WESTERN SYDNEY
UNIVERSITY

Predictive Analytics Framework For Electronic Health Records with Machine Learning Advancements:

Optimising Hospital Resources Utilisation with Predictive and Epidemiological Models

---

Belal Alsinglawi

A thesis submitted in fulfilment of the degree of

Doctor of Philosophy

School of Computer, Data and Mathematical Sciences

Western Sydney University

January 2022

Abstract

Resources utilisation in hospitals is vital for hospital healthcare management systems. Managing hospital beds' availability and efficiency are essential for addressing challenges associated with having an overabundance of patients in ICU and hospital. Further, the aim is to avoid any issue of hospital beds scarcity, especially in uncertainties such as pandemics. Hospitals strive to manage their resources whilst improving patient's healthcare outcomes within dynamic hospital environments where unpredictable situations (internal or external) may occur at any time. One of the most hospital resources' demanding factors is inpatients' extended stay during their hospitalisations. The inpatient length of stay (LOS) in hospitals is often used to measure hospital efficiency. It is considered one of the most significant indicators for the consumption of hospital resources. A shorter stay during inpatient hospitalisations will reduce the cost per discharge and shift care from hospital inpatient to less costly post-acute settings. However, stays that are too short may reduce the quality of healthcare and lead to poorer patient outcomes. Contrarily, longer hospital lengths of stay are often due to complications and may be associated with a higher risk of adverse events.

Furthermore, the longer LOS can be due to factors unrelated to the patient's clinical condition, such as the delays in consulting or coordinating healthcare with other healthcare professionals who have a role in assisting the patient's recovery. Current methods (APACHE versions (I, II, III, IV) and SAPS, SOFA) are traditional LOS scoring systems and are examined in the literature to estimate patients' length of stay during hospitalisation. However, they suffer from drawbacks such as their inability to provide accurate estimates about patients' LOS, especially related to patient-centred health outcomes, and poor performance. In addition, they are not designed to capture the state of dynamic operations flows at hospitals and identities patterns from hospital electronic health records (EHRs). Additionally, conventional clinical information systems such as clinical decision support systems still operate in traditional inpatient LOS estimation. Therefore, there is no improved hospital bed planning and management. Consequently, this results in resource shortage and may result in the denial of new patients, negatively impacts hospital quality of health services.

To improve hospital resources utilisation, this thesis introduces a new length of stay predictive research framework to predict inpatient hospitalisation at the time of hospital admission using

state-of-the-art of machine learning models. The framework assesses different moods of hospital admissions, such as general, emergent, and intensive care-based admission, in the form of case studies, using discharge data of inpatients from hospital EHRs data (opensource and real hospital data). Then, it identifies the type of admission that is likely to consume hospital resources, especially the extended LOS (long LOS). Accordingly, the thesis focuses on the most resources demanding type of admission assessed based on the long LOS. The framework is capable of exploring and discovering diseases that are frequently depleted hospital resources by using the relevant exploratory data analysis (EDA) techniques. The EDA aimed to understand the property of the hospital EHRs data and discover patterns and associations through statistical analysis and visualisations.

In this research, two experimental approaches were followed to examine the best method that could potentially be used as a practical research framework to predict patients' length of stay more efficiently and make accurate health and clinical judgments by healthcare workers such as beds managers, clinicians, and decision-makers at hospitals. The first experimental approach examined the use of the machine learning ensemble regression-based techniques to estimate hospital LOS. The second experimental method utilised ensemble binary classification learning. The experiments unveil facts that the classification-based approach is more reliable and efficient for hospital resource utilisations. In addition, it can guide hospital healthcare workers to understand the admission factors that decide that the patients are likely to be predicted short or long stay. However, adopting a predictive model in CIS and CDSS goes beyond accuracy and the ability of the model to achieve high predicted results. The critical part of the predictive model is safety, and that machine learning models are expected to explain their predictions. This includes situations when the ML models cannot explain what is intended. For this purpose, the research framework exploited explainable machine learning tools to explain the inner working of LOS predictive models. Therefore, the beds' managers or hospital clinicians can establish their decision on clear ground and based on the individual admitted cases (patient-centred).

The final part of the thesis evaluated the situation of uncertainties such as pandemics. The novel COVID-19 stressed hospitals and exhausted their resources and staffing globally. This motivated the thesis to examine the epidemiological models and estimate cases that may require hospitalisations using an empirical portion of the population. The epidemiological simulation

showed that public health containment strategies might control the number of cases that need to be admitted to the hospital and avoid an overabundance of patients who may require to attend intensive care units. The research highlighted the importance of the predictive machine learning models to predict COVID-19 hospitalisations from open-access linked data sources. The findings support the usefulness of prediction in the influx of patients during pandemics and how machine learning predictive models can assist healthcare workers and improve hospital resources utilisation from a data-driven approach.

Declaration of Original Work

I certify that this thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any university and that, to the best of my knowledge and belief, it does not contain any material previously published or written by another person, except where due reference is made in the text.

Signed:  10/09/2021

Dedication

I dedicate this work to my parents, wife, and brothers and sisters for always supporting me in my study and life. Also, I dedicate it to my beautiful children (Judi and Jana) for always making me smile and motivated.

Acknowledgements

First of all, I feel proud and privileged to express my deep, sincere gratitude and indebtedness to my advisor Dr Omar Mubin of the School of Computer, Data and Mathematical Sciences, Western Sydney University, for all of the support and encouragement he gave me. His guidance, motivation and countless advice steered me in the right direction whenever he thought I needed it during my academic journey. It is a great honour to work under his supervision.

I would express my deepest thanks and sincere appreciation to Dr Mauricio Novoa of the School of Engineering, Design and Built Environment, Western Sydney University, and Dr Fady Alnajjar of the Department of Computer Science and Software Engineering, College of IT, United Arab Emirates University (UAEU) for their valuable suggestions and comprehensive advice during my study. I would like to thank the Al-Ain hospital research ethics committee, UAEU, for facilitating the ethics approval for the data collection.

Also, I would like to express my special appreciation to Dr Mohammed Alorjani, Dr Osama Alshari, and Dr Khaled KhierAllah of the Medical School of Jordan University of Science and Technology, for their valuable clinical and epidemiological inputs and assistance during my PhD journey. Finally, I would like to thank A/Prof Dongmo Zhang of Computer, Data and Mathematical Sciences, Western Sydney University, for making PhD journey enjoyable and for his continuous support of PhD and research students.

Table of Content

Contents

List of Tables

List of Figures

Publications

The following are the list of the publications outputs from the thesis.

Published Papers

1. **Alsinglawi, B.**, and Mubin, O. 2019. "Predictive analytics and deep learning techniques in electronic medical records: recent advancements and future direction". ***Workshops of the International Conference on Advanced Information Networking and Applications,*** p.907–914.
2. **Alsinglawi, B**., Alnajjar, F., Mubin, O., Novoa, M., Alorjani, M., Karajeh, O., and Darwish, O.. "Predicting Length of Stay for Cardiovascular Hospitalizations in the Intensive Care Unit: Machine Learning Approach". ***42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC-2020, Canada)***.
3. **Alsinglawi, B**., Alnajjar, F., Mubin, O., Novoa, M., Karajeh, O., and Darwish, O. 2020. "Benchmarking predictive models in electronic health records: sepsis length of stay prediction". ***Advanced Information Networking and Applications: Proceedings of the 34th International Conference on Advanced Information Networking and Applications (AINA-2020), 15-17 April 2020, Caserta, Italy*****,** p.258–267.
4. Kheirallah, K., **Alsinglawi, B**., Alzoubi, A., Saidan, M., Mubin, O., Alorjani, M., and Mzayek, F. 2020. "The effect of strict state measures on the epidemiologic curve of COVID-19 infection in the context of a developing country: a simulation from Jordan". ***International journal of environmental research and public health****, 17*(18), p.6530.
5. **Alsinglawi, B**., Mubin, O., Alnajjar, F., Kheirallah, K., Elkhodr, M., Al Zobbi, M., Novoa, M., Arsalan, M., Poly, T.N., Gochoo, M. and Khan, G., 2021. A simulated measurement for COVID-19 pandemic using the effective reproductive number on an empirical portion of population: epidemiological models. ***Neural Computing and Applications***, pp.1-9.
6. **Alsinglawi, B.**, Alshari, O., Alorjani, M., Mubin, O., Alnajjar, F., Novoa, M. and Darwish, O., 2022. An explainable machine learning framework for lung cancer hospital length of stay prediction. ***Scientific Reports*****,** 12(1), pp.1-10.

List of Abbreviations

General Abbreviations

AIHW	Australian Institute of Health and Welfare
AMI	Acute Myocardial Infarction
APACHE	Acute Physiology And Chronic Health Evaluation
AMI	Acute Myocardial Infarction
CAD	Coronary Artery Disease
CFS	Clinical Frailty Scale
CCU	Coronary Care Unit
CICU	Cardiac Intensive Care Unit
COVID-19	Coronavirus Disease of 2019
COPS	Comorbidity Point Score
CPOE	Computerized Provider Order Entry
CDSS	Clinical Decision Support Systems
EHR	Electronic Health Records
EMR	Electronic Medical Records
ED	Emergency Department
ER	Emergency Rooms
ESI	Emergency Severity Score
HREC	Human Research Ethics Committees
HF	Heart Failure
KTAS	Korean triage and acuity score
ICD	International Classification of Diseases
IRPM	Intelligent Remote Patient Monitoring
ICU	Intensive Care Unit
SAPS	Simplified Acute Physiology Score
SOFA	Sequential Organ Failure Assessment
SIRS	Systemic Inflammatory Response Syndrome
LAPS	Laboratory Acute Physiology
LOS	Length of Stay
LDAPPM	Learning, Data Analytics, Predictive and Personalized Medicine
LCGMH	Linko Chang-Gung Memorial Hospital
NPIs	Non-Pharmaceutical Interventions
MEWS	Modified Early Warning Score
PED	Pediatric Emergency Department

Technical Abbreviations

AI	Artificial Intelligence
ADASYN	Adaptive Synthetic
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AUC	Area Under Curve
AUROC	Area Under Receiver Operating Characteristic Curve
CIS	Clinical Information Systems
CI	Confidence Interval

CS	Clinical Signifiance
CPU	Central Processing Unit
DNN	Deep Neural Network
GBR	Gradient Boosting
GBR	Gradient Boosting Regressor
G.mean	Geometric Mean Score
ENN	Edited Nearest Neighbours
EDA	Exploratory Data Analysis
kNN	k-Nearest Neighbors
ML	Machine learning
MLP	Multi-layer Perceptron
MLNN	Multilayer neural network models
MLA	Machine-Learning Algodiagnostic
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Main Square Error
NaN	Non-Values
NN	Neural Network
TPR	True Positive Rate
FPR	False Positive Rate
TN	True Negative
FN	False Negative
TP	True Positive
FP	False Positive
R2	R-squared
RF	Random Forest
RRF	Regression Random Forest
RMSE	Root Mean Squared Error
RMSLE	Root Mean Square Logarithmic Error
ROC	Receiver Operating Characteristic
RFE	Recursive Feature Elimination
LSTM	Long Short-Term Memory
LR	Logistic Regression
IBA	Index Balanced Accuracy
LDA	linear discriminant analysis
SAE	Stacked Autoencoders
SMO	Sequential Minimal Optimization
SVM	Support Vector Machine
SGD	Stochastic Gradient Descent
STD (std)	Standard Deviation
SMOTE	Synthetic Minority Oversampling Technique
SHAP	Shapley Additive Explanations
SEIR	Susceptible Exposed Infectious Recovered
VAE	Variational Autoencoders
XGBoost/XGB	extreme Gradient Boosting
xAI	explainable Artificial Intellige

1. Chapter One: Thesis Introduction

1.1 Background

Hospitals face problems related to the availability of limited hospital resources, including the limited number of beds availability, staffing, and medical equipment in hospital settings [1]. In addition, the growing demands on healthcare facilities and the pressure to improve medical quality and patient outcomes [2] have a persistent need to address issues associated with limited hospital resources. These challenges appear in scheduling and resource allocation on patient wait time, clinic overtime, overcrowding, excessive delays, and concerns regarding the safety of critical patient care [3] in the emergency room, intensive care units (ICU) or general hospital admissions. One of the quality care factors used in the hospital setting is called inpatients length of stay (LOS). Hospital length of stay elucidates the time interval between hospital admission and discharge time from the hospital [4]. The length of hospital stay is a quality metric for the hospital to manage inpatient length of stay during their hospitalisation at different hospital facilities. It helps hospital beds managers, clinicians, and nurses track patient stay from admission to the hospital to hospital discharge. Improving hospital resources utilisation depends on the hospital's length of stay(internal factor). Therefore, reducing inpatient LOS while maintaining patient quality of care is vital in hospital settings.

There are differences in the average length of stay at public hospitals compared to the average length of stay at private hospitals. For instance, in the Australian hospital context, the average length of stay in intensive care units (ICU) was nearly four days in public hospitals. In contrast, it was just a little over two days in private hospitals, according to the report by the Australian Institute of Health and Welfare (AIHW) 2016-2017 [5]. Also, the same report showed that the average length of stay for emergency surgery is increasing by 2.7% and 3.6 % for public and private hospitals, respectively, each year. Further, the average LOS for acute care was slightly higher in public hospitals (2.4 days) than in private hospitals (1.9 days), according to the AIHW report. In the United States, patients admitted to private treatment centres, and public treatment centres have, on average, 40% and 18% shorter LOS compared to National Health Service (NHS) public hospitals [6].

On the other hand, the average LOS for the cases of other sub-acute, non-acute care, rehabilitation care and mental health stay in the public hospital was much higher than in private hospitals. These figures indicate a gap between the medical services offered in private hospitals and public hospitals, which eventually leads to a longer stay in the public hospitals. As a result, this will add more costs to patients, healthcare insurance companies, and the government health care system. Also, it adds more stress on public hospitals' management and healthcare assessment systems to manage the shortage of beds availability.

In clinical practice and hospital settings, the current patient monitoring scoring systems (LOS), such as the *APACHE II (acute physiology and chronic health evaluation),* are traditional systems used to predict the probability of length of stay in the ICU and hospital mortality [7]. APACHE II is rendered within the first 24 hours of the patients' admission to the ICU. APACHE II classifies the disease severity in a classification system (an integer from 0-71) and is calculated based on several measurements. APACHE III is the later version of the APACHE II, which aims to provide initial risk conformity for severely ill hospitalised patients within an independently defined patient group and in ICU admission [8]. APACHE III consists of a set of questions to predict hospital mortality, ICU mortality, hospital length of stay, and ICU length of stay [9], with the score (range 0 to 299) calculated based on measurement inputs. APACHE IV has similar discrimination to APACHE III, whereas the calibration in APACHE IV is better than in the case of APACHE III [10]. The SOFA (Sequential Organ Failure Assessment or Sepsis-related organ failure assessment) is a clinical score system implemented in the ICU to monitor patients and identify their organ function or the rated failure. The SOFA has six sub-scoring systems, including cardiovascular, nervous, respiratory, liver, renal, and coagulation systems, with the associated scores (0, +1, +2, +3, +4). However, these scoring systems are limited to specialised care in acute medical conditions such as ICU settings. Also, they suffer from poor performance; hence, they are not disease-specific prediction methods. Further, there is consensus on the most suitable techniques for LOS [11] in ICU and hospital settings.

The advancement in health informatic systems has shown an unprecedented amount of rich clinical data of patients. Electronic health records (EHRs) have generated feasible opportunities to help clinicians understand and address a broad range of questions in medicine [12]. The widespread adoption of EHRs in health informatics and clinical information and decision support systems has

opened the doors to improve the healthcare domain, support clinical decision-making, and provide better inpatients outcomes. Typically, EHRs contain the patient's medical history, diagnoses, medications, treatment plans, immunisation dates, allergies, radiology images, laboratory and test results [13], and all administrative medical data such as LOS. Within electronic health records, the length of stay is an important quality metric that aims to determine how long the patients can stay in the hospital from admission to hospital to discharge. The LOS is measured in months, weeks, days, and hours. LOS metric measure is essential for hospitals and medical centres to decide the type of diagnosis for each group within the particular medical department (e.g. ICU). In addition, the LOS relates to the distribution of resources and planning. The high demands on the EHRs in the medical field have encouraged many researchers [4, 14] to build predictive health analytic modellings from electronic health records (e.g. classification and regression methods). The LOS Predictive modellings play a significant role in easing the hospital's operational success in the hospital and emergency departments, ICU units and hospital facilities. Machine learning techniques facilitate the doors in the clinical field and help the researchers in the domain of hospital LOS to construct predictive medical models and health analytical tools that can support clinical decision making and healthcare systems in hospital settings.

Health data furnishes rich information generated from multiple resources in hospital settings, including uncovering health implications, risk factors, heterogeneous and noisy data records, making health decisions such as diagnosis, triage, and treatment hard to predict. Besides its heterogeneity, EHRs are longitudinal, which comprise clinical information of various types of data recorded over time [15]. As a result, these issues add more complexity to medical prediction tasks, especially in the daily hospital routine (e.g. LOS prediction) carried by medical professionals such as physicians, beds managers, nurses, etc. Also, the complexity appears in the rapid growth of electronic medical data, which leads to more difficulties human medical experts are facing [16] in various predictive tasks such as LOS. Researchers obtain access to the EHRs data through agreed conditions such as the ethical permissions by hospitals or governments responsible for providing a secondary use of the EHRs data.

Machine learning (ML) can work effectively with EHRs in an automatic and collective process [17]. Hence, it helps the researchers and data scientists in the medical domain collect massive data, including patients' electronic health records, behavioural data, temporal data, and time-series data;

therefore, better intelligent diagnostic predictions will be achieved. For example, predicting the number of days patients are required to stay at hospitals would greatly benefit hospitals, patients, and caregivers and prevent late or early discharge for particular medical problems requiring prolonged LOS. In prediction learning modellings, whether the specific medical case can be best predicted by supervised learning or unsupervised models, or even a hybrid approach, it is necessary to deal with the complex medical case that requires appropriate data representation prior to the prediction process. Previous research tells us that identifying the baseline [18] predictive learning model in the length of stay prediction task is rendered a significant task in LOS healthcare assessment systems. Certainly is considered a research challenge, especially to benchmark algorithms for hospital resources utilisation that works best with the particular LOS prediction task.

1.2 Problem Statement and Research Motivation

The demand for medical services is increasing rapidly and globally. Handling bed availability, identifying the most demanding hospitalisations and managing bed occupancy and hospital resources via the length of stay create persistent needs for physicians, nurses, clinicians, beds managers, and hospital management and clinical decisions [19]. There are many attempts in the literature review with prediction algorithms to achieve reliable outcomes of the length of stay globally [18]. However, no authentic research attempts to provide an effective and research practical framework [18] for predicting the length of stay during hospital admissions, ICU, and emergency departments from benchmarking and explainable approaches. Further, the literature did not examine the external factors such as a pandemic that can stress hospital resources in the near future. Machine learning algorithms can determine how long such a particular admission may need to stay at the hospital and occupy bed space and explain the prediction outcomes of non-data specialist people such as beds managers [20]. Moreover, in EHRs, modelling data categorisation plays a dynamic role in exploiting the prediction method in a particular LOS prediction task.

Nonetheless, there are few studies in the literature review to address the challenge of benchmarking the appropriate predictive models in LOS prediction tasks and predicting future hospitalisations during uncertainties such as a pandemic. Consequently, finding the best predictive measurement methods with the highest reliable predictive explained outcomes to predict the length of stay tasks in ICU admissions and emergency rooms remains a research challenge. Developing a

methodological research framework that considers the appropriate data category and data representation is vital in forecasting the length of stay. Therefore, building a proper framework to implement and baseline prediction algorithms based on the type of hospital admission to predict the length of stay is deemed an imperative requirement in clinical information systems and for hospital resource utilisation. The predictive research framework will be a useful research tool for health professionals and researchers in healthcare assessment to facilitate the mission and provide better support for decision-making and improved patient outcomes. Further, it will benefit decision-makers during the time of uncertainties to predict the likelihood of future hospital stays for healthcare systems and assist hospital manger in planning ICU and hospital beds in more proactive and efficient ways. Indeed, this will help regulate the predictive tasks in personalised and healthcare recommendations and standardise prediction tasks in healthcare and clinical settings based on the data input.

The COVID-19 pandemic stressed hospitals worldwide. For instance, during the first wave of the pandemic in 2020, Spain was one of the highest countries in Europe, and suffered from high infected hospitalised rates. The beds occupancy by COVID-19 of acute care hospitals beds reached 100% by March 28th 2020, and 105% by April 6th. Additional beds were placed in improvised wards such as physical therapy gyms, corridors libraries, and tents outside the main hospitals. Moreover, the ICU beds occupancy reached 300% on the 6th of April from the same year [21]. This massive stress on hospitals in Madrid where some hospitals repurposed postanesthesia care units, pediatric ICUs and cardiac and coronary care units to manage the overabundance of COVID-19 during the peak time.

The epidemiological modelling tool can predict the number of near-future hospitalisation and ICU beds needed during an outbreak or a larger pandemic such as COVID-19 in public health settings. The use of hospitalisation and ICU forecasted rates for calibration allows for reliable forward prediction of ICU needs and bed occupancy during the outbreak, especially when the effects of government measures are not yet effective. Therefore, it allows the evaluation of the effects of a timely adaptive response easily. Moreover, it can also help to assess the effectiveness and impact of policy measures taken to control the outbreak by calibrating the effects to incoming data daily [22]. This motivates this research to assess the epidemiological curves and their projection on measuring the ratio of future hospitalisations. Also, it guides healthcare professionals and public

health policymakers to apply appropriate public health measures to control the spread of communicable diseases within infected communities. Most importantly is to avoid the damaging impact of the outbreak on the healthcare system, including the hospital resources, protecting the healthcare system from potential failure and keeping hospital beds occupancy and resources under control.

To this end, predicting hospital length of stay has always been a challenging hospital operational task. Managing scarce hospital resources plays a significant role in effectively operating hospitals' operation rooms, specialised doctors' rooms, emergency rooms and ICU rooms. Forecasting hospital length of stay allows the hospital to deliver high-quality care and provide improved patient health outcomes. Examining the external factors that directly impact the healthcare systems and hospitals by forecasting future numbers of hospitalised cases using epidemiological models has a significant benefit in efficiently preparing and predicting LOS in hospital settings via the proactiveness approach.

1.3 Research Questions

In the view of above, the thesis addressed the following research questions:

1. How can machine learning algorithms assist in constructing a thorough methodological predictive length of stay approaches research to facilitate hospital resources utilisation in emergency, chronic, acute and pandemic situations?
2. How can the predictive outcomes of machine learning make data sense for healthcare workers via clinical information systems?
3. How can epidemiological models project future hospitalisations and assist hospital management proactively to avoid the scenarios of patients' overabundance during pandemics?

While the first question addresses challenges assisted with benchmarking predictive models in electronic health records for hospital resources utilisation, the second question addresses challenges in the predictive outcomes of machine learning models for the LOS predictions tasks. Thus, both questions address information systems' data challenges via electronic health records. The first two questions are essential to deal with the challenges of benchmarking LOS predictive models and explain outperforming models' working inners. The third question is an integral

research question in the thesis to assess the factors influencing hospital bed occupancy during pandemics and apply suitable public health measures to protect the healthcare system.

1.4 Research Objectives and Scope

This research utilises various data mining techniques, including feature selection, feature engineering, missing values handling, class-balancing methods and model tuning techniques, to provide a comprehensive predictive length of stay research framework for the clinical information systems to improve hospital resources utilisation. Moreover, the research exploits the stochastic epidemiological models to assess and forecast the external factors that exhaust hospital resources with the plethora of patients and prepare the hospital management to proactively manage staffing, hospital resources, and bed occupancy more efficiently. The main objectives of the thesis are as follows:

1) To assess predictive models in the context of predicting length of stay in emergency admissions.
2) To assess the robustness of regression predictive methods against classification (binary) approaches and justify which prediction technique suites the context of LOS for hospital resources utilisation.
3) Predict LOS in the state of limited data input and deal with imbalanced input for a better data representation in ICU predictive tasks.
4) To validate the predictive LOS models on a real hospital dataset via an inclusive research framework.
5) To assess the external factors (public health measures) in pandemics and how the future forecasted hospitalisation can demand hospital resources.

The objectives are studied deeply and addressed in each chapter. For instance, objective 1 is studied and addressed in chapter 2. Objectives 2-5 are studied and addressed accordingly in Chapters 3-6, respectively.

1.5 Research Contributions

Each chapter in the thesis contributes toward the objectives of the thesis. Therefore, the contribution of the thesis per each chapter is as follows:

- Development and feature engineering (emergency department) for benchmarking predictive LOS framework for the emergency department to assess the predictive models and evaluating their performance in emergency inpatient hospitalisations. Chapter 2 provides a fundamental background for the thesis by assessing predictive (ensemble learning models) robustness in emergency admission, including emergency rooms. Moreover, the chapter addresses the model's black box issue by utilising a predictive outcomes explanation approach that provides an insightful task for decision-makers in healthcare assessment settings.
- ICU features extraction and development of benchmarking predictive LOS architectures models for optimising LOS hospital resources utilisation in intensive care units to address the challenge of the suitability of prediction tasks in acute prediction settings (ICU). The proposed case studies are selective to accomplish the aims of the chapters. Two predictive approaches (classification and regression) were followed to determine which direction is more appropriate for the LOS prediction task and for better decision-making by ICU beds managers and hospital professionals.
- Designing, developing, and utilising data class balancing methods and features selection approaches to deal with the issue of imbalance predictive LOS tasks in the ICU setting, especially with the lack of data input in the case of LOS diseases-focused tasks such as lung cancer. Chapter 4 provides a comprehensive methodological framework that addresses these issues. The proposed architecture is an algorithmic contribution to optimise the workflow of hospitals' management systems within the clinical information systems to support health decisions by healthcare workers in ICU settings. The chapter provides a methodological architecture to explain the predictive outcomes of machine learning ensemble models. They are explainable for doctors to support their decision for patient safety and improved health outcomes. Further, it assists junior doctors who are practising in ICU settings. Finally, it facilitates the mission for beds managers to manage bed occupancy efficiently based on patients' profiles and predicted explained ranked admission features.
- Validating the proposed methodological framework (ensemble learning models) in real hospital data to justify the usefulness of the predictive proposed LOS approaches in this thesis in real hospital data. Designing an algorithmic LOS predictive contribution using

multiple data prediction procedures to attest, verify and evaluate the proposed framework in LOS-ICU predictive tasks. A Complete EHRs features mining and extraction were performed to capture the whole patient's profile in this chapter. The framework was evaluated in patients' multiple inputs (complete patient EHRs profile, clinical profile, general admission profile, laboratory profile, and medication profile). This approach ascertains that the prediction framework can perform well in different EHRs data inputs. The predictive ensemble models provided insightful knowledge and directions for decisions for healthcare professionals via the complete patient EHRs profile. Explainable predictive model approaches were evaluated to explain the outcomes of the prediction models for the healthcare workers from the whole state of the ICU hospitalisations and for patient-centred predictive tasks. The explainer is a boundless explainable AI approach for providing detailed predictive information about the patients, the cohort of patients, or the whole state of the ICU hospitalisation at particular admission times. It supports beds managers, clinicians and nurses' decisions in safe and trusted inpatient LOS outcomes.

- Assessment of public health measures (non-pharmaceutical interventions "NPIs") in time of pandemics and uncertainties to evaluate their impact on future hospitalised cases via the development of epidemiological stochastic models. This chapter examined the external factors (NPIs) on hospital resources utilisation and the projection of the potential future hospitalised cases. The simulated models and outcomes in this chapter 6 are health guidance for hospital managers, clinicians and nurses to follow a proactive approach during pandemics with insightful results from the epidemiological stochastic models. The epidemiological stochastic models captured the different scenarios through simulations from real reported COVID-19 daily cases and forecasted the future hospitalised cases. The findings of these simulations in the chapter are supported with the proposed LOS-COVID-19 predictive machine-learning architecture using linkage COVID-19 dataset (from real hospitals). The data is large, and the ensemble models confirmed their robust performance to predict (multi-label) COVID-19 categorical LOS.

1.6 Research Outlines

The structure of the thesis is based on the contribution of each chapter to the overall aims of the thesis. This includes the literature review and the methods that are suitable for each chapter. The

flowchart Figure 1 represents the thesis structure, and Table 1 demonstrates the place of each method against the research objective of each chapter in the thesis.

For this thesis, we use words (framework, research framework, approach, methodology, methodological research/ research framework) interchangeably. In this thesis, we define the predictive framework/approach as a methodological predictive LOS approach. The thesis has two branches of contribution. The first one is the common acute, semi chronic, chronic diseases, which are usual day-to-day affairs of the hospital (LOS), Chapters 2-5. The second one is the extreme case of a pandemic, which means more patients per day basis (epidemiological curves and LOS) in chapter 6 of the thesis.

Chapter 2 introduces the research background about the emergency department length of stay. Then it provides a literature review of the work related to this chapter. The chapter provides a discussion about the challenges of LOS in the context of the emergency department. The research question and objectives of this chapter are discussed and also connected to the main research questions and outcomes of this thesis. The research method and preliminary study of the whole thesis are introduced in this chapter. Finally, the results are discussed, where the main finding of this chapter is the introduction of LOS benchmarked predictive models in ED rooms to be further explored in the following chapters.

Chapter 3 provides two case studies to compare the suitability of LOS prediction methods for the context of hospital resources utilisation. The case studies (case studies 1 and 2) are disease-based and aimed to assess which predictive approaches (classification vs regression) are more decision-oriented and more suitable for LOS prediction within hospital resources utilisation and the healthcare professionals in ICU settings. First, the chapter initiates with the background and discusses the aims of the benchmarking classification vs regression. Then, the chapter has two subsections (literature review and methods) based on the disease-focus of each case study for the ICU-LOS. Next, the data description is provided and followed for both case studies. Finally, chapter 3 is concluded with a conclusion and future work section.

Chapter 4 provides a comprehensive framework for the LOS-ICU predictive models with a data-driven approach. First, the chapter evaluates different methods to select the most performing variables for LOS-ICU prediction via the variables selection procedure. Then, it studies the class balancing method in the case of imbalanced datasets or data input in LOS predictive tasks. Finally,

the chapter focuses on a disease to assess its proposed LOS predictive framework aligned with the thesis' aims and expectations. The chapter is structured as follows: background, literature review, methods, result, discussion, and research implication with an explainable-AI approach, and finally, the conclusion.

Chapter 5 provides a thorough LOS-ICU predictive framework and validates the thesis' LOS contribution to a real hospital dataset. The chapter is structured with background, literature review, the chapter's research questions and objectives, methods, results, discussion, and the practical implication of the chapter's contribution and conclusion. Chapter 5 comprehensively accumulates the previous chapters' contributions and provides a practical framework for researchers in clinical information systems and hospital resources utilisation to further advance the research in this area. The chapter provides the practical research framework for machine learning models to be used and safely adopted in hospital settings by healthcare workers.

Chapters 2-5 follow the best research practice in applied machine learning research in health analytics, where the insights from the predictive outcomes greatly benefit healthcare decision-makers and healthcare workers. Further, they ease the workflow of hospital operations and optimise resources allocation in more efficient and safer ways.

Chapter 6 assesses the external public health measure and policies that can directly impact hospital resources utilisation during the pandemic (COVID-19). These measures are the strict lockdown and curfew, the public health measures in various settings in homogeneous communities where the public health measures are clearly projected or in the case that they are not clearly projected. The chapter provides two case studies; the first case study evaluates the curfew and how it helps the healthcare sector, particularly hospitals, forecast hospitalised cases. In comparison, the second case study assesses the different measures based on real public health policies in two different settings. Both case studies use simulated epidemiological models and country population or empirical portion of the population to forecast and project future hospitalised cases. The simulations from case studies support the strict non-pharmaceutical interventions via public health measures to protect healthcare systems and prevent any potential shortage of ICU stays, which eventually guide hospital decision-makers to prepare the beds' occupancy proactively. This chapter is supported by a LOS predictive case study for COVID-19. The case study of LOS aligned with the thesis' overall aims and objectives to provide an exhaustive framework that puts internal

and external factors in the fundamental motivation of the research to address actual research and practical challenges for hospital resources utilisation.

