UNIVERSITY OF TWENTE.

Kurtosis

N. Boersen

Optimizing Resource Allocation for Outpatients: Machine Learning-Based Length-of-Stay Predictions and Patient Scheduling

Supervisors
Dr. ir. A.A.M. Spil, Dr. A. Asadi, Msc V.J.J. van Ham

2023

Master Thesis

COLOPHON

MANAGEMENT
Behavioural, Management and Social Sciences
Industrial Engineering and Management

DATE
June 22, 2023

REFERENCE
Master's Thesis IEM

VERSION
1.0

STATUS
Final

PROJECT
Master's Thesis IEM

STUDENT NUMBER
1841025

AUTHOR(S)
N. Boersen

TELEPHONE
+31 6 4172 1365

EMAIL
info@niekboersen.com

POSTAL ADDRESS
P.O. Box 217
7500 AE Enschede

WEBSITE
www.utwente.nl

FILENAME
Boersen_MA_BMS.pdf

Preface

It is with pleasure and a sense of accomplishment that I present this master thesis, which focusses on the realms of increasing bed occupancy at the daycare facility at the small regional hospital. This research has been a journey of exploration, learning about numerous new topics, and, above all, meeting new people. Finally, I had the opportunity to apply the knowledge I have acquired over the past few years in practise.

The healthcare industry continues to improve to provide better and more sophisticated care to patients. In my opinion, helping this industry with my knowledge was not only a professional challenge, but also a personal commitment to making a positive impact on people's lives. This thesis allowed me to contribute to the field by developing a comprehensive methodology that can actually enhance hospital operations, resource allocation, and patient scheduling.

I would like to express my gratitude to all the highly skilled and kind professionals who have supported me throughout this journey. Their guidance, expertise, and willingness to share their knowledge have been invaluable. I am particularly grateful to my supervisors Ton Spil and Amin Asadi for guiding me through the process. Their expertise and guidance helped shape this master thesis. I especially liked the trust towards the end in the making the thesis to a good end. I would also like to extend my gratitude to Vincent van Ham, my supervisor from Kurtosis. His support and advice, particularly in navigating hospital practises, kept my thesis grounded in practicality and relevance. His knowledge and expertise have greatly contributed to shaping the direction of this thesis and ensuring its practical applicability. I am grateful to have him as support and the opportunity to work alongside him throughout this journey. At the hospital, I met numerous people who helped me throughout the process and whom I all want to thank. Someone I want to thank in particular is Ellen van Zalen. She was always willing to explain everything to me in the hospital every day, which greatly helped. Lastly, I want to thank my friends and family for their support throughout this journey. Their encouragement, understanding, and willingness to think along have been a constant source of motivation.

This thesis marks the end of my study and my student life, which I thoroughly enjoyed. I started with Creative Technology at the Faculty of Electrical Engineering, Mathematics, and Computer Science. I am convinced that this study was an amazing start to my scientific education. After a gap year, I decided to switch to the master Industrial Engineering & Management at the Faculty of Behavioural, Management and Social Sciences, which fitted my interests even better. I hope that the completion of this thesis is the beginning of a career in the field of healthcare optimisation.

Once again, I express my gratitude to all those who have supported me throughout the journey. This thesis is a testament to our collective efforts, and I am honoured to have had such an amazing time.

Niek Boersen

Abstract

With global healthcare demands intensifying, optimizing healthcare processes and resource allocation is vital for sustaining hospital competitiveness and meeting healthcare needs. This article proposes a machine learning-based approach to enhance bed occupancy via length-of-stay predictions and patient scheduling.

The literature study encompasses two systematic literature reviews on length-of-stay predictions and patient scheduling. The first review underlines the growing adoption of machine learning techniques in length-of-stay predictions. It further notes a shift in predictive modelling trends, where these techniques display beneficial performance when compared to traditional methods. Secondly, the literature review on patient scheduling highlights integer linear programming as a commonly used method.

The chosen methodology comprises two stages: machine learning-based length-of-stay predictions and ILP patient scheduling. Initially, various machine learning models are evaluated to identify the most effective performers. An integrated approach, combining regression and classification, is employed to ensure accurate and reliable length-of-stay predictions. The outputs of the regression model are validated using an independent operating classification model to verify the obtained results.

Artificial neural networks emerged as the superior regression model, achieving an R^2score of 0.776 and a mean average error of 55.1. For classification, Random Forest exhibited the highest average accuracy of 77.20%. In the subsequent stage, the ILP patient scheduling method demonstrated remarkable effectiveness in optimizing bed occupancy, with the potential to significantly increase it from 1.47 to 3.33 or even higher. Additionally, a user-friendly graphical user interface was developed to seamlessly integrate all models and provide valuable support to hospital planners seeking to enhance bed occupancy rates.

This study presents valuable contributions to both theory and practice by introducing a comprehensive and innovative approach. It introduces a novel validation method that combines regression and classification, which has not been explored in the existing literature. Furthermore, the integration of length-of-stay predictions from patients on the waiting list as a basis for patient scheduling represents a significant advancement that has not been previously investigated. These research findings fill critical gaps in the current knowledge and offer promising avenues for future advancements in healthcare optimization.

List of figures

List of tables

List of abbreviations

Abbreviations	Explanation
LOS	Length of Stay
RR	Ridge Regression
RF	Random Forest
SVM	Support Vector Machines
DT	Decision Tree
LR	Linear Regression
GB	Gradient Boosting
KNN	K-Nearest-Neighbour
NN	Neural Network
COTG	Health product declaration code (Type of operation)
AT	Anaesthesia technique
ASA	Metric to determine whether someone is healthy enough to tolerate surgery and anaesthesia.
Specialisations	
ORT	Orthopedics
PLA	Plastic surgery
CHI	Surgery
PYN	Pain management
KNO	Otorhinolaryngology
URO	Urology
CAR	Cardiology
DER	Dermatology

1 Introduction

The efficient allocation of healthcare resources is crucial in modern healthcare systems. Resources such as medical personnel, equipment, and facilities have limitations, and the growing demand for healthcare services further emphasises the importance of optimal resource allocation. Effective bed allocation is a critical aspect within hospitals, as it directly impacts the treatment capacity of patients. The availability of beds determines the number of people who can receive medical care, which makes the efficient utilisation of these resources of the utmost importance. Optimising patient allocation in the same number of beds improves bed occupancy rates within hospitals, maximising the utilisation of these important resources.

Healthcare care provided in western European countries is highly advanced and complicated and is becoming more advanced and expensive every year. This is no different in the Netherlands, and expenses are taking up more of the gross domestic product every year, rising from 10.0% in 2000 to 14.5% in 2021. (*Zorguitgaven; Kerncijfers*, 2022)*. The Netherlands National Institute for Economic Policy Analysis (CBP) estimates that this figure will increase to 18% of the total gross domestic product in 2060 (CPB, 2022).

The increase in these figures can be attributed in part to the increase in hospital expenses, as noted by Vonk et al. (2020). The increase in hospital expenses is due to various factors, including intensification of care, improved diagnostic techniques leading to earlier care, more costly care, and demographic ageing, which will result in higher healthcare costs for the government in the future (Vonk et al., 2020). In addition to increasing costs, there is also the problem of decreasing the number of medical personnel. According to the Dutch Social Economic Council (SER), additional 700,000 medical employees will be required in the healthcare sector over the next 20 years to maintain the current level of quality of care (Sociaal-economische Raad, 2020). In 2020, one in seven members of the working population is employed in healthcare care, with a projected increase to one in four by 2040 if no intervention is made (Raad, 2021). Rising costs and a shortage of medical personnel will pose a challenge for both local authorities and the government. Partial solutions to these challenges lie in a better use of existing resources. Research on sophisticated planning and capacity management is necessary to improve hospital efficiency, allowing the successful treatment of more patients with the same resources.

With the increasing demand for healthcare services, the hospital is under pressure to optimise its resources, including bed occupancy. Bed occupancy is the fraction of patients per bed per day and is a metric that indicates how well resources are allocated. Scheduling multiple patients in a single bed requires an optimised planning strategy. The current patient scheduling process faces several challenges that result in inefficient use of scarce resources. This research aims to elucidate these challenges and provide information on the urgent need for a more effective bed occupancy planning strategy in a small regional hospital.

In the hospital, the daycare department plays a critical role as a dedicated recovery area for patients who have undergone various medical operations. Currently, patients who arrive from the operating room are moved to the A2X department to recover from surgery. The department is equipped with a limited number of beds. Daycare departments only hospitalise outpatients, which are patients who are expected to leave the same day as arriving. The length of stay experienced by these patients can show significant variations, influenced by factors such as type of operation, individual recovery rates, specific medical conditions, and different types of anaesthesia. Patients arriving at the daycare department occupy a bed. It is imperative to allocate beds efficiently to ensure optimal patient care and resource utilisation. Furthermore, the time of patients in the hospital encompasses four distinct phases. The preoperative phase involves the arrival of patients and the necessary preparations for their procedures. Subsequently, the operation phase occurs, where the medical procedure is performed. After the operation, the patients enter the recovery phase, during which their vital signs are closely monitored until consciousness is restored. Finally, patients progress to the department phase, where they recover from their respective operations with the diligent care provided by the medical staff. The sum of time of all the above-mentioned phases is called the length of stay of a patient. In Figure 1, a schematic overview of the length of stay structure is shown.

Length of stay

Preoperative time	Operation time	Recovery time	Department time

Figure 1. Schematic overview of the length of stay

The existing planning process begins with the patient being placed on a waiting list by a physician or the administration department. The waiting list contains all patients who are currently waiting for a medical procedure. Subsequently, the operating room (OR) planning department reviews the list and selects patients for surgery. However, OR planners face significant challenges in determining which patients can be planned to increase bed occupancy, as they lack crucial information on the expected length of stay for individual patients.

The lack of reliable length-of-stay predictions poses several issues for the planning department. First, it hinders their ability to accurately forecast bed availability with potential bottlenecks in patient flow and increased waiting times for surgery. This, in turn, could result in lower patient satisfaction and compromise the overall quality of care provided by the hospital. Second, a lack of accurate length-of-stay predictions can lead to suboptimal resource allocation. OR planners often schedule surgeries for patients with short stays as if they need a bed for the whole day. In addition, they can schedule two patients with long LOS on a single bed, resulting in capacity problems at the end of the day. Such an inefficient allocation of resources can worsen the already stretched capacity in the day care department of the hospital, affecting the hospital's ability to meet the growing demand for healthcare services in the region. Third, the manual nature of the current bed allocation process not only increases the likelihood of human error, but also imposes a considerable workload on the OR Planning department. Since employees are scarce in the healthcare sector, efficient working processes are crucial to ensure high job satisfaction rates.

Given these challenges, it is evident that the small regional hospital needs to gather more insiders in its current planning process, as well as provide them with tools to make informed decisions. Implementing a data-driven prediction model for length-of-stay predictions could address these issues by allowing the OR Planning department to make these informed decisions regarding bed allocation based on accurate patient recovery times. By incorporating such a model into the hospital planning process, the small regional hospital can optimise its resources and ensure that it continues to provide high-quality healthcare services to the growing and ageing population of the Netherlands.

To address the challenge of increasing bed occupancy, an integrated system of length-of-stay predictions and patient scheduling will be developed. In order to tackle the problem systematically, a main research question and sub-research question are established. This approach will help break down the complex problem into smaller, manageable pieces, facilitating the development of a tool for hospital planners to increase the bed occupancy.

Main research question:

How can patient scheduling algorithms and optimisation techniques be integrated with length-of-stay predictions to effectively allocate resources and maximise bed occupancy?

Sub-research questions:

Current Bed Allocation Process and Management
- Who is responsible for managing the bed allocation process and what factors are considered when making allocation decisions?
- What are the current work processes for arranging beds for patients?

Length-of-stay predictions
- What factors influence the length of stay for patients and how can they be predicted or estimated?
- What models or algorithms are used to generate accurate length-of-stay predictions for outpatients, what are their capabilities, and how do they compare?
- How can length-of-stay predictions be incorporated into the bed allocation decision-making process, and how can they help solve bed allocation decision problems?
- How can the robustness and reliability of length-of-stay predictions be ensured?

Patient scheduling
- What are the most used techniques in patient scheduling?
- How can patient scheduling be used to increase bed occupancy?

Implementation
- How can the developed system be implemented in the small regional hospital?
- What challenges will be faced during implementation?

The research question and sub-research questions serve as a guide of the research process. The answers are obtained through a thorough examination of the relevant literature, data analysis, model development and evaluation, and scheduling model. In addition, a practical tool will be developed to enable the planning department to effectively implement the research findings in their real-world planning workflow.

2 Literature review

2.1 Introduction

The literature review will examine the current state of knowledge on the predictions of length of stay and patient planning. The review aims to identify gaps in the existing literature and provide information on potential solutions that can improve the efficiency of bed allocation processes. The review will explore the various factors that influence bed occupancy, length of stay, allocation of resources, and planning models. By analysing the relevant literature, the review will provide a basis for the proposed solution, a tool that helps OR planners increase bed occupancy at the small regional hospital.

The final product will be an integrated system of patient care and length-of-stay predictions. In this chapter, two systematic reviews of the literature will be conducted according to the PRISMA guidelines (Moher et al., 2009). The first systematic literature is on prediction of length of stay for hospitalised patients. Studies will be evaluated and the findings will be summarised to serve as a solid foundation for this research. The second systematic review of the literature is on patient planning. Various approaches are known in the field of patient planning. The systematic review of the literature will reveal the methodologies that occur most frequently and will help guide research.

In general, the literature will provide a logical progression of the topics. The review will serve as the foundation for the proposed data-driven solution, which aims to develop a prediction model for LoS and a planning system to improve the allocation of beds in the hospital.

2.2 PRISMA Systematic literature review | Length of Stay Predictions

This section presents a meticulous and comprehensive literature review focusing on the predictions of length of stay. It offers a concise summary of significant findings and notable advancements in this area of research.

2.2.1 Introduction

To ensure optimal levels of care, healthcare systems have started to place a growing emphasis on effective resource management and forecasting. The desired result is to minimise associated costs and improve patient care (Garg et al., 2012). Effective resource management can be aided by accurate prediction of the length of stay of patients. The length of stay is a healthcare metric that is used to determine the duration, in terms of days or hours, that a patient is expected to stay in the hospital after a single admission event (Huntley et al., 1998). Accurate predictions of length of stay are crucial for hospitals, as they help with proper management of hospital capacity, quality, and efficiency (Tibby et al., 2004; Weissman et al., 2007). On the contrary, inaccurate predictions can lead to extended hospitalisation of patients, which can result in dissatisfaction for both the patient and healthcare workers (Lequertier et al., 2021).

Length-of-stay predictions can be challenging because the population groups of patients are heterogeneous (Huntley et al., 1998; Schmidt, Geisler, Spreckelsen et al., 2013). There is a substantial body of literature that has investigated various methods of predicting length of stay, including physician assessments. Unfortunately, physician evaluations are poorly reliable in many cases due to the lack of background information on patients and heterogeneous opinions of healthcare professionals (Durstenfeld et al., 2016; Nassar & Caruso, 2016). Therefore, it is crucial to develop models that can empower the predictions of medical personnel.

Many articles describe techniques such as regression (statistical) techniques (Combes et al., 2014; Grampurohit & Sunkad, 2020) and artificial intelligence (AI) methods (Bacchi et al., 2022; Kadri et al., 2022; Mansoori et al., 2023; Mekhaldi et al., 2020).

Lequertier et al. (2021) conducted a systematic review of the literature on the use of these different techniques and found that the prediction of length of stay is moving towards the use of more sophisticated methods such as machine learning. However, the validity of these models is difficult to verify due to the challenges in reproducing the research findings (Lequertier et al., 2021). Machine learning techniques are superior for complex pattern findings and in cases with a large number of input data. (Bzdok et al., 2018; Bacchi et al., 2022).

Machine learning has increased in recent years due to technical advancements (Lequertier et al., 2021), availability of big data, and their outstanding performance compared to other models (Kuwajima et al., 2020; Le et al., 2011). Predicting length of stay of a patient is a regression problem for which machine learning can be a fitting solution (Ray, 2019). For this reason, the focus will be on state-of-the-art machine learning techniques to successfully predict the length of stay at the small regional hospital.

In the remainder of this chapter, the aim is to explore the field of length-of-stay predictions for hospital patients. An elaborate review of the existing literature will be conducted over the last couple of years. The literature will be reviewed according to the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). By conducting a comprehensive search across multiple databases, including Scopus, PubMed, and Web of Science. The objective of the review is to examine different approaches and methodologies used in the prediction of the length of stay of patients. This includes identifying the factors that influence LOS, information about model use, and evaluation methodologies.

The systematic review of the literature will revolve around predictions of length of stay for hospitalised patients. The research question will be

What are the most widely used methodologies using machine learning algorithms to predict length of stay in hospital patients?

With this research question, the systematic review of the literature aims to identify and analyse relevant studies that have explored the topic. To perform a systematic review of the literature, multiple databases will be used. Databases are known to cover a wide range of research articles in the field of healthcare and machine learning. Using a systematic approach, the review aims to minimise bias and ensure a thorough examination of the existing literature.

The analysis of the studies will analyse the use and evaluation of methodologies of different predictive models for the length of stay. This will involve evaluating the strengths and limitations of various machine learning algorithms, along with their performance metrics and validation techniques. The results of the systematic review of the literature will provide a solid foundation for this research.

2.2.2 Data Sources and Search Methods

The systematic review of the literature conducted in this study adhered to the PRISMA guidelines (Moher et al., 2009). To ensure a complete analysis, a thorough search was performed in multiple databases, including Scopus, PubMed, and Web of Science.

Carefully constructed queries, as presented in Appendix X, were used to capture relevant studies related to length-of-stay predictions. These queries incorporated key concepts and terms relevant to the research objectives, with the aim of covering a wide range of perspectives and approaches.

2.2.3 Study Selection

The PRISMA study selection was carried out independently by the researcher and was validated by Supervisor. The titles and abstracts of the selected publications were reviewed by an independent reviewer. At first, the titles will be scanned and held against the inclusion and exclusion criteria. Consecutively, studies that passed the title scanning will be evaluated based on their abstract. The studies that will be included in the review were read and analysed in full text.

2.2.4 Inclusion and exclusion criteria

The objective of the study is to conduct a systematic review of the literature to explore the field of length-of-stay predictions for hospital patients. Throughout the search, strict adherence to predefined inclusion and exclusion criteria was maintained to ensure quality.

To gain an complete understanding of the potential to predict length of stay, a deliberate decision was made not to focus solely on studies related to the daycare department. This approach allows for a broader exploration of the topic, taking into account the diverse population of patients who may not be limited to the daycare setting. Consequently, no articles were excluded on the basis of specialisation. To ensure the inclusion of studies with sufficient statistical power and state-of-the-art methodologies, specific criteria were established. Articles considered for inclusion needed to have a sample size that exceeded 100 patients and be published between January 2015 and January 2023. Furthermore, to maintain relevance to the quality standards of the population in the hospital, only articles from western countries were included. This criterion was designed to ensure that the selected studies reflect a similar standard of quality and care. For a comprehensive overview of the inclusion and exclusion criteria, together with an explanation for each criterion, see Table 1. Table 1 provides transparency and clarity on the selection process.

Table 1. Inclusion and exclusion criteria - systematic review of the literature LOS

Criteria	Description
Inclusion	
Study Design	Review articles, editorials, original studies,
Patient population	Adult, hospitalised patients
Outcome measure	Studies evaluating length of stay as primary or secondary outcome measure
Prediction methodology	Studies that develop or evaluate prediction models specifically designed to predict the length of stay of hospitalised patients. Includes both retrospective and prospective models.
Language	Studies published in English, as language restrictions may impact the feasibility of data extraction and analysis.
Geographical Location	Studies conducted in well-developed countries.
Publication date	Studies published from January 2015 to January 2023, to ensure relevance and accessibility.
Statistical methods	Studies employing a wide range of statistical analysis to enhance the results.
Exclusion	
Population	Studies that have a too specific population group. E.g. specific disease.
Western countries	To match healthcare quality standards in the Netherlands.
Duplicate studies	Studies identified through database searches and manual
Data Completion	Studies with incomplete or insufficient data to evaluate the development or evaluation of prediction models will be excluded.
Full text availability	Studies for which full text is not available will be excluded
Sample size	Studies with very small sample sizes (n<100) may not provide robust findings and will be excluded.

2.2.5 Data extraction

To ensure a systematic and structured approach to data extraction, a data extraction form was designed. This section presents a detailed overview of the data extraction form. The data extraction form was designed to capture essential information from the articles included in the study.

The evaluation of the articles during the full text screening involved evaluating general elements and task-specific elements related to the length-of-stay predictions. The comprehensive approach enabled the extraction of data to address the research objectives. This includes information on the predictive model used, performance evaluation measures, and factors that influence stay duration. Table 2 displays a comprehensive overview of the information collected. The full data extraction table can be found in Appendix A.

Table 2. Data extraction table LOS systematic literature review

Information	Description
General aspects	
Study identification	Author(s), year of publication, title, journal, DOI, study design.
Study Characteristics	Country, sample size, funding source.
Study Population	Patient demographics (e.g., age, gender), medical speciality, type of admission
Prediction model details	Model development or evaluation, model type, predictor variables, model performance measures.
Validation cohorts	Description of independent validation, sample size, patient characteristics, setting, and results.
Length of stay outcomes	Definition of length of stay, units of measurement, median/mean length of stay, IQR, proportion of short/long stays, classification classes.
Missing data	Description of missing data, methods used to handle missing data.
Task-specific aspects	
Predictor Variables	Detailed list of predictor variables used in the prediction model.
Medical specialisation	Description of the speciality or department.
Model results	Description of the process for model development, variable selection, model fitting techniques, feature engineering.
Model Evaluation	Performance measures of the prediction model, calibration measures
Model Validation	Model validation and calibration results.
Feature selection	The selection process of the included prediction variables.
Limitations	Discussion of the limitations of the prediction model, potential bias, and generalisability concerns.

2.2.6 Search strategy

The search strategy used incorporates a selection of synonyms to achieve a comprehensive coverage of relevant studies. To help construct the query, a synonym table, Table 3, was created as a reference. Considering various terms and variations, the search strategy aims to include a wide range of literature relevant to the research topic. The search query was established using the AND and OR operators. The final query can be found in Appendix A.

Table 3. Synonyms table LOS

Search terms	Synonyms
Prediction	Forecast, Estimation, Projection, Modelling, Prognostication
Length of stay	Duration, Hospital stay, Inpatient stay, LOS
Hospital	Medical Centre, Healthcare Facility, Clinic, Institution
Patients	Individuals, Subjects, Participants, Medical Cases
Predictive Models	Machine Learning Models, Statistical Models, and AI Models
Factors	Variables, Predictors, Covariates, and Features
Outcome	Result, End Point, Event, Dependent variable
Analysis	Examination, Evaluation, Assessment, Study

2.2.7 Quality Assessment

The systematic review of the literature in this study used PubMed, Scopus, and Web of Science databases. It is important to note that other databases could have been included to minimise the risk of missing relevant sources. The exclusion of certain databases increases the probability of overlooking valuable information. Furthermore, it is essential to consider the possible influence of the researcher's personal opinion during the article selection process. When interpreting the findings of the systematic review of the literature, these two aspects must be recognised and taken into account to ensure an unbiased analysis.

2.2.8 PRISMA flow diagram

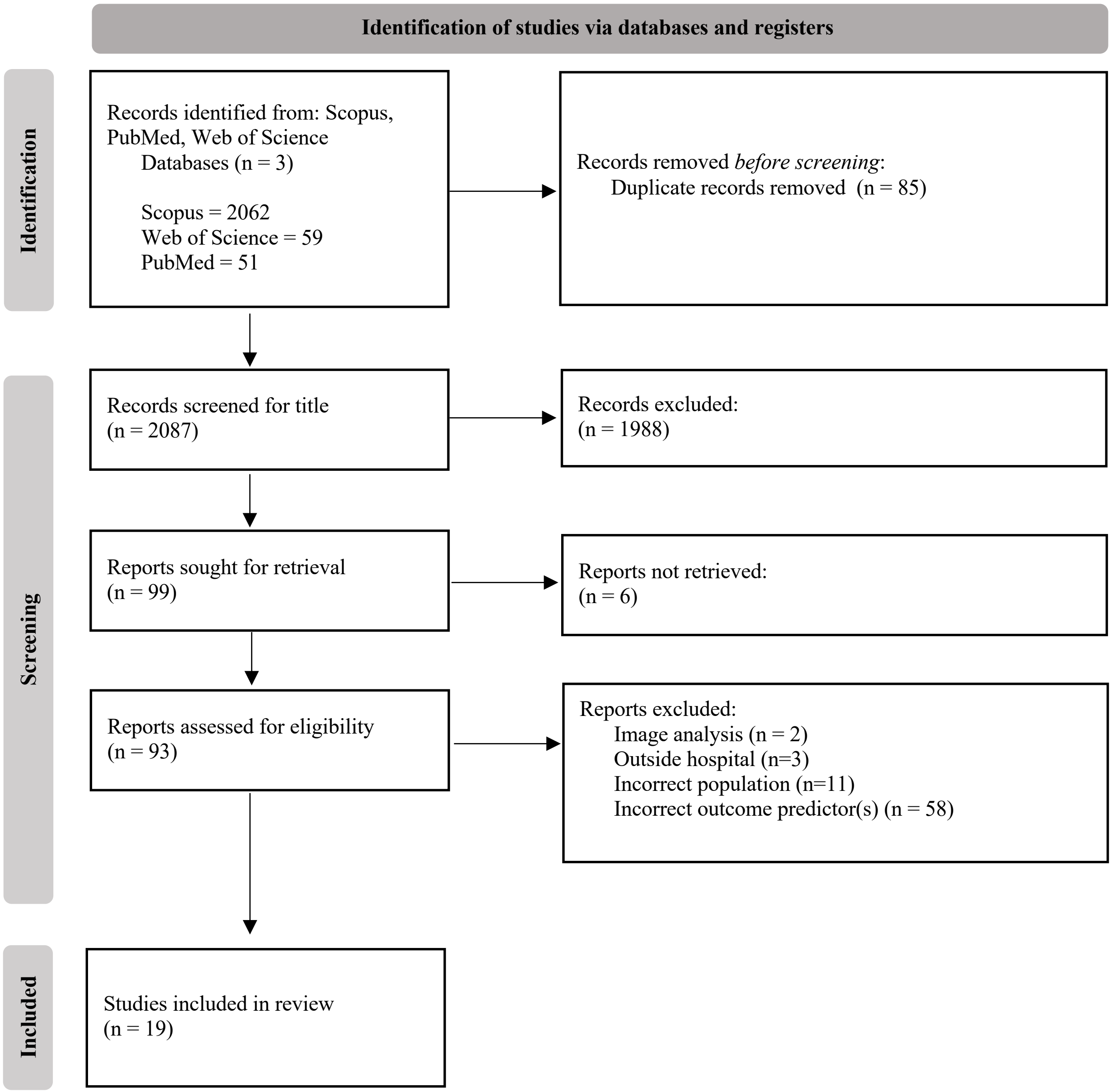

Figure 2. PRISMA flow diagram LOS

2.2.9 Data Synthesis

The data synthesis section offers an overview of the characteristics and findings of the included studies. This section examines the general characteristics and task-specific characteristics, to provide an overview of the findings in the articles. By analysing and summarising these characteristics, the section presents a clear and concise overview of the key findings within the included articles.

2.2.9.1 General Characteristics

Table 4 provides an overview of the general characteristics extracted from the articles included in the literature review. The table discusses various general aspects such as country, study population, approach, outcome, handling of missing data, validation cohorts and best models utilised in the studies. By providing the frequency and percentage distribution of each characteristic, the table offers a concise summary of the key features found within the included articles.

Table 4. General characteristics table systematic review of the literature LOS

Characteristics		**n**	**%**
Country	USA	11	68,75%
	Australia	2	12,50%
	Taiwan	1	6,25%
	Unknown	1	6,25%
	Germany	1	6,25%
	Multiple	1	6,25%
Study Population	Non-specific	7	41,80%
	Specific	10	58,20%
Approach	Classification	9	56,25%
	Regression	3	18,75%
	Both	4	25,00%
	None	1	6,25%
Outcome LOS	Short stay, Long stay	4	22.22%
	Extended stay	2	11.11%
	discharge, no discharge	1	55,60%
	Days	9	50,00%
Missing data	Unspecified	8	47.06%
	Dropped Missing	3	17.65%
	Extra-class	1	5.88%
	Imputation	1	5.88%
	Multiple	2	11.76%
Validation cohorts	Cross-validation/Hold-out	6	40,00%
	Hold-out	5	33,33%
	Cross-validation	1	6,67%
	Not specified	5	33,33%
Best Model	Logistic regression	3	18.75%
	Neural Network	2	12.50%
	Random Forest	1	6.25%
	Combination	2	12.50%
	Bayesian Model	1	6.25%
	XGBoost	1	6.25%
	Empirical Logistic Discrete Hazard Model	1	6.25%
	One-class JITL-ELM	1	6.25%
	Naive Bayesian Model	1	6.25%
	LightGBM	1	6.25%
	Positive Unlabelled Learning	1	6.25%

The distribution of studies in different countries reveals a notable concentration of research in the United States, which reaches 68,75% of the included studies. The finding suggests a strong presence of research in the United States. The other included studies vary from different continents, where only Heim et al. (2019) are located in Europe. Another remarkable included article is that of Barsasella D et al. (2022), researched in Taiwan. The article is included because Taiwan is a well-developed country and the article is insightful.

The characteristics of the study population indicate a wide range of research fields. Approximately 58,20% focused on a specific population. Most of the included studies focused on specific procedures such as total knee arthritis (Navarro et al., 2018), or the ICU department (Ma et al., 2020). The other 41,80% did not have a specific procedure and included a variety of patients in their length-of-stay predictions.

