10 days	0.81	0.52	0.63	439
accuracy			0.76	2495
macro avg	0.77	0.71	0.72	2495
weighted avg	0.76	0.76	0.75	2495

Custom Accuracy: 0.76

Figure 4.4: Classification Report and Confusion Matrix for CatBoost Classifier on Dataset B (Multi-Class Classification Task).

The feature importance analysis conducted using CatBoost on Dataset A (Figure 4.5) revealed several key factors affecting length of stay. Ranked in descending order of importance, the most significant features included: the overall average length of stay for same-service hospitalizations within the previous 30 days (*ba_specialty_prev_hosp_avg_los*); the number of recent transfers across wards (*ba_recent_transfers_count*), possibly indicative of increasing medical complexity (*i.e.*, patients needing specialized units); the surgery hospitalization area (*b_hosp_area_G*); the number of bed movements during hospitalization (*ba_movements_count*), including those within the same ward; low-intensity care level (*ba_intenscare_L*), suggesting that critical conditions tend to result in shorter stays due to a focus on stabilization before transfer; age in range 80-89 years (*p_age_range_80-89*), indicating a higher risk of prolonged stays for elderly inpatients, possibly due to age-related vulnerabilities or comorbidities; the average LOS for same-patient hospitalizations in the prior year (*ba_patient_prev_hosp_avg_los*); and the need for a multidimensional geriatric assessment (*cs_is_uvm_req*), typically associated with frail individuals and elderly.

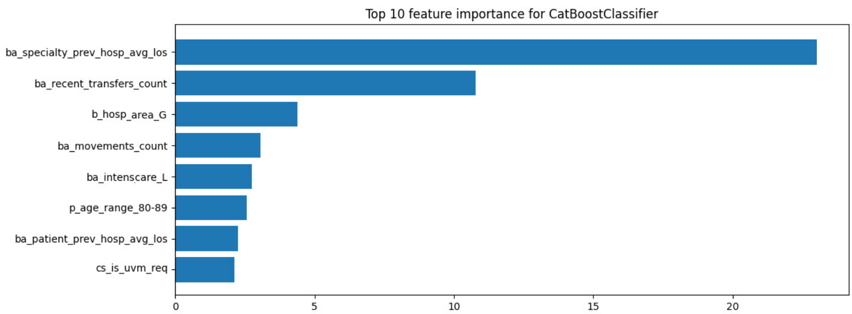


Figure 4.5: Feature Importance for CatBoostClassifier on Dataset A (Multi-Class Classification Task).

Interestingly, upon applying the same analysis to Dataset B, PCA components —derived from diagnosis text embeddings— started to emerge as significant predictors (Figure 4.6). This result highlights an important trade-off: while PCA offers a valuable tool for mitigating the curse of dimensionality, it can also lead to a less interpretable model. This limitation arises from the transformation of the original features into a new set of linearly combined variables, making it challenging to map the transformed representations back to their original interpretable elements.

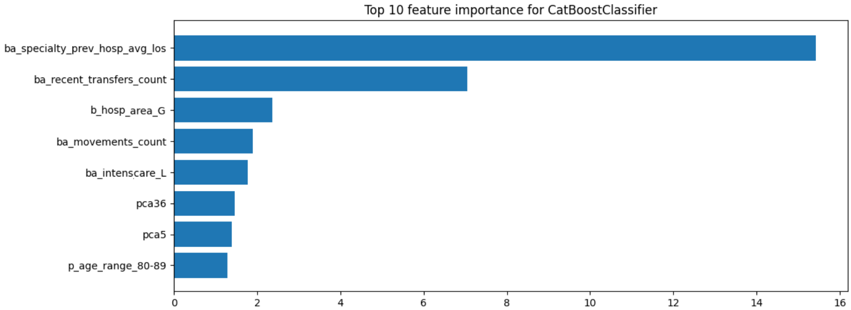


Figure 4.6: Feature Importance for CatBoostClassifier on Dataset B (Multi-Class Classification Task).

4.5.3 Binary Classification Models

In line with the multi-class classification task, AUROC and AUPRC (Area Under the Precision-Recall Curve) served as the primary evaluation metrics for the binary classification models. Ranging from 0 to 1, AUROC effectively captures the trade-off between true and false positives across all possible thresholds. Conversely, AUPRC prioritizes the identification of positive samples, making it particularly advantageous in scenarios involving imbalanced datasets.

For Dataset A, LogisticRegression provided the most accurate predictions for prolonged length of stay (accuracy 86.61%, F1-score 64.62%, AUROC 90.54%), followed by VotingClassifierSoft and CatBoostClassifier. On the other hand, when analyzing Dataset B, VotingClassifierSoft (accuracy 86.53%, F1-score 64.48%, AUROC 91.67%), CatBoostClassifier, and StackingClassifier demonstrated superior predictive capabilities for pLOS (Table 4.4).

Additionally, to assess the alignment between predicted and actual pLOS risks, we employed calibration curves for the three best-performing models (Figure 4.7). In binary classification tasks with probabilistic outputs, calibration curves offer a visual assessment of predicted probabilities compared to true class frequencies. An ideal model would exhibit a diagonal calibration curve, signifying perfect concordance between predicted and observed probabilities. This facilitates critical evaluation of model reliability, enabling the selection of models that generate trustworthy probability estimates for informed decision-making.

Table 4.4: Results of Tuned Models on Datasets A and B (Binary Classification Task).

	Dataset A				Dataset B			
Model	**Acc. ↑**	**F1 ↑**	**A.ROC ↑**	**A.PRC ↑**	**Acc. ↑**	**F1 ↑**	**A.ROC ↑**	**A.PRC ↑**
Log. Regression	**0.866**	**0.646**	**0.905**	**0.794**	0.866	0.654	0.911	0.798
Voting Soft	0.863	0.636	0.905	0.789	**0.865**	**0.645**	**0.917**	**0.808**
CatBoost	0.867	0.654	0.903	0.785	0.867	0.654	0.912	0.799
Stacking	0.867	0.654	0.903	0.785	0.867	0.654	0.912	0.799
GB	0.861	0.636	0.901	0.782	0.863	0.650	0.908	0.793
LGBM	0.861	0.642	0.898	0.777	0.862	0.642	0.907	0.791
AdaBoost	0.862	0.642	0.896	0.778	0.853	0.634	0.888	0.769
RF	0.863	0.635	0.893	0.777	0.857	0.584	0.904	0.788
XGB	0.842	0.606	0.884	0.751	0.862	0.655	0.901	0.785
KNN	0.848	0.545	0.884	0.743	0.862	0.516	0.874	0.726
MLP	0.827	0.618	0.867	0.733	0.862	0.625	0.874	0.745
DT	0.799	0.569	0.719	0.423	0.862	0.558	0.713	0.412

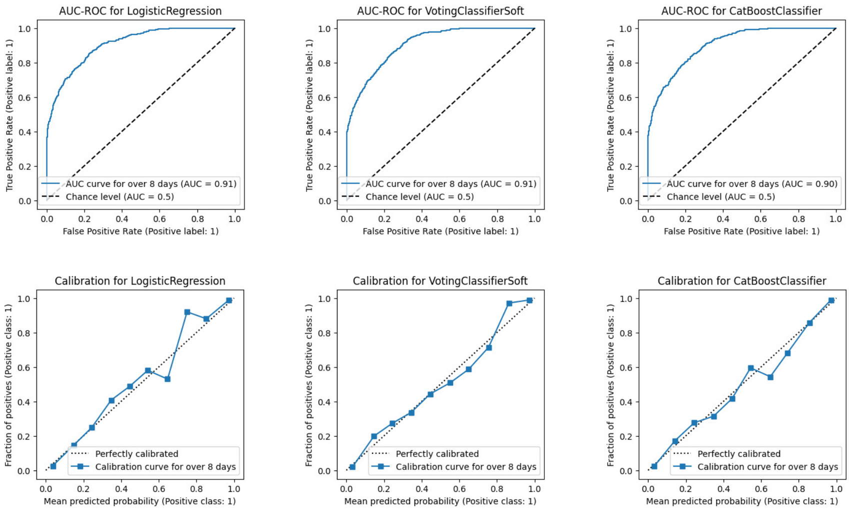


Figure 4.7: AUROC and Calibration Plot for LogisticRegression, VotingClassifierSoft, and CatBoostClassifier on Dataset A (Binary Classification Task).

In Dataset A, both VotingClassifierSoft and CatBoostClassifier proved well-calibrated, with points on their calibration plots clustering closely around the ideal diagonal line. However, VotingClassifierSoft revealed a propensity to overestimate the probabilities of pLOS in high-risk patients, whereas CatBoost exhibited a slight underestimation bias.

When considering Dataset B, CatBoost showed a reduced tendency to underestimate the likelihood of pLOS in high-risk inpatients (Figure 4.8).

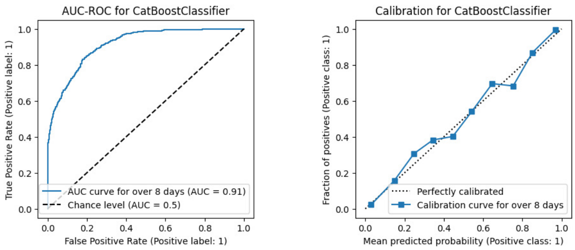


Figure 4.8: AUROC and Calibration Plot for CatBoostClassifier on Dataset B.

4.6 Discussion

While prior research has demonstrated success in predicting LOS for specific patient cohorts, such as individuals undergoing heart failure treatment [114], joint replacements [112], stroke interventions [115], and cesarean sections [116], this area remains relatively unexplored for broader, heterogeneous inpatient populations. The present work addresses this gap by considering a wider spectrum of diagnoses and conditions. By avoiding the limitations inherent in single-ward studies (*e.g.*, ICU [93, 100]), which may be constrained by internal dynamics, our methodology achieved encouraging generalizability for real-world implementation.

Our findings reinforce existing evidence that ensemble models consistently outperform individual algorithms for specific tasks [117]. Through the aggregation of multiple base learners, ensemble models effectively mitigate biases and improve overall predictive capabilities. Specifically, tree-based ensemble models allow for converting complex models into transparent decision rules. By employing tools such as SHAP values, Partial

Dependence Plots, and Individual Conditional Expectation Plots, practitioners can gain a deeper understanding of the model's decision-making process, which is pivotal for the acceptance and the adoption of ML-driven systems in clinical settings. Conversely, in our scenario, algorithms like AdaBoost and Multi-Layer Perceptron (MLP) exhibited lower performance, confirming that model selection should be guided by empirical evaluation rather than theoretical assumptions. Systematic performance comparisons remain essential for identifying the optimal model for a given task.

The comparative analysis of different variations of the same data source proved the instrumental role of integrating text extracted from diagnoses in capturing subtle nuances not represented in structured data alone. This is particularly relevant to leveraging the expertise of medical staff and the outcomes derived from physical examinations. Although public healthcare systems differ significantly from private or partially-public settings, limiting direct performance comparisons, our results align with prior investigations [118, 119, 120], which have demonstrated that text-derived features may enhance predictive performance.

Further analysis identified factors significantly associated with prolonged stays, including the average LOS for previous admissions in the same service and the number of transfers within the current hospitalization. When incorporating text-derived features, the need for a multidimensional geriatric assessment emerged as another key indicator for identifying patients at high risk of pLOS. These findings underscore the potential of machine learning to identify clinically relevant predictors of LOS and pLOS, potentially informing patient management and resource allocation strategies.

Lastly, our models leverage readily available data from institutional Electronic Health Records (EHRs), collected within the first 24 hours of hospitalization. This ensures that bed managers can swiftly employ them as decision-making aids upon patient admission, improving the timeliness of resource planning.

Despite its strengths, our study has certain limitations. First, its retrospective nature, relying on historical data, may introduce potential selection bias. Second, to prioritize real-world applicability and early prediction, we excluded less readily accessible data like vital signs and laboratory results, which could limit model accuracy. Incorporating more granular clinical data, such as specific medical interventions, might provide valuable insights and reduce prediction errors. On the other hand, the delayed availability and inconsistent collection of such information across departments (e.g., long-term care, rehabilitation) could constrain the model's applicability. Moreover, although clinical parameters are essential for patient monitoring, their direct impact on long-term outcomes like prolonged LOS (measured in days) might be less pronounced. Finally, while our study focused on data from a single public hospital to ensure a controlled environment and detailed analysis, its monocentric design restricts external validation. However, this limitation aligns with the documented influence of structural, organizational, and administrative factors on LOS, suggesting that a tailored approach, guided by our core framework, might be more appropriate than a one-size-fits-all solution. Nevertheless, the machine learning algorithms leveraged, including ensemble methods and text-derived features, are readily adaptable to diverse healthcare contexts. Additionally, the selected features are commonly available in most hospital information systems. This combination creates a lightweight yet valuable template that can be applied in diverse settings, acknowledging healthcare system diversity while providing an adaptable tool for LOS prediction. As additional Italian hospitals join this research, a more exhaustive assessment of the model's generalizability will be possible.

As a final consideration, the integration of Continuous Learning techniques represents a promising avenue for enhancing LOS prediction over time. By automatically labeling new hospital stays using their actual discharge dates, sourced from official administrative records such as the Hospital Discharge Register (Scheda di Dimissione Ospedaliera, SDO), the system could benefit from periodic retraining of models with fresh, labeled data. This approach would not only augment predictive performance but also mitigate the effects of data drift caused by evolving hospital workflows, patient demographics, clinical practices, and unforeseen events, thereby creating a robust, future-proof asset for optimizing resource allocation across diverse hospital environments.

4.7 Final Remarks

This study provides evidence that ensemble-based prediction models outperform traditional techniques in forecasting LOS and identifying general inpatients at high risk of prolonged LOS across diverse services and wards. While the individual machine learning algorithms employed are well-established, our novel contribution lies in the unique combination of these methods for LOS prediction. Our core methodology's reliance on readily available EHR data, coupled with algorithms that do not necessitate resource-intensive procedures or specialized hardware, promotes its practical integration into various healthcare settings and workflows, serving as a second-opinion tool to support both medical staff and healthcare management in their daily tasks.

A notable achievement was the preliminary embedding of our inference model into the EBMS employed at Ospedale di Sassuolo. By leveraging a combination of RESTful APIs and HL7 messaging, this integration significantly enhanced the user experience for bed managers involved in the experimental phase. Beyond providing a comprehensive visual representation of bed occupancy and availability across departments, the augmented version of the EBMS delivers predictive insights into expected LOS for each patient, particularly for those at high risk of prolonged hospitalization. Early identification of such patients allows healthcare providers to implement targeted interventions proactively, including closer monitoring and timely discharge planning, potentially mitigating the incidence of protracted stays. This preemptive approach can contribute to smoother patient flow, increased bed availability, a lower rate of rescheduled interventions, improved patient satisfaction, and, ultimately, reduced overall healthcare expenditures. The successful deployment of this technology within existing hospital infrastructure is particularly noteworthy given the context of Italy's digital landscape. As reported in the European Commission's recent report on the digital decade, the adoption of new technology stacks in Italy remains below 5%, attributed to a generally weak technological ecosystem and limited investment [121]. Our results emphasize the priority for hospitals to embrace adaptation and innovation to meet the demands of an aging population with chronic conditions, while containing costs and optimizing resources to ensure the sustainability of public healthcare services for future generations.

Chapter 5

Prediction of Prolonged Length of Stay in Emergency Departments

Overcrowding in Emergency Departments (EDs) is a pressing concern driven by high patient demand and constrained resources. Prolonged Length of Stay (pLOS), a major contributor to this congestion, can lead to adverse outcomes, including patients leaving without being seen, suboptimal clinical care, increased mortality rates, provider burnout, and escalating healthcare costs. This chapter investigates the application of various Machine Learning (ML) algorithms to predict both LOS and pLOS in EDs.

By identifying patients at high risk of pLOS, healthcare professionals can proactively implement strategies to expedite care, optimize resource allocation, and ultimately improve patient outcomes and ED efficiency, promoting a more effective and sustainable public healthcare delivery.

5.1 Introduction

Emergency Departments (EDs) play an essential role within the public health infrastructure, functioning as the primary point of access for emergency and urgent medical care on a continuous basis (24/7). Staffed by highly qualified physicians, EDs provide comprehensive evaluation and treatment for a wide range of illnesses and injuries.

The inherent unpredictability of emergencies necessitates a flexible approach within these departments, as activities cannot be predetermined. Patients may self-present or arrive by emergency vehicle, requiring a diverse array of treatments to address their individual needs. As a result, ED overcapacity poses a major global challenge to healthcare systems, primarily due to the widening gap between the ever-increasing demand for emergency services and the limited resources available. This imbalance is further compounded by an ageing population and rigorous cost-containment policies.

The effects of ED overcrowding are multiple, ranging from extended wait times and delayed treatment to ambulance rerouting and higher expenditures for hospitals. Additionally, overcrowding is associated with an escalating incidence of patients leaving the ED without physician consultation due to dissatisfaction with prolonged wait times [122, 123]. Addressing these challenges requires a comprehensive approach, combining effective triage systems and strategic resource allocation. Over the past two decades, EDs have been implementing initiatives like specialized Fast-track models [124, 125, 126], and Short-Stay Observation (SSO) units [127, 128]. Despite these interventions have demonstrated benefits, there remains substantial potential to further optimize ED resource management through the integration of accurate Length of Stay (LOS) prediction tools at the triage stage, an innovation that is currently lacking in many healthcare systems.

ED-LOS, defined as the time elapsed from a patient's arrival in the ED to their departure (encompassing discharge, transfer to another healthcare setting, or admission to a medical ward), is influenced by numerous variables. These include hospital organization, staffing levels, triage procedures, the need for consultations or advanced imaging [129], bed availability [130], and patient characteristics like age, sex, and comorbidities [131]. Prolonged LOS (pLOS) in EDs is not merely symptomatic of overcrowding but also a contributing factor, leading to deterioration of clinical conditions for waiting patients [132, 133], higher mortality rates [134], and severe provider burnout [135, 136]. The development of predictive models to forecast ED-LOS upon triage completion represents an emerging area of research. The application of analytics and Machine Learning (ML) techniques holds significant promise for early identification of patients at elevated risk for pLOS, enabling targeted interventions to prevent complications, streamline care processes, and ensure timely transitions between care settings. Such a proactive approach not only helps alleviate the adverse effects of overcrowding but also safeguards the long-term effectiveness and sustainability of EDs within the public health network. The ojective of this study was to design ML-based models for predicting ED-pLOS in general patient populations. Multiple algorithms were compared to identify the most effective models, examining the impact of different LOS thresholds. Additionally, key features influencing predictions were analyzed.

5.2 Related Works

Recent applications of ML models for predicting LOS in healthcare settings have evidenced their potential in dealing with the inherent complexities of Emergency Departments, where conventional statistical techniques often struggle due to their sensitivity to outliers and reliance on linearity assumptions.

Rahman *et al.* [137] utilized a decision tree encompassing 33 attributes to identify patients at risk of prolonged ED-LOS (defined as LOS exceeding 4 hours) in an Australian public hospital, achieving an accuracy of 85%. Similarly, Etu *et al.* [138] employed gradient boosting to predict ED-pLOS for COVID-19 patients in a Michigan hospital, attaining an accuracy of 85% as well. D'Etienne *et al.* [139] developed a two-step predictive model for early detection of extended ED-LOS (>16 hours), demonstrating 67.8% accuracy. Walsh *et al.* [140] applied Artificial Neural Network (ANN) to forecast ED disposition for infants with bronchiolitis, considering patient demographics and medical tests, reporting an accuracy of 81%. Benevento *et al.* [141] evaluated various regression techniques for projecting waiting times in two EDs located in northern-central Italy. Their approach incorporated data on new arrivals and queue status to improve prediction accuracy, reaching a Mean Absolute Error (MAE) ranging from 20.8 to 26.2 minutes in one hospital and 46 to 53.1 minutes in the other. Ricciardi *et al.* [142] explored multiple ML algorithms to predict prolonged ED-LOS (> 3 hours) in a university hospital in Salerno, Italy, with Random Forest performing with the highest accuracy (74.9%).

The importance of feature selection in predicting ED-LOS has been recognized in other works. For example, Azari *et al.* [143] proposed a framework that integrates data mining and class imbalance learning to estimate prolonged LOS (>14 hours). Similarly, Naemi *et al.* [144] emphasized the critical role of handling missing values and data skewness in LOS prediction, for both regression and classification tasks, yielding an AUC-ROC between 66% and 82%, respectively.

Despite extensive efforts in developing predictive models for prolonged length of stay in EDs, a key challenge remains the lack of a standardized definition for pLOS, with thresholds varying considerably across studies (3-24 hours). This inconsistency hinders the comparability of research findings, making it difficult to draw definitive conclusions about the effectiveness of different prediction models. Our study aims to address this gap by applying ML models to predict pLOS across multiple thresholds, thereby assessing the impact of choosing the most appropriate cut-offs. Notably, our approach leverages readily available data obtained after nursing assessment, such as demographics, triage codes, ICD-9 codes, and ED status metrics.

5.3 Materials and Methods

5.3.1 Data Selection and Inclusion Criteria

This retrospective study was conducted in collaboration with Ospedale di Sassuolo S.p.A., a secondary-level hospital situated in Emilia-Romagna, Italy, whose Emergency and Admission Department records an average of 40,000 annual visits.

We examined a dataset of 32,967 encounters, representing 24,829 unique patients discharged between November 2022 and January 2024. All visitors had a minimum documented length of stay in the ED of one hour. Data included information on patient demographics, clinical features, and access details. A summary of patient characteristics is presented in Table 5.1.

To mitigate potential biases and ensure data quality and integrity, we excluded patients deceased upon arrival, visitors who left without being seen (0.27% of the dataset), and sixteen outliers with stays exceeding the 99.95th percentile of the LOS distribution (corresponding to 54 hours).

Historical information from each patient's prior encounters, along with several fundamental ED metrics, was integrated to enrich the original dataset, including:

- *Patient-specific factors*: the number of previous visits within the last month (particularly those requiring SSO or involving critical

Table 5.1: Patient Characteristics for ED-LOS Prediction.

#	Feature	Total	Type
Patient demographics			
1	Seniority (age divided into 10-year groups)	32,967	Categorical
2	Gender (M/F)	32,967	Categorical
3	From outer province? (Y/N)	32,967	Boolean
4	From outer administrative district? (Y/N)	32,967	Boolean
Information from current access			
5	Admission month (1-12)	32,967	Categorical
6	Admission day of week (1-7)	32,967	Categorical
7	Admission on weekend (Y/N)	32,967	Boolean
8	Admission on working day (Y/N)	32,967	Boolean
9	Admission during night-time (11PM-5AM)? (Y/N)	32,967	Boolean
10	Access type	32,967	Categorical
11	Arrival mode	32,967	Categorical
12	Treatment	32,967	Categorical
13	Specialty	32,967	Categorical
14	Fast-track? (Y/N)	32,967	Boolean
Clinical information			
15	Acceptance reason	32,967	Categorical
16	Main problem	32,967	Categorical
17	Main problem category	32,967	Categorical
18	Main diagnosis (ICD-9)	32,967	Categorical
19	Main diagnosis notes (text)	32,967	Text
20	Secondary diagnosis (ICD-9)	32,967	Categorical
21	Diagnosis count	32,967	Numeric
22	Suspected COVID-19? (Y/N)	32,967	Boolean
23	Suspected violence? (Y/N)	32,967	Boolean
24	Work-related injury? (Y/N)	32,967	Boolean
25	Triage code (arrival)	32,967	Categorical
26	Assessment count	32,967	Numeric
ED metrics			
27	Current patients count	32,967	Numeric
28	Current patients count w/ RED triage code	32,967	Numeric
29	Current patients count w/ ORANGE triage code	32,967	Numeric
30	Current patients count w/ GREEN triage code	32,967	Numeric
31	Current patients count w/ BLUE triage code	32,967	Numeric
32	Current patients count w/ WHITE triage code	32,967	Numeric
33	Overall accesses count (prior 24 hr.)	32,967	Numeric
34	Overall average LOS (prior 12 hr.)	32,967	Numeric
35	Overall average LOS w/ RED triage code (prior 12 hr.)	32,967	Numeric
36	Overall average LOS w/ ORANGE triage code (prior 12 hr.)	32,967	Numeric
37	Overall average LOS w/ GREEN triage code (prior 12 hr.)	32,967	Numeric
38	Overall average LOS w/ BLUE triage code (prior 12 hr.)	32,967	Numeric
39	Overall average LOS w/ WHITE triage code (prior 12 hr.)	32,967	Numeric
Information from previous accesses			
40	Patient prev. accesses count (prior 30 days)	32,967	Numeric
41	Patient prev. accesses average LOS (prior 30 days)	32,967	Numeric
42	Patient prev. accesses total LOS (prior 30 days)	32,967	Numeric
43	Patient prev. accesses w/ SSO count (prior 30 days)	32,967	Numeric
44	Patient prev. accesses w/ RED triage code count (prior 30 days)	32,967	Numeric

conditions) and the average and total length of stay during previous visits.

- *Department context*: the number of patients present at the time of admission (total and by triage code) and the total number of patients admitted in the preceding 24 hours.
- *Temporal trends*: the average LOS for all patients admitted to the ED within the last 12 hours (total and by triage code).

The inclusion of department context and temporal trends was particularly aimed at capturing potential patterns within the ED environment.

Short-Stay Observation posed a methodological challenge. Patients undergoing SSO necessitate extended monitoring, typically ranging from 6 to 44 hours [145]. As these individuals remain under ED supervision, their LOS encompasses the observation period, thereby inflating the overall metric. Although SSO instances comprise only a small portion of the dataset (3%, with a mean LOS of 16 hours), excluding them could introduce bias, as they do not qualify as statistical outliers. Furthermore, despite nearly 75% of SSO patients enter the observation period within four hours from access, indicating early decision-making regarding their disposition, the necessity for SSO cannot be reliably determined at the time of admission. Guided by these considerations, we retained Short-Stay Observations while omitting any variable directly indicative of SSO status.

A similar challenge arose with Fast-track accesses, which represent a considerable portion of the dataset (19%). These pathways are specifically designated for patients exhibiting mild symptoms and lower acuity, resulting in shorter LOS (less than four hours). While their inclusion was debated due to the potential skewing of the average LOS, we opted to retain them, as Fast-track routes are integral to ED operations. Furthermore, despite a shorter average length of stay, individual Fast-track patients can still experience significant variation in their respective LOS. Including these patients allows for a more comprehensive analysis of factors influencing LOS in the ED.

The triage codes analyzed in this study adhered to the Italian five-tier numerical priority system (Table 5.2), ranging from 1 (most critical) to 5 (non-urgent), with a corresponding color-coded scheme (1=Red, 2=Orange, 3=Light Blue, 4=Green, 5=White). This system provides indications of maximum waiting times for care pathway initiation, which vary by priority code, with immediate access granted for emergencies and up to 240 minutes for non-urgent cases. Notably, these waiting times represent only a portion of the overall length of stay.

Table 5.2: ED Priorities and Maximum Waiting Times.

Priority Code	Color Code	Designation	Definition	Maximum Wait Time
1	Red	Emergency	Interruption or compromise of vital functions.	Immediate Access
2	Orange	Urgency	Risk of vital function impairment, stable condition with potential deterioration or severe pain.	15 min.
3	Light Blue	Deferrable Urgency	Stable condition without risk of deterioration, discomfort affecting the general state, often requiring complex care.	60 min.
4	Green	Minor Urgency	Stable condition without risk of deterioration, typically requiring simple mono-specialist, diagnostic or therapeutic services.	120 min.
5	White	Non-Urgent	Non-urgent problem with minimal clinical relevance.	240 min.

5.3.2 Descriptive Statistics

An initial quantitative analysis of the data revealed several key findings. Pediatric patients (0-5 years old) were the most frequent visitors, followed by adults in two age groups (28-33 and 50-60 years). Gender distribution among admissions was relatively balanced. Nearly 27% of patients originated from outside the hospital's province, with a further 11% residing within the province but outside the immediate administrative district. Admissions followed a distinct daily pattern, with peak volumes occurring during morning hours (8:00-11:00 and 12:30-14:30). Weekly distribution was fairly consistent, with a slight increase on Mondays. A seasonal trend was also evident, with admissions peaking in November, December, and January. Nearly 20% of admissions were expedited through Fast-track processes, particularly in pediatrics and obstetrics. Triage assessments revealed that the majority of patients presented with low (61.8%) or deferrable (26.8%) criticality. Almost 9% of patients required secondary assessment, and approximately 10% had a recent history of ED visits within the previous month. Roughly 7% of patients required hospitalization after assessment.

Subsequent examination identified notable relationships between the length of stay in ED and other covariates, including:

- *Admission Status and Disposition.* Patients with protected discharge (*e.g.*, transfers to nursing or rehabilitation facilities, support through home care services, or detailed post-discharge care plans) experienced the shortest LOS, whereas those awaiting hospitalization tended to have prolonged stays as a result of extended observation periods or delays in bed availability. Subjects transported by ambulance or medical cars generally recorded longer stays, suggesting a potential correlation between mode of arrival and the need for more extensive evaluation and treatment. Interestingly, individuals admitted from correctional facilities exhibited shorter LOS, possibly due to heightened safety concerns.

- *Fast-Tracking.* Fast-track admissions demonstrated significantly shorter LOS (1.3 ± 1.0 hour) compared to standard admissions (3.7 ± 3.8 hours). This confirms that implementation of Fast-track pathways is an effective strategy to reduce patient wait times and potentially enhance overall hospital efficiency. Specifically, the pediatric fast track was highly active during the weekend, when family pediatricians are off duty.