Chapter 7 is the conclusion and the research summary of the thesis and the future directions.

1.7 Summary of Thesis Methods.

The following table summarises the predictive methods that are used in this thesis. In addition, the table refers to each method and explains which research objective is studied. The thesis follows a top-down approach (reverse pyramid). It starts with general ED (chapter 2), specific diseases (chapter 3,4), real data for the hospital ward (chapter 5), and pandemic (chapter 6).

The scope of this thesis is primality health analytical methods using data-driven approaches with machine learning algorithms, particularly ensemble learning classification and their in-hospital robustness length of stay predictive tasks for clinical information systems and improving hospitals resources utilising with the data and machine learning approach. Further, the thesis provides a unique opportunity for healthcare workers and researchers to understand the inner workings of the machine learning prediction, how the LOS prediction was made, and which hospital features from the EHRs. The other unique aspect of this thesis is the examination of stochastic epidemiological models to forecast and estimate future cases of communicable diseases in the situation of pandemics such as COVID-19 as a case study to evaluate and verify the effectiveness of these models to project future cases and enrich clinical information systems with public health insights from undiscovered trends in public health data and therefore assist hospital and healthcare systems to look at the possibilities of the problem of influx of the patients and therefore plan hospitals resources ahead.

Table 1 Summary of Thesis Methods

Method	**Research Objective(s)**	**Chapter**	**Research Objective & Publications / Papers Output**
Random Forest Regressor "RFR" [23]	2	3	(Publications 1, 2)
Gradient Boosting Regressor "GBR" [24]	2	3	
Stacking Regressor "SR" [25])	2	3	
Deep Neural Network "Regression" [26]	2	3	
Logistic Regression "LR" [27]	2, 3, 4	3,4,5	2: (Publication 3) 1-5: (Publications 1,3,6,7)
Random Forest "RF" [23]	1, 2, 3, 4, 5	2,3,4,5, 6	
eXtreme Gradient Boosting "XGB or XGboost" [28]	1, 3,4,5	1,4,5, 6	1-5: (Publications 6,7)
Gradient Boosting Machines "GB" [29]	1,3,4,5	1,4,5,6	
Stochastic Gradient Descent "SGB" [30]	2	3	2: (Publication 3)
K-Nearest Neighbors "kNN" [31]	2	3	
Decision Tree "DT" [32]	2	3	
Gaussian NB [33]	2	3	
Support Vector Machine "SVM" [34]	2,3	3,4	
Multi-layer perceptron neural network "MLP" [35]	4	5	(Publications 7)
SEIR Chapter 6" [36]	5	6	5:(Publications 4, 5)
Extended SEIR "Chapter 6"	5	6	

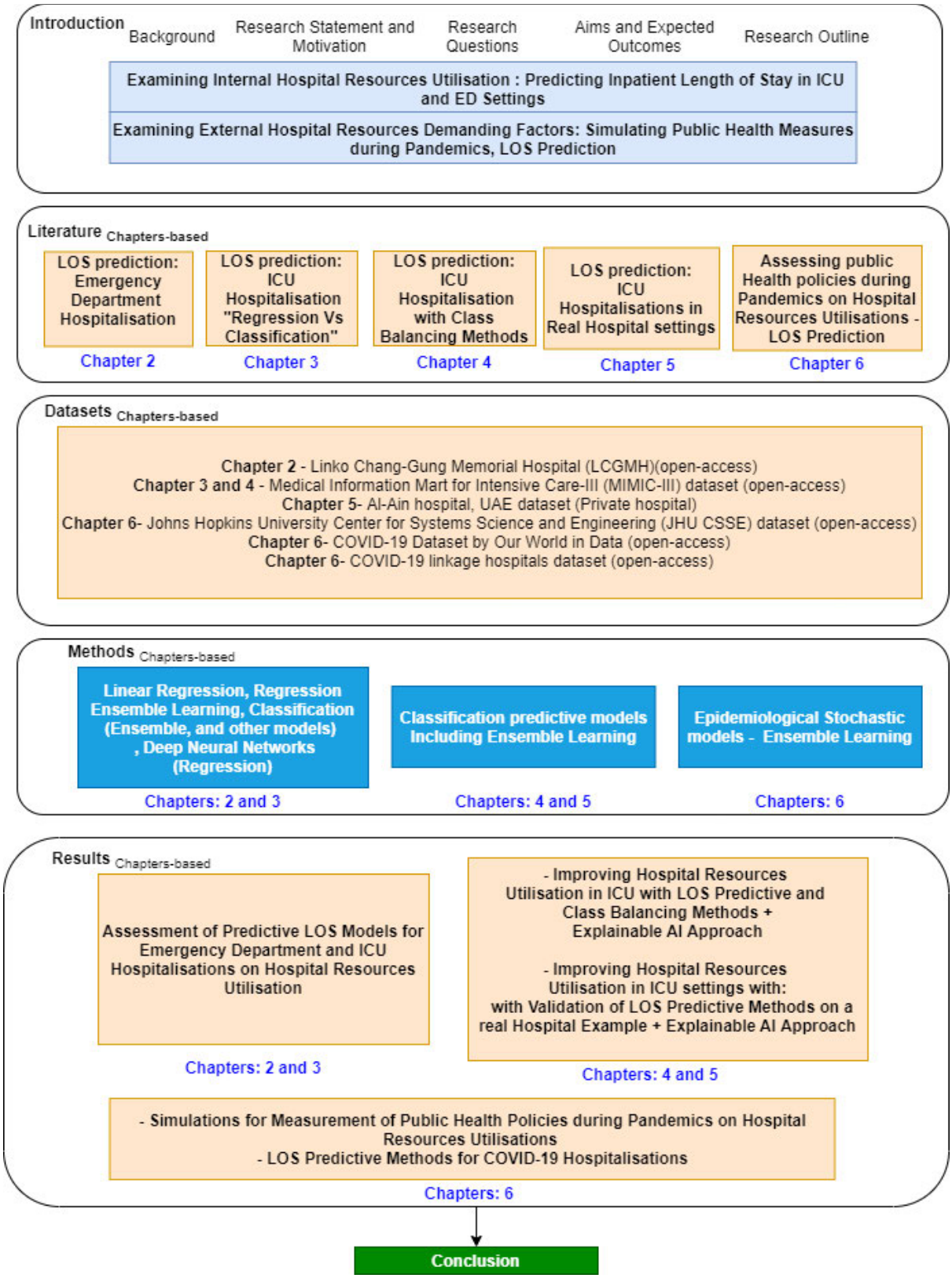


Figure 1. Flow Chart of the Thesis Structure

1.8 Ethics Thesis Statement

All datasets used in this research, including the open-source and the real hospital dataset, are approved via the ethics Western Sydney University approval number H15311. Ethics application approval and ethics amendments are available in the Appendix of this chapter.

Appendix

WSU-REDI and HREC ethics approval - 2019

WESTERN SYDNEY
UNIVERSITY

HUMAN RESEARCH ETHICS COMMITTEE

9 October 2019
Doctor Omar Mubin
School of Computing, Engineering and Mathematics

Dear Omar,

Project Title: "Predictive Analytics Framework for Electronic Health Records with Machine learning advancements"

HREC Approval Number: H13511
Risk Rating: Negligible

I am pleased to advise the above research project meets the requirements of the National Statement on Ethical Conduct in Human Research 2007 (Updated 2018).

Ethical approval for this project has been granted by the Western Sydney University Human Research Ethics Committee. This HREC is constituted and operates in accordance with the National Statement on Ethical Conduct in Human Research 2007 (Updated 2018).

Approval of this project is valid from 9 October 2019 until 9 October 2021.

This protocol covers the following researchers:

Omar Mubin, Belal Alsinglawi, Mauricio Novoa

Summary of Conditions of Approval
1. A progress report will be due annually on the anniversary of the approval date.

2. A final report will be due at the expiration of the approval period.

3. Any amendments to the project must be approved by the Human Research Ethics Committee prior to being implemented. Amendments must be requested using the HREC Amendment Request Form.

4. Any serious or unexpected adverse events on participants must be reported to the Human Research Ethics Committee via the Human Ethics Officer as a matter of priority.

5. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the Committee as a matter of priority.

6. Consent forms are to be retained within the archives of the School or Research Institute and made available to the Committee upon request.

7. Approval is only valid while you hold a position or are enrolled at Western Sydney University. You will need to transfer your project or seek fresh ethics approval from your new institution if you leave Western Sydney University.

8. Project specific conditions:
There are no specific conditions applicable.

Please quote the registration number and title as indicated above in the subject line on all future correspondence related to this project. All correspondence should be sent to humanethics@westernsydney.edu.au as this email address is closely monitored.

Yours sincerely

[redacted]

Professor Elizabeth Deane
Presiding Member,
Western Sydney University Human Research Ethics Committee

WSU-REDI and HREC ethics Amedmends approval – 2021 for using Al-Ain Hospital dataset

WESTERN SYDNEY
UNIVERSITY

Ethics Reference: H13511
Expiry Date: 9 October 2021

HUMAN RESEARCH ETHICS COMMITTEE

10 March 2021

Doctor Omar Mubin
School of Computer, Data and Mathematical Sciences

Dear Omar,

RE: Amendment Request to H13511

I wish to formally advise you that the Human Research Ethics Committee has approved your request to amend your approved research protocol H13511 "Predictive Analytics Framework for Electronic Health Records with Machine learning advancements".

The approved amendments are:

Add a new dataset from Al-Ain Hospital, Al Ain, UAE and College of Information Technology (CIT), United Arab Emirates University, Al Ain, UAE.

Project specific approval conditions:

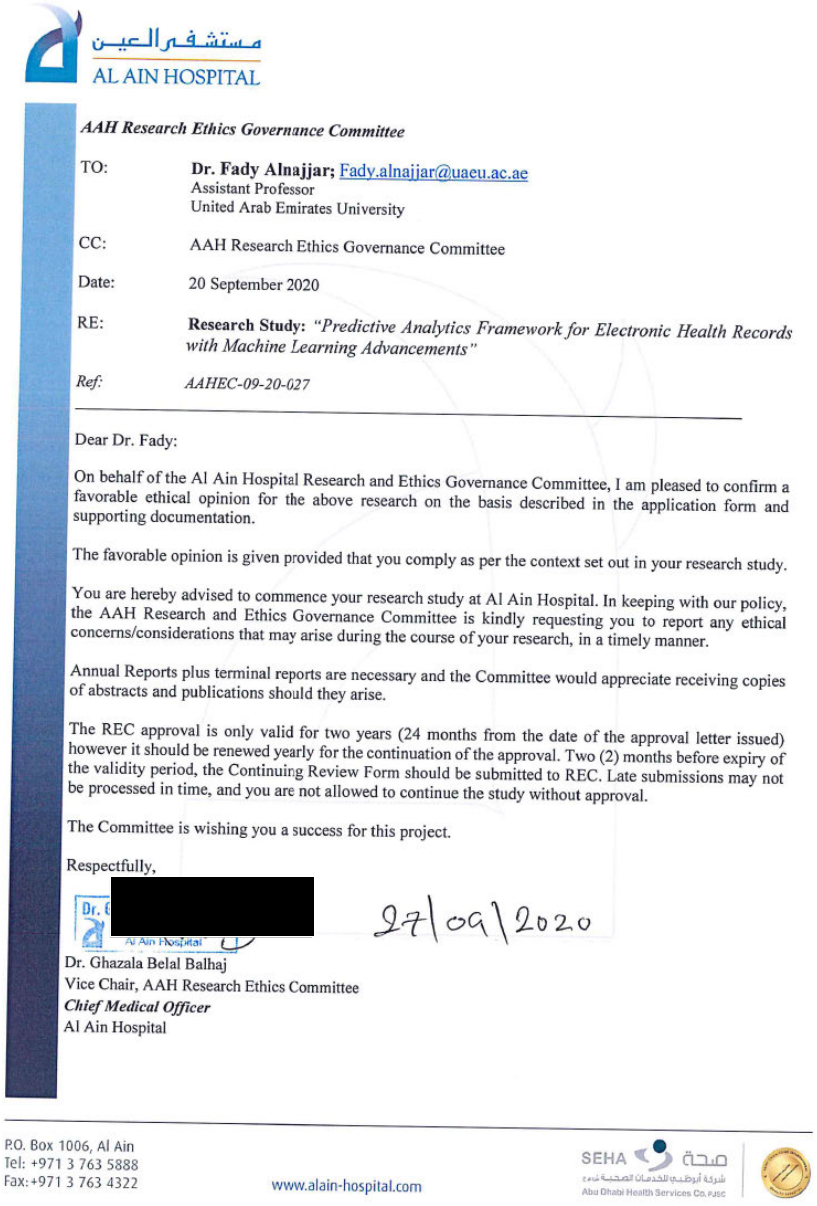
Please quote the registration number and title as indicated above in the subject line on all future correspondence related to this project. All correspondence should be sent to humanethics@westernsydney.edu.au as this email address is closely monitored.

Regards

[redacted]

Dr Geir Henning Presterudstuen
Presiding Member,
Western Sydney University Human Research Ethics Committee

Al-Ain ethics approval number: AAHREC-09-20-027 via UAEU



مستشفى العين
AL AIN HOSPITAL

AAH Research Ethics Governance Committee

TO: **Dr. Fady Alnajjar;** Fady.alnajjar@uaeu.ac.ae
Assistant Professor
United Arab Emirates University

CC: AAH Research Ethics Governance Committee

Date: 20 September 2020

RE: **Research Study:** *"Predictive Analytics Framework for Electronic Health Records with Machine Learning Advancements"*

Ref: *AAHEC-09-20-027*

Dear Dr. Fady:

On behalf of the Al Ain Hospital Research and Ethics Governance Committee, I am pleased to confirm a favorable ethical opinion for the above research on the basis described in the application form and supporting documentation.

The favorable opinion is given provided that you comply as per the context set out in your research study.

You are hereby advised to commence your research study at Al Ain Hospital. In keeping with our policy, the AAH Research and Ethics Governance Committee is kindly requesting you to report any ethical concerns/considerations that may arise during the course of your research, in a timely manner.

Annual Reports plus terminal reports are necessary and the Committee would appreciate receiving copies of abstracts and publications should they arise.

The REC approval is only valid for two years (24 months from the date of the approval letter issued) however it should be renewed yearly for the continuation of the approval. Two (2) months before expiry of the validity period, the Continuing Review Form should be submitted to REC. Late submissions may not be processed in time, and you are not allowed to continue the study without approval.

The Committee is wishing you a success for this project.

Respectfully,

[redacted] 27/09/2020

Dr. Ghazala Belal Balhaj
Vice Chair, AAH Research Ethics Committee
Chief Medical Officer
Al Ain Hospital

P.O. Box 1006, Al Ain
Tel: +971 3 763 5888
Fax:+971 3 763 4322

www.alain-hospital.com

SEHA صحة
شركة أبوظبي للخدمات الصحية ش.م.ع
Abu Dhabi Health Services Co. PJSC

All datasets in the thesis are approved via the ethics approval (Amedmends of H13511 approval on 17 Aug- 2021)

WESTERN SYDNEY
UNIVERSITY

Ethics Reference: H13511
Expiry Date: 9 October 2021

HUMAN RESEARCH ETHICS COMMITTEE

18 August 2021

Doctor Omar Mubin
School of Computer, Data and Mathematical Sciences

Dear Omar,

RE: Amendment Request to H13511

I wish to formally advise you that the Human Research Ethics Committee has approved your request to amend your approved research protocol H13511 "Predictive Analytics Framework for Electronic Health Records with Machine learning advancements".

The approved amendments are:

Add additional open-source datasets from FigShare (i.e., electronic health records (emergency department) dataset), GitHub repository (i.e., Our world in data, Johns Hopkins University Center for Systems Science and Engineering, and Data Science for COVID-19 dataset), and Kaggle (via Google Datasets; i.e., COVID-19 Hospitals Treatment Plan).

Project specific approval conditions:

Please quote the registration number and title as indicated above in the subject line on all future correspondence related to this project. All correspondence should be sent to humanethics@westernsydney.edu.au as this email address is closely monitored.

Regards

[redacted]

Associate Professor Gabrielle Weidemann
Presiding Member,
Western Sydney University Human Research Ethics Committee

2. Chapter Two- Hospital Resources Hospitalization Assessment with Machine Learning: Predicting Length of Stay for Emergency Admissions.

2.1 Chapter Summary

The chapter aims to assess the length of stay prediction in ED hospitalisation. A predictive methodological guide was constructed to capture the state of emergency admission and predicted LOS using the machine algorithms. A comparison between the classification model was conducted to evaluate the suitable approach in the context of emergency-based admission. The research ED-LOS architecture utilised predictive algorithms to forecast the LOS emergency department for the most appropriate modelling techniques. Also, the architecture considered the machine learning models explainer approach. The explainer method explains the black box of predictive methods and makes the predictive outcomes readable and understandable to the non-machine learning expertise. An experimental methodology evaluates the architecture, and the procedure uses an open sources dataset to perform practical experiments. In due course, this enables to baseline each prediction ED based on hospitalised length of stay as well as helps to improve the emergency department care delivery and supports the medical decision for the medical professional and healthcare researchers in ED settings. The chapter serves as an exploratory that initiating the research to explore different predictive machine learning techniques to guide the research analysis for the remaining part of the thesis. Therefore, this chapter starts with classification approaches, and it is not disease-focused.

2.2 Introduction

ED crowding has been described as a patient safety issue and a global public health problem [37]. The majority of patients visiting emergency departments could face long waiting times because of the crowdedness in public hospitals in Australia. Multiple factors increase the waiting times in ED. One of these factors is called ‘access block’ [38]. The access block happens when the patients are admitted to the emergency department and need an emergency room bed that delays inpatients leaving the ED due to the lack of bed availability and capacity in emergency rooms [39]. In 2017-

2018, eight million patients presented to Australian public hospitals' ED with an increase of 3.4% than the previous year [39]. In the United States, the medical-related ED visits resulting in admission attributed to 49.5 per 1,000 population in 2016 [40].

Furthermore, the waiting time for emergency department care (the proportion of presentations completed within 4 hours) increased from 67% to 72%, according to the access block indicator [5]. When the patients arrive at the emergency department, they are assigned to the triage category. According to the report, AIHW [39], 77% of inpatients admitted to an emergency department in Australia were assessed as urgent (Triage 4) or semi-urgent (Triage 4). Many factors extend or reduce waiting time during the emergency department visits, such as the availability of beds, nurses' staff, and medical staff. Also, there are other medical-related factors such as delays in diagnostic imaging and test results [41]. In addition, resource utilisation is critical for hospital resource utilisation, especially in emergency departments, to ensure the effectiveness of emergency care. Therefore, identifying the right length of stay assessment and procedures in emergency care systems (clinical information systems) reduces the length of stay at the emergency department and reduces waiting time for inpatients admission to the ED room. To this end, reducing patients' length of stay at the emergency department is deemed crucial to improve the quality of medical and health care services in ED rooms in hospitals. Overall, reducing ED length of stay at the hospitals is a challenging global task.

The number of ER visits is growing every year, and these numbers are creating pressure on ED facilities globally, especially in public hospitals. Crowding at ED is common and associated with increased costs and adverse patient outcomes [42]. The average length of stay in ED settings is increasing every year on a global scale; therefore, the extended waiting time for patients has a negative impact on patients and may reject new patients in the overabundance of patients in ED rooms. As a result, handling beds availability and identifying and managing the length of stay in ED creates persistent needs for the physicians, nurses, beds managers, ED hospital management, and clinical decision-makers in hospitals. Many attempts in the literature review the state-of-art prediction algorithms to achieve the performance in predicting the length of stay at hospitals [18]. However, there is no comprehensive research attempt to provide a research framework (guide) [18] for predicting the length of stay at ED using the advancement of machine learning prediction algorithms. The lack of research attempts appears in the fewer research studies that examined

machine learning in clinical information systems to determine the likelihood of patients needing to stay during ED admission. Therefore, building a robust LOS predictive approach is important for improving beds occupancy and managing hospital resources and ED staffing. Moreover, in EHRs, modelling LOS plays a significant role in exploiting the prediction method in a particular ED LOS prediction task.

Consequently, studying the predictive benchmarking approaches to examine the relevant prediction modelling methods with the robust prediction performance to predict the ED length of remains a research challenge. Developing a methodological research guide that considers the appropriate data category and data representation is vital in predicting ED length of stay. Therefore, establishing the appropriate ED LOS architecture to facilitate utilising the machine learning models to baseline their ED length of stay prediction tasks is deemed imperative research need for the clinical information systems. The predictive research guide can guide the development of a software tool in the future. The tool will be useful for ED beds managers, ED medical professionals and hospital researchers in the EHRs field to assess and facilitate the predictive tasks for more related research problems associated with LOS with ED, eventually providing a better support decision making in the ED context. Further, it will benefit the patients in the hospital and reduce waiting time at the ED hospital admissions by providing a feasible predictive health system in various ED admission types. Indeed, this will help regulate the EHRs, personalised healthcare recommendations, and standardised LOS ED prediction tasks in healthcare and clinical settings globally.

2.3 Literature Review

This section will discuss the most recent works from the literature, which have applied predictive modelling to forecast the length of stay of the most common prediction medical tasks at the emergency department. The literature review in this section emphasises the use of the common machine learning models to predict LOS in ED/ER. In addition, the literature review covers the research trend related to the focus of this thesis, particularly hospital length of stays that considered the various types of emergency admissions from a data-driven approach. Both factors were being taken into consideration: the short length of stay and prolonged health of stay. Therefore, all hospital admissions unrelated to the LOS emergency prediction focus were excluded in this chapter.

Predicting ED length of stay elucidates the time interval between ED hospital admission and discharge time from the ED. In other words, it means predicting the total admission time to stay at the hospital for a patient. Forecasting length of stay (predicting LOS) is crucial in the emergency department's various predictions tasks and general hospital admission. It is considered an important metric to assess the quality of healthcare and planning capacity at the hospital. Also, it is identified as the key performance for monitoring hospital care quality and managing patients' expectations [4]. Furthermore, determining the cost and post-hospital admission, which can decide the medical staffing cost and resources management [12], is vital in health clinical support decision systems. Further, failing to identify non-emergent admissions can increase healthcare expenditures and will eventually surge high-cost patients [43] during hospital admission.

Research projects [44, 45] have endeavoured to address LOS prediction in the emergency department. Related types of hospital admission to the ED considering LOS as a prediction problem using regression analysis models. Other research [46-48] evaluated the classification models in the LOS prediction problems. These models were examined in various clinical settings to forecast length stay tasks and find the best predictive technique in common medical cases at an emergency department. Each predictive medical case has different affecting factors that may influence the average length of stay (short LOS, medium LOS, and long LOS or prolonged) in the emergency type of admission.

The literature review studies interested to examine the LOS for asthma and chest related medical admissions to ED. For instance, Barnes et al. [49] proposed an approach to support automated real-time demand capacity management with decision tree learning to predict patients' discharge prioritisation (likelihood of discharge by 2 pm and midnight each day for inpatient length of stay). Their study compared the random regression forest (RRF) performance against logistic regression in the clinical information system to predict 2 pm and midnight discharges. Their approach showed that the model had significant sensitivity ($P < 0.01$), and lower specificity ($P < 0.01$). However, the model did not perform well in the early discharge task. Per the study, clinicians forecasting outperformed the authors' studied model to predict near future and aggregate prediction metrics-based tasks. Levin et al. [50] proposed an ED predictive model called e-triage using the Random Forest (RF) algorithm to resolve the problem of the differentiation of the heterogeneous ESI level 3 patients. The problem appears in ESI, where the ESI scoring system cannot classify all patients

to the correct ESI level 3 ED admissions (in chest pain admission, shortness of breath). The reported results indicated that e-triage accurately classified patents at ESI level 3 with AUC 0.73 to 0.92. However, the study did not treat some ED – EHRs data errors, which could lead to lower accuracy in the prediction model. Also, the e-triage did not consider to validate the tool (e-triage) performance.

Studies that evaluated the length of stay at ED by focusing on vehicles accidents, major poisoning, comorbidities and infectious diseases related admissions include but are not limited to sepsis, pediatric admission and critical care.

Zhang et al. [51] implemented logistic regression (LR) and multilayer neural network models (MLNN), which incorporated natural language processing (NLP) for free text variables to predict hospital admissions (injures, poisoning, comorbidities) or transferred to the emergency department. The proposed approach included structured variables achieved AUC = 0.824, 95% CI 0.818-0.830 for logistic regression and AUC = 0.823, 95% CI 0.817-0.829 for MLNN. In the case of free-text information generated, the model achieved AUC = 0.742, 95% CI 0.731-0.753 for logistic regression and AUC = 0.823, 95% CI 0.817-0.829 for MLNN. In the case of free AUC = 0.753, (95% CI 0.742-0.764) for MLNN. In the case of the combined approach, for both models (structured variables and free text variables), the results were AUC = 0.846, 95% CI 0.839-0.853 for logistic regression and AUC = 0.844 95% CI 0.836-0.852 for MLNN. However, the study did not treat missing values in the structured variable model, and the model ignored the order of the words in the text analysis, which resulted in a potential source of bias in the predictions. Work by [52] examined the use of ANN classifier to predict the average length of stay at the emergency department. Categorical variables such as patient age, sex, treatment unit and mode of arrival were used in the study. The results showed that the model was able to predict LOS with an average accuracy of 80%; however, the model could not achieve higher accuracy due to inaccurate input variables.

Taylor et al. [53] applied a big data-driven machine learning method (random forest, LR, CART, and K-means clustering in data pre-processing) to evaluate the performance of the ML in Big data prediction tasks in-hospital mortality of ED patients with sepsis. The results showed that the machine learning approaches in particular random forest outperformed the existing clinical decision rules (CDRs), where RF achieved AUC: 0.86, 95% CI = 0.82 to 0.90 as the best results

compared to CURB-65 (CDRs) score with AUC: 0.73, 95% CI = 0.67–0.80). McCoy et al. [54] explored the machine learning algodiagnostic (MLA) potential to predict the length of stay of sepsis patients and improve sepsis management and patients outcomes. The goal of the study was to identify patients with sepsis earlier. The results indicated that the MLA algorithm improved hospital length of stay from 3.35 days to 2.95 days with a 4.8 % reduction in hospital LOS. However, the study did not measure the effectiveness of the proposed approach in a scalable medical system. Also, the study did not consider the appropriate disease coding schema such as the International Classification of Diseases (ICD) coding for sepsis severity. Mao et al. [55] evaluated the performance of machine learning algorithms (InSight, gradient tree boosting) on two datasets (University of California, San Francisco- UCSF and MIMIC-III) to predict the length of stay for sepsis patients at the emergency department. LOS was binned in labels with (0-2, 3-5, 6-8, 9-11,12+) days, including ICU stay. Some ICD codes were considered in the study (sepsis, serve sepsis, septic shock). The reported result showed that the XGBoot model predicted sepsis and severe sepsis with an AUROC curve of 0.92 (95% CI 0.90 to 0.93) and 0.87 (95% CI 0.86 to 0.88), respectively, four hours before onset. Also, their model achieved septic shock prediction with an AUROC of 0.96 (95% CI 0.94 to 0.98). These results showed that their model could work effectively with vital signs input measurements in ED admitted sepsis and missing EHRs data. However, the authors did not discuss the type of the predicted model and did not provide more details on the technical side of their model. However, the authors did not provide more details about the machine learning algorithms were applied in their study.

Sundén-Cullberg et al. [56] implemented logistic regression on four categories of body temperature for patients with fever who were admitted to the emergency department. The main object of the study was to examine the prognostic value of fever in the emergency department in septic patients who are then subsequently admitted to the ICU. Statistical tests such as Chi-square, Wilcoxon signed-rank, and Kruskal-Wallis were used to evaluate the performance of the predictive algorithm. However, the study did not treat all missing values in the records or the missing values of attribute body temperature.

Vermeulen et al. [57] implemented Modified Poisson Regression models to analyse the association of overall improvement in ED LOS for patients with acute myocardial infarction (AMI), asthma and pediatric and adult upper limb fractures. The reported results showed that median LOS at the

ED improved by up to 26% (63 min) in the improved hospitals and deteriorated by up to 47% (91 min) in the unimproved sites. However, the predictive approach did not examine the crowd at the emergency department, which may negatively influence the length of stay at the hospital.

Work by Hosseininejad et al. [45] evaluated the prolonged emergency department length of stay for 10 medical ED cases / ED admission, including Gastroenterology, Respiratory, Neurosurgery, Cardiovascular using multivariate binary logistic regression. The method was able to identify 10.2% of the total number of the study sample size (1581 patients) with prolonged LOS in the ED with the ability to predict release after 6 hours from admission. However, their approach did not follow up with patients admitted to the ED to determine the outcome of patients with prolonged LOS. Also, it did not examine factors and causes of disposition after 6 hours from admission, where it remained the unaddressed problem. Another work by Azari et al. [58] applied logistic regression and imbalance learning methods to predict the prolonged length of stay (> 14 hours) for various ED admissions, including chest pain and abdominal pain. The results found that a length of stay of more than 14 hours had 10% of the ED visits. However, further data training is needed to deal with the noisy ED data. Hong et al. [59] evaluated the performance of three machine learning models (R, q XGBoost, and DNN) to predict patient disposition by using the emergency department information. Three datasets were studied and examined: 1) triage information, 2) patient history, 3) using a complete set of variables. The researchers included 972 variables from rich EHRs datasets containing the main medical diagnosis (heart disease or diabetes-related variables). The results showed that XGBoost had the best prediction results in the complete set of the variable dataset (AUC: 0.924 (CI 95%: 0.922–0.927)). However, the study reported that they did not evaluate the appropriateness of individual clinical decisions where further studies are needed for a better-standardised metric for hospital admissions.

Combes et al. [60] compared eight linear regressions (Logistic Model Tree (LMT), Multi-class alternating decision tree using the LogitBoost strategy (LADTree), Decision tree (C4.5 - J48), Decision tree with naive Bayes classifiers at the leaves (NBTree), Random Forest (RF), Decision/regression tree using information gain/variance and pruning it using reduced-error pruning (REPTree), Multilayer Perceptron (MP), and SVM) to predict LOS in PED. The overall results indicated that the linear regression modules were able to predict 75% of the cases (variables) with an error of ± 2 hours. However, the authors stated that there are problems

associated with basic linear regression methods that cannot discover non-linear relationships between other variables in their study. Barak-Corren et al. [61] applied Logistic Regression and Naive Bayes Classifiers on EMR pediatric data to predict hospitalisation after admission and evaluate the impact of the models on overall pediatric ED resource utilisation. The models identified 73.4% of the hospitalisations with 90% accuracy.

Some research did not specify the medical diagnosis, such as Chaou et al. [62], who used a multivariate accelerated failure time to analyse the influence factors of the collected covariates on the patient LOS. The results (statistical significance; $p = 0.649$) indicated that the patients with higher acuity (triage level I vs. level V) stayed longer in the emergency department. Further, the type of diagnostic activities which were given in the ED (e.g. observations, laboratory tests, etc.) had the highest impact on the emergency department LOS. However, the study ignored some of the possible correlates in the model, such as important medical diagnoses.

Work by Graham et al. [63] studied the crowding prediction problem. It exploited the potential of machine learning techniques (Logistic regression, decision trees, and gradient boosted machines "GBM") to predict ED admissions, improve patient flow, and prevent overcrowding at the emergency department in two major hospitals in Northern Ireland. The results showed that the GBM performed better than any model with an accuracy of 80.31%, AUC-ROC = 0.859 in comparison with 80.06%, AUC-ROC = 0.824 for decision tree and 79.94%, AUC-ROC = 0.849. However, the study did not investigate the importance of significant variables from patient EHRs data to improve accuracies, such as heart rate and pre-existing conditions. Gligorijevic et al. [64] developed a deep text model based on the deep learning algorithm bi-directional recurrent neural network (bi- RNN). The model was fed with nominal (structured data) with medical text (unstructured data). The model achieved an AUC of 88% as the best result compared to models (LR with AUS = 54.91%, MLP with AUC = 56.13%)

Work by Davood [65] investigated the problem of predicting patients' admission in emergency operations admission cases using demographic and clinical information using the minimum predictor variables with the use of machine learning model logistics such as Regression and Artificial Neural Networks. Both LR and ANN models showed similar results with 82.5% and 83.0% for a trained dataset, respectively, and 82.0% and 82.1% for the tested dataset. However,

the study did not investigate significant factors that could improve the proposed model's accuracy, such as the unclassified structure of clinical text information.