Regarding the methodological approach, included studies can be categorised into three groups: classification, regression, and a combination of both. The most prevalent approach was classification, observed in 56,25% of the included studies. The study by Banga et al. (2017) did not use a prediction method, but rather a statistical approach. The findings of this study focus on procedure-related variables, which are independent predictors, but are not used to make predictions. It suggests that the approach, classification, or regression, is correlated with the main objective of the article.

In terms of outcome measures, the included studies demonstrated a single focus on the duration of stay as the most important outcome factor. There are also studies that combine costs and length of stay(Navarro et al., 2018; Ramkumar et al., 2019), which is irrelevant in the case of hospital, as the focus is on LOS and planning. Unfortunately, no included studies incorporated their findings into a patient planning algorithm to increase bed occupancy.

Addressing missing data is a critical aspect in developing machine learning models for length-of-stay predictions. Missing data cannot be handled by most machine learning models and must be addressed during the data pre-processing phase. Approximately 47,06% of the studies excluded this pre-processing step. Based on the characteristics table with studies that included this step, dropping the missing records from the data set was the most common approach.

The validation cohorts used to assess model performance showed variations. Cross-validation in combination with a holdout approach was the validation cohort used the most frequently (40%). There were differences in the percentages for the hold-out approach. Most studies had a train-test division of 80:20, where Zeng (2022) used a division of 99: 1. For classification approaches, all studies included the accuracy as an evaluation metric in some studies accompanied by the F1 score, sensitivity, specificity and the ROC curve with AUC. The validation metric chosen most (60%) for the regression approach is the mean squared error.

Finally, the methodologies of the machine learning models constructed and evaluated. Approximately 47% of the studies deployed multiple models and evaluated their performance to select the best working model for the specific case. Studies that included only one model based their choice on related work. The machine learning models that were used the most frequently in the analysed studies included logistic regression (n = 9), XGBoost (n = 5), Random Forest (n = 4) and neural networks (n = 3).

2.2.9.2 Studies of the systematic literature

During the systematic review of the literature, two systematic reviews of the literature were identified. The first is by Lu et al. (2015) and the second is by Bacchi et al. (2022). The two reviews serve as an important reference in the field and contribute valuable information to the current study's research.

The systematic review by Lu et al. dates from 2015 which can be seen in the investigated approaches. The articles reviewed mainly focus on conventional statistical techniques, and no state-of-the-art machine learning techniques were discussed. The most used technique is Ordinary Leas Squares, which is a technique not seen in the review by Bacchi et al. The performance of the models is also significantly lower (R^2 score 0.3 -0.6) than the studies presented by Bacchi et al. (0.63 – 0.973) The reviewed literature is focused on the determinants in the predictions of length of stay. Based on the included articles, a conceptual framework was constructed that describes the factors that influence the length of stay. Figure 3 presents this conceptual framework of Lu et al. (2015).

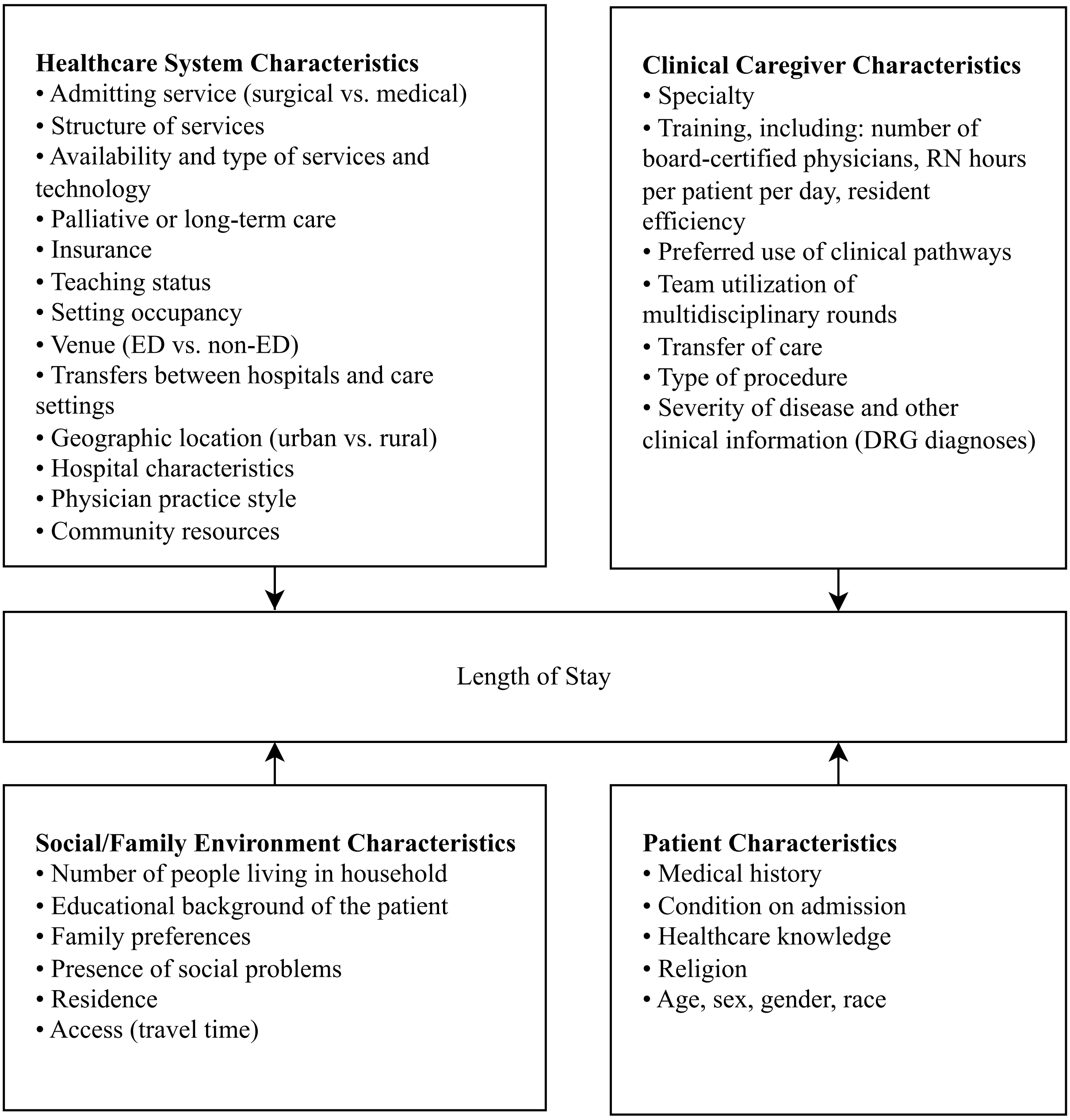

Figure 3. Influencing factors of LOS in patients

Lu et al. (2015) primarily investigated conventional techniques for predicting length of stay, while Bacchi et al. (2022) specifically focused on machine learning techniques to predict length of stay patients. A total of 21 articles were included in the review, which examined various medical specialities and patient populations. The machine learning models in the studies included support vector machines, (artificial) neural networks, Bayesian networks, decision tree algorithms, random forest and logistic regression models. The results of these studies varied widely in outcome measures and validation cohorts. The two outstanding results achieved an accuracy of 80% and an AUC of 0.94 using Random Forest (Daghistani et al., 2019). Another study showed an accuracy of 87.4% and an AUC of 0.905 using ANN(Launay et al., 2015).

The two systematic reviews of the literature conducted by Lu et al. (2015) and Bacchi et al. (2022) offer valuable information on predicting the length of stay of patients. The review by Lu et al. focused on conventional statistical techniques and identified determinants of length of stay. The review by Bacchi et al. (2022) explored the state-of-the-art machine learning techniques and found varying results in different models. Right now, there is no consensus on which model to use for which specific situation. The studies presented by Bacchi et al. (2022) achieved higher performance measures compared to the results presented by Lu et al. These findings highlight the potential of machine learning approaches to predict the length of stay of patients.

2.2.9.3 Classification

The most used approach (n=9) is the classification in the prediction of the length of stay. All studies provide different classification classes. The study by Bacchi et al. (2020) categorises patients into two groups, similar to Arora et al. (2022). The first group includes patients with stays less than two days, while the second group includes patients with stays longer than two days. The two studies by Arora et al. (2021, 2023) both employed different classification criteria. One study focused on stays of less than 2 days and more than 2 days, and the other classified as normal or extended stay.

Various models are used to perform the classifications for patients. Bayesian networks (Cai et al., 2016), XGBoost (Chen, 2021), positive unlabelled learning techniques (Arora et al., 2022), and support vector machines (Bacchi et al., 2020). The non-linear weighted extreme gradient boost technique proved by Chen (2021) achieved the highest accuracy of 87.3%.

The techniques have relatively high accuracy, but it is hard to compare the studies based on precision because they all have different study populations and categories. The studies underscore the potential of classification models to be used as information tools. Furthermore, at this point, there is no golden standard on categories for classifications, as all included studies report different categories. The classification categories depend greatly on the use case and the available data.

2.2.9.4 Regression

The second approach identified in the studies involves the use of regression analysis, and all included studies reported their predictions of the length of stay in terms of days. Within the study populations, there are differences and the populations differ significantly. Heim et al. (2018) investigated the factors that influence length of stay in patients with severe odontogenic infections, where Muhlestein et al. (2018) developed a model to predict brain tumour patients with LOS. In addition to the different study populations, there were also significant differences in the predictor variables used. Muhlestein et al. used preoperative pneumonia, sodium abnormality, weight loss, and race as key predictors, which are not seen in other studies. Siddiqa et al. (2022) analysed a large data set from New York hospitals and compared various machine learning models to predict the length of stay. They found that random forest regression outperformed the other models evaluated. In this study, an accuracy of 92% was achieved. Zeng (2022) constructed a regression prediction model using an unseen method: LightGBM. He achieved an R^2 of 96% and a relatively low MSE of 2.231. The study used a large data set that contains more than two million patient records, suggesting that a larger data set improves accuracy.

Studies contribute to the advancement of length-of-stay prediction methodologies. They also highlight the importance of accurate predictions for resource allocation, cost management, and patient care. The findings of these studies showcase the possibilities of machine learning techniques for the prediction of stay duration.

2.2.9.5 Ensemble Models

The third approach is to use ensemble methods when multiple predictions are made. The studies by Karnuta et al. (2020) and Navarro et al. (2018) focus on predicting length of stay, discharge position, and hospital costs using machine learning techniques. Both studies underscore the importance of cost prediction in the healthcare setting. Two promising neural networks were constructed to provide valuable information to healthcare providers, patients, and insurance companies on the length of stay of patients.

Similarly, Barsasella et al. (2021) aimed to predict two outcomes for patients; the length of stay and mortality among patients with Type 2 diabetes and hypertension. A combination of XGBoost for length of stay predictions and a logistic regression model for mortality predictions presented the best performance. Another study by Cai et al. (2015) developed a Bayesian network that provides real-time predictions for length of stay, mortality, readmission, all at once by using EHRs. The model estimated the probabilities that a patient would be home, hospitalised or alive within the next seven days. The researchers achieved an average accuracy of 82% with an AUROC of 0.84.

Studies by Karnuta et al., Navarro et al., Barsasella et al., and Cai et al. show the ability to accurately forecast multiple outcome variables using machine learning techniques. Accurate predictions for length of stay, costs, discharge positions, and readmissions enable improved resource management in the future.

The use of ensemble models offers decision makers a comprehensive set of information for informed decision making. Despite their potential advantages, the specific implementations of these ensemble models are not discussed in this article. The absence of a detailed discussion on implementation methodologies leaves room for further exploration and investigation for future research.

2.2.10 Conclusions

Length-of-stay predictions have been an emerging field of research in the last couple of years, garnering increasing attention and publication frequency. In particular, there has been a shift from traditional statistical methods to machine learning approaches in this domain. Geographically, most of these studies originated in the United States.

A comprehensive analysis of the existing literature reveals significant variations in multiple aspects of these studies. Sample sizes, model selection, specialisation, featured variables, validation cohorts, and outcome measures vary significantly. This diversity underscores the complexity and multifaceted nature of the length-of-stay predictions. Despite the growing popularity of machine learning techniques, a consensus on a definitive choice of models, handling missing data, and feature selection has yet to be established. The lack of standardisation can be attributed to the absence of a universally superior model for all scenarios. Consequently, the approach involves a custom solution for each specific case and the performance of experiments with multiple models to determine the most effective solution.

It is worth noting that there is a scarcity of studies that explore the combined use and research of multiple outcome predictors or models. In addition, the discussion and subsequent implementation of these approaches in most studies is lacking. Furthermore, none of the reviewed studies followed up the use of predicted results in planning systems, which was envisioned as a potential application in this research. The systematic review of the literature by Bacchi et al. (2022) provides an elaborated overview of studies that contain predictions about length of stay using machine learning techniques. The work did not include two crucial aspects: validation cohorts and handling missing data. The handling of missing data is crucial information to other researchers as it highly influences the outcomes of the models. The validation cohorts are also not discussed. Validation cohorts such as hold-out technique and cross-validation provide insights in the validity and power of the methodology. The literature research conducted, does include these aspects and showed a mixture of approaches. Most studies used a hold-out of 80/20 and a fivefold cross-validation to empower the results.

The conceptual framework proposed by Lu et al. (2015) serves as an initial reference for the feature selection procedure. The dataset obtained from the RCH includes variables that are identified as significant within the framework. Specifically, variables such as speciality, type of procedure, age, sex, physicians involved, and admission condition (ASA) are considered crucial by Lu et al. (2015). In particular, these variables are present within the RCH dataset. On the other hand, additional variables such as weight, length, second physician, and anaesthesia technique, which are not explicitly mentioned, exist within the data set and will be examined for relevance as part of the research.

In response to the research question "What are the most effective machine learning algorithms for predicting length of stay in hospital patients?", the review highlights the absence of a definitive answer. Instead, the findings emphasise the need for continued exploration and experimentation to identify the most effective algorithms for different scenarios.

The field of length-of-stay predictions has witnessed significant growth and attention in recent years. It slowly shifts towards machine learning approaches, as shown by the number of publications. However, a great deal of diversity is observed in how these studies approach problems, as the field of research is relatively new. Unfortunately, no specific research is focused on the exact same topic as the main research question. Therefore, this research needs to experiment with innovative methodologies to ensure robustness and reliability.

2.3 PRISMA systematic literature review | Patient scheduling

This chapter presents a systematic literature on patient planning in a hospital setting, summarising key findings and advancements in the field.

2.3.1 Introduction

In this research, patient scheduling and operating room scheduling are treated as synonymous terms, since the primary focus is on patients in the daycare department who have undergone an operation. Research focusses on exploring strategies for efficient patient care within operating room schedules. It will recognise the critical importance of effective operating room management in hospitals.

Effective management of operating rooms is critical for hospitals, as ORs are both costly to maintain and generate significant revenues (Marjamaa et al., 2008). However, managing ORs can be challenging due to conflicting preferences among stakeholders, such as management and physicians (Cardoen et al., 2001). Physicians strive primarily for high patient satisfaction, where hospital management is focused on cost-effectiveness.

To properly manage operating rooms, hospitals must consider the mix of treated patients. Admissions can be divided into two categories: scheduled and nonscheduled (Adan & Vissers, 2002; Demeulemeester et al., 2013). Scheduled patients are selected from a waiting list, while non-scheduled patients are emergency admissions (Adan & Vissers, 2002). Within this group of patients, there are still many different care needs. You have various specialisations that then also perform different procedures. To allocate patients to operating rooms, patient scheduling, several decisions need to be made. First, the number of slots per surgical group must be determined, and second, patients must be assigned to these slots (Testi & Tànfani, 2009). In the process of making these decisions, planners always need to take into account the availability of resources, especially beds.

The planning of operating rooms depends on the number of patients and resources available, including beds (Adan & Vissers, 2002). The scarcity of beds is a limiting factor in the freedom of planners. According to Robb et al, up to 62.5% of cancelled general operating room procedures at a large university teaching hospital were attributed to the absence of available beds (Robb et al., 2004). To increase efficiency and bed occupancy, it is crucial to adequately handle resources. In addition to cancelled operations, bed shortages can cause delays in scheduled inductions. When postpartum beds are full, patients are blocked in the upstream labour and delivery areas, preventing new admissions (Wang et al., 2019).

In the remainder of this chapter, the aim is to explore the field of patient planning in the daycare facility by assessing the literature in a systematic approach. The literature will be reviewed according to the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The objective of the review is to examine the different approaches and methodologies used for patient scheduling in daycare facilities.

The systematic review of the literature will focus on the scheduling of ILP patients. The research question will be:

What different methodologies can be applied to optimise patient scheduling in daycare departments?

With this research question, the systematic review of the literature aims to identify and analyse relevant studies that have explored the topic. To conduct the systematic literature review, multiple databases will be used. The databases are known to cover a wide range of research articles in the field of healthcare. Using a systematic approach, the review aims to minimise bias and ensure a thorough examination of the existing literature. The studies will be analysed for their methodologies, results, and evaluation methods to obtain a comprehensive overview of the available literature.

2.3.2 Data Sources and Search Methods

The systematic review of the literature was performed according to the PRISMA guidelines (Moher et al., 2009). A comprehensive search was carried out in the databases of Scopus, PubMed, and Web of Science.

The queries will be constructed and presented in Appendix X, and the databases will search with the constructed query to capture relevant studies related to patient scheduling. These queries incorporated key concepts and terms pertinent to the research objectives, with the aim of encompassing a wide range of perspectives and approaches.

2.3.3 Study Selection

The selection of the PRISMA study was carried out independently by the researcher and validated by the Supervisor. The titles and abstracts of the selected publications were independently reviewed by the researcher. At first, the titles will be scanned and held against the inclusion and exclusion criteria. Consequently, studies that passed the title scan will be evaluated based on their abstract. The studies that will be included in the review underwent full text reading and analysis.

2.3.4 Inclusion and exclusion criteria

The objective of this study is to conduct a systematic review of the literature in order to explore the field of patient scheduling. To ensure a comprehensive review, well-defined inclusion and exclusion criteria are established.

Given the diverse patient population in the daycare department, it is crucial to obtain a holistic understanding of the possibilities of scheduling patients. Therefore, the decision was made to include articles from all specialisations, as it allows for a more comprehensive analysis. However, if the articles exclusively focus on specific diseases or conditions, they will be excluded to maintain the relevance of the study to the broader population of patients.

For a detailed and transparent overview of the inclusion and exclusion criteria, refer to Table 5. The table provides clarity and transparency in the study selection process, ensuring a systematic and comprehensive review of the literature.

Table 5. Inclusion and exclusion criteria systematic literature review LOS

Criteria	Description
Inclusion	
Study design	Articles that discuss scheduling optimisations for outpatients
Patient Population	Studies focussing on daycare departments or similar outpatient settings.
Optimisation	Studies proposing or applying planning algorithms or approaches for patient scheduling.
Outcome Measures	Studies evaluating relevant outcomes such as resource utilisation, waiting times, patient satisfaction, or cost effectiveness.
Constraints	Studies that discuss constraints, variables, or factors considered in ILP models for scheduling patients in the daycare department.
Publication Type	Peer-reviewed journal articles, conference papers, and reputable academic sources.
Exclusion	
Inpatient setting	Studies focussing solely on inpatient scheduling or scheduling in other healthcare settings without direct relevance to daycare departments.
General Scheduling	Studies focussing on general scheduling methods or approaches without specifically addressing patient scheduling in daycare departments.
Theoretical	Studies primarily discussing theoretical aspects of ILP without practical application or evaluation in the scheduling of patients in the daycare department.
Language	Studies not published in English, as language limitations may affect the comprehension and synthesis of findings.
Accessibility	Studies that are not available in full text format or are not available through reliable sources.
Population Specificity	Studies that have a too specific population group not representative of outpatients.
Duplicate Studies	Studies identified through database searches and manual screening that are duplicates of studies already included.
Data Insufficiency	Studies with incomplete or insufficient data to evaluate scheduling or their application in scheduling patients in the daycare department.

2.3.5 Data Extraction

To ensure a systematic and structured approach to data extraction, a data extraction form was designed and used. Each article that met the defined inclusion criteria underwent a title selection, followed by a screening of the abstract. Articles that remained within the inclusion criteria were subjected to full-text screening.

During the evaluation of these articles, both general elements and task-specific elements related to patient scheduling were considered. The data extracted will be a solid foundation for this research. Table 6, presents the information collected related to the general aspects and the task-specific aspects. The table containing the results of the data extraction can be found in Appendix B.

Table 6. Extraction table systematic literature review Patient Planning

Information	Description
General aspects	
Authors	The author(s) of the study.
Title	The title of the research article.
Year	The year of publication.
Country	The country where the study was conducted.
Task-specific aspects	
Main Objective	The primary aim or goal of the study.
Study Population	The population or group of patients involved in the study.
Input Variables	The variables used as input in the patient scheduling model.
Data	The type and source of data used in the study.
Intervention/Approach	The approach or intervention used for patient scheduling.
Findings	The main results and findings of the study.
Implementation	Details about how the patient scheduling system was implemented.
Discussion/Implications	Discussion of the implications of the study and potential impact.
Software used	The software or tools used for patient scheduling.
Conclusions	The overall conclusion or summary of the study.
Appointment type	The type of appointments considered in the study (e.g., outpatient, inpatient).
Validation	The process of validating the patient scheduling model or system.
Future directions	Suggestions for future research or improvements in patient scheduling.

2.3.6 Search strategy

The search strategy employed incorporates a selection of synonyms to achieve a comprehensive coverage of relevant studies. To help construct the query, a synonym table, Table 7, was created as a reference. Taking into account various terms and variations, the search strategy aims to encompass a broad range of literature relevant to the research topic. The search query was established using the AND and OR operators. The final query can be found in Appendix B as well as the employed search query.

Table 7. Synonyms table systematic literature review Patient Planning

Term	Synonyms
Optimisation	Mathematical Optimization, Integer Programming, Integer Optimization, Optimization Models, Optimization Techniques, Mathematical Models, Operations Research, Combinatorial Optimization, Decision Optimization, Heuristic Optimization, Metaheuristic Optimization, Constraint Programming
Patient Scheduling	Appointment Scheduling, Planning, Scheduling Optimization, Scheduling, Patient Planning, Patient Appointment, Patient Booking, Scheduling Efficiency, Scheduling Algorithms, Scheduling Models, Scheduling Systems, Scheduling Strategies, Scheduling Policies
Daycare Department	Daycare, Day Care, Day Treatment, Day-care, Outpatient Department, Ambulatory Care Center, Outpatient Care, Outpatient Clinic, Clinic, Outpatient Treatment, Same-Day Care, Same-Day Treatment, One-Day Care, One-Day Treatment, Single-Day Care, Single-Day Treatment, Short Stay, Minor Procedures, Non-Admitted Patients, Brief Intervention, Transitional Care, Ambulatory Services, Same-Day Surgery, Walk-In Clinic

2.3.7 Quality assessment

The systematic review of the review of the literature in this study used PubMed, Scopus, and Web of Science databases. It is important to note that other databases could have been included to minimise the risk of missing relevant sources. The exclusion of additional databases increases the probability of overlooking valuable information. Furthermore, it is essential to consider the potential influence of the researcher's personal opinion during the article selection process. When interpreting the findings of the systematic review of the literature, these two aspects must be acknowledged and taken into account to ensure unbiased analysis.

2.3.8 PRISMA flow diagram

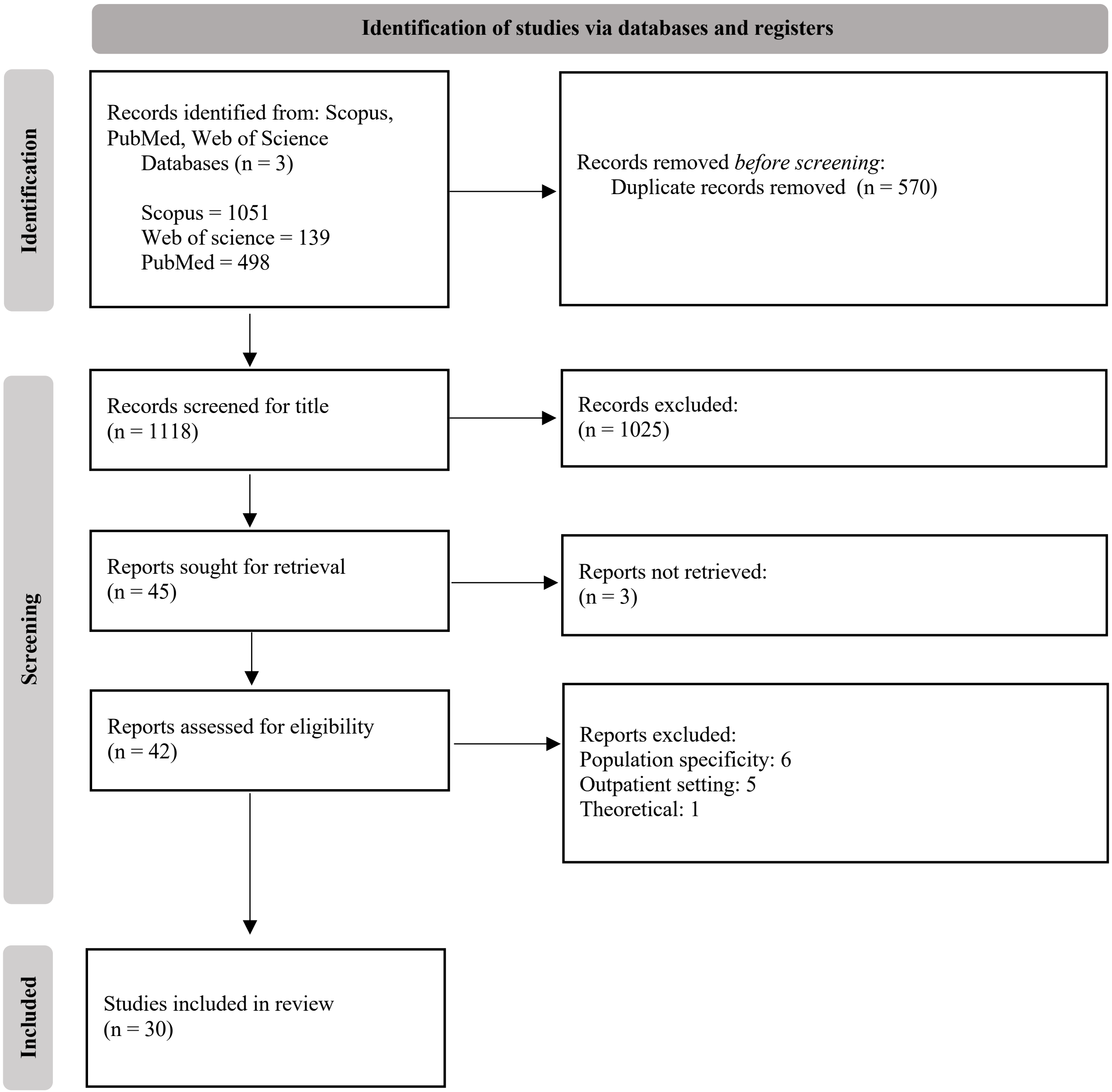

Figure 4. PRISMA flow diagram Patient Planning

2.3.9 Data Syntheses

The data synthesis section offers an overview of the characteristics and findings of the included studies. This section examines the general characteristics and task-specific characteristics to provide an overview of the findings in the articles. By analysing and summarising these characteristics, the section presents a clear and concise overview of the key findings within the included articles.

2.3.9.1 General Characteristics

Table 8, provides an overview of the general characteristics extracted from the articles included in the literature review. The table presents various general aspects such as country, type of appointment, approach, and validation. By providing the frequency and percentage distribution of each characteristic, the table offers a concise summary of the key features found in the included articles.

Table 8.Synthesis of general statistics literature review Patient planning

Characteristics	Feature	n	%
Country	USA	8	25,81%
	Hong-Kong	5	16,13%
	The Netherlands	4	12,90%
	Canada	4	12,90%
	China	3	9,68%
	Germany	2	6,45%
	Jordan	2	6,45%
	South-Korea	1	3,23%
	Tunisia	1	3,23%
	Unspecified	1	3,23%
Appointment type	Single-appointment	26	83,87%
	Multi-appointment	5	16,13%
Approach	(M)ILP	10	32,26%
	Simulation	5	16,13%
	Heuristic	3	9,68%
	Mathematical modelling	3	9,68%
	SIP	2	6,45%
	Simulation with combination	2	6,45%
	Data analysis, regression	1	3,23%
	Greedy heuristic	1	3,23%
	MDP	1	3,23%
	MOPSO, MO-PASS	1	3,23%
	SAA	1	3,23%
	Simulated annealing	1	3,23%
Software Used	Unspecified software	12	38,71%
	CPLEX	4	12,90%
	Combination	4	12,90%
	Matlab	3	9,68%
	Arena	3	9,68%
	Python	2	6,45%
	IVE Xpress 8.6, Simpy	1	3,23%
	Gurobi	1	3,23%
	Microsoft Visual Basic	1	3,23%
Validation	Simulation	14	45,16%
	Unspecified	8	25,81%
	Sensitivity analysis	5	16,13%
	Validated with real data, experiments	3	9,68%
	Comparing models	1	3,23%

The analysis of the general characteristics table deduced from the data extraction table reveals noteworthy findings. First, in terms of country representation, it is evident that most of the studies focused on scheduling originated in western countries. The United States had the highest representation with 25.81% of the articles, followed by Hong Kong (16,13%), the Netherlands (12.90%), and Canada (12.90%). The observation suggests significant research interest in patient scheduling within Western nations, possibly driven by factors such as healthcare infrastructure, research funding, or academic institutions.