- *Main problem.* Conditions such as withdrawal syndrome, depression, suicide attempts, and substance overdose were linked to longer stays. Notably, substance abuse-related admissions had markedly longer LOS (6.4 ± 6.8 hours), reflecting the need for intensive detoxification, withdrawal management, or behavioral health complications requiring extended monitoring. Admissions due to self-harm and suicide attempts also showed longer LOS, likely due to the severity of the condition, the necessity of specialized care, and the involvement of social support systems. In contrast, gynecological and obstetric admissions were resolved more swiftly.

- *Triage Code and Patient Acuity.* The mean length of stay varied according to triage categories. Patients assigned a White triage code reported expeditious stays, with an average duration of 2.5 ± 2.2 hours. Those with a Green triage code had slightly longer stays, averaging

2.7 ± 2.5 hours. As the triage severity increased, so did the duration of ED stay. Encounters classified under the Red triage code had a mean LOS of 3.7 ± 4.6 hours, Light Blue admissions averaged to 4.0 ± 4.3 hours, and Orange triage cases exhibited the longest mean LOS of 4.8 ± 5.6 hours. These findings, though seemingly counterintuitive, reflect the complex dynamics of ED operations. While Red and Orange triage cases receive immediate medical attention, their overall LOS tends to be longer due to the complexity of their conditions, which often require extensive diagnostic procedures and treatments. Conversely, White and Green triage cases, despite potentially experiencing longer waiting times for initial assessment, generally result in shorter overall LOS, as they typically require fewer resources and interventions after evaluation.

- *SSO and Hospitalization.* As anticipated, undergoing Short-Stay Observation increased LOS, with the majority of patients associated with Orange and Light Blue triage codes. Furthermore, while patients subsequently admitted to Pediatrics and Obstetrics departments exhibited shorter ED-LOS, those admitted to Long-Term Care experienced longer stays in the ED, possibly due to bed availability shortage, complex medical needs, and bureaucratic processes involving social work assessments and coordination with other facilities.
- *Age.* Younger patients (0-6 years) experienced shorter stays, whereas LOS increased slightly for patients aged 45 and older. This pattern likely reflects the complexity of medical conditions and comorbidities associated with advancing age.
- *Reassessments and Complexities.* Cases requiring reassessment predictably resulted in longer stays in the ED due to waiting times for test results, availability of specialists, changing clinical conditions, and unforeseen complications. Patients with multiple ED visits within the previous 30 days (so-called "*frequent flyers*") were also associated with higher LOS, suggesting the presence of chronic conditions, complex medical cases, or social factors contributing to recurrent visits to the ED and extended stays.

The analysis indicates that ED-LOS is significantly influenced by three key components: individual patient characteristics, admission-related factors, and ED operational variables. Understanding these relationships offer valuable guidance for improving both patient care delivery and resource management.

5.3.3 Datasets and Model Development

Building upon the experiments in Chapter 4, this investigation utilized a dual-dataset methodology. Dataset A consisted exclusively of structured data, including patient demographics, clinical details, and admission-related information. Dataset B incorporated supplementary features derived from free-text triage notes documented by nursing personnel. The analysis of triage notes —aimed at extracting potentially valuable information— involved removing common stop words and non-specialized terms, tokenizing the text into manageable units, and generating embeddings using a pre-trained Italian BERT base model (XXL uncased variant). To reduce computational complexity, Principal Component Analysis (PCA) was applied to the embeddings in Dataset B, limiting dimensionality to 100 primary components. Standardization of numerical variables and one-hot encoding of categorical features into binary indicators were achieved through a preprocessing pipeline, ensuring consistent feature scaling prior to model development.

A total of twelve classification algorithms was utilized to forecast ED-LOS as a multi-class target, categorized into three intervals (1-4 hours, 5-8 hours, >8 hours). The same classifiers were subsequently used to predict pLOS at four distinct thresholds (4, 6, 8, and 12 hours). To enhance predictive accuracy and mitigate biases associated with individual models, ensemble methods such as Voting and Stacking were also implemented. A detailed description of the models used can be found in Appendix C.

Both datasets A and B were randomly partitioned, with 80% (n = 26,373) allocated for training and 20% (n = 6,594) reserved as a hold-out validation set. Model evaluation employed a five-fold cross-validation strategy to identify top-performing classifiers, followed by systematic hyperparameter optimization using a grid search approach. Classification performance was assessed using a comprehensive suite of metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUROC). Final model performance was evaluated on the independent test set.

Classification Models.

Adhering to international guidelines for patients presenting to the Emergency Department, the Italian Ministry of Health recommends establishing a clinical diagnosis within a maximum of four hours [145]. Additionally, for patients requiring hospitalization, the total length of stay in the ED should not exceed eight hours from the time of initial evaluation [146]. These directives apply even in cases of complex clinical presentations and aim to facilitate effective management of the diagnostic-therapeutic pathway. In alignment with these benchmarks, our study stratified ED admissions into three distinct LOS categories:

- Group 1: LOS $\leq$ 4 hours (26,392 admissions)
- Group 2: 5 $\leq$ LOS $\leq$ 8 hours (5,205 admissions)
- Group 3: LOS > 8 hours (1,370 admissions)

Binary Classification Models.

To further analyze ED stays, we investigated binary classification models to distinguish between "short" and "prolonged" LOS. Four pLOS thresholds, commonly employed in the literature, were considered: 4, 6, 8, and 12 hours.

The significant disproportion in the dataset (Table 5.3) between the majority class (patients with a "short" length of stay) and the minority class (patients with a prolonged LOS) presented a potential challenge. In such scenarios, models are prone to overfitting to the majority class, which can hinder their ability to correctly identify instances in the minority class, ultimately resulting in skewed and potentially misleading accuracy metrics.

Table 5.3: Different Thresholds Considered for ED-pLOS Definition.

Threshold	**Patients with LOS > cutoff (%)**	**Mean LOS (hours)**	**Std. Dev. (hours)**	**Total encounters**
4 hours	6,575 (20%)	8.00	5.58	32,967
6 hours	2,735 (8.3%)	11.71	7.15	32,967
8 hours	1,370 (4,2%)	16.05	8.00	32,967
12 hours	751 (2.3%)	21.05	7.79	32,967

Both ADASYN (ADAptive SYNthetic sampling) and SMOTE (Synthetic Minority Over-sampling TEchnique) techniques were evaluated to balance the class distribution by synthesizing new data points for the minority class. However, no significant improvement was observed following oversampling —likely due to the inherent ability of ensemble methods to effectively mitigate the effects of class imbalance. Consequently, these techniques were excluded from the final evaluation. For methodological consistency, the prediction of pLOS employed the same algorithms and performance metrics as those used in the multi-class LOS analysis.

5.4 Results

5.4.1 Classification Models

When applied to Dataset A (Table 5.4), the StackingClassifier demonstrated superior performance in predicting ED-LOS (accuracy 88.60%, F1-score 87.86%, AUROC 94.16%). The LGBMClassifier, XGBClassifier, and CatBoostClassifier followed closely behind. Similar performance patterns were observed with Dataset B, where StackingClassifier again emerged as the top performer (accuracy 87.64%, F1-score 86.77%, AUROC 93.51%).

Table 5.4: Results of tuned models on datasets A and B (multi-class classification task).

	Dataset A				**Dataset B**			
Model	**Acc. ↑**	**F1 ↑**	**A.ROC ↑**	**A.PRC ↑**	**Acc. ↑**	**F1 ↑**	**A.ROC ↑**	**A.PRC ↑**
Stacking	**0.886**	**0.879**	**0.942**	**0.820**	**0.876**	**0.868**	**0.935**	**0.802**
LGBM	0.880	0.871	0.940	0.811	0.872	0.862	0.935	0.798
XGB	0.878	0.868	0.939	0.810	0.874	0.864	0.934	0.800
CatBoost	0.875	0.866	0.935	0.805	0.876	0.867	0.933	0.797
Voting Soft	0.877	0.865	0.938	0.812	0.873	0.860	0.933	0.802
GB	0.874	0.864	0.932	0.786	0.873	0.862	0.930	0.784
MLP	0.860	0.858	0.911	0.750	0.858	0.855	0.912	0.745
RF	0.867	0.851	0.929	0.789	0.849	0.823	0.909	0.745
LR	0.866	0.849	0.911	0.733	0.865	0.847	0.909	0.731
DT	0.863	0.848	0.910	0.727	0.864	0.849	0.909	0.725
KNN	0.844	0.813	0.864	0.676	0.834	0.796	0.844	0.643
AdaBoost	0.842	0.812	0.909	0.695	0.842	0.812	0.909	0.695

Legend: LGBM (LightGBM), XGB (XGBoost), GB (Gradient Boost), MLP (Multi-Layer Perceptron), RF (Random Forest), LR (Logistic Regression), DT (Decision Tree), KNN (K-Nearest Neighbors).

While accuracy is a less reliable metric for imbalanced datasets, as in the current case, the Area Under the Precision-Recall Curve (AUPRC) effectively captures the trade-off between precision and recall, a critical consideration when modeling infrequent yet impactful events like extended stays. Notably, the inclusion of embedded representations derived from free-text notes did not result in performance improvement. This unexpected finding may be attributed to constraints in BERT's ability to capture the nuances of medical language specific to ED triage notes. These limitations could manifest as misinterpretations of abbreviations, medical jargon, and local conventions within the notes, hindering the model's capacity to fully grasp their semantic meaning. Existing biomedical language models such as BioBERT or ClinicalBERT, which have demonstrated effectiveness through fine-tuning, are not currently trained on Italian corpora. Moreover, an analysis of a representative sample of several hundred triage notes revealed partial redundancy, as much of the information contained within the unstructured text was already present in Dataset A. Consequently, incorporating this information would likely offer minimal incremental predictive power and could instead introduce noise into the modeling process.

Feature importance analysis using the XGBoost classifier on Dataset A identified several key factors influencing ED-LOS (Figure 5.1). As expected, ED occupancy emerged as a significant contributor to extended LOS. Treatment modality was also influential: general treatment, typically involving more extensive procedures, contrasted with Fast-track processing, which generally correlated with shorter stays due to its focus on less severe cases. Temporal factors, including admission hour and day of the week, were found to affect LOS, potentially due to variations in staffing levels, availability of diagnostic services, and overall ED workload. Patient age was a critical determinant, with pediatric cases frequently receiving prioritized treatment, which impacted LOS. Specific diagnoses and conditions, such as orthopedic issues, gastrointestinal complaints, and lower back pain, required extensive evaluation, resulting in longer stays. Cardiac-related conditions, like angina, also necessitated thorough assessment, contributing to prolonged LOS. Triage levels were another important factor, with higher severity categories (*e.g.*, orange, red) typically entailing more comprehensive care, thus increasing LOS. Insurance status, particularly coverage by the Italian National Institute for Insurance against Accidents at Work (INAIL) correlated with specific injury types and treatment protocols, leading to extended stays. Patients suspected of COVID-19 infection also experienced longer LOS due to additional testing and isolation measures.

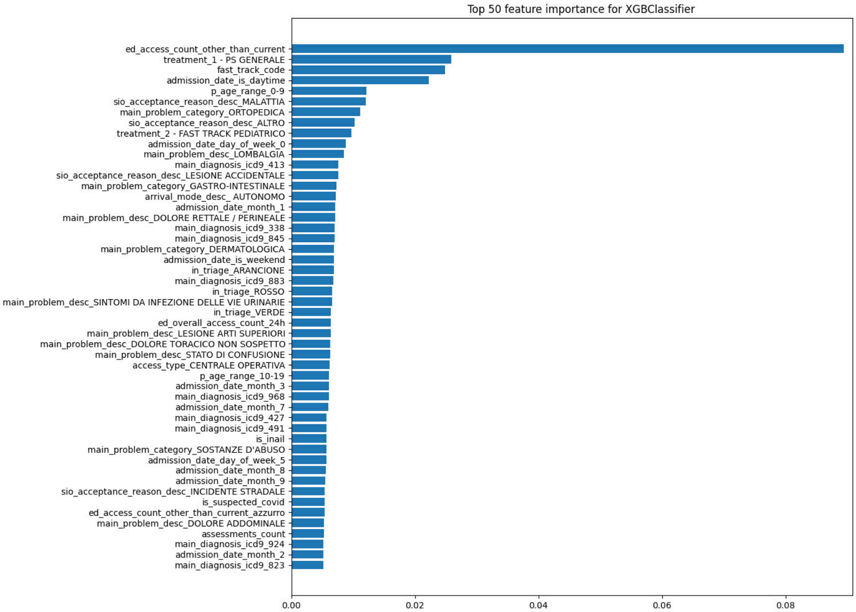


Figure 5.1: Feature Importance for XGBClassifier on Dataset A (Multi-Class Classification Task).

5.4.2 Binary Classification Models

Predicting ED-pLOS involves a binary classification approach derived from the multi-class classification task of ED-LOS: length of stay is converted into a binary variable, distinguishing between short or typical LOS and extended or prolonged LOS. For instance, using an average LOS threshold (e.g., 4 hours), the task focuses on predicting whether a patient's LOS will exceed this threshold (pLOS) or not.

When assessing prediction of pLOS using a four-hour cutoff on Dataset A, the XGBClassifier exhibited superior performance, achieving the highest AUPRC, F1-Score, and AUROC. Given the moderately imbalanced dataset

(20% positive class), the XGBClassifier proved optimal (Table 5.5).

Table 5.5: Results of Top-Performing Models on Dataset A (4h Cutoff).

Model	Acc. ↑	Precision ↑	Recall ↑	F1 ↑	AUROC ↑	AUPRC ↑
XGB	**0.906**	**0.826**	**0.670**	**0.740**	**0.938**	**0.843**
CatBoost	0.906	0.834	0.659	0.736	0.938	0.842
Stacking	0.906	0.834	0.659	0.736	0.938	0.842
LGBM	0.904	0.821	0.660	0.732	0.938	0.841
Voting Soft	0.904	0.833	0.646	0.728	0.937	0.836

As the pLOS cutoff extended to six hours, class imbalance intensified (8.3% positive class). The LGBMClassifier achieved the highest F1-Score, closely approaching the top performance in AUPRC and AUROC (Table 5.6).

Table 5.6: Results of Top-Performing Models on Dataset A (6h Cutoff).

Model	Acc. ↑	Precision ↑	Recall ↑	F1 ↑	AUROC ↑	AUPRC ↑
LGBM	**0.958**	**0.860**	**0.585**	**0.696**	**0.963**	**0.808**
XGB	0.957	0.868	0.565	0.684	0.964	0.812
CatBoost	0.955	0.848	0.563	0.677	0.961	0.806
Stacking	0.955	0.848	0.563	0.677	0.961	0.806
Voting Soft	0.955	0.870	0.539	0.666	0.963	0.807

For the eight-hour cutoff, characterized by severe class imbalance (4.2% positive class), the XGBClassifier again excelled, yielding the best F1-score and AUROC, while maintaining a competitive AUPRC (Table 5.7).

Table 5.7: Results of Top-Performing Models on Dataset A (8h Cutoff).

Model	Acc. ↑	Precision ↑	Recall ↑	F1 ↑	AUROC ↑	AUPRC ↑
XGB	**0.980**	**0.892**	**0.602**	**0.719**	**0.978**	**0.808**
LGBM	0.980	0.887	0.602	0.717	0.977	0.811
CatBoost	0.978	0.883	0.551	0.679	0.975	0.804
Stacking	0.978	0.883	0.551	0.679	0.975	0.804
GB	0.978	0.898	0.544	0.677	0.969	0.784

In the twelve-hour cutoff scenario, the dataset became extremely imbalanced, with only 2.3% of instances belonging to the positive class. Despite all top models achieving high accuracy (0.991), this metric was deemed unreliable for model selection. AUPRC was prioritized as the most critical indicator, reflecting sensitivity to minority class prediction. XGBClassifier once again demonstrated superior overall performance (Table 5.8).

Table 5.8: Results of Top-Performing Models on Dataset A (12h Cutoff).

Model	Acc. ↑	Precision ↑	Recall ↑	F1 ↑	AUROC ↑	AUPRC ↑
XGB	**0.991**	**0.896**	**0.687**	**0.777**	**0.991**	**0.857**
LGBM	0.991	0.925	0.660	0.770	0.991	0.862
Voting Soft	0.991	0.951	0.647	0.770	0.990	0.848
MLP	0.990	0.815	0.707	0.757	0.985	0.819
CatBoost	0.990	0.914	0.640	0.753	0.987	0.832

The XGBClassifier consistently demonstrated robust performance across all pLOS thresholds (4, 6, 8, and 12 hours). While other models, such as LGBMClassifier and VotingClassifierSoft, performed well in specific scenarios, the robustness of XGBClassifier, particularly in handling highly imbalanced datasets, is noteworthy and suggests an optimal balance between identifying patients at risk of prolonged stays (high recall) and reducing unnecessary interventions for patients not truly at risk (high precision). In addition, calibration curves were generated for the XGBClassifier at each cutoff to assess the alignment between predicted probabilities and observed outcomes. Given that a diagonal line represents perfect calibration (*i.e.*, predicted probabilities correspond exactly to observed event frequencies), XGBClassifier proved well-calibrated for the four-hour cutoff in Dataset A, with plot points closely clustering around the ideal diagonal line (Figure 5.2).

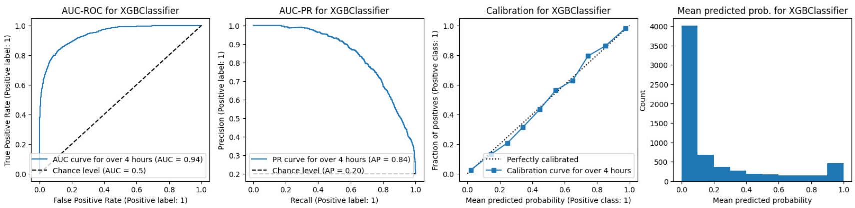


Figure 5.2: AUROC, AUPRC, and Calibration Plot for XGBClassifier on Dataset A (4h Cutoff).

However, as the threshold increased, the calibration curve worsened

(Figure 5.3). Overestimation of mid to high-range probabilities became evident at the eight-hour cutoff (Figure 5.4), escalating to severe miscalibration at the twelve-hour cutoff (Figure 5.5). These findings indicate that, in the examined context, the XGBClassifier is more reliable for short-term (4- and 6-hour horizons) than for longer-term resource planning.

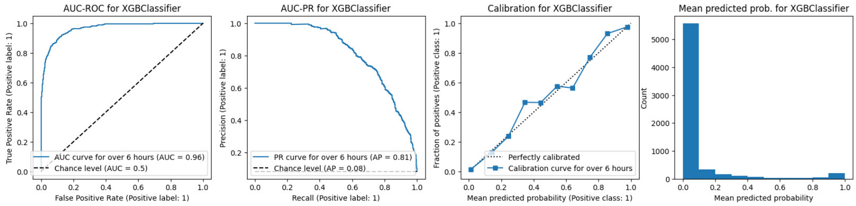


Figure 5.3: AUROC, AUPRC, and Calibration Plot for XGBClassifier on Dataset A (6h Cutoff).

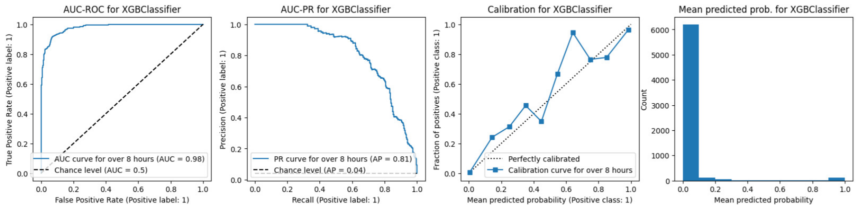


Figure 5.4: AUROC, AUPRC, and Calibration Plot for XGBClassifier on Dataset A (8h Cutoff).

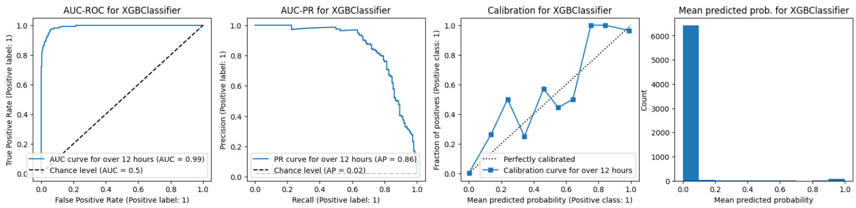


Figure 5.5: AUROC, AUPRC, and Calibration Plot for XGBClassifier on Dataset A (12h Cutoff).

Consistent with the findings from the multi-class classification task, incorporating embeddings from triage notes (Dataset B) did not improve the results.

5.5 Discussion

This study employed a comprehensive machine learning approach to identify factors influencing ED-LOS and to develop predictive models. A comparative analysis of twelve algorithms revealed the superiority of ensemble models in accurately forecasting LOS across three distinct categories (1-4 hours, 5-8 hours, >8 hours) and predicting the likelihood of prolonged stays at four predefined thresholds (4, 6, 8, and 12 hours). The combination of multiple individual algorithms into ensemble methods reduces bias while enhancing predictive performance [147, 148, 149]. In particular, ensemble models built upon decision trees offer clear visibility into their reasoning process, supporting their potential integration into clinical practice, where understanding the rationale behind predictions is essential.

While nearly all ensemble models exhibited high accuracy for the investigated ED, the XGBoost classifier demonstrated superior consistency and a well-balanced trade-off between precision and recall. The inclusion of queue-based indicators (both global and triage code-specific), representing real-time and historical ED status, significantly bolstered the models' predictive capacity, underscoring the importance of considering operational contexts in LOS forecasting. Unexpectedly, embeddings derived from free-text triage notes had minimal influence on model performance in this ED setting, contrasting our previous findings on hospital admissions [150]. This discrepancy may be partly attributed to the inclusion of ICD-9 codes as features in the ED models, as these codes likely capture relevant diagnostic information, diminishing the added value of textual embeddings. Our prior hospital LOS models did not include diagnosis codes, rendering free-text notes more impactful.

Although this study focused on a specific ED, the underlying methodology exhibits potential for broader applicability across diverse settings with minimal site-specific adjustments. Notably, while Italian guidelines recommend static triage code-based thresholds – primarily focused on initial nursing assessment and representing only a limited portion of the ED experience – our model's ability to forecast the entire LOS offers a more comprehensive and adaptive strategy for ED management.

This research acknowledges certain limitations. The retrospective nature of the analysis may introduce selection bias, and the limited feature set, despite favoring readily available data, excludes potentially valuable information such as vital signs and laboratory test results. While incorporating these additional data points could refine predictions, it might reduce the model's transferability, creating a trade-off that requires careful consideration. Additionally, the study's single-center nature, drawing exclusively from one emergency department, limits the scope of external validation. It is important to note, however, that ED-LOS is a complex metric influenced by institution-specific variables—including operational workflows, regional governance, and available resources. Consequently, comprehensive generalizability may not be entirely feasible or even optimal in this context.

Future efforts should prioritize model explainability, as ensemble models are often perceived as "black boxes" due to their complexity, and such opacity is not an option in clinical decision-making. While many tree-based ensemble models (*e.g.*, Random Forest and XGBoost) provide built-in feature importance scores to identify influential predictive variables, advanced interpretability techniques, including SHAP (SHapley Additive exPlanations), Partial Dependence Plots (PDPs), and Individual Conditional Expectation (ICE) plots, offer more granular insights. SHAP values decompose individual predictions by quantifying feature-specific contributions, helping clinicians understand the factors driving model outputs (*e.g.*, patient age, diagnosis, admission type); PDPs visualize the marginal effect of features on predicted outcomes, while ICE plots extend this analysis by uncovering heterogeneous effects across individual patients.

To ensure consistent explainability across model of varying complexity, additional strategies merit consideration. Model calibration should present probabilities and risks in clinically familiar formats, such as confidence intervals or probability thresholds, to integrate seamlessly with existing decision-making processes. Natural language explanations, potentially generated by local language models, could translate model insights into accessible clinical interpretations—for instance, clarifying that advanced age and prior ICU admission history are the primary determinants of a predicted extended stay. Additionally, regular training sessions for healthcare professionals would both demystify model mechanics and generate valuable feedback for improving explanatory frameworks and clinical utility.

5.6 Final Remarks

According to the National Agency for Regional Health Services [151], Italian Emergency Departments recorded a total of 18.27 million visits in 2023, representing a 6% increase compared to 2022. As patient volumes rise, staff efficiency diminishes, resulting in extended waiting times. Moreover, overcrowding in EDs, often exacerbated by organizational inefficiencies and personnel shortages [152], contributes to patients leaving without receiving medical attention, elevated patient mortality and morbidity rates, as extensively documented in the literature [153, 154, 155].

Predictive analytics, particularly when integrated into the triage process, offer a viable solution to alleviate these challenges. Early identification and monitoring of patients at risk for prolonged LOS allows ED administrators to initiate timely interventions, such as adjusting staffing levels and refining patient triage and assessment protocols. For example, the ability to anticipate which patients would benefit from SSO units can improve operational planning.

In the contemporary landscape of constrained public funding, optimizing ED operations, preserving assets, and containing costs are critical priorities. The proposed data-driven framework has the potential to enhance resource management and foster proactive, targeted actions, particularly for patients with severe conditions.

Chapter 6

MondrIAn: A First-Class Data-Driven Platform To Support Collaborative Bed Management

Many of the experimental investigations presented in this dissertation would not have been possible without the systematic collection and aggregation of clinical data, previously fragmented across diverse repositories — both electronic and paper-based— characterized by significant constraints in cross-system compatibility and integration.

This chapter illustrates a concrete example of digital transformation implemented within an Italian public healthcare institution. The analysis involved a preliminary assessment of the operational context and established protocols, followed by a re-engineering of patient pathway management processes. This revision was orchestrated to accommodate evolving care and organizational needs, particularly notable in the post-pandemic landscape, while enhancing procedural transparency and information dissemination.

Above all, this analysis documents an enriching journey that began two years ago and is still ongoing.

6.1 Background

Hospital bed management represents a critical determinant of healthcare delivery efficiency and institutional economic performance. The inherent complexity of managing bed availability and utilization has been primarily amplified by demographic shifts, particularly population aging [156], and dramatically highlighted during the COVID-19 pandemic, which exposed the vulnerabilities in resource allocation systems when confronted with sudden demand fluctuations.

Challenges in bed management stem from its extensive interplay with multiple hospital departments and its influence on patient flow dynamics. Ineffective management manifests in emergency departments congestion, surgical procedure cancellations, suboptimal patient placements, and prolonged hospitalizations [157]. These inefficiencies can trigger adverse events, including increased medication errors, hospital-acquired infections, and elevated mortality rates [158, 159]. Additionally, delayed discharges impede proper patient flow [160, 161] from the emergency department and throughout the hospital [162, 163], creating bottlenecks that compromise care quality and patient safety.

The financial impact of bed mismanagement is substantial [164]. Length of stay (LOS) variations account for approximately 85-90% of hospital cost differentiation among patients [165]. Notably, each day exceeding the expected stay increases total costs by 10-12%, with intensive care unit oversights being even more costly—approximately three times the standard daily care expenses.

In addition, the organizational structure of healthcare institutions, traditionally characterized by departmentalization, often results in fragmented communication and decision-making processes, hindering the systemic approach necessary for optimal bed utilization [166]. The complexity is further compounded by the need for coordinated communication across physically separated care teams, particularly in planning care progression and discharge procedures.

While expanding operational capacity through infrastructure investment presents one solution pathway, it proves neither sustainable nor indefinitely scalable [167].

The recent adoption of Electronic Bed Management Systems (EBMS) has proven to be a significant advancement in addressing these challenges [168, 169, 170]. Such systems facilitate internal communication between medical personnel, allow for efficient bed allocation, and support coordinated discharge planning. This technological approach aligns with the principle of "think global, act local", acknowledging that effective hospital bed management requires both system-wide perspective and unit-level execution.

However, successful implementation requires transformative shifts in organizational culture, business models, and structural processes [171, 172]. To explore these dimensions, we conducted a collaborative study across five Italian hospitals located within the same geographical region. The synergy of combined effort, collective expertise, and decades of on-the-ground experience led to the development of the first *cooperative* EBMS, called MondrIAn.

This research is significant for two primary reasons. To our knowledge, it represents the first initiative to involve multiple public healthcare facilities focusing on co-bed management. Secondly, it demonstrates how digital transformation, beyond process optimization, can generate unprecedented opportunities, leading to outcomes that surpass the sum of individual contributions.

6.2 Related Works

The evolution of patient flow management systems within hospitals reflects a broader, ongoing digital transition in healthcare operations. Long-standing approaches have historically relied primarily on paper records, whiteboards, and scheduled meetings to monitor bed availability and patient journeys, resulting in information silos and demanding considerable effort to maintain, particularly during unplanned events.

In recent years, healthcare institutions worldwide have increasingly adopted digital tools to address these limitations.

Electronic Patient Journey Boards (EPJBs) represent an early advancement in the visualization of patient care progress [173, 174]. Conventional patient journey boards —typically positioned at nurses' stations— necessitated manual updates, whereas their electronic counterparts offer real-time insights. Building upon the EPJB concept, some institutions have introduced electronic hospital capacity dashboards (e-Dashboards) to facilitate more comprehensive monitoring of patient pathways [175]. Despite these contributions to better information display, the literature examining their effectiveness in hospital-wide patient flow management remains limited, as these devices often operate in isolation and lack seamless integration with other systems.

A different approach involves embedding discharge planning tools directly into Electronic Health Record (EHR) systems. Standard EHR messaging frameworks and discharge order entries, though essential, frequently lack the capability for real-time, integrated communication among clinical teams. By incorporating tools such as electronic discharge checklists [176] and asynchronous reporting features, clinicians can track discharge readiness and address potential barriers in a timely manner. Nonetheless, these tools often provide only passive communication capabilities, limiting their efficacy. Additionally, legacy EHR systems may prevent the integration of external tools.