Lit et al. [44] implemented a logistic regression module and generalised linear models to evaluate the benefit of adding automated laboratory and comorbidity measures at the point of admission (emergency department/ nonemergency admission). Two score systems were used (LAPS and COPS) in the Linear regression modules. However, the study indicated that the source of information was limited to only a single healthcare dataset which undermines the maximum benefits of adopting such a prediction system.

Table 2 represents a review of selected related works in the literature review where LOS Application refers to the domain of the study area in the ED length of stay in hospital admission. Dataset feature represents the nature of the study, whether it has been carried on the private dataset (local hospital) or public datasets (openly available for public access). The sample study shows the number and the volume of participant elements in each study. The prediction error and scoring system column is the appropriate measurement unit that was examined in each study. The medical condition column indicates the nature of the clinical and medical problem that the study addressed. The prediction algorithm refers to the predictive modellings which were used in the study. The model’s performance baselining on the public datasets column indicates whether the authors evaluated their model on public datasets. Finally, the results column shows the main findings of the study.

Table 2. Represents a summary of LOS – ED Hospital admission related works

Work	LOS Application	Dataset (Private[1] or Public)	Study Sample	Prediction Error OR Scoring System	Medical Condition (Problem)	Prediction Algorithm	Models Performance Baselining on Public Dataset(s)	Results
Barak-Corren et al. [61], 2017	Pediatric Emergency Department	Private	59,033 patient visits (11,975 hospitalized, 47,058 discharged)	AUC	Pediatric Emergency Admission	**Classification** + **Regression** (Logistic Regression + Naive Bayes Classifier)	x	Identified 73.4% of the hospitalizations with 90% specificity and (AUC = 0.91)
Lit et al. [44], 2010	Emergency Department	N/A	155,474 unique patients (61 ± 19 years)	LAPS COPS	Medical Diagnosis not Specified	**Supervised Multiple Models** (Multiple Regression & Multiple Classification)	x	- 4.5 ± 7.7 days (Mean LOS) - 2.8 days (median LOS)
Hosseininejad et al. [45], 2017	Emergency Department	Private	1581 patients	ESI	Multiple Medical Cases Admissions	**Regression** (Multivariate Binary Logistic Regression)	x	- 6 hours (disposition) - Only 10.2% of prolonged cases were detected
Vermeulen et al. [57], 2015	Emergency Department	Private	N/A	Canadian Triage Scoring system	- Patients with acute myocardial infarction (AMI) - Asthma and paediatric	**Regression** (Poisson Regression Models)	x	ED LOS (Median 63 min)

[1] Private: unavalaible or inaccessible

					- Adult upper limb fractures			
Azari et al. [58], 2015	Emergency Department)	N/A	N/A	F1-measure	Multiple Medical Cases Admissions	**Regression** (Logistic regression)+Imbalance learning methods	x	Predicted >14 Hours (Precooled LOS)
Gul and Guneri [52], 2015	Emergency Department)	Private	1500 ED patients	Average classification error	Vehicles Accidents, injures (Major)	**Classification** (ANN)	x	The average accuracy of 80%

After evaluating the previous works from the literature. Table 2 summarises the literature review comparison based on the chapter focus. Therefore, the following findings are derived and discussed based on the ED literature review.

Benchmarking predictive models in the ED LOS context in the predictive tasks via electronic health data appears like a difficult target. The lack of publicly available datasets creates transparency and reproducibility of reported results [18]. The ED LOS literature review comparison (Table 2) showed that far fewer works benchmarked length of stay from publicly available datasets. Most of the studies used their private datasets which are obtained from either local hospitals in the metropolitan areas or regional areas or even from institutional hospital admissions datasets and, hence, are smaller in size. Generally, private medical institutions do not tend to share or make their patients' data openly available due to the nature of patients' data sensitivity related concerns [66]. However, recent trends in the institutions' policies lead to a lack of universal reference agreed upon benchmarking for the ED length of stay in clinical settings.

Consequently, most of the evaluated studies on private datasets that used predictive algorithms are not justified by external parties (specialists in domain knowledge such as researchers, medical practitioners such as doctors, Nurses, etc.). Therefore, there is a lack of adequate research attempts to baseline the best predictive algorithms to predict the length of stay from hospital ED data. For this purpose, there is a need to do more research on open datasets to verify the effectiveness of predictive modellings on emergency admissions LOS, which will help evaluate the performance of these predictive analytics techniques in various EHRs inputs. Therefore, a better understanding is needed of ED LOS settings from a data-driven approach to support the adoption of clinical information predictive systems. Eventually, it is important to achieve ED-LOS predictive research guide with the most appropriate health analytics techniques to suit hospital resources utilisation in emergency rooms.

2.4 Research Question

Based on the LOS-ED prediction research challenges, the following research questions will be studied in this chapter. The research questions are an integral part of the thesis research questions (1 and 2):

1.a Can we benchmark LOS predictive tasks in emergency admission mode?

2.a Further explains the predictive result of the outperforming model(s) in clinical information systems?

The research question will evaluate the LOS predictive tasks using the predictive machine ensemble learning models. The outperforming model will be further utilised to explain the inner workings of the machine learning model and whether these models are able to be feasible and provide insights for clinical information systems and healthcare professionals to improve resources utilisation in hospital management systems in ED settings.

2.5 Research Expected Outcome

This chapter aims to derive ED data predictions insights to support the decision making of hospital professionals via ED-LOS prediction. The proposed methods in this chapter assess which approach is more practical and explainable in LOS-ED clinical information systems. The outcomes of the predicted models aim to guide ED healthcare professionals, especially ED doctors, to which machine learning models are more feasible in practical implementation and, therefore, can be used in emergency admissions data predictive tasks. For this purpose, open-source emergency data is employed to assess the proposed architecture in this chapter. Furthermore, the research predictive guide benchmark of the models captures the state of ED and explains the inner working of the outperforming model for healthcare workers at ED, especially ED doctors. This approach served as a preliminary research investigation in this thesis and evaluated predictive models in various hospital facilities. Thus, the thesis is motivated and driven by performing models from this chapter.

2.6 Methods

2.6.1 Preliminary Study

This preliminary study attempts to initialise a pilot study in this thesis for the length of stay predictive models using ED factors. The data used in this study was collected from a de-identified dataset used in a previously published paper [62]. The dataset used in this study is shared as open access for people. The dataset describes the ED inpatient hospitalisations at the Linko Chang-Gung Memorial Hospital (LCGMH) in Taiwan. To use the dataset in the thesis, research ethics approval is associated with the use of this dataset in this thesis. The ethics approval statement is attached to the appendix of chapter one.

The case study in this chapter aims to assess the suitability of predictive models in ED and benchmark machine learning classification methods that can be applied to predict the patient's length of stay. This study uses Python programming language 3.6 version for predictions and results explainability. Also, R software was used in limited data visualisation tasks. The

predictive machine learning LOS tasks were performed with Python, and a selected desired model for prediction explainability was further analysed using the explainable artificial intelligence (xAI) methods called SHAP [67].

2.6.2 Dataset Description and Data Pre-processing

The data consists of 106206 records, as shown in Table 3. The attributes consist of three categories: patient's details, diagnosis, and emergency department status. The distribution of ED-LOS (duration) is per Figure 2.

Table 3. Emergency department dataset

label	Description	Value
consult	Consultation provided	0,1
obs	Admission/Observation provided	0,1
critical	Critical condition	0,1
triage	Taiwan Triage	(1,2,3,4)
Specialty	Medical condition(Adult non-trauma, Trauma, Pediatric non-trauma)	0,1,2
consult	Consultation provided	1,0 (yes, no)
agegp	Age group	(<20: 1 / 20-40: 2 / 40-60:3 / 60-80: 4/ >80: 5)
EKG	Electrocardiography	0,1
lab	laboratory tests	0,1
ct	Computed Tomography	0,1
xray	X-ray	0,1
transfer	Transferred patients	0,1
triage_physician	Times of triage to physician	Numerical
physician_discharge	Physician to discharge	Numerical
shi0t_triage	Shift during which the patient arrived	0,1,2 (day, eve, night)
censore	ED daily census > 95 percentile	1
results		
weekend	Weekend	0,1
crowded	Crowded	0,1
week0	Weekday(sat,sun,mon,tue,wed,thu,fri)	1,2,3,4,5,6,7

The choice of the LCGMH hospital emergency dataset is due to important reasons that facilitated the experimental part of this study. Firstly, the dataset contains important emergency department attributes necessary for decision-making in clinical settings, especially the LOS variable, measured in hours and days. Hence, the nature of ED LOS in hours and days is crucial for the successful implementation of the prediction tasks. Therefore, this LOS made the dataset a favourable choice besides the "triage" variable, which is one of the most important metrics

to assess patient cases at the time of admission. Other important factors for patients discharging from ED include “triage physician, physician discharge, transfer, etc.”. Secondly, The Emergency department dataset is an open source for researchers, given the usefulness of the data ED features that this dataset contains. Finally, it made a prefered choice because it is available online and easy to access via the provided web link for the dataset.

The dataset is used for the empirical aspect of the aims of this chapter. Therefore, it does not focus on the ED LOS regulation and the specific context of Taiwan ED practices. However, the same study provided the dataset reports an ED-LOS median of 1.46 hours for ED average LOS at LCGMH hospital. Therefore, in this case study, we refer to the typical (standard) LOS at ED with 4 hours per Figure 4 and based on the definition of ED LOS range (4 hours to 48 hours), according to the literature [68]. Thus, the LOS variable (duration) was processed and discretised into 3 labels: typical ED LOS (4 hours), 4 hours to 48 hours label, and the third label are 48 hours + label, as shown in Figure 5.

Missing values check was performed by R software, and it was observed that all data entries in the dataset were clean and did not report any missing values per Figure 3.

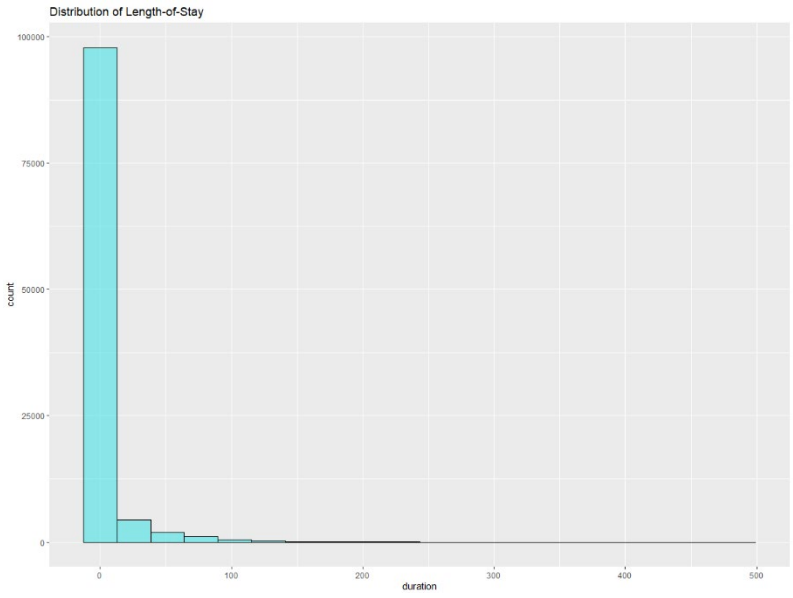


Figure 2. LOS distribution

The recommended triage in Australia [69] is different from the Taiwanese one [70]. In Australia, the triage consists of five levels, while in Taiwan, the triage consists of only four levels. However, the goal is to find the predicted correlation between duration (LOS) and other attributes, perhaps the impact of triage on LOS. Therefore, it was left for the selected predictive model to rank the predictor variables into the target (duration: LOS).

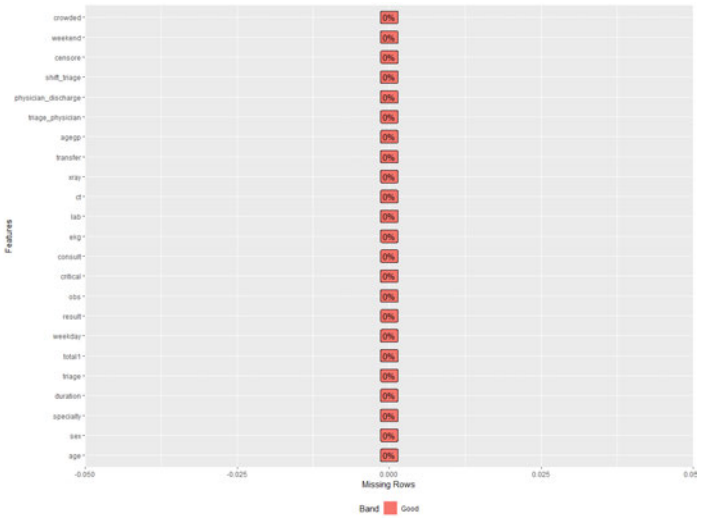


Figure 3. Missing value report for the ED dataset in this study

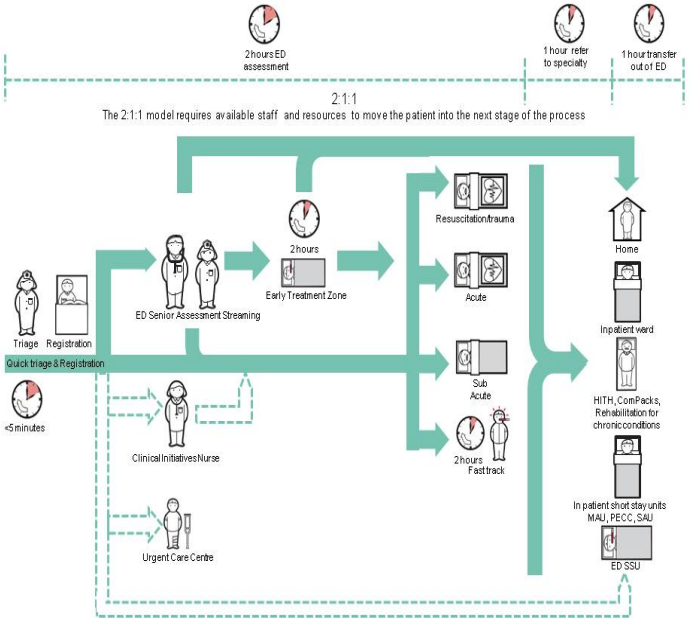


Figure 4. The ideal Emergency Department patient journey [71] in NSW, Australia.

The diagnosis in this data was determined by some medical tests and images, such as CT scans, lab tests, electro cardiology and x-ray. Unfortunately, these diagnoses are incomplete and require a specific medical description of the patient's status. Moreover, it was found that the crowd in this emergency department was less than 0.05, which is very low if compared with the overcrowded ED in other countries such as Australian hospitals [72].

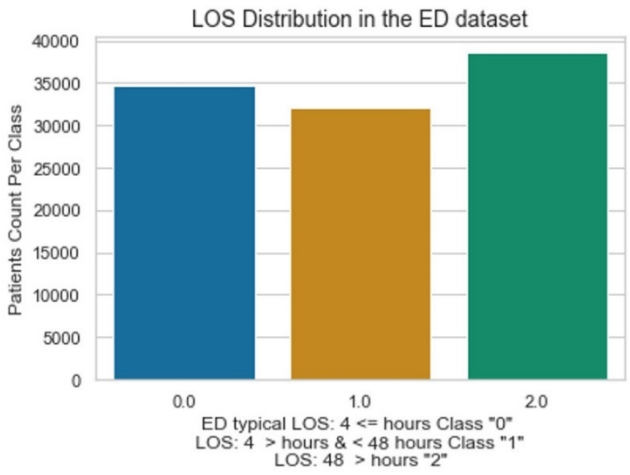


Figure 5. LOS-ED Labels

2.6.3 Predictive Analysis Results

This section discusses the steps involved in the predictive analytical stage of this chapter. First, the experimental setup used Python 3.6. Therefore, the variables selection includes all variables anticipated in the prediction analysis Table 3. Second, features transformation was applied to the dataset, resulting in a set of extracted/ transformed features for the prediction stage (Figure 6).

Prediction Results

In this preliminary study, ensemble-based learning predictive methodological architecture was utilised (Figure 7) to facilitate the mission and assist in choosing the best prediction ensemble learning model based on the variables input. The duration of the LOS variable was set as the target during the prediction process, while the other variables (triage, results, obs, critical, triage_physicia, etc.) are the independent variables.

During the transformation of the variable, categorical variables were transformed into ordinal variables (0, 1), as seen in Figure 6. In the next step, we have split data into a training set (70%) and a testing set (30%) to predict the LOS of stay at the LCGMH hospital dataset of this study per Figure 7. Three ensemble models (Random Forest “RF”, Gradient Boosting “GB”, and XGB) were utilised in this case study. The model’s descriptions are explained in Chapters 4-5 of this thesis.

According to the LOS proposed predictive architecture Figure 7, the dataset is split into training and testing sets. First, the base learners of models are trained on a subset of the dataset (Ds), and then a single prediction per model is added to the meta-model. The meta-model generates

the final predicted model (m) on the test set of the dataset. Finally, it achieves the final predicted model (P) based on each model's specific approach to reach the final predicted value of each classifier. The classifiers are evaluated using the evaluating metrics (Accuracy, Precision, Recall and F1-Score). The winning model further explains prediction results in the model's prediction explainer stage (SHAP) or (SHAP). Chapter 4 explains more details about the SHAP explainable AI method.

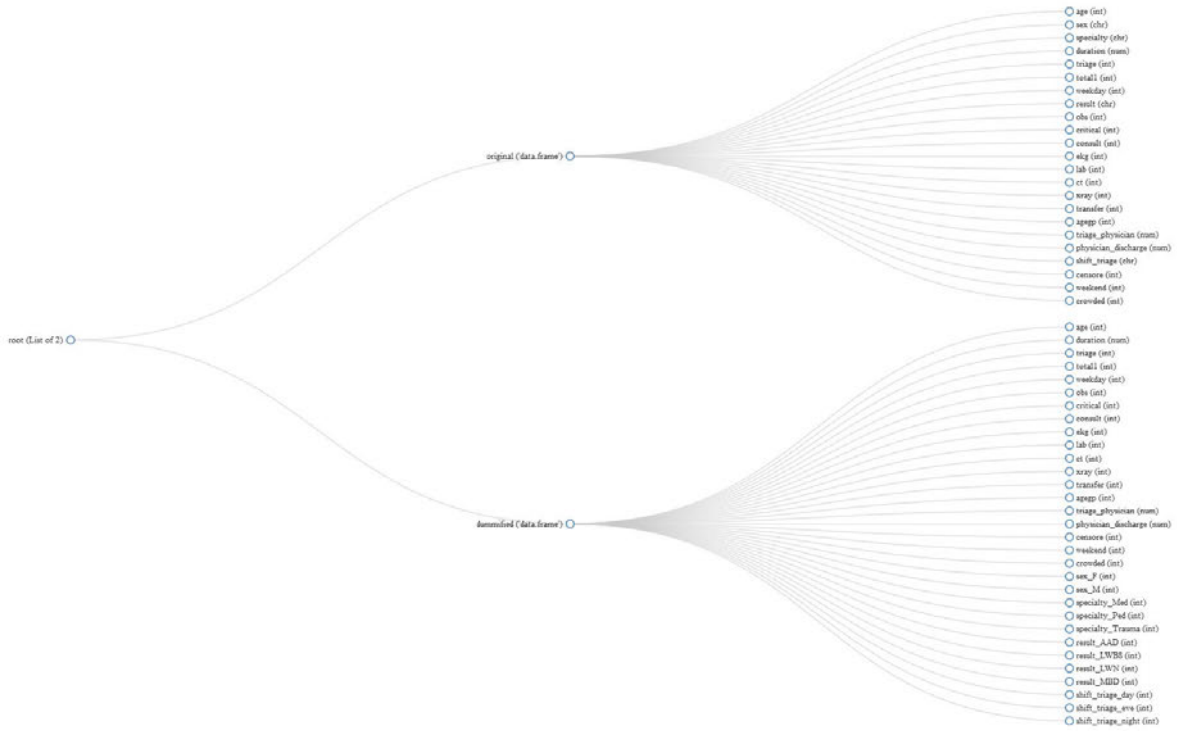


Figure 6. Features Transformation procedure in pre-processing stage with (one-hot encoding) method

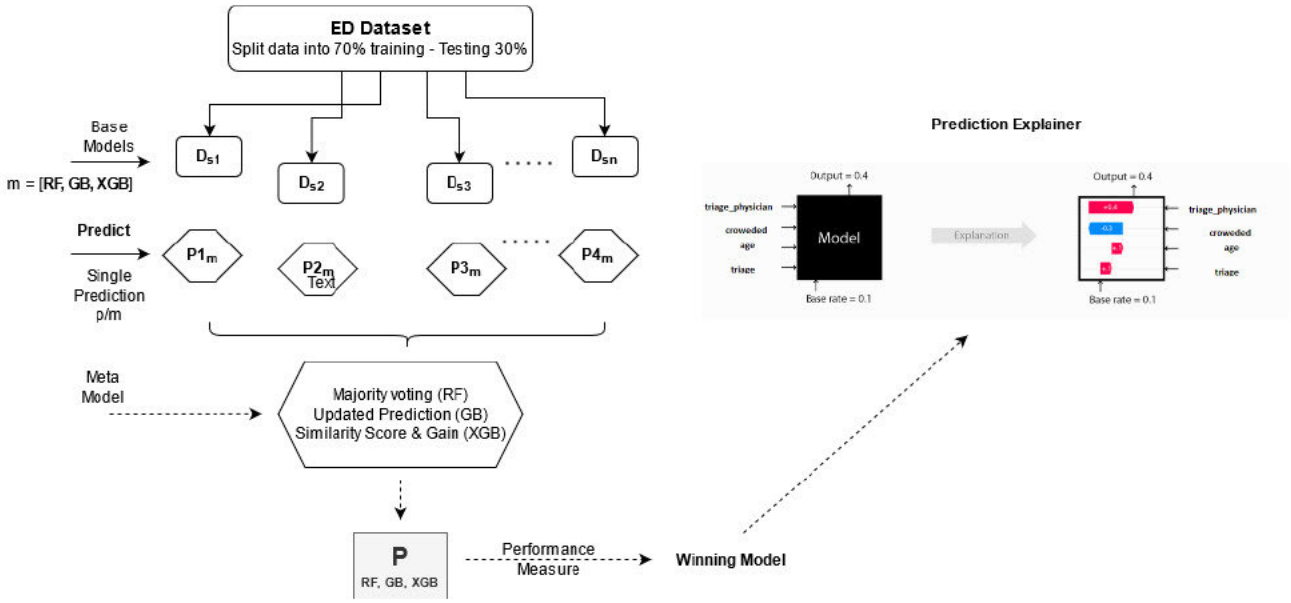


Figure 7. ED-LOS Predictive Architecture (with Prediction explainer)

The three models (RF, GB, XGB) in the case study reported high predicted outcomes. While Random forest achieved a weighted Recall of 94%, the GB and XGB achieved 98% of recall predictive values. From the prediction models' summary and results (Table 4, and Figure 8), three models achieved desired results. This is explained by the association between dependent (duration) and independent variables. Also, the data discretisation approach into three labels helped group LOS (duration variables) into a clear data projection categorisation. This approach results in high-achieving and reliable predictions. Hence, predictions in the dynamic and instant-decision environment must be safe, trusted, and robust. The confusion matrix (Figure 8) justifies the desired performance of the ensemble algorithms.

Table 4. Results of the predictive models in ED-LOS prediction architecture

Duration (LOS)	Precision			Recall			F1-Score		
	RF	GB	XGB	RF	GB	XGB	RF	GB	XGB
Label 0: < 4 hours	0.93	0.96	0.97	0.96	0.99	0.99	0.95	0.98	0.98
Label 1: > 4 hours < 48 hours	0.90	0.97	0.97	0.91	0.95	0.96	0.91	0.96	0.96
Label 2: > 48 hours	0.99	1.00	0.99	0.95	0.98	0.98	0.97	0.99	0.99
AVG Accuracy	RF: 0.94, GB: 0.98, XGB: 0.98								

Weighted Prediction	0.94	s0.98	0.98	0.94	0.98	0.98	0.94	0.98	0.98

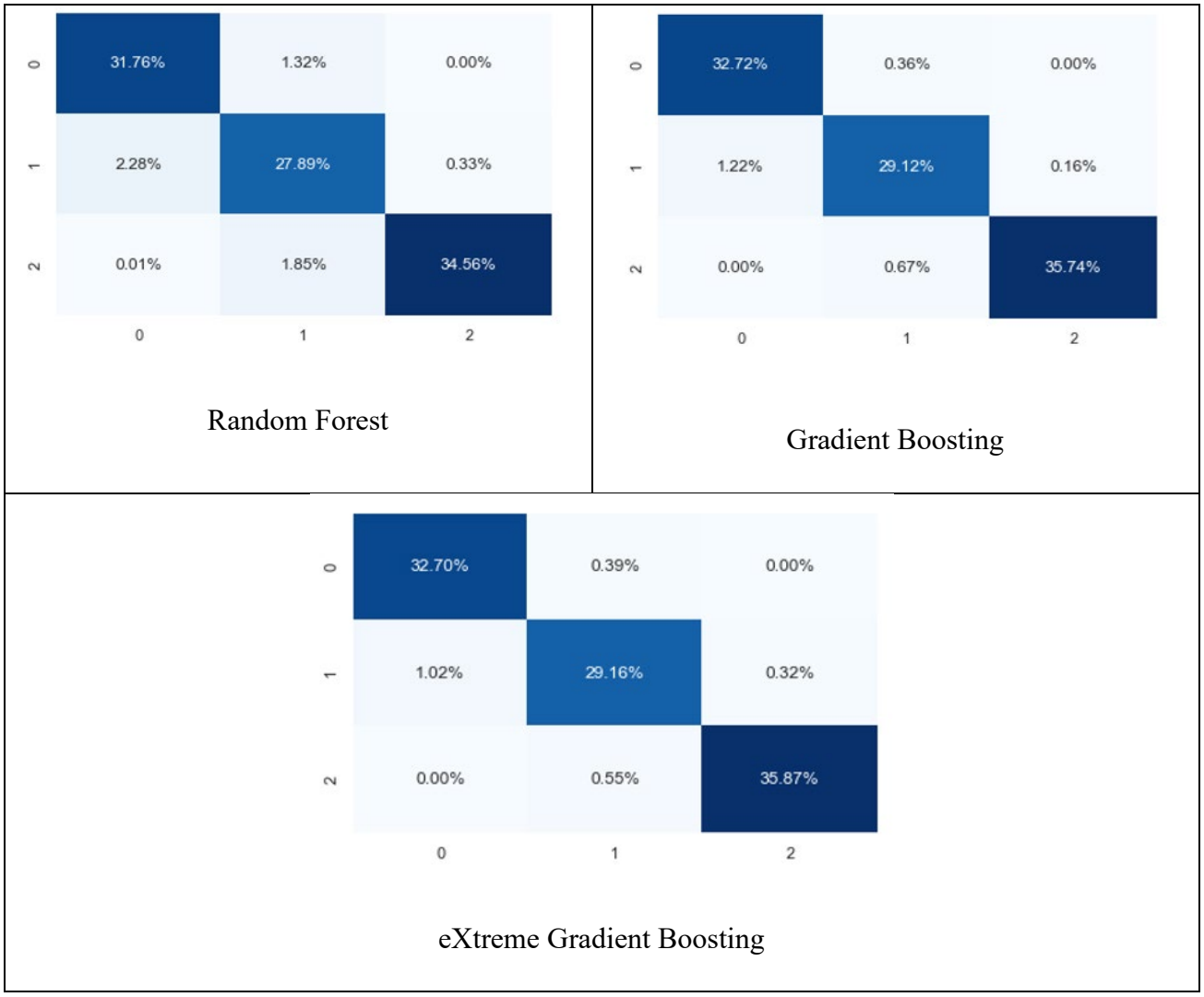


Figure 8. Multi-label confusion matrix (MCM) for the three predictive models in the pilot study

In ED settings, the prediction outcomes of prediction models must be clarified. They must be able to show that such a prediction algorithm can make a definitive decision, trustworthy, reliable, and explainable. The three models are suitable for the context of ED-LOS; therefore, explaining the inner working of the winning model is an essential task to show how predictive models are able to rank features to perform final predicted outcomes. For this purpose, we utilised a SHAP (SHapley Additive exPlanations) [67] which is a game-theoretic approach to explain the output of any machine learning model. The SHAP explains the output of predictive models, and they work best with tree ensemble methods. Therefore, the SHAP explainer was built for the winning model (XGB). The SHAP does not support GB. Hence, XGB is the choice

of the winning model to be explained further. The black box of the XGB model Figure 7 is explained in Figure 9.

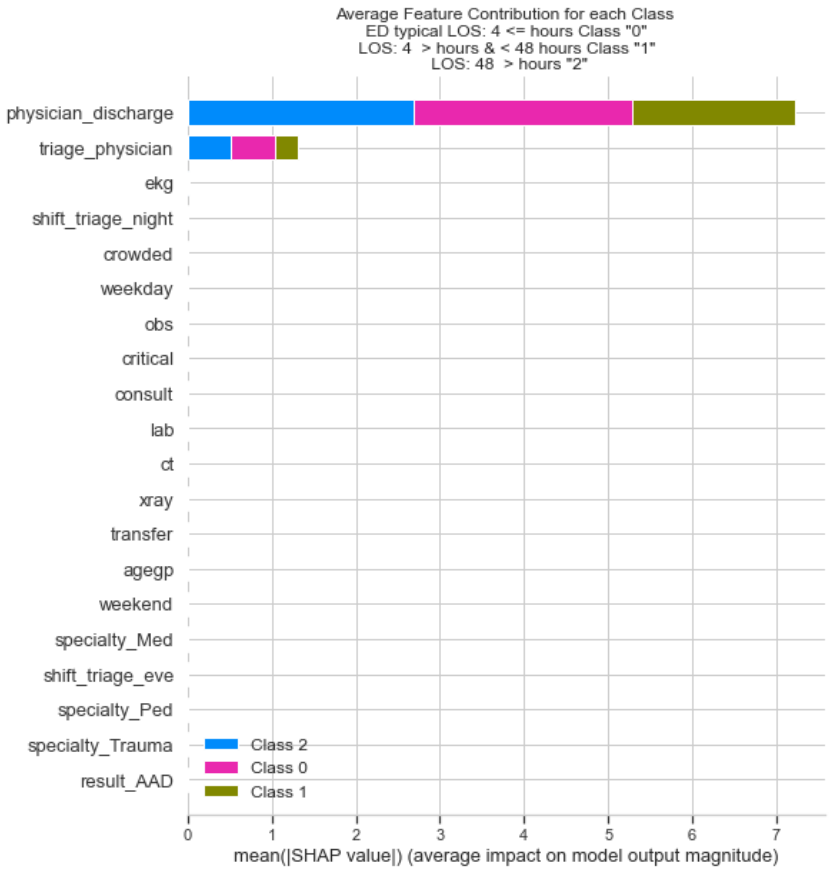


Figure 9. SHAP explainer for XGB model (ED-LOS)

It was noticed from Figure 9 that XGB relied mostly on physician discharge as the most significant variable to make class prediction decisions (0, 1 or 2). However, triage physicians also came as the second important variable for making the LOS prediction. This confirms what doctors do in real ED practices. Therefore, XGB is more suitable for clinicians making critical decisions in ED settings, which supports their decision based on the inpatient admitted case into the ED. Thus, this is one of the direct research implications of implementing the ED-LOS predictive architecture in ED settings.