Regarding the type of appointment, the vast majority of studies (83,87%) focused on single appointment scheduling. The statistic emphasises the prominence of single-appointment scheduling in the literature and the potential need for further research on multi-appointment scheduling. The research ahead will focus on single appointment scheduling, as they target outpatients who only stay one day at the hospital.

Analysing the various approaches employed in the study, it is notable that a significant portion (32.26%) utilised (Mixed) integer linear programming techniques. Simulation-based approaches also played an important role, accounting for 16.12% and two studies used a combination of ILP and simulation to schedule. Other approaches, such as heuristics, mathematical modelling, and various optimisation algorithms, were used in smaller proportions. It can be concluded that a wide range of methodologies are being employed in the field of patient scheduling. The variety of methodologies can be driven by the need for different tools for specific problems as well as the personal preference of researchers. The same can be concluded about the choice of software usage. Among the different software packages used, CPLEX (12.90%) and a combination of several software packages (12.90%) are the most frequently deployed. Additionally, Matlab by Mathworks and Arena by Rockwell Automation are also being used frequently with 9.68%. There is also a significant proportion (38,71%) that does not specify the software packages being used, which makes replication harder.

Regarding the validation techniques, the most significant proportion (45.16%) used simulation as a validation technique. The articles implemented their scheduling approach within simulation software to assess performance and effectiveness. Simulation allows for the creation of virtual environments to mimic real-world scenarios, providing a platform to evaluate the scheduling algorithms and the impact on system efficiency. Furthermore, 25.81% of the studies did not specify the validation methodology. The information gap makes it challenging to assess the validity of the approach taken. Among other studies, sensitivity analysis of the ILP and comparison of the results with real-world data were the approaches to validate the results.

2.3.9.2 Task-specific

General characteristics provide information on geographical locations, methodologies, validation techniques, and software usage. In this section, the articles will be analysed in depth to find similarities, abnormalities, and outstanding articles to achieve a deeper understanding of how other researchers approach patient planning.

A subset of the articles included in the review focus on optimisation and improvement of scheduling efficiency in outpatient services. The articles by Lü and Zhang (2023), Feng et al. (2023), Kuiper et al. (2023), Belien et al. (2023), Mahdavi et al. (2023), and Song and Zhao (2021) show a particular interest in the field of improving patient schedules using an ILP or mathematical modelling. Collectively, they point toward improving current appointment scheduling methods. The main objectives of these articles focus on both increasing patient satisfaction, reducing waiting times, and effectively managing work schedules.

Meanwhile, the studies of Bovim et al. (2022, Gao et al. (2022), Wing and VanBerkel (2022), and Khaaled et al. (2022) revolve around the theme of resource utilisation and management. Most of the resources are considered beds, and the availability of nurses is used in the most optimal way. The emphasis of these articles also is on reducing waiting times, suggesting that this is a major challenge in outpatient scheduling.

Several other papers, such as the paper by Yang et al. (2017), and Forghani and Masoumi (2017), Lin (2015), explore the field of adaptive appointment scheduling, hybrid systems, and resource allocation. Adaptive appointment scheduling systems use real-time data and advanced algorithms to adjust appointment schedules based on various factors such as patient needs, provider availability, and unexpected events. The more advanced planning schedules are more state-of-the-art approaches and are only being researched in more recent articles. It is an additional challenge to incorporate the mentioned unexpected events as no-shows as described in the study of Tohidi et al. (2021). The articles published by Aslani et al. (2021), Anvaryazdi et al. (2020), and Tohdi et al. (2021) also incorporate the element of uncertainty and stochastic variations. These more models show progress towards a more realistic and robust model.

Another remarkable approach by Luo et al. (2012) focusses on handling scheduling interruptions. They offer a unique perspective compared to other articles. When considering interruptions within appointment scheduling, effectiveness increases significantly compared to models that do not incorporate interruptions. Another noteworthy study is that of Schafer et al. (2019), which focusses on patient-to-bed assignments. They address the problem of having multiple stakeholders within their appointing methodology and outperform other scheduling algorithms.

Across all studies, many different input variables are used in the constructed models. The most common input variables that occur are service times (n=11), arrival times (n = 9) and hospital capacity (n=7). An important note is that all arrival and service times are not generated by probability distributions. The purpose of this research is to implement real-time, personally generated timings for the service time of a patient.

The articles presented in the systematic literature review have a scientific approach, in which reality is modelled and an algorithm is evaluated. Unfortunately, as shown in the extraction table in Appendix X, only 16% of the articles proposed a practical implementation of the researched scheduling methodology. The other articles do not discuss the practical implementation at all or point out that additional research is required. In the articles where implementation is discussed, is on a small scale, only being deployed at a single hospital, department, or clinic.

2.3.10 Conclusions

The systematic review of the literature on patient scheduling provides valuable information in the field. Analysis of general characteristics provided an overview of the methodologies, validations, and software packages used. Task-specific analysis provided information about objectives, outstanding articles, input variables, and implementation strategies.

The most frequently occurring approaches used a combination of ILP and CPLEX to achieve the set goal. Therefore, ILP and CPLEX will be used as a basis for patient scheduling. It is important to note that the objectives differ slightly from the objective of this research, since it follows a consecutive approach. The combination of first predicting the length of stay of the patients followed by scheduling the patients accordingly was not found as a methodology in the articles. Combining the two methodologies bridges a gap in the literature to see whether or not the approach at hand is feasible. Furthermore, the actual implementation of this consequential approach is also not discussed in the literature and adds direct value to the hospital planning system. Taken together, the systematic review of the literature provided rich insights that will be used and evaluated in the following chapters, in which the methodologies will be tested.

2.4 Conclusion literature

The systematic literature on patient scheduling and length of stay predictions for outpatients has provided comprehensive information on the field and laid the foundation for this research. Analysis of the literature revealed key findings and trends, as well as shed light on challenges, methodologies, and approaches.

In the last few years, the attention has shifted towards machine learning approaches to predict the length of stay of hospital patients. The relatively new approaches demonstrate superior predictive capabilities compared to traditional methodologies. While machine learning models are often criticised for their lack of transparency, commonly referred to as a "black-box" characteristic, the enhanced performance they offer outweighs this limitation. Consequently, it is highly likely that machine learning will surpass traditional approaches, as evidenced by the increasing frequency of its publications.

It was evident that there is no consensus on which machine learning model to implement in the case of length-of-stay predictions. Therefore, the experimentation and evaluation of state-of-the-art machine learning is crucial to find out which model is most suitable. The input variables used in different studies also varied significantly and the choice of which to use was mainly decided by the availability of the data. The literature review by Lu et al. (2015) presented a detailed exploration of the factors associated with the length of stay. The results of the study will be a solid foundation for selecting relevant characteristics in the selection process of this research.

The scheduling of patients in daycare departments revealed the complexity of managing resources, considering the diverse needs and uncertainties of patients. Furthermore, the literature revealed that there is a wide range of optimisation objectives, suggesting tailer-made solutions for each specific case. The single specific objective of this investigation is to increase bed occupancy and was not found among the included studies. However, the main objectives were clear: improve patient care while adhering to cost-conscious strategies.

Comprehensive literature reviews on length-of-stay predictions and patient scheduling have significantly enriched this research, providing foundational insights. These reviews expand the scientific knowledge base by introducing critical information on the management of missing data and validation cohorts, aspects not covered comprehensively in previous systematic literature reviews, thus substantiating the scientific contribution of this study.

The aim of this study is to develop an integrated approach that tackles the overarching challenge of increasing bed occupancy in healthcare facilities. Taking advantage of the knowledge and methodologies of reviews in the literature, this research aims to achieve improvements in healthcare operations. These improvements include increased efficiency, patient satisfaction, and cost-effectiveness. This research seeks to contribute innovative solutions to the complexities of patient care and resource optimisation to advance healthcare management.

3 Methodology

The bed allocation process at hospital is currently managed by a planning department consisting of five team members. They assign patients to operating slots based on the operation room schedule, with a limitation being the number of hospital beds available. The current procedure is hindered by insufficient knowledge of the length of stay of patients, usually resulting in one patient per bed. In exceptional cases, after consulting physicians, multiple patients might be scheduled per bed. This lack of information poses challenges for planners in optimally allocating beds. This research aims to establish an integrated system to assist planners in the process and increase bed occupancy. The next chapter outlines the methodology developed to achieve this research's objectives.

3.1 Introduction

The conclusions of systematic literature research provided the insight that there is no single machine learning model outperforming another. Therefore, state-of-the-art machine learning models will be developed and evaluated to find out which one is outstanding in this specific scenario.

Initially, patient records are collected containing demographic information, medical evaluations, and length of stay. The collected data will be preprocessed before performing exploratory data analysis and feature selection. Exploratory data analysis provides information on the data set that will provide information about the population at hand. The knowledge will lead to the identification of valuable features in length of stay predictions. To develop an accurate prediction model, a selection of state-of-the-art machine learning models will be developed and tuned. The models that will be included are random forest, gradient boosting, support vector machines, decision tree, K-nearest neighbours, XGBoost, logistic regression, ridge regression, and neural networks. To validate the accuracy of the prediction model, two different approaches were implemented: a classification approach and a regression technique. The classification approach involved categorising patients into groups according to their length of stay. On the other hand, the regression technique involves predicting the total number of minutes a patient will be in the hospital and will occupy a bed.

The combination of these two techniques will be used to validate the results, after which they will be corrected. The output of the classification model, indicating the probability of belonging to a specific class, will be compared with the output of the regression model to validate the results. The performance of both approaches and all models will be evaluated using various metrics such as accuracy, precision, R^2, recall, and F1 score, and the best performing models are selected for further optimisation and implementation. In addition to machine learning to predict the length of stay, an ILP will be formulated to schedule patients. The ILP model integrates the length-of-stay predictions to create schedules that contain sequentially plannable patients. In addition to length-of-stay predictions, the input includes the planning horizon and the speciality for which the planners will be conducting the planning. The aim of the approach is to increase bed occupancy by providing a guide schedule for the planner. The last step is to build a graphical user interface for planners to implement the system into their workflow.

Sequential steps of the process are shown in Figure 5. The initial step in the system involves evaluating patients on the waiting list at the small regional hospital. Demographic information, medical evaluations, and other available data are collected for these patients. The information collected serves as input for the prediction model, which uses two machine learning techniques to predict the length of stay of each patient. The results of the prediction models undergo a trustworthiness and usability evaluation. If necessary, corrective measures are implemented to improve the reliability of the planning process. In the subsequent step, the length-of-stay predictions of patients on the waiting lists are obtained and used as input for the patient scheduling ILP. The output of the model will be a schedule that can be used by operation room planners. The schematic diagram in Figure 5 provides a comprehensive overview of the system as a whole. To achieve the envisioned system, several steps need to be taken. In this chapter, the steps to achieve the system will all be worked out in-depth.

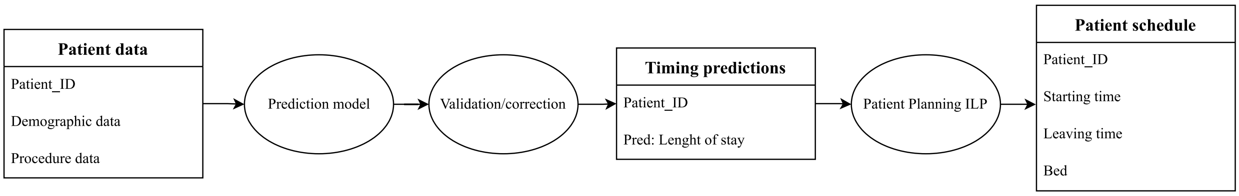

Figure 5. Schematic overview of envisioned system

3.2 Data collection and pre-processing.

The data collected will be extracted from the HiX electronic patient records (EPR) system by Chipsoft. The HiX data management system has the ability to export selected data to be analysed in R and Python. No privacy-sensitive information, for example, name, date of birth, or images, will be saved during the data collection phase to prevent misuse. The exported data set contains columns; see Appendix X for column names, with information on patients at the A2X department.

Pre-processing the data set involves several steps. First, duplicate entries will be removed. Second, patients who have undergone multiple operations within one year will be excluded, as these multiple medical procedures may be correlated. The third step is to address the missing values in the database. In many cases, the patient’s records are incomplete. Missing values can occur, for example, due to the absence of a medical assessment or errors in the administrative process. To address numerical missing values, the "imputation" method will be utilised. This imputation method was recognised as a reliable and effective approach to handling missing data during the literature review, as demonstrated by Muhlestein et al. (2017). In case a record is incomplete, an imputation algorithm fills the empty cell with the mean of the column. Missing values for categorical columns will be handled by adding a new category called ‘missing’, a methodology found in the systematic review of the literature. Subsequently, outliers outside a 99% confidence interval will be removed for the total time, weight, length, and age of the columns to decrease the complexity of the model. The result is a data set that contains 17,545 patient records.

Additionally, two methodologies will be applied in the data preprocessing phase to enable the models to comprehend the data. The first methodology is to scale the numerical variables according to the Min-Max scaling process. By scaling the values to a range between 0 and 1, Min-Max scaling preserves the relative relationships and proportions between the data points. This can be advantageous in certain algorithms that rely on absolute values and relationships, such as the distance-based algorithm k-nearest-neighbours. In Equation 1, the formula for the scaling the values accordingly is depicted.

$$Scaled\ value = \frac{Value - min}{max - min} \tag{1}$$

The second methodology hot-one encoding. It is necessary when dealing with categorical data because most machine learning algorithms work with numerical data and cannot directly process categorical variables. Figure 6 presents a visualisation of one hot encoding.

{Procedure A, Procedure B, Procedure C, Procedure D}

Patient ID	Procedure_A	Procedure_B	Procedure_C	Procedure_D
ID_0	1	0	0	0
ID_1	0	1	0	0
ID_2	0	0	1	0
ID_3	0	0	0	1
...	...	...	...	...

Figure 6. One-hot Encoding Visualisation

3.3 Data Analysis

Upon completion of the preprocessing, the data will be subjected to exploratory data analysis (EDA) techniques to extract insights from the dataset. EDA will facilitate the identification of trends, patterns, and relationships between variables, as well as the detection of outliers. Identification of outliers is important for several reasons. Outliers can represent critical events or special patients. For example, obese patients or patients with complications with an extended stay. Inclusion of outliers can confuse the model during training and can increase the complexity of the model. An increase in complexity can then lead to a decrease in the potential prediction quality. Therefore, outliers outside of a 95% confidence interval will be removed from the dataset.

An overview of the descriptive statistics will be presented to grasp insights about the data set at hand. It is also crucial to display the descriptive statistics for reproducibility to compare datasets. In addition to descriptive statistics, the EDA will consist of tables and visualisations to provide insight into the different probability distributions of length of stay.

The distribution of length of stay will be depicted by a box plot to present the information extracted from the EPR. In addition, the data will serve to uncover correlations between features, which will be displayed using correlation plots. This information will be invaluable during the subsequent feature selection phase. In addition to providing information on population characteristics and correlations within the population, the current use of beds at small regionalwill also be quantified. The bed occupancy can be calculated according to Equation 2.

$$BO = \frac{Total\ number\ of\ patients}{Total\ number\ of\ used\ beds} \tag{2}$$

Bed occupancy can be assessed daily or monthly to analyse patterns and trends. The current status of bed occupancy will be compared with the results of the schedules, allowing an evaluation of the impact of these schedules on bed occupancy.

3.4 Feature Selection

It is crucial to identify the features that show a correlation with the length-of-stay of patients. Incorporating features that have a meaningful correlation can substantially enhance the predictive performance of the model. The hospital-sourced dataset comprises ten features, including factors such as the type of procedure and the attending physician. The study by Liu et al. (2015) provided a solid framework of influential features. A significant proportion of the available features are present in the framework and will therefore be used. Other factors, which are not present in the framework will be evaluated using two techniques.

To determine the significance of the numerical features, a correlation plot will be used. The correlation plot visualises the linear relationship between pairs and features. The correlation coefficient ranges from -1 to 1. The correlation coefficient serves as an indicator of the strength of the relationship between variables. A higher absolute value identifies a stronger correlation, whereas values closer to 0 indicate a weaker association.

The categorical variables are evaluated using a technique known as permutation. The technique involves the construction of a temporary neural network to predict the duration of stay. Subsequently, a baseline performance will be established considering all features. Next, an algorithm will be employed to randomly shuffle one of the features while keeping the remaining features intact. This procedure generates a modified data set in which only one feature is randomly rearranged. The neural network will then be re-trained using the modified dataset and its performance will be evaluated. The extent of performance degradation serves as an indication of the importance of the feature. When the algorithm is applied to all features, a comprehensive insight into their correlations can be obtained.

The permutation feature importance test is a model-agnostic technique, which means that it can be applied to neural networks. Neural networks need to be developed in a later phase as well, so the model can also be used to determine the importance of the feature. The permutation feature importance technique also evaluates the numerical features which can also be used and interpreted. The technique offers a valuable tool for selection, interpretation, and understanding of the contribution of individual features to the overall performance of the model.

3.5 Model Development

The literature research provided information on the methodologies used in developing machine learning models for length-of-stay predictions. This section will dive into the workings of machine learning algorithms and present a comprehensive overview on how the different models are developed.

3.5.1 Machine learning

Machine learning techniques have made significant progress over the last two decades, and today they are the most rapidly growing technological field (Jordan & Mitchell, 2015). The availability of more powerful computers and vast data sources has allowed the use of machine learning techniques. Machine learning has various applications across a wide range of disciplines, with common applications including classification, regression, and clustering (Kourou et al., 2015).

The healthcare industry is also adopting the prevalence of machine learning techniques. Disease diagnoses and detection (Fatima & Pasha, 2017), treatment prognosis (Kourou et al., 2015), disease risk prediction, and health monitoring (Saleem & Chishti, 2020) all areas where machine learning is currently being implemented.

Machine learning, which is a subset of artificial intelligence, is concerned with the task of learning from data samples and relates this task to the broader concept of inference (Bishop & Nasrabadi, 2006; Mitchell, 2006; Witten & Frank, 2002). When building a machine learning model to complete tasks, a learning phase must be completed. The learning phase consists of two phases. In the first phase, unknown dependencies of a system will be estimated (Kourou et al., 2015). That is, the correlations between the input variables and the output variables need to be revealed. In the second phase, these found dependencies will be used to predict new outputs for the system (Kourou et al., 2015). The learning phase can use labelled, unlabelled, and partially labelled data, and the specific approach used is classified as supervised, unsupervised, or partially supervised machine learning, respectively (Ang et al., 2016). Each method has its own strengths and weaknesses, and the choice of approach will depend on the specific needs and characteristics of the data being analysed. When the learning phase is complete, the model enters the model testing phase where the performance of the model is evaluated on unseen data (Sarker et al., 2021). In Figure 7, the general structure of a machine learning-based predictive model is shown considering both the training and the testing phase. After the testing phase, the results will be evaluated by comparing the predictions with the ground truth.

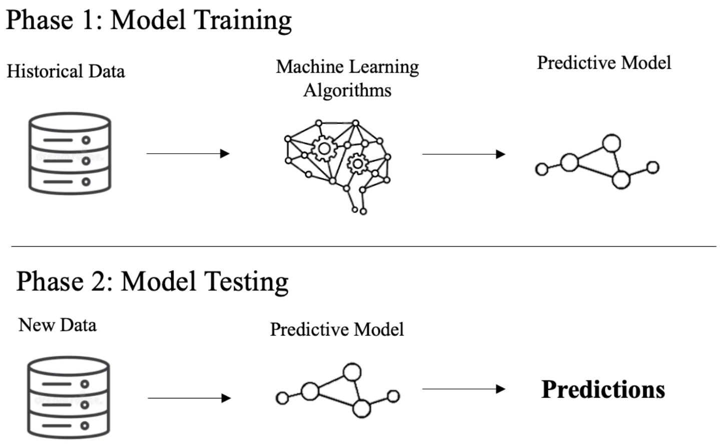

Figure 7. General structure of a machine (Sarker et al., 2021)

There exists a wide range of algorithms that are utilised to achieve desired objectives, such as linear regression, logistic regression, Decision Trees (DT), Random forest (RF), Naïve bayes (NB), Support vector machines (SVM), Gradient Boosting (GB), Artificial Neural Networks (ANN), K-Nearest Neighbours (KNN). The systematic review of the literature did not find a consensus on which model to use. Therefore, the machine learning models used in the systematic literature review of Bacchi et al. (2021) and the systematic literature review on length-of-stay predictions will be developed. These models are Random Forest (RF), Gradient Boosting (GB), Support Vector Machines (SVM), Decision Tree (DT), K-Nearest neighbour (KNN), XGBoost, Logistic regression (LR), Ridge Regression (RR), and Neural Networks (NN). The study conducted by Sarker et al. (2021) provides a valuable resource for understanding the inner workings of all machine learning models used in the thesis.

For the model development, two separate approaches are constructed, namely classification and regression. The development of both approaches is essential, as the results of the two independently operating models will be used for validation. Both approaches are types of supervised learning, as the target value, the length of stay, is known. According to Bacchi et al. (2021), these are the two types of predictions that are being used most. To designate the predictions of the length of stay as a classification problem, four different groups are specified.

The patient stays 0-2 hours (class 0), 2-4 hours (class 1), 4-6 hours (class 2) and 6 hours or more (class 3). The implementation of these exact groups is based on current estimates from planners. A schematic overview of the methodology is shown in Figure 8.

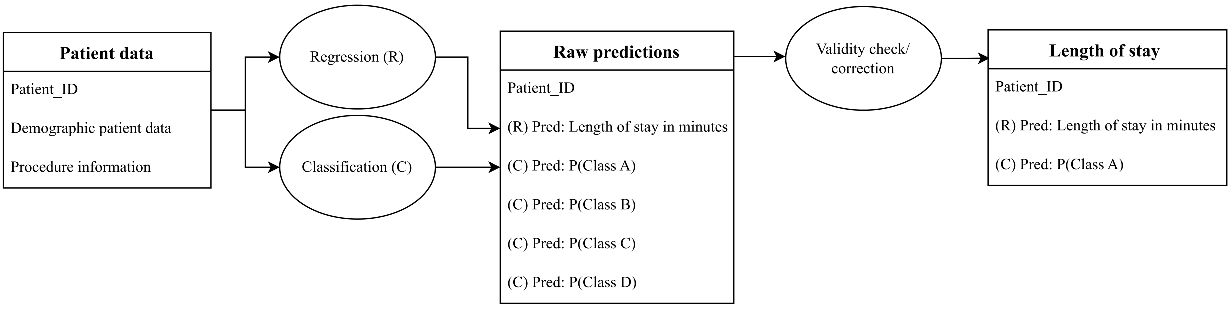

Figure 8. Overview of methodology of length of stay predictions

The proposed methodology comprises several sequential steps aimed at generating accurate length-of-stay predictions. Patient data and procedure information is inputted into both a regression model and a classification model. The regression model produces a numerical value representing the length of stay in minutes, while the classification model generates probabilities indicating the probability that a patient belongs to a specific class. For example, patient X can be assigned to class A (0-2 hours) with a probability of 0.90. The combination of these outputs is then utilised during the validation and correction step, as outlined in Section 3.6. The corrected and validated results subsequently serve as input to the ILP patient scheduling process.

3.5.2 Parameter Tuning

Machine learning models are derived by training a model on historical data. Training machine learning models comes with a pre-set number of parameters that alter the behaviour of the model. These parameters are termed hyperparameters (Lavesson & Davidsson, 2006; Probst et al., 2018). Hyperparameter tuning is a crucial process for optimising the performance of machine learning algorithms (Weerts et al., 2020; Yang & Shami, 2020). As the performance of many machine learning algorithms depends on their hyperparameter settings, it is essential to explore and select optimal values for these parameters to achieve qualitative results. The objective is to identify optimal settings and minimise time and cost through a few sequential queries for parameter tuning (Nguyen, 2019). With hyperparameter optimization, the goal is to find global optimum $\mathbf{x}^{\star}$of an unknown black box function $f$where $f(\mathbf{x})$ can be evaluated for any arbitrary $\mathbf{x} \in \mathcal{X}$ (Cho et al., 2020). The mathematical notation is depicted in Equation 3.

$$\mathbf{x}^{\star} = \underset{\mathbf{x} \in \mathcal{X}}{\arg max} \, f(\mathbf{x}) \tag{3}$$

Given the numerous models to be developed and evaluated, along with extensive search spaces and multiple parameters for each model, the implementation automation of hyperparameter tuning is required. Bayesian optimisation has been recognised as a technique for addressing various design problems in the realm of parameter optimisation (Shahriari et al., 2016). According to Shahriari et al. (2016), the Bayesian optimisation technique has demonstrated its superiority over human experts in terms of both quality and speed of tuning. This claim is further supported by the findings of Snoek et al. (2012). The purpose of Bayesian optimisation is to discover the global optimum of the function f(x) by constructing a probabilistic model for f(x), which represents an unknown function to be optimised. Bayesian optimisation uses the probabilistic model to make informed decisions on where to evaluate the function next within parameter space X, effectively accounting for uncertainty (Nguyen, 2019). The algorithm's implementation is facilitated by the Bayesian Optimisation class from Keras (Chollet & others, 2015). A total of n=30 iterations of Bayesian optimisation were performed for each model.

The search space was deliberately defined with a wide range by setting a conservative lower bound and an expansive upper bound. The objective is to encompass a wide range of parameter configurations, taking into account complexity and computational efficiency. The upper and lower bounds were established to ensure a thorough exploration of feasible values within the practical range. It should be noted that the potential values might not have been explicitly included in the search space. The complete search space can be found in Appendix C accompanied with explanation of the parameters, and, as an example, the Random Forest search space is displayed in Table 8.

Table 8. Hyperparameter tuning search space for Random Forest

Hyperparameter	**Type**	**Range**
Number of estimators	Integer	[10,200]
Maximum features	Categorical	{auto, sqrt, log2}
Maximum depth	Integer	[10,100]
Minimum sample split	Integer	[2,10]
Minimum sample leaf	Integer	[1,4]
Bootstrap	Categorical	{True, False}

3.6 Model validation and evaluation

Accurate validation and evaluation of predictive machine learning models is paramount in scientific research. It enables the researcher to assess the reliability and performance of the machine learning models constructed. In the first part of this section, the validation methodology of the models will be explained. Later, the evaluation of the different models will be explained.

Validation of machine learning models will be two-fold. The first technique is to divide the data into train, validation, and testing data. The systematic review of the literature revealed that a significant number of articles used an 80/20 division for train test data. Therefore, the constructed training set contains 80% of the records and is used to train the models. Within this 80%, 10% of the training set is used as validation data to enhance the model during the training process. The last 20% of the data is reserved as test data to evaluate performance. The structure of the different data sets can be seen in Equations 4a, 4b, 4c.

$$D = D_{\text{train}} \cup D_{\text{val}} \cup D_{\text{test}} \tag{4a}$$
$$\mathrm{d}_1 = D_{\text{train}} \cap D_{\text{val}} \tag{4b}$$
$$\mathrm{d}_2 = D_{\text{val}} \cap D_{\text{test}} \tag{4c}$$

The total data set D contains all records of all patients. Set d_1 consists of the records used for model training and enhancement. Set d_2 contains records that are utilised for evaluating the model's performance after training. Set d_2 contains data that are used to evaluate model's performance based on the validation process. It helps to evaluate the model's ability to generalise beyond the training data.

The second technique is the k-fold cross-validation procedure. Cross-validation is the most widely used data resampling method to estimate the prediction error of the models (Berrar, 2018). In principle, cross-validation is repeatedly splitting the data into different subsets k and for each subset k, estimating the parameters and evaluating (Emmert-Streib & Dehmer, 2019) to prevent overfitting (Simon, 2007). The k-fold cross-validation will be used on dataset D which is portioned into k disjoints where k refers to the number of subsets (Berrar, 2018). The models are applied to all subsets of k, and the average performance in all subsets of k is the cross-validated performance. Cross-validation can take a considerable amount of time for higher values of k. The systematic review of the literature showed varying values for k, but since many models must be validated, a $k = 3$ will be used for computational reasons.

The validated models will then be assessed based on visualisations and performance indicators. These methods differ according to prediction technique classification (C) or regression (R). Different metrics serve distinct purposes to provide valuable information. The systematic review of the literature provided generally used validation metrics. The same validation metrics will be used in this research to compare with other articles. In addition, additional performance indicators will be provided to allow other researchers to compare to their findings. For the classification task, accuracy is the most common measure used to identify the proportion of correctly classified instances among all instances. Precision and recall are additional performance indicators that provide more insight. The recall measures the proportion of true positive classifications among all positive classifications, whereas the recall measures the proportion of true positive classifications among all positive instances. The harmonic balance between the two is the F1 score. Evaluation of the model should focus on its ability to identify and classify instances of both classes, rather than solely emphasising the majority class. Additionally, for the best performing model, the ROC curve will be constructed, and the AUC will be calculated per class to provide better insight into the model. The AUC was proven to be a better evaluation metric for classification evaluation (Hossin & Sulaiman, 2015) and provides information on the performance of different classes. The AUC can be deducted from the ROC curve and can be calculated according to Equation 5.

$$AUC = \frac{S_p - \frac{n_p(n_n+1)}{2}}{n_p n_n} \tag{5}$$

Where S_p is the sum of all positive ranked examples while n_p and $n_n$denote the number of positive and negative examples respectively (Hossin & Sulaiman, 2015).

For the regression task, three types of performance indicators are used. The R^2 score is a common metric that measures the proportion of variance in the dependent variables that is explained by the independent variables. The R^2 score is accompanied by the MSE, which is the average of the squared differences. This is a useful parameter for comparing models in their performance, but not as insightful for humans. Therefore, MAE is also included. This is the mean absolute error, which in this case will indicate the average number of minutes when a prediction is off. In addition to the evaluation metrics mentioned above, a visual evaluation will also be included. By plotting the true length of stay against the predicted length of stay for regression, a graph will show how well the model is performing.