Other recent and remarkable developments include the implementation of web-based Kanban solutions for bed management optimization [165]. Originally conceived in manufacturing to improve resource and task flow efficiency, the Kanban framework has been adapted to healthcare settings [177, 178], incorporating visual representations of bed allocation to support workflow prioritization. Although the application of this methodology has shown promising results in small- to medium-sized hospitals, its effectiveness remains contingent upon strict adherence to established protocols by staff, and may struggle in complex, dynamic environments, particularly when dealing with inter-departmental task dependencies and the potential oversimplification of patient data.

Electronic Bed Management Systems (EBMSs) address key communication and operational challenges associated with traditional, paper-based approaches [179, 168]. Multidisciplinary teams, especially in critical care settings, require synchronized and readily accessible evidence to make informed bed management decisions. EBMSs provide centralized access to bed status indicators, patient flow data, and resource allocation details, allowing staff to prioritize patient care and reduce operational inefficiencies caused by fragmented information and manual processes [169]. Nevertheless, the recent pandemic has emphasized the interconnected nature of healthcare systems, with unprecedented demand for hospital beds highlighting the imperative for coordinated, region-wide strategies.

These considerations lay the groundwork for introducing a *Cooperative Electronic Bed Management System (CEBMS)*—a novel paradigm that leverages a distributed platform for real-time, collaborative bed management among proximate public healthcare institutions, extending the capabilities of traditional EBMSs and fostering regional coordination in patient care delivery.

6.3 Methods and Intervention

6.3.1 Overview of Patient Flow and Bed Management

In healthcare, "patient flow" denotes the systematic progression of patients through sequential care stages, departmental units, and provider interactions within the continuum of care. Achieving effective patient flow hinges on comprehensive management protocols that ensure appropriate bed allocation based on clinical requirements, minimizing delays and unnecessary diversions due to bed shortages [180, 160].

Bed management serves as an essential, equilibrating mechanism between bed demand and supply, engaging all organizational levels—from operational to strategic. Through bed management optimization, healthcare facilities can mitigate congestion while sustaining operational efficiency, requiring continuous policy refinement to align with strategic objectives and institutional goals. The responsibilities of a bed manager include real-time reporting and coordination of patient volume, bed occupancy rates, and discharge forecasts in conjunction with anticipated admissions [181]. Discharge forecasting, in particular, is fundamental, as bed availability remains highly sensitive to patient admission and discharge dynamics. While timely bed turnover facilitates seamless new admissions when inpatients are ready for discharge, non-clinical factors —including family unpreparedness and pending diagnostic results— may extend length of stay, further complicating coordination efforts.

Emergency bed management introduces additional complexities. Given the unpredictability of arrival rates, with historical averages and seasonal experience often serving as primary decision-making tools, throughput inefficiencies in Emergency Departments (ED) emerge as primary contributors to hospital capacity constraints, leading to prolonged wait times, heightened patient dissatisfaction, and potential transfers to alternative facilities [182], underscoring the importance of accurate discharge prediction.

Such complex interaction of interrelated processes requires a holistic, rather than compartmentalized, analytical approach.

6.3.2 Case Study

The implementation of a novel Electronic Bed Management System was assessed across five hospitals within the province of Modena, Italy.

The study commenced in March 2021 in partnership with Ospedale di Sassuolo S.p.A., a healthcare facility operating under a hybrid public-private management model. The hospital spans approximately 40,000 square meters, accommodates 246 beds (of which 9 in the intensive care unit), employs 800 staff members, and primarily serves a population of over 110,000 residents in the Sassuolo health district, with an average of 1,290 monthly discharges.

In March 2022, the investigation was extended to encompass four additional public hospitals within a proximity network, organized under a "hub-and-spoke" model governed by the Local Health Authority of Modena[1]. The participating facilities included Ospedale di Carpi (279 beds, including 16 in intensive care), Ospedale di Mirandola (130 beds, 4 in intensive care), Ospedale di Pavullo (123 beds), and Ospedale di Vignola (95 beds). Collectively, this network provides 873 available beds throughout the Modena province[2], equivalent to 2.88 beds per 1,000 inhabitants [183]. This figure is marginally below both the regional benchmark of Emilia-Romagna (2.94 beds per 1,000 inhabitants) and the national average (3.1 beds per 1,000 inhabitants [184]), yet closely aligned with the World Health Organization's (WHO) recommended target of three beds per 1,000 inhabitants [185].

6.3.3 Study Design

The assessment of existing bed management protocols involved pairing two researchers with at least one bed manager per facility, accumulating 200 hours of on-site review. The evaluation comprised 88 hours dedicated to interviews, analysis, and requirements gathering, complemented by 112 hours of operational shadowing during the initial implementation phase.

[1] Azienda Sanitaria Locale (ASL) di Modena.

[2] Neither the "Azienda Ospedaliera-Universitaria Policlinico di Modena" nor the "Nuovo Ospedale Civile di Baggiovara" is included in the total bed capacity calculation.

Scheduled visits were conducted across all departments integral to each hospital's bed management team, with approximately ten hours allocated per facility for observation and inquiry. The qualitative analysis framework included structured interviews with ten hospital bed management administrators and three information technology managers. Supplementary interviews with clinical and administrative staff provided insights into their experiences with the discharge process and allowed for comprehensive evaluation of existing tools and communication strategies used to share information on discharge barriers and patient readiness.

This methodological approach facilitated the identification of both common patterns and divergent practices across the healthcare facilities under study.

The requirements' gathering phase was carried out through collaborative discussions with hospital management and bed management specialists, leading to a development strategy that prioritized operational efficiency while minimizing personnel overhead.

6.3.4 Settings and Context

Each of the hospitals under investigation employs two dedicated bed managers, working in rotating shifts, to manage bed requests. The staff members are responsible for coordinating bed allocation across various pathways —from routine admissions and scheduled procedures to emergency cases— while simultaneously monitoring projected patient discharges. Bed allocation requests primarily originate from internal wards, emergency care, and external referrals (such as other regional healthcare facilities, nursing homes, and residential care homes for the elderly). Although operating independently, all hospitals participating in this study utilize the same administrative software platforms to manage ED patient flow and admission-discharge-transfer (ADT) workflows. However, each facility maintains its own distinct system instance.

Upon receiving a request, managers evaluate both current bed availability and anticipated discharges. However, during the observational period of the study, delays in ADT system updates were repeatedly noted, with beds occasionally remaining marked as unavailable after patient discharge, leading to deferrals in sanitization procedures. Interviews revealed that the bed request workflow is generally perceived as time-consuming and procedurally intensive by nursing staff, often subordinated to more urgent operational priorities. Despite the availability of Electronic Medical Records (EMRs), bed requests are frequently processed manually through forms transcribed by administrative personnel, resulting in limited transparency and necessitating multiple calls for bed availability confirmation. In cases involving inter-facility transfers, where additional communication is required to coordinate patient movements effectively, the process becomes even more cumbersome.

Initial comparative analysis highlighted organizational differences among hospitals. The Ospedale di Sassuolo implemented a centralized control room, where bed managers maintain phone contact with various departments, acting as a central hub for bed requests. Conversely, other hospitals adopted a hybrid approach, wherein designated areas supervise non-critical beds, semi-critical beds, emergency care requests, and external transfers. In these facilities, bed managers routinely visit departments, personally assessing availability and identifying potential critical issues.

Historically, Ospedale di Sassuolo utilized a spreadsheet-based system for managing bed allocations, which, over time, evolved to provide a more accurate map of bed assignments across wards. Internal and external transfer requests were similarly documented using dedicated spreadsheets. At the conclusion of each business day (approximately 7:30 p.m.), a printed copy of the current bed status report was provided to the on-duty emergency room medical staff, allowing them to promptly assess bed availability in urgent situations. Other hospitals implemented a decentralized system of interconnected spreadsheets, each maintained independently by ward managers, though structural variations existed across facilities. At the time of investigation, bed management procedures relied on labor-intensive manual reconciliation of bed assignment spreadsheets through comparison with ADT data, which also over-complicated historical analysis and limited process improvement. Additionally, a few departments leveraged manual whiteboards to track bed utilization and patient details, leading staff to frequently circumvent the official ADT system in favor of informal (and potentially outdated) communication channels for obtaining updates. These practices encouraged information silos, hindering real-time visibility and restricting occupancy awareness to a ward-by-ward perspective, which ultimately compromised hospital-wide bed management efficiency.

All contributing hospitals consistently employed operation management indicators to monitor bed flow, including:

- *Hospital occupancy rate*: the percentage ratio between the number of patient-days and the number of bed-days within a given period.
- *Mean length of hospital stay*: the ratio between the total number of patient-days and the total number of discharges within a specified period, including deaths.
- *Replacement interval index*: the average time a bed remains unoccupied between the discharge of one patient and the admission of another.
- *Bed turnover* the average number of patients per hospital bed, with intensive care beds being an exception.

6.3.5 Limitations of the Current Approach

The research identified several critical operational inefficiencies in bed management processes across the participating healthcare institutions, as evidenced by interviews with bed managers, clinicians, and IT managers.

A primary challenge involved the labor-intensive task of maintaining near real-time synchronization of bed occupancy spreadsheets, wherein staff were required to cross-reference ADT and ED systems twice daily. Additionally, during off-hours, bed management responsibilities shifted to ED physicians and department coordinators, requiring bed managers to manually re-enter overnight changes upon their return—an effort typically taking one to two hours. Despite these efforts, manual updates frequently proved unreliable, necessitating direct departmental communication for verification.

Preservation of historical data presented another substantial challenge. Bed managers were tasked with creating daily backup copies of occupancy spreadsheets, with any oversight risking the loss of critical records. Furthermore, reconstructing patient flow movements was particularly complex, especially in cases involving infections, as it required consultation of multiple versions of spreadsheets—a process both time-consuming and susceptible to error.

Limited coordination between wards and bed managers often led to assignment errors, typically discovered hours later. Although bed managers were expected to be notified of each bed change, urgent situations frequently disrupted communication. Technical limitations further complicated the situation, as shared spreadsheets could become temporarily inaccessible when left open by ward coordinators, forcing bed managers to work on a local copy of the locked file.

Existing tools, including conventional whiteboards and spreadsheets, lacked the comprehensive information needed to capture a complete view of patient care and progress. Space constraints within spreadsheets prevented the inclusion of crucial information (*e.g.*, admission diagnoses), prompting the use of various color-coding and font conventions to convey critical information such as infections or pending discharges. Unfortunately, these expedients varied across institutions and even among individual bed managers, potentially leading to misinterpretation. In addition, structural departmental bed rearrangements, even when transient, necessitated extensive spreadsheet reformatting, adding to the administrative burden.

Critical metrics, like length of stay and discharge timing, were updated manually on a daily basis. Moreover, discharge planning lacked standardization and information was not consistently transferred when patients moved between wards, necessitating data duplication. In a multidisciplinary discharge process, the availability of shared and accessible information is vital for efficient management. Systems that rely on manual, paper-based, or spreadsheet-based methods prove demanding, impeding timely communication while imposing administrative burdens that limit efforts to streamline hospital operations. All the stakeholders consistently indicated that real-time access to information would facilitate more expedient and precise decision-making while reducing procedural overlaps.

Patient safety considerations emerged as another prominent theme, with particular emphasis on infection surveillance capabilities, especially concerning COVID-19 transmission.

IT managers identified organizational culture as the primary barrier to digital transformation, particularly due to the need to alter some established workflows. Concerns emerged about potential resistance from bed management staff, who might perceive digital shift as a loss of control over the process. Additionally, considering both the diverse technological background of the upcoming EBMS's intended users and the primary focus of healthcare professionals on patient care, an intuitive and efficient user interface (UI) was identified as a priority, with redundant data entry deemed unacceptable.

Building on these insights, development objectives for implementing a functional EBMS centered on three primary requirements:

- Enhanced visual clarity in presenting clinical information.
- Real-time dissemination of both clinically and operationally relevant information across multidisciplinary teams.
- System configurability to accommodate local requirements.

A preliminary low-fidelity prototype was designed and presented to key stakeholders (hospital and bed managers) for evaluation and iterative refinement. Feedback was systematically incorporated into subsequent modifications, following an evidence-based design methodology. Leveraging the in-depth local knowledge of multidisciplinary teams, the EBMS was developed to support both clinical and unit processes.

6.4 Implementation of the EBMS

6.4.1 System Design and Basic Features

A responsive web-based application was identified as the optimal implementation model, offering seamless accessibility across a heterogeneous array of devices, including workstations, laptops, and tablets.

The application architecture follows the Model-View-Controller (MVC) pattern, with a hierarchical data model in which rooms contain one or more beds, wards comprise collections of rooms, and each ward is associated with a specific facility.

Users can be assigned to one or more predefined groups (*e.g.*, Bed Manager, Hospital Administration, Emergency Room Staff, Ward Coordinators, IT Manager), each having distinct authorization levels. This approach restricts access, visibility, and functionality based on role-specific permissions. Furthermore, platform login is fully integrated with the organization's central authentication system, eliminating the need for users to manage separate credentials.

Extensive efforts were devoted to interface design, iteratively developed in collaboration with bed managers and hospital administrators, to ensure a user-friendly, intuitive, and visually engaging experience.

Upon their first access, users are introduced to primary interactive areas and key functionalities via a brief on-screen tutorial.

The homepage provides a detailed, high-fidelity visual map of wards and bed locations, which supports visual memory and facilitates user orientation. Ward layouts are pre-configured to reflect actual architectural floor plans, yet modifications can be made through the EBMS administration interface to accommodate updates or changes in infrastructure.

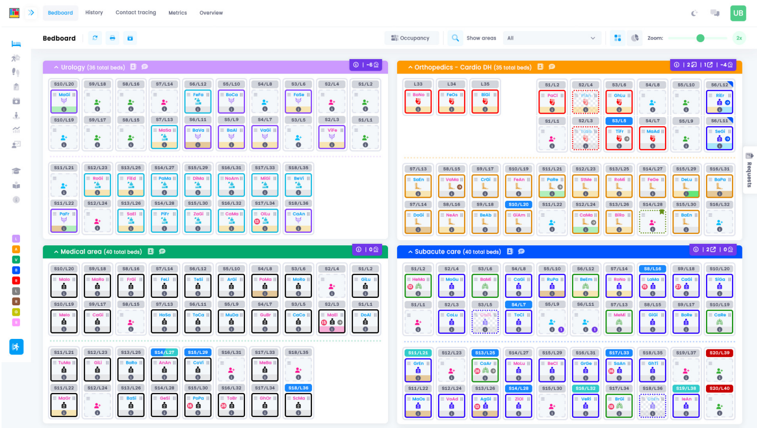


Figure 6.1: Map of Wards and Bed Locations.

Initial data population of the dashboard, along with subsequent updates, is ensured through integration with the Hospital Information System (HIS), leveraging HL7 interfaces connected to ER and ADT systems, as well as regional patient registries.

Each ward area displays one or more beds, color-coded by gender—blue for male, pink for female, and green for gender-neutral. For multi-bed rooms, gender is automatically determined by the first assigned patient —in compliance with single-sex room policies— and dynamically adjusted upon discharge events.

The application employs a comprehensive set of icons and chromatic elements to guide user attention, supplemented by explanatory tooltips and contextual information accessible through hover interactions, thereby eliminating the need for users to memorize visual conventions.

Occupied beds, represented by squares outlined in a color corresponding to the occupant's gender, encapsulate essential information within a compact visual format, including discipline-specific icons, patient initials, length-of-stay indicators (triggered upon exceeding ward-specific thresholds), and various status markers (Figure 6.2). Visual cues highlight specific conditions such as day hospital (DH) status, "bed blocker" designations with explanations for delayed discharge, and pending transfers to other wards or external facilities (*e.g.*, other healthcare institutions or elder care facilities).

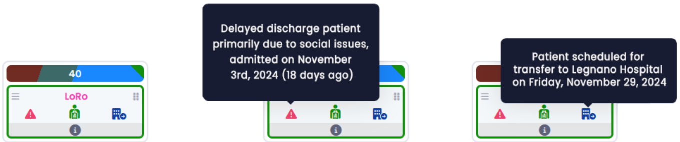


Figure 6.2: Color-Coded Bed Squares Showing Patient Data, Medical Specialty, and Transfer Status (Details).

Bed headers display both room and bed identifiers, with multiple dynamic chromatic variations to indicate specific alert conditions (Figure 6.3). Examples include infection status, isolation requirements, bedridden or end-of-life inpatients, and patients under short-stay observation (SSO). Likewise, the bed footer features a multiple color-coded bar signaling relevant information. For example, as a patient's anticipated discharge date approaches, the footer progressively intensifies in green, aiding rapid identification of imminent discharges.

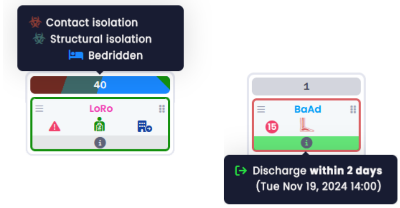


Figure 6.3: Color-Coded Bed Header and Footer Indicating Patient Risk Alerts and Discharge Disposition (Details).

A comprehensive overview of the patient's clinical status and conditions is accessible via a descriptive pop-up, triggered by briefly hovering over the bed or tapping it on touch-enabled devices (Figure 6.4). This supplementary view includes the inpatient's full name and its Medical Record Number (MRN), admission diagnosis, bed management annotations, requests for multidimensional assessment, immunization status, and details regarding admission from other healthcare facilities, districts, or regions.

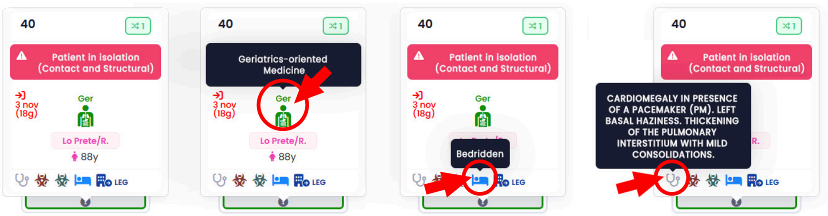


Figure 6.4: Patient Status Accessible via Hover or Tap (Details).

Unoccupied beds are depicted as rectangles containing a "+" symbol, indicating readiness for patient assignment. Bed allocation can occur automatically (via ADT system synchronization) or manually. In case of manual allocation, the assignment can be performed in one of two ways: by searching for the patient in the hospital registry and entering a minimal set of clinical information, or by selecting the patient from a specific worklist (Figures 6.5 and 6.6) and assigning them to the desired bed (either by dragging the patient directly onto the target bed or by specifying the bed manually).

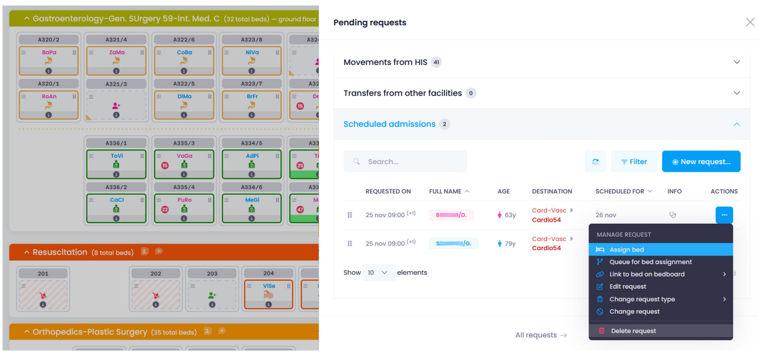


Figure 6.5: Patient Assignment from Scheduled Admissions List.

Manual allocation is commonly used for transfer requests, such as those originating from another facility or the emergency department.

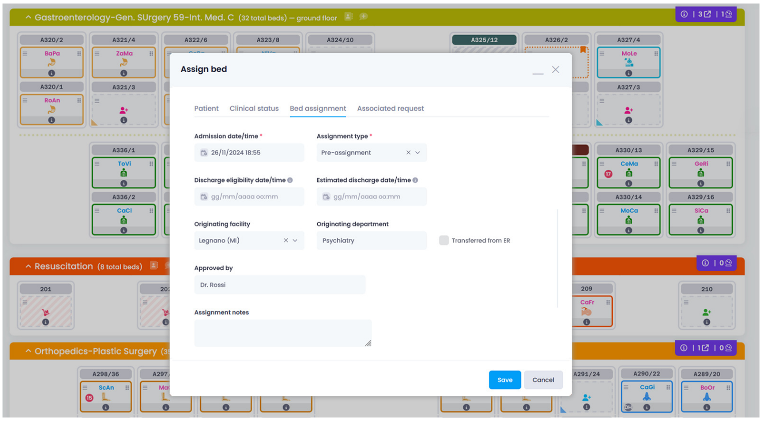


Figure 6.6: Patient Assignment Modal.

Similarly, transferring a patient to another bed within the same unit or to a different department is managed automatically through ADT integration, or by simply dragging the patient to the new location. Automated validation protocols support optimal patient placement, flagging potential incompatibilities and high-risk scenarios during both initial assignments and subsequent relocations. The system alerts against incompatible patient combinations, from mixed-gender rooms to placing "clean" individuals with those under isolation for infection control (Figure 6.7).

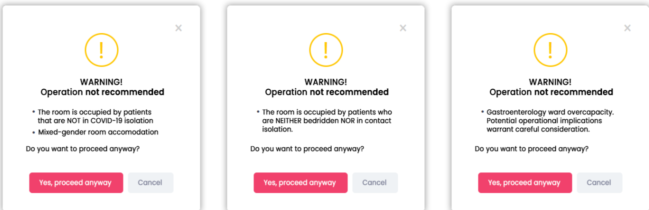


Figure 6.7: Automated Alerts for High-Risk Bed Assignments (Details).

This systematic approach minimizes reliance on arbitrary decisions, thereby reducing the likelihood of adverse events associated with subprime bed management assignments.

Each hospital ward is distinguished by a designated name and color code. The ward header displays essential summary metrics, including the number of occupied and available beds, the count of patients admitted from other units (referred to as "temporary in-area placements"), the number of patients formally under the ward's care but temporarily accommodated elsewhere (namely "temporary out-of-area placements"), and the total projected daily discharges (Figure 6.8). Detailed ward analytics are accessible through an interactive pop-up, allowing users to view occupied beds stratified by medical specialty, patient gender, and clinical risk assessment. Additionally, an always-available modal view presents occupancy levels across the entire hospital, organized by unit or specialty (Figure 6.9).

Figure 6.8: Ward Summary with Bed Occupancy Metrics (Details).

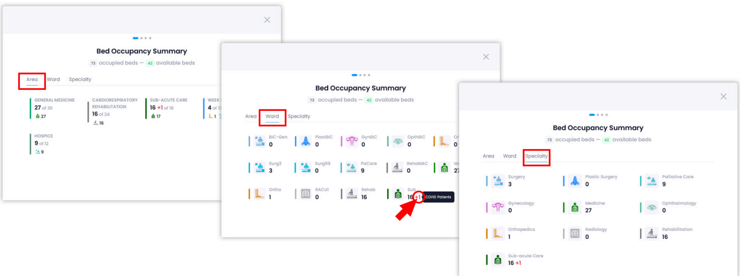


Figure 6.9: Occupancy Summary by Area, Ward, and Specialty (Details).

6.4.2 Key Functionalities and UI Design

The following section outlines the key functionalities implemented in accordance with stakeholder requirements. Many of these features are accessible through a contextual menu interface (Figure 6.10), allowing both individual bed-level and room-level interactions, thereby ensuring clear, immediate usability while optimizing the overall user experience.

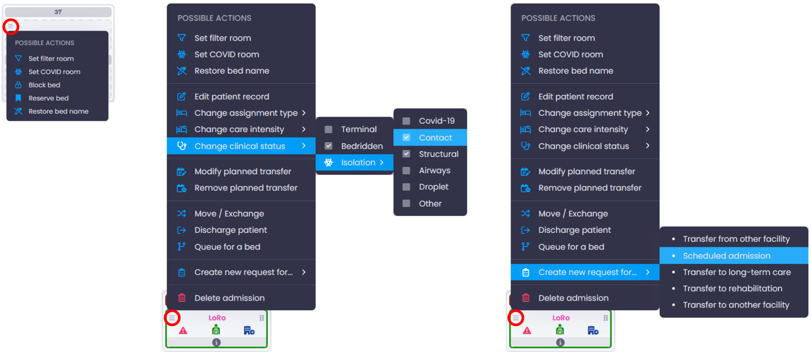


Figure 6.10: Quick Actions Contextual Menu (Details).

Blocking, Reserving, Pre-assigning, and Queueing

The system enables bed managers to lock, reserve, or pre-assign beds. Blocking a bed indicates its unavailability (Figure 6.11), which may be necessary for structural reasons —such as during periods of low patient influx (*e.g.*, in the summer), when certain wards may reduce bed availability to facilitate room renovations— or to designate the bed for caregivers or patient isolation requirements. Once blocked, a bed remains unassignable until its status is explicitly reversed by the bed manager.

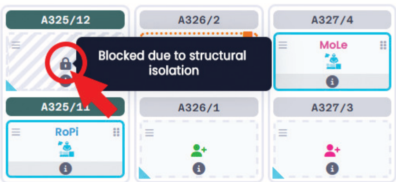


Figure 6.11: Blocked Beds Prevent Assignment.

Reservations apply when the future bed's occupant is not yet identified, but allocation is required for specific purposes (Figure 6.12), such as emergency room needs (in line with recent directives from the Italian Ministry of Health [186]) or SSO.

Figure 6.12: Bed Reservation for Future Allocation.

Pre-assignment, in contrast, serves to earmark a bed for patients whose admission, while not yet formalized, is anticipated within hours (for instance, patients in transit via ambulance from other healthcare facilities). In such cases, the bed is flagged as assigned but rendered in semi-transparency on the UI, indicating that it is not yet physically occupied (Figure 6.13). Once the patient is formally admitted, the pre-assignment automatically transitions to confirmed status, and the bed displays as fully occupied. If admission does not occur, the pre-assignment can be canceled.

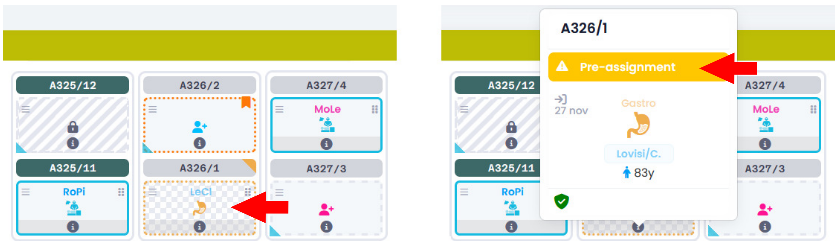


Figure 6.13: Pre-Assigned Bed Status for Anticipated Admissions.

The system also supports bed booking, allowing multiple patients to be queued for a currently occupied bed (Figure 6.14). Queueing proves particularly valuable for managing post-discharge bed availability, facilitating seamless reassignment to waiting patients. Upon confirmation of discharge and subsequent bed availability, the system automatically assigns the bed to the first patient in the queue. While patients may be queued for different beds to reflect various placement options, multiple queuing of the same patient for a single bed is not allowed.

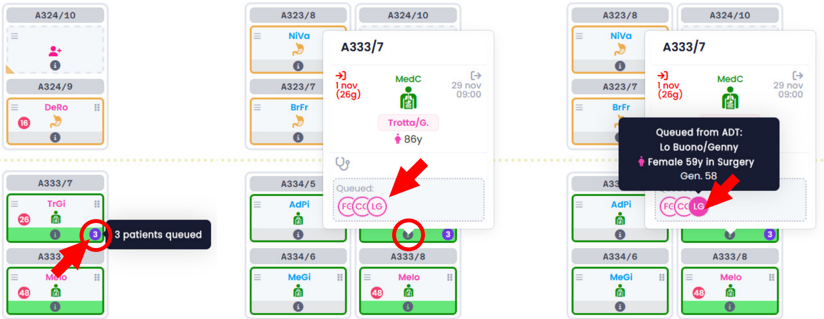


Figure 6.14: Bed Queuing for Efficient Post-Discharge.

Blocking, reserving, pre-assigning, and queueing beds are advanced features that empower bed managers to operate proactively, surpassing the capabilities of traditional ADT systems while providing real-time visibility into short-term occupancy trends. This transparency facilitates informed decision-making across departments, benefiting administration, ward coordinators, and emergency department physicians.

LOS Prediction and Patient Flow Forecast

The system equips bed managers with the ability to assign anticipated discharge dates to individual patients, enabling the dashboard to proactively highlight beds nearing availability as these dates approach. Integration with machine learning algorithms for LOS predictions further supports bed managers by suggesting the probability that a patient's stay will fall within one of three ranges: 1-3 days, 4-10 days, or more than 10 days.

The EBMS also includes a dedicated planning form to facilitate detailed admission and discharge forecasting over a 5-day horizon, categorized by department, discipline, and patient gender (Figure 6.15). Admission forecasts can be pre-populated by retrieving data from specific waitlists (*e.g.*, scheduled admissions) or manually adjusted by bed managers. Discharge projections are automatically inferred from the estimated discharge dates set for each patient. This functionality provides a valuable, albeit indicative, assessment of the anticipated balance between bed demand and supply for the upcoming days, segmented by department and, more granularly, by specialty within each department. A negative balance indicates a potential bed shortage, whereas a positive balance suggests bed availability.