2.6.4 Research Implication of ED-LOS

The proposed methodological architecture can differentiate the predicted labels with high predicted outcomes and the ability of the proposed architecture to explain the inner working of the predictive models for non-machine learning practitioners in easy, flexible and trustworthy outcomes that align with the decision in real scenarios in ED. On the other hand, the remaining attributes were negligibly ranked by XGB. This can confirm that the XGB did not consider

them as strong predictors to build the decision based on them. Eventually, making sense of predictive machine learning tasks via the explainable machine learning approach for healthcare ED assessments systems can enrich ED clinical information systems and digital healthcare patients-physicians portals to support decision-making and thus improve healthcare outcomes and minimise healthcare outcomes human errors. Hence, ED is dynamic and rapid decision-making environment; the machine learning data-driven applications (CIS) can provide great potential to support ED nurses and physicians to allow them to make a clear and better decisions when clinical insights of machine learning outcomes and xAI are integrated to them in the CIS-ED systems. This will increase the efficiency in the workplace, especially at different times of day, week, month and year and in the variation of hospitalised cases and during uncertainties such as pandemics.

2.7 Summary of the Preliminary Study and Limitations

This chapter conducted a preliminary case study on ED available dataset. The chapter constructed a predictive proof of concept predictive LOS-ED methodological architecture. The predictive models are ensemble learners, including Random Forest, Gradient Boosting, and eXtreme Grading Boosting. These prediction models showed high prediction results on the state of ED. LOS-ED methodological architecture provides an explainable AI approach that explains the inner working of the predictive results and makes the machine learning models explainable, and the prediction outcomes trusted from a decision-making perspective, especially for ED doctors. The triage factor did not give any strong correlation with the predicted variable; therefore, it did not reveal any strong predictor variable.

The experiments were studied LOS in emergency admission mode based on hourly window time (e.g. first 4 hours, 48 hours etc.). The label of the typical ED LOS (4 hours), and the labels less than 48 and the labels greater than 48 were constructed from duration (LOS label). The reported finding showed a weak ranking for other important variables such as (CT scans, shift triage etc.) between the dependent variable (duration, referred to as the LOS in the dataset) and the independents.

The limitation of this study is the fact that the usage of the proposed architecture was limited to the dataset. Therefore, it was impossible to investigate other important variables related to patient profile, such as biomarkers at the time of ED admission, clinical, medication, and many more.

The proposed research guide help identifies the suitableness of prediction models for LOS tasks for hospital resources utilisation. Furthermore, the machine learning ensemble models are proven to be suitable for LOS; therefore, they anticipate desired outcomes based on the thesis' aims and expected outcomes. The explainable AI method is able to explain the decisions and the outcomes of the prediction models. Therefore, they are considered to describe outperforming models in later chapters.

3. Chapter Three: Benchmarking Predictive Models for Hospital Resources Utilisation: Length of Stay Prediction

3.1 Chapter Background

Machine learning (ML) has shown steady growth in healthcare in recent years. However, impediments are currently slowing down its contribution to digitised healthcare in hospital settings. One of the main hindrances is the absence of universally accepted benchmarks in evaluative predictive models in the field [73]. The purpose of this study is to address challenges associated with benchmarking predictive models within clinical information systems that are an integral part of the digital health area and healthcare. Outcomes of the research will assist in setting parameters and norms to frame ML predictive models for better use in clinical information systems.

Predicting hospital LOS using data-driven models has practical advantages over conventional methods [14]. One of them is to discover patterns and find insights from electronic health data. Literature review shows that most LOS studies focus on a specific type of disease, relying on one or a limited number of ML predictive models that drive research tasks. Most studies utilised ML models using their private health institutional datasets in isolation. That meant that it was not possible to evaluate those separate datasets against common standards and public predictive benchmarks [73], particularly relating to hospital resources utilisation. Because of this void, the research was structured to examine predictive models on the open-access database (MIMIC-III) and benchmark their performance with two purposes:

1) To create a research opportunity for health analytic researchers that could advance ML towards improving hospital management and clinical information systems, and
2) To optimise hospital resources utilisation and staffing using ML models.

In this chapter, two case studies on common hospital and intensive care units (ICUs) admission conditions were carried out to benchmark ML predictive models with the use of electronic health and hospitalisations by disease records that impose big demands on resources. Two auxiliary questions were instrumental in answering the main query in this section:

1.b Are there robust predictive and regression classifications in the domain of LOS predictive tasks? If so, what are those?

1.c Which predictive experimentation approach (classification or regression) provides the most decision insight for healthcare professionals working with clinical information systems?

Both above-listed research questions are intrinsic to the thesis research questions (number 1 in the introduction chapter).

3.2 Case Study (1): Classification of LOS Prediction Approach based on Electronic Health Records on Sepsis

Case Study Summary

Sepsis is considered a life-threatening organ dysfunction caused by a dysregulated host response to infection. Sepsis is leading to approximately six million deaths worldwide yearly. Forecasting sepsis LOS is challenging for hospitals globally. Although there are many attempts to improve LOS predictions, there is still a lack of baseline and prediction metrics that can assist in enhancing the management of hospital systems. The study applied a new research architecture based on the relation between the time involved in LOS and the identification of outperforming algorithms for sepsis LOS-ICU prediction. Outcomes of this deployment contribute to better predictive modelling and information visualisation for hospital clinical information systems in the form of ensemble and random forest (RF) models that outdo other classifications intending to predict the LOS for sepsis from electronic medical records in ICU hospitalisations.

3.2.1 Introduction and Literature Review

The majority of sepsis deaths occur in impoverished and developing countries. Reports show that sepsis negative patient outcomes are increasing there [74]. In the United States, national data on electronic medical records (EMR) indicates that an average of 5.9% (approximately 1.67 million) of all hospitalisations and an estimated 260,000 deaths are sepsis-related cases per year. In Australia, there were 36,434 sepsis cases in both public and private hospitals between 2016 and 2017. That data confirmed sepsis among the 20 principal and most common types of hospital diagnosis and admissions in the country [75]. These figures also allow seeing the economic impact of sepsis on medical costs [74]. The cost of sepsis management in U.S. hospitals was $24 billion in hospital expenses ($18,244 per hospitalisation), ranking it the highest among admissions for all diseases there. Generally, in Australia, there are 18,000

patients who suffer from sepsis every year. An estimated $846 million are required to be treated in our ICUs [76].

Sepsis management, particularly LOS, creates a lot of pressure on hospital resources' utilisation globally [77]. In the U.S., the average LOS of sepsis hospitalisation is 75% greater than most other conditions [78]. Several disease severity scoring systems related to sepsis have been developed and adapted in hospitals to identify the severity of this type of patient. The Sequential Organ Failure Assessment (SOFA), The Modified Early Warning Score (MEWS), and SIRS criteria are the most common ones [79]. However, these scoring systems suffer from a lack of accuracy to identify sepsis diagnosis severity, and they require manual tabulation at the bedside [55].

ML improvements are becoming more practical and feasible approaches to resolve the burden on healthcare and the hospital system, such as unnecessary inpatient hospital stays, patients' future outcomes, risk of re-admission, and the cost beyond inpatients hospitalisations [80]. Predictive analytics can also play a significant role in ameliorating these encumbrances by anticipating the number of days patients are likely to stay in a hospital based on their health conditions and diagnosis for admission. ML can eventually improve patients' health outcomes and reduce unnecessary healthcare expenditure in hospital management systems.

LOS modellings play a significant role in facilitating hospitals' operational success during various admissions, including sepsis hospitalisations, predicting the risk of mortality and chronic conditions. Forecasting LOS is crucial for sepsis management since, if successful, it can help to estimate the total amount of time that a patient will stay. This is an appraisal that can further assist in administering human resources and organising areas under great stress in a hospital, such as acute Medicine, ICU, Emergency Department (ED) and General Hospital Admissions. ML predictive analysis on LOS is especially fitting for assessing and decision making on hospital planning capacity, managing patients' expectations and clinical support, monitoring hospital performance and quality of care [4], identifying costs, and post-hospital admissions, that affect medical staffing and resources management [12].

A well-developed ML predictive model can also identify both life-threatening diseases that are difficult to treat and also increase healthcare expenditures; and eventually, also distinguish what admissions convert into high-cost patients [73] during hospitalisation. Accordingly, this case study examined the best three outperforming machine learning classifiers' algorithms and benchmarked the most accurate prediction classification models that predict sepsis LOS from

EHRs in ICU hospital settings. The outcomes of this case study aimed to introduce a robust LOS predictive model for sepsis diagnosis in ICU hospitalisation.

In the past, several methods have been used to address the problem of hospital management and LOS prediction of sepsis by using forecasting models such as regression analysis models [81, 82]. Other researchers have addressed LOS with models of prediction problem classification [46-48]. For instance, McCoy et al. [54] explored the potential of machine learning algodiagnostic (MLA) to predict the length of stay of sepsis patients and improve sepsis management and patient outcomes. The goal of the study was to identify patients with sepsis with earlier intervention and avert its development. The results indicated that the MLA algorithm was able to improve the average hospital LOS from 3.35 days to 2.95 days and a reduction to 4.8% of human and capital resources. Mao et al.[55] evaluated the performance of ML algorithms (InSight, gradient tree boosting) on two datasets (University of California, San Francisco- UCSF and MIMIC-III) to predict LOS for sepsis patients at their emergency department. LOS was binned in labels with (0–2, 3–5, 6–8, 9–11, 12+) days, including ICU stay. However, the study did not examine the performance of their proposed algorithm in a real environment.

Shimabukuro et al. [83] evaluated a ML approach to predict the LOS for severe sepsis using randomised clinical trials. The study used both allocations (control, experimental) in hospital and ICU stays. The results showed that their approach reduced LOS from 13.0 to 10.3 days (26% decrease). However, the study did not elaborate on the type of prediction classifier, features engineering, and the prediction model selection since it was a clinical trial. The population of patients was a small sample size (142 patients), with a total of 38 having severe sepsis.

Burdick et al. [84] implemented an ensemble machine learning algorithm (XGBoost) to predict severe sepsis onset within 48 hours. Two datasets were studied (The Dascena Analysis Dataset "DAD" and the Cabell Huntington Hospital Dataset "CHHD") to evaluate the performance of the prediction method. The DAD was used for training and testing models, whereas CHHD was used for external validation. The model results showed the potential for great benefit for severe sepsis LOS prediction since it reduced LOS by 32.3%.

Common to most of the literature reviewed is that the publications had limitations due to LOS predictions that were based on private medical institutions' databases. Furthermore, there were far fewer attempts to benchmark LOS predictions on sepsis, considering its model run time as

an essential vital element in real-time settings. The model run time is the time required to build the predictive model. However, the shortest is better, especially if the clinicians and caregivers are more engaged in such clinical decisions. Therefore, we searched for information with more ample scope for research from within open datasets and EHRs to verify the effectiveness of predictive machine learning models, which would help evaluate the performance of these predictive analytic techniques in various EHRs inputs and at a global scale. Admittedly, forecasting LOS is still a challenge for many researchers because of the intrinsic difficulty of the task, which remains an open research question for forecasting LOS in the hospital management system.

3.2.2 Method

The method used in my research comprises the description of the dataset choice, the steps toward the predictive algorithms benchmarking, and the outperforming model. This method aims to compare the prediction capabilities of ensemble machine learning and the conventional traditional machine learning classifiers to predict sepsis length of stay in binary prediction problems. Then, its goal is also to develop a robust predictive architecture that can be a valuable tool to baseline the sepsis LOS prediction models in ICU admissions in the hospital management system.

3.2.2.1 Data Description

The MIMIC-III (Medical Information Mart for Intensive Care III and the later version MIMIC-IV) [85] is used in this study to assess the performance of these algorithms (machine learning and deep learning). MIMIC-III is a large and freely available dataset that is maintained by the Massachusetts Institute of Technology (MIT)'s Laboratory for Computational Physiology. It consists of large records of de-identified health-related data from over 40,000 patients who stayed in ICU hospitals [85] between 2001 and 2012 at the teaching hospital of Harvard Medical School (BIDMC) in Massachusetts, USA. The database contains many types of information, such as demographic age, patients' vital signs, laboratory and test results, medications, health and medical procedures, mortality data and caregivers' notes. MIMIC-III is widely used in the research field and supported by many researchers. Therefore, it is highly suited for medical prediction tasks and particularly for benchmarking clinical predictive tasks. Figure 10 provides an overview of the MIMIC-III database.

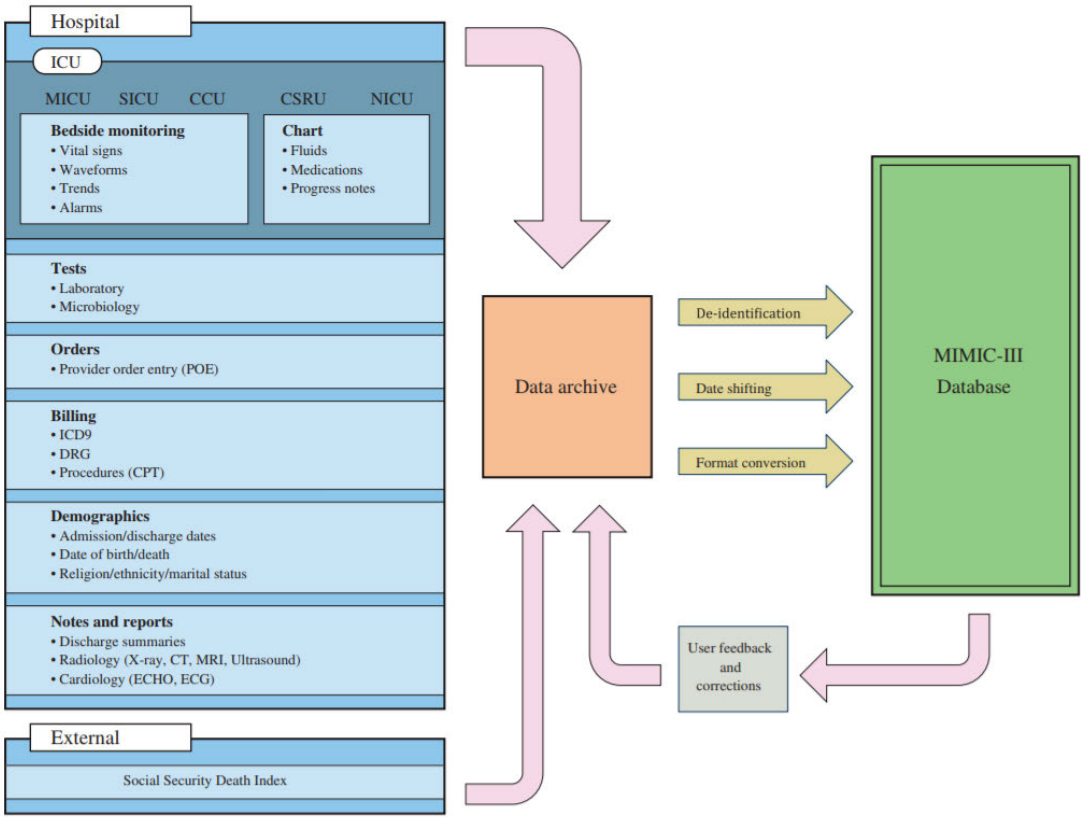


Figure 10. An Overview of MIMIC-III database construction [85]

MIMIC-III contains hospital LOS associated data such as diagnostic codes, survival data, bedside information, demographics, notes and reports (e.g., discharge summaries) and more. All of this information, including categorical variables scale data, was extracted from the MIMIC-III database using an SQL server and stored in the research database (using SQL stored procedures).

In the MIMIC-III database, all distinct sepsis hospitalisations were identified as compromising (1783 cases, "Male: 958/Female: 825") with no exclusion for age groups. The mean LOS in the dataset is 10.17 days, and the age mean for sepsis admitted patients is 63.77. Figure 11 shows the feature extraction, sepsis cohort selection, outliers removals, and missing values treatment procedure during the data processing following the MIMIC-III extractor tool [86]. Figure 12 shows the proposed predictive architecture for sepsis length of stay in hospital management systems.

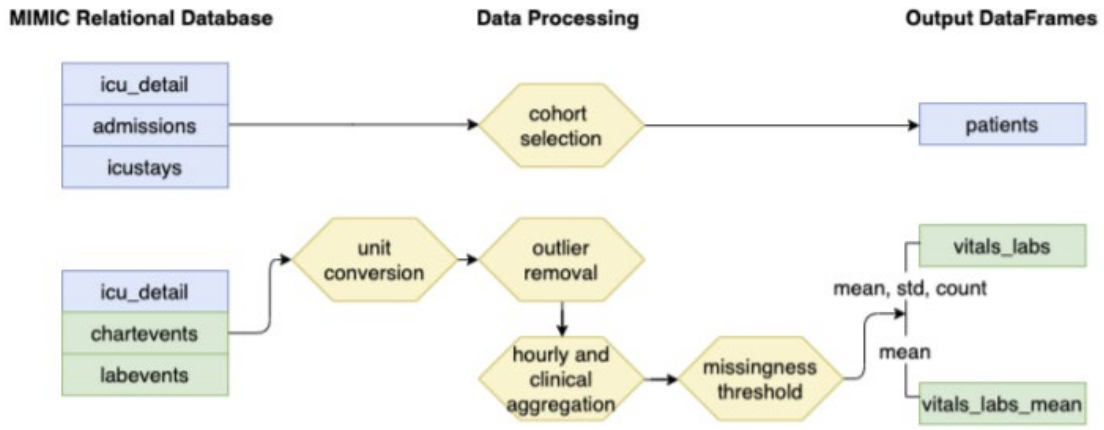


Figure 11. Features extraction and data processing steps [86]

3.2.2.2 Data Preparation and Cleaning

During data preparation, feature extraction and features selection, variables were extracted from MIMIC-III tables "ADMISSIONS", "PATIENTS", "ICUSTAYS", and "DIAGNOSES_ICD". These variables are admission and demographic information. Clinical information extracted from Tables ("D_Labitems", "Microbiologyevents", "Prescriptions", "Transfers", "CharEvent"). The variable (HAMD_ID) represents unique admission events to the ICU in the MIMIC-III database, used to identify each admission to the ICU. Table 5 shows the extracted variables from tables that were included in this study. First, data blending and combining of all selected variables were performed using merge and join DataFrames with Pandas in Python in one table (sepsis_mimic). Next, an "Impute Missing Values" technique was used to replace missing values with specific values that have meaning to sepsis admitted cases. Afterwards, the decision was to extract 24 variables (Table 5), and then a newly modified table was constructed (md_sepsis_mimic).

Table 5. Describes extracted sepsis variables from MIMIC-III.

Variables type	Variables
Admission and demographic variables	(HAMD_ID, Age, Length_of_stay "LOS", Admission_Type, Admission_location, Admission_Diagnosis, Insurance, Religion, Marital_Status, Ethnicity, LOS_group)
Clinical variables	("itemid" variables from chartevents.csv: Body_Tempreature, pH, Glucose, Heart_rate, Blood_pressure), Drug_code, Procedure_Event, Transfer_Event, Lab_item, MicroLabs_Value, DRG_CODE, DRG_MORTALITY, Chart_Event)

3.2.2.3 Features Engineering and Pre-processing

During features engineering, a technique called 'one-hot-encoding' was used to convert five categorical variables (such as gender, admit type, admit location, etc.) from (md_sepsis_mimic) into values (0 or 1) so that it can improve the prediction for the selected machine learning models. The encoded categorical variables were added to the extended dataset (md_sepsis_mimic.csv), generating 37 variables (predictor variables). The preprocessing data stage is an essential step in the prediction models. Lastly, data scaling "normalisation" was applied to get scaled features from the dataset, which will be passed to the prediction models.

3.2.2.4 Models Description

The choice of models (Figure 12) was decided according to the nature of the data (labelled) which was extracted from the MIMIC-III repository. The sepsis LOS-ICU predictive architecture considered classification models for performance evaluation of LOS prediction for sepsis hospitalisation. The classifiers were used as follows: (1) Logistic Regression (LR) (chapter 4) [27] (2) Random Forest (RF) (chapter 4) [23] (3) Stochastic Gradient Descent (SGD) [30] (4) K-Nearest Neighbors (kNN) [31] (5) Decision Tree (DT) [32] (6) Gaussian NB (7) [33] Support Vector Machine (SVM) [34]. The predictive architecture compared 7 models against each other since the target feature (LOS) is a binary label (0: Short LOS, 1: Long LOS). Python programming language was used to build the predictive architecture for the sepsis LOS-ICU algorithm. The predictive programming architecture helped to develop the benchmarking model, using the Sklearn machine learning library to implement all classifiers. The LOS in MIMIC-III is a continuous variable. However, the sepsis LOS-ICU predictive architecture considered predicting LOS for sepsis as a binary classification problem [73] or binary prediction outcomes [49]. Therefore, LOS was discretised (binned) in ICU into two groups (Short LOS < 7 days, and Long LOS ≥ 7 days) according to previous studies that have grouped LOS into a binary variable or two buckets such as [73] [49, 87].

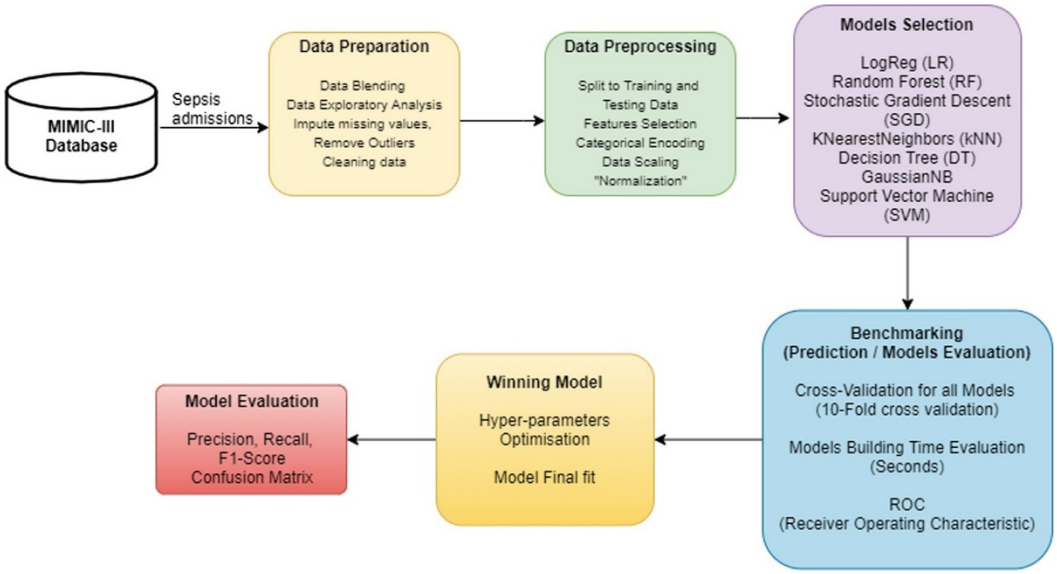


Figure 12. Proposed predictive architecture for sepsis length of stay in hospital management systems.

3.2.2.5 Models Benchmarking and Evaluation

Prior to the training and evaluation processes, the dataset was divided into training (66%) and testing (34%). The trained models were not exposed to the test dataset during the training phase, and any predictions were made on the dataset. Hence, it is designed to be indicative of the performance of the prediction models in general.

Cross-validation is a more sophisticated approach than using training and testing datasets since it tests the harnessing of each predictive machine learning algorithm. Cross-validation divides the dataset into a number of equally sized groups called folds. In cross-validation, each model is trained on all folds except one which is left for testing. Afterwards, this process was repeated until each fold got the chance of being left out and acted as a testing dataset. In the predictive benchmarking architecture, the 10 folds were used in cross-validation for each model (9 folds for training and one for testing).

To measure the models' prediction performance, the ROC [88] (also called AUCROC "Area Under the Receiver Operating Characteristics") was used as the primary measure of the model prediction. Consequently, the best three prediction models will be pointed out. Thereafter, based on the compression between the prediction results and models building time (Figure 13), the outperforming algorithm was selected for the next step. Finally, the hyperparameter tuning stage was applied once the best predictive algorithm was selected, as it was crucial in model performance optimisation. Therefore, grid-search with the cross-validation (GridSearchCV)

method was used to fit the model, display the best hyperparameters, and evaluate the performance.

3.2.3 Results and Discussion

In this section, we represent and discuss the predictive results of the LOS sepsis predictive research framework. The discussion highlights the importance of the predictive results in comparison with previous LOS sepsis research attempts. Hence, the results are the first attempt in the literature to predict LOS sepsis and benchmarked ML machine learning models in intensive care units, considering performance benchmarked predictive models and time to build the model.

The prediction models were trained to predict sepsis LOS. The prediction results showed that the ensembles of machine learning models (RF, DT, SGD) had relatively higher accuracy than other models. It is noted that the Logistics Regression (LR) was ranked as the second-best ML algorithm alongside the ensemble models. Figure 13 provides a comparison between the predictive models. The figure shows a tradeoff between the prediction accuracy and each model's building (exaction) time. The results show that the RF is the best model in terms of accuracy; however, the model building time (0.103001 s) was the second least in building time. The Gaussian NB model is ranked number 1 amongst all model building time models (0.018998 s). However, this model had the worst accuracy (0.67) compared to the other classifiers.

The task is predicting the probability of binary outcomes (Short LOS "<7 days" and Long LOS " >7 days" or "0, 1"); therefore, the ROC is used as a tool to achieve this task. The ROC [89] is calculated by plotting the Sensitivity (True Positives/(True Positives + False Negatives)) against Specificity (True Negatives/(True Negatives + False Positives)).

The Random Forest model outperformed other models with ROC = 0.93. The Decision Tree and Logistic Regression follow this, showing ROC of 0.90 and 0.89, respectively. The other models showed good ROC above 0.80 except for the kNN classifier (ROC = 0.73). This indicates that all models can predict both classes (short LOS and long LOS) with relatively good ROC scores.

In the clinical diagnosis domain, the best outperforming models were implemented with the highest ROC. Therefore, the choice is to select RF as the winning algorithm to predict LOS sepsis in an ICU setting and help to improve resources utilisation in hospital management

systems. The RF hyperparameters optimisation was practised by choosing the best features by importance (e.g. blood pressure, heart rate, pH, age, drug code, DRG_MORTALITY admission_type, procedure_event, transfer_event, MicroLabs_value). Hence, grid-search with cross-validation (cv = 5 folds) was applied using the "GridSearchCV" technique from the Sklearn library, which helped to narrow down the range of hyperparameters (Figure 13).

To evaluate the performance of the winning model (RF), the classification report from Sklearn was used to accomplish this task. Table 6 shows the RF model's robustness in LOS binary prediction problems according to the predicted outcomes (Precision, Recall, and F1-Score). Confusion Matrix (Figure 15) was used to evaluate the predicted labels (short LOS: 0, long LOS: 1) again the true labels (short LOS: 0, long LOS: 1). There were diagonal of squares that represented the true positive (T_P) and the true negative (T_N). This showed that the labels LOS are correctly predicted and classified. The off-diagonal squares illustrated the incorrect classifications (misclassified) for the four LOS groups. Also, it described false positive (F_P) and false-negative (F_N).

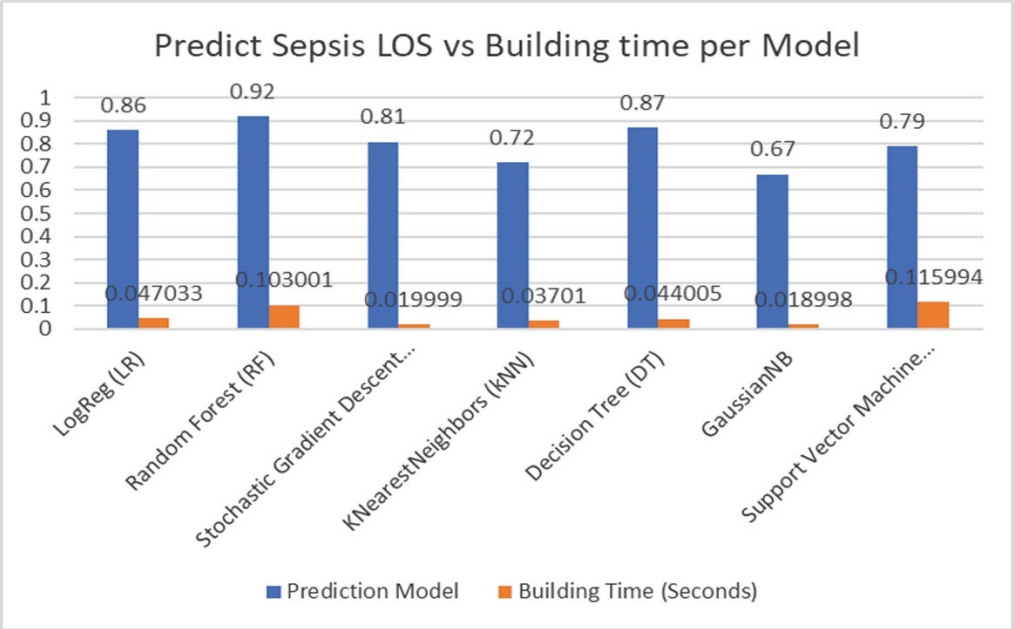


Figure 13. A comparison between the accuracy of each predictive model and the building time for each model.

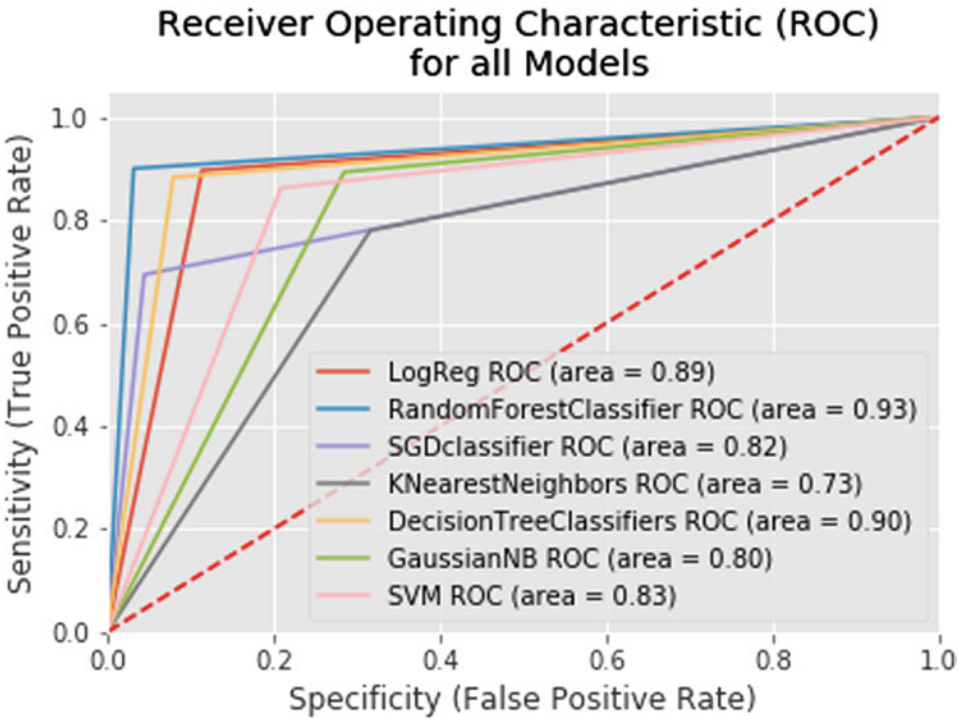


Figure 14. The ROC values for all models for LOS sepsis prediction.

Table 6. Precision, Recall, and F1-Score for

Random Forest (sepsis Predict LOS).

Model	LOS label	Precision	Recall	F1 score
RF	Short LOS	0.91	0.96	0.93
	Long LOS	0.95	0.90	0.92

There are some limitations associated with the current case study. Firstly, the study did not examine all clinical factors associated with sepsis due to the number of extracted variables in the dataset. The limitation is linked to the data collection from MIMIC-III providers. Hence, the variables input to the LOS sepsis predictive models were scoped to what has been provided in the dataset. We consider this limitation is outside of our control. Therefore, the associated clinical vitals, labs, and medications that directly impact sepsis hospitalisations will be examined in future research. Since the focus was to predict sepsis inpatients for all admissions in the MIMIC III dataset, it is worth evaluating the prediction models into specific age group admissions, such as in pediatric hospitalisations and the elderly age group. Secondly, the prediction models for sepsis of various types, including septic shock, will be examined and evaluated. Finally, another important feature to consider is the performance of deep learning

models to be assessed on sepsis datasets with complex features or big-sized data where deep neural network approaches are needed.