On the basis of the performance indicators, two separate models will be chosen. The classification model will be chosen on the basis of the accuracy. The accuracy is the average performance of the model and will therefore be the best indicator to indicate the best-performing model. For regression machine learning models, the decision will be based on the R^2 score. The R^2score, a statistical measure that indicates the model's ability to explain variability, will be a reliable metric to choose the best-performing model. The R^2 score will reflect superior predictive abilities to capture underlying patterns and trends related to length of stay. In Table 9, an overview of the performance indicators for classification and regression will be presented.

Table 9. Overview of Performance Indicators for Classification and Regression Models

Task	Metrics	Formula	Explanation
Classification	Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Proportion of correctly classified instances among all instances
	Precision	$\frac{TP}{TP + FP}$	Proportion of true positive classifications among all positive classifications
	Recall	$\frac{TP}{(TP + FN)}$	Proportion of true positive classifications among all actual positive instances
	F1 score	$2 * \frac{(Precision * Recall)}{(Precision + Recall)}$	Harmonic mean of precision and recall
Regression	R^2	$1 - \frac{RSS}{TSS}$	Coefficient of determination. Proportion of variance in the dependent variable that is explained by the independent variable(s)
	MSE	$\frac{1}{n} * \sum (Y_i - \hat{Y})^2$	Mean Squared Error. Average of the squared differences between the actual and predicted values
	MAE	$\frac{1}{n} * \sum \lvert Y_i - \hat{Y} \rvert$	Mean Absolute Error.

Note: In these formulas, TP = True Positives, TN = True Negatives, FP = False Positives, FN = False, RSS stands for "Residual Sum of Squares", and TSS stands for "Total Sum of Squares".

3.7 Correction and validation

The reliability and accuracy of the length-of-stay predictions generated by the machine learning algorithms are of utmost importance to ensure effective planning. In Section 3.6, the validation of machine learning models was extensively discussed. However, for individual predictions, it is essential to implement additional precautions measures to prevent underestimation.

For individual predictions, it is necessary to assess the quality. In case of a correct prediction or an overestimation, no problem would occur as the patient is discharged before the next patient is scheduled. Underestimations can impose a significant logistical problem since there is no bed available for the newly arrived patient. In both systematic literature reviews, no literature was found that addresses this exact problem. In the systematic review of the length-of-stay literature, the length-of-stay predictions are not actively implemented in planning systems. In the systematic review of planning literature, only probability distributions on the length of stay are used to plan patients.

To account for the uncertainty in extended length of stay, two layers of validation are implemented. The first layer entails that the output of the two models must agree. For example, if the classification model predicts class 0-2 hours and the regression model predicts 90 minutes, the patient will be included as a candidate for the patient planning model. If the output of the two models does not agree for a certain patient, the patient will be excluded from the scheduling as the risk of an extended length of stay is too high. By implementing this strategy, the goal is to reduce the probability of an underestimation and therefore prevent logistical problems. The novel approach of combining two independent operating models needs to be tested in depth to assess the performance.

The second layer of validation is related to the classification model. The output of the classification model is given by the probability that a patient belongs to a specific class. The certainty of a class must exceed the accuracy of a patient being in a class with 90%. The 90%

is conservatively chosen, as this novel technique is unverified by the literature and is not physically evaluated in this research. When the envisioned system is implemented or is simulated, the 90% can be adjusted accordingly to a different accuracy score to include more patients within the pool of schedulable patients. The exclusion of patients by these two layers of validation induces the risk that the patient is not planned at all. Therefore, the proposed model schedules can only serve as a guide for human planners to increase bed occupancy in places where possible.

3.8 Patient scheduling

The previous section elaborated on the prediction methodologies, validation, and evaluation of models. The output of the prediction methods, discussed in Section 3.5, is validated and corrected, as explained in Section 3.7. The output of these models serves as an input for patient scheduling. A systematic review of the literature revealed the prevalence of (mixed) integer linear programming algorithms for patient planning. In this research, the same methodology is used as in the reviews. The patient scheduling ILP formulation consists of parameters, decision variables, an objective function, and constraints.

Integer linear programming is a mathematical optimisation technique that solves problems with linear objective functions and constraints (Vielma, 2015). ILPs formulate the problem with decision variables, an objective function, and constraints. It uses specialised algorithms, such as branch- and bound or cutting-plane methods, to find optimal integer values by exploring the search space in a systematic manner. ILP is useful for complex optimisation problems that involve discrete decisions in various real-world applications.

The objective of ILP in this research to maximise the number of patients planned, in line with the main objective of this study. There are several inputs to the ILP. The first input is the number of beds available to plan patients. The number of beds limits the outcome of the ILP. The second input concerns the available time that the model can use to plan patients. The opening and closing times can be adjusted by the planners according to the available time in the hospital. The third input comprises the output set of plannable patients from the waiting list, accompanied by their corresponding length of stay predictions.

ILP is subject to practical constraints, including the requirement that each bed can accommodate only one patient at a time and that a new patient can only be planned once the previous patient has been discharged. Additionally, patients can only be planned within the hospital opening hours. The planning horizon is determined by the specified start and end time of the model, typically encompassing a few hours set by the availability on the department. Solving the ILP results in a schedule with patients that can be scheduled consecutively to increase bed occupancy.

3.9 Software and Tools

Accomplishing the envisioned system requires programming. In this study, the Python programming language was used in combination with R. In this section, the various libraries used will be explained.

For data pre-processing and analysis in R, the following libraries are used: data.table (Wickham et al., 2020), dplyr (Wickham et al., 2020), and ggplot2 (Wickham, n.d.). These libraries provided functionalities for data manipulation, filtering, summarization, and visualisation. Machine learning tasks were performed using various libraries in Python. Scikit-learn (Pedregosa et al., 2011) provided a wide range of machine learning algorithms and evaluation metrics. Matplotlib (Hunter, 2007) was used for generating data visualisations. The Keras library (Chollet et al., 2015) was used to construct and train neural networks. The hiplot library (Haziza et al., n.d.) allowed interactive visualisation and analysis of high-dimensional data. The Keras_tuner library (O'Malley et al., 2019) facilitated hyperparameter tuning for Keras models. XGBoost (Chen & Guestrin, 2016) was used for gradient boosting. The scikit-optimise library (Skopt Contributors, 2020) provided tools for Bayesian optimisation. For ILP scheduling, the literature review revealed that CPLEX was the commonly preferred methodology; however, due to the researcher's proficiency in Python programming, the PuLP library developed by Mitchell and Stuart (2011) is used to solve the ILP. Pandas (McKinney, 2010) was used for data manipulation and analysis. NumPy (Harris et al., 2020) provided support for numerical computations. The TKinter library (Python Software Foundation, n.d.) was used to create a graphical user interface for the system. The PIL library (Pillow Contributors, 2022) facilitated image processing and display in the graphical user interface. The selection of libraries is based on their functionalities and the researcher's personal preference, ensuring compatibility with the research objectives and requirements.

3.10 Conclusions

Based on the comprehensive methodology employed, this research aims to increase bed occupancy at the small regional hospital by accurately predicting patient length of stay and efficient patient scheduling. The methodology used in this study encompasses the key steps in achieving the objective. The methodology involves steps such as data collection, preprocessing, exploratory data analysis, feature selection, and the development and evaluation of state-of-the-art machine learning models. The two-layer validation and correction process ensure the reliability of the predictions. Patient scheduling will be accomplished using an ILP algorithm to increase bed occupancy and resource allocation. The next chapter will dive into the results achieved by following the methodology.

4 Results

The results section of this study presents the findings and results of the research, highlighting important discoveries, analyses, and conclusions. It provides a detailed overview of the research process, including exploring and analysing data, evaluating machine learning models, and developing a patient scheduling tool.

4.1 Exploratory data analysis

The exploratory data analysis can be divided into two sections. In Section 4.1.1, the population at hand will be evaluated on their statistics. The distributions over specialisations and over the different time phases will be evaluated. In Section 4.1.2, correlations are explored among various factors that relate to patient stay, shedding light on possible associations and underlying patterns.

The dataset at hand is pre-processed. The removal of duplicates and outliers resulted in the data set that will be used. The dataset contains 21,545 patient records collected for 21,421 unique patients (mean age 50.7 ± 17.05 years, with 51.12% females) are collected from the A2X department between 2018 and 2022. In the next section, the dataset will be studied in depth.

4.1.1 Patient exploration

Understanding the distribution of the length of stay in different medical specialities is vital for understanding the patient population in the hospital. This analysis allows for an assessment of the impact of each specialisation on the overall flow of patients within the department. To provide a visual overview of this distribution, boxplots are utilised as a powerful tool. The box plots offer a clear depiction of the distribution of stay durations for each medical speciality. This enables the identification of variations in length of stay and the detection of potential patterns or outliers. Figure 9 shows the box plots that show the length of stay distribution for each specialisation.

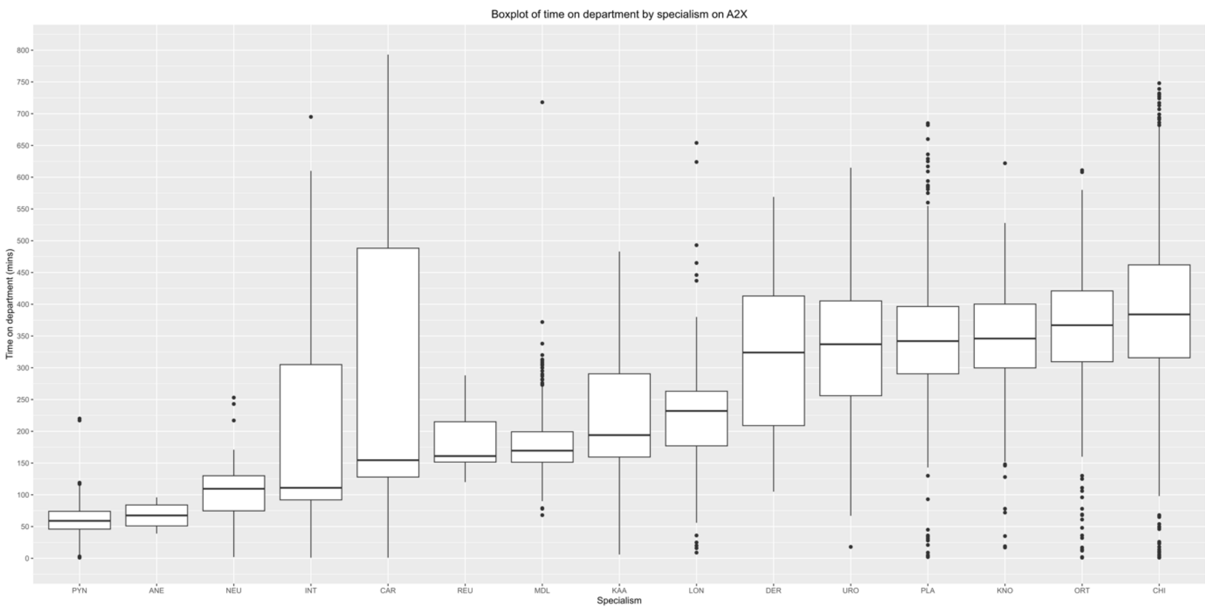

Figure 9. The time on the department per specialisation

Examination of the boxplot reveals significant variations in the median and interquartile range (IQR). In particular, the mean of the length of stay for each specialty ranges from μ=60.3 to μ=402.2, indicating a wide and diverse distribution of the duration of stay. A remarkable observation is the specialisation in pain. Pain specialisation has a very low median and a small interquartile range, which may suggest easier predictability. A remarkable observation is found in the pain specialisation, where a notably low median and a small interquartile range are observed. This suggests a potential for easier predictability in length of stay for patients in this specialisation. This information is valuable as patients with short length of stay of can be scheduled consecutively, enabling more efficient planning processes.

Figure 10 presents an analysis of the frequency of the thirty medical procedures that occur according to their medical procedure code (COTG). There is a significant difference in the amount a specific procedure is conducted at the hospital. The number of patients with a specific COTG procedure performed influences the prediction capabilities of the models. The availability of a larger volume of data for a specific procedure allows for more extensive information and enhances the predictive capabilities of the model for those particular procedures.

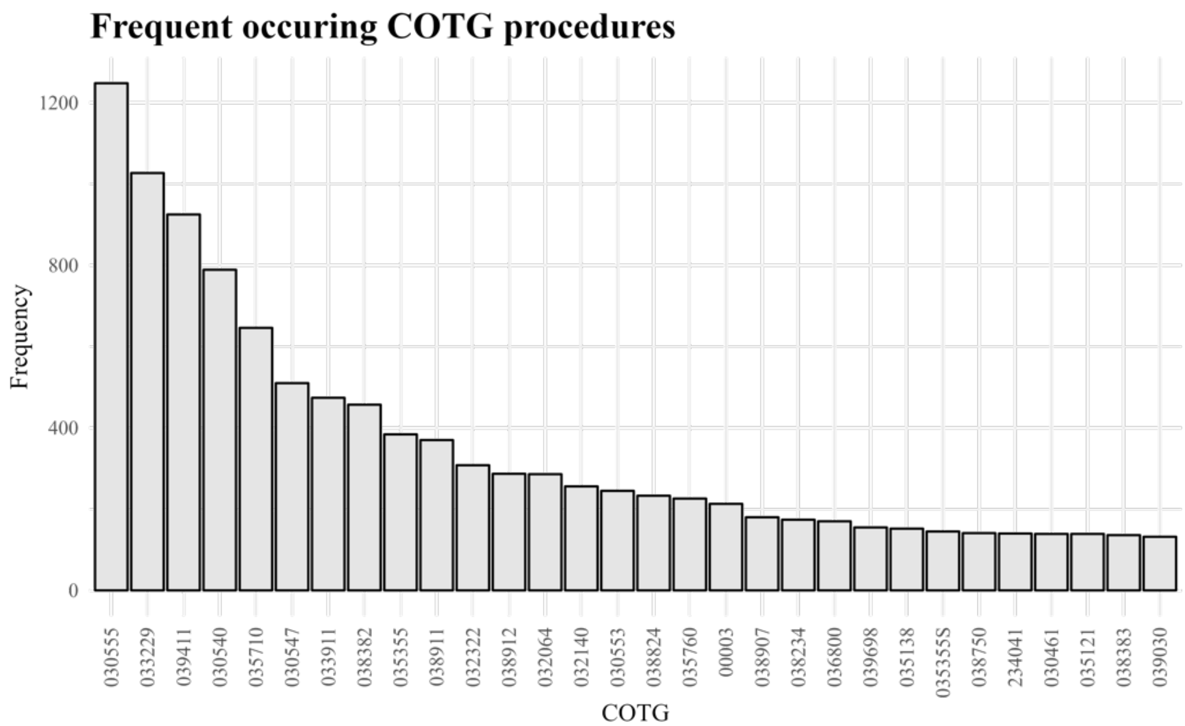

Figure 10. Frequency histogram of most COTG codes

As an example, COTG 030555 exhibits an annual frequency of approximately 1200 occurrences, while other procedures have only a few instances per year. The availability of the amount of data for specific procedures plays a crucial role in the predictive accuracy of the model. Insufficient support for certain procedures within the training set may lead to reduced information availability. As a consequence, the quality of the predictions for these specific procedures can be negatively affected.

The objective of this research is to increase bed occupancy at small regional hospital. Therefore, it becomes essential to examine the current state of bed occupancy. To gain insight into the present state, the data from 2022 was utilised to calculate the average occupancy of beds. Equation 2 described in Section 3.3 was used to calculate the bed occupancy. The bed occupancy is calculated on daily and montly basis together with the mean. Figure 11and Figure 12 present bed occupancy per day and per month, respectively, providing visual representations of bed occupancy.

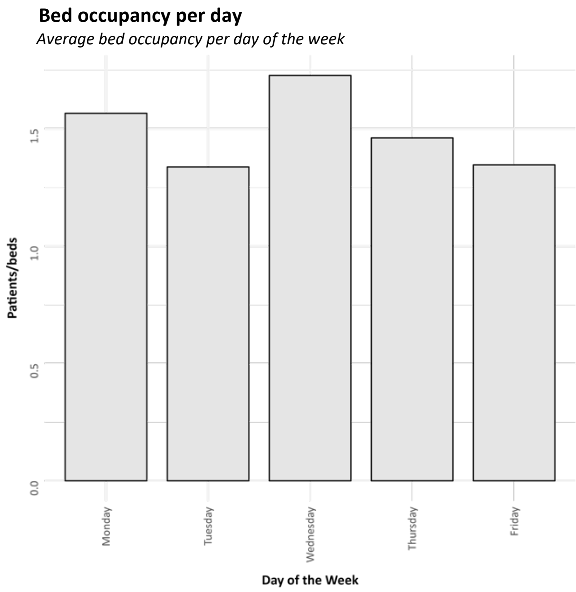

Figure 11. Bed occupancy per week

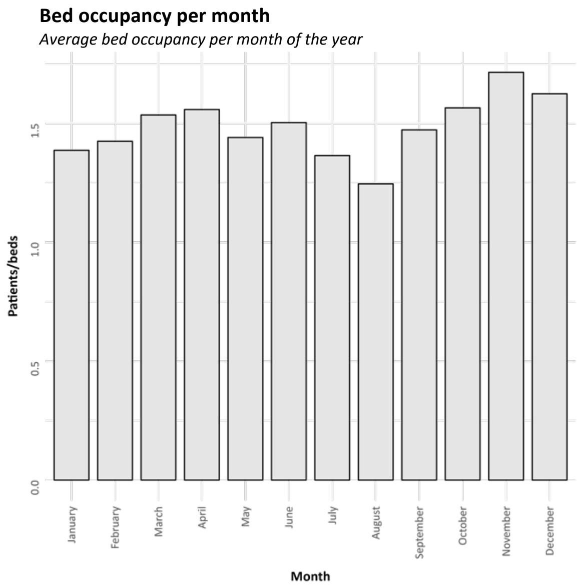

Figure 12. Bed occupancy per month

In 2022, the mean occupancy of the bed was found to be $\mu = 1.47$. It should be noted that the occupancy of the bed of 1.47 is highly influenced by the pain department. The pain department has already achieved a remarkable level of advancement in efficiently scheduling patients consecutively. The efficient scheduling can be attributed to the relatively short length of stay for patients in this specialised field. By analysing weekly bed occupancy, no significant trend is observed. However, when examining the monthly bed occupancy, a slight trend towards higher bed occupancy becomes apparent. This trend can be attributed to the efforts of the hospital to improve bed occupancy through improved planning and efficiency measures. Understanding the distribution of the length of stay is crucial to see whether there are patients that can be planned sequentially. The length of stay is presented as a histogram in Figure 13.

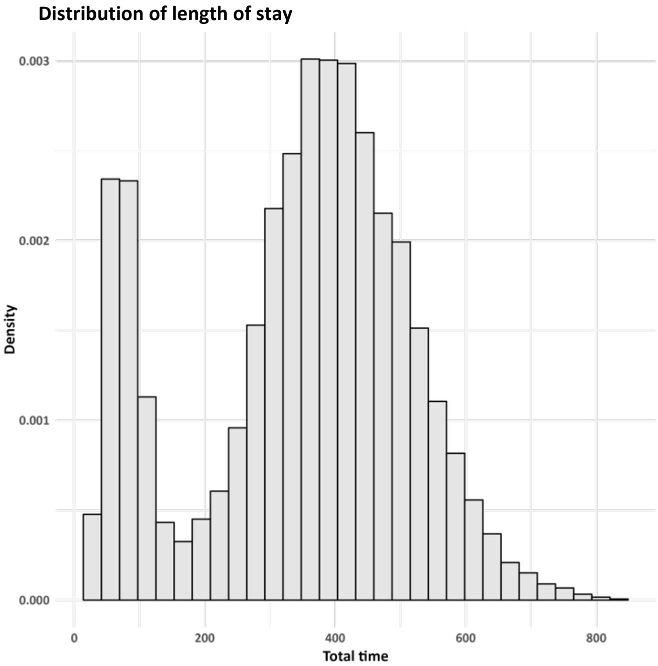

Figure 13. Probability distribution of length of stay

An interesting observation found in the distribution of time within the total time is the existence of two distinct peaks. The first peak represents patients who are discharged from the hospital in a few hours. The second peak represents a substantial group of patients who experience a considerably longer stay. The two different peaks can be attributed to the type of procedure in combination with the anaesthesia technique used. Patients under general anaesthesia have a longer stay in the daycare department due to the medication used. Variation in length of stay highlights an interesting pattern that can be used to improve bed occupancy. Especially patients at the first peak can be used to increase bed occupancy due to the short length of stay.

4.1.2 Feature selection

Continuing from Section 4.1.1, the focus now shifts toward exploring correlations among factors associated with length of stay. The objective is to uncover valuable information on the associations and patterns that influence length of stay. In this section, rigorous analysis and statistical techniques are used to find the correlation between variables and length of stay. The data set contains two types of variables, numerical and categorical. The influence of numerical variables on the length of stay can be deducted via a correlation map. In case of the categorical variables, only permutation techniques are utilized to identify significant contributing variables. Figure 14 depicts a correlation map of the available numerical data.

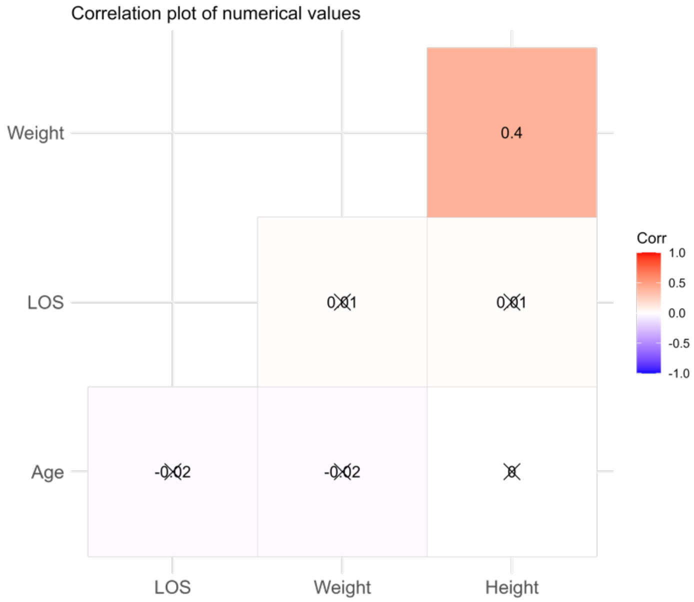

Figure 14. Correlation map of numerical data

The constructed correlation plot provides a visual representation of the correlations among the variables. Upon analysing the correlation plot, it becomes evident that the correlations involving the variable length of stay are insignificant, as indicated by the crosse. As expected, there is a significant correlation between weight and height. These results suggest that there is no direct correlation between these variables and the length of stay. Although there is no direct influence on stay length, indirect correlations cannot be evaluated using this method. To evaluate the indirect influence of numerical variables and the influence of categorical variables, the permutation technique as described in Section 3.4, is used. A temporary neural network was constructed and the permutation technique was executed. Figure 15 presents the result of this technique.

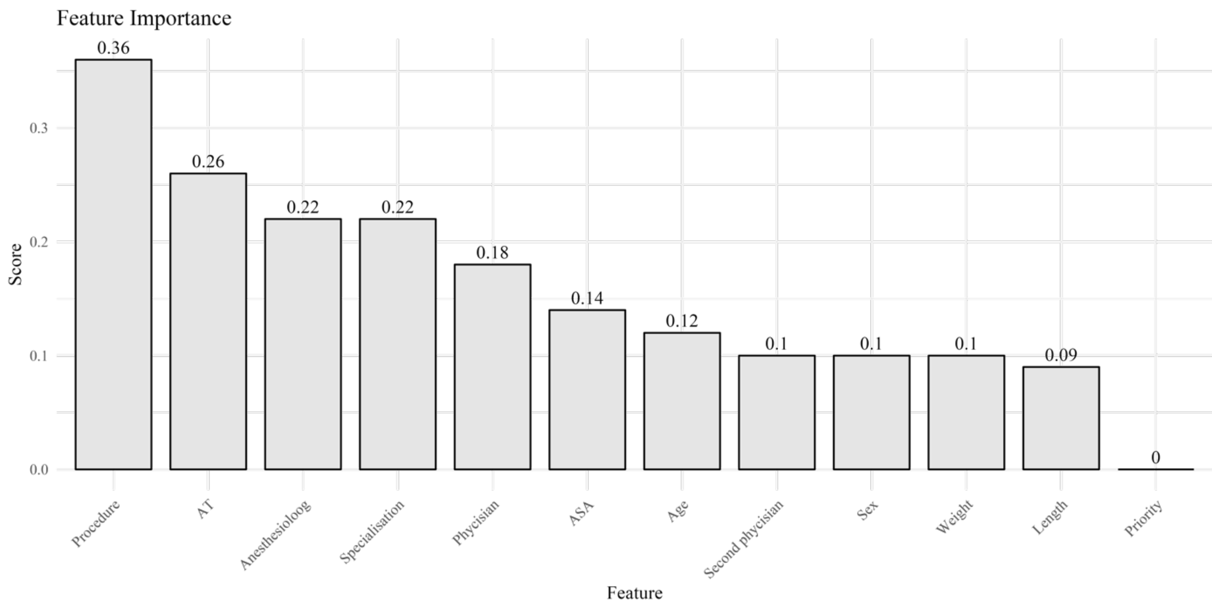

Figure 15. Feature importance calculated by the permutation technique

The results of Figure 15 show that all variables except 'priority' negatively influence the performance of the model when the data in the column are randomised. As opposed to the results of the correlation plot, the results of the permutation technique from Figure 14, display that the numerical values are correlated with the length of stay. No significant direct correlation was found in Figure 14, but it can now be concluded that they indirectly influence the performance of the model. For this reason, all variables will be included in the data set to be used during the model development phase. Table 10 shows all included features.

Table 10. Features included in the LOS prediction model

Features	
Sex	*Physician*
Weight	*Second physician*
Length	*Physical status patient (ASA)*
Age	*Procedure code (COTG)*
Specialisation	*Anaesthesia technique (AT)*

4.2 Results of the models

The methodology of Chapter 3, the machine learning models to be tested, are outlined. This section presents the results obtained from the evaluation of these models, which are shown in Table 11. The table presents the performance metrics for both classification and regression tasks. These include accuracy, precision, recall, R^1 score, R^2 score, mean squared error, and mean absolute error.

Table 11. Overview of the results of machine learning models

Methods	**Classification**				**Regression**		
	Accuracy	**Precision**	**Recall**	**R1**	**R2**	**MSE**	**MAE**
Random forest	0,772	0,714	0,657	0,672	0,780	5696	55,789
Gradient Boosting	0,762	0,705	0,681	0,686	0,790	5679	56,069
Support Vector Machines	0,757	0,696	0,664	0,673	0,770	5954	56,169
Decision tree	0,756	0,706	0,649	0,662	0,750	6472	59,491
KNN	0,750	0,689	0,641	0,654	0,757	6280	58,203
XGBoost	0,772	0,719	0,686	0,697	0,784	5588	55,864
Linear Regression	0,765	0,708	0,673	0,682	0,744	5737	55,873
Ridge regression	0,759	0,698	0,654	0,653	0,776	5798	56,023
Neural Networks	0,769	0,710	0,663	0,653	0,788	5656	55,597

Based on the results presented in Table 11, several conclusions can be drawn. In terms of classification tasks, the models were evaluated on the basis of accuracy. The Random Forest outperformed the other models with an accuracy score of 0.772. This indicates that the Random Forest model is well suited to accurately predict the target variable in the classification task.

In the context of regression tasks, models were evaluated based on their performance using the R^2 score. In particular, the neural network model demonstrated decent performance, with an R^2 score of 0.788. This score indicates a robust predictive capacity of the neural network to predict the length of stay. A noteworthy metric is the mean average error associated with the regression task. For the neural network, the MAE was calculated to be 55.597. The value represents the average deviation, measured in minutes, between the predicted and actual stay. The relatively low MAE underscores the model's capacity to provide reliable estimations.

Another remark is that all models show similar performance. One plausible explanation for the comparable performance of these models could be that the data set used lacks distinct patterns or relationships that can be effectively captured by any of the models. Another explanation could be that the data set is too small, limiting the ability of the models to demonstrate their full potential.

4.2.1 Classification

The previous section highlighted the two best performing models for the classification and the regression task based on accuracy and the R^2 score. In this section, the performance of the classification model will be evaluated in depth.

The best-performing classification model is, as stated, Random Forest with an accuracy of 0.772. The accuracy score is in this case not the only important factor; it is also important to look at the precision per class. The results of the precision per class can be found in Appendix D. Table 12, presents the findings for the Random Forest model per class.

Table 12.Precision, recall, and F1 score per class for Random Forest

Class	Precision	Recall	F1 score	N
A	0,9286899	0,9807356	0,9540034	571
B	0,544	0,3541667	0,4290221	192
C	0,6027944	0,3719212	0,4600152	812
D	0,7802632	0,9198552	0,8443284	1934
Accuracy	0,7720148	0,7720148	0,7720148	0,7720148
Macro avg	0,7139369	0,6566697	0,6718423	3509
Weighted avg	0,7504212	0,7720148	0,7505192	3509

Analysis reveals that precision scores vary between different classes. Class A exhibits a high precision score and extremely high recall, which is advantageous because it indicates that these procedures can be planned consecutively with high reliability. On the contrary, the remaining classes exhibit lower precision scores, indicating that the predictions for these classes are poor. The knowledge obtained suggests that prioritising the planning of individuals of class A is advisable. Class A consists primarily of patients with shorter stays and demonstrates a higher level of reliability in the predictions.