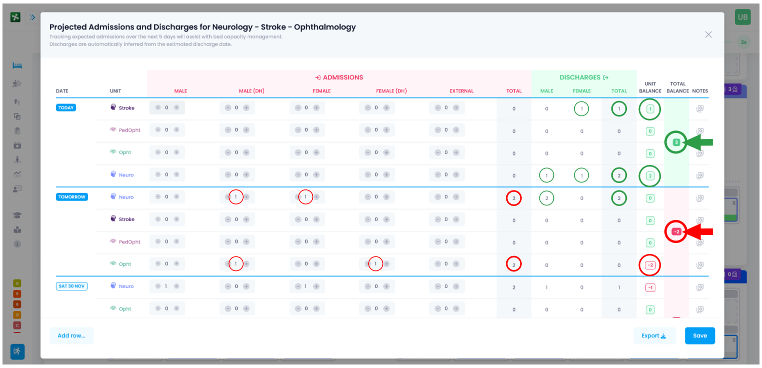


Figure 6.15: 5-Day Bed Capacity Planning and Demand Forecast.

Once planning is confirmed, an informative tooltip, accessible from the header of each department, provides a concise summary of patient flow (both admissions and discharges) for the current date and the subsequent two days, further detailed by gender (Figure 6.16).

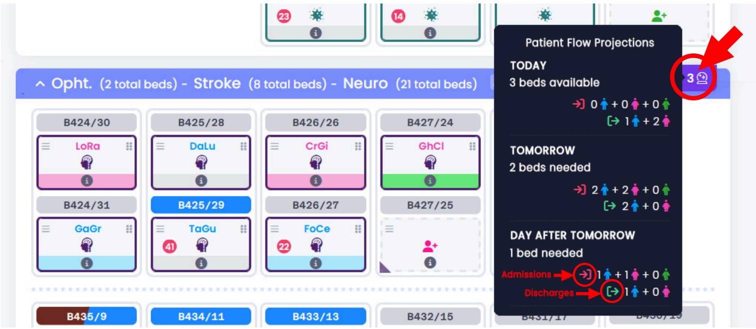


Figure 6.16: 3-Day Patient Flow Summary by Department.

Contact Tracing

Contact tracing in healthcare settings plays a fundamental role for infection control and prevention. By promptly identifying and isolating individuals exposed to infected patients, contact tracing helps to prevent further transmission, protect healthcare workers, and safeguard other patients. During the COVID-19 pandemic, contact tracing has exhibited exceptional utility in mapping transmission patterns, especially given the virus's potential for pre-symptomatic and asymptomatic spread. Implementation of robust contact tracing protocols has enabled healthcare institutions to sustain essential services and prevent the emergence of infection clusters within hospital units, while protecting vulnerable patient populations and preserving workforce capacity through early intervention strategies.

The developed Electronic Bed Management System includes comprehensive contact tracing functionality to support safer patient flow. The system tracks inpatient movements and enables detailed contact tracing for both currently admitted and recently discharged patients, including the identification of secondary-level contacts. Specifically, bed managers can access a list of *direct contacts* for a given patient —defined as individuals who shared the same room during the patient's stay, including intra-department and inter-department transfers— as well as a graph of *secondary contacts*, which represent patients who shared a room with the patient's direct contacts. In the tracing graph, the patient being tracked is displayed in bold and shaded light blue. Blue lines indicate connections to direct contacts, while edges in different colors represent relationships with secondary contacts (Figure 6.17).

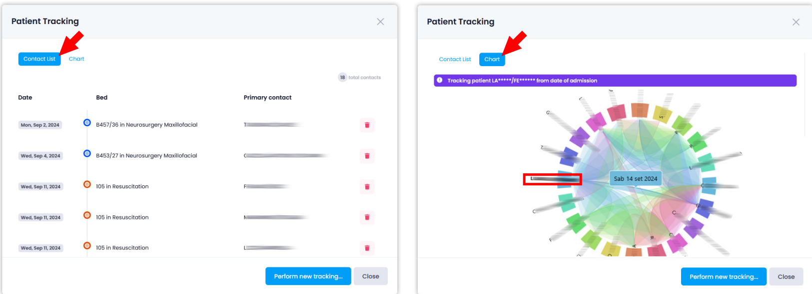


Figure 6.17: Patient Contact Tracing (Summary and Visualization).

Hovering over a connection between two patients reveals the date of their initial contact, providing a clear and intuitive visualization of exposure pathways.

Timeline

A major limitation of paper-based systems and spreadsheet-based solutions, as consistently emphasized during interviews with bed managers, lies in the difficulty of performing comparative analyses of bed allocations and assignments over different points in time (*e.g.*, the past three days). This complexity becomes increasingly pronounced as the temporal interval for comparison grows.

The implemented system addresses this challenge by enabling highly effective temporal analysis, extending the capability to examine arrangements weeks or even months in the past. Using automated processes, the EBMS captures comprehensive snapshots of hospital-wide bed status at four predetermined daily intervals (11:00, 14:00, 17:00, and 19:00). Additionally, the system accommodates manual snapshot generation, allowing bed managers to document specific situations of interest. Collectively, these sequential snapshots form a continuous temporal timeline. The UI enables retrospective examination through intuitive navigation tools: stakeholders may traverse the timeline using chronological scrolling or direct date selection via an integrated calendar.

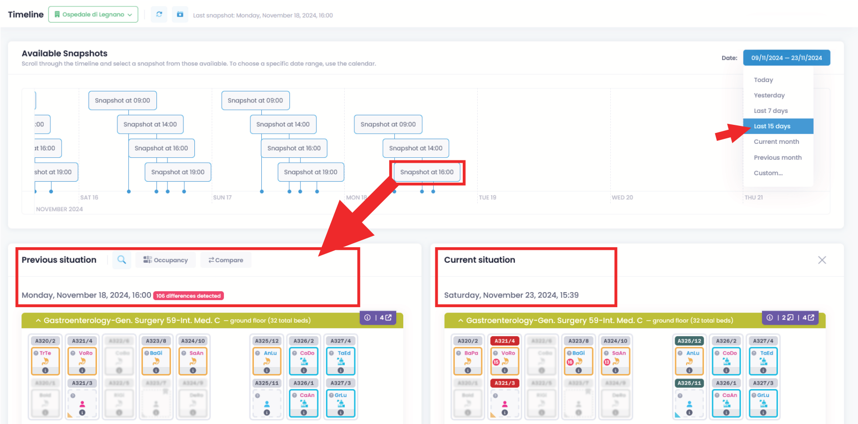


Figure 6.18: Comparative Timeline: Prior vs. Current State.

Specific historical configurations can be accessed through their corresponding snapshot, allowing users to search for individual patients, review summary statistics of bed occupancy for a given day, or compare historical and current configurations to identify differences (Figure 6.18). For instance, even if a patient's bed assignment remains unchanged, their clinical status may have evolved. As noted by a user during the testing phase, the functionality "effectively operates as a time machine."

The solution's analytical depth and temporal flexibility represent a significant advancement over traditional management methods, offering enhanced operational insights and administrative accountability. With its automated documentation and comparative capabilities, the system provides robust support for both operational decision-making and retrospective evaluations, including applications in dispute resolution and legal proceedings.

Copilot in Action

A key operational challenge faced by the hospital prior to the EBMS implementation involved bed allocation during the absence of bed managers, particularly during night shifts and weekends. Throughout these periods, Emergency Department physicians and unit coordinators had to rely solely on the most recent situation recorded by bed managers in spreadsheets and distributed via printed materials.

The EBMS introduces a sophisticated copilot functionality, enabling bed managers to activate an "Out of Office" (OoF) mode at the end of their shifts. This modality grants all platform stakeholders —including hospital administrators, ED residents, and ward coordinators— access to real-time bed status updates derived from the ADT system, with a maximum latency of 15 minutes. While OoF mode is active, the dashboard automatically synchronizes with ADT-reported data, ensuring continuity and coherence in bed occupancy visualization. During this interval, manual adjustments to bed allocations (such as transfers, discharges, or clinical data updates) are restricted until the copilot mode is deactivated.

Importantly, any additional patient information previously documented by the bed management team (*e.g.*, isolation requirements for infectious diseases) is retained, including pre-allocations or queued patients. However, beds marked as blocked will automatically be unblocked if the ADT system reports them as assigned to a patient.

The authority to activate or deactivate the "Out of Office" mode is reserved exclusively for bed managers within the central management team.

This functionality greatly enhances confidence among hospital staff, who can rely on a continuously updated dashboard even in the absence of bed managers, thereby reducing the need for interdepartmental communications or on-call staff consultations. Furthermore, it eliminates the historical administrative burden on bed managers, who previously needed to reconcile their spreadsheets with any changes occurred during their absence.

Inter-Facility Transfer Management

A core responsibility of the bed manager is to coordinate both internal hospital admission requests (such as scheduled procedures) and external facility transfers (both incoming and outgoing). Previously managed through individual spreadsheets, this administrative process has been integrated into the EBMS, facilitating comprehensive tracking of:

- Transfers from external facilities
- Scheduled admissions
- Long-term care transfers
- Rehabilitation transfers
- Transfers to other facilities

Each new request is linked to a specific patient, identified through the regional registry, and includes relevant clinical and social contextual information. Submitted requests populate a centralized list from which bed managers can allocate patients to available beds at the appropriate time. Requests pending beyond a predetermined number of days are flagged in red, with the waiting period visibly marked to draw the bed manager's attention to potential delays (Figure 6.19).

Prior to MondrIAn's implementation, the five hospitals involved in the pilot study relied on email communication for patient transfer requests, using manually completed paper forms that were scanned and attached. Stakeholders widely recognized the need for an integrated system to manage inpatient transfer requests, particularly for patients awaiting rehabilitation programs or long-term care placement.

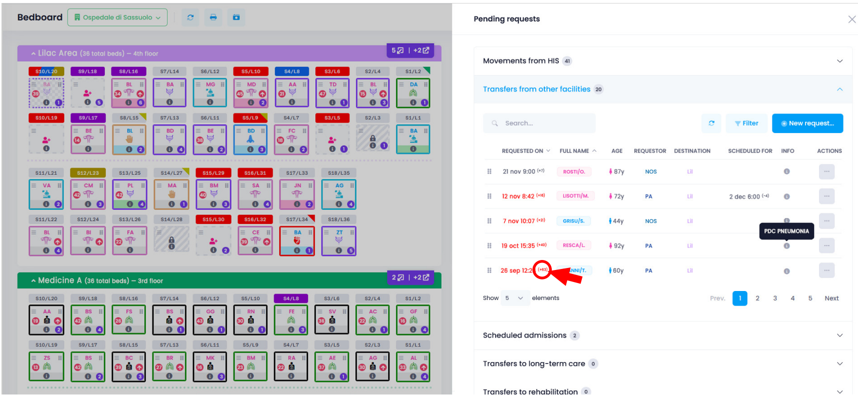


Figure 6.19: Centralized Request Management with Delay Monitoring (in Days).

The implemented system fosters collaborative management of patient requests. For local public health authorities or multi-site healthcare organizations, it functions as an effective tool for coordinating patient transfers and temporary placements across facilities. When a patient is transferred, the system automatically updates bed occupancy status by releasing the bed in the originating facility and assigning a new bed in the receiving facility. Real-time updates are reflected on each hospital's dashboard, ensuring that all involved bed managers have access to up-to-date information. Details of the transferred patient remain accessible through the "Discharge/Transfer History" page of the originating facility for reference.

In scenarios where a patient is transferred without immediate bed assignment at the destination facility, the system releases the originating bed while adding the patient to a waitlist for the destination facility. The receiving bed manager can then manually assign the patient to a bed as soon as one becomes available. For patients transferred without prior bed assignment in the originating facility, the system still records the transfer accurately, though the patient's record will only be accessible on the "Discharge/Transfer History" page of the receiving facility.

The standardization of inter-facility communication protocols represents a progression toward integrated bed management practices, establishing a cohesive framework for cross-institutional collaboration.

Real-Time Metrics

The system autonomously collects and analyzes data on a daily basis to generate key periodic indicators, including average and median length of stay, bed occupancy rate, turnover interval, bed blockers ratio, percentage of urgent cases discharged within four days, and the proportion of patients originating from intensive or semi-intensive care units. Additionally, it monitors variations in infection rates (Figure 6.20).

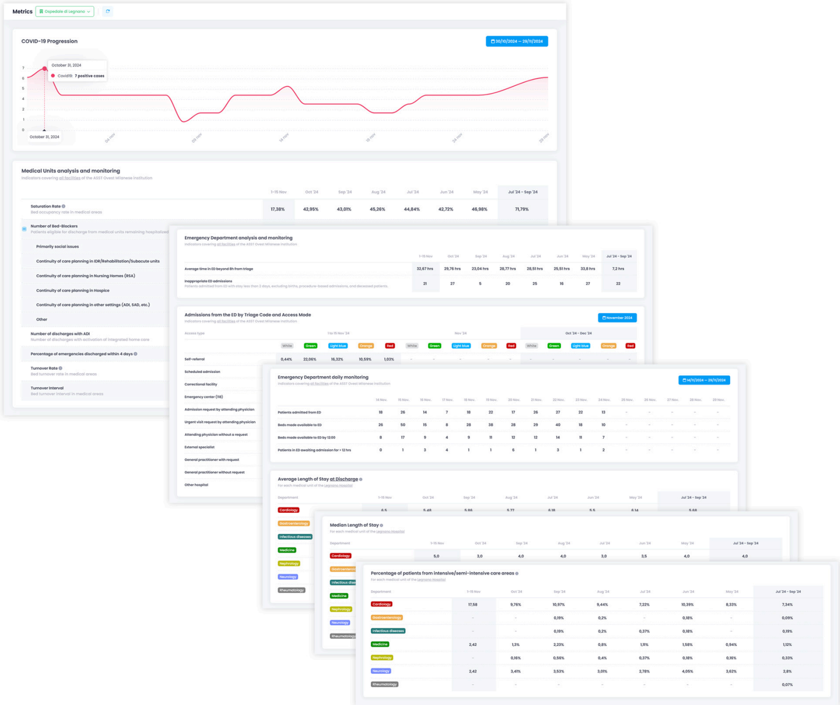


Figure 6.20: Automated Key Performance Indicators and Infection Monitoring.

All collected data is automatically processed for statistical reporting and integration into regional data flows. Occupancy reports are distributed to designated mailing lists at pre-set intervals, including non-working days, thereby reducing the need for on-call staff involvement in data monitoring.

Hospital management benefits from access to a high-impact dashboard connected to all organizational facilities and updated in real time (Figure 6.21), with the option for projection on large displays stationed in strategic locations. This interface provides a comprehensive and easily interpretable overview, supporting detailed analysis ranging from macro-level organizational trends to facility-specific and departmental metrics.

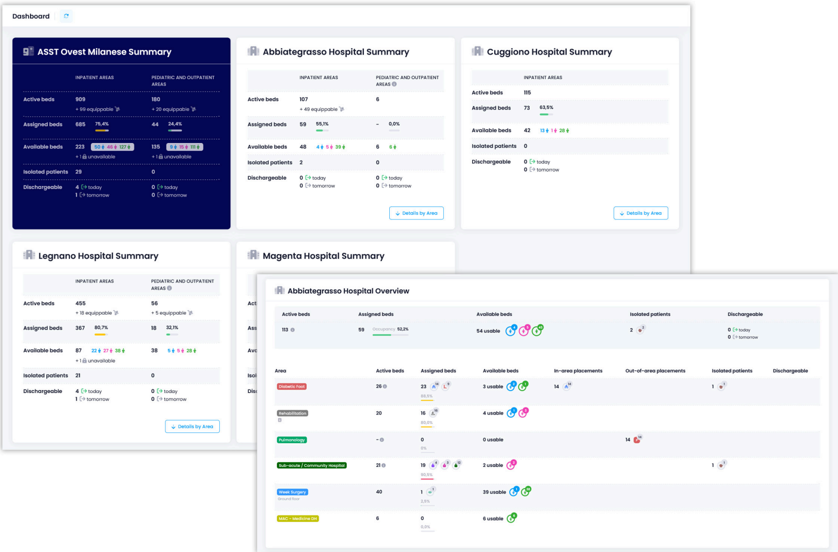


Figure 6.21: Hospital-Wide Real-Time Analytics Dashboard Displaying Multi-Facility Metrics.

Given the importance of inter-hospital collaboration and shared metrics for improving regional coordination in patient care, as highlighted by the COVID-19 pandemic, this global interactive dashboard stimulates strategic cooperation, bridging communication gaps that can severely affect patient admissions, transfer times, and clinical outcomes.

User Testing and Iterative Development

The deployment of the system in the production environment (detailed in Appendix D) was preceded by a "live-testing" phase during which bed managers were shadowed for one week, working in parallel with one of the authors. Over the first two days, the author observed the actions performed by the bed manager on the original spreadsheet and replicated them within the EBMS. During the subsequent three days, roles were reversed: the bed manager operated the EBMS while the author mirrored their actions on the spreadsheet, providing implementation support as needed. The purpose of this approach was twofold: to identify any system components not performing as intended, and to assess the efficacy of the UI through repeated simulation of bed managers' daily tasks. Users were also encouraged to share their perceptions of the tool's utility and usability during use.

Following any platform modifications, functionality tests were conducted in a controlled environment using delayed, real patient data. System behaviour was further monitored through periodic evaluations in both test and production environments, allowing for the resolution of any remaining issues. On-site, real-time feedback from clinical staff was instrumental in refining the EBMS.

The development process adhered to Continuous Integration/Continuous Delivery (CI/CD) principles, facilitating streamlined code integration and deployment. Continuous Integration (CI) involves the automated and frequent merging of code changes into a shared repository, while Continuous Delivery/Deployment (CD) consists of a structured process for integration, testing, and delivery. This methodology enabled rapid features iteration and efficient incorporation of user feedback.

Bi-weekly releases were communicated to bed managers via email, utilizing clear and accessible language supported by visual documentation to ensure ease of understanding.

Results

A retrospective before–after comparison was carried out to evaluate hospital performance indicators through the examination of electronic records following the adoption of the CEBMS. The experimental phase, initiated in March 2021 at Ospedale di Sassuolo S.p.A., culminated in the first software release in mid-March 2022. The pilot project was subsequently expanded to the remaining four secondary hospitals within the province of Modena (Carpi, Mirandola, Pavullo, Vignola).

The assessment methodology incorporated a concise, targeted questionnaire administered to two bed managers per facility, each having regularly operated the EBMS for a minimum of twelve months. The survey invited users to share their experiences with the system, including the ease of use of the platform, the accuracy of data collected, and the overall helpfulness of the implemented features.

In response to the question, "On a scale from 1 to 5, how easy is it to use the software? (1 = very difficult, 5 = very easy)", 60% of respondents rated it as "Very easy", while 40% found it "Quite easy." These results validate the investment in designing an intuitive interface, particularly considering the significant paradigm shift from previous management methods and the varying levels of technological proficiency among primary stakeholders.

When asked, "Has adopting the software affected your typical workday?" 80% answered "Significantly improved", 10% "Slightly improved", and the remaining 10% "Unchanged." To the question, "Can you estimate the amount of time saved daily (on average) due to the software use?" 40% reported "1–2 hours", 20% "More than 2 hours", 10% "Up to 30 minutes", while the remaining 30% indicated "Not sure." Excluding the "Not sure" responses and averaging across the remaining 70% —using midpoint estimates for time ranges (*e.g.*, 90 minutes for "1–2 hours" and 15 minutes for "Up to 30 minutes") and assuming 2.5 hours for "More than 2 hours"— the average time saved amounted to approximately 1.125 hours per day. With a 5-day workweek totaling 40 hours, this equates to an average weekly saving of about 5 hours and 40 minutes (22 hours per month), or approximately 14% increase in time efficiency.

On the question "Do you believe the software adoption has impacted interdepartmental communication?" 50% answered "Significantly improved", 30% "Slightly improved", and 20% indicated no change.

In response to "On a scale from 1 to 5, to what extent do you think the software could help in anticipating discharges? (1 = no help, 5 = significant help)", 50% indicated "Slight help", 20% "Significant help", and 30% reported "No help."

Finally, when asked "Have you encountered any bed allocation errors in the past month?" 40% answered "None", another 40% reported "1–5", and the remaining 20% indicated "10–15". Given the considerable monthly discharge volume (averaging 1,175 discharges per hospital), these results highlight the system's high accuracy in managing bed assignments.

The most commonly recognized tangible benefits included "Simplifies and accelerates bed allocation" (100%), "Increases visibility and awareness of departments status" (100%), "Reduces information retrieval time" (90%), "Facilitates and improves discharge planning" (50%), and "Reduces unnecessary phone calls" (30%). Most valued features comprised "Visual bed dashboard" (100%), "Contact tracing" (90%), "Real-time metrics monitoring" (80%), "Planned discharge management" (70%), and "Bed-map history consultation (timeline)" (60%).

Building on the question regarding the software's potential to assist in anticipating patient discharges, the study analyzed the average LOS (ALOS) across participating facilities over two years following the experimental phase (data officially submitted to the Regione Emilia-Romagna [187]). Table 6.1 presents the average hospital stay, calculated as the total number of bed days divided by total admissions, for the years 2022 to 2024 across participating facilities. The temporal analysis benefits from the Italian Council of Ministers' termination of the COVID-19 emergency status on March 31, 2022 [188] (initially declared January 31, 2020), thus eliminating pandemic-related confounding variables from the dataset.

Table 6.1: Average Length of Stay (ALOS) in Days.

Hospital Facility	**2022**	**2023**	**2024***
Ospedale Sassuolo	5.8	5.9	5.8
Ospedale Carpi	6.8	6.8	6.4
Ospedale Mirandola	8.8	8.7	8.4
Ospedale Pavullo	7.7	7.3	7.4
Ospedale Vignola	7.3	7.1	6.7

* Data up to October 2024.

A comparison of data from 2022 with subsequent years reveals an overall reduction in the average LOS. Using an estimated daily inpatient cost of €592 [189], the projected annual savings based on 2023 discharge data (Table 6.2) amount to approximately €243,164.

Table 6.2: Annual Discharges.

Hospital Facility	**2022**	**2023**	**2024***
Ospedale Sassuolo	15,330	15,894	10,204
Ospedale Carpi	10,927	11,351	7,767
Ospedale Mirandola	4,182	4,486	2,940
Ospedale Pavullo	3,688	3,640	2,423
Ospedale Vignola	3,714	4,027	2,827

* Data up to October 2024.

This estimate is conservative, as it does not account for potential fixed or variable costs unique to each facility, omits possible variations in case complexity or departmental units (*e.g.*, ICU beds may incur costs 3-4 times higher than standard beds), and assumes a uniform distribution of LOS reductions across all patients.

Additionally, the system implementation yielded a significant reduction in processing times for inter-facility transfer requests among the pilot study institutions.

6.5 Discussion

The adoption of Electronic Bed Management Systems represents a recent advancement over traditional methods such as paper records, spreadsheets, and physical whiteboards, which are susceptible to transcription errors, data loss, and restricted access.

Previous studies have explored the utility of EBMSs in healthcare settings. For example, a cloud-based EBMS implemented by Gauteng health services [190] in South Africa employed IoT sensors to monitor bed occupancy, allowing staff to determine bed availability. Noonan *et al.* [179] collaborated with an Irish hospital to deploy an EBMS that provided a comprehensive view of patient and bed statuses, enabling data analysis to enhance patient flow. However, these systems primarily focused on single-hospital applications, lacking a collaborative design for inter-hospital integration. The EBMS proposed in this study addresses this gap by offering real-time insights into bed availability and infection status across multiple sites, fostering cooperative bed management across geographically distributed teams. The system facilitates resource coordination within individual settings and across healthcare networks, allowing clinical personnel to prioritize patient care over administrative tasks. The implementation of comprehensive request management functionalities expedites patient handoffs between hospitals, increasing the efficiency of inter-facility care transitions.

Additionally, the CEBMS improves communication and clinical handover processes, which are fundamental to effective bed management [181]. An embedded messaging application favors direct and immediate interactions, accommodating both synchronous and asynchronous exchanges while ensuring communication occurs over a secure, dedicated channel.

MondrIAn's integration into the discharge planning process proved remarkably effective. The platform aids team members in quickly identifying discharge-ready patients, promoting timely preparation and alleviating hospital congestion. Moreover, MondrIAn incorporates machine learning models to forecast length of stay across three categories (1–3 days, 4–10 days, or more than 10 days), augmenting the accuracy of anticipated discharge dates and supporting better resource planning. This is particularly valuable, as research indicates that advancing discharge times by even one hour can significantly reduce overcrowding and access delays [191].

The developed CEBMS seamlessly integrates with existing ADT and ED systems through globally accepted medical IT standards, such as HL7 messaging protocols, extending rather than superseding their functionalities [192].

This study identifies several key advantages. First, it introduces an innovative framework for real-time communication to core stakeholders regarding bed occupancy and discharge readiness. Effective communication, as emphasized by the WHO International Patient Safety Goals, is a critical factor in addressing the complexities of bed management [181]. Poor communication has consistently been identified as a significant contributor to prolonged LOS and operational inefficiencies [193]. Second, to the best of our knowledge, this study is the first to advocate for collaborative bed management across a network of healthcare facilities, providing open shared metrics that have proven invaluable during recent regional emergencies. Such an approach has the potential to alleviate surgery postponements and expedite bed turnover.

The cooperative nature of the proposed EBMS also presents a compelling opportunity to leverage Federated Learning techniques. By enabling each participating healthcare facility to periodically retrain LOS prediction models using local data —incorporating actual discharge dates from the Hospital Discharge Register (Scheda di Dimissione Ospedaliera, SDO) as automatic labels upon patient release— the system engenders a collaborative learning environment. This strategy ensures that models remain relevant across the network of hospitals over time by mitigating data drift through continuous updates, while maintaining strict compliance with data protection regulations by sharing model weights instead of sensitive patient data.

The provincial-scale adoption of the system facilitated global monitoring of performance indicators and alignment with institutional goals. Notable improvements following MondrIAn's implementation include enhanced visibility for ED physicians, who traditionally lacked access to ADT systems, and greater operational transparency through strategically placed visual dashboards across healthcare facilities. Additionally, the system's capability to swiftly reconstruct direct patient contacts has strengthened infection control protocols by enabling prompt implementation of protective measures.

Despite its promising outcomes, this study acknowledges certain limitations. Continuous optimization during the pilot phase may have delayed the full realization of benefits. Moreover, while reductions in overall average LOS were observed, the variability in impact across departments was not extensively analyzed. Finally, the potential influence of other organizational or external factors on observed improvements also warrants consideration. Nevertheless, the system's scalability and adaptability were validated through its further adoption by other several public hospitals in 2023, corroborating its efficacy in diverse healthcare settings.

6.6 Final Remarks

Hospital bed management presents complex operational challenges influenced by multifaceted constraints, including resource limitations, fluctuating admission rates —particularly in emergency and readmission scenarios— and varying levels of care intensity across ordinary, scheduled, and emergency cases. The COVID-19 pandemic has exposed the inadequacies of traditional planning methodologies in rapidly evolving healthcare environments, underscoring the necessity for more adaptive strategies.

Ineffective patient flow management can cause serious disruptions, resulting in prolonged emergency department wait times and ward overcrowding. These circumstances may lead to overstretched staff and services, an increased likelihood of errors, and adverse effects on employees and patients. Furthermore, unnecessary hospital stays generate elevated operational costs while exposing patients to hospital-acquired infections, depression, or physical deconditioning.

Electronic Bed Management Systems have emerged as practical tools, offering real-time tracking of admissions, discharges, and patient pathways. Through digital management of bed supply and demand, these systems facilitate streamlined communication among multidisciplinary teams and ensure coordinated care delivery.

This research introduces the first Cooperative EBMS (CEBMS), designed for collaborative bed allocation across multiple healthcare facilities within a public network. By leveraging a unified platform, hospitals can efficiently transfer patients between facilities while simplifying information management. Through consolidated metrics on bed occupancy, availability, pending requests, and infection status, bed management teams and hospital administrators can make informed decisions, ensuring that patients receive appropriate care in a timely manner and are seamlessly transitioned throughout the healthcare system.

Provincial-level adoption and consistent utilization of the CEBMS have yielded improvements in both quantitative and qualitative performance indicators, including reduced ALOS and enhanced inter-departmental and inter-facility communication. Manual processes previously reliant on paper records or spreadsheets have been automated or semi-automated, requiring minimal time and improving efficiency and safety. Advanced ML techniques further support predictive modeling for discharge planning and optimal bed allocation across interconnected facilities.

This integrated approach to bed management offers a robust model for contemporary healthcare systems worldwide. By maximizing resource usage while maintaining high standards of care, hospitals are better equipped to confront the growing complexities of modern healthcare demands.

Implementation of the MondrIAn platform has significantly improved transparency in patient journey, contributing to digital transformation and operational sustainability across provincial healthcare networks. As of 2023, the system has expanded to include six additional public hospitals in northern Italy: four within ASST Ovest Milanese (Legnano, Magenta, Cuggiono, Abbiategrasso), two within ASST Bergamo Ovest (Treviglio and

Romano), and the ASST Ospedale Metropolitano Niguarda in Milan.

Chapter 7

Conclusion

The evolution of public healthcare systems represents one of the most significant achievements in modern society. The Italian healthcare model, in particular, has effectively supported the country's high life expectancy — 83 years in 2022 [194], making it the third highest in the European Union— while operating with a comparatively lower *per capita* expenditure than the EU average. At the same time, maintaining this balance has become increasingly difficult due to mounting pressures from an aging population, a rising chronic disease burden, and resource constraints. With health spending at 9.4% of GDP —compared to the EU average of 11% [194]— concerns about long-term sustainability are intensifying.

As digital transformation (DT) in public healthcare emerged globally as both an opportunity and a challenge during the COVID-19 pandemic, which exposed the sector's vulnerability to unprecedented demand, many countries accelerated digital adoption in response. Italy, however, continues to face structural barriers, including limited public funding, regional disparities in digital infrastructure, and varying levels of digital literacy among healthcare professionals and patients, all of which compromise the efficacy of DT initiatives and amplify existing disparities in healthcare access.