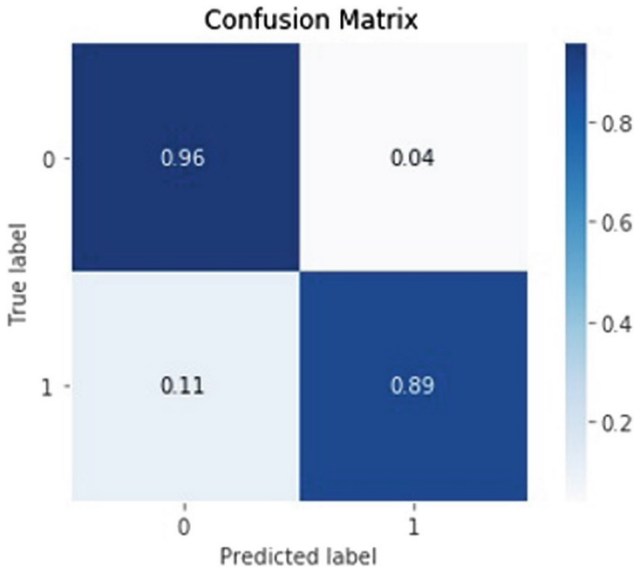


Figure 15. The Confusion Matrix for Random Forest model.

3.3 Case Study (2): Predicting LOS with a Regression Approach for Cardiovascular (Heart Failure) Hospitalisations in ICUs

Case Study Summary

This case study contributes to the field of predictive modelling for electronic health records for hospital management systems. The prediction of cardiovascular LOS based on hospitalisation at the time of patients' admission to a coronary care unit (CCU) or the cardiac intensive care unit (CICU) is deemed a challenging task for hospital management systems worldwide. Several studies have examined the LOS predictive analytics for cardiovascular inpatients in ICU. However, these researches scarcely utilise machine learning models to predict the probability of heart failure patients and LOS in ICU hospitalisation. Uniquely, this case study used a predictive research architecture to predict LOS for heart failure diagnoses from EHRs using state-of-art machine learning models, particularly ensembles regressors and deep learning regression models. Outcomes of the case study showed that the gradient boosting regressor (GBR) for LOS and EHRs outweighed the other proposed models in this study. Furthermore, GBR and its added-on Staking Regressor method showed results of a higher R^2 value. The Random Forest Regressor (RFR) was the fastest model to train in the performance measure of

training speed. RFR outcomes suggest that the deep learning-based regressor did not obtain better results than the traditional regression model in this study.

3.3.1 Introduction and Literature Review

In a traditional hospital management system, comprehensive transitional care interventions for hospitalisations reduce the risk of re-admissions and the risk of mortality. Fittingly, proper predictive analytic methods can ameliorate pressures on treatment and direct impact on hospital resources utilisation. Exemplarily, heart failure (HF) is a clinical syndrome identified by distinctive symptoms such as breathlessness, ankle swelling, and fatigue caused by a structural and functional cardiac abnormality. HF results in reduced cardiac output and or elevated intracardiac pressures of the heart at rest or during stress [90]. Generally, in the world, HF patients or sufferers amount to 1% to 2% of the total population [91]. Heart failure hospitalisations are costly for hospitals since these increase healthcare expenditure. Globally, this represents as much as 1% to 2 % in costs for the healthcare system [92]. The forecast for HF cost by 2030 is that it will amount to $3.5 billion in Australia [93] and $69.7 billion in the U.S.A [94]. These figures show how hospital systems (private and public) will experience increasing costs and pressure to treat HF inpatients [95]. Consequently, it has a direct impact on hospital resources utilisation.

ML prediction models can play a significant role to facilitate different stages of hospitalisation, such as predicting inpatients' LOS, risk of mortality, risk of re-admission or unplanned admission [80, 96]. Several recent studies have explored the effectiveness of prediction models to address the problem of inpatient LOS for cardiovascular hospitalisations. However, only a few have examined the effectiveness of prediction models for HF. For instance, in statistical analysis approaches, Omar and Guglin [97] implemented univariate analysis to determine the short LOS (7 < days) and longer LOS (7 days) for HF patients. Almashrafi et al. [95] utilised a multivariate regression model to determine the prolonged length of stay for congestive heart failure (HF) patients. The observational study by Durstenfeld et al. [98] used a statistical generalised estimating equation in their prediction for the analysis of variance (ANOVA) to compare the predicted and actual LOS for the observations. However, sample sizes of HF patients in these studies are small, and some of these approaches did not assess whether accuracy improved closer to discharge, such as in [98].

Using ML-based methods, Tsai et al. [99] compared the performance of artificial neural networks and linear regression LR for the length of stay of HF inpatients. Other ML predictive

approaches such as [49, 100, 101] examined the LOS for heart failure patients using decision tree-based approaches. The literature review on studies relating to HF LOS prediction focused on predicting the risk of re-admission and mortality. Whereas very few studies attended to the problem of HF predictive LOS, most of them only used statistical inference approaches to determine the LOS of HF hospitalisations. There is a great need for more studies that apply ML techniques. Ideally, this is the direction of this case study. ML should offer hospital management systems' solutions, improve patients' health outcomes, discharge planning for HF, and improve resource utilisation. The key contribution of my case study is to introduce a practical predictive architecture for LOS benchmarking of the regression models versus the deep learning model for heart failure hospitalisations from the hospital electronic medical records in ICU settings.

This chapter analyses and compares the prediction abilities of the regression ensembles-based machine learning method against the deep neural networks and finds the optimal prediction method for the heart failure LOS regression prediction problem in ICU-based hospitalisations.

3.3.2 Method

The methods of this case study used a predictive framework for HF LOS prediction using ICU electronic medical records data. It followed steps towards building both a predictive framework to baseline the outperforming predictive model for HF LOS admissions and a performing model where top features are passed during the model tuning stage. Relevant models' evaluation metrics were chosen to examine the performance of each model.

3.3.2.1 Data Description and Data Preparation

The dataset used in this case study is based on the data extraction procedure in case study 1. The process_mimic.py was used during the data preparation process [64] to mine the cardiovascular inpatient variables. Five tables (CSV file) were merged, which contained Vitals, laboratory test, Demographic with variables from MIMIC-III tables ("ADMISSIONS", "PATIENTS", "ICUSTAYS"), and selected variables were combined using merge and join (DataFrames) with Pandas in Python in one final table ("HF_MIMIC1_4v"). For missing values, a technique called "Impute Missing Values" [102] was applied to replace missing values with specific values that have meaning to heart failure admitted cases from the dataset.

3.3.2.2 Data Preprocessing

Pearson correlation test was implemented to determine the correlation between the independent variable and the output variable. The features' inputs for the candidate models (25 independent

variables) were selected according to the correlation between the independent variables (LOS) and the dependent variables. To convert categorical variables, (one-hot-encoding "0,1") was used for variables (such as gender, admit type, admit location, etc.). This technique is essential to improve results for prediction models. The data scaling "normalisation" technique was applied to get scaled features from the dataset, which will be passed to the prediction designs.

3.3.2.3 Candidates Models

The chosen models (Figure 16) demonstrated the nature of the length of stay data type (scale / continuous), which was extracted from the MIMIC-III repository. Further, regression-based models were considered for the goals of this case study. Accordingly, the regression-based models were evaluated, including Random Forest Regressor [23], Gradient Boosting Regressor [24] and Stacking Regressor [25] and the best regressor model in this case study was assessed with the Deep Neural Network model [26].

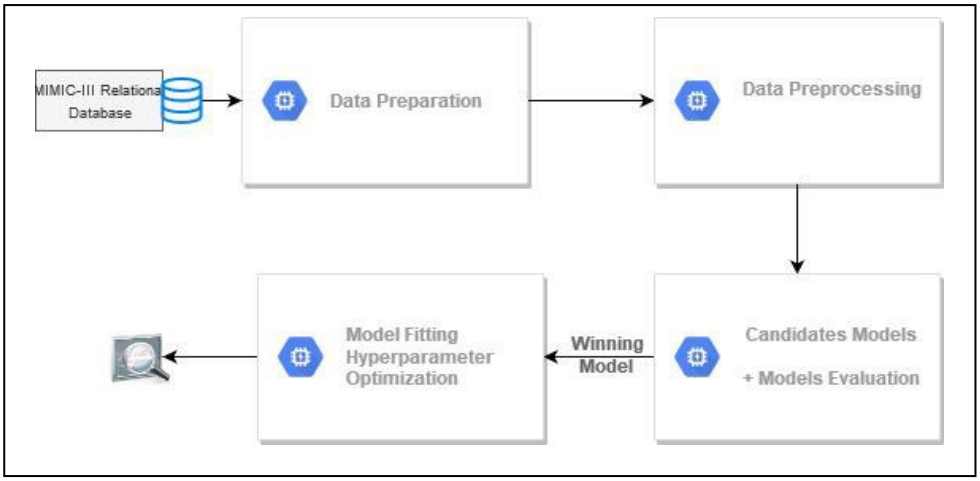


Figure 16. Proposed predictive framework for HF length of stay in ICU hospital management systems

3.3.2.4. Candidates Models Evaluation

Model evaluations are important in prediction tasks to examine the performance of each model and report the outperforming model(s). In this case study, R-squared (R2 or the coefficient of determination) was used as a regression evaluation metric. Thus, R^2 indicated how much variation of a dependent variable was explained by the independent variable(s) in the regression models. Also, the mean average error MAE (equation 1) was used to measure the difference between continuous variables' regression evaluation metrics. It is worth noting that the closer R was to the value 1.0, the better the prediction model performance was. Regression models

were evaluated against each other (stage 1), and their performance was assessed and compared with the DNN (regression) model (stage 2).

$$MAE = \frac{\sum_{i=1}^{n} |LOS_{pred} - LOS_{obsrv}|}{n} \quad (1)$$

where LOS_{pred} is the predicted LOS, LOS_{obsrv} is the observed LOS, and n is the sample size (n = HF patients).

3.3.2.5 Model Fitting and Hyperparameters Optimization

Hyperparameters or tuning are vital for the success of a fitting model. Hyperparameters determine the skill for the selected model, learned from data, and the ML expert manually configures them. Grid search parameter tuning and random search parameter tuning [103] are common algorithm tuning considered the last step before obtaining predictions results. In this case study, the choice of model tuning was based on the winning model.

3.3.3. Results and Discussion

All distinct HF hospitalisations were identified (1592) based on ICD-9. Table 8 shows selected variables (demographic and vital) from the HF patients' characteristics in the study. Male inpatients' HF hospitalisations (52%) were slightly higher than female inpatients' HF hospitalisations (48%). The LOS patients' mean and median are 67.74 and 62, respectively. The minimum LOS is 0.13 days, and the maximum LOS is 93.94 days. All prediction models were implemented using Python programming language. Scikit- learn [104] library was used during the regression models building and Keras [105] for the deep learning model. The HF dataset was divided into training (64%) and testing (34%). Regression models were compared with deep learning models using the evaluation metrics.

Three regression models were considered (Random Forest Regressor, Gradient Boosting Regressor, and the Stacking Regressor). The regressors were built and evaluated, including exaction time. The results in Figure 17 showed that the three models (GBR, Stacking Regression, RFR) showed relatively close R^2 and MAE: R^2:0.81, 0.81, and 0.80 and MAE: 2.00, 1.92, and 1.98, respectively. The stacking regressor was the slowest in prediction execution speed (11.85 seconds), and GBR was the fastest in execution time. GBR and Stacking had the best R^2. The model training and evaluating time is vital in real-time settings, such as predicting clinical and medical cases. Therefore, the GBR was the winning Regressor model.

The Deep Neural Networks (DNN) approach was implemented using Keras with TensorFlow backend using three layers (25 dense layers). Some functions were configured during DNN,

such as "linear" activation and Stochastic Gradient Descent (SGD) for the optimiser. Figure 18 illustrates the DNN prediction results, whereas Table 7 compares the prediction results for the regression models vs the DNN model. It was noticed that regression models showed overall better results compared to DNN. Generally, DNN performs well in larger data sizes and high dimensional data due to its automatic learning feature benefit. Therefore, GBR was fitted into the model tuning stage due to its overall desired performance. Gradient Boosting Regressor performed well compared to other regression models as well as the DNN. Model refinement was obtained by using GridSearchCV from Scikit-learn to get the best estimators (hyperparameters) for the GBR model. One of the advantages of the GBR is that it was robust to the overfitting problem; hence, a larger number of n_estimatrors = 200 can result in better performance. The second tuned parameter is the max_depth = 3, where tuning this parameter to the performance depends on the iteration of the input value. The default value is 3. The depth of the maximum iteration was set to three individual regression estimators. The third tuned parameter was the loss function, and set it to 'ls', 'lad', where ls is least squares regression, and the lad is least absolute deviation. Then, GBR was fitted on the predicators (Figure 19). Then the model achieved improved and attained better results (R^2: 0.84±0.0.7) after fitting the GBR with the best hyperparameters and the top features.

Table 7. Regression vs DNN Prediction results

Model	***R2***	***MAE***	***Time***
RFR	0.8±0.08	1.98±0.16	0.95 sec
GBR	0.81±0.07	2.0±0.14	0.85 sec
Stacking Regression	0.81±0.08	1.92±0.15	11.84 sec
DNN Regression	0.77±0.06	2.30±0.18	4.11 sec

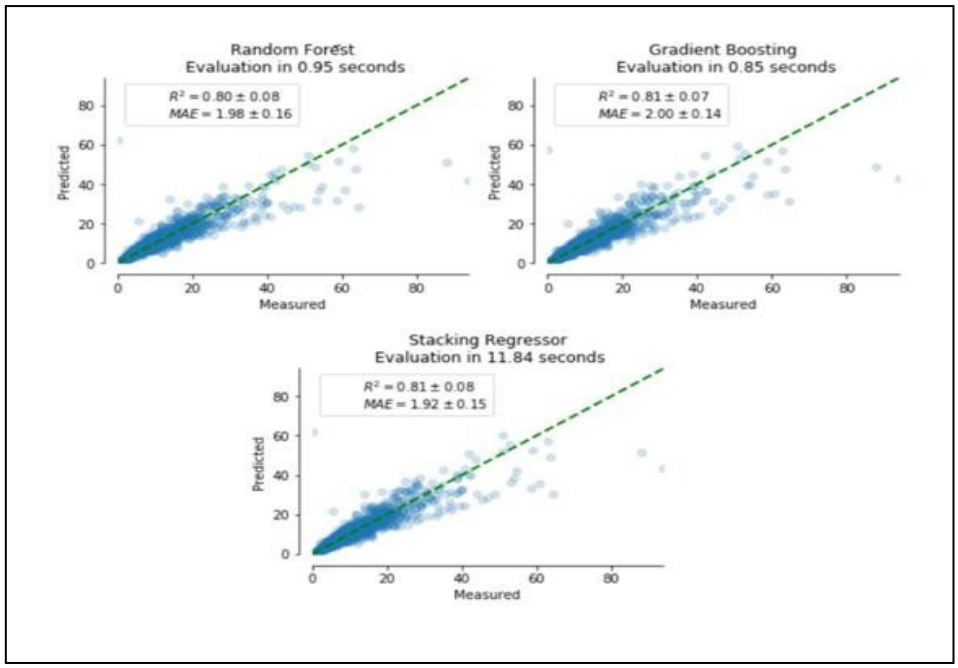


Figure 17. Single Regression predictors vs Stacked predictors

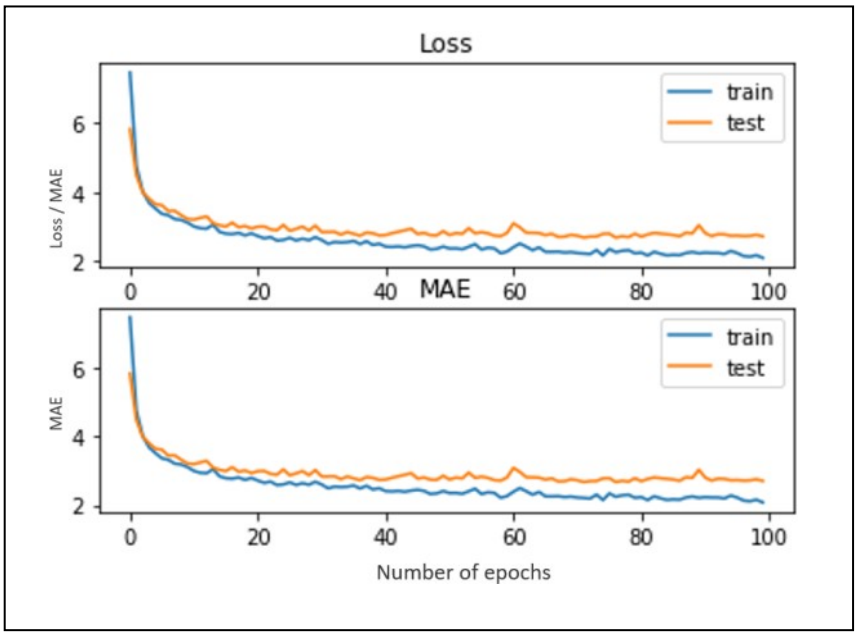


Figure 18. Loss vs MAE for the DNN model on training and testing HF sets.

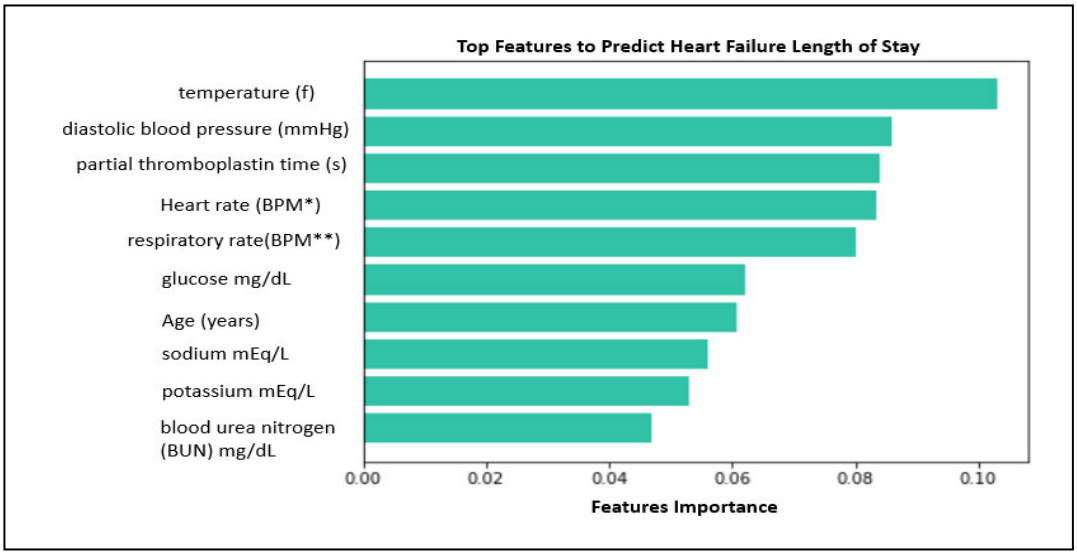


Figure 19. Best 10 features (predictors) for GBR model. * beats per minute, ** breaths per minute

Table 8. Patient characteristics

HF				
	(Mean/ Median)	***Std***	***Min***	***Max***
Demographics				
LOS	(10.77 / 8.13) days	9.16	0.13 days	93.94 day
Age	67.74/62	13.17	0	89
Weight	Mean: 80.84	9.79	34.59	127.59
Gender	Female: 765 (48%)	Male: 827 (52%)		
Vital				
Temperature (F)	Mean: 97.83	3.41	33.9	100.48
Heart Rate	Mean: 92.27	22.19	35.5	161.41
Respiratory Rate	Mean: 18.02	3.40	8.4	27.8
Pulse Oximetry	Mean: 97.09	1.60	57.85	99.83

The goal of this case study was to develop ML predictive LOS approach in relation to HG patients. However, the case study had limitations associated with it. Particularly, it did not consider LOS factors to predict LOS, such as whether other medical comorbidities [106] are associated with HF inpatient (presence of one or more additional health conditions often co-occurring with the primary condition), which may need further research exploration in a future study to measure its impact on the LOS or the extended LOS. Also, the post-hospitalisation intervention was not addressed, and its effect on heart failure LOS was not examined.

3.4 Chapter Conclusion and Future Direction

This chapter studied problems of predicting LOS in ICU-hospital settings using classification and regression approaches. Two LOS case studies aimed to benchmark the LOS predictive models in both predictive methods. Their results allow us to summarise valuable conclusions and recommend the following future direction.

LOS binary predictions with a classification approach (case study 1) for sepsis LOS in ICU proposed a predictive architecture based on ensemble-based models such as (RF, DT) alongside the LR provided higher ROC scores. Furthermore, the Random Forest (RF) model outperformed all other models in the proposed predictive architecture. Therefore, the RF can be implemented as a promising predictive tool to improve clinical information systems for hospital resources for sepsis hospitalisations by correctly categorising and predicting the (Short and Long) LOS.

The regression approach (case study 2) aimed to develop ML LOS predictive framework to predict HF for patients' hospitalisation in ICUs using electronic medical records' historical data. The GBR regressor outperformed all other models in this study with (R2 0.8±0.08). On the other hand, the Deep Learning Model (Deep Learning Regression) did not report better results than regression models. Hence, it will be commendable to examine DNN on an extensive EHRs dataset with a large volume of heart failure hospitalisations and compare it with regression models in a future study. Moreover, in future work, the proposed predictive HF LOS framework will be validated on a real-world external dataset to evaluate the performance of the proposed model on other data sources. Thus, the case study contribution provides new insight to healthcare decision-makers in hospital management systems, clinical AI research, and the clinical practitioner to predict patients' future outcomes and determine the costly HF hospitalisations using the artificial intelligence (AI) approaches. Further, it facilitates the mission to clinical AI researchers towards a more comprehensive framework to predict the common LOS heart failure hospitalisations from the clinical diagnosis and improve resources utilisation using machine learning advancements.

While predictive regression approaches determine which admission variables matter most, the predictive classification approaches, primarily the predictive ensemble methods, are more understandable for decision-making rules, especially in clinical information systems. The classification approach can categorise predicted outcomes into 0: short, long, or 1: long LOS, per the first case study; therefore, the classification (binary) is a direct query response for

healthcare professionals in the hospital such as beds managers, nurses or clinician. Consequently, healthcare decision-wise, the classification approach is more practical and suitable for clinical information systems, especially in dynamic and uncertain environments where managing resources efficiently is deemed crucial for hospitals. Hence we are not limited to particular predictive machine learning models. However, this study motivates researchers in clinical information systems to examine other predictive models in a more comprehensive approach to compare different classifiers such as Naïve Bayes, Multilayer Perceptron (MLP) with other boosting and bagging ensembles classifiers (CatBoost, LogitBoost, XGboost, LightGBM etc.). Another important note is assessing and evaluating these classifiers' performance evaluation in different prediction problems such as multi-class/multi-label classification and imbalanced classification problems besides the binary predictive tasks.

The research will take the opportunity to explore and benchmark predictive LOS classification models in more thoroughly data-driven, implementable, and explainable methods in Chapters 4-5.

4. Chapter Four: Class Balancing Methods for Predicting Inpatient Length of Stay in the Intensive Care Unit: An Explainable Predictive Framework for Lung Cancer Hospitalisations (Case Study)

4.1 Chapter Summary

This chapter introduces a predictive length of stay framework for lung cancer patients using machine learning models. The framework proposed here deals with imbalanced datasets for classification-based approaches using electronic healthcare records (EHRs). The framework offers the capability of predicting inpatients hospitalisations and LOS in intensive care unit settings while using the MIMIC-III dataset. Succinctly, performance evaluation for ML models and outperforming predictive algorithms were measured using sensitivity, specificity, AUC, IBA, G.mean, and mean accuracy. Random Forest Model outperformed other models and achieved desired predicted results during the proposed framework's three phases. Oversampling methods (SMOTE and ADASYN) with clinical significance features selection achieved the highest AUC results (98 % with CI%95: 95.3%-100%, and 100% respectively). The combination of oversampling and undersampling methods achieved the second-highest AUC results (98%, with CI%95: 95.3%-100%, and 97%, CI%95: 93.7%-100% SMOTE-Tomek, and SMOTE-ENN respectively). Undersampling methods reported the least important AUC results (50%, with CI%95: 40.2%-59.8%) for both methods (ENN and TomekLinks). Using ML explainable technique called SHAP, we were able to explain the output of the predictive model (RF) with SMOTE class balancing technique and understand the most significant clinical attributes (features) that contributed to predicting lung cancer LOS with RF model.

4.2 Background

The number of services offered in hospital settings is the measure for hospital resources utilisation to quantify and describe hospital services by individuals to avoid, cure health problems, promote maintenance of patients' health and well-being, and obtain information about personal health status and prognosis [107]. Hospital resources management could

increase throughput and offer better inpatients' quality for treatments and health outcomes [108]. However, the health condition under which clinicians work influences assiduously their own decisions. Hence, decision-making within working groups is hardly autonomous; therefore, doctors rely on their peers or healthcare workers. In contrast, doctors working independently can be less impacted by their own decision since other human interference does not exist. However, working in dynamic, challenging environments and conditions may affect their decisions [109].

Consequently, it is crucial to find alternative approaches to allocating hospital resources and managing patients in hospitals fairly and effectively. Thus, resources utilisation in intensive care units is vital for hospital management and hospital healthcare assessment systems. Managing hospital bed availability and efficiency are obligatory for addressing challenges associated with having an overabundance of patients in ICU and hospitals and with avoiding any issue of ICU beds scarcity, especially in uncertainties such as pandemics [110]. Most significantly, minimising the risk associated with acquired infection during ICU hospitalisation, risk of mortality [111], and medical complications for vulnerable patients are key aspects. Improved hospital resources and planning have the potential to mitigate and minimise these risks [112, 113]. Therefore, a lower ICU length of stay (LOS) is necessarily associated with lower total hospital charges. Consequently, hospital resources are well-managed, and better outcomes are achieved for the patients [114].

Traditional LOS calculations are currently in use, such as ICU APACHE versions I,II, III, IV, SAPS [7, 8, 10, 115], and SOFA [116]. These methods use patients' features and/or ICU features to estimate the inpatient LOS during hospital admission. However, they suffer from poor performance; hence, they are not disease-specific prediction methods. Further, there is consensus on the most suitable techniques for ICU LOS [11]. The vast majority of hospital management use electronic healthcare records (EHRs) to facilitate their daily operational and medical procedures and LOS determination. The EHRs healthcare assessment systems store data associated with patients' encounters, such as demographics, diagnosis, laboratory tests, prescriptions, radiological images, clinical notes, and many more [117, 118].

Electronic health records are integral components of Clinical Information Systems (CIS). The CIS is one of the approaches to support the process of hospital resources utilisation and management. The CIS-EHRs systems are digitised computer applications that aim to support doctors with their decision and hospital workflow and ease the mission in a dynamic and

challenging environment, where uncertainties from internal and external factors can occur anytime under different circumstances. CIS-EHRs systems still aid hospitals worldwide in minimising medical and staffing errors and improving patients' safety and health outcomes in prediction medicine. Furthermore, it helps the hospital to reduce costs associated with daily operational factors in efficient ways. Clinical decision support systems are one of the typical implementations of CIS. The CDSS systems are classified into two categories: knowledge-based systems where computer inputs, and the systems where rules obtain the information to produce decisions. The non-knowledge-based CDSS is established upon artificial intelligence models (AI-CDSS), where knowledge and information are extracted from pattern discovery techniques, data mining, and prediction insights to determine the outcomes. The applications of AI-CDSS in healthcare are broad from disease management, precision medicine, diagnosis, drug discovery, and control [119].

4.3 Literature Review

The literature review on predictive LOS for lung cancer patients using ML models revealed a lack of studies specific to this area. Necessarily, a review is needed to scrutinise cancer-based studies with a statistical perspective and the use of ML models in the context of hospital resources' utilisation and healthcare quality while focusing on LOS-cancer-based studies within CIS and AI-CDSS. Numerous statistical-based studies examined the risk factors for LOS in hospitals. However, conventional statistical models have limitations in processing multiple unprocessed variables and in their application to real healthcare data. This procedural deficiency has led scientists to adopt ML in the development of prediction models [120].

ML models and healthcare analytics have proven to be powerful tools to recognise EHRs patterns [121]. ML predictive algorithms assume that there are data and associations and relationships between clinical variables (predictors/independents) and the target variables (dependents). Hence, ML models can predict the ICU LOS remaining time to ML algorithms to predict inpatient LOS in the context of healthcare assessment systems in clinical settings. At the same time, non-regression ML algorithms have great importance in improving research outcomes. Many of these algorithms are expected to deal with a large number of variables in sophisticated and non-linear ways, producing very efficient complex predictions [122, 123]. Based on these techniques, predictive models may assist healthcare systems in identifying clinically significant risks or identifying unique and unusual risk predictors [124]. ML algorithms have been in use in medical imaging and genomics; however, their use to model

clinical outcomes is less well-established [121]. Whether in the ICU or otherwise, ML predictions are essential to hospital LOS. These algorithmic techniques are broadly generalisable, and scientists can build ensembles based on them to predict many other clinical outcomes.

In the literature, state-of-the-art ML models (ensemble methods) [50, 125] were studied in the context of emergency department LOS prediction. Multivariate analysis-based studies [73, 126] were examined in the ICU predicting LOS. Recent attempts are applied deep learning-based regression techniques such as Bayesian Neural Network (BNN) [127] and Short Long Term Memory (LSTM) for time-series prediction [43]. In many studies, predicting LOS with regression-based predictive models is studied extensively [11, 128-131]. At the same time, most of these studies are focused on emergency departments (ED) or cardiovascular-related admission to ICU units or patients who stayed in ICU after the surgical or medical intervention using classification approach such as [132].

The limited number of cancer-based studies assessed the predictive models. For instance, Best et al. [133] evaluated multivariate regression to predict inpatients' length of stay complications after lobectomy for lung cancer at three different treatment healthcare facilities. Key clinical variables were used in their study to evaluate the model performance. Their study reported that the body composition on the preoperative chest and computed tomography is an independent predictor of the LOS (4 days) and the postoperative complications after lobectomy for lung cancer. Nevertheless, the study suffered from limitations associated with the nature of the data collection procedure, such as missing imaging data and potential selection bias.

Pompili et al.[134] examined the logistic regression model to assess whether quality of life (QoL) scales are associated with postoperative length of stay. The findings reported that the model achieved an AUC of 0.762 for preoperative patients. The QoL was associated with the prolonged postoperative hospital stay for lung cancer patients and enhancement of patient healthcare quality in the context of recovery after surgery. The study endured issues related to data collection that underestimated the overall outcomes of the study.

Dong et al. [135] analysed the effectiveness of oxygen desaturation (EOD) and heart rate to predict major postoperative cardiopulmonary complications for non-small-cell lung cancer patients using binary logistic regression. Their proposed approach enabled to predict LOS for the ROC-AUC: 0.750, 95% CI= 0.668–0.831. A similar study by Li et al. [136] applied multivariate logistic regression to estimate the effects of pulmonary fissure completeness on

postoperative cardiopulmonary complications and hospital length of stay in patients undergoing video-assisted thoracoscopic surgery lobectomy for early-stage non-small-cell lung cancer. The reported results showed the multivariate logistic regression could better predict outcomes for the pulmonary length of hospital stay following video-assisted thoracoscopic surgery, and lobectomy for early-stage non-small-cell lung cancer with mean = 14.0 days; 95% CI = 13.4–14.7 days. However, literature on lung cancer LOS is not evaluated on multiple predictive models (benchmarking), and factors, such as the technical side of the models, are not discussed or described appropriately. This lacking rendered the proposed models unable to perform well in different data inputs. Furthermore, the proposed studies lacked clinical data insights, and there was no attempt to examine models within clinical settings and, most importantly, in relation to clinical decision support systems to manage hospital beds efficiently.