In addition to evaluating precision scores for different classes, the performance of the classification model can be further assessed using the ROC curve. ROC curves provide information on the model's ability to classify patients correctly across different thresholds. By plotting the true positive rate against the false positive rate, the ROC curve visualises the trade-off between sensitivity and specificity. The constructed ROC curve in 18 shows the performance of the model in all four classes, as well as macro and micro averages. The macro-average calculates the average performance across all classes, treating each class equally. The micro-average aggregates the total true positives, false positives, and false negatives across all classes to calculate the overall performance.

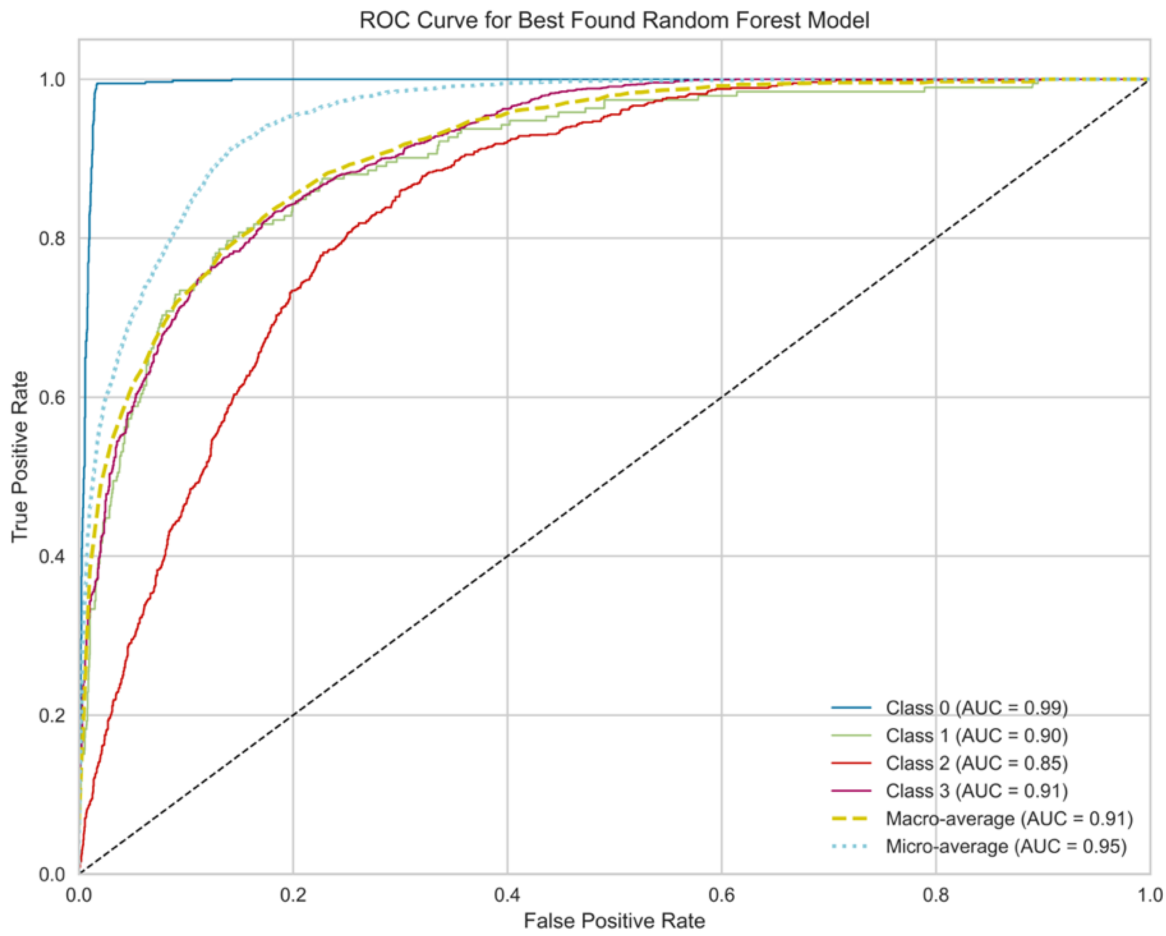

Figure 16. ROC curve for random forest with classes separated

Analysing the ROC curve for the classification model reveals the individual area under the curve (AUC) scores. Class A demonstrates a high AUC of 0.99, indicating a strong discriminative ability to accurately classify instances within this class. Class B exhibits an AUC of 0.9, suggesting a good level of discriminatory power. Class C and D show AUC values of 0.85 and 0.91, respectively, indicating moderate discriminative abilities. In addition to the class-specific AUC scores, the macro-average AUC is found to be 0.91. This score represents the overall discriminative power of the model across all classes. The micro-average AUC of 0.95 indicates a strong overall performance of the classification model in accurately classifying instance scores across the entire dataset.

4.2.2 Regression

Having completed the evaluation of the classification model, the attention now turns towards the regression task. Among the regression models, the neural network achieved the highest R^2 score = 0.788 and the lowest MSE = 5656 with MAE = 55,59. The R^2 score means that the model was able to explain 78.8% of the variance in the duration of stay predictions. The training history of the neural network is shown in Figure 17.

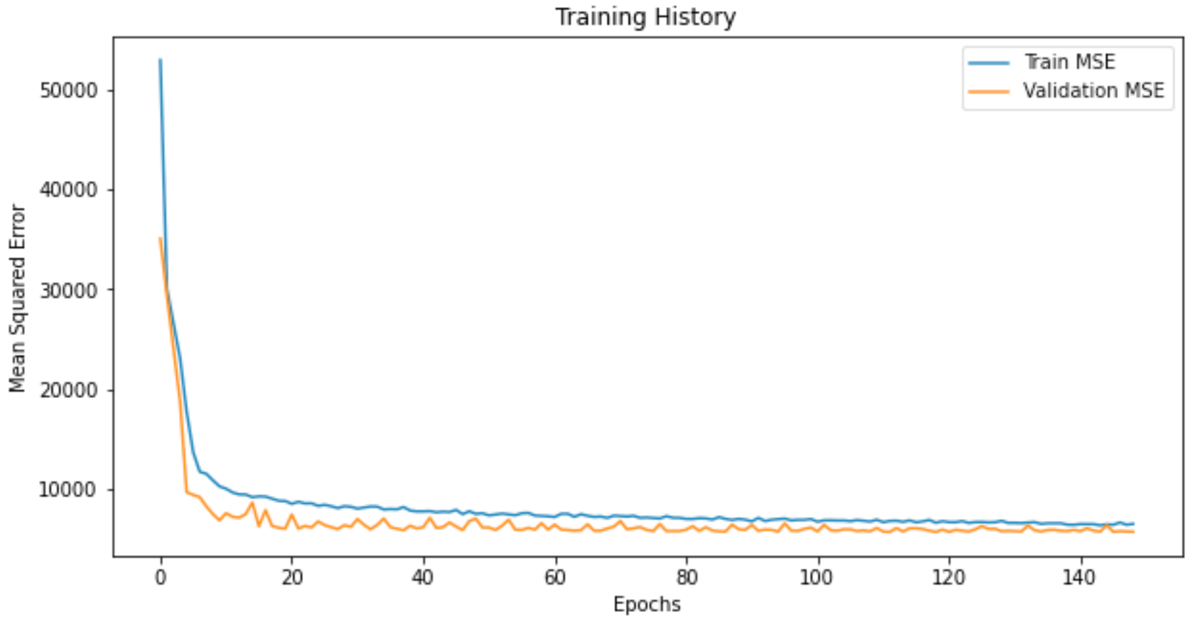

Figure 17. Training History Neural Network

The training history of the neural network is depicted through a visualisation. The figure shows the MSE values plotted against the number of training epochs. Initially, the MSE exhibits relatively high values, indicating the initial difficulty of the model in making accurate predictions. However, as training progresses, a gradual decrease in MSE is observed. This decline signifies the learning process of the model. The convergence of the MSE further suggests the effective capture of underlying patterns and relationships by the model. A remarkable finding illustrated in Figure 17 is that MSE training and MSE validation are closely related. This observation implies that the model is not overfitting, as the validation MSE closely mirrors the observed training MSE. Furthermore, the training history demonstrates a smooth trajectory without significant fluctuations, indicating an appropriate learning rate. The accuracy of the predictions can be assessed by visually comparing the predicted values against the true values of the length of stay. Figure 18 shows the plot that illustrates this relationship.

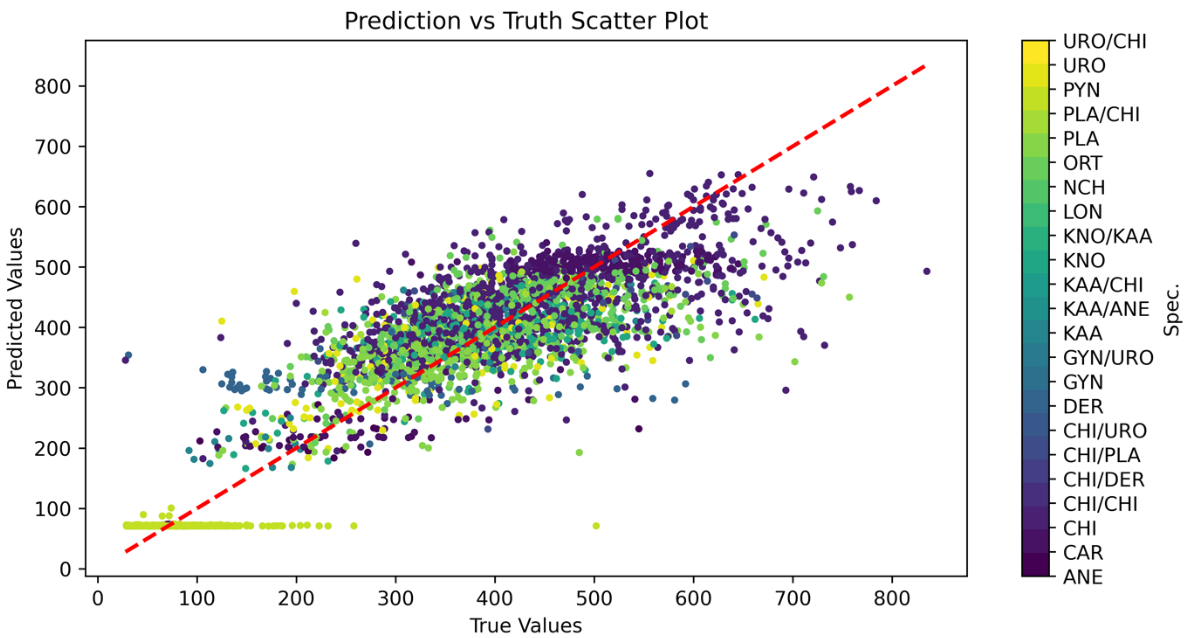

Figure 18. Predicted LOS plotted against true LOS per specialisation

In this plot, the closer the points align with the diagonal dotted line, the more accurate the prediction. A closer proximity to the actual value signifies better prediction performance. In general, the model shows reasonable performance, which is also confirmed by the R^2 score. It must be stated that there are also many predictions that differ significantly from the true length of stay. Predictions that are underestimated are a significant issue when planning patients. Section 4.2.3, dives into the validation and correction methods.

A remarkable observation can be found within the pain (PYN) specialisation, specifically in the pain department. Section 4.1.1 suggested accurate predictions as the specialisation has a small IQR. Figure 18 contradicts these assumptions. The model always predicts the same length of stay for each patient. When diving into this, it happens that the data from patients in the pain specialisation are all similar. There is only one physician and only a few procedures conducted in the small regional hospital. This affects prediction capabilities as the model cannot distinguish between patients.

The model is unable to always produce a perfect prediction of the length of stays of patients. The findings in Figure 18 underscore the importance of further validation and corrections. In the next section, the correction and validation methods are implemented and evaluated.

4.2.3 Correction and validation

In the hospital, ensuring the accuracy or avoidance of overestimation in the model's predictions is crucial. To avoid overlap and streamline operations, hospital physicians schedule one patient in the morning and another in the afternoon. This careful scheduling approach ensures that no logistical conflicts occur. The developed tool needs to take into account additional correction and validation to achieve the same level of safety. If the system underestimates, it could lead to a situation in which the next patient arrives before the current patient sharing the same bed is discharged. This overlap between patients can cause several problems. First, the operation of the second scheduled patient must be delayed due to the inaccessibility of a prepared bed. Second, the extended waiting time leads to patient dissatisfaction. Lastly, delayed patients can exceed hospital operating hours, resulting in undesirable overtime for nurses. Therefore, it is imperative to incorporate additional safety measures before building the schedule. The research aims to increase the occupancy of the beds in the hospital, focussing on short-stay patients with a high probability of having a short stay. These patients are patients classified as class A, indicating a predicted stay duration of 0-2. The prediction capabilities of the RF model have an accuracy of 0.90 percent. Therefore, these predictions will not be corrected if the neural network outputs a length-of-stay prediction under 120 minutes.

In addition to patients of class A, patients of class B will also be considered. These patients have a short stay of 2-4 hours; however, since the prediction accuracy within this class is significantly lower, 0.54, a correction will be implemented. Correction involves adding extra minutes to the regression method to account for uncertainty. The extra number of minutes added to the expected length of stay contributes to a higher rate of correct/overestimated predictions. Figure 19 shows the relationship between the minutes added and the percentage correct/overestimated.

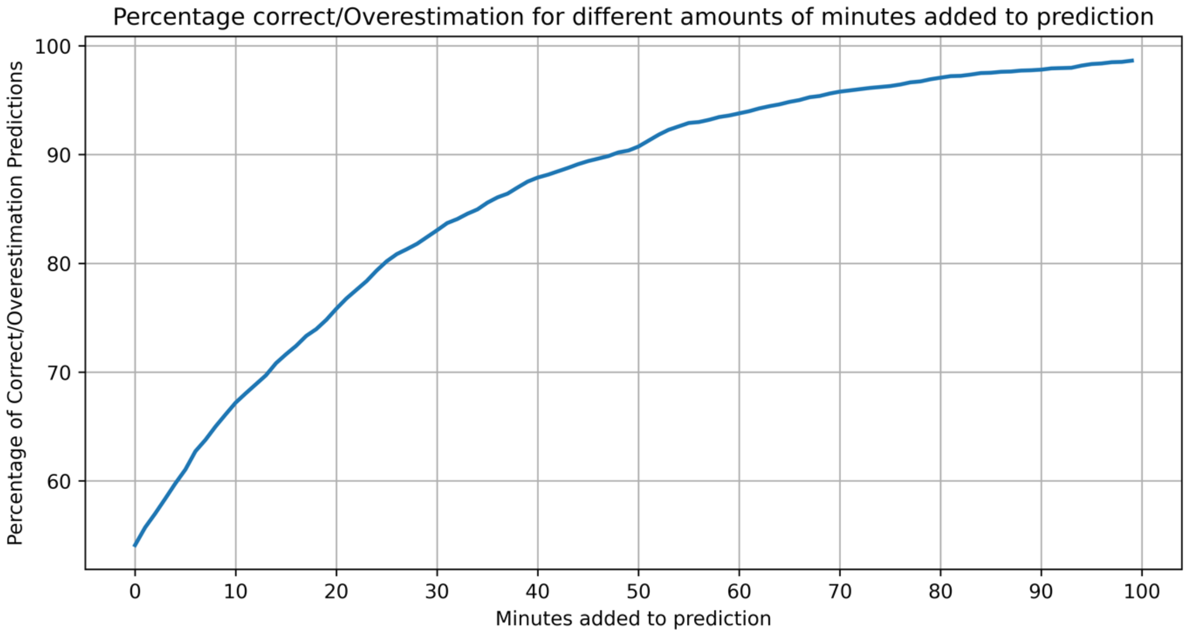

Figure 19. Percentage correct/overestimated against minutes added to the prediction

Adding 50 minutes to the regression prediction yields an impressive 90.4% accuracy for correct predictions or overestimations. This high percentage ensures a reasonable margin to avoid any potential overlap between patients. Figure 20 presents a scatterplot that illustrates the number of correct predictions or overestimations for adding 50 minutes.

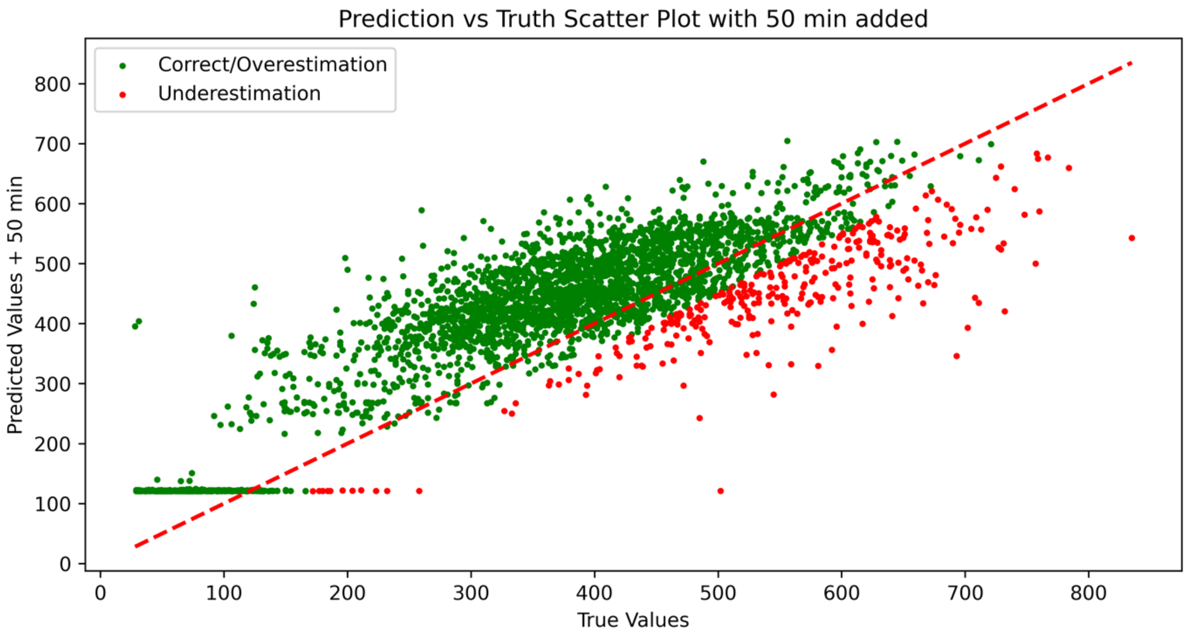

Figure 20: Adding minutes to the predictions results in conservative predictions.

Classes C and B are not used in this research for two reasons. The first reason is the relatively low predictability of these classes, specifically 0.602 and 0.780, respectively, which hinders accurate planning. The second reason pertains to the research objective of increasing bed occupancy, in which patients with longer stays do not contribute to this goal. Although the results and predictions of these patients may still be useful, they are not relevant for this study.

Patient scheduling relies on the output of the prediction models. Patients in class A are scheduled for 120 minutes as a security measure, irrespective of the output of the regression model. Class B patients are included only if their RF output falls within the Class B range. Additionally, a safety buffer of 50 minutes is added to prevent the occurrence of aforementioned issues. Patients whose regression and classification outputs do not align will be excluded from the scheduling method. The average occupancy of beds in 2022 was 1.47 patients per bed per day. The developed system operates dynamically, scheduling short-term patients in real time. Consequently, comparing the system's impact becomes challenging, as it relies on the specific circumstances of the current waiting list. Therefore, based on the current waiting list, all patients will be included in the system to assess its capacity to schedule patients beyond the average occupancy of 1.47.

4.3 Patient scheduling

To optimise patient scheduling based on length of stay information, an integer linear programming (ILP) model has been formulated. The ILP model efficiently assigns patients to available beds, using the predictions generated by the length-of-stay prediction models. This section will begin with the formulation of the ILP that contains the indices, parameters, decision variables, and the objective function. The formulation will be followed by a detailed description of the ILP.

Sets of indices:

- B: Number of beds available in the hospital.
- P: Set of patients to be scheduled.

Sets of parameters:

- L_i: The length of stay for patient i.
- O_i: opening time of the block.
- C: closing time of the block.

Decision variables:

$$x_{i,j} = \begin{cases} 1, if\ patient\ i\ is\ assigned\ to\ bed\ j \\ 0, otherwise \end{cases}$$

$$S_i = starting\ time\ of\ patient\ i$$

Objective Function:

$$\text{Maximize:} \quad \sum_{i \in P} \sum_{j \in B} x_{i,j}$$

Where $x_{i,j}$ is the binary decision variable for patient *i* on bed *j*.

Subject to:

$$\begin{aligned} &S_i - S_k - L_k + M(1 - y_{ik}) \geq 0 \\ &S_k - S_i - L_i + M(1 - y_{ik}) \geq 0 \\ &x_{i,j} + x_{k,j} - 1 \leq M(1 - y_{ik}) \end{aligned} \tag{1-3}$$

$$\textstyle\sum_{i:S_i \leq t < S_i + L_i} x_{i,j} \leq 1, \;\; \forall j \in B \tag{4}$$

$$S_1 \geq O \tag{5}$$

$$S_i + L_i \leq C, \;\; \forall i \in P \tag{6}$$

The ILP formulation starts with sets of indices. The first set described is set B, which contains the available beds in the hospital in the form of (bed 1, bed 2, bed 3, etc). The second set P, which all the patients accompanied by their length-of-stay prediction. The length of set P is equal to the length of the waiting list. The parameters exhibited by the ILP are L_i, and S_i.The first parameter is L_i where L is the predicted length-of-stay for patient i. The last parameter that can be set is C which is the closing time of the department. The decision variable $x_{i,j}$ equals 1 if patient i on bed j, 0 otherwise. The objective function $\sum x_{i,j}$,$\forall i, j$ aims to maximise the number of patients being scheduled within the set constraints. S_i is the starting time of the patient, which can be chosen within the openings hours.

There are six constraints to the ILP to ensure its feasibility. Constraint 1-3 ensures that only one patient is assigned to each bed at a given time, preventing double occupancy. There are three components. The equation $S_i - S_k - L_k + M(1 - y_{ik}) \geq 0$ is essentially saying: if patient i starts after patient k concludes, then y_{ik} should be 0 . If this isn't the case, the $M(1 - y_{ik})$ term will make the constraint always true, effectively making it non-binding. Similarly, $S_k - S_i - L_i + M(1 - y_{ik}) \geq 0$ establishes that if patient k begins after patient i finishes, then y_{ik} should also be 0. If not, the big-M term neutralizes the constraint. Lastly, the third equation $x_{i,j} + x_{k,j} - 1 \leq M(1 - y_{ik})$ serves as a gatekeeper. If y_{ik} is 1 (indicating overlap), then either $x_{i,j}$ or $x_{k,j}$ must be 0, preventing both patients from being scheduled on bed j at overlapping times. Constraint 4 monitors the active bed occupancy at any given time t. For every moment, it checks the aggregate of patients occupying a bed and mandates that this number must not transcend the total available beds in set B.

Furthermore, Constraints 5-6 ensure that the first patient's starting time is within the operating hours and that the last patient does not exceed the closing time. for each patient falls within hospital operating hours. When implementing the formulated ILP in all its aspects, an optimal patient scheduling solution can be obtained, taking into account time sequencing, operational constraints, and resource limitations. To illustrate the practical application of the approach, an example of the scheduling process will be presented in Section 4.4.

4.4 Patient Planning Tool

Integration of the prediction models, correction, and validation tools, along with the Patient scheduling ILP, is required to increase bed occupancy. Due to the lack of coding proficiency among planners, a custom tool has been developed to address this limitation. Known as the Patient Planning Tool, the tool provides a graphical user interface to meet the needs of hospital planners. In this section, the functionalities, workings, and usability will be elucidated.

As shown in Figure 5, the only input to the system is the current waiting list. The system automatically predicts the length of stay for all patients and applies the validation and correction methods described in Section 11.5. Sequentially, the system plans patients with short-term stay predictions consecutively. The output is a patient schedule that planners can use as a guide to see which patients can be planned on the same day.

In Figure 21, the workflow for using the Patient Planning Tool is illustrated. The initial step involves exporting the waiting list, which can be performed by the planners themselves through the HiX. Subsequently, the tool can be launched and, using the button 'Upload waiting list', the file can be selected and uploaded to the software. The planner is then required to provide the following information:

- Available beds: The number of beds available on the specific day for which the schedule is created.
- Opening time of the block: The initiation time, expressed as an integer, indicating the earliest possible scheduling time for the first patient.
- Closing time of block: The termination time, represented as an integer, indicating the latest time the last patient must have left the hospital.
- Choose the specialisation: The specialisation for which the planner is formulating the schedule. Each planner plans a distinct specialisation.

The subsequent step involves pressing the "Schedule Patients" button, triggering the automated patient planning process based on their respective length-of-stay predictions.

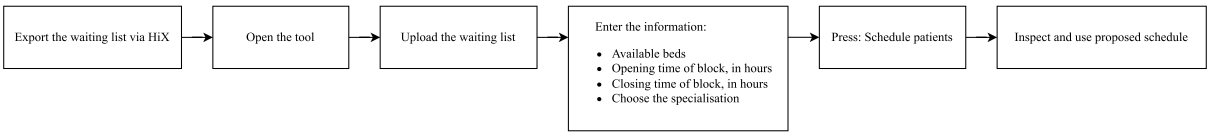

Figure 21. Visualisation of the workflow of the Patient Planning Tool.

The software produces a schedule showing patients arranged sequentially, considering opening and closing times. It is crucial to note that the schedule should be regarded as a guide, as additional constraints may need to be considered during the planning process. In Figure 22, the graphical user interface of the software is shown. In Section 4.5, a scenario is given based on the current waiting list to provide insight into achievable improvements in bed occupancy.

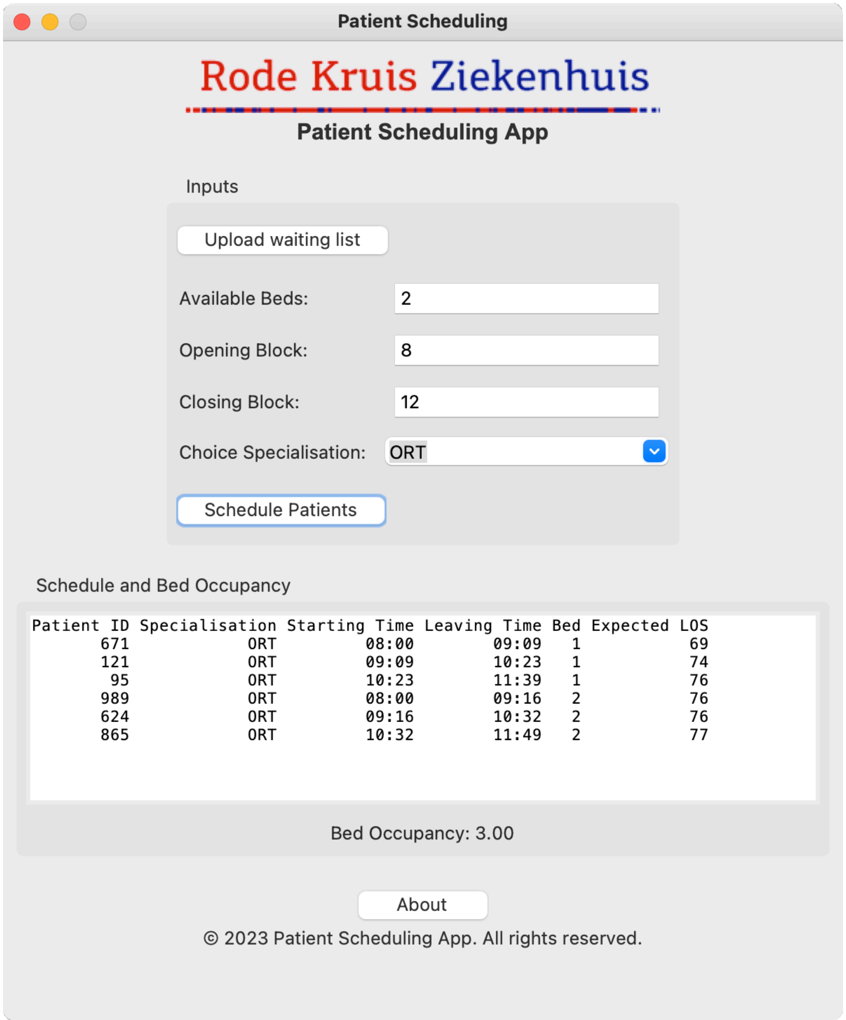

Figure 22. Graphical user interface of the Patient Planning Tool

4.5 Practical improvement

In the dynamic environment of the Red Cross Hospital, the constantly evolving waiting list poses challenges in accurately assessing the achievable increase in bed occupancy. In this section, the performance of the system will be evaluated using a simulated waiting list generated from historical data.

To create the synthetic waiting list, a random sample of patients will be selected from 2022. The fictional waiting list comprises 700 patients, reflecting the current size of the waiting list at the small regional. The distribution of the length of stay of the patients is depicted in Figure 23.