This thesis explored how DT and machine learning (ML) can address critical challenges within Italian public healthcare. Following an introductory discussion of DT fundamentals and complexities in Chapter 1, we presented several practical implementations that were developed and validated through partnerships with regional public healthcare institutions, demonstrating the potential of new technologies to reshape healthcare delivery.

Chapter 2 introduced a neural network for the automated detection of SARS-CoV-2 test outcomes from self-administered swabs, with results made accessible via a RESTful API for integration with the Electronic Health Record system (FSE). Though seemingly distinct from the other chapters, this preliminary investigation exemplified how AI could accelerate emergency interventions and alleviate the burden on healthcare personnel, aligning with the broader research objective of leveraging ML to enhance operational efficiency and diagnostic precision in healthcare.

In Chapter 3, we developed ML models to predict hospital length of stay using both structured and unstructured data. The successful incorporation of these models into existing Electronic Bed Management Systems (EBMSs) illustrated the practical applicability of ML solutions in daily hospital workflows. Building on this, Chapter 4 expanded the research to forecast prolonged LOS in Emergency Departments, tackling the critical issue of overcrowding that affects both emergency-urgency services and inpatient wards.

Chapter 5 introduced MondrIAn, the first Cooperative EBMS, which represents a remarkable step toward collaborative bed management across healthcare facilities. By combining real-time territorial bed occupancy data with predictive ML models from Chapter 3, the system promotes inter-hospital collaboration and expedites patient transfers. Its adoption across multiple hospitals in northern Italy has yielded measurable improvements in operational standards and resource utilization, setting a benchmark for regional healthcare systems.

Collectively, our implementations have shown particular promise in automating routine tasks to alleviate administrative workloads, optimizing resource allocation through predictive analytics, simplifying inter-facility collaboration and communication, and supporting data-driven decision-making in clinical and administrative contexts.

Despite these positive outcomes, this work faces several limitations that suggest avenues for further research. While the developed models were trained on robust datasets, variations in documentation standards across different hospitals and administrative regions may limit their generalizability. Broader multi-center collaborations would help validate and refine the proposed approaches, minimizing biases introduced by local idiosyncrasies and protocols. Future efforts should also prioritize the development of inter-regional digital collaboration frameworks to promote nationwide integration and interoperability.

The decision to focus on early prediction necessitated the exclusion of certain data streams, such as vital signs and laboratory results, which were less readily accessible. Nonetheless, even considering their delayed or inconsistent availability (potentially leading to missing data or bias), incorporating these variables could still improve model accuracy, particularly through time-series approaches that capture dynamic patient status during hospitalization.

Although ensemble methods yielded reasonable predictive performance in forecasting LOS and pLOS, their inherent complexity could hinder clinical interpretation. Existing explainability tools have been applied only on a limited scale, underscoring the need for advanced explainable AI (XAI) techniques —such as SHAP or integrated gradients— to elucidate algorithmic decisions and strengthen clinical trust. Probabilistic outputs should also be presented in formats familiar to clinicians (*e.g.*, confidence intervals or clearly defined risk thresholds) and could be supplemented with real-time natural language explanations (*e.g.*, by leveraging Local Large Language Models) to facilitate the translation of model findings into actionable insights.

Another important consideration is that rapid shifts in healthcare needs —influenced by emergent diseases, regulatory changes, or demographic trends— may necessitate frequent retraining or adjustment of machine learning models. Continual learning methodologies, which empower systems to autonomously label new data, offer a promising strategy for regular updates to help maintain relevance and accuracy over time. The incorporation of federated learning techniques could further improve performance while preserving patient privacy by enabling the exchange of model parameters without centralizing sensitive information. The consistent application of anonymization and pseudonymization practices remains paramount for safeguarding patient confidentiality throughout this process.

From an operational standpoint, although preliminary adoption of ML-based insights in EBMSs demonstrated technical feasibility, large-scale deployment would require more extensive acceptance testing, staff education, and comprehensive real-world impact assessments to address the practical and ethical challenges inherent in clinical settings.

Finally, two additional experiments worthy of investigation include predicting in-hospital mortality and readmission rates. Precise forecasting of in-hospital mortality could assist in timely interventions and resource allocation for critically ill patients, potentially improving survival outcomes. Similarly, projecting readmission risk within a defined timeframe could enable proactive post-discharge care management, reducing the likelihood of recurrent hospitalizations and associated healthcare costs. These additions would contribute to a more holistic approach to patient flow management and enhance the clinical utility of the proposed framework.

Extending the scope of these AI-driven solutions beyond hospital-based resource management to areas such as preventive care and early-stage disease risk prediction holds further potential to alleviate high-acuity admissions and lessen the load on inpatient and emergency services.

Italy's Recovery and Resilience Plan, allocating €16.1 billion for healthcare system improvements by 2026, positions digital transformation as a critical pathway toward securing the performance, sustainability, and resilience of the Italian public healthcare system, aiming to preserve societal welfare over time. The strategic deployment of AI and emerging technologies will be instrumental in achieving these objectives.

Publications and Achievements

The research presented in this thesis has led to publications in international conferences.

Notably, the work on MondrIAn received the Best Scientific Communication award at the "47th National Congress of ANMDO (Italian National Association of Hospital Management)", held in Bologna, Italy, in June 2022. A comprehensive list of my publications is provided in Appendix A. Many of these papers were co-authored with fellow Ph.D. students and researchers from the department where I conducted my research. I am deeply grateful for their invaluable collaboration and the work we accomplished together.

Bibliography

[1] WHO, "Coronavirus disease (covid-19) pandemic. health topics.," 2019. https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [Accessed: (2024/10/15)]. 9

[2] M. K. Barai and S. Dhar, "Covid-19 pandemic: Inflicted costs and some emerging global issues," *Global Business Review*, vol. 25, no. 3, pp. 812–831, 2024. 9

[3] M. Ikram and Y. Sayagh, "The consequences of covid-19 disruption on sustainable economy in the top 30 high-tech innovative countries," *Global Journal of Flexible Systems Management*, vol. 24, no. 2, pp. 247–269, 2023. 9

[4] P. Malliet, F. Reynès, G. Landa, M. Hamdi-Cherif, and A. Saussay, "Assessing short-term and long-term economic and environmental effects of the covid-19 crisis in france," *Environmental and Resource Economics*, vol. 76, no. 4, pp. 867–883, 2020. 9

[5] A. D. Kaye, C. N. Okeagu, A. D. Pham, R. A. Silva, J. J. Hurley, B. L. Arron, N. Sarfraz, H. N. Lee, G. E. Ghali, J. W. Gamble, *et al.*, "Economic impact of covid-19 pandemic on healthcare facilities and systems: International perspectives," *Best Practice & Research Clinical Anaesthesiology*, vol. 35, no. 3, pp. 293–306, 2021. 9

[6] S. Barello, L. Palamenghi, and G. Graffigna, "Burnout and somatic symptoms among frontline healthcare professionals at the peak of the italian covid-19 pandemic.," *Psychiatry research*, vol. 290, p. 113129, 2020. 9

[7] H. Leite, C. Lindsay, and M. Kumar, "Covid-19 outbreak: implications on healthcare operations," *The TQM Journal*, vol. 33, no. 1, pp. 247–256, 2020. 9

[8] L. Cobianchi, F. Dal Mas, A. Peloso, L. Pugliese, M. Massaro, C. Bagnoli, and P. Angelos, "Planning the full recovery phase: an antifragile perspective on surgery after covid-19," *Annals of Surgery*, vol. 272, no. 6, pp. e296–e299, 2020. 9

[9] D. Golinelli, E. Boetto, G. Carullo, A. G. Nuzzolese, M. P. Landini, and M. P. Fantini, "Adoption of digital technologies in health care during the covid-19 pandemic: systematic review of early scientific literature," *Journal of medical Internet research*, vol. 22, no. 11, p. e22280, 2020. 9

[10] S. L. Connolly, E. Kuhn, K. Possemato, and J. Torous, "Digital clinics and mobile technology implementation for mental health care," *Current psychiatry reports*, vol. 23, no. 7, p. 38, 2021. 9

[11] S. Ghosh, M. Hughes, I. Hodgkinson, and P. Hughes, "Digital transformation of industrial businesses: A dynamic capability approach," *Technovation*, vol. 113, p. 102414, 2022. 10

[12] E. Parliament, "Regulation (eu) 2021/694 of the european parliament and of the council of 29 april 2021 establishing the digital europe programme and repealing decision (eu) 2015/2240," 2021. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32021R0694 [Accessed: (2024/10/15)]. 10

[13] I. Iyamu, A. X. Xu, O. Gómez-Ramírez, A. Ablona, H.-J. Chang, G. Mckee, M. Gilbert, *et al.*, "Defining digital public health and the role of digitization, digitalization, and digital transformation: scoping review," *JMIR public health and surveillance*, vol. 7, no. 11, p. e30399, 2021. 11, 16

[14] C. Drago, A. Gatto, and M. Ruggeri, "Telemedicine as technoinnovation to tackle covid-19: A bibliometric analysis," *Technovation*, vol. 120, p. 102417, 2023. 11

[15] P. Galetsi, K. Katsaliaki, and S. Kumar, "Values, challenges and future directions of big data analytics in healthcare: A systematic review," *Social science & medicine*, vol. 241, p. 112533, 2019. 11

[16] Z. Wang, N. Luo, and P. Zhou, "Guardhealth: Blockchain empowered secure data management and graph convolutional network enabled anomaly detection in smart healthcare," *Journal of Parallel and Distributed Computing*, vol. 142, pp. 1–12, 2020. 11

[17] A. I. Stoumpos, F. Kitsios, and M. A. Talias, "Digital transformation in healthcare: technology acceptance and its applications," *International journal of environmental research and public health*, vol. 20, no. 4, p. 3407, 2023. 11

[18] C. Guo and J. Chen, "Big data analytics in healthcare," in *Knowledge technology and systems: Toward establishing knowledge systems science*, pp. 27–70, Springer, 2023. 12

[19] Z. Dlamini, F. Z. Francies, R. Hull, and R. Marima, "Artificial intelligence (ai) and big data in cancer and precision oncology," *Computational and structural biotechnology journal*, vol. 18, pp. 2300–2311, 2020. 12

[20] R. Bernardi and M. Exworthy, "Clinical managers' identity at the crossroad of multiple institutional logics in it innovation: The case study of a health care organization in england," *Information Systems Journal*, vol. 30, no. 3, pp. 566–595, 2020. 12

[21] K. Schwab, *The fourth industrial revolution.* Crown Currency, 2017. 12

[22] G. Ford, M. Compton, G. Millett, and A. Tzortzis, "The role of digital disruption in healthcare service innovation," *Service business model innovation in healthcare and hospital management: Models, strategies, tools*, pp. 57–70, 2017. 12

[23] F. Dal Mas, M. Massaro, P. Rippa, and G. Secundo, "The challenges of digital transformation in healthcare: An interdisciplinary literature review, framework, and future research agenda," *Technovation*, vol. 123, p. 102716, 2023. 12

[24] P. Apell and H. Eriksson, "Artificial intelligence (ai) healthcare technology innovations: the current state and challenges from a life science industry perspective," *Technology Analysis & Strategic Management*, vol. 35, no. 2, pp. 179–193, 2023. 12

[25] S. Y. Guraya, "Transforming simulation in healthcare to enhance interprofessional collaboration leveraging big data analytics and artificial intelligence," *BMC Medical Education*, vol. 24, no. 1, p. 941, 2024. 13

[26] V. Jahmunah, V. K. Sudarshan, S. L. Oh, R. Gururajan, R. Gururajan, X. Zhou, X. Tao, O. Faust, E. J. Ciaccio, K. H. Ng, *et al.*, "Future iot tools for covid-19 contact tracing and prediction: a review of the state-of-the-science," *International journal of imaging systems and technology*, vol. 31, no. 2, pp. 455–471, 2021. 13

[27] A. Darwish, A. E. Hassanien, M. Elhoseny, A. K. Sangaiah, and K. Muhammad, "The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 4151–4166, 2019. 13

[28] D. V. Dimitrov, "Medical internet of things and big data in healthcare," *Healthcare informatics research*, vol. 22, no. 3, pp. 156–163, 2016. 13

[29] R. D. Garcia, G. Ramachandran, and J. Ueyama, "Exploiting smart contracts in pbft-based blockchains: A case study in medical prescription system," *Computer networks*, vol. 211, p. 109003, 2022. 13

[30] P. V. Kakarlapudi and Q. H. Mahmoud, "A systematic review of blockchain for consent management," in *Healthcare*, vol. 9, p. 137, 2021. 13

[31] M. Elhoseny, A. Abdelaziz, A. S. Salama, A. M. Riad, K. Muhammad, and A. K. Sangaiah, "A hybrid model of internet of things and cloud computing to manage big data in health services applications," *Future generation computer systems*, vol. 86, pp. 1383–1394, 2018. 13

[32] C. G. Kochan, D. R. Nowicki, B. Sauser, and W. S. Randall, "Impact of cloud-based information sharing on hospital supply chain performance: A system dynamics framework," *International Journal of Production Economics*, vol. 195, pp. 168–185, 2018. 14

[33] D. Ivanov, A. Dolgui, and B. Sokolov, "Cloud supply chain: Integrating industry 4.0 and digital platforms in the "supply chain-as-a-service"," *Transportation Research Part E: Logistics and Transportation Review*, vol. 160, p. 102676, 2022. 14

[34] J. Craig and V. Petterson, "Introduction to the practice of telemedicine," *Journal of telemedicine and telecare*, vol. 11, no. 1, pp. 3–9, 2005. 14

[35] P. J. Heinzelmann, N. E. Lugn, and J. C. Kvedar, "Telemedicine in the future," *Journal of telemedicine and telecare*, vol. 11, no. 8, pp. 384–390, 2005. 14

[36] A. Khodadad-Saryazdi, "Exploring the telemedicine implementation challenges through the process innovation approach: A case study research in the french healthcare sector," *Technovation*, vol. 107, p. 102273, 2021. 14

[37] J. J. Moffatt and D. S. Eley, "Barriers to the up-take of telemedicine in australia–a view from providers," *Rural and remote health*, vol. 11, no. 1, pp. 116–121, 2011. 14

[38] S. Hermes, T. Riasanow, E. K. Clemons, M. Böhm, and H. Krcmar, "The digital transformation of the healthcare industry: exploring the rise of emerging platform ecosystems and their influence on the role of patients," *Business Research*, vol. 13, no. 3, pp. 1033–1069, 2020. 15

[39] F. Santarsiero, G. Schiuma, D. Carlucci, and N. Helander, "Digital transformation in healthcare organisations: The role of innovation labs," *Technovation*, vol. 122, p. 102640, 2023. 15

[40] M. Herrmann, P. Boehme, T. Mondritzki, J. P. Ehlers, S. Kavadias, and H. Truebel, "Digital transformation and disruption of the health care sector: Internet-based observational study," *Journal of medical internet research*, vol. 20, no. 3, p. e104, 2018. 15

[41] R. Agarwal, G. Gao, C. DesRoches, and A. K. Jha, "Research commentary—the digital transformation of healthcare: Current status and the road ahead," *Information systems research*, vol. 21, no. 4, pp. 796–809, 2010. 15

[42] B. Meskó, Z. Drobni, É. Bényei, B. Gergely, and Z. Győrffy, "Digital health is a cultural transformation of traditional healthcare," *Mhealth*, vol. 3, 2017. 15

[43] S. Kraus, F. Schiavone, A. Pluzhnikova, and A. C. Invernizzi, "Digital transformation in healthcare: Analyzing the current state-of-research," *Journal of Business Research*, vol. 123, pp. 557–567, 2021. 15

[44] G. Schiuma, E. Schettini, F. Santarsiero, and D. Carlucci, "The transformative leadership compass: six competencies for digital transformation entrepreneurship," *International Journal of Entrepreneurial Behavior & Research*, vol. 28, no. 5, pp. 1273–1291, 2022. 15

[45] C. Appleby, J. Hendricks, J. Wurz, C. Shudes, M. Shukla, and C. Chang, "Digital transformation: From a buzzword to an imperative for health systems," *Deloitte Insights*, 2021. https://www2.deloitte.com/us/en/insights/industry/health-care/digital-transformation-in-healthcare.html [Accessed: (2024/10/16)]. 15

[46] A. Urbinati, D. Chiaroni, V. Chiesa, and F. Frattini, "The role of digital technologies in open innovation processes: an exploratory multiple case study analysis," *R&d Management*, vol. 50, no. 1, pp. 136–160, 2020. 15

[47] M. S. Rahman, M. A. Islam, M. A. Uddin, and G. Stea, "A survey of blockchain-based iot ehealthcare: Applications, research issues, and challenges," *Internet of Things*, vol. 19, p. 100551, 2022. 15

[48] U. Zaman, Imran, F. Mehmood, N. Iqbal, J. Kim, and M. Ibrahim, "Towards secure and intelligent internet of health things: A survey of enabling technologies and applications," *Electronics*, vol. 11, no. 12, p. 1893, 2022. 15

[49] M. Massaro, "Digital transformation in the healthcare sector through blockchain technology. insights from academic research and business developments," *Technovation*, vol. 120, p. 102386, 2023. 15

[50] J. Frizzo-Barker, P. A. Chow-White, A. Charters, and D. Ha, "Genomic big data and privacy: challenges and opportunities for precision medicine," *Computer Supported Cooperative Work (CSCW)*, vol. 25, pp. 115–136, 2016. 16

[51] G. Westerman, D. Bonnet, and A. McAfee, "The nine elements of digital transformation," *MIT Sloan Management Review*, vol. 55, no. 3, pp. 1–6, 2014. 16, 17

[52] P. Gray, O. A. El Sawy, G. Asper, and M. Thordarson, "Realizing strategic value through center-edge digital transformation in consumer-centric industries.," *MIS Quarterly Executive*, vol. 12, no. 1, 2013. 16

[53] M. Duplaga, "The acceptance of e-health solutions among patients with chronic respiratory conditions," *Telemedicine and e-Health*, vol. 19, no. 9, pp. 683–691, 2013. 16

[54] C. Gjellebæk, A. Svensson, C. Bjørkquist, N. Fladeby, and K. Grundén, "Management challenges for future digitalization of healthcare services," *Futures*, vol. 124, p. 102636, 2020. 16

[55] G. Elia, A. Margherita, and G. Passiante, "Digital entrepreneurship ecosystem: How digital technologies and collective intelligence are reshaping the entrepreneurial process," *Technological forecasting and social change*, vol. 150, p. 119791, 2020. 16

[56] M. J. Sousa, A. M. Pesqueira, C. Lemos, M. Sousa, and Á. Rocha, "Decision-making based on big data analytics for people management in healthcare organizations," *Journal of medical systems*, vol. 43, pp. 1–10, 2019. 16

[57] B. Solis, "The six stages of digital transformation maturity," *Custom Research by Altimeter Group on behalf of Cognizant, Altimeter*, 2016. 17

[58] I. M. of Economy and F. (MEF), "The national recovery and resilience plan (nrrp).," 2021. https://www.mef.gov.it/en/focus/The-National-Recovery-and-Resilience-Plan-NRRP/ [Accessed: (2024/10/15)]. 17

[59] R. Emilia-Romagna, "Bollettino ufficiale regione emilia-romagna n° 23 del 2 febbraio 2022, p. 110 ("deliberazione della giunta regionale 17 gennaio 2022, n. 33")," 2022. https://bur.regione.emilia-romagna.it/dettaglio-inserzione?i=ccec7a9e8ef34e73bc189a57275c9bd4 [Accessed: (2024/09/30)]. 22

[60] R. Emilia-Romagna, "Comunicato stampa rer del 17 gennaio 2022," 2022. https://notizie.regione.emilia-romagna.it/comunicati/2022/gennaio/covid-da-mercoledi-19-gennaio-in-emilia-romagna-sara-possibile-l2019autotesting-per-inizio-e-fine-isolamento-in-caso-di-positivita-lo-potranno-fare-i-cittadini-con-assistenza-sanitaria-nella-regione-e-che-abbiano-gia-fatto-la-terza-dose-boster-utilizzando [Accessed: (2024/09/30)]. 22

[61] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific reports*, vol. 10, no. 1, p. 19549, 2020. 25

[62] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Zha, W. Liang, C. Wang, K. Wang, *et al.*, "Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography," *Cell*, vol. 181, no. 6, pp. 1423–1433, 2020. 25

[63] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, and K. Wang, "Automated detection of covid-19 through convolutional neural network using chest x-ray images," *Plos one*, vol. 17, no. 1, p. e0262052, 2022. 25

[64] F. Mehboob, A. Rauf, R. Jiang, A. K. J. Saudagar, K. M. Malik, M. B. Khan, M. H. A. Hasnat, A. AlTameem, and M. AlKhathami, "Towards robust diagnosis of covid-19 using vision self-attention transformer," *Scientific Reports*, vol. 12, no. 1, p. 8922, 2022. 25

[65] M. Rezaei and M. Azarmi, "Deepsocial: Social distancing monitoring and infection risk assessment in covid-19 pandemic," *Applied Sciences*, vol. 10, no. 21, p. 7514, 2020. 25

[66] G. Jignesh Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Face mask detection using transfer learning of inceptionv3," in *Big Data Analytics: 8th International Conference, BDA 2020, Sonepat, India, December 15–18, 2020, Proceedings 8*, pp. 81–90, Springer, 2020. 25

[67] R. Emilia-Romagna, "Comunicato stampa rer del 19 gennaio 2022," 2022. https://salute.regione.emilia-romagna.it/notizie/regione/2022/gennaio/sperimentazione-autotesting-per-inizio-e-fine-isolamento-in-caso-di-positivita-al-virus [Accessed: (2024/09/30)]. 26

[68] OpenCV, "Opencv - open computer vision library." https://opencv.org/ [Accessed: (2024/09/30)]. 28

[69] OpenCV, "Cascade classifier training." https://docs.opencv.org/3.4.15/dc/d88/tutorial_traincascade.html [Accessed: (2024/09/30)]. 28, 29

[70] Ultralytics, "Yolov5." https://github.com/ultralytics/yolov5 [Accessed: (2024/09/30)]. 31

[71] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020. 31

[72] M. Lalinia and A. Sahafi, "Colorectal polyp detection in colonoscopy images using yolo-v8 network," *Signal, Image and Video Processing*, vol. 18, no. 3, pp. 2047–2058, 2024. 31

[73] M. Balasubramani, C.-W. Sung, M.-Y. Hsieh, E. P.-C. Huang, J.-S. Shieh, and M. F. Abbod, "Automated left ventricle segmentation in echocardiography using yolo: A deep learning approach for enhanced cardiac function assessment," *Electronics*, vol. 13, no. 13, p. 2587, 2024. 31

[74] N. Iriawan, A. A. Pravitasari, U. S. Nuraini, N. I. Nirmalasari, T. Azmi, M. Nasrudin, A. F. Fandisyah, K. Fithriasari, S. W. Purnami, Irhamah, *et al.*, "Yolo-unet architecture for detecting and segmenting the localized mri brain tumor image," *Applied Computational Intelligence and Soft Computing*, vol. 2024, no. 1, p. 3819801, 2024. 31

[75] B. Aldughayfiq, F. Ashfaq, N. Jhanjhi, and M. Humayun, "Yolo-based deep learning model for pressure ulcer detection and classification," in *Healthcare*, vol. 11, p. 1222, MDPI, 2023. 31

[76] darkpgmr, "Darklabel - video/image labeling and annotation tool." https://github.com/darkpgmr/DarkLabel [Accessed: (2024/09/30)]. 31

[77] "Oecd indicators on healthcare resources: Hospital beds by sector between 2001 and 2021," 2024. https://stats.oecd.org/ [Accessed: (2024/01/08)]. 50

[78] "Health at a glance 2023 oecd indicators," 2023. https://doi.org/10.1787/7a7afb35-en [Accessed: (2024/01/08)]. 50

[79] I. M. of Economy and Finance, "Monitoraggio della spesa sanitaria, rapporto n. 7," 2020. https://www.rgs.mef.gov.it/VERSIONE-I/attivita_istituzionali/monitoraggio/spesa_sanitaria/2020/ [Accessed: (2024/01/08)]. 50

[80] I. M. of Economy and Finance, "Monitoraggio della spesa sanitaria, rapporto n. 10," 2023. https://www.rgs.mef.gov.it/VERSIONE-I/attivita_istituzionali/monitoraggio/spesa_sanitaria/ [Accessed: (2024/01/08)]. 50

[81] A. Rojas-García, S. Turner, E. Pizzo, E. Hudson, J. Thomas, and R. Raine, "Impact and experiences of delayed discharge: A mixed-studies systematic review," *Health Expectations*, vol. 21, no. 1, pp. 41–56, 2018. 50

[82] B. A. Marfil-Garza, P. F. Belaunzarán-Zamudio, A. Gulias-Herrero, A. C. Zuñiga, Y. Caro-Vega, D. Kershenobich-Stalnikowitz, and J. Sifuentes-Osornio, "Risk factors associated with prolonged hospital length-of-stay: 18-year retrospective study of hospitalizations in a tertiary healthcare center in mexico," *PloS one*, vol. 13, p. 11, 2018. 50

[83] S. Ackroyd-Stolarz, J. R. Guernsey, N. J. Mackinnon, and G. Kovacs, "The association between a prolonged stay in the emergency department and adverse events in older patients admitted to hospital: a

retrospective cohort study," *BMJ Quality & Safety*, vol. 2011, no. 20, pp. 564–569, 2011. 50

[84] J. A. Heit, M. D. Silverstein, D. N. Mohr, T. M. Petterson, W. M. O'Fallon, and L. J. Melton, "Risk factors for deep vein thrombosis and pulmonary embolism: a population-based case-control study," *Archives of internal medicine*, vol. 160, no. 6, pp. 809–815, 2000. 50

[85] K. Hauck and X. Zhao, "How dangerous is a day in hospital?," *A model of adverse events and length of stay for medical inpatients. Medical care*, pp. 1068–1075, 2011. 50

[86] I. B. Molloy, B. I. Martin, W. E. Moschetti, and D. S. Jevsevar, "Effects of the length of stay on the cost of total knee and total hip arthroplasty from 2002 to 2013. the journal of bone and joint surgery," *American*, vol. 99, p. 5, 2017. 50

[87] A. Marshall, C. Vasilakis, and E. El-Darzi, "Length of stay-based patient flow models: recent developments and future directions," *Health care management science*, vol. 8, pp. 213–220, 2005. 51

[88] M. Bo, G. Fonte, F. Pivaro, M. Bonetto, C. Comi, V. Giorgis, L. Marchese, G. Isaia, G. Maggiani, E. Furno, Y. Falcone, and G. C. Isaia, "Prevalence of and factors associated with prolonged length of stay in older hospitalized medical patients," *Geriatrics & gerontology international*, vol. 16, no. 3, pp. 314–321, 2016. 51

[89] J. Evans, D. Kobewka, K. Thavorn, G. D'Egidio, E. Rosenberg, and K. Kyeremanteng, "The impact of reducing intensive care unit length of stay on hospital costs: evidence from a tertiary care hospital in canada," *Canadian Journal of Anesthesia*, vol. 65, no. 6, pp. 627–635, 2018. 51

[90] K. Stricker, H. U. Rothen, and J. Takala, "Resource use in the icu: short-vs. long-term patients," *Acta Anaesthesiologica Scandinavica*, vol. 47, no. 5, pp. 508–515, 2003. 51

[91] Istat, "Annual italian statistics," 2022. last accessed 2024/01/08. 51

[92] S. Barnes, E. Hamrock, M. Toerper, S. Siddiqui, and S. Levin, "Real-time prediction of inpatient length of stay for discharge prioritization," *Journal of the American Medical Informatics Association*, vol. 23(e1), e2-e10, 2016. 51

[93] C. Gholipour, F. Rahim, A. Fakhree, and B. Ziapour, "Using an artificial neural networks (anns) model for prediction of intensive care unit (icu) outcome and length of stay at hospital in traumatic patients," *Journal of clinical and diagnostic research: JCDR*, vol. 9, p. 4, 2015. 51, 65

[94] K. Stone, R. Zwiggelaar, P. Jones, and N. Mac Parthaláin, "A systematic review of the prediction of hospital length of stay: Towards a unified framework," *PLOS Digital Health*, vol. 1, p. 4, 2022. 51

[95] H. M. Zolbanin, B. Davazdahemami, D. Delen, and A. H. Zadeh, "Data analytics for the sustainable use of resources in hospitals: predicting the length of stay for patients with chronic diseases," *Information & Management*, vol. 59, no. 5, p. 103282, 2022. 51

[96] R. N. Mekhaldi, P. Caulier, S. Chaabane, A. Chraibi, and S. Piechowiak, "Using machine learning models to predict the length of stay in a hospital setting," in *World conference on information systems and technologies*, (Cham), pp. 202–211, Springer International Publishing, April 2020. 51

[97] G. Hesselink, M. Van Den Bogaert, R. P. Akkermans, and Y. Schoon, "Risk factors for prolonged length of stay of older patients in an academic emergency department: a retrospective cohort study," *Emergency medicine international*, vol. 2019, 2019. 51

[98] C. P. Launay, A. Kabeshova, A. Lanoé, J. Chabot, E. J. Levinoff, and O. Beauchet, "Age effect on the prediction of risk of prolonged length hospital stay in older patients visiting the emergency department: results from a large prospective geriatric cohort study," *BMC geriatrics*, vol. 18, no. 1, pp. 1–6, 2018. 51

[99] B. Thompson, K. O. Elish, and R. Steele, "Machine learning-based prediction of prolonged length of stay in newborns," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1454–1459, IEEE, 2018. 51

[100] X. Ma, Y. Si, Z. Wang, and Y. Wang, "Length of stay prediction for icu patients using individualized single classification algorithm," in *Computer methods and progr in biomedicine, 186, 105224*, AMS, 2020. 51, 65