Lung cancer patients amount is rough to 27% of ICU admissions [137, 138]. These patients suffer substantially worse ICU outcomes compared to patients with other types of cancer. Additionally, most lung cancer-based studies reported descriptive statistics or regression analysis about the hospitalisation characteristics, such as the median or mean and p-Value [139]. Predicting LOS cancer-based studies [140-142] are less prevalent in the literature review within the context of inpatient admission to ICU in the LOS predictive classification approach. LOS lung cancer-based machine learning studies with a classification-based focus are scarce, and there were no ML studies that examined the LOS predictive models for lung cancer in ICU hospitalisations. Further, the search did not uncover any relevant studies that examined class-balancing methods with ML techniques to predict cancer LOS tasks, especially cancer-based studies in the ICU healthcare context.

Table 9. A comparison between literature studies in the context of lung cancer LOS

Work	**Study Focus**	**Prediction Model**	**Benchmarking Predictive Models**	**Class-Balancing Method(s)**	**Models' explainability**	**Clinical Diagnostic Code**	**Features Selection Method**	**Number of distinct events (Patients)**	**Evaluation methods for Winning Model(AUC, ROC, etc.)**
Best et al. [133]	Lung Cancer	Regression (multivariate regression)	x	x	x	N/A	Spearman correlation	958	N/A
Pompili et al.[134]	Lung Cancer	Logistic Regression	x	x	x	N/A	Backward elimination	250	(AUC 0.762, R^2= 0.14).
Dong et al. [135]	Lung Cancer (non-small cell)	Binary Logistic Regression	x	x	x	N/A	Manually	171	(ROC-AUC: 0.750, 95% CI= 0.668–0.831).
Li et al. [136]	Lung Cancer	Multivariate logistic regression	x	x	x	N/A	Manually	528	(mean = 14.0 days; 95% CI = 13.4–14.7 days)

4.4 Research Questions and Expected Outcomes

The literature review revealed an area for development that is inedited in predictive ML research with the potential to set a sound framework and process to predict lung cancer patients' LOS at the time of ICU admission and their hospitalisation resources records and management. Accordingly, it is important to examine data input factors, including the statistical methods that treat problems associated with data representation. In that way, the healthcare system and hospitals can have better predictive analysis and decision making and improve the utilisation of clinical information systems in hospital settings. Therefore, the research questions that motivate this chapter are:

> **1.d** How to evaluate predictive machine learning models in the state of limited LOS cancer patients' data input?
>
> **2.b** How to treat LOS unbalanced data representation in clinical information systems?

Both research questions are an integral part of the thesis' main research questions. Where a) refers to the first research question and b) refers to the second research question in chapter 1.

Suitably, the experimental research in this chapter explores the LOS prediction as a health assessment metric for resource utilisation in ICU settings with ML classification approaches.

The key contributions in this chapter are to introduce a doable data-driven framework to predict the Length of Stay for unexplored research topics (lung cancer) admitted patients to the ICU. The study provides a practical framework to deal with an imbalanced classification problem in EHR datasets. Hence, the problem is deterministic for machine learning models' performance in healthcare analytics, particularly electronic medical records. The research framework will examine the problem using different six class-balancing algorithms. Thereafter, our proposed framework deliberates the EHR data's dimensionality hardened issue by focusing on clinically significant attributes (Lung Cancer diagnosis) as input features. Furthermore, we utilise the features selection method Recursive features elimination (RFE) in the Lung Cancer LOS to eliminate the worst performing features and select the subset of features associated with the target predicted LOS class. Thus, the optimal features selection method will be evaluated further against the six class-balancing methods to achieve the desired predicted outcomes of LOS lung cancer. Finally, our new predictive approach utilises the explainable machine learning approach (SHAP) that fits the outperforming classifier with the clinically appropriate class balancing method in the context of binary class prediction problems.

4.5 Method

4.5.1 Data Description and Features Extraction

The method used in this research was based on a dataset of MIMIC-III (Medical Information Mart for Intensive Care III, v1.4) that was available to conduct LOS for the lung cancer experiment using the proposed framework further here below. The MIMIC-III dataset compromised de-identified health-related data associated with over 40,000 patients who stayed in ICU between 2001 and 2012 at the Harvard Medical School's teaching hospital (BIDMC) in Massachusetts, USA [85]. The MIMIC-III is a relational database consisting of data tables on patients who stayed at the ICU BIDMC hospital. The patient's details such as demographic age, patient vital signs, laboratory and test results, medications, health, and medical procedures are linked by a unique admission ID (HADM_ID) amongst all database tables (EHRs). The dataset has great advantages for this research since it is freely available for researchers worldwide, and it contains a diverse and substantial population of ICU patients. Also, the dataset comprises high temporal resolution data such as electronic documentation, laboratory results, bedside monitor trends, and waveforms. The previous chapter (chapter 3) used the data description of MIMIC-III, and more details on the data attributes and characteristics are discussed there.

The inclusion mechanism in this experiment considered only ICU hospitalised patients. All patients who died in the hospital were excluded from the inclusion protocol at the first screening. Further, all events with missing unique patient stay ID (HDAM_ID and ICUSTY_ID)[2] dropped from the inclusion criteria. An additional inclusion criterion was applied comprising diagnosis codes for lung cancer hospitalisations (162. x) identified by the International Classification of Diseases (ICD-9). Accordingly, 119 lung cancer patients were included in this experiment from the whole MIMIC-III dataset. Figure 20 reveals the inclusion protocol for lung cancer patients in this study.

[2] HDAM_ID and ICUSTY_ID: a unique ID to link unique ICU stays with each HDAM_ID

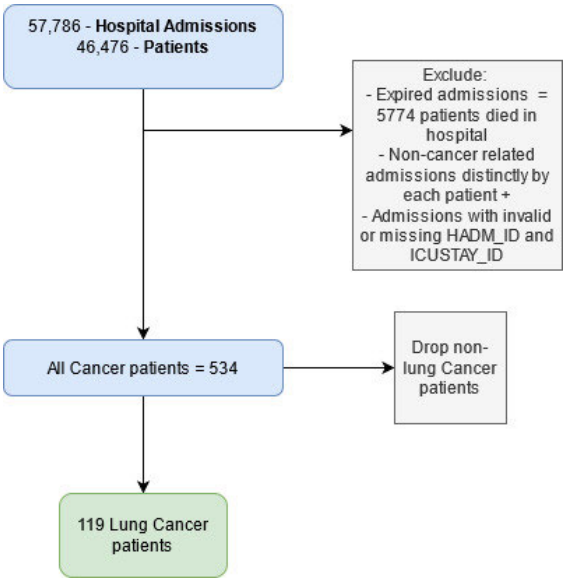


Figure 20. Inclusion protocol for lung cancer patients from MIMIC-III dataset

The number of attributes associated with lung cancer patients comprised a set of complete blood count, differential, white blood count (WBC), vital signs, laboratory tests, demographics, and medications, as seen in Table 10. These variables were extracted from previous works [131, 143] for further data processing. A clinical oncologist involved in this experiment affirmed the attributes' selection with relevance to lung cancer diseases. Eventually, the number of features that are inserted is 75 attributes. Other non-significant clinical variables to lung cancer LOS were dropped from the table according to the inclusion criteria (Figure 20). These variables are basophils, eosinophils, polymorphonuclear leukocytes, surface culture, daily weight, lymphocytes, monocytes, triglycerides, @0.9SodiumChloride, acetominophen, albuterol, aspirin, atenolol, epoetin, hydrochlorothiazide, ipratropium, levetiracetam, lisinopril, neostigmine, oxycodone, pantoprazole, phenylephrine, phytonadione, ranitidine, statin, trazodone, zolpidem, INS_Government, INS_Medicaid, INS_Medicare, and INS_Private. The LOS distribution in the lung cancer case study is 85.58% for Short LOS and 14.42% for Long LOS. The majority of the admitted cases of the population are senior adults, 80% (aged 65+) based on the inclusion criteria. Adults (Middle age: (> 35 years old & < 65) cases were 20%, and there were observations in the MIMIC-III dataset for the young age category (> 14 years old & < 36), and similarly, no observations were available from the dataset for children age category (< 14 years old). The data showed that the population of admitted cases were (65.72%: male) and (34.73%: female).

Table 10. Lung cancer LOS attributes characteristics (features by group, type, mean, standard deviation (std), min value, max value, and feature type)

Feature	mean	std	min	max	Type	*P*-value
Complete Blood Count Information						
RBCs	3.491681	0.346278	2.49	4.6	Continuous	0.536
WBCs	11.23081	4.080906	1	22.2	Continuous	0.862
Platelets	239.9916	90.13732	23	491	Continuous	0.323
Hemoglobin	10.55518	1.127045	7.6	13.2	Continuous	0.259
Hemocrit	32.04676	3.797873	22.5	41.5	Continuous	0.613
Differential Information						
Bands	1.260504	1.52639	0	11	Continuous	0.100
Neutrophils	81.83613	4.391396	46	93.5		0.977
Vitals Information						
Temperature (F)	97.79863	1.86576	82.25	100.3	Continuous	0.983
Heart rate	91.54513	16.08313	56	155	Continuous	0.080
Respiratory rate	19.1562	3.865609	10.4	27.82353	Continuous	0.367
Systolic blood pressure	112.818	14.42322	76.82143	150.1667	Continuous	0.195
Diastolic blood pressure	59.06757	8.691874	41	78.22222	Continuous	0.885
Pulse oximetry	97.06236	1.345939	91.83333	99.69231	Continuous	0.575
Labs Information						
troponin	0.068109	0.02044	0.01	0.235	Continuous	0.671
BUN	19.89356	9.919206	5	81.5	Continuous	0.932
INR	14.94034	0.841213	12.6	18.5	Continuous	0.395
PTT	29.93655	4.387553	19.8	50.1	Continuous	0.515
creatinine	0.863025	0.324706	0.3	2.9	Continuous	0.902
glucose	135.0999	26.96387	74	205	Continuous	0.519
sodium	137.9699	3.94132	120	147	Continuous	0.821
potassium	4.155406	0.436217	2.9	5.15	Continuous	0.976
chloride	103.3396	4.385687	85	112	Categorial	0.263
PEEP_Set	4.892157	1.079598	0	9.333333	Continuous	0.923
tidal_volume	497.8605	67.35754	0	700	Continuous	0.174
anion_gap	13.13025	1.426835	8.5	18	Continuous	0.293
Inspired_O2_Fraction	50.67107	5.678632	30	83.33333	Continuous	0.997
Demographic Information						
GENDER	0.394958	0.490909	-	-	Binary	0.966
Admission Type						0.440
ADM_ELECTIVE	0.420168	0.495673	-	-	Categorial	
ADM_EMERGENCY	0.571429	0.496964	-	-	Categorial	
ADM_URGENT	0.008403	0.09167	-	-	Categorial	
Age Category						0.703
AGE_middle_adult	0.168067	0.375507	-	-	Categorial	
AGE_senior	0.831933	0.375507	-	-	Categorial	
Medications Information						
amiodarone	0.042017	0.201476	-	-	Binary	0.614
ampicillinsulbactam	0.033613	0.180994	-	-	Binary	0.444
atropine	0.016807	0.12909	-	-	Binary	0.592

calciumgluconate	0.210084	0.409091	-	-	Binary	0.439
carvedilol	0.008403	0.09167	-	-	Binary	0.706
cefazolin	0.235294	0.425976	-	-	Binary	0.342
cefepime	0.016807	0.12909	-	-	Binary	0.592
ceftriaxone	0.033613	0.180994	-	-	Binary	0.452
clonazepam	0.02521	0.157426	-	-	Binary	0.004**
clopidogrel	0.016807	0.12909	-	-	Binary	0.592
dextrose	0.848739	0.708767	-	-	Binary	0.014
diazepam	0.033613	0.180994	-	-	Binary	0.452
digoxin	0.016807	0.12909	-	-	Binary	0.110
diltiazem	0.042017	0.201476	-	-	Binary	0.614
diphenhydramine	0.218487	0.414967	-	-	Binary	0.398
enoxaparin	0.008403	0.09167	-	-	Binary	0.706
fentanyl	0.02521	0.157426	-	-	Binary	0.277
fentanylcitrate	0.218487	0.414967	-	-	Binary	0.253
fluconazole	0.02521	0.157426	-	-	Binary	0.277
fondaparinux	0.008403	0.09167	-	-	Binary	0.706
furosemide	0.12605	0.33331	-	-	Binary	0.080
glucagon	0.142857	0.351407	-	-	Binary	0.371
haloperidol	0.008403	0.09167	-	-	Binary	0.008**
heparin	0.739496	0.440766	-	-	Binary	0.572
hydralazine	0.042017	0.201476	-	-	Binary	0.390
hydromorphone	0.361345	0.482421	-	-	Binary	0.368
insulin	0.436975	0.498109	-	-	Binary	0.176
levofloxacin	0.10084	0.30239	-	-	Binary	0.168
levothyroxine	0.10084	0.30239	-	-	Binary	0.642
metoclopramide	0.05042	0.219736	-	-	Binary	0.761
metoprolol	0.327731	0.471371	-	-	Binary	0.528
metronidazole	0.117647	0.323552	-	-	Binary	0.294
midazolam	0.109244	0.313264	-	-	Binary	0.037
nitroglycerin	0.042017	0.201476	-	-	Binary	0.614
nitroprusside	0.008403	0.09167	-	-	Binary	0.008**
norepinephrine	0.042017	0.201476	-	-	Binary	0.614
ondansetron	0.352941	0.479905	-	-	Binary	0.686
phenytoin	0.05042	0.219736	-	-	Binary	0.118
piperacillin	0.05042	0.219736	-	-	Binary	0.118
potassium_y	0.378151	0.486976	-	-	Binary	0.853
prednisone	0.067227	0.251473	-	-	Binary	0.993
propofol	0.193277	0.396538	-	-	Binary	0.030
vancomycin	0.142857	0.351407	-	-	Binary	0.145

** Pearson correlation is significant at the 0 01 level (2-tailed)

4.5.2 Pre-processing

Data pre-processing was deemed an essential task in the data mining process. Generally, datasets suffer from missing values, outliers, or raw data that require further processing and features redundancy [144]. Several steps were performed to process and extract (lung cancer LOS) before evaluating the prediction stage.

4.5.2.1 Data Imputation

Null function from the Pandas library in Python [145] was utilised in this experiment to verify and eliminate records with frequent missing values for each admission. This decision was coordinated with the clinical oncologist and made only in the records with cells (entries) that suffer from many missing values that cannot be replaced by handling missing values techniques Figure 23. Any event (lung cancer admission) that lacked clinical insights was disregarded to avoid any negative impact on the prediction models' performance and the overall research aim. An imputation method to treat missing values with entries based on the variable median [143] was performed in the case of missing values that did not cause any absence of each event (also referred to as admission).

4.5.2.2 Discretisation of the Target Class (LOS)

Discretization is the process of transferring numeric/continuous variables into nominal/categorical variables (bins). Several artificial intelligence studies such as [146, 147] practised continuous variables transformation into nominal/categorical variables to be examined in various machine learning and statistical methods. Healthcare and clinical decision support system studies [148, 149] binned the continuous variables into nominal target variables. In hospital healthcare systems, binning (continuous) LOS into nominal and categorical variables is accompanied by advantages for healthcare caregivers to maximise hospital resource utilisation [150, 151]. This can be achieved by binning LOS continuous variable into classes (labels) to help healthcare workers initially predict patients' future stays at hospital admission. The binning process for discretising LOS variables into categories (labels) is studied to facilitate LOS prediction methods. Previous research works categorised length of stay into different labels. For Instance, Zebin et al. [87] grouped short LOS to 0-7 days and the long LOS to >7 days. Similarly, Allard et al. [152] categorised long LOS to > 7 days and short LOS to below 7 days. In this experiment, LOS continuous variable is binned into a **binary learning** LOS approach based on previous studies using the discretised scaled label "LOS" into two labels in Figure 21 as follows:

1) Label zero (0) for a short length of stay (0-6 days), and
2) Label one (1) for a long length of stay (7 + days).

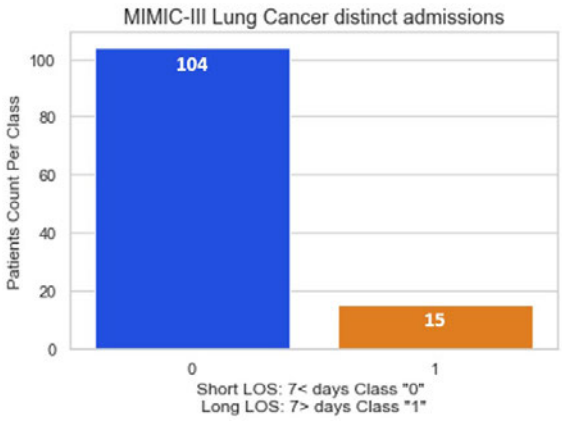


Figure 21. Number of patients per class (majority-short LOS “0”, minority-short LOS “1”)

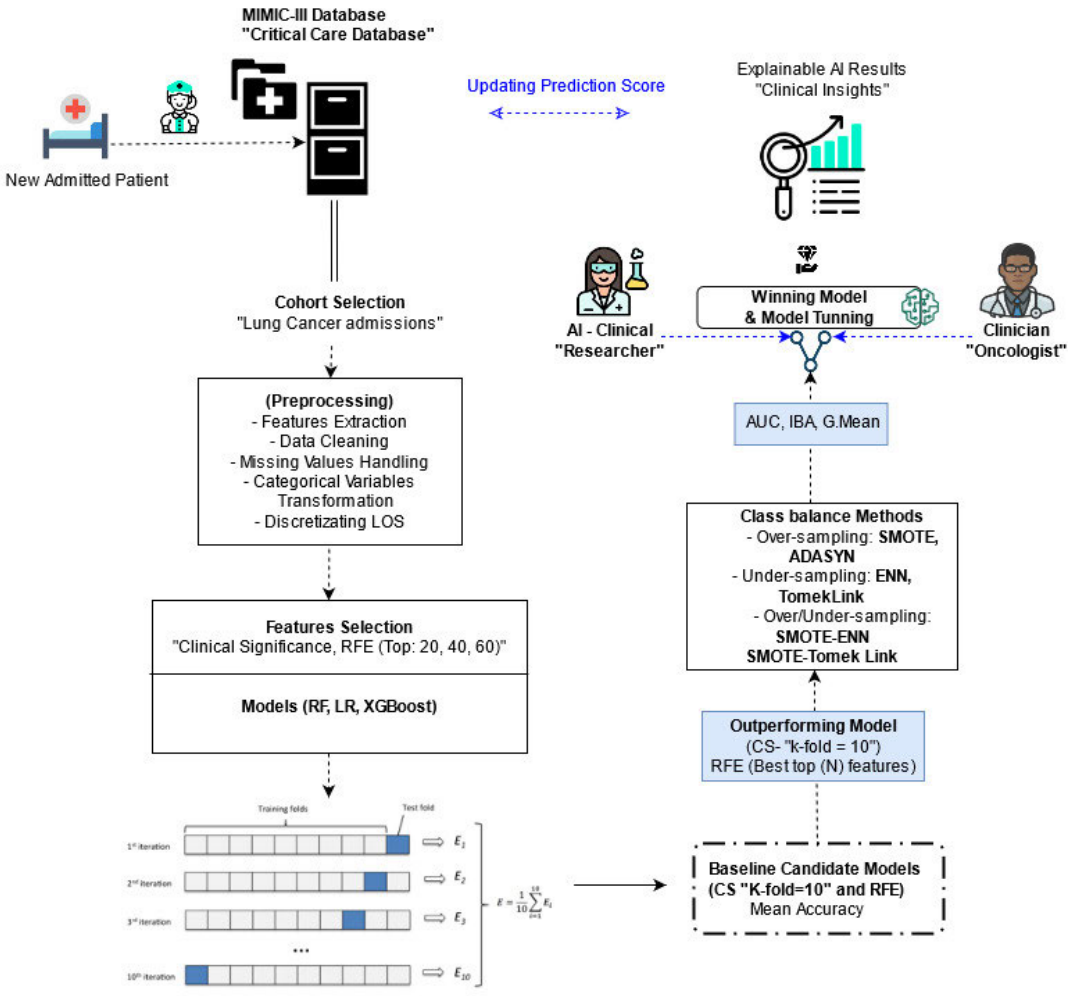


Figure 22. Lung cancer LOS predictive framework in ICU settings

4.5.3.3 Categorical Variable Transformation

The one-hot-encoding or “nominal encoding” method [131] was implemented in this chapter to transform (independent) categorical variables before building the LOS predictive models.

This method aims to transform categorical to nominal and binary attributes that improve machine learning models' performance.

4.5.3 Features Selection Techniques

The feature selection procedure is substantial in feature engineering to identify and select a subset of input variables (attributes) most relevant to the target class. In this experiment, two feature selection techniques were considered. The first is Clinical Significance (CS) and the second one is Recursive Feature Elimination (RFE). The following subsections discuss both approaches to the features selection processes.

4.5.3.1 Features Selection with Clinical Significance

The non-clinical significant variables were disregarded from the inclusion criteria. The inclusion and exclusion decision for the significant clinical variables was affirmed by the clinical oncologist. This decision was necessary for the study. Firstly, it ensures all clinical features of lung cancer are considered, and the non-important features are eliminated. Therefore, the feature selection (CS) puts the patients on the length of stay prediction perception from a clinical perspective. Secondly, it helped to reduce the features' dimensionality and improve machine learning models' performance in the baselining stage. The disadvantage of the approach is that it may leave weak associations between independent and dependent variables and impact the performance's predictive models. AI-based cancer studies [153, 154] utilised features selection with the CS approach in machine learning predictive tasks. Table 10 shows variables' selection with a clinical significance approach.

4.5.3.2 Features Selection with Recursive Feature Elimination (RFE)

The principle of the RFE technique [155] is based on selecting features recursively. This is achieved by removing a smaller set of attributes per loop. This process occurs recursively, and the weakest features are eliminated at the end. The features are ranked by the model's coefficient (coef) or feature importance. The optimal set of features is attained using cross-validation. RFE has been utilised in cancer-based studies such as in [156-158]. RFE is achieved using the algorithm in Figure 23.

```
Pre-processing data
Input: load extracted dataset MIMIC-III (d^e):
Output: cleaned & processed dataset (d^p)
Function (CTM):
   For each row R in d^e:
        If R in d^e contains:
          -invalid & expired admissions
```

-missing records > threshold value (th_v)
-non-cancer events (e)
end if
End For
dataset (d^c) ← Return and save R in the cleaned dataset

Data Discretization and Transformation
Input: load dataset (d^c)
Output: Discretized dependent (y: LOS) – Transformed set of independents features x= [x_1, x_2, . . .x_n]
Function (DLOS):
For each row R_y in y:
If y≥0 & <cut_point:
Label R_r = 0
else:
Label R_r = 1
end if else
end For
Function (TLOS):
While (F **is not** Continuous):
For each Categorical Feature F in d^c:
Transpose F← f = $(f_1, f_2, f_3, \ldots, f_n)^T$
end For
Break while
Features selection with Recursive Feature Elimination (RFE)
Inputs:
dataset d^c
Set of x features F= [x_1, x_2, . . .x_n]
Ranking model M(D, F)
Procedure:
"Recursive loop"
For i in [1 to x]:
Rank F with M(D, F)
f^* ← last ranked feature in F
R (x – i +1) ← f^*
F ← F - f^*
Output:
Optimal Ranked Features (R)
where:
CTM is the function to clean & treat missing values
DLOS is the function to discretise the continuous LOS
TLOS is the function to transform categorical features
(th_v) is the threshold of missing values in each R= 70%
cut_point: the cut point of 7+ days
Fis a categorical feature with ($f_1, f_2, f_3, \ldots, f_n$) layered f distinct values

Figure 23. Pseudocode for LOS lung cancer research framework

4.5.4 Class-Balancing Methods

In binary class predictive tasks, one class may dominate other classes. This occurs when there are many classes, and one class has the vast majority of the observations for the target predicted within itself (e.g., short LOS). A problem arises when the distribution of classes is skewed

(biased) and may still persist regardless that class distributions show minor, intemperate or imbalanced skew. Particularly, imbalanced data create extensive challenges to ML modelling, especially when ML models are meant to deal with assumptions of an equal number of samples for each class. Ignoring this problem may cause errors in minority classes. Hence, ML models might become prone to classification errors and ignore the minority class observations. Therefore, treating this issue is vital to ensure that ML predictive models provide reliable results in electronic medical records (EMR) domains. In this experiment, to overcome these challenges with binary predictive tasks, this experiment refers to the majority class as short LOS, and the minority is mentioned as long LOS. It also employed six class balancing techniques and compared their performance in a binary class predictive task. The six class-balancing techniques are described as the following:

4.5.4.1 SMOTE

The Synthetic Minority Oversampling Technique (SMOTE) is an oversampling technique applied in imbalanced datasets in classification problems. The SMOTE is an over-sampling method in which the minority class involves creating synthetic elements of minority class examples based on the existing ones. It picks up a point from the minority class and calculates the nearest neighbours by the Euclidean distance between data points in the feature space.

4.5.4.2 ADASYN

The Adaptive Synthetic (ADASYN) algorithm [159] is an oversampling method similar to the SMOTE technique. It works by generating many samples for a given feature vector x_i. The x_i is proportional to a number of the nearby samples, and it does not belong to the same class as x_i. This helps to deal with outliers, especially when generating new synthetic training examples.

4.5.4.3 Edited Nearest Neighbours (ENN)

The under-sampling with Edited Nearest Neighbours (ENN) method applies the nearest-neighbours algorithm. It edits the dataset and removes the samples from the dataset that do not agree enough with their neighbourhood [160]. In ENN, the nearest neighbours are computed, and if the selection criterion is not satisfied, the sample is removed. This process of removing noise from samples ensures that each sample in a class is under-sampled.

4.5.4.4 Tomek Links

Tomek Links Algorithm [161] is an under-sampling method that detects a pair of observations (two different samples from different classes x and y) near to teach others. The pairs are called Tomek links. The Tomek link is defined for any sample *z*.

4.5.4.5 SMOTE+ENN

The SMOTE+ENN algorithm [162] is a hybridisation technique (combination of over and undersampling) where SMOTE helps to do extensive data cleaning. Further, the misclassification caused by NN (nearest-neighbours) samples is removed [163] in both classes (short LOS and long LOS). This results in achieving a clear, concise separation between classes.

4.5.4.6 SMOTE-Tomek

The SMOTE+Tomek algorithm [162] is a hybridisation technique (combination of over and undersampling) that combines the oversampling (SMOTE) and the undersampling (Tomek-Links) techniques in order to achieve optimised performance for the classifier. The SMOTE+Tomek first applies SMOTE to the minority class (e.g. long LOS) to a balanced distribution, then the examples from the majority class (short LOS) in Tomek links are identified and removed.

The six class balancing methods are achieved in the following algorithmic steps in Figure 24:

Input: dataset (D); majority sample (S_{maj}) "S-LOS", and minority sample (S_{min}) "L-LOS"

Output: balanced dataset

Oversampling

SMOTE (X_{new}) steps:

1. Identify feature vector (x_i) and its nearest neighbor's (x_{zi}), where u is a number, and it is randomly chosen from U (0,1).
2. Take the two difference between (x_{zi}) and (x_i).
3. Multiply the difference ($x_{zi} - x_i$) with Random number (u) between (0, 1).
4. Identify a new point on the line segment by Adding the random to feature vector
5. Repeat the process for the identified features.
6. The new generated SMOTE synthetic example is obtained as follows:

$$X_{new} = x_i + u * (x_{zi} - x_i)$$

ADASYN (S_i) steps:

1. Calculate the ratio of majority to minority ($d = \frac{m_s}{m_l}$). where d ∈ (0,1], (m_s) is the minority class and (m_l) is the majority class.
2. Initiate the algorithm when ($d < d_{th}$), where d is lower than a certain threshold (d_{th}).
3. Calculate the total number of the synthetic data sample to generate the minority class (G) to generate:
$$G = (m_l - m_s) * \beta$$
where λ ∈ (0,1] (random number) and $(x_{zi} - x_i)$ is the difference in *n* denominational space.
4. Find *K* (nearest neighbors) for each $x_i \in$ *minority_class* according to Euclidean distance in *n* denominational space, then calculate the ratio (r_i):
$$r_i = \frac{\Delta_i}{K}$$
where (Δ_i) is the number of samples in the *K* nearest neighbors of x_i that belongs to the majority class. $r_i \in (0,1]$
5. Normalise (r_i) values in which the sum of all r_i values are equal to:
$$r_i = \frac{r_i}{\sum r_i}, \sum \hat{r}_i = 1$$
where $\hat{r}_i$ is the density distribution.
6. Calculates the number *G* of synthetic data examples that are required to be generated for each minority example x_i.
$$g_i = \hat{r}_i * G$$
7. For each minority class data sample (x_i for each neighborhood, generate $g_i$synthetic data sample;
1) Randomly choose one minority data sample (example) within the neighborhood (x_{zi}.
2) the new generated synthetic example is achieved with the following equation (ADASYN):

$$S_i = x_i + (x_{zi} - x_i) * \lambda$$

Undersampling

ENN (*N*) steps:

1. Obtain k nearset neighbour of x_i, $x_i \in N$
2. Remove x_i if number of neighbours from another class is dominant.
3. Repeat process for every majority instance of subset N

Tomek Link for sample (*z*) steps:

1. Tomek links identification
d(x,y) < d(x,z) and *d(x,y) < d(y,z)*
where *d(.)* is the distance between the two samples
2. Borderline and noise examples removal between the two samples

Over/Under-sampling

SMOTE-ENN steps:

1. Over-sampling using SMOTE

2. cleaning using ENN

SMOTE-Tomek steps:

1. Over-sampling using SMOTE
2. cleaning using Tomeklink

Figure 24. Pseudocode for the six class balancing methods

Table 11. The benefits and drawbacks of

the class balancing technique in predictive machine learning tasks

Technique	Benefits	Drawback
SMOTE	-Increases the feature availability to each class and prevents information loss. - May overcome the issue of overfitting of datasets [164]. -Improved performance in low-dimensional data [165].	- Increases training time and memory to hold training data [166] (computational cost).
ADASYN	Enhances the learning about the distribution of the sample in a more efficient way [167] - In minority and majority classes, it does not sacrifice one class in the preference for another [159].	- Risk of generating many false positives due to the generated synthetic data may be very similar to the majority class.
ENN	Reducing the number of training data samples and improving storage and model run time. - Removes unwanted overlaps between classes [166].	Ignore useful information which might be important in building a rule model.
TomekLinks	- Treat outliers efficiently. - Removes unwanted overlaps (points) between classes [164].	- low performance in the binary class prediction - Many samples are removed if the decision boundary is unclear.
SMOTE-ENN	Good performance in small datasets [162] - Adjust class distribution [166].	Can remove more examples in-depth
SMOTE-Tomek	Good performance in small datasets [162].	introducing artificial minority class examples too deeply in the majority class may lead to overfitting

4.5.5. Predictive Methods

The research target predictive class is binary; therefore, it is a classification problem. Suitable predictive classification algorithms are selected considering their robustness in binary prediction problems. Subsequently, the overall performance, including mean accuracy, precision, sensitivity, specificity, f1-score, IBA, G-mean and AUC scores, are evaluated. Finally, the model tuning stage was applied to the outperforming model.

4.5.5.1 Random Forest

Random Forest (RF) algorithm [23] is an ensemble learning model and classification-based method. The RF model works by generating random subsets from the original dataset (bootstrapping). Then, in each node in the decision tree, only a random set of features are to be considered for deciding the best split. After that, a decision tree model is fitted on each of the subsets. The final output (prediction) is achieved by calculating the average predictions from all decision trees. To summarise, the model operates by randomly selecting data points and features and then building multiple trees (forests). The RF classifier was appropriated in this study for the LOS lung cancer predictive framework with Gini Index (IG [168]) is implemented:

$$I_G = 1 - \sum_{i=1}^{C} (P_i)^2 \qquad (1)$$

where *Pi* is the proportion of samples that belongs to a class (C) for a particular node.

4.5.5.2 XGBoost

The eXtreme Gradient Boosting (XGBoost) algorithm [28] is an ensemble-based learning (Boosting) model. The XGBoost is an implementation of the gradient boosted decision trees [29] designed for performance and speed. It uses more regularised model formalisation to control the overfitting, giving it better performance [28].