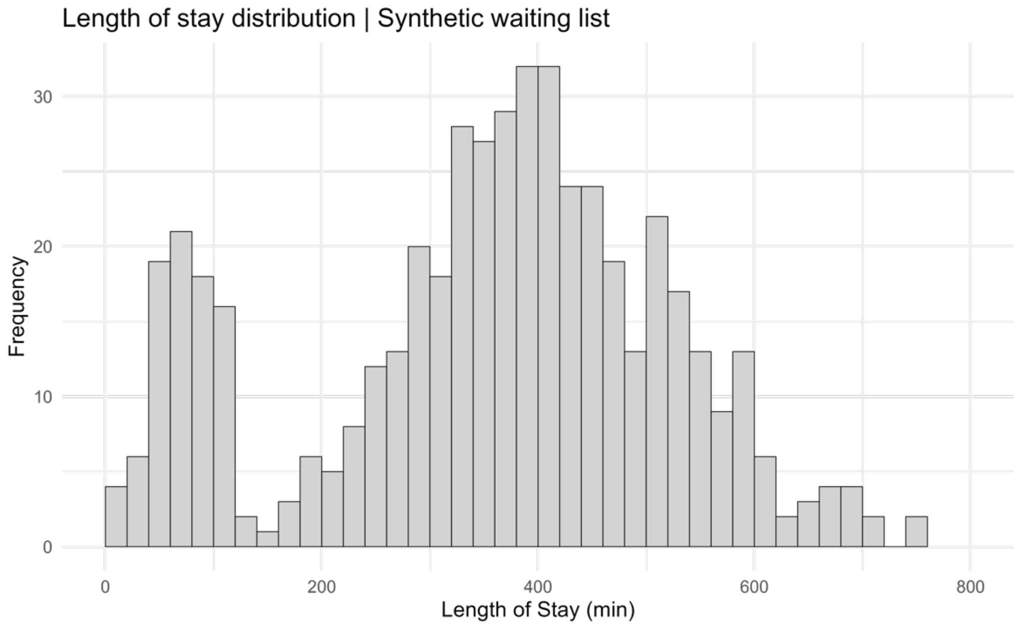

Figure 23. Length of stay distribution fictional waiting list

The length of stay distribution in the generated waiting list closely resembles the actual distribution observed in the small regional hospital, as shown in Figure 13. The data set provides a reliable representation of the waiting list at the hospital. Therefore, the data set will be used to accurately calculate the potential increase in bed occupancy.

To assess potential improvements, a one-day planning horizon will be utilised. To obtain a representative overview of the potential increase in bed occupancy per bed per day, the number of beds will be set at one. The department timings will be set from 08:00 to 21:00, aligning with the actual department opening hours. The dataset includes the following medical specialisations, which will be utilised for scheduling patients to evaluate the potential increase in bed occupancy: ORT, PLA, CHI, PYN, KNO, URO, CAR, DER. Due to validation and correction, 51% of patients are considered unplannable due to uncertainty in their prediction. By running the scheduling software, the schedules and bed occupancy per specialisation can be deducted. The statistical results are shown in Table 13. An example schedule is presented in Appendix E.

Table 13. Achievable bed occcupancy rates based on synthetic waiting list

Specialisation	**Bed occupancy**
ORT	3.0
PLA	3.0
CHI	3.0
PYN	26.0
KNO	3.0
URO	3.0
CAR	4.0
DER	4.0

The data presented in Table 13 indicates the following occupancy rates: ORT, PLA, and CHI all have an occupancy of 3.0. CAR and DER exhibit a bed occupancy of 4.0, while PYN demonstrates a significantly higher occupancy rate of 26.0. The high occupancy rate in the PYN department can be attributed to the short stay of patients with this specialisation. It is important to note that the current bed occupancy rate attained may not accurately reflect the potential achievable rates. Various practical constraints, such as personnel availability and medical resources, impose limitations on the maximum bed occupancy that can realistically be achieved.

The mean bed occupancy rate equals 6.125 and without the PYN specialisation, a mean bed occupancy of 3.333 is achieved. Another noteworthy point to mention is that on average there are still 109.62 minutes left before the department closes. This extra time allows for possible underestimations by the system or patients who may need to stay longer. Both bed occupancy rates exceed the average bed occupancy rate of 1.47 found in Section 4.1. The results clearly indicate that the implementation of this tool leads to a substantial increase in bed occupancy in the small regional hospital.

4.6 Conclusions

In this section, the findings of Chapter 4 will be concluded. Section 4.1 focusses on exploratory data analysis, which provided valuable information on the study population. Building on the knowledge from exploratory data analysis, Section 4.2 delves into the development, tuning, and evaluation of various machine learning models. These models were constructed with hyperparameter tuned and evaluated to ensure optimal performance. Subsequently, in Section 4.3 and Section 4.4 two scheduling ILPs were formulated to be utilised within the tool presented in Section 4.4. To evaluate the tool in a practical scenario, Section 4.5 demonstrated achievable bed occupancy rates. This tool serves as a valuable resource for planners, providing guidance to help increase bed occupancy. The sections are discussed in consecutive order.

The exploratory data analysis revealed significant patterns and correlations within the patient population. The correlation graph and the permutation graph provided valuable information on the correlated features. Remarkably, all available variables demonstrated varying degrees of influence on the length of stay and were therefore included in the models. The most significant correlating characteristic was, as expected, the medical procedure. The results align with the findings presented in the literature review conducted by Lu et al. (2015).

The model development, tuning, and evaluation process described in Section 4.2 yielded the best-performing models. The Random Forest classification model achieved an accuracy of 0.772 while the regression neural networks resulted in an R^2 score of 0.778. Visual representations of the performance of the model are illustrated in Figures 18 and 19. Given that these models will be implemented in a practical tool, it is crucial to establish validation and correction procedures. An innovative combination of regression and classification techniques has been implemented to provide reliability. The validated and corrected output is the input for the patient scheduling ILP.

The development of the planner tool focusses solely on the length-of-stay predictions for patients who successfully passed the validation check. Subsequently, the length-of-stay predictions for included patients were corrected by adding 55 minutes to achieve a correct/overestimation rate of 91.8%. The patient scheduling ILP, which is based on length-of-stay predictions, was incorporated into the tool. The tool was developed with a user-friendly graphical interface to facilitate its use by hospital planners.

The practical improvement section featured a scenario based on a synthetic waiting list. The results reveal a significant improvement in bed occupancy across all departments, underscoring the effectiveness and potential of the tool.

The results chapter presents crucial findings on the underlying models and planning systems. These findings have culminated in the development of software for hospital planners. The results chapter provided important results on all the underlying models and planning systems that ended up in the software for hospital planners. It is safe to say that a significant contribution can be made to increasing bed occupancy in the small hospital.

5 Discussion

This paper presents a comprehensive methodology to increase the occupancy of beds at the small regional hospital. The proposed approach uses length-of-stay predictions and integer linear programming techniques. The primary objective of this study is to develop a data-driven approach that can improve the work flow of hospital planners, leading to optimised resource allocation and improved patient care. Within this chapter, an in-depth discussion will be presented, focussing on the findings derived from the research. The strengths and limitations of the methodology will be examined and shed light on its effectiveness and potential drawbacks. Additionally, the impact of the findings and their implications for practical implementation will be worked out.

Furthermore, a discussion will be conducted to explore the strengths and limitations of the methodology. In addition, the underlying assumptions made during the study will be examined and the potential impact of the results will be evaluated. In the last section, future research directions aim to stimulate further scientific research and encourage collaboration among researchers to investigate the possibilities of the proposed methodology.

5.1 Results

The results can be divided into two sections, the length-of-stay predictions and the ILP model. First, the results of the length of stay will be discussed, followed by the results of the ILP model. The length of stay predictions is divided into two sections, regression and classifications.
In the case of regression, the neural network approach outperformed the other models, where in the case of classification, the Random Forest model outperformed.

5.1.1 Length of Stay Predictions

In the case of regression, a R^2 score of 0,776 was achieved. The score is compared to other studies found in the systematic review of the literature relatively low. The difference can be attributed to the fact that these studies are using a retroperspective approach, whereas this research is making prospective predictions. Retrospective predictions by definition outperform prospective predictions as there are valuable features available to make the predictions. Another noteworthy study is that by Siddiqa et al. (2022), who reported an R^2 score of 0.92. At first, the model seems to outperform the constructed model from this research. Unfortunately, the R^2 score in this study cannot be compared to the found R^2 score for several reasons. The first reason being that the study population of Siddiga et al. is primarily focused on heart transplant patients. A more specific study population narrows the variability within LOS. Another study that outperforms the results from this study is the of Zeng X (2022), reporting a R^2 of 0.96. The study conducted has the same generic population, but the higher R^2 can be explained by two key differences. The first difference is that the study used a data set containing information from more than two million patients, including many more features that are not available at the RCH. The second difference is that this study is also retrospective, including the total cost of the feature. By including these features, the model can be trained more accurately as total costs have a significant correlation with the length of stay of the patient. In addition to R^2, an MAE of 55.5 minutes was achieved.

The classification case approached the length of stay predictions by categorising the patients into different groups. These groups were 0-2 hours, 2-4 hours, 4-6 hours, and more than 6 hours. The best performing model is the random forest model with an accuracy of 77.20%. This accuracy is directly in line with the studies presented in the systematic review of the literature. The lowest reported accuracy was by the study of Arora et al. (2021) with an accuracy of

70.30%. They classified the patients into ten different groups, all representing one day. The longer the patient's stay, the lower the accuracy of the model. The same can be seen from the results of this study. The study by Karnuta et al. (2020) reported an accuracy of up to 91.80%, which seems significantly higher than the results from this study. The accuracy values of the study by Karnuta et al. (2020) cannot directly be compared with the results of this study. The reported 91.80% is the precision for a specific class of patients and should therefore be compared with the accuracy achieved in only one class of this research. In class 0-2 hours, this research achieved an accuracy of 92.86% and is therefore comparable. Overall, it can be stated that the accuracy of the classification model is in line with the results found in the systematic review of the literature. One difference is that the model of this research uses a prospective approach, whereas other studies use a retrospective approach, which makes the length-of-stay predictions considerably more straightforward. An important note for both models is that the models are tuned using a predetermined search space during the hyperparameter search. There might be a scenario where the optimal parameters of one of the models were located outside the search space.

In Section 4.1, the length-of-stay probability distribution is presented. At this point, patients are divided into four classes. These classes are selected in consultation with hospital planners. Narrower intervals in the classes enable more precise planning, but the certainty within the classes decreases. As shown in Figure 13, an alternative option is to create two classes that align with the two distinct peaks. Experimentation with two classes achieved a 96.01% accuracy, indicating the potential for further research to explore its impact on system performance.

5.1.2 Patient scheduling

While the ILP formulation offers numerous advantages in solving the complex optimisation problem, it is important to be aware of potential risks. The first risk is that the system cannot find an optimal solution. When there are not enough plannable patients on the waiting list, the software does not provide a schedule. On the other hand, when there are many plannable patients, for example, ten thousand, ILP cannot solve the problem within the given time. In that case, the software will not output a schedule or a sub-optimal schedule. At this point, the software is not capable of providing information to the planner about what is going wrong.

Furthermore, it is important to note that the patient scheduling ILP formulation may not account for all real-world constraints. For example, the algorithm does not consider scenarios where a physician is physically unable to treat multiple patients sequentially due to limitations.

An additional important consideration is the desired probability threshold from the classification model. Currently, a minimum certainty of 0.9 is set as a requirement for patients to be included in the scheduling tool. This conservative threshold is deliberately chosen to avoid possible logistical errors. However, a drawback of this option is that a substantial portion of patients, approximately 50% based on the current waiting list, are excluded from scheduling. During the implementation phase, it is important to fine-tune the certainty threshold of 0.9 to strike a balance between minimising exclusions and minimising the risk of logistical errors.

5.2 Validation and correction

Validation and correction methodologies were not found in the systematic review of the length of stay literature nor in the systematic review of the patient planning literature. In the systematic review of the planning literature, only probability distributions on the length of stay are used to plan patients. Using the probability distributions for length of stay, use statistics to account for patients who have an extended length of stay.

A statistical assessment on the probability of extended stay duration without the use of probability distributions is challenging. The reason is that every patient is unique and therefore has no probabilities available that the prediction of length of stay is correct. Patients who pass the two validation layers will be considered plannable patients and will be scheduled by the ILP. The fact that they pass the two layers of validation does not ensure a correct prediction of the length of stay. As shown in the results section, the average MAE is 55.5 minutes. 55.5 minutes MAE form a considerable problem in the planning of patients if the actual values of the regression model are used. Therefore, two methodologies were proposed to correct for this error. The first is a custom loss function as described in the methodology section. Forcing the model to penalise underestimations harder than overestimations, shifts towards outputs where underestimation is less common. Approaching the correction in this manner requires coding knowledge and is therefore not being used. The second methodology used a strategy in which additional minutes are added on top of the regression output that is used to plan patients. The first methodology is more accurate as it takes into account all features within the data set, but requires quite a lot of coding knowledge to change over time, which cannot be done by the RCH. The second methodology is less sophisticated but understandable to planners and, therefore, implemented in the system.

Validation and correction methods implemented within the system play a crucial role in minimising risks and ensuring that the length of stay remains within acceptable limits. By incorporating these methods, the system acts with an additional layer of protection against potential errors or inaccuracies that could lead to logistical issues. However, it is important to note that these validation and correction methods are based on certain assumptions and pre-defined parameters. Although they have been designed to be robust and effective, it is essential to acknowledge that they may not cover every possible scenario, and logistical issues can still occur. However, by incorporating these validation and correction methods, the system shows a proactive approach in minimising the risk of these logistical problems.

5.3 Usability

The developed system demonstrates usability in certain aspects, while also presenting challenges. This section discusses the usability of the system, addressing specific points related to length-of-stay predictions and patient scheduling.

The developed software is directly testable by planners. The software will directly influence the workflow of the planners and increase bed occupancy at the RCH. Predictions provide planners with information on the expected lengths of stay of patients, and the schedule is a guide to plan them. However, it should be noted that the planning tool can only be used as a guide. The reason for this is that the software is not validated at this point. More research is required to assess the reliability of the tool. In Chapter 6, a strategy for developing an implementation strategy is discussed to successfully implement the tool.

An important aspect to note is that the tool only takes into account the total length of stay of the patients and the availability in the department. However, the specific timings for the start of the operation are not set by the tool. Manually taking into account the exact time stamps of when a patient needs to be operated is required. For example, for a specific procedure, the patient is required to arrive one hour before the operation. The planner must allocate a dedicated slot in the operating room schedule for the patient undergoing the specific procedure. Planners are assumed to take into account this limitation.

5.4 Limitations

The findings presented in this research are subject to certain limitations that must be taken into account. These limitations include conservative adjustment, limited data availability, and retraining in case of changes. Understanding these limitations is crucial when interpreting the findings of this research. Each of these limitations will be addressed individually.

Patient planning mode validation and correction: The patient planning model implemented in this research is conservatively tuned. The constraints that the regression and classification results must align and that the classification model must output a probability greater than 0.9, induce results that are not optimal.

Limited data availability: The development of the models is based on data from approximately 17000 patients. The sample size may not provide sufficient information on infrequent procedures or unique patient cases. Acquiring a larger and more diverse dataset, possibly from other hospitals, would improve the generalisability of the findings.

Changes within the hospital: The implementation of length-of-stay prediction models in this research introduces potential retraining. With the addition of new doctors, existing models no longer produce predictions, as the model is trained only on the current situation. Similarly, changes in procedures or the participation of new anaesthetists require retraining and revaluation of the models to ensure their functioning. Additionally, new policies that have a significant impact on the length of stay require retraining of the models.

These limitations provide valuable information on areas that require further attention and refinement. Addressing these limitations would enhance the reliability, flexibility, and applicability of the system.

5.5 Future research

The research provided valuable information and has effectively demonstrated the efficacy of the methodology. However, several areas can be further investigated to ensure continuous progress in the field of hospital resource allocation and patient scheduling. The areas will be presented in this chapter as possibilities for advancement and refinement are explored. In the following sections, the points will be addressed in a systematic way.

Collaboration with multiple hospitals to enhance predictive capabilities: To further improve the predictive capabilities of the models, future research should consider collaborating with multiple hospitals. By including data from various hospitals, models can be trained on a more diverse and representative dataset. This collaboration would provide a broader understanding of patient characteristics, treatments, and results. In the end, this will improve the accuracy and generalisability of the predictions.

Multi-objective patient scheduling: Currently, integer linear programming (ILP) for patient scheduling is focused solely on planning patients with short-stay predictions to increase bed occupancy. However, to establish a more robust planning methodology, future research should aim to develop a multi-objective ILP. This advanced approach would seek to strike a balance between maximising the number of patients scheduled and optimising the use of the department. By considering patients with varying lengths of stay, this multi-objective approach ensures that all patients, regardless of their predicted stay duration, are included in the scheduling process.

Incorporating specific time stamps: At this point, the system only includes arrival times and departure times. An important area of exploration would involve incorporating precise time stamps for the start of the operation and the departure of the operating room. By accounting for these specific timings, the tool could further enhance its scheduling capabilities and improve the workflow of planners. With this version, planners are required to manually estimate the start time of each operation. Although this task is feasible, as it relies on predetermined preoperative times in most cases. By incorporating operation start times into the scheduling process, planners gain valuable insight into precisely reserving the appropriate slot for each patient. Integration of start times enables more accurate scheduling and improves overall efficiency of the system.

Real-world implementation and data acquisition: The complete methodology proposed in this research, combining various independent components, is relatively novel and has not been verified by the existing literature. Future research should focus on implementing the developed tool in real-world hospital environments or conducting extensive simulations. This practical implementation will facilitate the collection of new data, allowing for a comprehensive evaluation of the methodology's performance under real conditions. By acquiring real-world data, researchers can further refine the models and validate their effectiveness in optimising patient scheduling and resource allocation.

Comprehensive prediction of treatment phases: Although the current focus lies on predicting the length of stay, future research should explore the possibility of predicting all phases of patient treatment. This involves the prediction of the preoperative time, operation time, recovery time, and department time. When the prediction capabilities are extended to encompass all treatment phases, a more comprehensive and accurate schedule can be generated. The timings can be implemented within the patient scheduling ILP to exactly determine the time stamps of the patient. This holistic approach provides a complete overview of patient flow and allows more precise resource allocation and efficient scheduling.

Encouraging appropriate discharge practises: Currently, there is no incentive for patients to be discharged earlier, as beds are readily available. However, this lack of urgency can affect the accuracy of length-of-stay predictions. Future research should emphasise the importance of informing hospital personnel, particularly nurses, about the importance of discharge of patients as soon as they are physically capable. Encouraging appropriate discharge practises will contribute to more accurate data sets and improved predictions.

Simulations to evaluate the planning system: To comprehensively assess the effectiveness of the planning system, future research should conduct simulations based on the predictions of length of stay. These simulations would provide valuable information on the implications and effects of the proposed planning system in the real world. By simulating various scenarios and evaluating the results, researchers can fine-tune the system, identify potential bottlenecks, and optimise resource allocation strategies.

Incorporating additional features: Although current models use available features related to procedure and patient information, future research should explore the inclusion of additional relevant features. For example, incorporating drug usage data, comorbidity information, or sociodemographic factors could enhance predictive precision. The study by Lu et al. (2015) provided an overview of features that influence the length-of-stay. By capturing more of these characteristics, the length-of-stay prediction model would gain a broader perspective, resulting in more accurate predictions and better patient scheduling.

Excluding infrequent procedures for improved accuracy: The current system aims to predict the length of stay for each possible combination of patients and procedures. The reason for this is that planners can generally use the model. However, due to the infrequency of certain procedures, the accuracy of the predictions can be compromised. Future research should investigate which procedures can be excluded from the prediction and scheduling process without reducing its usability.

Dynamic corrections and adjusted loss functions: Currently, a fixed correction of 55 minutes is added to the length of stay predictions to coop with underestimations. However, future research should explore the possibility of implementing dynamic corrections, potentially tailored to specific procedures or specialisations. When considering the variability and specific characteristics of each procedure, more precise adjustments can be made. This can in the end result in improved prediction accuracy. Furthermore, researchers should further investigate the use of adjusted loss functions in machine learning models. By penalising underestimations more severely, the models can be fine-tuned to favour overestimations, enhancing the overall reliability of the predictions.

In conclusion, future research should focus on collaborative efforts with multiple hospitals to improve the predictive capabilities of the models. Implementing a multi-objective patient scheduling ILP, conducting real-world implementations or simulations, and including additional features. Conducting future research would lead to the development of a more comprehensive and accurate planning system.

5.6 Scientific and Practical Contributions

This section highlights the scientific and practical contributions of this research. The focus will be on the novel combination of patient scheduling and length-of-stay predictions and the integration of regression and classification models. For practical implementation, the focus will be on the dedicated tool for planners at the RCH.

This research yields two notable scientific contributions. First, it introduces a novel methodology that combines patient scheduling and length-of-stay predictions. Through an extensive literature review, no other articles were found to have employed such an approach. Second, this research integrates regression and classification models within the patient scheduling context. It emphasises the cruciality of accurately making length-of-stay predictions. The additional validation and correction layers were developed to test the proposed methodology in a real-world situation, which is not seen in the literature until now.

In terms of practical contributions, this research offers noteworthy insights. Few studies identified in the systematic literature review have actually implemented the findings derived from their research. Additionally, the dynamic approach of using patients on the waiting list, as demonstrated, is an innovative and practical application. The development of a dedicated tool, which can be used directly by OR planners to optimise bed occupancy, further exemplifies the contributions made by this research. The scientific and practical applications can be considered significant as they are novel and have not been seen in the literature.

6 Managerial recommendations

This article presents a comprehensive methodology to increase bed occupancy at the small regional hospital. The methodology combines length-of-stay predictions for outpatients with dynamic patient scheduling by a technique called integer linear programming. The study offers valuable information to optimise the workflow of operating room planners and increase bed occupancy at the RCH. In this chapter, the managerial recommendations are given. A short description of the challenges and the developed tool are provided, followed by an implementation strategy.

The hospital planners complete their task by carefully selecting patients from the waiting list and assigning them a designated day and bed for the next medical procedure. During the selection and allocation process, planners often face challenges in efficiently allocating resources due to limited knowledge about the expected length of stay for patients. As a result, planners tend to adopt a conservative approach to planning patients. This approach aims to prevent logistical issues that may arise when a consecutive patient arrives before the previous patient has left the bed.

To enhance planner workflow, a user-friendly tool has been developed to generate schedules with consecutively plannable patients. The intuitive tool simplifies the process by allowing the upload of the waiting list, followed by easily fillable parameters that the planning department can complete. The generated schedule is based on validated and corrected length-of-stay predictions for the patients. In the next section, the steps for integrating and implementing the tool to enhance the bed occupancy are described.

6.1 Implementation strategy

The success of implementing the tool is based on effective participation of stakeholders. Hospital management, planning department, physicians, and nursing staff need to participate to gain their support and participation in the implementation process. Detailed explanations must be provided to explain the functionalities, benefits and implementation process of the tool, highlighting how the tool aligns with the hospital's objective of increasing bed occupancy. In addition to stakeholder participation, the tool should be developed alongside its uses according to future research presented in Section 5.5.

Before full-scale implementation, a pilot phase will be conducted with a subset of patients and planners. The pilot phase will provide an opportunity to evaluate the tool's effectiveness, gather feedback from planners, and identify any necessary refinements. Planners' input regarding usability and impact on workflow will be collected and analysed to fine-tune the tool for optimal performance. Following the successful pilot phase, the tool will be deployed for full-scale implementation. Communication and training sessions will be conducted to inform all planners about the official adoption of the tool and its incorporation into their routine practises. Clear guidelines and support will be provided to ensure a smooth transition and maximise the tool's benefits.

To ensure a smooth transition, comprehensive training sessions must be conducted with the planning department. It is crucial that users understand the limitations of the system and get used to working with the tool. Planners will be familiarised with the features and functionalities of the tool and need to understand the output schedules in depth. Additionally, collaboration between the planning department and tool developers will be crucial to seamlessly integrate the tool into the existing workflow. New processes and guidelines must be defined to incorporate the tool effectively. The involvement of the planning department is required to refine their workflows and ensure that the tool becomes an integral part of their daily operations.

After implementation, the tool's performance must be monitored to gather more information on the performance and gather ongoing feedback from the planners. Regular assessments will be conducted to evaluate the impact of the tool on bed occupancy. Additionally, the impact on the workflow of the planners need to be monitored closely. Based on the findings, continuous improvement and refinements must be made to enhance the functionality and usability of the tool. At this point, patients are excluded from the scheduling as their prediction certainty is too low. As more data is collected, the minimum prediction certainty can slowly be lowered to allow more patients to be included in the scheduling. The lowering of the required certainty must be discussed between the planning department, management, and the developer. Periodic reports from planners and the developer need to be established to inform all stakeholders on the implementation of the tool. The reports will highlight the improvements in bed occupancy, efficiency, and flaws that can be reported to management. Ongoing technical support and assistant need to be available to the planning department to ensure a successful implementation.

Following the steps in this implementation plan, the small regional hospital can successfully integrate and utilize the developed tool. The implementation of the tool will lead to optimized patient scheduling and result in increased bed occupancy.

7 Conclusions

This research developed a complete system to help increase bed occupancy at the small regional hospital. Using a combination of state-of-the-art machine learning techniques and integer linear programming, patients are efficiently scheduled. In the current situation, hospital planners plan patients in a very conservative way. The reason for this is that it is unknown how long patients stay in the hospital. The system outputs a schedule that can be used as a guideline, with patients that can be be sequentially planned without exceeding the opening times of the department. Patient schedules are constructed with additional safety measures in place to avoid logistical problems. The safety measures can be slowly relaxed by planners as they observe that the system is working as intended.

This research methodology is based on two systematic reviews of the literature on patient planning and length-of-stay predictions. Unfortunately, there is no consensus on the ideal machine learning model for length-of-stay predictions. Consequently, various state-of-the-art machine learning models were developed and evaluated to determine the optimal performer using data from the small regional hospital. Ten different machine learning models were developed and hyperparameter tuned for both classification and regression tasks. The models were subsequently tested, and their evaluation metrics were analysed. Among the classification models, the Random Forest model achieved the highest accuracy of 77.20%, outperforming the other machine learning models. For the regression task, neural networks proved to be the best choice, achieving an R^2 score of 0.776.

The classification model serves as a validation for the regression model by providing probabilities that patients fall within specific time frames. In addition, a correction was implemented to prevent the model from underestimating the length of stay. With these safety precautions in place, the model only underestimated 9.6% of the predictions. It can be considered an acceptable margin, especially considering that the model overestimates half of the patients. In the end, the overestimations and underestimations will cancel each other out, and with the extra safety precautions in place, the probability on logistical issues is neglectable. The result of the software is a robust schedule that enables sequential patient scheduling, thus increasing bed occupancy at the small regional hospital.

Reflecting on the research question in this research, 'How can patient scheduling algorithms and optimisation techniques be integrated with length-of-stay predictions to effectively allocate resources and maximise bed occupancy?' It can be concluded that a combination of length-of-stay predictions and ILP patient scheduling can maximise bed occupancy. Equipping planners with length of stay predictions, certainty scores, and suggestions for patient schedules significantly improves their ability to plan multiple patients with limited resources. Although the average bed occupancy in the past year was 1.47, the practical use case showed bed occupancy rates of 3.0 and higher. Implementing the system will significantly increase bed occupancy in the small regional hospital by consecutively scheduling patients.

This article proposes an elaborate methodology to dynamically schedule patients from the waiting list, with the ultimate goal of increasing bed occupancy. A robust framework has been developed for direct use hospital planners, to increase bed occupancy and ultimately improving overall patient care.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Zheng, X. (n.d.). TensorFlow: Large-scale machine learning on heterogeneous systems.

Adan, I. J. B. F., & Vissers, J. M. H. (2002). Patient mix optimisation in hospital admission planning: a case study. International Journal of Operations & Production Management, 22(4), 445–461. doi:10.1108/01443570210420430

Alrefaei, M. H., & Diabat, A. (2015). Modelling and optimization of outpatient appointment scheduling. RAIRO Operations Research, 49(3), 435–450. doi:10.1051/ro/2014041

Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2016). Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 13(5), 971–989. doi:10.1109/TCBB.2015.2478454

Anvaryazdi, S. F., Venkatachalam, S., & Chinnam, R. B. (2020). Appointment scheduling at outpatient clinics using two-stage stochastic programming approach. IEEE Access: Practical Innovations, Open Solutions, 8, 175297–175305. doi:10.1109/access.2020.3025997

Arjannikov, T., & Tzanetakis, G. (2021, July 27). Cold-start hospital length of stay prediction using positive-unlabeled learning. 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). Presented at the 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Athens, Greece. doi:10.1109/bhi50953.2021.9508596

Arora, A., Demb, J., Cummins, D. D., Callahan, M., Clark, A. J., & Theologis, A. A. (2023). Predictive models to assess risk of extended length of stay in adults with spinal deformity and lumbar degenerative pathology: development and internal validation. The Spine Journal: Official Journal of the North American Spine Society, 23(3), 457–466. doi:10.1016/j.spinee.2022.10.009

Aslani, N., Kuzgunkaya, O., Vidyarthi, N., & Terekhov, D. (2021). A robust optimization model for tactical capacity planning in an outpatient setting. Health Care Management Science, 24(1), 26–40. doi:10.1007/s10729-020-09528-y

Auguie, B. (n.d.). gridExtra: Miscellaneous Functions for “Grid” Graphics.

Bacchi, S., Gluck, S., Tan, Y., Chim, I., Cheng, J., Gilbert, T., Menon, D. K., Jannes, J., Kleinig, T., & Koblar, S. (2020). Prediction of general medical admission length of stay with natural language processing and deep learning: a pilot study. Internal and Emergency Medicine, 15(6), 989–995. doi:10.1007/s11739-019-02265-3

Bacchi, S., Tan, Y., Oakden-Rayner, L., Jannes, J., Kleinig, T., & Koblar, S. (2022). Machine learning in the prediction of medical inpatient length of stay. Internal Medicine Journal, 52(2), 176–185. doi:10.1111/imj.14962

Bakker, M., & Tsui, K.-L. (2017). Dynamic resource allocation for efficient patient scheduling: A data-driven approach. Journal of Systems Science and Systems Engineering, 26(4), 448–462. doi:10.1007/s11518-017-5347-3

Banga, A., Mohanka, M., Mullins, J., Bollineni, S., Kaza, V., Ring, S., … Torres, F. (2017). Hospital length of stay after lung transplantation: Independent predictors and association with early and late survival. The Journal of Heart and Lung Transplantation: The Official Publication of the International Society for Heart Transplantation, 36(3), 289–296. doi:10.1016/j.healun.2016.07.020

Barsasella, D., Bah, K., Mishra, P., Uddin, M., Dhar, E., Suryani, D. L., … Syed-Abdul, S. (2022). A machine learning model to predict Length of Stay and mortality among diabetes and Hypertension inpatients. Medicina (Kaunas, Lithuania), 58(11), 1568. doi:10.3390/medicina58111568

Berrar, D. (2018). Cross-validation. 1, 542–545.