[101] E. Rocheteau, P. Liò, and S. Hyland, "Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit," in *Proceedings of the conference on health*, (and learning), pp. 58–68, inference, April 2021. 51

[102] T. A. Daghistani, R. Elshawi, S. Sakr, A. M. Ahmed, A. Al-Thwayee, and M. H. Al-Mallah, "Predictors of in-hospital length of stay among cardiac patients: a machine learning approach," *International journal of cardiology*, vol. 288, pp. 140–147, 2019. 51

[103] E. P. Rotar, J. P. Beller, M. E. Smolkin, W. Z. Chancellor, G. Ailawadi, L. T. Yarboro, M. Hulse, S. J. Ratcliffe, and N. R. Teman, "Prediction of prolonged intensive care unit length of stay following cardiac surgery," *Seminars in Thoracic and Cardiovascular Surgery*, vol. 34, pp. 172–179, March 2022. 51

[104] M. T. Chuang, Y. H. Hu, and C. L. Lo, "Predicting the prolonged length of stay of general surgery patients: a supervised learning approach," *International Transactions in Operational Research*, vol. 25, no. 1, pp. 75–90, 2018. 51

[105] T. A. Trunfio, A. Scala, C. Giglio, G. Rossi, A. Borrelli, M. Romano, and G. Improta, "Multiple regression model to analyze the total los for patients undergoing laparoscopic appendectomy," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, pp. 1–8, 2022. 51

[106] R. Gohil, M. Rishi, and B. H. Tan, "Pre-operative serum albumin and neutrophillymphocyte ratio are associated with prolonged hospital stay following colorectal cancer surgery," *British journal of medicine and medical research*, vol. 4, no. 1, p. 481, 2014. 51

[107] Y. Y. Jo, J. Han, H. W. Park, H. Jung, J. D. Lee, J. Jung, H. S. Cha, D. K. Sohn, and Y. Hwangbo, "Prediction of prolonged length of hospital stay after cancer surgery using machine learning on electronic health records: retrospective cross-sectional study," *JMIR medical informatics*, vol. 9, p. 2, 2021. 51

[108] T. A. Trunfio, A. Borrelli, and G. Improta, "Is it possible to predict the length of stay of patients undergoing hip-replacement surgery?," *International Journal of Environmental Research and Public Health*, vol. 19, no. 10, p. 6219, 2022. 51

[109] M. Olivato, N. Rossetti, A. E. Gerevini, M. Chiari, L. Putelli, and I. Serina, "Machine learning models for predicting short-long length of stay of covid-19 patients," *Procedia Computer Science*, vol. 207, pp. 1232–1241, 2022. 52

[110] A. J. Zeleke, P. Palumbo, P. Tubertini, R. Miglio, and L. Chiari, "Machine learning-based prediction of hospital prolonged length of stay admission at emergency department: a gradient boosting algorithm analysis," *Front. Artif. Intell.*, vol. 6, 2023. Article 1179226. 52

[111] G. D'Onofrio, A. D'Amore, F. Onofaro, E. Caputi, A. Napoli, M. Triassi, and M. R. Marino, "Prediction of hospital length stay for patients undergoing mastectomy," *Stud Health Technol Inform.*, vol. 29, pp. 261–264, June 2023. PMID: 37387012. 52

[112] V. Di Matteo, T. Tommasini, P. Morandini, V. Savevski, G. Grappiolo, and M. Loppini, "Machine learning prediction model to predict length of stay in patients undergoing hip or knee arthroplasties: Results from a high volume single center multivariate analysis," *Preprints*, vol. 2023110915, 2023. 52, 65

[113] I. M. of Health, "Rapporto annuale sull'attività di ricovero ospedaliero," 2020. last accessed 2024/01/08. 57

[114] M. Sud, B. Yu, H. C. Wijeysundera, *et al.*, "Associations between short or long length of stay and 30-day readmission and mortality in hospitalized patients with heart failure," *JACC: Heart Failure*, vol. 5, no. 8, pp. 578–588, 2017. 65

[115] R. Chen, S. Zhang, J. Li, *et al.*, "A study on predicting the length of hospital stay for chinese patients with ischemic stroke based on the xgboost algorithm," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, pp. 1–10, 2023. 65

[116] E. Montella, M. R. Marini, M. Majolo, E. Raiola, G. Russo, G. Longo, A. Lombardi, A. Borrelli, and M. Triassi, "Regression and classification methods for predicting the length of hospital stay after cesarean section: a bicentric study," in *Proceedings of the 6th International Conference on Medical and Health Informatics*, (New York, NY, USA), pp. 135—-140, Association for Computing Machinery, 2022. 65

[117] W. E. Muhlestein, D. S. Akagi, J. M. Davies, and L. B. Chambless, "Predicting inpatient length of stay after brain tumor surgery: developing machine learning ensembles to improve predictive performance," *Neurosurgery*, vol. 85, no. 3, p. 384, 2019. 65

[118] J. Chrusciel, F. Girardon, L. Roquette, D. Laplanche, A. Duclos, and S. Sanchez, "The prediction of hospital length of stay using unstructured data," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 351, 2021. 66

[119] L. Y. Jiang, X. C. Liu, N. Nejatian, *et al.*, "Health system-scale language models are all-purpose prediction engines," *Nature*, pp. 1–6, 2023. 66

[120] D. Zhang, C. Yin, J. Zeng, X. Yuan, and P. Zhang, "Combining structured and unstructured data for predictive models: a deep learning approach," *BMC medical informatics and decision making*, vol. 20, no. 1, pp. 1–11, 2020. 66

[121] ANSA, "Ue, imprese europee in ritardo sulla trasformazione digitale," 2024. https://www.ansa.it/osservatorio_intelligenza_artificiale/notizie/progetto_solaris/2024/07/02/ue-imprese-europee-in-ritardo-sulla-trasformazione-digitale_8baf5d15-b807-41c6-89d2-ea75e9e14a86.htm [Accessed: (2024/07/03)]. 68

[122] J. L. Phillips, B. E. Jackson, E. L. Fagan, S. E. Arze, B. Major, N. R. Zenarosa, and H. Wang, "Overcrowding and its association with patient outcomes in a median-low volume emergency department," *Journal of clinical medicine research*, vol. 9, no. 11, p. 911, 2017. 70

[123] C. M. Fernandes, A. Price, and J. M. Christenson, “Does reduced length of stay decrease the number of emergency department patients who leave without seeing a physician?,” *The Journal of emergency medicine*, vol. 15, no. 3, pp. 397–399, 1997. 70

[124] E. Manno, M. Pesce, U. Stralla, F. Festa, S. Geninatti, M. F. Balzarro, and B. Gelain, “Specialized fast track: A sustainable model to improve emergency department patient flow,” *J. Hosp. Adm*, vol. 4, p. 40, 2015. 70

[125] M. Sanchez, A. J. Smally, R. J. Grant, and L. M. Jacobs, “Effects of a fast-track area on emergency department performance,” *The Journal of emergency medicine*, vol. 31, no. 1, pp. 117–120, 2006. 70

[126] J. Considine, M. Kropman, E. Kelly, and C. Winter, “Effect of emergency department fast track on emergency department length of stay: a case–control study,” *Emergency Medicine Journal*, vol. 25, no. 12, pp. 815–819, 2008. 70

[127] S. Daly, D. A. Campbell, and P. A. Cameron, “Short-stay units and observation medicine: a systematic review,” *Medical Journal of Australia*, vol. 178, no. 11, pp. 559–563, 2003. 70

[128] C. W. Baugh, A. K. Venkatesh, and J. S. Bohan, “Emergency department observation units: a clinical and financial benefit for hospitals,” *Health care management review*, vol. 36, no. 1, pp. 28–37, 2011. 70

[129] P. Yoon, I. Steiner, and G. Reinhardt, “Analysis of factors influencing length of stay in the emergency department,” *Canadian Journal of Emergency Medicine*, vol. 5, no. 3, pp. 155–161, 2003. 70

[130] D. Liew, D. Liew, and M. P. Kennedy, “Emergency department length of stay independently predicts excess inpatient length of stay,” *Medical Journal of Australia*, vol. 179, no. 10, pp. 524–526, 2003. 70

[131] G. Sariyer, M. G. Ataman, and I. Kiziloğlu, “Factors affecting length of stay in the emergency department: a research from an operational viewpoint,” *International Journal of Healthcare Management*, vol. 13, no. sup1, pp. 173–182, 2020. 70

[132] J. F. Kenny, B. C. Chang, and K. C. Hemmert, "Factors affecting emergency department crowding," *Emergency Medicine Clinics*, vol. 38, no. 3, pp. 573–587, 2020. 70

[133] J. M. Pines, R. T. Griffey, and D. C. Cone, "What we have learned from a decade of ed crowding research," *Academic Emergency Medicine*, vol. 22, no. 8, pp. 893–915, 2015. 70

[134] G. Savioli, I. F. Ceresa, N. Gri, G. Bavestrello Piccini, Y. Longhitano, C. Zanza, and M. A. Bressan, "Emergency department overcrowding: understanding the factors to find corresponding solutions," *Journal of personalized medicine*, vol. 12, no. 2, p. 279, 2022. 70

[135] S. Gaeta, T. Edwards, S. Bourenane, C. E. Gonzalez, K. McFarland, and D. Milling, "Emergency department surge and overcrowding: An interdisciplinary solution for an institutional issue," *Journal of Clinical Oncology*, vol. 37, no. 27, 2019. 70

[136] K. Phillips, M. Knowlton, and J. Riseden, "Emergency department nursing burnout and resilience," *Advanced emergency nursing journal*, vol. 44, no. 1, pp. 54–62, 2022. 70

[137] M. A. Rahman, B. Honan, T. Glanville, P. Hough, and K. Walker, "Using data mining to predict emergency department length of stay greater than 4 hours: Derivation and single-site validation of a decision tree algorithm," *Emergency Medicine Australasia*, vol. 32, pp. 416–421, 2020. 71

[138] E. E. Etu, L. Monplaisir, S. Arslanturk, S. Masoud, C. Aguwa, I. Markevych, and J. Miller, "Prediction of length of stay in the emergency department for covid-19 patients: A machine learning approach," *IEEE Access*, vol. 10, pp. 42243–42251, 2022. 71

[139] J. P. d'Etienne, Y. Zhou, C. Kan, S. Shaikh, A. F. Ho, E. Suley, and H. Wang, "Two-step predictive model for early detection of emergency department patients with prolonged stay and its management implications," *The American Journal of Emergency Medicine*, vol. 40, pp. 148–158, 2021. 71

[140] P. Walsh, P. Cunningham, S. J. Rothenberg, S. O'Doherty, H. Hoey, and R. Healy, "An artificial neural network ensemble to predict disposition and length of stay in children presenting with bronchiolitis,"

European Journal of Emergency Medicine, vol. 11, no. 5, pp. 259–264, 2004. 71

[141] E. Benevento, D. Aloini, and N. Squicciarini, "Towards a real-time prediction of waiting times in emergency departments: A comparative analysis of machine learning techniques," *International Journal of Forecasting*, vol. 39, no. 1, pp. 192–208, 2023. 71

[142] C. Ricciardi, M. R. Marino, T. A. Trunfio, M. Majolo, M. Romano, F. Amato, and G. Improta, "Evaluation of different machine learning algorithms for predicting the length of stay in the emergency departments: a single-centre study," *Frontiers in Digital Health*, vol. 5, 2024. 71

[143] A. Azari, V. P. Janeja, and S. Levin, "Imbalanced learning to predict long stay emergency department patients," in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 807–814, 2015. 71

[144] A. Naemi, T. Schmidt, M. Mansourvar, A. Ebrahimi, and U. K. Wiil, "Quantifying the impact of addressing data challenges in prediction of length of stay," *BMC Medical Informatics and Decision Making*, vol. 21, pp. 1–13, 2021. 71

[145] I. M. of Health – General Directorate of Health Planning, "Linee di indirizzo nazionali sull'osservazione breve intensiva (obi)," 2019. https://www.salute.gov.it/imgs/C_17_pubblicazioni_3142_allegato.pdf [Accessed: (2024/01/08)]. 74, 79

[146] I. M. of Health – General Directorate of Health Planning, "Linee di indirizzo nazionali sul triage," 2019. https://www.salute.gov.it/imgs/C_17_pubblicazioni_3145_allegato.pdf [Accessed: (2024/01/08)]. 79

[147] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, 2011. 86

[148] S. Wang and X. Yao, "Multiclass imbalance problems: Analysis and potential solutions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 4, pp. 1119–1130, 2012. 86

[149] G. Valentini and F. Masulli, "Ensembles of learning machines," *Neural Nets: 13th Italian Workshop on Neural Nets, WIRN VIETRI 2002 Vietri sul Mare, Italy, May 30–June 1, 2002, Springer Berlin Heidelberg*, vol. 13, pp. 3–20, 2002. 86

[150] P. Perliti Scorzoni, A. Giovanetti, F. Bolelli, and C. Grana, "Sustainable use of resources in hospitals: A machine learning-based approach to predict prolonged length of stay at the time of admission," in *In Proceedings of 12th IHIET International Conference*, 2024. 86

[151] AGENAS, "Accessi in pronto soccorso e implementazione dm 77/2022 per una migliore presa in carico dei pazienti," 2024. https://www.agenas.gov.it/comunicazione/primo-piano/2413-22-aprile-%E2%80%93-evento-agenas-accessi-in-pronto-soccorso-e-implementazione-dm-77-2022-per-una-migliore-presa-in-carico-dei-pazienti/ [Accessed: (2024/01/08)]. 88

[152] O. Bashkin, S. Caspi, R. Haligoa, S. Mizrahi, and R. Stalnikowicz, "Organizational factors affecting length of stay in the emergency department: initial observational study," *Israel journal of health policy research*, vol. 4, pp. 1–7, 2015. 88

[153] M. Sartini, A. Carbone, A. Demartini, L. Giribone, M. Oliva, A. M. Spagnolo, and M. L. Cristina, "Overcrowding in emergency department: causes, consequences, and solutions — a narrative review," *Healthcare*, vol. 10, no. 9, 2022. 88

[154] H. R. Rasouli, A. A. Esfahani, M. Nobakht, M. Eskandari, S. Mahmoodi, H. Goodarzi, and M. A. Farajzadeh, "Outcomes of crowding in emergency departments; a systematic review," *Archives of Academic Emergency Medicine*, vol. 7, no. 1, p. 52, 2019. 88

[155] K. Payne, D. Risi, A. O'Hare, S. Binks, and K. Curtis, "Factors that contribute to patient length of stay in the emergency department: A

time in motion observational study," *Australasian Emergency Care*, vol. 26, no. 4, pp. 321–325, 2023. 88

[156] S. Abedian, E. Bitaraf, and M. Askari, "Advantages of a web-based real-time bed-management system for hospital admission monitoring in iran," in *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth*, pp. 536–540, IOS Press, 2018. 90

[157] R. A. Baru, E. A. Cudney, I. G. Guardiola, D. L. Warner, and R. E. Phillips, "Systematic review of operations research and simulation methods for bed management," in *IIE Annual Conference. Proceedings*, p. 298, Institute of Industrial and Systems Engineers (IISE), 2015. 90

[158] W. J. Thomas, K. E. Guire, and G. G. Horvat, "Is patient length of stay related to quality of care?," *Journal of Healthcare Management*, vol. 42, no. 4, pp. 489–507, 1997. 90

[159] A. J. Singer, H. C. Thode Jr, P. Viccellio, and J. M. Pines, "The association between length of emergency department boarding and mortality," *Academic Emergency Medicine*, vol. 18, no. 12, pp. 1324–1329, 2011. 90

[160] G. Kerpedzhiev, T. Manner-Romberg, O. Meindl, and C. Regal, "Towards a maturity model: Bed management capabilities in hospitals," 2019. 90, 93

[161] B. Wertheimer, R. E. Jacobs, E. Iturrate, M. Bailey, and K. Hochman, "Discharge before noon: effect on throughput and sustainability," *Journal of hospital medicine*, vol. 10, no. 10, pp. 664–669, 2015. 90

[162] S. Khanna, D. Sier, J. Boyle, and K. Zeitz, "Discharge timeliness and its impact on hospital crowding and emergency department flow performance," *Emergency Medicine Australasia*, vol. 28, no. 2, pp. 164–170, 2016. 90

[163] S. Khanna, J. Boyle, N. Good, and J. Lind, "Early discharge and its effect on ed length of stay and access block," in *Health informatics: building a healthcare future through trusted information*, pp. 92–98, IOS Press, 2012. 90

[164] F. Landeiro, K. Roberts, A. M. Gray, and J. Leal, "Delayed hospital discharges of older patients: a systematic review on prevalence and costs," *The Gerontologist*, vol. 59, no. 2, pp. e86–e97, 2019. 90

[165] H. A. L. Rocha, A. K. L. d. C. Santos, A. C. d. C. Alcântara, C. S. S. d. C. Lima, S. G. M. O. Rocha, R. M. Cardoso, and J. R. Cremonin, "Bed management team with kanban web-based application," *International Journal for Quality in Health Care*, vol. 30, no. 9, pp. 708–714, 2018. 90, 92

[166] J. Stein, D. Murphy, C. Payne, D. Clark, W. Bronstein, D. Tong, B. Castle, and S. Shapiro, "A remedy for fragmented hospital care," *Harvard Business Review*, 2013. https://hbr.org/2013/11/a-remedy-for-fragmented-hospital-care [Accessed: (2024/10/27)]. 90

[167] P. Landa, M. Sonnessa, E. Tànfani, and A. Testi, "Multiobjective bed management considering emergency and elective patient flows," *International Transactions in Operational Research*, vol. 25, no. 1, pp. 91–110, 2018. 90

[168] F. Tortorella, D. Ukanowicz, P. Douglas-Ntagha, R. Ray, and M. Triller, "Improving bed turnover time with a bed management system," *JONA: The Journal of Nursing Administration*, vol. 43, no. 1, pp. 37–43, 2013. 90, 92

[169] A. Jadwin, "Maximizing patient flow & reducing inpatient hospital los: Incremental steps to create culture change," *Oncology Issues*, vol. 28, no. 4, pp. 26–31, 2013. 90, 92

[170] G. Winti, S. Rodenhurstii, and B. D. Frankliniii, "Streamlined ward working-live bed status, multidisciplinary handover and electronic vital sign documentation.," *British Journal of Healthcare Computing & Information Management*, 2010. 90

[171] A. Hanelt, R. Bohnsack, D. Marz, and C. Antunes Marante, "A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change," *Journal of management studies*, vol. 58, no. 5, pp. 1159–1197, 2021. 91

[172] J. Reis, M. Amorim, N. Melão, and P. Matos, "Digital transformation: a literature review and guidelines for future research," *Trends and Advances in Information Systems and Technologies: Volume 1 6*, pp. 411–421, 2018. 91

[173] D. Rolls, S. Khanna, N. Lloyd, A. Reeson, R. Jayasena, C. McCormick, and S. Hakkennes, "Before-after evaluation of patient length of stay in a rehabilitation context following implementation of an electronic patient journey board," *International Journal of Medical Informatics*, vol. 134, p. 104042, 2020. 91

[174] K. W. Clark, S. Moller, and L. O'Brien, "Electronic patient journey boards a vital piece of the puzzle in patient flow," *Australian Health Review*, vol. 38, no. 3, pp. 259–264, 2014. 91

[175] D. A. Martinez, E. M. Kane, M. Jalalpour, J. Scheulen, H. Rupani, R. Toteja, C. Barbara, B. Bush, and S. R. Levin, "An electronic dashboard to monitor patient flow at the johns hopkins hospital: communication of key performance indicators using the donabedian model," *Journal of medical systems*, vol. 42, pp. 1–8, 2018. 91

[176] A. Keniston, L. McBeth, J. Pell, K. Bowden, A. Metzger, J. Nordhagen, A. Anthony, J. Rice, M. Burden, *et al.*, "The effectiveness of a multidisciplinary electronic discharge readiness tool: prospective, single-center, pre-post study," *JMIR human factors*, vol. 8, no. 4, p. e27568, 2021. 92

[177] I. S. d. Oliveira, E. d. F. A. Lima, R. I. C. d. Silva, K. Crozeta, I. C. B. Dias, and C. C. Primo, "Beds management in the urgency and emergency using kanban," *Research, Society and Development*, vol. 9, no. 10, p. e1329108354, 2020. 92

[178] T. G. Esteves Silva, M. F. Souza Bezerra da Silva, L. C. Nery Vieira, C. Agra Pimentel, and M. Andreotti Musetti, "Contributions of electronic and traditional kanban to hospital bed management," *Revista Producao Online*, vol. 21, no. 3, pp. 818–836, 2021. 92

[179] F. Noonan, J. Nogales, C. Doyle, E. Broderick, and J. Walsh, "Bed management system development.," in *HEALTHINF*, pp. 376–383, 2023. 92, 119

[180] R. Freitas and C. Araújo, "The digital transformation applied to bed management in hospitals," *RAM. Revista de Administração Mackenzie*, vol. 25, no. 2, p. eRAMR240099, 2024. 93

[181] B. M. Porto, S. X. Terra, and D. A. Philippi, "A qualitative analysis of the main factors in bed management in high complexity hospital emergencies," *Revista de Gestão e Secretariado*, vol. 14, no. 2, pp. 1682–1717, 2023. 93, 120

[182] A. Cameron, K. Rodgers, A. Ireland, R. Jamdar, and G. A. McKay, "A simple tool to predict admission at the time of triage," *Emergency Medicine Journal*, vol. 32, no. 3, pp. 174–179, 2015. 93

[183] R. Emilia-Romagna, "Posti letto per 1000 abitanti," 2024. https://applicazioni.regione.emilia-romagna.it/ReportERHome/stats/flusso/13/macroambito/1/ambito/1/scheda/3 [Accessed: (2024/11/01)]. 94

[184] OECD, "Oecd statsdata explorer hospital beds by sector," 2024. https://data-explorer.oecd.org/vis?lc=en&fs[0]=Topic%2C1%7CHealth%23HEA%23%7CHealthcare%20resources%20and%20equipment%23HEA_RES%23&pg=0&fc=Topic&bp=true&snb=20&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_HEALTH_REAC_HOSP%40DF_BEDS_SECT&df[ag]=OECD.ELS.HD&df[vs]=1.0&dq=ITA..10P3HB.._T....&pd=2015%2C&to[TIME_PERIOD]=false&vw=tb [Accessed: (2024/11/01)]. 94

[185] B. J. Finkelstein and L. B. Junior, "A capacidade de leitos hospitalares no brasil, as internações no sus, a migração demográfica e os custos dos procedimentos," *Jornal Brasileiro de Economia da Saúde*, vol. 12, no. 3, pp. 273–280, 2020. 94

[186] I. M. of Health – General Directorate of Health Planning, "Linee di indirizzo nazionali per lo sviluppo del piano di gestione del sovraffollamento in pronto soccorso," 2019. https://www.salute.gov.it/imgs/C_17_pubblicazioni_3143_allegato.pdf [Accessed: (2024/11/08)]. 106

[187] R. Emilia-Romagna, "Scheda di dimissione ospedaliera - efficienza - degenza, accessi medi e peso medio - degenza media - icp," 2024. https://applicazioni.regione.emilia-

romagna.it/ReportERHome/stats/flusso/10/macroambito/4/ambito/1/scheda/1 [Accessed: (2024/11/05)]. 118

[188] P. del Consiglio dei Ministri Dipartimento della Protezione Civile, "Coronavirus: il 31 marzo si chiude lo stato di emergenza," 2022. https://www.protezionecivile.gov.it/it/notizia/coronavirus-il-31-marzo-si-chiude-lo-stato-di-emergenza/ [Accessed: (2024/11/08)]. 118

[189] M. della Salute, *Rapporto annuale sull'attività di ricovero ospedaliero - DATI SDO 2022*. Ministero della Salute, ex Direzione generale della programmazione sanitaria - Ufficio, 2022. 118

[190] G. Health, "Gauteng health introduces electronic bed management system," 2016. https://www.gov.za/speeches/gauteng-health-introduces-electronic-bed-management-system-19-apr-2016-0000 [Accessed: (2024/11/09)]. 119

[191] S. Khanna, J. Boyle, N. Good, and J. Lind, "Unravelling relationships: Hospital occupancy levels, discharge timing and emergency department access block," *Emergency Medicine Australasia*, vol. 24, no. 5, pp. 510–517, 2012. 120

[192] K. Ward-Stockham, O. Omonaiye, P. Darzins, C. Kitt, E. Newnham, N. F. Taylor, and J. Considine, "Understanding the influences on hospital discharge decision-making from patient, carer and staff perspectives," *BMC Health Services Research*, vol. 24, no. 1, p. 1097, 2024. 120

[193] F. Borges, E. Bernardino, M. M. Stegani, and N. S. Tonini, "Performance of nurses in the bed management service of a teaching hospital," *Revista brasileira de enfermagem*, vol. 73, no. 4, p. e20190349, 2020. 120

[194] OECD, E. O. on Health Systems, and Policies, *Italy: Country Health Profile 2023*. OECD Publishing, 2023. 125

[195] D. S. Foundation, "Django - the web framework for perfectionists with deadlines." https://www.djangoproject.com/ [Accessed: (2024/09/30)]. 157

[196] HL7, "Health level 7 international." https://www.hl7.org/ [Accessed: (2024/11/20)]. 180

Appendix A

List of Publications

This Section contains the list of research papers published during the Ph.D. period, as well as pre-prints which are currently under review.

[1] Anita Giovanetti, Laura Canalini, and Paolo Perliti Scorzoni. A Compact Deep Ensemble for High Quality Skin Lesion Classification. In *2021 21st International Conference on Image Analysis and Processing (ICIAP)*, pages 510–521. Springer, August 2022.

[2] Paolo Perliti Scorzoni, Anita Giovanetti, Federico Bolelli, and Costantino Grana. Sustainable Use of Resources in Hospitals: A Machine Learning-Based Approach to Predict Prolonged Length of Stay at the Time of Admission. In *2024 12th International Conference on Human Interaction and Emerging Technologies (IHIET)*, pages 603–613. AHFE Open Access, August 2024.

[3] Paolo Perliti Scorzoni, Anita Giovanetti, Federico Bolelli, and Costantino Grana. Machine Learning-Based Prediction of Emergency Department Prolonged Length of Stay: A Case Study from Italy. In *2025 8th International Conference on Intelligent Human Systems Integration (IHSI)*, pages 173–183. AHFE Open Access, February 2025.

[4] Paolo Perliti Scorzoni, Anita Giovanetti, Federico Bolelli, and Costantino Grana. Optimizing Resource Allocation in Public Healthcare: A Machine Learning Approach for Length-of-Stay Prediction. In *2024 3rd*

International Workshop on Artificial Intelligence for Healthcare Applications (AIHA), as part of the 27th International Conference on International Conference on Pattern Recognition (ICPR). Springer, May 2025.

Appendix B

Deployment of the SARS-CoV-2 Image Classifier

The framework employed for implementing the ecosystem supporting the trained neural network is Django [195]. The rationale for this choice can be summarized as follows:

- Extremely versatile ("batteries-included") and widely adopted framework.
- Seamless integration with Django REST Framework to expose microservice-oriented logic (RESTful APIs).
- Ability to execute custom logic in both interactive and batch modes (using commands).

B.1 Utility Commands

Several utility commands have been implemented for basic operations:

find_duplicates

Detects duplicate images (through file hashing[1]) within a folder and subsequently removes them.

```
python manage.py find_duplicates <folder> [--dup-delete]
```

The `--dup-delete` parameter forces deletion of duplicates (default setting is to rename duplicate files).

swabs_detection

Performs massive detection and classification of images contained in a source folder. Any duplicate (or previously processed) images are discarded.

```
python manage.py swabs_detection <source_folder>
   [--weights] [--metadata_in] [--metadata_out]
   [--use-db] [--device]
```

The `--use-db` parameter can be specified to persist metadata of the processed images within the application database, in conjunction with the `--metadata_in` parameter for indicating an optional reference file containing ground-truth labels associated with the images to be imported. Additionally, an output file can be generated (using the `--metadata_out` parameter) containing, for each input image: the number of swabs detected, the predicted class, and a flag indicating whether the predicted class matches the ground-truth label (using the input reference file, if present). To leverage a specific device for model execution, the `--device` parameter can be used (possible values: 0, 1, 2, ..., 'cpu').

B.1.1 Basic Front-End

A minimalist web application has been developed to facilitate sample testing.

[1]The "imagehash" library (https://github.com/JohannesBuchner/imagehash) was utilized in place of standard cryptographic hashing algorithms (such as MD5 or SHA-1) to allow for greater flexibility in comparison. With MD5 and SHA-1, even imperceptible changes in the original image would result in entirely different hash values.

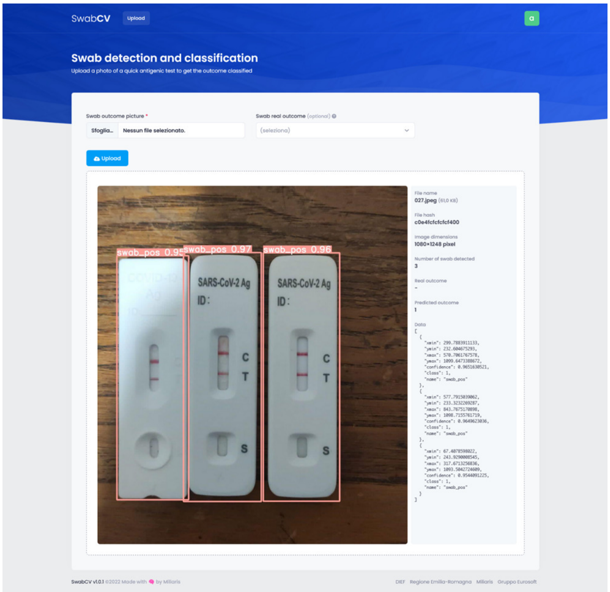


Figure B.1: Image Classification Through the Front-End Interface.