Considering dataset (d^c) with m features and n of examples, where $d^c = [(x_i, y_i)]$ ($x_i \in R^m$, $y_i \in R$, $i = 1,2, \ldots, n$), the XGBoost model can be described as the following [169]:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) , f_k \in F (i = 1,2, \ldots, n) \qquad (2)$$

$F = \{f(x) = w_{q(x)}\}(q: R^m \rightarrow \{1,2, \dots T\}, w \in R^T)$ is the CART decision tree structure set, q is the tree structure of the sample map to the leaf nodes, T is the number of leaf nodes, and w is the real score of leaf nodes.

When constructing the XGBoost model, finding the optimiser is necessary to establish an optimal model. Therefore, the objective function of the XGBoost is divided into an error function L term, and a model complexity function Ω. Then, the objective function is written as the following:

$$obj = L + \Omega \quad (3)$$

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (4)$$

$$\Omega = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (5)$$

where γT is the regular term of $L1$, $\frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$, is the regular term of $L2$. γ and λ are adjustment parameters to prevent the model from overfitting.

Now the objective function is expressed as:

$$Obj^{(t)} = \sum_{i=1}^{n} \left(y_i - \left(\hat{y}_i^{(t-1)} + f_t(x_i)\right)\right)^2 + \Omega \quad (6)$$

where $\hat{y}_i^{(t-1)}$ is the predicted value of t-1th model and $f_t(x_i)$ is the new function added at tth time. The *Obj* is a scoring function that is used as an evaluation model, noting that the smaller the *Obj* value, the better the model effect.

4.5.5.3 Logistic Regression

The logistic regression (LR) [80, 170] is a statistical model that uses the logistic function to predict dependent variables from the independents used. It is used in machine learning in predictive binary tasks (classification). The logistic function is formulated as the following:

$$logistic(n) = \frac{1}{1+exp\,(-n)} \quad (7)$$

Some literature review studies scrutinised ensemble-based models (e.g., RF) in predicting LOS in clinical settings, such as [50, 125]. The LR is currently in used LOS, such as [96, 171] predictive problems. The XGboost model has not been examined in the available literature with LOS predictive tasks to the best of our knowledge.

4.5.5.4 Comparison of the Predictive Models

In this chapter, the RF model (Bagging) is assessed and compared to other prominent classifiers such as XGBoost (Boosting) and Logistic Regression. The outperforming model is to be

selected as the winning model for the LOS lung cancer framework evaluation in class-balancing and model clinical explanation. Table 12 compared RF, XGBoost, and LR via their pros and cons.

Table 12. LOS Lung cancer models' description

Model	Model's Advantages	Model's Disadvantages
Random Forest (RF)	- Robust against overfitting in decision trees. - Works well with categorical and continuous variables. - Desired predictive performance in many contexts.	- Complexity (creates a lot of trees) to make a final predictive decision. - Computational cost (longer training time)
Xtreme Extreme Gradient Boosting (XGBoost)	- Less feature engineering requires (can handle missing values and normalisation). - Less prone to overfitting. - Speed and performance.	- Difficult in interpretation. - May lead to overfitting if hyperparameters are not tuned correctly.
Logestic Regression (LR)	- Easy to implement and fast to train. - Interpreted model's coefficients as an indication of feature importance.	- Suffers from overfitting in high-dimensional datasets. - Not able to work with nonlinear problems

4.5.6 Predictive Methods Performance and Evaluation

A set of evaluation metrics was used to evaluate the predictive models' performance. The accuracy (mean cross-validation "k-fold" accuracy, Figure 22), Index Balanced Accuracy (IBA), Geometric Mean Score (GMS), Precision, Sensitivity, Specificity, F1-score, and Area Under the Curve (AUC) are used in this research. Following formulas are applied during the three main phases of framework performance evaluation.

Accuracy:

Donates the ratio of the correct predictions to the total of a number of predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

where TP: True Positive, FP: False Positive, FN: False Negative, TN: True Negative.

The accuracy does not describe the predictive story in imbalance-class datasets. Therefore, other metrics are used to evaluate imbalanced data, such as Precision, Sensitivity, G.mean and IBA.

Precision:

Refers to the number of positive classifications that are actually correct or called positive predicted value '**PPR'**.

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

***Sensitivity (Recall)*:**

Measures the proportion of actual positives that is well classified or called the true positive rate '**TPR'** or Recall.

$$Sensitivity = \frac{TP}{TP+FN} \quad (10)$$

***Specificity*:**

Measures the proportion of actual negatives that is well classified (true negative rate '**TNR'**)

F1-Score:

It can be interpreted as the weighted average of precision and recall. F1-Score = 1 is the best possible value, and F1-Score close to 0 is the worst value.

$$F1 - Score = \frac{2*Preision*Recall}{Precision+Recall} \quad (11)$$

Index Balanced Accuracy (IBA)*:*

The Generalised IBA [172] is used to weigh a suitable measure to evaluate an imbalanced datasets' performance. The weight factor assists in favouring those results with better classification rates for the minority class. The following formula calculates the IBA:

$$IBA_{\alpha}(M) = (1 + \alpha.Dom).M \quad (12)$$

where *Dom* is the dominance; *Dom = TPR – TNR* within the range [-1, +1]

Dom is used to estimating the relation between TPR and TNR. The closer Dom is to 0, the more balanced both individual rates are achieved.

It is weighted by $\alpha \geq 0$ to reduce its influence on the results of the particular metric M

$1 + \alpha . Dom$ is the weighting factor

Area Under the ROC Curve (AUC)

The AUC measures the quality of the model's predictions regardless of what classification threshold is chosen. It represents the area under the ROC curve plots (TPR vs. FPR):

$$\text{TRP} = \frac{TP}{TP+FN}, \qquad \text{FPR} = \frac{FP}{FP+TN} \tag{13}$$

where TPR is the true positive rate, and FPR is the false positive rate.

Geometric Mean Score (G.Mean):

The G.mean [173] aims to maximise each of the classes' accuracy while keeping the accuracy balanced.

$$G.mean = \sqrt{TPR * TNR} \tag{14}$$

This study refers to accuracy in the models' benchmarking (baselining) performance evaluation, whereas Precision, Sensitivity, Specificity, AUC, IBA, and G.mean refer to class-balancing performance evaluation metrics.

4.6. Results

4.6.1 Experiments Setup

The setup for the experiments on LOS lung cancer framework evaluation was conducted on a computer with (Intel core i7 8th Gen), 8 core CPU with a speed of 1.90GHZ, and 16 GB RAM. Python 3.6 was used for machine learning models and for building, executing, and evaluating all framework steps, including development and deployment.

4.6.2 Baseline Stage with Cross-Validation

The first phase of the experiments was the benchmarking stage (models-baselining) with cross-validation (k-fold=10). In this phase, three proposed predictive models (RF, XGBoost, and LR) were assessed on feature selection methods (CS) and the RFE with three varieties (Top 20 features, Top 40 features, and Top 60 features). The outcome of the first phase was the model's mean accuracy => of 85%. The second phase incorporates evaluating the performance of the candidate model using six class balancing methods. The research framework integrates all these

phases in the pipeline and fits the outperforming model for further clinical interpretations using the SHAP machine learning interpretability (third phase).

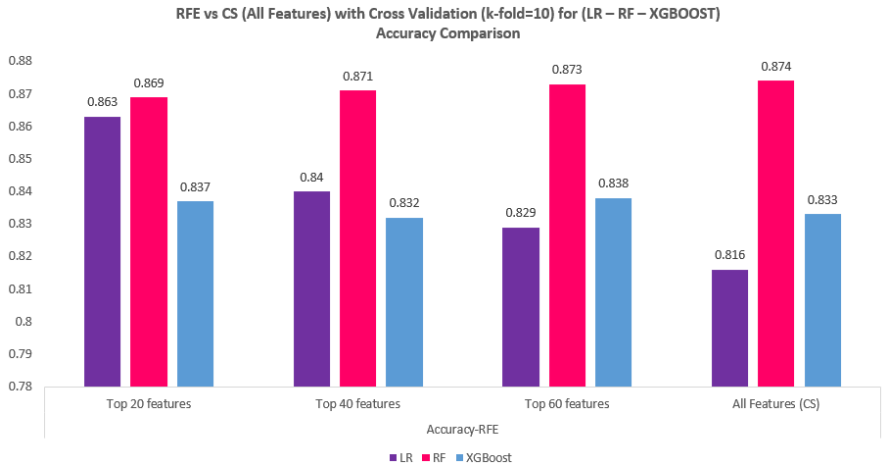


Figure 25. Accuracy comparison between RFE and CS (All Features) with cross-validation (k-fold=10) for (LR – RF – XGBoost)

A total of 119 unique lung cancer patients met the inclusion criteria (Figure 20) in this experimental sample. The cohort selection was verified on the proposed models (RF, XGBoost, and LR) using cross-validation (k-fold=10) with the mean accuracy comparison and standard deviation (std) error in the mean performance as the performance evaluation metrics. Baselining analysis results are reported in Figure 25 and Figure 26. As seen in Figure 25, the RF achieved the best predictive results with (k-fold =10, mean accuracy 87.4%) by ensuing the CS feature selection procedure. Moreover, the RF classifier attained with RFE (top 60 features) the highest mean accuracy (87.3%) amongst the RFE model-based top features' selection procedure.

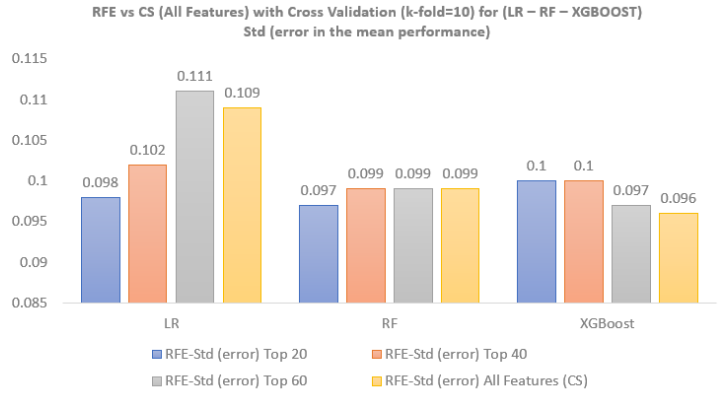


Figure 26. RFE vs CS (All Features) with cross-validation (k-fold=10) for (LR – RF – XGboost)

Std (error) in the mean performance

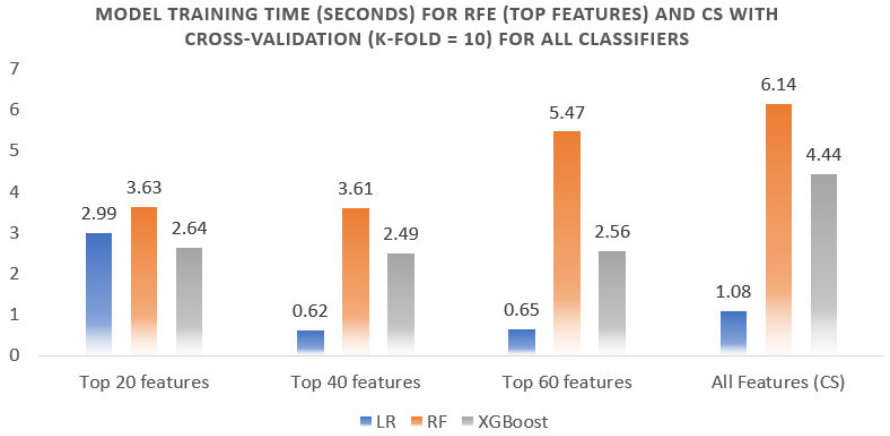


Figure 27. Time to execute model with cross-validation (k-fold = 10)

The XGBoost classifier showed a minor fluctuation, with mean accuracy ranging from 82.3% to 82.8% using two different feature selection methods (RFE and CS). In contrast, the logistic regression classifier's mean accuracy performance retreated as more features were built up in each RFE evaluation with (k-fold = 10) metric. This trend was affirmed within all features (CS), where the LR achieved the lowest mean accuracy (81.6%).

The std error in the mean performance for RF remained stable (9.9%) with RFE top features (40, 60) and all features (CS), respectively. However, the RFE (Top 20 features) evaluation had the least std error in the mean performance (9.7%). XGboost classifier showed an improved trend in the reported std error in the mean performance. While it recorded 10% of the std error in RFE (Top 40 features), it achieved an optimized (std = 9.6%) in the CS feature selection procedure with (k-fold = 10). LR classifier acquired relatively higher std error in the mean performance compared to other RF and XGboost.

While LR is the fastest model to train, the XGboost and RF needed more time to report their cross-validation results (k-fold=10) (Figure 27). Hence, RF and XGboost are more computational costly models according to the data input and number of features in this study. In analysing reported results during the baseline phase, RF is designated for further detailed analysis in the next phase (class-balancing) of performance evaluation.

4.6.3 Class Balancing Results

Various class-balancing methods were exploited in the second phase to compare the RF model performance on each technique. Six class balancing methods were fitted into the RF classifier to evaluate their performance (AUC). Clinical significance (CS) with all features (75 features) is used in the lung cancer subset. Selection procedure RFE with the top 60 of the RF model-based features selection technique is exploited in this chapter to predict the short and long length of stay with imbalanced data.

Table 13 demonstrates the lung cancer dataset before applying imbalanced data (WCB) and utilising various class balancing techniques. The second column in the table illustrates the short LOS percentage to the long LOS per each class-balancing approach. Better predictive performance is anticipated to be achieved from the balanced data. Specific metrics (section 4.5.6) were being used during classification performed on balanced dataset evaluation measures.

Table 13. Percentage of short LOS to long LOS in the lung cancer dataset, benefits, and drawbacks of each method

Approach for LOS Predictive Models Evaluation	**Percentage of Short LOS to Long LOS per approach**
Without class balancing (WCB)	87.4% - 12.6%
Over-Sampling	
SMOTE	50% - 50%
ADASYN	49.66% - 50.34%
Under-Sampling	
ENN	88.1 % - 11.9 %
TomekLinks	89.88% - 10.12%
Combination of Over-and Under-Sampling	
SMOTE-ENN	46.3% – 53.7 %
SMOTETomek	50%-50%

The RF model successfully predicted the short LOS and the long LOS with the highest reported predicted IBA score 100% and 100% for sensitivity and specificity using the class-balancing

oversampling technique (ADASYN) in both CS and RFE subsets. SMOTE obtained an IBA score of 96% with 98% sensitivity and specificity as the second most desired result in class-balancing with the oversampling approach (for CS and RFE) subset. Undersampling methods followed an opposite trend, while they attained a 0% IBA score for ENN and TomekLinks, respectively, following (CS and RFE) in the feature selection procedures. Furthermore, the specificity dropped drastically, and both class-balancing methods reached the lowest specificity scores (11% and 4%) correspondingly in both approaches (CS and RFE).

Table 14. A Comparison between Class-Balancing Methods using Clinical Significance Features Selection (CS) and Random Forest Model.

Method	Precision	Sensitivity	Specificity	F1-Score	AUC	IBA	G.mean
Oversampling							
SMOTE-CS	98%	98%	98%	98%	98%	96%	98%
ADASYN-CS	100%	100%	100%	100%	100%	100%	100%
Undersampling							
ENN-CS	80%	89%	11%	85%	50%	0%	0%
TomekLinks-CS	92%	96%	4%	94%	50%	0%	0%
Combination of over-and under-sampling							
SMOTETomek-CS	98%	98%	98%	98%	98%	96%	98%
SMOTE-ENN-CS	97%	97%	98%	97%	97%	94%	97%

The combination of over and undersampling methods (SMOTETomek and SMOTE-ENN) reported the same results in the CS and RFE approaches. SMOTETomek gained high desired predicted outcomes with 96% IBA score and 98% for sensitivity and specificity evaluation metrics. The SMOTE-ENN showed compacted results for both classes (short and long LOS) with 94%: IBA, 97% sensitivity, and 98% specificity, respectively.

Table 15. A comparison between Class-Balancing Methods using Features Selection (RFE=60) method and Random Forest Model.

Method	Precision	Sensitivity	Specificity	F1-Score	AUC	IBA	G.mean
Oversampling							

SMOTE-CS	98%	98%	98%	98%	98%	96%	98%
ADASYN-CS	100%	100%	100%	100%	100%	100%	100%
Undersampling							
ENN-CS	80%	89%	11%	85%	50%	0%	0%
TomekLinks-CS	92%	96%	4%	94%	50%	0%	0%
Combination of over-and under-sampling							
SMOTETomek-CS	98%	98%s	98%	98%	98%	96%	98%
SMOTE-ENN-CS	97%	97%	98%	97%	97%	94%	97%

The confusion matrix table (Figure 28) was used in this study for further evaluation and described the RF classification model's predictive outcomes on the testing dataset. In Figure 28, TP is donating the percentage of the number of samples that have been correctly predicted. The TN refers to the percentage of the outcomes where the RF correctly predicted the true negative (short LOS/long LOS). In contrast, FP denotes the percentage of the RF predicted outcomes. The model incorrectly predicted the positive class, and the FN is the outcome when the RF model incorrectly predicted the negative class (short LOS/long LOS).

As seen (Figure 28), the oversampling method (ADASYN) successfully predicted the TN and TP 56.52 and 43.48%, respectively, for the short LOS and long LOS classes. Further, ADASYN did not commit any false predictions (FP or FN). Similarly, SMOTE (oversampling) efficaciously predicted (short/long) LOS classes with TP: 44.44% and TN: 53.33%. The RF-SMOTE rate in the FP was minimal, as well for FN (0%). Nonetheless, the undersampling class-balancing methods reported unreliable predicted results with TN 89.47% and 95.83% for ENN and Tomeklinks, respectively. TP percentages for both class balancing methods were very poor (0%) for each of them. On the other hand, the combination with over/undersampling methods such as SMOTE-ENN and SMOTE-Tomek revealed desired results and that TN and TP rates were equal (48.89%) in the SMOTE-Tomek and 41.94%-54.84% in the SMOTE-ENN. Both techniques reported low rates for FN with (1 prediction) for both models and (0%) in the case of FP.

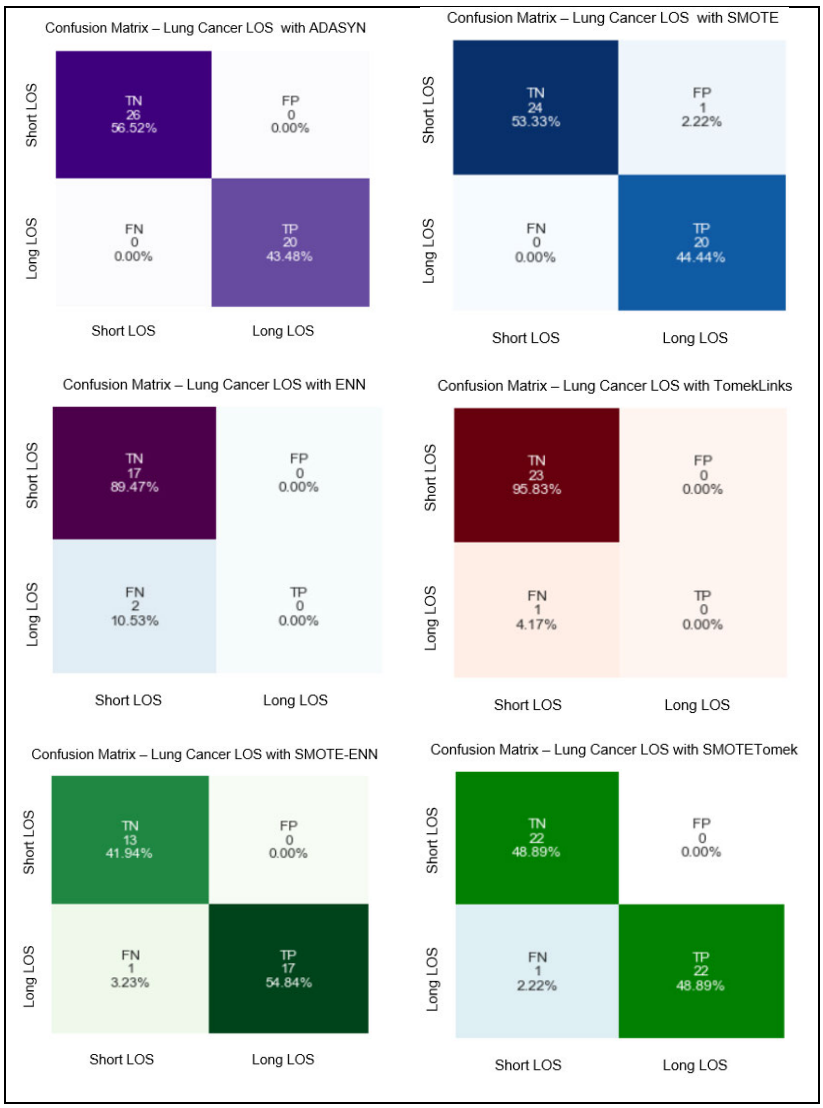


Figure 28. Confusion matrix for class-balancing techniques for lung cancer LOS with CS.

4.7. Discussion

The experiments on ML algorithms in this chapter demonstrated that ML application on LOS for ICUs of lung cancer patients could help set a sound framework and process for LOS and ICU admission with quantifiable benefits in hospitalisation resources records and management. Furthermore, the study's outcomes present several important findings on classifiers, class-balancing methods, oversampling and their combination and balancing techniques.

Three classifiers (Random Forest, XGBoostand Logistic Regression) were found to be excellent for benchmarking through cross-validation methods (k-fold = 10). The XGBoost classifier results showed a mean accuracy performance that fluctuated among different RFE top feature sets and CS feature selection methods. Nevertheless, its logistic regression was the

fastest model to train, whereas the RF and XGBoost models both needed more time to solve their prediction outcomes. Further, it was observed that the RF results were the most computational costly model among all three of them. The RF showed resistance to any changes in the features' selection varieties, such as the CS and RFE with the various top features' approaches. Unlike other models, RF improved with the most features when added to each (RFE) feature selection performance attesting with cross-validation (k-fold = 10). Thus, RF had a robust performance and reported stable results with different feature selection methods. The RF was nominated as the winning model for the class-balancing stage.

The six class-balancing methods with RF classifier were tested and evaluated. This experiment aims to examine the most robust class-balancing approaches (oversampling, undersampling, or the combination of both). AUC is reported to measure the quality of the model's prediction in each class balancing method (how well the RF model can distinguish between LOS classes, "short vs long"). CS outcomes (AUC scores for six class-balancing methods) are observed as the same as RFE. Hence, the CS approach is referred to as discussing and reporting results with class-balancing (AUC) performance measures. Oversampling reported the best AUC scores (100% and 98%) for ADASYN and SMOTE. Correspondingly, the combination of both "SMOTETomek-CS" and "SMOTE-ENN-CS" came up as the second-best approach with 98% and 97%, respectively. As opposed to the oversampling or the combination approach, the undersampling presented the weakest AUC results (50%) for both TomekLinks and ENN. It indicated that the two methods are incapable of differentiating between the two classes (short LOS and long LOS), resulting in lower and undesired performance. Subsequently, undersampling methods are not suitable for predicting inpatients' length of stay. Eventually, TomekLinks and ENN were assuredly disregarded from LOS predictions in binary class problems.

The oversampling and combination of oversampling and undersampling presented high predicted AUC results for the short LOS and the long LOS. Both class-balancing approaches are considered for further clinical explanation to evaluate their clinical insights with the clinical oncologist. Further, to assess their feasibility and the clinical insights, they may induce the hospital resources' utilisation and hospital healthcare assessment systems in ICU.

The class balancing technique (ADASYN) results were the most successful predicted outcomes from the confusion matrix (Figure 28) on the test dataset. ADASYN distinguished two classes (short LOS and long LOS), where the RF did not report any false positive or false negative

predictions. The second most crucial result concerning the class balancing methods came from SMOTE, which showed the RF desired ability to efficiently differentiate between the two classes with only one minor false-positive prediction. Undersampling class methods (ENN and TomekLinks) produced weak predictive outcomes and unreliable performance, where both techniques provided high true negative ratios and zero outcomes for the true positive. Both methods reported noticeable false negative for how RF incorrectly predicts the positive class following both approaches. The SMOTE-ENN and SMOTE-Tomek are combined between class under/oversampling techniques, whereas their testified outcomes (true-positive and false-negative) are desired with minor incorrect predictions.

4.7.1 Explanations with Class-Balancing Using SHAP

Model explainability refers to how a human can consistently predict the model's results [174]. In the machine learning domain, the higher the explainability of a certain model, the better it is for someone to understand and comprehend the predictions that have been made. In a healthcare context, particularly in hospital clinical decisions or healthcare assessment systems, the length of stay continuum is important in decision-making [175]. In this chapter, the local explanation was followed to analyse the RF that is viewed as a black-box predictive model [176]. The local explanation approach determines what variables (lung cancer features) explain the specific prediction (LOS: short or long) of the Random Forest using the class balancing methods, as seen in Figure 29. The Random Forest prediction outcomes with the four class balancing methods (Figure 29) were unlocked and explained using the SHapley Additive exPlanations (SHAP) [177]. The SHAP works by explaining the prediction of instance x by computing each feature's contribution to the prediction. It is also referred to as a method to explain individual predictions. A TreeExplainer function [67] using the TreeSHAP [177] algorithm was exploited to visualise and explain the Random Forest (ensemble) tree model's output.

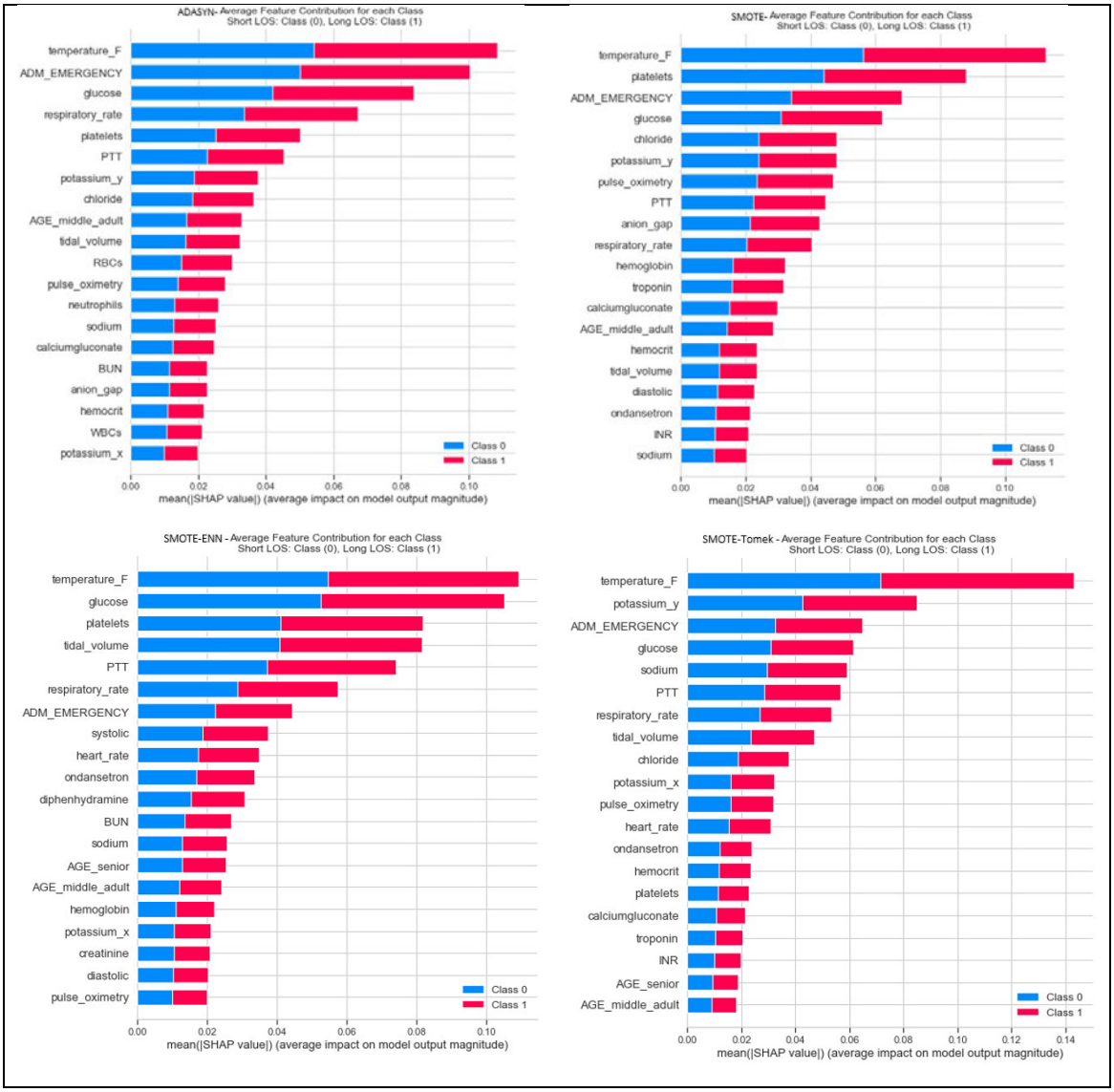


Figure 29. SHAP (mean value; the impact of each model's features on the model output magnitude for selected class-balancing methods with RF

The Random Forest SHAP explanations for the dependable class balancing algorithms (SMOTE, ADASYN, SMOTE-ENN, and SMOTETomek) are depicted in Figure 29. Top features such as temperature (F), Emergency Admission' ADM_EMERGENCY', Glucose, and respiratory rate (respiratory_rate) were highly explained and are the most highly ranked features. This confirms how the SHAP using RF could rank clinical variables based on features' importance with clinical soundness. While RF-SHAP (SMOTE-ENN) ranked (systolic) variable in the top features, the diastolic came in the least in features by importance in the list.

In terms of the SHAP (RF) model explainability for the class balancing techniques (Figure 29), it has been observed that the SHAP explanation of the SMOTE is more definitive in the real practice and contented for clinicians. Therefore, using the SHAP ranking (mean SHAP value) in this experiment can judge that the sequence of data for SMOTE and RF classifier features

by importance is more reliably related to the situation of the patients. For instance, patients sharing common demographic, diagnostic and laboratory features are supposed to require a similar level of resource utilisation; therefore, the SMOTE is expected to be efficiently able to quantify and standardise resource utilisation for patients during their hospital stay. Further, the provision for the LOS for the patients is achieved based on the SMOTE ranking of the lung cancer clinical variables. Thus, this verifies the RF model's suitability to predict and SMOTE reasonability for lung cancer LOS prediction in ICU.

4.7.2 Research Implication

Modelling LOS was chosen in this study because it acts as a primary reason for increasing cost. Accurate modelling of this outcome can help healthcare systems identify risk factors for unnecessary hospital days of stay, potentially reduce waste, provide more efficient allocation of medical resources, and potentially improve health care for patients. Such models could be used to build an application into the background of EHRs to determine predicted outcomes automatically. The model provides patients and families with information to aid in planning for work absences or care about discharge. Moreover, non-clinicians can utilise the predictions. For example, beds managers could ensure that adequate numbers of beds are available in intensive care units.

The lung cancer LOS prediction framework has several clinical research implications. Firstly, it can help the clinicians to apply more procedures and actions based on the SMOTE-RF scale (prediction), such as evaluating the patients' severity at the time of admission to determine the LOS. Secondly, the framework helps detect early clinical patient problems and critical situations requiring urgent intervention to accelerate medical decisions and treatments. Thirdly, it guides clinicians in the proper direction of performing the correct procedure and predicts the clinical course (anticipation) during ICU admission. Last but not least, the framework helps the junior doctors practising in the ICU to manage the patients' hospitalisation based on superior performance based on the methodological and experimented classifier (RF) and the appropriate class balancing method (SMOTE).

4.7.3 Study Limitations

Although the proposed framework in this chapter has many advantages, it also has some limitations to address. Firstly, external validation was not performed for the proposed framework on another medical dataset with similar characteristics to the current study due to the lack of accessibility to other hospital data. This prevented the study from verifying the

proposed lung cancer LOS predictive framework on real-world hospital data and attesting the class balancing technique performance, especially the SMOTE. The study reports the necessity for AI researchers and clinicians involved in AI studies to take further research steps while collecting data to consider enough numbers for admitted patients based on various diagnoses that demand ICU resources and hospital resources. This step should be liaised with clinical expertise to assist filter variables clinically significant to the selected diagnosis or disease. Furthermore, the clinical attributes need to be similar in their characteristics within both datasets that need to be examined in the proposed predictive framework in this study. Thus, the external validation should consider a real hospital dataset with similar attributes and characteristics as examined in the current study.