Biau, G., & Scornet, E. (2016). A random forest guided tour. TEST, 25(2), 197–227. doi:10.1007/s11749-016-0481-7

Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning (Vol. 4, Issue 4). Springer.

Bovim, T. R., Abdullahu, A., Andersson, H., & Gullhav, A. N. (2022). Integrated master surgery and outpatient clinic scheduling. Operations Research for Health Care, 35(100358), 100358. doi:10.1016/j.orhc.2022.100358

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. Nature Methods, 15(4), 233–234. doi:10.1038/nmeth.4642

Cai, X., Perez-Concha, O., Coiera, E., Martin-Sanchez, F., Day, R., Roffe, D., & Gallego, B. (2016). Real-time prediction of mortality, readmission, and length of stay using electronic health record data. Journal of the American Medical Informatics Association, 23(3), 553–561. doi:10.1093/jamia/ocv110

Cardoen, B., Demeulemeester, E., Beliën, J., Glouberman, S., & Mintzberg, H. (2001). Managing the Care of Health and the Cure of Disease—Part I: Differentiation. Health Care Management Review, 201(3), 921–932. doi:doi:10.1016/j.ejor.2009.04.011

Chen, Y. (2021). Prediction and Analysis of Length of Stay Based on Nonlinear Weighted XGBoost Algorithm in Hospital. Journal of Healthcare Engineering, 2021. doi:10.1155/2021/4714898

Cho, H., Kim, Y., Lee, E., Choi, D., Lee, Y., & Rhee, W. (2020). Basic Enhancement Strategies When Using Bayesian Optimization for Hyperparameter Tuning of Deep Neural Networks. IEEE Access, 8, 52588–52608. doi:10.1109/ACCESS.2020.2981072

Chollet, F., & others. (2015). Keras.

Combes, C., Kadri, F., & Chaabane, S. (2014). PREDICTING HOSPITAL LENGTH OF STAY USING REGRESSION MODELS: APPLICATION TO EMERGENCY DEPARTMENT.

Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics, 19(1), 270. doi:10.1186/s12859-018-2264-5

CPB. (2022). Zorguitgaven, ons een zorg? 2–3.

Daghistani, T. A., Elshawi, R., Sakr, S., Ahmed, A. M., Al-Thwayee, A., & Al-Mallah, M. H. (2019). Predictors of in-hospital length of stay among cardiac patients: A machine learning approach. International Journal of Cardiology, 288, 140–147. doi:10.1016/j.ijcard.2019.01.046

Dehghanimohammadabadi, M., Rezaeiahari, M., & Seif, J. (2023). Multi-Objective Patient Appointment Scheduling Framework (MO-PASS): a data-table input simulation–optimization approach. Simulation, 99(4), 363–383. doi:10.1177/00375497221132574

Delignette-Muller, M. L., & Dutang, C. (2015). {fitdistrplus}: An {R} Package for Fitting Distributions. Journal of Statistical Software, 64(4), 1–34. doi:10.18637/jss.v064.i04

Demeulemeester, E., Beliën, J., Cardoen, B., & Samudra, M. (2013). Operating Room Planning and Scheduling. In B. T. Denton (Ed.), Handbook of Healthcare Operations Management: Methods and Applications (pp. 121–152). Springer New York. doi:10.1007/978-1-4614-5885-2_5

Durstenfeld, M. S., Saybolt, M. D., Praestgaard, A., & Kimmel, S. E. (2016). Physician predictions of length of stay of patients admitted with heart failure. Journal of Hospital Medicine, 11(9), 642–645. doi:10.1002/jhm.2605

Emmert-Streib, F., & Dehmer, M. (2019). Evaluation of Regression Models: Model Assessment, Model Selection and Generalization Error. Machine Learning and Knowledge Extraction, 1(1), 521–551. doi:10.3390/make1010032

Fan, X., Tang, J., Yan, C., Guo, H., & Cao, Z. (2021). Outpatient appointment scheduling problem considering patient selection behavior: data modeling and simulation optimization. Journal of Combinatorial Optimization, 42(4), 677–699. doi:10.1007/s10878-019-00487-x

Fatima, M., & Pasha, M. (2017). Survey of Machine Learning Algorithms for Disease Diagnostic. Journal of Intelligent Learning Systems and Applications, 09(01), 1–16. doi:10.4236/jilsa.2017.91001

Feng, H., Jia, Y., Zhou, S., Chen, H., & Huang, T. (2023). A dataset of service time and related patient characteristics from an outpatient clinic. Data, 8(3), 47. doi:10.3390/data8030047

Fu, Y., & Banerjee, A. (2021). A stochastic programming model for service scheduling with uncertain demand: An application in open-access clinic scheduling. Operations Research Forum, 2(3). doi:10.1007/s43069-021-00089-6

Gao, Y., Zhang, Q., Lau, C. K., & Ram, B. (2022). Robust appointment scheduling in healthcare. Mathematics, 10(22), 4317. doi:10.3390/math10224317

Garg, L., McClean, S. I., Barton, M., Meenan, B. J., & Fullerton, K. (2012). Intelligent patient management and resource planning for complex, heterogeneous, and stochastic healthcare systems. IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans, 42(6), 1332–1345. doi:10.1109/TSMCA.2012.2210211

Grampurohit, S., & Sunkad, S. (2020). Hospital Length of Stay Prediction using Regression Models. 2020 IEEE International Conference for Innovation in Technology (INOCON), 1–5.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020).

Array programming with NumPy. In Nature (Vol. 585, Issue 7825, pp. 357–362). Springer Science and Business Media LLC. doi:10.1038/s41586-020-2649-2

Heim, N., Berger, M., Wiedemeyer, V., Reich, R., & Martini, M. (2019). A mathematical approach improves the predictability of length of hospitalization due to acute odontogenic infection: A retrospective investigation of 303 patients. Journal of Cranio-Maxillofacial Surgery, 47(2), 334–340. doi:10.1016/j.jcms.2018.12.002
Heim, N., Berger, M., Wiedemeyer, V., Reich, R., & Martini, M. (2019). A mathematical approach improves the predictability of length of hospitalization due to acute odontogenic infection: A retrospective investigation of 303 patients. Journal of Cranio-Maxillofacial Surgery, 47(2), 334–340. doi:10.1016/j.jcms.2018.12.002
Hills, R., & Kitchen, S. (2007). Satisfaction with outpatient physiotherapy: A survey comparing the views of patients with acute and chronic musculoskeletal conditions. In Physiotherapy Theory and Practice (Vol. 23, Issue 1, pp. 21–36). Informa UK Limited. doi:10.1080/09593980601147876

Hua, L., Dongmei, M., Xinyu, Y., Xinyue, Z., Shutong, W., Dongxuan, W., … Ying, W. (2023). Research on outpatient capacity planning combining lean thinking and integer linear programming. BMC Medical Informatics and Decision Making, 23(1), 32. doi:10.1186/s12911-023-02106-6

Huang, Y.-L., Hancock, W. M., & Herrin, G. D. (2012). An alternative outpatient scheduling system: Improving the outpatient experience. IIE Transactions on Healthcare Systems Engineering, 2(2), 97–111. doi:10.1080/19488300.2012.680003

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science and Engineering, 9(3), 90–95. doi:10.1109/MCSE.2007.55

Huntley, D. A., Cho, D. W., Christman, J., & Csernansky, J. G. (1998). Predicting length of stay in an acute psychiatric hospital. Psychiatric Services, 49(8), 1049–1053. doi:10.1176/PS.49.8.1049/

J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. The BMJ, 372. doi:10.1136/BMJ.N71

Jerbi, B., & Kamoun, H. (2011). Multiobjective study to implement outpatient appointment system at Hedi Chaker Hospital. Simulation Modelling Practice and Theory, 19(5), 1363–1370. doi:10.1016/j.simpat.2011.02.003

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science (New York, N.Y.), 349(6245), 255–260. doi:10.1126/science.aaa8415

Kadri, F., Dairi, A., Harrou, F., & Sun, Y. (2022). Towards accurate prediction of patient length of stay at emergency department: a GAN-driven deep learning framework. Journal of Ambient Intelligence and Humanized Computing, 1–15. doi:10.1007/s12652-022-03717-z

Karnuta, J. M., Churchill, J. L., Haeberle, H. S., Nwachukwu, B. U., Taylor, S. A., Ricchetti, E. T., & Ramkumar, P. N. (2020). The value of artificial neural networks for predicting length of stay, discharge disposition, and inpatient costs after anatomic and reverse shoulder arthroplasty. Et al [Journal of Shoulder and Elbow Surgery], 29(11), 2385–2394. doi:10.1016/j.jse.2020.04.009

Khawaled, Y., Alrefaei, M., Smadi, M., & Alawneh, A. (2022). Designing an outpatient-appointment scheduling using ahp and simulated annealing. International Journal of Applied Mathematics, 35(2). doi:10.12732/ijam.v35i2.6

Koo, A. B., Elsamadicy, A. A., Lin, I.-H., David, W. B., Sujijantarat, N., Santarosa, C., … Matouk, C. C. (2020). Predictors of extended length of stay following treatment of unruptured adult cerebral aneurysms: A study of the National Inpatient Sample. Journal of Stroke and Cerebrovascular Diseases: The Official Journal of National Stroke Association, 29(11), 105230. doi:10.1016/j.jstrokecerebrovasdis.2020.105230

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V, & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8–17.

Kuiper, A., Mandjes, M., de Mast, J., & Brokkelkamp, R. (2023). A flexible and optimal approach for appointment scheduling in healthcare. Decision Sciences, 54(1), 85–100. doi:10.1111/deci.12517

Kuwajima, H., Yasuoka, H., & Nakae, T. (2020). Engineering problems in machine learning systems. Machine Learning, 109(5), 1103–1126. doi:10.1007/s10994-020-05872-w

Launay, C. P., Rivière, H., Kabeshova, A., & Beauchet, O. (2015). Predicting prolonged length of hospital stay in older emergency department users: use of a novel analysis method, the Artificial Neural Network. European Journal of Internal Medicine, 26(7), 478–482. doi:10.1016/j.ejim.2015.06.002

Launay, C. P., Rivière, H., Kabeshova, A., & Beauchet, O. (2015). Predicting prolonged length of hospital stay in older emergency department users: use of a novel analysis method, the Artificial Neural Network. European Journal of Internal Medicine, 26(7), 478–482. doi:10.1016/j.ejim.2015.06.002

Lavesson, N., & Davidsson, P. (2006). Quantifying the Impact of Learning Algorithm Parameter Tuning.

Le, Q. V., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J., & Ng, A. Y. (2011). Building high-level features using large scale unsupervised learning. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 8595–8598. doi:10.48550/arxiv.1112.6209

Leeftink, A. G., Vliegen, I. M. H., & Hans, E. W. (2019). Stochastic integer programming for multi-disciplinary outpatient clinic planning. Health Care Management Science, 22(1), 53–67. doi:10.1007/s10729-017-9422-6

Lequertier, V., Wang, T., Fondrevelle, J., Augusto, V., & Duclos, A. (2021). Hospital Length of Stay Prediction Methods: A Systematic Review. Medical Care, 59(10), 929–938. doi:10.1097/MLR.0000000000001596

Lin, C. K. Y. (2015). An adaptive scheduling heuristic with memory for the block appointment system of an outpatient specialty clinic. International Journal of Production Research, 53(24), 7488–7516. doi:10.1080/00207543.2015.1084060

Liu, L., & Liu, X. (1998). Block appointment systems for outpatient clinics with multiple doctors. The Journal of the Operational Research Society, 49(12), 1254–1259. doi:10.1057/palgrave.jors.2600631

Lu, M., Sajobi, T., Lucyk, K., Lorenzetti, D., & Quan, H. (2015). Systematic review of risk adjustment models of hospital Length of Stay (LOS). Medical Care, 53(4), 355–365. doi:10.1097/MLR.0000000000000317

Luo, J., Kulkarni, V. G., & Ziya, S. (2012). Appointment scheduling under patient no-shows and service interruptions. Manufacturing & Service Operations Management: M & SOM, 14(4), 670–684. doi:10.1287/msom.1120.0394

M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. In International Journal of Data Mining & Knowledge Management Process (Vol. 5, Issue 2, pp. 01–11). Academy and Industry Research Collaboration Center (AIRCC). doi:10.5121/ijdkp.2015.5201

Ma, X., Si, Y., Wang, Z., & Wang, Y. (2020). Length of stay prediction for ICU patients using individualized single classification algorithm. Computer Methods and Programs in Biomedicine, 186(105224), 105224. doi:10.1016/j.cmpb.2019.105224

Mansoori, A., Zeinalnezhad, M., & Nazarimanesh, L. (2023). Optimization of Tree-Based Machine Learning Models to Predict the Length of Hospital Stay Using Genetic Algorithm. Journal of Healthcare Engineering, 2023.

Marjamaa, R., Vakkuri, A., & Kirvelä, O. (2008). Operating room management: why, how and by whom? Acta Anaesthesiologica Scandinavica, 52(5), 596–600. doi:10.1111/J.1399-6576.2008.01618.X

McKinney, W. (2010). {D}ata {S}tructures for {S}tatistical {C}omputing in {P}ython. In S. van der Walt & J. Millman (Eds.), {P}roceedings of the 9th {P}ython in {S}cience {C}onference (pp. 56–61). doi:10.25080/Majora-92bf1922-00a

Mekhaldi, R. N., Caulier, P., Chaabane, S., Chraibi, A., & Piechowiak, S. (2020). Using Machine Learning Models to Predict the Length of Stay in a Hospital Setting.

Mitchell, S., O'Sullivan, M., & Dunning, I. (2011). PuLP: A Linear Programming Toolkit for Python. The University of Auckland, Auckland, New Zealand, 65. http://www.optimization-online.org/DB_FILE/2011/09/3178.pdf

Mitchell, T. M. (2006). The discipline of machine learning (Vol. 9). Carnegie Mellon University, School of Computer Science, Machine Learning

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Plos Medicine. doi:10.1371/journal.pmed.1000097

Muhlestein, W. E., Akagi, D. S., Davies, J. M., & Chambless, L. B. (2019). Predicting inpatient length of stay after brain tumor surgery: Developing machine learning ensembles to improve predictive performance. Neurosurgery, 85(3), 384–393. doi:10.1093/neuros/nyy343

Nassar, A. P. J., & Caruso, P. (2016). ICU physicians are unable to accurately predict length of stay at admission: a prospective study. International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care, 28(1), 99–103. doi:10.1093/intqhc/mzv112

Navarro, S. M., Wang, E. Y., Haeberle, H. S., Mont, M. A., Krebs, V. E., Patterson, B. M., & Ramkumar, P. N. (2018). Machine Learning and Primary Total Knee Arthroplasty: Patient Forecasting for a Patient-Specific Payment Model. Journal of Arthroplasty, 33(12), 3617–3623. doi:10.1016/j.arth.2018.08.028

Nguyen, V. (2019). Bayesian Optimization for Accelerating Hyper-Parameter Tuning. 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 302–305. doi:10.1109/AIKE.2019.00060

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., & others. (2019). Keras {Tuner}.

Otten, M., Dijkstra, S., Leeftink, G., Kamphorst, B., Meierink, A. O., Heinen, A., … Boucherie, R. J. (2023). Outpatient clinic scheduling with limited waiting area capacity. The Journal of the Operational Research Society, 74(2), 540–561. doi:10.1080/01605682.2021.1978347

M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. The BMJ, 372. doi:10.1136/BMJ.N71

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in {P}ython. Journal of Machine Learning Research, 12, 2825–2830.

Probst, P., Bischl, B., & Boulesteix, A.-L. (2018). Tunability: Importance of Hyperparameters of Machine Learning Algorithms.

Raad, S.-E. (2021). Advies Aan de slag voor de zorg : een actieagenda voor de zorgarbeidsmarkt. www.ser.nl

Rahman, M. M., Kundu, D., Suha, S. A., Siddiqi, U. R., & Dey, S. K. (2022). Hospital patients' length of stay prediction: A federated learning approach. Journal of King Saud University - Computer and Information Sciences, 34(10), 7874–7884. doi:10.1016/j.jksuci.2022.07.006

Ramkumar, P. N., Navarro, S. M., Haeberle, H. S., Karnuta, J. M., Mont, M. A., Iannotti, J. P., Patterson, B. M., & Krebs, V. E. (2019). Development and Validation of a Machine Learning Algorithm After Primary Total Hip Arthroplasty: Applications to Length of Stay and Payment Models. Journal of Arthroplasty, 34(4), 632–637. doi:10.1016/j.arth.2018.12.030

Ramkumar, P. N., Navarro, S. M., Haeberle, H. S., Karnuta, J. M., Mont, M. A., Iannotti, J. P., Patterson, B. M., & Krebs, V. E. (2019). Development and Validation of a Machine Learning Algorithm After Primary Total Hip Arthroplasty: Applications to Length of Stay and Payment Models. Journal of Arthroplasty, 34(4), 632–637. doi:10.1016/j.arth.2018.12.030

Ray, S. (2019). A Quick Review of Machine Learning Algorithms. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 35–39. doi:10.1109/COMITCon.2019.8862451

Robb, W. B., O'Sullivan, M. J., Brannigan, A. E., & Bouchier-Hayes, D. J. (2004). Are elective surgical operations cancelled due to increasing medical admissions? Irish Journal of Medical Science, 173(3), 129–132. doi:10.1007/BF03167925

Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

Saleem, T. J., & Chishti, M. A. (2020). Exploring the Applications of Machine Learning in Healthcare. International Journal of Sensors, Wireless Communications and Control, 10(4), 458–472. doi:10.2174/2210327910666191220103417

Schäfer, F., Walther, M., Hübner, A., & Kuhn, H. (2019). Operational patient-bed assignment problem in large hospital settings including overflow and uncertainty management. Flexible Services and Manufacturing Journal, 31(4), 1012–1041. doi:10.1007/s10696-018-9331-0

Schmidt, R., Geisler, S., & Spreckelsen, C. (2013). Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources. BMC Medical Informatics and Decision Making, 13, 3. doi:10.1186/1472-6947-13-3

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2016). Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE, 104(1), 148–175. doi:10.1109/JPROC.2015.2494218

Shehadeh, K. S., Cohn, A. E. M., & Epelman, M. A. (2019). Analysis of models for the stochastic outpatient procedure scheduling problem. European Journal of Operational Research, 279(3), 721–731. doi:10.1016/j.ejor.2019.06.023

Siddiqa, A., Abbas Zilqurnain Naqvi, S., Ahsan, M., Ditta, A., Alquhayz, H., A. Khan, M., & Adnan Khan, M. (2022). Robust length of stay prediction model for indoor patients. Computers, Materials & Continua, 70(3), 5519–5536. doi:10.32604/cmc.2022.021666

Simon, R. (2007). Resampling Strategies for Model Assessment and Selection BT - Fundamentals of Data Mining in Genomics and Proteomics (W. Dubitzky, M. Granzow, & D. Berrar (eds.); pp. 173–186). Springer US. doi:10.1007/978-0-387-47509-7_8

Simon, R. (2007). Resampling Strategies for Model Assessment and Selection BT - Fundamentals of Data Mining in Genomics and Proteomics (W. Dubitzky, M. Granzow, & D. Berrar (eds.); pp. 173–186). Springer US. doi:10.1007/978-0-387-47509-7_8

Sociaal-economische Raad. (2020). Zorg voor de toekomst - Over de toekomstbestendigheid van de zorg.

Srinivas, S., & Khasawneh, M. T. (2017). Design and analysis of a hybrid appointment system for patient scheduling: an optimisation approach. International Journal of Operational Research, 29(3), 376. doi:10.1504/ijor.2017.084344

Srinivas, S., & Ravindran, A. R. (2020). Designing schedule configuration of a hybrid appointment system for a two-stage outpatient clinic with multiple servers. Health Care Management Science, 23(3), 360–386. doi:10.1007/s10729-019-09501-4

Testi, A., & Tànfani, E. (2009). Tactical and operational decisions for operating room planning: Efficiency and welfare implications. Health Care Management Science, 12(4), 363–373. doi:10.1007/s10729-008-9093-4

Tibby, S. M., Correa-West, J., Durward, A., Ferguson, L., & Murdoch, I. A. (2004). Adverse events in a paediatric intensive care unit: relationship to workload, skill mix and staff supervision. Intensive Care Medicine, 30(6), 1160–1166. doi:10.1007/s00134-004-2256-y

Tohidi, M., Kazemi Zanjani, M., & Contreras, I. (2020). A physician planning framework for polyclinics under uncertainty. Omega, (102275), 102275. doi:10.1016/j.omega.2020.102275

Vielma, J. P. (2015). Mixed Integer Linear Programming Formulation Techniques. In SIAM Review (Vol. 57, Issue 1, pp. 3–57). Society for Industrial & Applied Mathematics (SIAM). doi:10.1137/130915303

Vonk, R. A. A., Hilderink, H. B. M., Plasmans, M. H. D., Kommer, G. J., & Polder, J. J. (2020). Toekomstverkenning zorguitgaven 2015-2060.

Wang, J., & Fung, R. Y. K. (2015). Adaptive dynamic programming algorithms for sequential appointment scheduling with patient preferences. Artificial Intelligence in Medicine, 63(1), 33–40. doi:10.1016/j.artmed.2014.12.002

Wang, X., Liu, N., Truong, V.-A., & Anderson, B. R. (2019). Integrated Scheduling and Capacity Planning With Considerations for Patients' Length-of-Stays. Production and Operations Management. doi:10.1111/poms.13012

Weerts, H. J. P., Mueller, A. C., & Vanschoren, J. (2020). Importance of Tuning Hyperparameters of Machine Learning Algorithms. http://arxiv.org/abs/2007.07588

Weissman, J. S., Rothschild, J. M., Bendavid, E., Sprivulis, P., Cook, E. F., Evans, R. S., Kaganova, Y., Bender, M., David-Kasdan, J., Haug, P., Lloyd, J., Selbovitz, L. G., Murff, H. J., & Bates, D. W. (2007). Hospital workload and adverse events. Medical Care, 45(5), 448–455. doi:10.1097/01.mlr.0000257231.86368.09

Wickham, H. (n.d.). ggplot2: Elegant Graphics for Data Analysis. https://cran.r-project.org/package=ggplot2

Wing, J., & Vanberkel, P. (2022). Simulation optimisation for mixing scheduled and walk-in patients. Health Systems (Basingstoke, England), 11(4), 276–287. doi:10.1080/20476965.2021.1943010

Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record, 31(1), 76–77.

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing, 415, 295–316. doi:doi:10.1016/j.neucom.2020.07.061

Yeon, N., Lee, T., & Jang, H. (2010, December). Outpatients appointment scheduling with multi-doctor sharing resources. Proceedings of the 2010 Winter Simulation Conference. Presented at the 2010 Winter Simulation Conference - (WSC 2010), Baltimore, MD, USA. doi:10.1109/wsc.2010.5679023

Zeng, X. (2022). Length of stay prediction model of indoor patients based on light gradient boosting machine. Computational Intelligence and Neuroscience, 2022, 9517029. doi:10.1155/2022/9517029

Zhu, H., Chen, Y. (frank), Leung, E., & Liu, X. (2018). Outpatient appointment scheduling with unpunctual patients. International Journal of Production Research, 56(5), 1982–2002. doi:10.1080/00207543.2017.1355574

Zorguitgaven; kerncijfers. (2022, July 7). https://www.cbs.nl/nl-nl/cijfers/detail/84047NED

Appendix A

Table 1. Data extraction table Length of stay predictions

Reference	Country	Study Population	Prediction Models	Classification/ Regression	N	Length of Stay Outcomes	Missing Data	Statistical Methods	Clinical Utility	Predictor Variables	Model Evaluation	Validation Cohorts	Limitations	Best Model	Best Result	Optimization Methods	Features Selection
Arjannikov & Tzanetakis (2021)	USA	Non-specific	Positive Unlabed Learning	Classification	2343569	Long Stay (>4), Short Stay (<4)	Unspecified	Unspecified	Not discussed	Clinical Characteristics, Demographic	Accuracy, Confusion Matrix	CV=10	Unspecified	PL	ACC 0.73	Unspecified	Unspecified
Arora et al. (2021)	USA	Age ≥50 with Elective Lumbar or Thoracolumbar Instrumented Fusions	LR	Classification	8866	Extended LOS	Unspecified	CHI-square, Fishers	Not discussed	Comorbidities, Demographics, Operative Information	AUC, Sensitivity, Specificity	Hold-out, 80/20	Unspecified	LR	AUC=0.77, Sensitivity=77%, Specificity=0.68	Unspecified	Unspecified
Bacchi et al. (2020)	Australia	Non-specific	ANN, CNN, LR, RF	Classification	313	More than 2 Days, Fewer than 2 Days	Unspecified	Unspecified	Not discussed	Demographic data, Patient data Hospital data	AUC, F1 Score, MAE, MSE, NPV, PPV, Sensitivity, Specificity	CV=5, Hold-out	Single Study, Small Sample Size	NN	0.82	Grid Search	Unspecified
Banga et al. (2017)	USA	Lung Transplantation	None	None	12647	Prolonged Hospital LOS	Unspecified	Cox Proportional Hazards Analysis, Kaplan-Meier Curve, Multivariate Logistic for Features	Not discussed	Donor, Operative Variables, Recipient	None	Not specified	Retrospective	None	None	None	Logistic Regression
Barnes et al. (2016)	USA	Non-specific, Inpatient Medical Unit	RF, LR	Classification, Regression	8000	Regression for Minutes, Classification for Stay Categories	Unspecified	McNemar Test and Youden's Index	Demographic and Clinical Variables, Length of Stay, and more	Demographic data, Patient data Hospital data	Accuracy	Larger Dataset, Training on Own Data	Comparison of Predictions from Model to Clinicians	RF	Accuracy Measure (P > .10)	Unspecified	Unspecified
Barsasella et al. (2022)	Taiwan	Diabetes and Hypertension Inpatients	GBM, LR, SVM, XGBOOST	Classification, Regression	58618	LOS in Days, Mortality Probability	Removed Missing Data	Unspecified	Data-driven Decision-making,	Demographic data, Patient data Hospital data	AUC, AUPR, CV Score, MAE, Precision, Recall, RMSE, R2, Test Score	CV=10, Hold-out	Class Imbalance, Only One Data Source	RF, XGBOOST	Random Forest (RF): MAE 0.027, RMSE 0.401, R2 0.591	Hyperparameter Search	Unspecified
Cai X et al. (2015)	Australia	Non-specific	Bayesian Network	Classification, Regression	32634	Mortality, Readmission, and Length of Stay	In Hospital, at Home, Dead	Unspecified	Demographic Information, Patient History, Ward Type, and more	Accuracy, AUROC	ACC, AUROC	Hold-out	Bayesian Model	Bayesian Model	Accuracy: 0.8	Best-First Search	Unspecified
Chen (2021)	USA	Non-specific	XGBoost	Classification	114209	Days	Unspecified	Non-specified	Not discussed	Unclear	ACC, F1, Kappa, ROC, RMSE	CV=3, Hold-out, 70/30	Unspecified	XGBoost	ACC=0.822, F1=0.8501, Kappa=0.8122, RMSE=1.823	Grid-search	Unspecified
Heim et al. (2019)	Germany	Odontogenic Infections	Semiparametric Logistic Discrete Hazard Model	Regression	303	Days	Unspecified	Regression Analysis, Time-to-event Models	Promote Transparency Regarding Costs and Patients	Age, Antibiotics, Diabetis, Gender, Localisation, Spreading of Infection	Not mentioned	Not specified	Unspecified	Emiparametric Logistic Discrete Hazard Model	States Probability per Patient	Unspecified	Unspecified
Karnuta et al. (2020)	USA	Shoulder Replacement	ANN	Classification, Regression	111147	Cost, Days (High, Low, Medium), Discharge Disposition	Imputation	Non-specified	Assist Management Making Choices	Age, Arthoplasty Type, Diagnosis, Hospital, Income	AUC, ROC	Hold-out	Potential Bias Older Data, Single Database, Black Box	ANN	Accuracy 70.3-91.8%, AUC 0.72-0.89	Unspecified	Unspecified
Koo et al. (2019)	USA	Unruptured Adult Cerebral Aneurysms	Logistic Regression	Classification	46880	Fewer than 5 Days, More than 5 Days	Unspecified	Logistic Regression Analysis, Multivariate Logistic Regression	Not discussed	Choice of Procedure (Open Surgical vs. Endovascular), Demographics,	Not specified	Not specified	Not specified	Logistic Regression Analysis, Multivariate Logistic Regression	Importance of Variables	Unspecified	Unspecified
Ma et al. (2020)	Multiple	ICU Patients	iForest, JITL-ELM, K-means Clustering, One-class ELM, PCA	Classification	4000	More than 10 Days, Within 10 Days	Averaging Methods, Interpolation	Isolation Forest Algorithm, K-means Clustering, One-class ELM, PCA	Not discussed	Physiological Indicators of ICU Patient	AUC, ACC, G-Mean, Lift Value, Precision, Sensitivity, Specificity, Miss Rate	Not specified	Not specified	One-class JITL-ELM	AUC = 0.8510, Accuracy= 0.82, G-	Unspecified	Unspecified
Muhlestein et al. (2019)	USA	Craniotomy for Brain Tumor	Linear Classifiers, Naive Bayes, NN, RuleFit,	Regression	41222	Days	Imputation, Missing Categorical as 'Missing'	Mann-Whitney U Test, Partial Dependence Plots, Permutation Importance,	Describes Potential Use Cases	Patient data, Hospital data	RMSLE	CV=5, Hold-out	Validation Due to ML, Missing Important Predictors	SVM, Two Gradient Boosted, Combined Using a Elastic Net to Create Ensemble Model	RMSLE 0.555	Hyperparameter Search	Unspecified
Navarro et al. (2018)	USA	Primary Total Knee Arthroplasty	Naive Bayesian Model	Classification, Regression	141446	Classification Costs, Regression LOS	Unspecified	Unspecified	Not discussed	Age, Comorbidity Scores ("Risk of Illness" and "Risk of Morbidity"), Gender, Race	ACC, AUC	Hold-out	Not specified	Naive Bayesian Model	AUC 0.7822 for LOS, 0.7382 for Inpatient Costs	Unspecified	Unspecified
Rahman et al. (2022)	USA	Non-specific	Lasso Regression, Linear Regression, Ridge Regression	Regression	92753	Days	Dropped Missing	Data Visualisation, Correlation Analysis	Not discussed	Patient data, Hospital data	MAE, MSE, R2	Hold-out, 70/30	Not specified	Linear	MAE 1.389, MSE 2.0320, R2 0.873	Unspecified	All Available
Siddiqa et al. (2022)	USA	Non-specific	DT, LR, MLR, RF, RR, XGBOOST	Regression	2,3E+07	Days	Dropped Missing	Bivariate Analysis, Mutual Information Regression, Univariate Analysis	Manage Hospital Resources	Patient data, Hospital data	MSE, R2	CV=10, Hold-out, 80/20	Unspecified	RF	MSE 5, R2 0.92	Unspecified	Importance Scores
Zeng (2022)	USA	Non-specific	LR, lGB, RF, RR, XGBOOST	Regression	2343569	Days	Dropped Missing	Correlation Analysis, Data Visualisation	Not discussed	Not explicitly mentioned	MSE, R2	CV=10, Hold-out, 99/1	Unspecified	LightGBM	MSE 2.231, R2 0.960	Unspecified	Correlation Analysis