Through a dedicated interface, users can submit a picture (optionally accompanied by its corresponding label) and receive the classification result. This visual feedback enhances user understanding of the model's detection capabilities. In addition to displaying the processed image (where the identified swab is framed by a bounding box), the application provides several details related to the uploaded file (name, size, width and height in pixels), the number of swabs detected, the predicted outcome, and additional metadata for each detected antigen test (bounding box, predicted class, confidence level).

B.1.2 API

The developed set of APIs allows the model to be invoked following a microservice architecture. To invoke the APIs, an authorization token associated with a valid user is required[2]. The available endpoints include:

[2]Both can be easily created through the Django administration console.

Paginated list of imported swabs

`GET /api/v1/swabs/`

To navigate to a specific page of the list, the `page` parameter can be used.

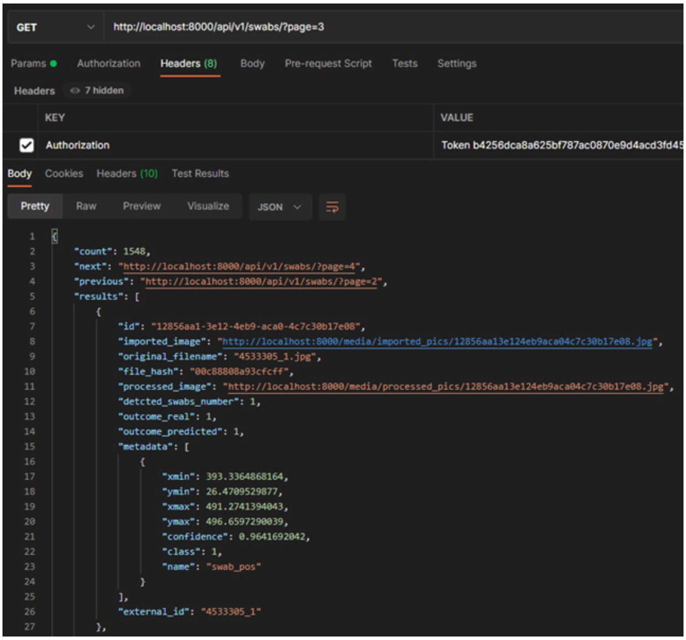


Figure B.2: REST API Endpoint for Listing Specimen Tests.

Details of a previously imported swab

`GET /api/v1/swabs/<id>`

The `<id>` path parameter must correspond to an existing identifier (otherwise, a 404 error is returned). Attributes of the returned object (in JSON format) include:

- Internal image identifier (id)
- URL for viewing the uploaded image

– Name of the uploaded file
– File hash
– URL for viewing the processed image
– Number of swabs detected in the image
– Ground-truth label (if provided as input)
– Classification outcome
– Additional metadata (bounding-box coordinates, identified class, confidence level, etc.)

Notably, assuming direct integration between the APIs and FSE, the number of swabs detected in the image can be used by the caller (FSE) to condition upstream operations (*e.g.*, in case of no swab or more than one swab detected, the citizen may be prompted to upload a proper image).

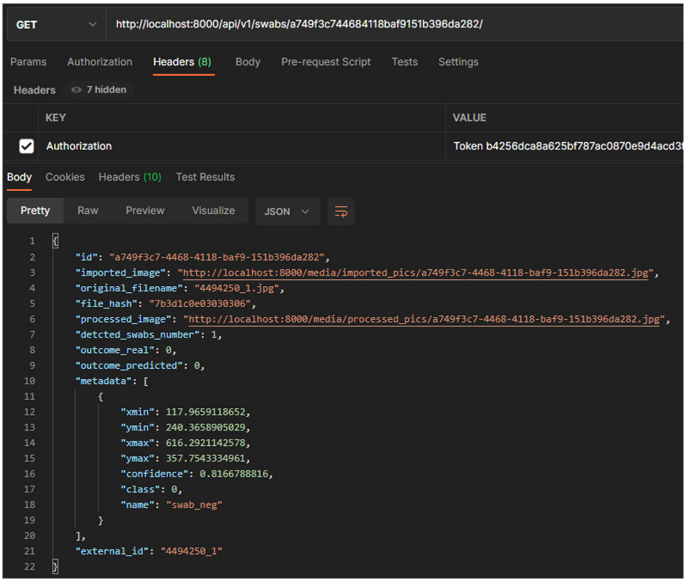


Figure B.3: REST API Endpoint for Specimen Test Details.

Importing a new swab (image file and optional label)

`POST /api/v1/swabs`

If the image is to be accompanied by its corresponding label (*i.e.*, the outcome declared by the citizen during the submission phase to FSE), the `outcome_real` parameter can be used (permitted values: 0 = Negative / 1 = Positive).

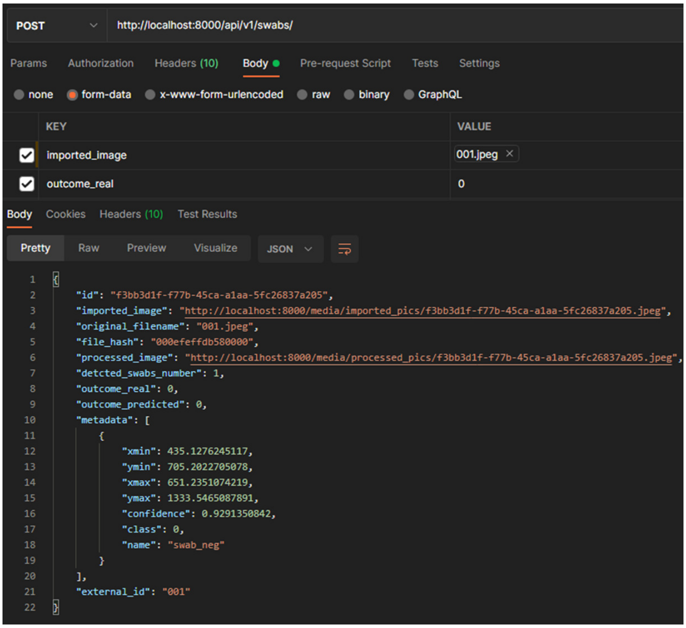


Figure B.4: REST API Endpoint for Uploading Specimen Test (Inference).

Appendix C

Supplementary Model Insights for LOS and pLOS Prediction

Some details regarding the predictive models discussed in the Chapters 4 and 5 have been placed in this Appendix to maintain the natural reading flow of the thesis.

C.1 Summary of ML Algorithms

This section presents a concise overview of the machine learning algorithms utilized for predicting Length of Stay (LOS) and prolonged Length of Stay (pLOS) in Chapters 4 and 5. These algorithms range from simple baselines (DummyRegressor/Classifier) to complex ensembles and neural networks, reflecting the varying degrees of complexity in hospital and emergency department data. Linear models (Linear Regression, Logistic Regression, Ridge, Lasso, ElasticNet) retain a degree of interpretability that is particularly valuable in clinical contexts. In contrast, tree-based methods (Decision Trees, Random Forests, Gradient Boosting, XGBoost, LGBM, CatBoost) and ensemble techniques (AdaBoost, Stacking, Voting) are better suited for capturing the non-linear relationships frequently observed in patient conditions. Instance-based (KNeighbors) and kernel-based (SVR, SVC) methods are typically employed to model complex data structures, while neural networks (MLPRegressor/Classifier) can capture high-order interactions through multiple layers of abstraction. By leveraging this diverse set of algorithms, our analyses aimed to determine which approaches could best address challenges such as class imbalance in pLOS prediction and variable correlations in public healthcare datasets.

Analyses were primarily conducted using the Scikit-learn library in Python.

C.1.1 Regression Models

- **DummyRegressor:** A baseline method that ignores the input features and always predicts a constant value (typically the training set mean), ensuring that any clinically relevant model outperforms a naive guess.

- **LinearRegression:** A classical model that fits a linear relationship between the predictors $\mathbf{x}$ and the outcome y by minimizing the residual sum of squares. In its simplest form, it is expressed as:

  $$y = \beta_0 + \beta^T \mathbf{x},$$

  with parameters estimated by:

  $$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta^T \mathbf{x}_i\right)^2 .$$

  Its interpretability is advantageous in healthcare contexts, where a clear understanding of feature contributions is essential for clinical decision-making.

- **Ridge:** A regularized variant of linear regression incorporating an ℓ_2 penalty to mitigate multicollinearity. The optimization function is:

  $$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta^T \mathbf{x}_i\right)^2 + \lambda \|\beta\|_2^2 \right\},$$

  where the ℓ_2 penalty helps stabilize the model, which is beneficial when many correlated features (*e.g.*, comorbidities) are present.

- **Lasso:** Similar to Ridge but employing an ℓ_1 penalty to promote sparsity in the coefficients, facilitating feature selection. Formally:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta^T \mathbf{x}_i\right)^2 + \lambda \|\beta\|_1 \right\},$$

  with the penalty applied only to β. In the examined context, Lasso can aid in isolating key clinical predictors of LOS.

- **ElasticNet:** A hybrid of Ridge and Lasso that combines both ℓ_1 and ℓ_2 penalties to balance sparsity and stability. Its loss function is given by:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta^T \mathbf{x}_i\right)^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \right\}.$$

  Alternatively, it is sometimes formulated as:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta^T \mathbf{x}_i\right)^2 + \lambda \left[\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right] \right\},$$

  where $\alpha \in [0, 1]$ controls the mix between the ℓ_1 and ℓ_2 penalties. This approach is particularly effective in capturing interactions within patient data by selecting a subset of relevant features while controlling variance.

- **SVR (Support Vector Regression):** Extends the principles of Support Vector Machines (SVMs) to regression by fitting a function that deviates from the actual targets by no more than a pre-specified ε. It solves:

$$\min_{w,b,\xi,\xi^*} \quad \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*),$$

  subject to:

$$y_i - (w^T \phi(\mathbf{x}_i) + b) \leq \varepsilon + \xi_i,$$
$$(w^T \phi(\mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi_i^*,$$
$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \ldots, n$$

  The slack variables ξ_i, ξ_i^*, along with the regularization parameter C, allow for errors beyond the ε-insensitive zone.

- **KNeighborsRegressor:** A non-parametric, instance-based method predicting outcomes as the (possibly weighted) average of the k nearest neighbors in the feature space.
- **DecisionTreeRegressor:** A non-parametric model that partitions the feature space into regions by recursively splitting the data to minimize the variance (or another impurity measure) of the target within each node. It provides a simple, rule-based structure but risks overfitting if not pruned or ensembled.
- **GradientBoostingRegressor:** An additive model that builds trees sequentially, with each new tree fitted to the residual errors of the ensemble so far. The model prediction is given by:

$$f(\mathbf{x}) = \sum_m f_m(\mathbf{x}),$$

  where each f_m minimizes a differentiable loss function.
- **AdaBoostRegressor:** An ensemble method that iteratively fits weak regressors (often shallow trees) while increasing the weight of instances with larger errors. This approach can, for example, emphasize outlier cases, such as unexpectedly long hospital stays. The final prediction is a weighted sum of the base learners:

$$f(\mathbf{x}) = \sum_t \alpha_t \, h_t(\mathbf{x}).$$

- **RandomForestRegressor:** An ensemble of decision trees trained on bootstrapped samples with random feature selection. Predictions are aggregated by averaging:

$$f(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} f_t(\mathbf{x}),$$

  thereby reducing variance compared to a single tree. This averaging mechanism often yields robust performance in diverse healthcare datasets with many correlated predictors.

– **XGBRegressor:** An implementation of extreme gradient boosting that minimizes a regularized objective function:

$$L(\Theta) = \sum_{i=1}^{n} l(y_i, f(\mathbf{x}_i)) + \Omega(f),$$

where $\Omega(f)$ penalizes model complexity.
Extreme gradient boosting is known for its efficiency and strong performance on large, heterogeneous clinical datasets.

– **LGBMRegressor:** A gradient boosting framework that uses tree-based learning algorithms with a histogram-based approach for efficiency, optimized for speed and lower memory consumption.

– **CatBoostRegressor:** A gradient boosting method designed to handle categorical features natively. It employs techniques such as ordered boosting and target statistics to reduce overfitting.

– **MLPRegressor:** A multilayer perceptron (feedforward neural network) that models non-linear relationships through one or more hidden layers with non-linear activation functions. It is trained via backpropagation to minimize a loss (*e.g.*, mean squared error). This model can capture higher-order interactions among patient variables but may require careful hyperparameter tuning and a sufficiently large dataset for optimal performance.

– **StackingRegressor:** An ensemble technique that combines multiple base regressors by using their predictions as inputs to a meta-regressor. This stacked model is trained via cross-validation to improve overall prediction accuracy.

– **VotingRegressor:** An ensemble method that aggregates predictions from multiple distinct regression models by averaging their outputs.

C.1.2 Classification Models

– **DummyClassifier:** A baseline classifier that predicts a constant label (or follows a predefined distribution), providing a benchmark for evaluating other models. In our pLOS experiments (*e.g.*, LOS $>$ 8 days), it served as a lower bound for meaningful clinical performance.

- **LogisticRegression:** A linear model for binary (or multinomial) classification that models the probability of class membership using the logistic function:

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + \exp\left(-(\beta_0 + \beta^T \mathbf{x})\right)}.$$

For multinomial classification, the softmax function is used to generalize this formulation. Its coefficients can be interpreted to highlight how each patient feature affects the probability of pLOS.

- **SVC / LinearSVC:** Support Vector Classifiers seek the optimal hyperplane that maximizes the margin between classes. For a linear SVM, the decision function is:

$$f(\mathbf{x}) = w^T \mathbf{x} + b,$$

and classification is made by the sign of $f(\mathbf{x})$. Kernelized SVCs extend this approach to nonlinearly separable data by implicitly mapping inputs into a higher-dimensional space, which can be crucial when classifying patients with highly variable clinical trajectories.

- **KNeighborsClassifier:** A non-parametric method that assigns a class label to a new instance based on the majority label among its k nearest neighbors (using, for example, the Euclidean distance). This approach can identify local patterns within patient subpopulations (*e.g.*, specific comorbidities or triage groups).

- **DecisionTreeClassifier:** Constructs a tree by recursively partitioning the feature space according to criteria such as the Gini impurity. The Gini index at a node is defined as:

$$\text{Gini} = 1 - \sum_j p_j^2,$$

where p_j represents the proportion of samples belonging to class j.

- **GradientBoostingClassifier:** Similar to its regression counterpart, this method builds an ensemble of weak learners sequentially, with each subsequent classifier trained to correct the errors of the combined ensemble. The final decision function is given by:

$$f(\mathbf{x}) = \sum_m f_m(\mathbf{x}),$$

with the aggregated output typically transformed (*e.g.*, via a sigmoid or softmax) to yield class probabilities or labels.

– **AdaBoostClassifier:** An adaptive boosting method that iteratively reweights misclassified instances and combines weak classifiers into a strong ensemble using a weighted majority vote. For binary classification, the standard decision function is defined as:

$$f(\mathbf{x}) = \operatorname{sign}\left(\sum_t \alpha_t\, h_t(\mathbf{x})\right).$$

For multiclass classification, the aggregation mechanism may differ. By emphasizing misclassified samples at each iteration, this method can be particularly useful in cases where certain subsets of patients (like those with rare conditions) are repeatedly misclassified in standard models.

– **RandomForestClassifier:** An ensemble of decision trees (each trained on a bootstrap sample and a random subset of features) that predicts the class label based on majority voting. It often excels on healthcare data with numerous partially redundant features.

– **XGBClassifier:** Implements extreme gradient boosting for classification tasks by minimizing a regularized loss function, thereby combining many weak learners (typically decision trees) into a highly accurate classifier.

– **LGBMClassifier:** A gradient boosting classifier based on LightGBM that uses histogram-based algorithms for computational efficiency.

– **CatBoostClassifier:** A gradient boosting classifier optimized for categorical data, employing ordered boosting to prevent target leakage and improve generalization.

– **MLPClassifier:** A multilayer perceptron for classification that consists of one or more hidden layers with non-linear activation functions. It is trained via backpropagation to minimize a loss function (typically cross-entropy).

– **StackingClassifier:** An ensemble method that combines the predictions of several base classifiers by training a meta-classifier on their outputs. A stacking predictor generally performs on par with the best base-layer predictor and may even surpass it by integrating their diverse strengths. For example, in pLOS or ED flow classification, it can unify distinct perspectives (*e.g.*, linear, tree-based, and neural networks) into a more robust predictor. However, its training is computationally intensive.

- **VotingClassifierHard:** An ensemble that assigns the class label based on a majority vote among different classifiers (hard voting), where each classifier contributes an equal or weighted vote. Averaging or combining equally performant models can mitigate individual weaknesses.
- **VotingClassifierSoft:** Similar to hard voting, but instead of class labels, it averages the predicted class probabilities (soft voting) and assigns the class with the highest aggregated probability.

C.2 Hyperparameter Selection and Optimization Process

Hyperparameter optimization represents a critical yet often underreported aspect of developing effective healthcare prediction models. For both the hospital LOS prediction models described in Chapter 4 and the Emergency Department (ED) pLOS prediction models presented in Chapter 5, we implemented a systematic approach to hyperparameter tuning to yield optimal performance while mitigating overfitting.

The hyperparameter search space was formulated based on both domain-specific knowledge and established best practices in the machine learning literature. All optimization procedures were conducted utilizing a grid search methodology combined with stratified k-fold cross-validation ($k = 5$) to ensure robust performance estimation across the heterogeneous patient populations represented in our datasets. This involved defining a parameter grid —for instance, specifying the range of tree depths, learning rates, and number of estimators for ensemble models such as Random Forest and Gradient Boosted Trees— and exhaustively training and evaluating each combination of hyperparameters.

Tables C.1 to C.3 present the hyperparameter search grids explored during GridSearchCV optimization for hospital LOS and pLOS prediction, with optimal values indicated in bold typeface. Values for the Emergency Departments scenario exhibit minimal differences.

Table C.1: Hyperparameter Tuning Summary for Regressors.

Algorithm	**Hyperparameters Grid Search**
XGBoost	{'colsample_bytree': [0.3, **0.7**], 'gamma': [0, 0.01, 0.1], 'learning_rate': [**0.01**, 0.1, 0.5], 'max_depth': [4, **6**, 8], 'n_estimators': [100, 200, **500**], 'sampling_method': ['uniform', 'subsample', **'gradient_based'**], 'subsample': [**0.4**, 0.6], 'tree_method': [**'hist'**], 'eval_metric': [**'rmse'**], 'objective': [**'reg:squarederror'**]}
CatBoost	{'n_estimators': [100, 1000, **2000**], 'learning_rate': [**0.01**, 0.05]}
Ridge	{'alpha': [0.1, 1, 10, 50, 100, **250**, 500, 750, 1000]}
Gradient Boosting	{'n_estimators': [100, **500**, 1000], 'max_depth': [**3**, 5, 7], 'min_samples_split': [2, 5, **7**], 'learning_rate': [**0.01**, 0.1], 'loss': [**'sqrd_error'**, 'abs_error', 'huber', 'quantile']}
LightGBM	{'n_estimators': [**500**, 700, 1000], 'boosting_type': [**'gbdt'**], 'learning_rate': [0.01, **0.05**], 'objective': [**'regression'**]}
Elastic-net	{'alpha': [**0.1**, 1, 10, 100, 1000], 'l1_ratio': [**.01**, .1, .5, .9, .95, 1]}
SVR	{'kernel': [**'linear'**, 'rbf', 'poly'], 'degree': [**2**, 3], 'gamma': [**'scale'**, 'auto'], 'C': [0.01, 0.1, **1**], 'epsilon': [0.01, 0.1, 1, **2**]}
Lasso	{'alpha': [**0.1**, 1, 10, 50, 100, 250, 500, 750, 1000]}
Random Forest	{'n_estimators': [10, **400**, 500], 'criterion': [**'sqred_error'**, 'friedman_mse', 'abs_error', 'poisson'], 'bootstrap': [**True**, False], 'oob_score': [True, **False**]}
K-Nearest Neighbors	{'n_neighbors': [5, 10, **20**, 30], 'weights': ['uniform', **'distance'**], 'algorithm': [**'auto'**, 'ball_tree', 'kd_tree', 'brute']}
AdaBoost	{'n_estimators': [50, 100, **200**, 500, 1000], 'learning_rate': [**0.01**, 0.1], 'loss': [**'linear'**, 'square', 'exponential']}
Multi-Layer Perceptron	{'hidden_layer_sizes': [(512, 256, 128,)], 'activation': [**'relu'**], 'solver': [**'adam'**], 'batch_size': [64, **128**, 256, 512], 'max_iter': [**100**, 200, 300, 400],}
Decision Tree	{'criterion': [**'sqrd_error'**, 'friedman_mse', 'abs_error', 'poisson'], 'splitter': ['best', **'random'**]}

Table C.2: Hyperparameter Tuning Summary for Multi-Class Classifiers.

Algorithm	**Hyperparameters Grid Search**
XGBoost	{'colsample_bytree': [0.3, 0.7, **0.8**], 'gamma': [0, 0.01, **0.05**, 0.1], 'learning_rate': [0.01, **0.1**, 0.5], 'max_depth': [2, **3**, 4, 6], 'n_estimators': [100, **200**, 500], 'subsample': [0.4, 0.6, **0.9**], 'tree_method': [**'hist'**], 'objective': [**'multi:softprob'**]}
CatBoost	{'n_estimators': [1000, 2000, **3500**, 4000], 'learning_rate': [**0.01**, 0.05], 'depth': [**5**], 'loss_function': [**'MultiClass'**]}
Gradient Boosting	{'n_estimators': [100, **400**, 500, 1000], 'max_depth': [**3**, 5, 7], 'min_samples_split': [**2**, 5, 7], 'learning_rate': [0.01, **0.1**], 'loss': [**'log_loss'**, 'exponential']}
LightGBM	{'n_estimators': [**500**, 700, 1000], 'boosting_type': [**'gbdt'**], 'learning_rate': [0.01, **0.05**], 'objective': [**'multiclass'**], 'metric': [**'multi_logloss'**]}
SVC	{'kernel': ['linear', **'rbf'**, 'poly'], 'degree': [2, 3], 'gamma': ['scale', **'auto'**], 'C': [0.01, 0.1, **1.5**]}
Random Forest	{'n_estimators': [10, 100, **400**, 500], 'criterion': [**'gini'**, 'entropy', 'log_loss'], 'bootstrap': [**True**, False], 'oob_score': [**True**, False]}
K-Nearest Neighbors	{'n_neighbors': [5, 10, 20, 30, **40**], 'weights': [**'uniform'**, 'distance'], 'algorithm': [**'auto'**, 'ball_tree', 'kd_tree', 'brute']}
Logistic Regression	{'C': np.logspace(0, 4, 5), 'l1_ratio': np.linspace(0, 1, 5), 'penalty': ['l1', **'l2'**, 'elasticnet'], 'solver': ['lbfgs', 'newton-cg', 'newton-cholesky', 'sag', **'saga'**], 'multi_class': ['ovr']}
AdaBoost	{'n_estimators': [50, **100**, 200, 500], 'learning_rate': [**0.01**, 0.1]}
Multi-Layer Perceptron	{'hidden_layer_sizes': [(512, 256, 128,)], 'activation': [**'relu'**], 'solver': [**'adam'**], 'batch_size': [64, **128**, 256, 512], 'max_iter': [**100**, 200, 300, 400],}
Decision Tree	{'criterion': [**'gini'**, 'entropy', 'log_loss'], 'splitter': ['est', **'random'**]}

Table C.3: Hyperparameter Tuning Summary for Binary Classifiers.

Algorithm	**Hyperparameters Grid Search**
XGBoost	{'colsample_bytree': [0.3, 0.7, **1**], 'gamma': [**0**, 0.01, 0.05, 0.1], 'learning_rate': [0.01, **0.3**, 0.5], 'max_depth': [2, 3, 4, **6**], 'n_estimators': [100, **200**, 500], 'subsample': [0.4, 0.6, **1**], 'tree_method': [**'auto'**], 'objective': [**'binary:logistic'**]}
CatBoost	{'n_estimators': [1000, 2000, **3500**, 4000], 'learning_rate': [**0.01**, 0.05], 'depth': [**5**]}
Gradient Boosting	{'n_estimators': [100, **400**, 500, 1000], 'max_depth': [**3**, 5, 7], 'min_samples_split': [**2**, 5, 7], 'learning_rate': [0.01, **0.1**], 'loss': [**'log_loss'**, 'exponential']}
LightGBM	{'n_estimators': [**500**, 700, 1000], 'boosting_type': [**'gbdt'**], 'learning_rate': [0.01, **0.05**]}
Linear SVC	{'penalty': ['l1', **'l2'**], 'loss': ['hinge', **'squared_hinge'**], 'dual': [**'auto'**], 'C': [0.01, 0.1, **1.5**]}
Random Forest	{'n_estimators': [10, 100, **400**, 500], 'criterion': [**'gini'**, 'entropy', 'log_loss'], 'bootstrap': [**True**, False], 'oob_score': [**True**, False]}
K-Nearest Neighbors	{'n_neighbors': [5, 10, 20, 30, **40**], 'weights': [**'uniform'**, 'distance'], 'algorithm': [**'auto'**, 'ball_tree', 'kd_tree', 'brute']}
Logistic Regression	{'C': np.logspace(0, 4, 5), 'l1_ratio': np.linspace(0, 1, 5), 'penalty': ['l1', **'l2'**, 'elasticnet'], 'solver': ['lbfgs', 'newton-cg', 'newton-cholesky', 'sag', **'saga'**]}
AdaBoost	{'n_estimators': [50, **100**, 200, 500], 'learning_rate': [**0.01**, 0.1]}
Multi-Layer Perceptron	{'hidden_layer_sizes': [(512, 256, 128,)], 'activation': [**'relu'**], 'solver': [**'adam'**], 'batch_size': [64, **128**, 256, 512], 'max_iter': [**100**, 200, 300, 400],}
Decision Tree	{'criterion': [**'gini'**, 'entropy', 'log_loss'], 'splitter': ['best', **'random'**]}

Performance was evaluated using both error-based metrics (*e.g.*, mean absolute error, root mean squared error) and goodness-of-fit measures (*e.g.*,

R^2-score and adjusted R^2-score) in regression analyses. For the prediction of LOS as either a multiclass or binary outcome, performance was assessed using multiple classification metrics, including F1-score, area under the receiver operating characteristic (ROC) curve, and the area under the Precision-Recall curve. These evaluations served as the basis for model selection, guiding the choice among candidate models and informing subsequent hyperparameter tuning. In Chapter 5, this methodological framework was expanded to incorporate iterative refinements in learning rates and layer architectures for neural network–based classifiers, implementing random search in specific instances to traverse larger hyperparameter spaces more efficiently. Throughout the optimization process, we tracked not only predictive accuracy but also training time and model stability.

The final selected configurations were then retrained on the entire training set, ensuring robust parameter estimation prior to deployment within hospital ward and ED settings.

C.2.1 Evaluating Performance under Class Imbalance

In regression tasks —such as predicting LOS as a continuous variable— common error metrics (*e.g.*, MAE, MSE, RMSE, and $\mathbf{R^2}$) provide valuable insights into model performance and are inherently insensitive to issues of class imbalance, as they assess the quality of continuous predictions rather than categorizing outcomes. In contrast, classification tasks under imbalanced conditions require a broader set of evaluation metrics to fully characterize model performance. This was evident in our experiments on classifying prolonged LOS as a binary outcome (0 = *non*-pLOS, 1 = pLOS) in EDs, where a baseline classifier that solely predicted the majority class achieved high accuracy ($\tilde{8}0\%$) yet offered no meaningful information regarding pLOS identification. Instead, precision and recall were found to be more informative.

Precision, defined as the proportion of true positive predictions among all predicted positives, assesses the model's ability to avoid false positives. Conversely, recall (or sensitivity) measures the proportion of actual positive cases correctly identified and is essential for minimizing false negatives.

Our experiments revealed that precision was remarkably relevant for hospital ward admissions, with high precision scores in Gradient Boosting classifiers indicating the model's effectiveness in minimizing false positives. This is critical for preventing unnecessary resource allocation to patients incorrectly flagged as requiring prolonged stays. On the other hand, recall was more significant in the ED setting, where failing to detect a pLOS case could be detrimental, leading to inadequate resource allocation or contributing to overcrowding. Our best-performing models, such as XGBoost, maximized recall by correctly identifying most pLOS cases, albeit at the expense of some false positives.

The F1-score, representing the harmonic mean of precision and recall, emerged as the most balanced metric, especially for ward-level pLOS prediction, where both false positives and false negatives have significant operational implications. However, its usefulness varied depending on clinical objectives: in scenarios where false negatives were more costly (*e.g.*, preventing premature discharges), recall was prioritized over the F1-score.

Cohen's Kappa and the Matthews Correlation Coefficient (MCC) provided additional insights by accounting for chance agreement and class distribution skewness. In our ED experiments, MCC outperformed raw accuracy in distinguishing effective models from misleading ones. Log Loss (or Cross-Entropy Loss) demonstrated particular value in probabilistic modeling, as it penalized misclassifications more heavily when associated with low probability scores. In our ED dataset, lower log loss values correlated with well-calibrated probability scores, aiding clinicians in risk assessment.

Threshold-based evaluation metrics, such as the area under the ROC curve (AUC-ROC) and the area under the Precision-Recall curve (AUC-PR), offered deeper insights into model performance across a range of decision thresholds. While AUC-ROC was useful in relatively balanced scenarios (*e.g.*, hospital ward predictions with a nearly even LOS class distribution), AUC-PR served as a more reliable indicator in highly imbalanced settings like EDs, where prolonged stays were less frequent but critical. The Precision-Recall curve consistently demonstrated the superior performance of ensemble models, particularly XGBoost, LightGBM, and CatBoost, across various threshold settings.