Search query Length of Stay predictions

("length of stay" OR "hospital stay" OR "duration of stay" OR "inpatient stay" OR “Duration”) AND (("prediction" OR "predictive modeling" OR "forecasting" OR "prognostics") OR ("machine learning" OR "statistical models" OR "AI" OR “artificial intelligence” OR "predictive analytics")) AND ("hospital patients" OR "medical center patients" OR "inpatient population" OR "admitted individuals")

Appendix B

Table 2. Data extraction table Length of stay predictions

Reference	Country	Main Objective	Study population	Input Variables	Data	Intervention/ Approach	Findings	Implementation	Discussion/Implications	Software Used	Conclusion	Validation	Future directions
Hua et al. (2023)	China	Optimize outpatient service scheduling	Outpatient clinic in China	Scheduling variables, doctors, patients, appointments, services, working time	Hospital data	(M)ILP	Increased work efficiency and patient flow with outpatient service planning	Unspecified implementation	Unspecified implications	CPLEX	The proposed method of outpatient service scheduling effectively reduces waiting time, improves work efficiency, and enhances the overall quality of medical service.	Unspecified	Unspecified
Dehghani mohammadabadi et al. (2023)	Unspecified	Achieve efficient schedule balancing	Breast cancer patients	Processing time, no-shows, unpunctuality, emergencies	Cancer clinic	MOPSO, MO-PASS	Improved objective functions related to system throughput and avoiding overtime	Unspecified implementation	Future research required	Matlab	The MO-PASS framework is a practical and easy-to-implement solution that bridges the gap between existing models and algorithms in appointment scheduling, offering a viable solution for improving the level of service and considering patient flow and physician availability.	Simulation	Incorporating additional factors, specific assignment rules, multi-stage problems, evolutionary algorithms
Feng et al. (2023)	China	Achieve efficient schedule balancing	Non-specific	Consultation times, date, holiday indicator, gender, cancer indicator, distance	Data from Hangu clinic, China	Data analysis, regression	Showed relations between indicators and visit count	Unspecified implementation	Operational challenges, missing values, estimated service time	Python	The dataset reveals various patterns and characteristics related to the service time of outpatient consultations, highlighting the potential influence of variables such as previous visits, medical conditions, gender, appointment time, and patient address, providing valuable insights for future research in outpatient appointment scheduling.	Unspecified	More research on correlations, appointment scheduling
Kuiper et al. (2023)	The Netherlands	Design efficient appointment scheduling	Non-specific	Patients, service times, variation, no-shows, walk-ins	Unspecified	Mathematical modelling, ILP	Outperforms competing approaches, offers a general and easy-to-use solution	Webtool that produces optimal schedules instantaneously	Gap between theory and practice, optimization benefits, need for further research	Unspecified software	The proposed approach offers a general, easy-to-use, and superior solution for appointment scheduling in healthcare settings.	Simulation	Incorporating additional phenomena, heterogeneous patient populations, multi-stage processes
Otten et al. (2023)	The Netherlands	Maximize in-person consultations	Non-specific	Demand, capacity, trajectory, waiting area	Hospitals	(M)ILP	Improved outpatient scheduling with constraints	Tried with several clinics, output is a schedule	Lack of validation outcomes, deterministic early arrival times	Unspecified software	The cooperative simulation optimization approach enables outpatient clinics to effectively manage waiting room occupancy and deliver required appointments during the COVID-19 pandemic, considering capacity restrictions and optimizing the appointment mix.	Simulation	Including accompanying persons, evaluating interventions, actual scheduled patients, short-term control
Bovim et al. (2022)	Germany	Maximize resource utilization, minimize waiting times	OC consultations and surgeries	Capacity, patient flow, surgeon constraints	Orthopaedic Department at St. Olav's Hospital	(M)ILP	Improved coordination, resource utilization, patient throughput, decreased queues	Unspecified implementation	Importance of coordination, scheduling policies, flexibility, and patient calling lists	IVE Xpress 8.6, Simpy	The developed model provides blueprints for effective scheduling.	Simulation	Incorporating uncertainty, exploring mechanisms, variations in demand
Gao et al. (2022)	Hong-kong	Minimize costs, optimize appointment scheduling	Non-specific	Fixed cost, variable cost, capacity, service duration	Unspecified	(M)ILP	Appointment scheduling system	Unspecified implementation	Cost structure, trade-offs, effectiveness of proposed model	Matlab	The developed tool aids in decision making.	Simulation	Not discussed
Wing & Vanberkel (2022)	Canada	Balance waiting times and physician overtime	Emergency center	Priority, length, arrival rates	Nova Scotia Health Authority	Simulation	Tool to help planners	A tool to help planners	Optimization framework, challenges, uncertainties	Python	General rules of thumb are presented for scheduling.	Simulation	Effects on patient behavior
Khawaled et al.(2022)	Jordan	Minimize waiting times, maximize utilization	Outpatients	Priority, number of patients	Unspecified	Simulated annealing	Efficient appointing system with AHP-SA algorithm	Tool for decision makers	Unspecified discussion/implications	Unspecified software	The developed model assists with scheduling.	Simulation	Not discussed
Fan et al. (2021)	China	Optimize outpatient clinic operations	Dalian City Dermatology Hospital	Arrival interval, patient flow, waiting patience, service times	Dalian City Hospital	Simulation, MOCBA, GA	Improved efficiency and performance of outpatient services	Unspecified implementation	Unspecified discussion/implications	Combination	Optimizing outpatient clinic operations by considering patient preferences and implementing joint scheduling schemes significantly improves service efficiency and overall system performance.	Validated with real data, experiments	Including behavioral patterns of doctors
Fu & Banerjee (2021)	USA	Optimize same-day assignment in a clinic	Non-specific	Same-day requests, patient types, block length, throughput, arrival and service times	Unspecified	SIP	Improved cost and efficiency in optimizing same-day requests assignment	Can be implemented, uncertain how	Comparison, method application, sensitivity analysis, future research	CPLEX	The output of the SIP model provides sufficient information for the clinic manager to arrange appointments based on optimal values, accommodating patient requests and allowing earlier request senders more choices.	Sensitivity analysis	Not discussed
Tohidi et al. (2021)	Canada	Plan physicians in polyclinics under uncertainty	Physicians	Work schedules, arrival process, capacity, treatment times	Hospital	(M)ILP	Optimal solutions provided by robust problem solving algorithm	Discussed as a possibility	Effectiveness of proposed framework, impact of uncertainty, trade-offs, additional constraints	Unspecified software	The proposed physician scheduling framework, which incorporates uncertainty and corrective actions, results in lower costs and improved efficiency compared to a deterministic approach.	Simulation	Incorporating additional constraints and preferences, patient waiting time, other clinic dynamics
Aslani et al. (2021)	Canada	Develop robust capacity planning model	Unspecified	Uncertain demand, budget, uncertainty set	Seng Hospital Singapore	(M)ILP	Robust capacity planning model for outpatient clinics, addressing uncertainty	Unspecified implementation	Identification of critical time periods	Unspecified software	The proposed robust model provides a feasible capacity plan while minimizing costs and accommodating uncertainty in demand, and future research directions include multi-objective models and application in outpatient settings with multiple appointment types.	Simulation	Extending the model, multi-objective optimization, different clinic goals
Srinivas & Ravindran (2020)	USA	Balance performance measures, minimize waiting times	Family medicine clinic patients	No-show rate, service time variation, cost ratios, scenarios	Family medicine clinic in Pennsylvania	(M)ILP	Schedule configuration minimizing cost, balancing capacity and demand	Unspecified implementation	Impact of factors, importance of patient flow stages, limitations and future research	Combination	The proposed approach and insights drawn from the analysis provide valuable guidance for healthcare practitioners in designing effective appointment systems and minimizing costs in a clinic setting.	Sensitivity analysis	Incorporating patient availability and preferences, impact of walk-ins, clinics with three or more stages
Anvaryazdi et al. (2020)	USA	Minimize wait time, maintain efficient patient flow	Women's OBGYN clinics	Patient categories, providers, slots, new categories, service duration, targets, penalties, capacity restrictions	Unspecified	SIP	Reduced patients' indirect wait time, improved appointment scheduling efficiency	Clinic managers use the tool	Effectiveness of proposed model, integration of scheduling template, comparison, future research	CPLEX	The proposed two-stage stochastic programming and simulation models, along with the scheduling template, can improve appointment scheduling efficiency and reduce patient appointment delays in outpatient clinics.	Simulation	Not discussed
Shehadeh et al. (2019)	USA	Develop stochastic scheduling model for outpatient procedures	Non-specific	Waiting time, idle time, overtime cost	Unspecified	(M)ILP	Outperforms existing models for Stochastic Outpatient Procedure Scheduling	Unspecified implementation	Comparison, performance analysis, implementability, advantages	CPLEX	The proposed model demonstrates improved performance and implementability compared to existing models for the Stochastic Outpatient Procedure Scheduling Problem.	Comparing models	Including uncertainty, trade-offs between metrics, dynamic scheduling

Schäfer et al. (2019)	Germany	Optimize patient-bed assignments	Cardiology and gastroenterology patients	Time stamps, department, age, gender, care level	German hospital	Greedy heuristic	Improved patient-bed allocations, reduced overflow, optimized objectives	Applicable to large hospitals worldwide, not limited to German setting	Trade-offs, sensitivity analyses for objectives	Gurobi	The proposed decision model for hospital bed allocation, considering multiple stakeholders and addressing aspects not covered in previous research, outperforms other methods and improves stakeholder objectives.	Sensitivity analysis	Other heuristics, uncertainty modeling, additional stakeholders
Leeftink et al. (2019)	The Netherlands	Optimize scheduling for multi-disciplinary cancer clinics	Cancer patients	Population distribution, referral probabilities, performance measures, clinic capacity	UMCU	SAA	Improved efficiency and performance of multi-disciplinary clinics	Implemented in real-life situations in HPB clinic of UMCU	Impact of weight settings, trade-off, need for dynamic scheduling, incorporation of variability	Combination	The integrated optimization approach can help hospitals efficiently organize multi-disciplinary care systems and improve clinic performance.	Validated with real data, experiments	Additional sources of variability, impact of priority rules
Zhu et al. (2018)	Canada	Evaluate appointment scheduling systems	Non-specific	Arrival distributions, no-shows, punctuality	Unspecified	Heuristic	Heuristic policy outperforms currently adopted policy in reducing waiting costs	Unspecified implementation	Impact of patient unpunctuality on scheduling systems	Unspecified software	Implementing a heuristic policy that smoothes patient arrival flows and reduces appointment intervals can significantly improve efficiency and reduce waiting costs in appointment scheduling systems in healthcare settings.	Simulation	Incorporating additional factors, effective operation rules
Bakker & Tsui (2017)	The Netherlands	Develop resource allocation model for patient scheduling	Hospital patients	Resource allocation, specialist availability, durations, arrival patterns	Unspecified	Simulation	Improved service level, reduced wait times, higher resource utilization	Unspecified implementation	Generalizability challenges, customization, staff support	Unspecified software	The dynamic data-driven approach to specialist allocation shows promise in improving patient appointment scheduling and resource utilization, but further customization and validation are required for different hospital settings.	Simulation	More sophisticated models, synergies between surgery and appointment scheduling
Lin et al. (2017)	Hong-kong	Analyze integrated resource allocation and scheduling	Ophthalmology clinic	Appointments, patient classes, visitors, punctuality, procedure durations, schedule, parameters	Unspecified	Heuristic, Simulation	Improved system performance in terms of patient waiting time, resource overtime, congestion	Unspecified implementation	Impact of resource flexibility, choice of objectives and weights, applications and limitations	Microsoft Visual Basic	Integrated strategies can effectively improve system performance in healthcare clinics by optimizing resource allocation and appointment scheduling.	Unspecified	Exploring application in other clinics or environments
Srinivas & Khasawneh (2017)	USA	Propose hybrid appointment system for scheduling	Non-specific	OA ratio, no-show rate, CV of service time, patient calls	Unspecified	(M)ILP	Effective handling of system variations without impacting rejection rate and overtime rate	Unspecified implementation	Impact of system parameters, flexibility, need for further research	Unspecified software	The Hybrid Appointment System (HAS) can handle variations in system parameters and provide a balanced schedule that minimizes total loss for a clinic.	Sensitivity analysis	Arrival of walk-in patients, multi-objective optimization, computer simulation model
Lin (2015)	Hong-kong	Improve performance with adaptive appointment scheduling	Outpatient clinics in public hospitals	Patient class, waiting time, quota, staffing, distribution	Unspecified	Heuristic, MIP	Adaptive heuristic algorithm outperforms other methods in appointment scheduling	Unspecified implementation	Importance of patient class and waiting time information, balance between objectives	Combination	An adaptive heuristic approach effectively improves service performance in specialist outpatient clinics by reducing waiting times and congestion compared to traditional scheduling rules and mathematical formulations.	Sensitivity analysis	Appointment booking decisions, adaptive heuristic algorithm
Alrefaei & Diabat (2015)	Jordan	Optimal appointment system for outpatient department	Non-specific	Clinic hours, doctor hours, resources, patients, waiting time, doctor utilization	Unspecified	Simulation	Identification of appointment systems with good performance across multiple objectives	Unspecified implementation	Classification of appointment systems, ranking, application to outpatient scheduling	Arena	A systematic framework for selecting an appointment system that balances multiple objectives in an outpatient department clinic is presented.	Unspecified	Unspecified
Wang & Fung (2015)	Hong-kong	Develop adaptive algorithms for outpatient appointment scheduling	Non-specific	Patient preferences, revenue, mismatch, offer acceptance	Unspecified	MDP	Adaptive algorithms improve appointment systems, consider patient preferences and revenue	Unspecified implementation	Initialization, exploration vs. exploitation trade-off, effects of preferences, advantages	Matlab	The proposed adaptive algorithms provide effective approaches for sequential appointment scheduling, considering patient preferences and maximizing expected revenue.	Validated with real data, experiments	Exploring dependencies, optimizing revenue, exploration probabilities, enhancing algorithms
Luo et al. (2012)	USA	Develop framework for scheduling models considering interruptions	Non-specific	Interruption rate, patient arrival, service times, waiting costs, appointments	Unspecified	Mathematical modelling, ILP	Considering interruptions in appointment scheduling improves effectiveness	Unspecified implementation	Impact of interruptions, performance differences, benefits of flexibility	Unspecified software	Understanding interruptions in appointment scheduling is important for improving performance, and the developed framework provides insights into effective scheduling policies.	Unspecified	Analyzing interruptions, appointment scheduling policies
Y.-L. Huang et al. (2012)	USA	Reduce wait times and improve patient flow	Non-specific	Treatment time, patient arrival, no-shows, lateness, conflicts, overwriting	Unspecified	Simulation	Reduced patient wait time without significantly increasing physician idle time	Implemented in three clinics	Impact of treatment time estimates, importance of wait ratio, need for scheduling templates	Unspecified software	The patient scheduling approach effectively reduces patient wait time and improves patient flow without significantly increasing physician idle time, and it can be implemented successfully in clinics without additional workload on medical staff.	Simulation	Addressing patient no-shows, exploring cost-effectiveness, different specialties, ancillary services
Jerbi & Kamoun (2011)	Tunisia	Optimize appointment scheduling for outpatient department	Nephrology outpatient department	Scheduling rules, appointment rules, waiting times, doctor utilization	Appointment systems, rules, no-shows, walk-ins	Simulation, ILP	Optimal appointment schedule with specific scheduling and appointment rules	Unspecified implementation	Multi-objective approach, decision maker's preferences, potential extension	Arena	Management preferences through linear satisfaction functions resulted in the selection of an optimized appointment schedule that balances resource utilization and waiting times.	Simulation	More research resources in outpatient departments
Yean et al. (2010)	South-korea	Examine appointment scheduling for outpatient units	Department of ophthalmology	Service time distribution, walk-in patients, punctuality, no-show rate, ratios, resource capacity	EMR system, observations, interviews with doctors and nurses	Mathematical model	Appointment scheduling should be approached as a system problem	Unspecified implementation	Interdependency between patient flows, study limitations, practical considerations	Arena	Appointment scheduling for outpatient units with multiple doctors and shared resources should be derived as a system problem, considering the interdependency among patient flows, rather than relying on individually favored rules for each doctor.	Unspecified	Not discussed
Liu & Liu (1998)	Hong-kong	Implement block appointment system	Outpatient clinics involving multiple doctors	Number of doctors, service times, arrival times, no-shows	Consultation times, doctors' arrival pattern	Simulation	Identification of properties shared by best appointment schedules, development of simulation search scheme and suboptimal appointment rule	Unspecified implementation	Efficient frontier, myotic scheduling rule, application in public clinics	Unspecified software	The simulation-based appointment system, along with the myotic scheduling rule, can offer effective solutions for clinic operations, with further potential for improvement and application.	Unspecified	Impact of doctors' arrival patterns, making the system simpler

Search query patient scheduling

("Integer Linear Programming" OR "optimization" OR "Mathematical Optimization" OR "Integer Programming" OR "Integer Optimization" OR "ILP" OR "Optimization Models" OR "Optimization Techniques" OR "Mathematical Models" OR "Operations Research" OR "Combinatorial Optimization" OR "Decision Optimization" OR "Heuristic Optimization" OR "Metaheuristic Optimization" OR "Constraint Programming") AND
("Patient Scheduling" OR "Appointment Scheduling" OR "Planning" OR "Scheduling Optimization" OR "Scheduling" OR "Patient Planning" OR "Patient Appointment" OR "Patient Booking" OR "Scheduling Efficiency" OR "Scheduling Algorithms" OR "Scheduling Models" OR "Scheduling Systems" OR "Scheduling Strategies" OR "Scheduling Policies") AND
("Daycare Department" OR "Daycare" OR "Day Care" OR "Day Treatment" OR "Day-care" OR "Outpatient Department" OR "Ambulatory Care Center" OR "Outpatient Care" OR "Outpatient Clinic" OR "Clinic" OR "Outpatient Treatment" OR "Same-Day Care" OR "Same-Day Treatment" OR "One-Day Care" OR "One-Day Treatment" OR "Single-Day Care" OR "Single-Day Treatment" OR "Short Stay" OR "Minor Procedures" OR "Non-Admitted Patients" OR "Brief Intervention" OR "Transitional Care" OR "Ambulatory Services" OR "Same-Day Surgery" OR "Walk-In Clinic")

Appendix C

Model	Hyperparameter	Type	Range
Gradient Boosting	n_estimators	Integer	[100, 500]
	learning_rate	Real	[0.01, 0.3] (log-uniform)
	max_depth	Integer	[3, 10]
	min_samples_split	Integer	[2, 10]
	min_samples_leaf	Integer	[1, 4]
	max_features	Categorical	{auto, sqrt, log2}[3]
svm	svr__C	Real	[0.1, 10] (log-uniform)
	svr__kernel	Categorical	{linear[1], rbf[2]}
Decision tree	max_depth	Integer	[10, 50]
	min_samples_split	Integer	[2, 10]
	min_samples_leaf	Integer	[1, 4]
	max_features	Categorical	{auto, sqrt, log2}[3]
KNN	n_neighbors	Integer	[1, 20]
	weights	Categorical	{uniform, distance}[4]
	p	Integer	[1, 5]
XGBoost	n_estimators	Integer	[10, 300]
	learning_rate	Real	[0.001, 0.2] (log-uniform)
	max_depth	Integer	[1, 10]
	min_child_weight	Integer	[1, 10]
	gamma	Real	[0, 0.5]
	subsample	Real	[0.1, 1.0]
	colsample_bytree	Real	[0.1, 1.0]
Ridge regression	ridge__alpha	Real	[0.001, 10] (log-uniform)
	ridge__solver	Categorical	{svd[5], cholesky[6], lsqr[7], sparse_cg[8], sag[9], saga[10]}
Neural network	batch_size	Integer	[32, 512] (step 32)
	units_input	Integer	[32, 512] (step 32)
	activation_input	Categorical	{relu[11], elu[12], selu[13]}
	l2_input	Float	[1e-5, 1e-2] (sampling='log')
	dropout_input	Float	[0.1, 0.8] (step 0.1)
	num_layers	Integer	[1, 4]
	units_{i}	Integer	[16, 512] (step 16)
	activation_{i}	Categorical	{relu[11], elu[12], selu[13]}
	l2_{i}	Float	[1e-5, 1e-2] (sampling='log')
	dropout_{i}	Float	[0.1, 0.8] (step 0.1)
	learning_rate	Float	[1e-5, 1e-2] (sampling='log')
	decay	Float	[1e-8, 1e-4] (sampling='log')
Random Forest	Number of estimators	Integer	[10,200]
	Maximum features	Categorical	{auto, sqrt, log2}[3]
	Maximum depth	Integer	[10,100]
	Minimum sample split	Integer	[2,10]
	Minimum sample leaf	Integer	[1,4]
	Bootstrap	Categorical	{True, False}

(1) 'Linear' signifies that the SVM's kernel utilizes a linear function to create a decision boundary. (2) 'RBF' or Radial Basis Function is a non-linear kernel for SVM, projecting data to higher dimensions for class separability. (3) 'Auto', 'sqrt', and 'log2' are options for the maximum features parameter in tree-based models, where 'auto' uses all features and 'sqrt' and 'log2' use the square root or log base 2 of the feature number, respectively. (4) 'Uniform' and 'distance' are choices for the 'weights' parameter in KNN models. In the 'uniform' case, all points in each neighbourhood are weighted equally, whereas 'distance' weights points by the inverse of their distance, giving closer neighbours more influence. (5) 'SVD', or Singular Value Decomposition, and (6) 'Cholesky' are solvers for Ridge Regression that are particularly efficient for symmetric positive-definite matrices. (7) 'LSQR' or Least Squares QR is an iterative solver for Ridge Regression. (8) 'Sparse_CG' or Sparse Conjugate Gradient is an effective solver for large sparse linear systems. (9) 'SAG', or Stochastic Average Gradient, and (10) 'SAGA', or Stochastic Average Gradient Augmented, are optimization algorithms with particular benefits for large-scale machine learning problems, the latter offering improved support for non-smooth penalty functions. Furthermore, (11) 'relu', (12) 'elu', and (13) 'selu' are activation functions in the neural network. 'Relu' or Rectified Linear Unit introduces non-linearity without affecting the receptive fields of convolutions. 'Elu' or Exponential Linear Unit helps to mitigate the vanishing gradient problem, and 'selu' or Scaled Exponential Linear Unit modifies 'elu' to have self-normalizing properties in neural networks.

Appendix D

Precision, recall and accuracy per class for classification with RF.

Model	Class	precision.	Recall	f1-score	support
Random Forest	**A**	0,9286899	0,9807356	0,9540034	571
	B	0,544	0,3541667	0,4290221	192
	C	0,6027944	0,3719212	0,4600152	812
	E	0,7802632	0,9198552	0,8443284	1934
	accuracy	0,7720148	0,7720148	0,7720148	0,7720148
	macro avg	0,7139369	0,6566697	0,6718423	3509
	weighted avg	0,7504212	0,7720148	0,7505192	3509
Gradient Boosting	**A**	0,9245902	0,9877408	0,9551228	571
	B	0,560241	0,484375	0,5195531	192
	C	0,5534351	0,3571429	0,4341317	812
	E	0,7818017	0,8929679	0,8336954	1934
	accuracy	0,7620405	0,7620405	0,7620405	0,7620405
	macro avg	0,705017	0,6805567	0,6856257	3509
	weighted avg	0,7400687	0,7620405	0,7438048	3509
SVM	**A**	0,9259868	0,9859895	0,9550466	571
	B	0,5310345	0,4010417	0,4569733	192
	C	0,5470383	0,3866995	0,4531025	812
	E	0,7804766	0,8805584	0,8275024	1934
	accuracy	0,7571958	0,7571958	0,7571958	0,7571958
	macro avg	0,6961341	0,6635723	0,6731562	3509
	weighted avg	0,7364873	0,7571958	0,7413449	3509
Decision Tree	**A**	0,9296482	0,971979	0,9503425	571
	B	0,5895522	0,4114583	0,4846626	192
	C	0,5458716	0,2931034	0,3814103	812
	E	0,7608881	0,9214064	0,8334892	1934
	accuracy	0,7563408	0,7563408	0,7563408	0,7563408
	macro avg	0,70649	0,6494868	0,6624761	3509
	weighted avg	0,7292187	0,7563408	0,7288042	3509
KNN	**A**	0,9039088	0,971979	0,9367089	571
	B	0,5423729	0,3333333	0,4129032	192
	C	0,5392857	0,3719212	0,4402332	812
	E	0,7722147	0,885212	0,8248615	1934
	accuracy	0,7503562	0,7503562	0,7503562	0,7503562
	macro avg	0,6894455	0,6406114	0,6536767	3509
	weighted avg	0,7271675	0,7503562	0,731516	3509
XGBoost	**A**	0,9278689	0,9912434	0,9585097	571
	B	0,5869565	0,421875	0,4909091	192
	C	0,5555556	0,455665	0,5006766	812
	E	0,8071599	0,8743537	0,8394142	1934
	accuracy	0,7717298	0,7717298	0,7717298	0,7717298
	macro avg	0,7193852	0,6857843	0,6973774	3509
	weighted avg	0,756531	0,7717298	0,7613395	3509
Logistic Regression	**A**	0,9262295	0,9894921	0,9568163	571
	B	0,56	0,4375	0,4912281	192
	C	0,5646388	0,3657635	0,4439462	812
	E	0,7822762	0,8991727	0,8366611	1934
	accuracy	0,7651753	0,7651753	0,7651753	0,7651753
	macro avg	0,7082861	0,6729821	0,6821629	3509
	weighted avg	0,7431764	0,7651753	0,7464362	3509

Appendix E

Patient schedules based on synthethic waitinglist

Patient ID	Specialisation	Starting Time	Leaving Time	Bed	Expected LOS
470	ORT	08:00	10:39	1	159
27	ORT	10:39	15:26	1	287
411	ORT	15:26	20:13	1	287
280	PLA	08:00	10:42	1	162
344	PLA	10:42	14:51	1	249
423	PLA	14:51	19:10	1	259
485	CHI	08:00	09:46	1	106
382	CHI	09:46	13:49	1	243
444	CHI	13:49	17:57	1	248
375	PYN	08:10	08:22	1	12
149	PYN	08:22	08:34	1	12
72	PYN	08:34	09:00	1	26
6	PYN	09:00	09:28	1	28
243	PYN	09:28	09:59	1	31
36	PYN	09:59	10:34	1	35
262	PYN	10:34	11:09	1	35
479	PYN	11:09	11:46	1	37
468	PYN	11:46	12:24	1	38
163	PYN	12:24	13:03	1	39
460	PYN	13:03	13:43	1	40
173	PYN	13:43	14:23	1	40
128	PYN	14:23	15:08	1	45
169	PYN	15:08	15:57	1	49
67	PYN	15:57	16:49	1	52
207	PYN	16:49	17:42	1	53
187	PYN	17:42	18:36	1	54
156	PYN	18:36	19:31	1	55
14	PYN	19:31	20:26	1	55
2	KNO	08:00	10:32	1	152
319	KNO	10:32	13:28	1	176
39	KNO	13:28	17:54	1	266
236	URO	08:00	10:52	1	172
410	URO	10:52	14:38	1	226
162	URO	14:38	18:26	1	228
421	CAR	08:00	10:12	1	132
400	CAR	10:12	12:44	1	152
40	CAR	12:44	15:53	1	189
323	CAR	15:53	19:27	1	214
155	DER	08:00	10:15	1	135
182	DER	10:15	13:17	1	182
57	DER	13:17	16:24	1	187
81	DER	16:24	19:50	1	206