The selection of evaluation metrics for LOS and pLOS prediction should be guided by the relative costs of misclassification. If missing a pLOS case is more harmful than incorrectly flagging a *non*-pLOS patient, prioritizing recall to minimize false negatives is advisable. However, if timely discharge planning is the primary concern, precision may be favored to mitigate risks associated with premature discharge, including insufficient recovery time, increased complication rates, compromised patient safety, and a higher likelihood of readmission. More specifically, in scenarios focused on reducing unnecessary prolonged stays and optimizing bed turnover in hospital wards, emphasis should be placed on precision and the F1-score to prevent excessive false positives. Conversely, ED settings should prioritize recall and AUC-PR to maximize the detection of high-risk cases.

Ultimately, the choice and prioritization of metrics should align with specific clinical and operational objectives, ensuring that predictive models not only achieve statistical accuracy but also support effective decision-making in patient flow management.

C.3 Machine Learning Pipeline Design

In developing predictive models for LOS estimation, a flexible, general-purpose pipeline was designed to systematically process both structured and unstructured data. This modular framework, configurable through parameter settings, encompasses data preprocessing, feature selection, text vectorization, and model training, ensuring a standardized and efficient approach to managing diverse data sources. A Python implementation of the pipeline's construction is presented in Listing C.1.

```python
def create_pipeline(model, is_regression=True, use_diagnosis=False):
    CustomPipeline =
        ImbalancedPipeline if (apply_smote and not is_regression) else Pipeline

    numeric_transformers = CustomPipeline(steps=[('scaler', StandardScaler())])
    categorical_transformers = CustomPipeline(steps=[('onehot', OneHotEncoder())])

    if use_diagnosis:
        text_transformers_steps = []
        if vectorization_type == vectorization_type_tfidf:
            text_transformers_steps.append(('vectorizer', NLTKTokenizer()))
        elif vectorization_type == vectorization_type_fasttext:
            text_transformers_steps.append(('vectorizer', FasttextVectorizer(
   aggregation='mean')))
        elif vectorization_type == vectorization_type_bert:
            text_transformers_steps.append(('vectorizer', BertVectorizer()))
            if apply_pca_to_vectorization:
                text_transformers_steps.append(('pca', PCA(n_components=
   pca_n_components, random_state=random_seed)))
        text_transformers = CustomPipeline(steps=text_transformers_steps)

    transformers = [('num', numeric_transformers, numeric_features),
                    ('cat', categorical_transformers, categorical_features)]
    if use_diagnosis:
        transformers.append(('text', text_transformers, textual_feature))
```

```
    steps = [('preprocessor', ColumnTransformer(transformers))]

    if not skip_select_k_best_features:
        feature_selector_scorer =
            mutual_info_regression if is_regression else f_classif
        steps.append(('select_k_best',
            SelectKBest(score_func=feature_selector_scorer, k=kbest_size)))

    if apply_smote and not is_regression:
        steps.append(('oversampling', SMOTE(random_state=random_seed)))
        #steps.append(('oversampling', ADASYN(random_state=random_seed)))

    steps.append(('model', model))
    pipeline = CustomPipeline(steps=steps)

    return pipeline
```

Listing C.1: Pipeline Construction

During preprocessing, numerical variables are normalized to facilitate comparability across different scales, while categorical attributes are encoded for seamless integration into machine learning models. In addition, a feature selection mechanism identifies the most informative variables based on statistical relevance to optimize predictive performance. For classification tasks, a supplementary step involving the Synthetic Minority Oversampling Technique (SMOTE) can be applied to address class imbalances, thereby improving the model's ability to generalize across diverse patient populations.

The pipeline also provides three mutually exclusive text vectorization techniques, each employing a distinct tokenization strategy. The first approach utilizes the NLTK tokenizer in combination with Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. This method segments text into individual tokens, filters out common stopwords, and standardizes case to enhance consistency. TF-IDF is widely adopted in contexts where a deep semantic understanding is not essential, as its straightforward implementation and interpretability make it a practical choice when context-dependent linguistic nuances are of lesser importance. The NLTK tokenizer implementation is shown in Listing C.2.

```
class NLTKTokenizer(BaseEstimator, TransformerMixin):
    def __init__(self, stopwords=None, punct=None):
        self.stopwords = stopwords or words_to_exclude
        self.punct = punct or set(string.punctuation)

    def transform(self, X):
```

```python
        def tokenize(text):
            tokens = word_tokenize(text, language='italian')
            filtered_tokens = [word.lower() for word in tokens if word.lower() not in
     self.stopwords and word not in self.punct]
            return ' '.join(filtered_tokens)

        return [tokenize(doc) for doc in X]
```

Listing C.2: NLTK Tokenizer

The second approach employs the FastText vectorizer, which generates word embeddings based on a pretrained language model. Unlike TF-IDF, which treats words as discrete entities, FastText represents words as dense vectors in a high-dimensional space, preserving their semantic relationships. This method proves particularly effective in recognizing variations in word forms (*e.g.*, inflection and derivation) and capturing meaningful similarities between related terms; however, while FastText embeddings improve generalization, they do not dynamically adjust word meanings based on the surrounding context. The FastText vectorizer implementation is provided in Listing C.3.

```python
class FasttextVectorizer(BaseEstimator, TransformerMixin):
    def __init__(self, aggregation='mean'):
        self.model = fasttext.load_model('models/fasttext_cc.it.300.bin')
        self.agg_func = np.mean if aggregation == 'mean' else np.sum

    def transform(self, X):
        return np.array([
            self.agg_func(
                [self.model.get_word_vector(word) for word in text.split()],
                axis=0
            ) for text in X
        ])
```

Listing C.3: FastText Tokenizer

The third technique leverages BERT-based vectorization, integrating deep learning to generate context-aware representations of textual data. Through a pretrained transformer model, the BERT tokenizer encodes entire sentences, preserving syntactic structures and dynamically adapting word representations according to their usage within a given context. This enables a more nuanced understanding of linguistic variations, making BERT particularly well-suited for specialized domains, such as medical applications, where terminology often carries highly context-dependent meanings. The BERT vectorizer implementation is included in Listing C.4.

```
class BertVectorizer(BaseEstimator, TransformerMixin):
    def __init__(self):
        self.device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
        self.tokenizer = BertTokenizer.from_pretrained('bert-base-italian-xxl-uncased'
     )
        self.bert_model = BertModel.from_pretrained(
            "bert-base-italian-xxl-uncased", output_hidden_states=True
        ).to(self.device)

    def transform(self, X):
        tokenized_texts = X.apply(lambda x: self.tokenizer.encode(x,
     add_special_tokens=True))
        max_len = max(map(len, tokenized_texts))
        all_embeddings = []

        with torch.no_grad():
            for text, tk_txt in zip(X, tokenized_texts):
                padded_text = np.array(tk_txt + [0] * (max_len - len(tk_txt)))
                attention_mask = np.where(padded_text != 0, 1, 0)

                padded_text_tensor = torch.tensor(padded_text, dtype=torch.long).to(
     self.device).unsqueeze(0)
                attention_mask_tensor = torch.tensor(attention_mask, dtype=torch.long)
     .to(self.device).unsqueeze(0)

                outputs = self.bert_model(padded_text_tensor, attention_mask_tensor)
                embeddings = outputs.last_hidden_state[:, 0, :].cpu().numpy()
                all_embeddings.append(embeddings[0])

        return np.vstack(all_embeddings)
```

Listing C.4: BERT Tokenizer

Within the context of public healthcare, our evaluations indicate that BERT-based vectorization is the most effective method for text representation. Although TF-IDF and FastText offer simpler alternatives that may be suitable for less complex textual data, their inability to capture contextual dependencies limits their effectiveness in tasks requiring precise linguistic comprehension. Nonetheless, the modular design of the pipeline ensures that the most appropriate vectorization technique can be selected based on specific analytical requirements, providing adaptability across different applications.

Appendix D

Deployment of the MondrIAn CEBMS

D.1 HL7 Integration

Modern healthcare landscape is characterized by a multitude of vendor-specific systems, each tasked with managing distinct aspects of patient care and administration. From billing and medication management to clinical documentation and patient tracking, these disparate systems require seamless data exchange to optimize care coordination and system performance.

Health Level Seven (HL7) is a globally recognized suite of standards designed to facilitate the transfer of clinical and administrative health data between software applications [196]. Established by Health Level Seven International in 1987, these protocols operate at the application layer —the seventh layer of the Open Systems Interconnection model— and have been widely adopted by major organizations, including the American National Standards Institute (ANSI) and International Organization for Standardization (ISO). Primary specifications addressing core health data interoperability requirements include:

- *Version 2.x Messaging Standard*: Established in 1989, the HL7 Version 2 (HL7v2) messaging standard supports hospital workflows by defining electronic messages for a variety of administrative, logistical, financial, and clinical processes. Commonly referred to as "Pipehat", this standard uses a non-XML encoding syntax structured by segments (lines) with composite fields, components within composites, and subcomponents within components. The standard utilizes specific delimiters, such as the carriage return for segment separation, vertical bar (|) for field separation, caret (^) for component separator, ampersand (&) for subcomponent separator, hash (#) for truncation separator, and tilde (~) for repetition separator. Each segment begins with a three-character code identifying its type and specifying the category of information contained. Every message has "MSH" as its first segment, which denotes the message type and determines the subsequent segment sequence.

- *Version 3 Messaging Standard*: Developed between 1995 and 2005, HL7 Version 3 (HL7v3) represents a significant advancement in healthcare data standardization. Incorporating a formal methodology known as the Health Development Framework (HDF) and leveraging object-oriented principles, HL7v3 introduces the Reference Information Model (RIM), a static model that structures healthcare data with rigid types and XML-based messaging. Although more complex than its predecessor, HL7v3 offers a secure message format that supports comprehensive healthcare workflows.
- *Clinical Document Architecture (CDA)*: Based on HL7v3, the CDA standard provides a markup language for the structured exchange of clinical documents, including patient summaries, specialist reports, and discharge notes.
- *Fast Healthcare Interoperability Resources (FHIR)*: FHIR is a modern, extensible interoperability standard designed to simplify data exchange using HTTP-based RESTful protocols. It supports JSON and XML formats and incorporates OAuth for secure access and ATOM for efficient data querying.

The choice between HL7 standards depends on organization's specific requirements, existing infrastructure, and strategic objectives. For instance, FHIR's flexibility and ease of use make it suitable for modern web applications, while HL7v2 remains popular in legacy systems because of its simpler structure and lower implementation complexity.

Many healthcare institutions continue to utilize HL7v2 messaging due to its ability to support interoperability among critical healthcare systems, including Patient Administration Systems, Electronic Health Records, and Laboratory and Radiology Information Systems. This standard remains widely supported by all major health informatics vendors, ensuring broad integration. Additionally, the backward compatibility of HL7v2 across its versions allows for continuity as the standard evolves.

D.1.1 ADT Messaging System

HL7 defines a range of message types tailored to various healthcare functions, such as ACK (General Acknowledgment), ADT (Admit, Discharge, Transfer), MDM (Medical Document Management), ORM (Pharmacy/treatment Order Management), and QRY (Query).

ADT messages, integral to HL7v2, facilitate essential notifications for patient admissions, transfers, and discharges, supporting patient care alignment across healthcare providers. These messages inform relevant personnel of a patient's status, improving follow-up and communication, especially for patients with complex or chronic conditions. ADT messages help track frequent healthcare users, guiding them toward appropriate clinical and non-clinical interventions to reduce unnecessary emergency visits and hospital readmissions.

Key ADT message types include ADT-A01 (patient admission), ADT-A02 (patient transfer), ADT-A03 (patient discharge), ADT-A04 (patient registration), ADT-A08 (patient information update), and ADT-A12 (cancel patient transfer).

Each message type serves a distinct function, ensuring efficient information exchange for timely and coordinated care. For example, an ADT-A08 message, used for patient information updates, alerts systems to changes in a patient's address or name. The ADT-A04 message, related to patient registration, notifies systems when a patient arrives as an outpatient.

MondrIAn leverages HL7 standards (either Version 2 or Version 3) to achieve integration with hospital information systems. Notably, integration with the ADT system is bidirectional, enabling automatic notifications when bed managers determine to reserve beds in advance or perform patient pre-assignments.

Examples of relevant messages follow below.

```
MSH|^~\&|ACCEWEB|HITECH|MONDRIAN|MILIARIS|202411050848||ADT^A01^ADT_A01|373090602|
      P|2.5|||||8859/1|||
EVN||202411050848|||202411050848|
PID|||90616291^^^HITECH^PI^BDA^~166CC383^^^^SS^~20241007~CODICEFISCALE994^^^
      Ministero Finanze^NN^~20241007~^^^^PNT^^^^^^^^NPI^^^^&&^&~||ROSSI^MARIA||
      19570304|F|||^^MODENA^MO^^IT^BDL^^029033~&V. C. A. DALLA CHIESA&12^^MODENA^MO
      ^41100^IT^L^^036023~&V. C. A. DALLA CHIESA&12^^MODENA^MO^41100^IT^H^^036023~
      ^^^^^^M^^~||^PRN^PH^^^^^^^^^3381234567~^PRN^CP^^^^^^^^^3401234567~^NET^
      Internet^^^ORN^FX^^^^^^^^^|||||||||||||IT^^100^ITALIA||IT^^100^ITALIA|||N|
      MEF@20200806~NAR@20241007~||NAR|||||
PV1||I|5141^^^030701^^^^^^^33F1^Blocco NORD - Settore F - Piano 1^|1|^^^ADT^VN||
      CODICEFISCALE677^DOTTORE^MARIA^^CODICEFISCALE677^^^^^^^^NN||||||3|||^^^^^^^^
      ^^^^LR||2024052248^^^ADT^MR^030701|1||||||||||||||||||||||||202411050848||||
      ||V|
PV2|||||||||||||||||||||||||||||||||||||||3^RIC.PROGRAMMATO^SAS^|||||||||||
```

Listing D.1: Patient Admission Message (ADT-01).

```
MSH|^~\&|ACCEWEB|HITECH|MONDRIAN|MILIARIS|202411050828||ADT^A03^ADT_A03|373082692|
      P|2.5|||||8859/1|||
EVN||202411050828|||202411050800|
PID|||90641084^^^HITECH^PI^BDA^~319JM614^^^^SS^~20240917~CODICEFISCALE205^^^
      Ministero Finanze^NN^~20240917~^^^^PNT^^^^^^^^NPI^^^^&&^&~||ROSSI^MARIO||
      19630904|M|||^^MODENA^MO^^IT^BDL^^015146~&VIA P. GIARDINI&67^^MODENA^MO
      ^41100^IT^L^^036023~&VIA P. GIARDINI&67^^MODENA^MO^41100^IT^H^^036023~^^^^^^^M
      ^^~||^PRN^CP^^^^^^^^^059217079~^NET^Internet^^^ORN^FX^^^^^^^^^|||||||||||||IT
      ^^100^ITALIA||IT^^100^ITALIA|||N|MEF@20230104~NAR@20240917~||NAR|||||
PV1||4|2621^^^030701^^^^^^^33D0^Blocco NORD - Settore D - Piano T
      ^|1||||||||||3||||2024807549^^^ADT^MR
      |||||||||||||||||||||||||202410300700|202411050800||||2024807549^^^ADT^MR|V|
PV2|||||||||||||||||||||||||||||||||||||||3^RIC.PROGRAMMATO^VIG^|||||||||||
DG1|||||||||||||||||||||
```

Listing D.2: Patient Discharge Message (ADT-03).

```
MSH|^~\&|ACCEWEB|HITECH|MONDRIAN|MILIARIS|202411050847||ADT^A02^ADT_A02|373090304|
      P|2.5|||||8859/1|||
EVN||202411050847|||202411050847|
PID|||32180188^^^HITECH^PI^BDA^~169YP975^^^^SS^~20240611~CODICEFISCALE441^^^
      Ministero Finanze^NN^~20240611~^^^^PNT^^^^^^^^NPI^^^^&&^&~||ROSSI^MARIO
      ENRICO||19631109|M|||^^MEDOLLA^MO^^IT^BDL^^41036~&VIA CADUTI&160,&^^MEDOLLA^
      MO^41036^IT^L^^036021~&VIA CADUTI&160^^MEDOLLA^MO^41036^IT^H^^036021 ~^^^^^^M
      ^^~||^PRN^PH^^^^^^^^^3201234567~^PRN^CP^^^^^^^^^3201234567~^NET^Internet^
      EMAIL@LIBERO.IT~^ORN^FX^^^^^^^^^|||||||||||||IT^^100^ITALIA||IT^^100^ITALIA
      |||N|MEF@20230207~NAR@20240611~||NAR|||||
PV1||I|3713^^^030701^^^^^^^^|4||4001^325^757^030701^^^^^^^32B2^Blocco SUD -
      Settore B - Piano 2^||||||||3||||2024050126^^^ADT^MR||||||||||||||||||||||||
      |202411050847|202411050847||||||V|
PV2|||||||||||||||||||||||||||||||||||||||3^RIC.PROGRAMMATO^CAR^|||||||||||
```

Listing D.3: Patient Transfer Message (ADT-02).

```
MSH|^~\&|ACCEWEB|HITECH|MONDRIAN|MILIARIS|202411050848||ADT^A08^ADT_A01|373090650|
      P|2.5||||||8859/1|||
EVN|A01|202411050848||||202411050848|
PID|||90616291^^^HITECH^PI^BDA^~166CC383^^^^SS^^20241007~CODICEFISCALE994^^^
      Ministero Finanze^NN^^20241007~^^^^PNT^^^^^^^^NPI^^^^&&^&~||ROSSI^MARIA||
      19570304|F|||^^MIRANDOLA^MO^^IT^BDL^^029033~&V. PUNTA&35^^MIRANDOLA^MO^41037^
      IT^L^^036022~&V. PUNTA&35^^MIRANDOLA^MO^41037^IT^H^^036022~^^^^^^^M^^^||^PRN^
      PH^^^^^^^^^3381234567~^PRN^CP^^^^^^^^^3401234567~^NET^Internet^~^ORN^FX^^^^^^
      ^^^|||||||||||||IT^^100^ITALIA||IT^^100^ITALIA|||N|MEF@20200806~NAR@20241007
      ~||NAR|||||
PV1||I|5141^^^030701^^^^^^^33F1^Blocco NORD - Settore F - Piano 1^|1|2024107989^^^
      ADT^VN||CODICEFISCALE677^DOTTORE^MARIA^^CODICEFISCALE677^^^^^^^^NN|||||||3|||
      ^^^^^^^^^^^^LR||2024052248^^^ADT^MR^030701|1|||||||||||||||||||||||
      202411050848||||||2024107989^^^ADT^MRT|V|
PV2|||||||||||||||||||||||||||||||||||||3^RIC.PROGRAMMATO^MIR^|||||||||||
```

Listing D.4: Patient Data Modification Message (ADT-08).

D.2 Deployment Workflow

Deploying web applications in healthcare requires careful consideration of security, scalability, and integration with existing technological infrastructures. The MondrIAn Cooperative Electronic Bed Management System (CEBMS) is hosted in a data center of the Regione Emilia-Romagna. The application serves multiple healthcare institutions through a multi-tenant design, where each hospital maintains isolated data while sharing computational capacity. This approach optimizes resource utilization while safeguarding data segregation and security.

A multi-tiered architecture separates concerns between client interaction, business logic, and data persistence (Figure D.1).

The Django web framework provides the foundation for the application, offering built-in security features, authentication and authorization mechanisms, a first-class Object-Relational Mapper (ORM) for database interactions, and a template engine for dynamic content generation. While Django includes an embedded web server for development purposes, its use in production environments is strongly discouraged due to limitations in performance and security. MondrIAn employs Gunicorn for handling requests. Gunicorn is a Python Web Server Gateway Interface (WSGI) that supports concurrent processing through multiple parallel workers and integrates seamlessly with various web frameworks, including Django. In this deployment, Gunicorn is configured with Uvicorn workers to provide compatibility with the Asynchronous Server Gateway Interface (ASGI), allowing the system to handle asynchronous protocols effectively, such as those required for the chat application embedded within the CEBMS. However, as Gunicorn is not designed to handle all functionalities of a traditional HTTP server —such as serving static assets efficiently— the stack incorporates NGINX on top as a high-performance HTTP server. NGINX acts as a reverse proxy with SSL termination and a load balancer, efficiently routing client requests to backend servers and distributing traffic to ensure system reliability. User requests are first routed through NGINX, which then forwards them to Gunicorn, where they are processed by the Django application.

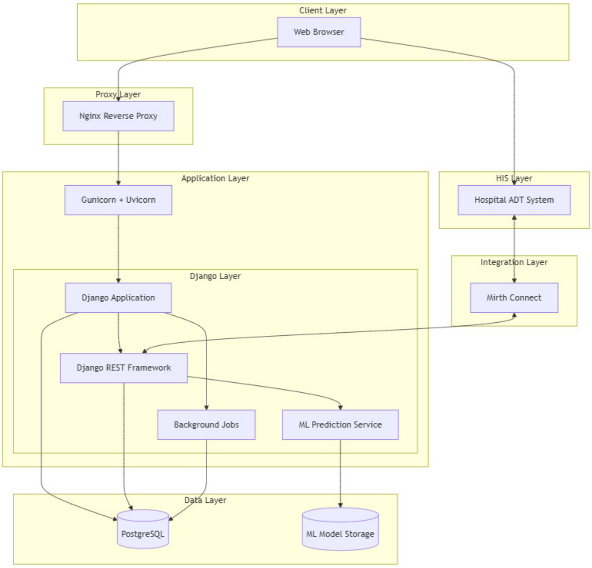


Figure D.1: Layered Architecture of the MondrIAn CEBMS.

Data persistence is handled by PostgreSQL, a highly reliable and scalable relational database management system (RDBMS). PostgreSQL is renowned for its robust data storage, transactional integrity, and advanced features such as table partitioning and native support for JSON data types, making it an ideal choice for complex data management needs. To further optimize system performance, REDIS is employed as an in-memory key-value data store. This caching mechanism significantly reduces database load and accelerates response times, ensuring a smooth user experience.

Python virtual environments are used to isolate dependencies, avoiding conflicts between packages and promoting maintainability throughout the system's lifecycle.

Security measures implemented encompass SSL/TLS encryption for all data in transit, regular security updates and patch management, network segmentation and firewall configuration, role-based access control (RBAC) for multi-tenant isolation. User authentication is managed via integration with the corporate LDAP system.

Business logic customization is supported through a dedicated layer responsible for synchronizing, processing, and mapping data across user interfaces and hospital information systems. This design facilitates a clear separation between the view and data layers while allowing adjustments to accommodate institutional differences in data aggregation and processing.

Integration with hospital ADT and ED systems is achieved through Mirth Connect, an open-source integration engine widely used in healthcare. Mirth Connect allows healthcare systems —such as EHR, HIS, and Laboratory Information Systems (LIS)— to communicate and exchange data by transforming and routing messages. It supports multiple versions of HL7 (*e.g.*, HL7v2) and can handle both inbound and outbound messages, ensuring full interoperability between platforms. It also provides tools to map and transform data from one representation to another, facilitating integration between systems that use different data formats (*e.g.*, converting HL7 messages to XML, JSON, or other formats as required by the receiving party) and employs a scripting language (JavaScript) to perform complex transformations, custom logic, and business rule applications on the data as it passes through. Mirth Connect can route messages based on conditions (*e.g.*, message type, patient identifier) to the appropriate destination system, organizing workflows into channels, each defining data sources, transformations, routing, and destinations. This modular structure allows healthcare organizations to manage complex integrations by configuring different workflows for different message types and data streams. Mirth Connect is highly customizable and can be scaled to manage large volumes of messages, which makes it suitable for both small clinics and large hospital networks. The MondrIAn platform leverages Django REST framework to expose RESTful API endpoints for system integration through Mirth protocols (Figure D.2).

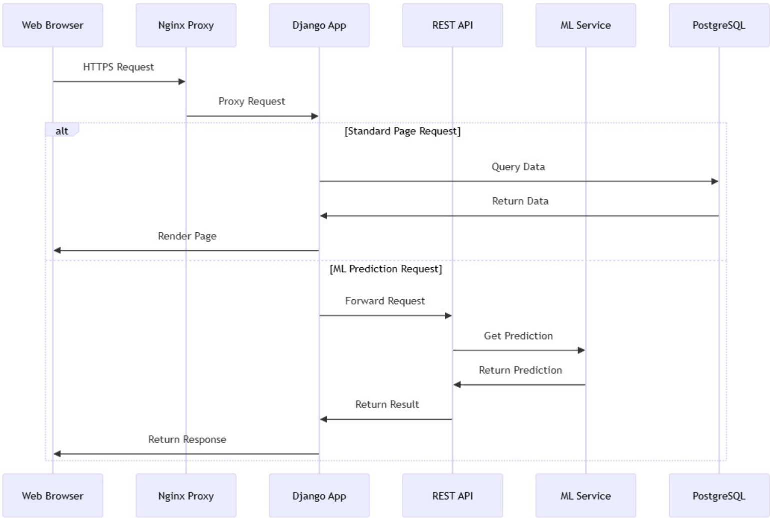


Figure D.2: Sequence Diagram of Standard and LM Prediction Requests in MondrIAn.

The Cooperative EBMS incorporates a machine learning model —implemented using the scikit-learn library— to predict the expected patient length of stay. Upon assigning a patient to a bed, this model is invoked through a dedicated API to estimate the most likely LOS category (1-3 days, 4-10 days, or over 10 days) based on a range of features, including patient demographics, clinical condition, and admission-related factors. This predictive capability empowers bed managers to proactively plan discharges, optimize bed allocation, and streamline patient flow. A scheduled task regularly updates predictions to account for changes in the patient's condition, thereby assisting the bed management team in refining discharge planning accordingly.

In this study, the BentoML framework was selected to serve the CatBoost ML model employed by MondrIAn. BentoML is an open-source platform designed to streamline the process of serving and deploying ML models. It offers robust support for Docker containerization and enables inference models to be exposed as RESTful APIs, ensuring seamless integration with web applications.

The ML deployment process involved several stages:

- *Model Development.* The CatBoost library, integrated within scikit-learn, was employed to develop and train the base model.
- *Service Encapsulation.* The serialized model was then integrated into the BentoML framework, which natively supports CatBoost, and wrapped within a service.

```python
import bentoml
import pickle

# Load the trained model
with open("los_model_v20240911.pkl", "rb") as f:
  model = pickle.load(f)

# Save the model to BentoML model store
bento_model = bentoml.catboost.save_model("los_model", model)
```

Listing D.5: Model Versioning and Serialization into Store.

- *API Endpoint Creation.* An API endpoint was created to handle incoming HTTP POST requests and serve model predictions.

```python
from bentoml import Service
from bentoml.io import JSON
import bentoml.catboost

# Load the model from BentoML model store
svc = Service(name="los_service")

@svc.api(input=JSON(), output=JSON())
def predict(input_data):
  # Perform prediction using the model
  return model.predict(input_data)
```

Listing D.6: Model Encapsulation within a Service.

- *Containerization.* The API service, along with all necessary dependencies and configuration files, was packaged into a Docker container, ensuring portability and scalability of the deployed model.

```
bentoml containerize los_service:latest
docker run -p 5000:5000 los_service:latest
```

– *Deployment.* The Docker container was deployed to expose the RESTful API interface for accessing model predictions, allowing the Django backend to send input data via HTTP requests and receive predictions in response.

```
import requests
from django.http import JsonResponse

def get_prediction(request):
  input_data = request.data  # input data is provided in the POST request
  response = requests.post("http://localhost:5000/los_predict", json=
   input_data)
  return JsonResponse(response.json())
```

Listing D.7: API Invocation.

This modular architecture decouples the model service from the application backend, enhancing maintainability and scalability.

BentoML's alignment with Machine Learning Operations (MLOps) principles facilitated the seamless implementation of the following best practices:

– *Model Management and Versioning.* Effective model management relies on structured repositories with built-in version control, enabling the storage of model iterations alongside associated metadata. Versioned models guarantee that updates can be controlled and rolled back if necessary, supporting Continuous Integration and Continuous Deployment (CI/CD) processes in machine learning.

– *Automated Deployment.* Generating deployable artifacts, such as Docker containers with pre-configured dependencies, ensures consistency and efficiency in MLOps pipelines. Automating these workflows allow for the rapid deployment of updated models, meeting the evolving requirements of data-driven systems.

– *API-Based Model Serving and Scalability.* Machine learning models can be deployed as RESTful APIs to facilitate integration with external applications and services. Deployment to scalable infrastructures, such as Kubernetes, secures high availability and efficient load management, leveraging the system's ability to handle production-scale traffic effectively.

- *Monitoring and Logging.* Continuous monitoring and logging are integral to maintaining model performance in production environments. Logs capturing inference times, system metrics, and input-output data provide valuable insights into operational behavior. Such capabilities enable the detection of issues like data drift or model degradation, ensuring timely interventions to maintain reliability and accuracy.
- *Continuous Integration (CI) and Model Retraining.* As new data becomes available or operational conditions evolve, machine learning models require frequent updates to remain relevant. Automated retraining workflows, integrated into CI/CD pipelines, streamline the process of incorporating new data and validating updated models. This approach delivers consistent performance and flexibility in response to changing operational requirements.
- *Flexible Deployment Options.* BentoML is compatible with various deployment targets, including cloud platforms, on-premises environments, and edge devices. This flexibility aligns with MLOps requirements for adaptable deployment strategies that can be customized based on the application context, data privacy concerns, and latency requirements.

By adopting these MLOps practices, MondrIAn ensures its model's robustness, scalability and adaptability, which are essential for accurate patient management in dynamic healthcare settings across interconnected facilities.