Secondly, the sample size in this study is relatively small due to the limited availability of the lung cancer diagnosis in the MIMIC-III 1.4 version dataset. This prevented evaluating the proposed models on a larger dataset for the same lung cancer diagnosis. Consequently, this did not exploit the advancement for deep learning techniques to predict lung cancer's LOS and find further clinical insights or associations between the clinical variables in the disease-centred approach using deep neural networks. Thirdly, the hyperparameter tuning procedure was not performed because of the small size of the lung cancer subset. This decision was made to avoid the risk of overfitting. Finally, the frame did not study the relative accuracy of the model compared to clinician estimates of LOS. Still, it is believed to be used as a guide for quality improvement initiatives. LOS indices, which compare expected to observed LOS, have been proposed as efficiency and hospital performance markers. Using patient-specific predicted LOS to measure expected LOS may improve the accuracy of such indices, allowing hospitals to generate more representative quality metrics and, in reimbursement schemes that incentivise quality care, avoid punishment for taking on higher-risk patients.

4.8. Chapter Conclusion

The chapter represents the potential of ML to predict the LOS of ICU cancer-based hospitalisation, in particular lung cancer patients, efficiently. Suitable class balancing methods were evaluated to deal with the imbalanced class problem, primarily challenging to the predictive modelling task because of the severely skewed class distribution in clinical health records data (clinical EHRs). The presented LOS research framework is the first in the literature review for lung cancer hospitalisation and predicting lung cancer patients' future days in ICU hospitalisations to the best of our knowledge. The framework provides a practical framework

to be exploited by clinical oncologists, hospital beds managers, and healthcare givers as a robust predictive and explainable artificial intelligence tool for lung cancer patients in ICU settings. Among the examined machine learning methods, features selection, and class balancing techniques, the Random Forest ensemble classifier has proven itself robust in different feature selection procedures (RFE or clinical significance).

Furthermore, the class balancing with oversampling such as ADASYN and SMOTE achieved the most outstanding AUC and G.mean results, followed by the over/and undersampling methods. However, undersampling methods did not achieve reliable results in terms of the AUC and D.mean metrics. The Random Forest and the outstanding class balancing methods are explained to non-artificial intelligence experts using the SHAP machine learning explainable method.

Future research aims to expand our framework and examine the most common cancers admitted to the ICU and group them based on the diagnoses that are likely to use more ICU resources. We will study the challenge of predicting LOS by focusing on patients (patient-centred) and examine the appropriate methods to achieve this challenge.

5. Chapter Five: Improving Hospital Resources Utilisation and Workflow with a Data-Driven Approach: Predicting Inpatient Length of Stay in ICU from Real Life Data

5.1 Chapter Summary

Managing hospital resources is vital for hospital resource management, especially in the intensive care unit (ICU). Efficient beds management leads to reducing hospital healthcare expenditure, increasing workflow effects, and improving patient outcomes. This research introduces predictive hospital-ICU length of stay to predict patients' stay at the time of admission. The framework uses electronic health records (EHRs) from real hospital database (Al-Ain Hospital, Al-Ain, UAE), and it contributes to the implementation of predictive models in clinical information systems (CIS).

Supervised predictive classification is used in this chapter to predict patients' length of stay who were admitted to ICU. A real hospital dataset was exploited to evaluate the performance of the machine learning approaches in the proposed ICU length of stay predictive framework. In addition, a set of performance measures (Accuracy, AUC, Sensitivity, Specificity, F1-score, Precision, Recall and more) were studied to assess the classifiers' ability to predict short and long ICU length of stay.

The introduced length of stay predictive framework can perform efficient predictive short and long classes with all machine learning models with the desired accuracy among the three stages within the framework. The XGBoost model was the best outperforming model with AUC 98% for predicting short and long LOS. The contribution in this chapter provides a practical potential for hospitals and ICU rooms to predict patients at the time of admission using machine learning models. Also, our work advances the research in clinical information systems for hospitals to provide a robust and trustworthy LOS predictive framework by utilising an explainable artificial (xAI) method to explain the predictive outcomes of the predictive models.

5.2 Background

The inpatient length of stay in hospitals is an indication often used to measure hospitals' efficiency. It is considered one of the most significant indicators for the consumption of hospital resource utilisation [178]. A shorter length of stay will reduce the cost per discharge and shift care from hospital inpatient to less costly post-acute settings [179]. The Australian National Health Performance Authority hospital report [180] revealed that information about the efficiency of hospitals and local health systems is associated with crucial factors to determine the hospital LOS. For instance, a shorter LOS is more efficient from a hospital perspective while making beds available more quickly to provide care for more patients and reduce healthcare costs associated with each patient. However, stays that are too short may reduce healthcare quality and lead to poorer patient outcomes. Contrarily, longer hospital lengths of stay are often due to complications and may be associated with a higher risk of adverse events. Furthermore, the longer LOS must be due to factors unrelated to the patient's clinical condition, such as the delays in consulting or coordinating healthcare care with other healthcare professionals who have a role in assisting the patient's recovery. Moreover, more extended stays can also occur if there are delays in assuring the patient is accepted into another health care service, such as aged care home, community care service and a rehabilitation facility, the report said [180].

Consequently, healthcare quality depends on the availability of an abundance of bed spaces and the ability to transfer patients to other hospital wards of moving them hours-by-hours and day-by-day. However, managing hospital bed availability and efficiency are mandatory to address frequent and working challenges in intensive care units (ICU), such as the issues associated with the overabundance of patients in ICU in the hospital. Moreover, to avoid any hurdles of ICU beds shortage, particularly in uncertainties like pandemics [110]. Most importantly, diminishing the risk of acquired infection during ICU hospitalisation, the risk of mortality [111] and medical complications for vulnerable patients are key aspects. An improved resources utilization could potentially mitigate and minimise these risks [112, 113]. Therefore, a lower ICU length of stay (LOS) with quality care is necessary to lower hospital charges. Subsequently, hospital resources are well-managed, and better outcomes are achieved for the patients [114].

The in-practice conventional ICU length of stay scoring hospital systems are adopted to estimate the ICU. For instance, the APACHE versions I, II, III, IV and SAPS [7, 8, 10, 115]

are currently used in the intensive care unit inpatients' length of stay. The APACHE and SAPS are benchmarking systems based on acute physiology and chronic health evaluation. While SOFA systems are used in predicting ICU mortality [116] and help to triage ICU admitted patients, it gives the estimation of ICU patients' stays based on risk categories. These methods rely on the patients' features and the ICU features to measure the patients' length of stay during hospital and ICU admissions. These methods' drawbacks include the inability to estimate the LOS such as SAPS or may require complex data collection; also, they are not disease-specific estimation methods, and there is consensus on the most suitable methods for ICU LOS [11]. For that reason, examining a reliable and accurate prediction mechanism for resource utilization in intensive care units is deemed an essential research task. It is vital to study the alternatives for the current traditional LOS scoring systems in ICU settings. It can be achieved by exploiting the advancement of artificial intelligence to provide more accurate methods to estimate inpatients' future length of stay than the current traditional systems in ICU settings.

Hospital healthcare management systems use hospital electronic healthcare records (EHRs) to facilitate their daily operational and medical procedures and the LOS inpatients' determination.

With substantial data stored in hospital management systems such as the hospital databases systems and the data linkage to clinical information systems (CISs) in ICU [181], the CISs allowed shortening the ICU length of stays without altering other patient outcomes [182]. Clinical decision support systems (CDSS) are integral components of the CISs [181]. The CDSS offers an excellent opportunity for researchers and hospital healthcare managers, healthcare givers, and government healthcare systems to exploit the EHRs data in a more sophisticated approach. The CDSS can be cost-effective for health systems through clinical intervention, decreasing hospital length of stay [183, 184]. Eventually, it helps hospitals in cost containment [185]. The Computerized Provider Order Entry CPOE is a form of CISs [181], where COPE implementation helped reduce the prolonged length of stay in ICU [186]. Consequently, the healthcare assessment systems store EHRs data associated with patients' encounters from demographics, diagnosis, laboratory tests, prescriptions, radiological images, clinical notes, and many more [117, 118]. The data generated from EHRs systems are large and yet exploited in scalable setups within hospitals and healthcare assessment systems such as the general CIS and the specified CDSS and CPOE systems globally.

5.3 Literature Review

Resources utilization in intensive care units is vital for hospital healthcare management systems (e.g. ICU-CIS). Hospital length of stay is an important metric to assess the quality of care in the ICU department and the hospital in general. This section discusses all relevant research studies that examined the LOS in ICU as a binary prediction problem as our research is motivated based on that. In this section, studies are compared according to specific criteria that fit with the aim of this research based on the study motivation and the research challenges in the domain of ICU-CIS length of stay for CDSS. The studies in the literature review are compared against the following criteria: 1) benchmarking machine learning models, 2) type of clinical information system, 3) prediction model(s), 4) models' explainability, 5) LOS prediction defined labels, 6) number of features and 7) evaluation methods/accuracy.

Ma et al. [126] studied extreme learning machines (ELM), which is based on the feedforward neural network (FNN), to create a personalised model for patients and determine the number of hospital stays in ICU. They compare just-in-time learning methods (JITL) with ELM to decide whether or not they can be discharged within 10 days. Their reported results showed that the combination of JITL and ELM showed an area under the curve (AUC = 0.8510) with a lift value of 2.1390, the precision of 1, and the G.mean is 0.7842 to predict if the patient is likely to stay with LOS <10 days or >=10 days. They compared one class JITL and ELM to one-class SVM (support vector machines), where the class JITL and ELM outperformed the one-class SVM with AUC 85.10% and AUC 46.47, respectively, for personalised patient care that is based on age categories. Their proposed solution did not examine the models' results' explainability of the predictions.

Work by Su et al. [187] examined three machine learning models, XGBoost, Logistic Regression (LR), and Random Forest (RF), against the ICU scoring system SOFA for predicting the length of stay of ICU-sepsis patients. Their binned LOS is into two labels (short LOS: ≤6days, or long LOS >6 days). They used the oversampling method SMOTE (chapter 4) to treat imbalanced data. The RF outdid other studied models, where it achieved (AUC=76%) compared to XGBoost (AUC=75%), LR (AUC 66%) and the baselined SOFA (AUC=62%). Their findings stated that the patients with long LOS-ICU (> 6 days) were older than 59.81± 16.59 years and that the ML models outperformed the traditional LOS scoring system (SOFA). Their study was limited to the data collection centre, and the regional factors in the data collection caused the biased nature of predictions. Moreover, the study was conducted only on

one specific disease (sepsis) and did not consider the robustness of the proposed models on other diseases in the dataset.

Staziaki et al. [188] conducted a study and compared the performance of the artificial neural network (ANN) and support vector machine (SVM) based models to predict LOS intensive care unit admission and extended LOS after trauma to the torso. They considered and compared the prediction of LOS on CT imaging (radiology reports), then on data input of the clinical parameters (age, sex, vital signs, clinical scores, and laboratory values) and the combination of (CT+clinical attributes) together. Their findings reported that the combination of CT and clinical data attributes (all features) significantly enhanced the prediction of both outcomes with either ANN or SVM. The SVM model (all features) predicted ICU-LOS (short LOS: ≤2 days, long LOS: > 2 days) admissions with (AUC = 87% ±0.03), and the ANN achieved AUC= 78% ±0.12. However, the study did not eliminate the noise in the data and cleaned non-trauma patients where these admissions can bias the predicted results. In addition, the radiologist was involved in adjudicating electronic radiology reports, and interpretation biases may impact the LOS predicted outcomes of trauma patients. Nonetheless, the authors intended to use convolutional neural networks to extract features used in their study directly from medical imaging without the need for any manual human labelling of the CT images.

Alghatani et al. [189] evaluated six classifiers to predict LOS (short LOS: ≤2 days, long LOS: > 2 days). The six classifiers (LR, RF, SVM, XGboost, LDA: linear discriminant analysis, KNN: k-nearest neighbour) were evaluated on eligible ICU admissions using MIMIC-III (v1.4) database [85]. A total of 33 features were used to predict the short and long LOS. The RF and XGboost classifiers outperformed other models (AUC = 69.78%, 69.69%) using the quantiles approach. They proposed their predictive approaches in a practical predictive framework called Intelligent Remote Patient Monitoring (IRPM). However, their system was limited to benchmarking the classifiers only on vital signs. Further, they did not explain the prediction decisions of the quantiles approach in an AI explainable approach.

Gentimis et al.[190] used ANN to predict the length of stay (short LOS: ≤5 days, and long LOS: > 5 days) using the MIMIC III database. They extracted 25 features from MIMIC-III tables (Admissions, CPT Events, ICU Stays, Services, Procedures ICD, and Diagnoses ICD). The reported results identified that ANN is able to predict LOS with 80%. However, the study

lacks important model performance metrics such as AUC, Sensitivity, or Specificity. These metrics are important to differentiate the model performance with accuracy and how likely the model can distinguish the decision boundaries to predict LOS short or LOS long effectively.

Steele and Thompson [191] utilised seven predictive models (Naïve Bayes, Bayesian Network, kNN, KStar, Locally Weighted Learning (LWL), C4.5 Decision Tree, and SVM) for LOS prediction. In addition, a real hospital dataset called HCUP Florida SID was used to predict the short LOS: <8 days and long LOS: ≥ 8 days. Bayesian Network (BN) achieved the best result among other predictive models with AUC=90%, and Naive Bayes ranked as the second most helpful model after BN. However, the study suffered from drawbacks. For example, it did not specify the nature of clinical, laboratory, vital signs collected to assess further models performance on more admission features considered a viable picture of patient's information to identify the short from the long LOS.

Table 16. Comparison criteria of related previous studies

Work	Location of Study	Benchmarking	CIS Approach	Prediction Model	Models' explainability	# LOS defined labels	Number of Features	Number of distinct events (Patients)	Evaluation methods for Winning Model(AUC, ROC, etc.)
Ma et al, [126]	China	✓	CDSS	ELM (FNN), SVM	x	2 (Short LOS: <10 days, Long LOS: >= 10 days)	360	4000 (records)	AUC (85.10%: JITL-ELM), Acc, lift, precision, G-mean, Sensitivity, Specificity
Su et al. [187]	China	✓	x	XGBoost, Random Forest, Logistic Regression	x	2 (Short LOS: ≤6days, or Long LOSL >6 days)	N/A	2224	AUC (76%: RF), F1-Score, Sensitivity, Specificity
Staziaki et al. [188]	USA	x	x	SVM, ANN	x	2 (Short LOS: ≤2 days, Long LOS: > 2 days)	N/A	840	AUC (87%: SVM), Acc
Alghatani et al. [189]	Saudi Arabia	✓	IRPM	LR, RF, SVM, XGboost, kNN, LDA	x	2 (Short LOS: ≤2 days, Long LOS: > 2 days)	12 (basline features) + 21 (engineered features)	45,254	AUC (69.78%: RF, 69.69%: XGBoost)
Gentimis et al.[190]	USA	x	x	ANN	x	2	25	50,000	Accuracy (80%: ANN)

						(Sort LOS: ≤5 days, and Long LOS: > 5 days)			
Steele and Thompson [191]	USA	✓	x	Naïve Bayes, Bayesian Network, kNN, KStar, Locally Weighted Learning (LWL), C4.5 Decision Tree, and SVM)	x	2 (Sort LOS: <8 days, and Long LOS: ≥ 8 days)	21	30,000	AUC (90%: BN, 89.9%: Naive Bayes)
Our Study*	Australia, UAE	✓	CDSS	XGBoost, RF, GB, LR, MLP (ANN)	✓	2 (Sort LOS: <7 days, and Long LOS: > 7 days)	475	1045	ROC (98%**: XGBoost)

*Please refer to the results and discussion in this work for more details on the achieved results ** XAL approach

Table 16 discussed the previous related studies based on the comparison of criteria, including Benchmarked ML models, CIS Approach, Prediction Model, Models' explainability, number of LOS classes, number of features, number of distinct events (patients), and evaluation methods for winning model. It was noticed that the related studies benchmarked LOS on different predictive models and setups to identify the most suitable or outperforming model(s). In addition, half of the studies identified the type of CIS systems in their study as the target healthcare management system to show where their proposed solution/framework fits into which category of CIS. Ensemble learners are favoured for the literature reviewed studies, and that is proven by their desired predicted health outcomes to identify short LOS from long LOS. The number of features used in each

study was subjected to the study design and goals. Most of the studies used the AUC metric to evaluate the predictive outcomes of the LOS for the short and long LOS. However, based on the literature review, it was confirmed that there were not any studies that explained the inner working of the predictive machine learning models. This is the most important inference from the comparison that machine learning models should explain how they make their decisions to clinical expertise or hospital managers (non-AI expertise). Then non-machine learning expertise can make their clinical or hospital administrative decisions when ML models can give them insights to support their decision and improve hospital workflow and resource utilisation. Therefore, models' explainability is identified as a research challenge that is not yet addressed in CIS systems. This motivates our research to study explainable AI approaches in LOS-ICU settings for ICU systems, besides the objectives of this research that are identified in section 5.4: Research expected outcomes.

5.4 Research Questions and Expected Outcomes

In the era of artificial intelligence and the use of data, machine learning is transforming clinical information systems, such as the studies of AI in CISs, in particular the CDSS systems, and Learning, Data Analytics, Predictive and Personalized Medicine (LDAPPM) [192]. The CDSS systems have been classified into two types: 1) knowledge-based; 2) non-knowledge based [185]. In CDSS, the non-knowledge based systems are the systems that require data sources, whereas CDSS leverages AI, ML and statistical learning to get insight and patterns from the data rather than being programmed to follow expert medical knowledge [193]. Although non-knowledge based systems are rapidly growing in using AI in hospital management systems and medicine, they are associated with challenges, including problems with understanding the logic of AI or ML use to produce recommendations (called alack-boxes). Therefore, they are yet to reach widespread implementations in CDSS and generally within CIS systems[185]. To this end, this raises research questions:

> **1.e** The first question is how to benchmark ICU-LOS predictive methods on real hospital data inputs and setups.
>
> **2.c** The second question is how the AI models can explain their decisions and the inner workings in ICU-LOS predictive tasks to unveil hidden knowledge in the data and provide trusted LOS predictions for the CIS and CDSS systems.

The first research question in this work will evaluate the various EHRs data inputs from the whole picture of inpatient hospitalisations (events) and the patient's entry profiles information, such as general admission, clinical, laboratory and medication. Then, we evaluate them in practical aspects within the LOS predictive framework. The second research question elevates to what extent AI models can be doable, provide correct predictions, minimise errors, and offer safer practical research on predictive LOS.

Both research questions in this chapter are an integral part of the thesis research questions 1 and 2.

The key contributions of this research are as follows:

- Proposed a practical data-driven predictive research framework to predict inpatient length of stay during hospital -ICU admission regardless of admission diagnosis code and types.
- This chapter provides models benchmarking approach within the LOS framework for improving LOS prediction tasks and improving hospital resource utilisation from a machine learning approach via clinical information systems.

- The proposed framework is a doable practical research implication that can easily fit into the CIS prediction pipelines.
- The proposed practical model's explainable prediction approach explains the inner workings of the predictive machine learning models and makes results more understandable for health workers in hospital settings.

The overall goal of this research is to introduce a practical predictive, and explainable ICU-LOS framework that can be a viable implementation for in-hospital healthcare assessments (e.g. CIS, CDSS, CPOE or even LDAPPM).

5.5. Methods

This paper discusses and evaluates the methods and algorithms used in the proposed predictive LOS framework for predicting LOS for patients during their admission to the hospital, particularly during their entry to the ICU department and discharging from the hospital. The machine learning algorithms used in this research are to predict the LOS of inpatients in a real hospital dataset. Therefore, this procedure is vital to assess the performance and validate the predictive models in real hospitalisation data. In this section, all steps involved in the predictive framework (Figure 31) are discussed thoroughly. Hence, the following section goes through each step of the framework meticulously.

5.5.1 Data Description and features extraction

Our retrospective study was conducted using electronic health records EHRs data from Al-Ain hospital. Al-Ain hospital is located in Al Ain City in UAE. Al-Ain hospital has a 402-bed acute care hospital with over 35 medical departments. In addition, the hospital's emergency and trauma centre serves the Al-Ain community by operating 24 hours a day, seven days a week, year-round. The extracted data from the Al-Ain hospital EHRs database is the period for all ICU admissions between 31/12/2017 and 3/4/2020. The nature of EHRs is de-identified, and removed all patients' details and identifiers to comply with data protection regulations in UAE and Australia. The study population is 1045 distinct patients admitted to Al-Ain hospitalisations within the data extraction period. The ethics committee approved this study of Al-Ain hospital and UAE University (ethics approval: AAHREC-09-20-027), and pre-existing amended approval by Western Sydney University (WSU) with the ethics number (H13511). Both ethics approvals for the Al-Ain hospital dataset and WSU ethics are available in chapter 1 Appendix.

The data consists of four tables: 1) The main ICU table contains general admission information such as admission units, discharge unit, admission type, primary diagnosis code and secondary diagnosis codes, gender, age at the time, and many more. 2) Clinical even table mainly consists of the clinical

procedure of the values associated with the experiments, such as body mass index, respiratory rate etc. 3) Laboratory information table consists of lab-obtained information from each patient during their stay at the hospital, such as insulin fasting, acetaminophen level, etc. 4) This table is the medication table for the patients during their stay at the hospital. Table 27 (Thesis Appendix) includes information related to medication such as amiodarone, atropine, etc. Table 28 (Thesis Appendix) provides more details about the attributes in each table and their descriptions.

The inclusion mechanism in this paper compromised all ICU hospitalisations at Al-Ain hospital. During the first screening, all patients who died in the hospital were excluded from the inclusion protocol (expired hospitalisations). Moreover, all hospitalisations with a high percentage of missing information (threshold is >70% of the whole patient admission EHRs record) were dropped from the inclusion criteria (Figure 30). This includes information such as admission information, clinical information, laboratory information, and medication information. This work used the international classification of diseases code ICD-10 [194] as the operational diseases classification coding system at Al-Ain hospital. The research framework follows two experimental scenarios. The first one is a combined development set (Admission+Clinical+ Laboratory + Medication) with all eligible patients (N=1045) that includes all of the dataset information. The second scenario is the subset of patients' profile information in four separate sets (General Admission information, Clinical information, Laboratory information, and Medication information). The figure depicts the inclusion criteria for the predictive research framework for Al-Ain hospital.

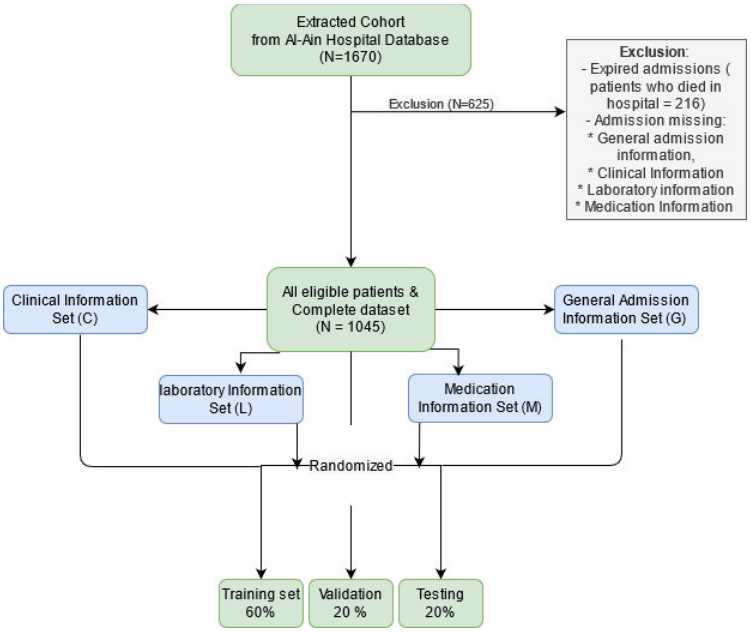


Figure 30. Inclusion protocol for the length of stay hospitalization for Al-Ain hospital dataset

All extracted features within Al-Ain hospital electronic medical records were qualified for the study features selections, and the total number of eligible features is 475.

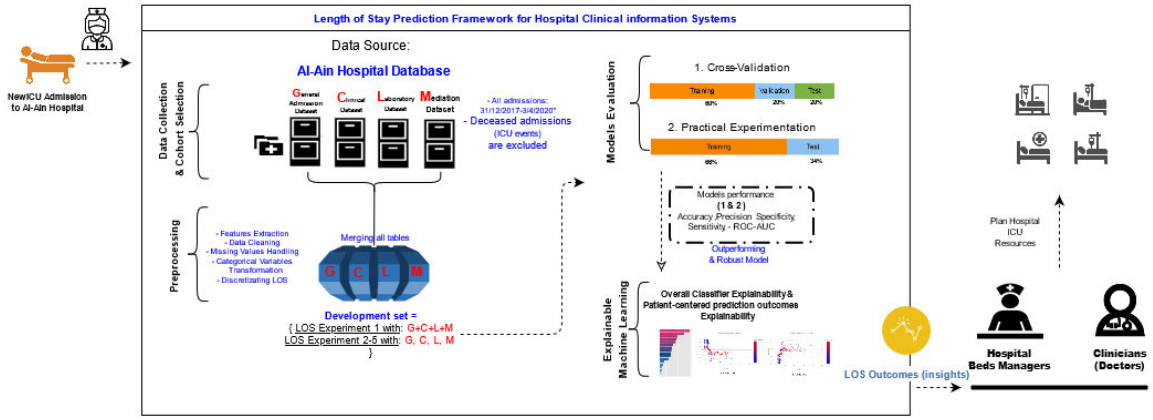


Figure 31. Predictive LOS framework on ICU hospitalisations from real Hospital dataset

5.5.2 Data Pre-processing and Discretisation

Data preprocessing is one of the essential tasks in the data mining process. Generally, EHRs datasets suffer from missing values, outliers, or raw data that require further processing and features redundancy [38]. Several steps were performed to process and extract all features that contributed to the patients' stay at Al-Ain hospital. The data extraction and mining in this chapter followed similar methods in Chapter 3 (case study 1 and case study 2) and 4 (section: 4.5). This chapter anticipates the similar steps that are carried out during the preprocessing stage for the proposed predictive lung cancer framework in chapter 4.

Data imputation was used to handle missing values, or values that contain NaN, or blanks for the four imported tables (Al-Ain dataset) in Pandas Dataframe [145]. The tables are G: general admission, C: clinical table, L: laboratory table, and M: medication table (Figure 31). Training machine learning models with datasets with many missing values, non-values (NaN) or blanks can drastically impact the machine learning quality, performance, and predicted outcomes. Therefore, a proper data imputation approach should be followed to handle the identified issues in the EHRs dataset (Al-Ain dataset). Thereby, any non-value represents that there were no entries for the specific procedures for the patient (clinical, laboratory procedure, medication or general information entry). Therefore, it was replaced with zero value (0) since the input was unavailable or possible due to the not applicable option. The null function from the Pandas library in Python was for this purpose. This decision was coordinated with the clinical ICU specialist, internal medicine clinician and a pathologist who

participated in this study[3]. The remaining features and their correspondence entries (patient records' data) are recorded or observed with the patient's information.

Categorical variables' transformation is necessary, especially when dealing with nominal or categorical variables. For example, the Al-Ain dataset has several nominal attributes that require further data representation, such as variables' transformation using the nominal encoding method [131]; it is also called the one-hot encoding method [195]. The one-hot encoding transforms categorical or nominal attributes into 0: No value or 1: Yes, when the value exists or does not. This approach aims to transform categorical to nominal and binary attributes that improve machine learning models' performance.

The data discretisation, transferring numeric/continuous variables into nominal/categorical variables (set of intervals) with minimal loss of information. The statistical rationale behind data discretisation is examined by statistical studies such [146, 147] that proposed methods to transfer continuous variables into nominal/categorical. For example, electronic health records (EHRs) and clinical decision support system (CDSS) studies [148, 149] binned the continuous variables into nominal target variables. In hospital CIS systems, binning the (continuous) length of stay into nominal/categorical is accompanied by advantages for healthcare caregivers to maximise hospital resource utilisation [150, 151]. This chapter implemented the same data discretisation approach in chapter 4 and supported similar studies to this research used in the literature review studies. For instance, Zebin et al. [87] grouped short LOS to 0-7 days and the long LOS to >7 days. Similarly, Allard [152] et al. categorised long LOS to > 7 days and short LOS to below 7 days. Table 16 provides more peer-reviewed literature studies that binned LOS into two labels similar to the approach in our research. Consequently, in this work, the LOS continuous variable is binned into a **binary class** LOS approach based on previous studies using the discretised scaled label "LOS" into two labels: 1) Label zero (0) for a short length of stay or (short LOS, 0-6 days), and label one (1) for a long length of stay (long LOS, 7+ days). Consequently, the LOS was binned into two labels and, therefore, this resulted in predictive LOS task with classification problem. The below algorithms (1 and 2) describes the steps involved in the data preprocessing and discretisation. The rationale of the algorithms is built the proposed pseudocode for LOS lung cancer framework in chapter 4.

Algorithms 1 and 2 (Figure 32) show the steps that demonstrate data cleaning, preprocessing, data discretisation and features transformation within the predictive LOS framework (Figure 31).

[3] Refer to the acknowledgment section of the thesis.

Algorithm 1-data cleaning and preprocessing Pseudocode

Input:
EHR // Hospital real ICU dataset (Al-Ain hospital dataset)
F // set of all of database eligible features, general admission features $f\_g=[f_{g1}, f_{g2},..., f_{gn}]$, clinical features $f\_c=[f_{c1}, f_{c2},..., f_{cn}]$, laboratory features $f\_l=[f_{l1}, f_{l2},..., f_{ln}]$, and Medication features $f\_m=[f_{m1}, f_{m2},..., f_{mn}]$
Output: Cleaned & Processed EHR dataset (d^{EHR})
CP (d^{EHR}):
  For each row F in d^{EHR}:
    If F in d^{EHR} contains:
      - drop ← invalid & expired admissions
      - drop ← missing records: G & C& L& M
  return dataset (d^{cEHR}) ← Return and save cleaned F

Algorithm 2-Data Discretisation and Transformation Pseudocode

Input:
T // target feature (LOS)
d^{cEHR} // cleaned EHR dataset
Output: Discretised dependent (T: LOS) & Transformed set of independents: $f\_g=[f_{g1}, f_{g2},..., f_{gn}]$, $f\_c=[f_{c1}, f_{c2},..., f_{cn}]$, $f\_l=[f_{l1}, f_{l2},..., f_{ln}]$, $f\_m=[f_{m1}, f_{m2},..., f_{mn}]$ in *F*
DLOS (T):
  For each row R_T in T:
    If y≥0 & <cut point:
      Label $R_T = 0$
    else:
      Label $R_T = 1$
    end if else
  return Discretised T
TLOS(d^{cEHR}):
  While (F **is not** Continuous):
    For each Categorical Feature F in d^{cEHR}:
      $F_{tranformed}$ ← (F ∈ $f\_g=[f_{g1}, f_{g2},..., f_{gn}]$, $f\_c=[f_{c1}, f_{c2},..., f_{cn}]$, $f\_l=[f_{l1}, f_{l2},..., f_{ln}]$, $f\_m=[f_{m1}, f_{m2},..., f_{mn}]$)^{Transpose}
    end For
  return transformed_d^{cEHR}
  Break while

CP donates the function to clean & process and treat missing values
DLOS is the function to discretise the continuous LOS
TLOS is the function to transform categorical features
Cut point: the cut point of 7+ days
F is all categorical features in d^{EHR}

Figure 32. Pseudocode for algorithms 1 and 2

5.5.3 Models Selection and Performance Evaluation

This section evaluates the machine learning models used to assess the predictive LOS framework on ICU hospitalisations from real hospital datasets. The models' implementation, model tuning, and performance evaluation used Python (version 3.6) and the Sklearn (version 0.22.1) package. This chapter refers to the same classifiers (mathematical models) used in chapter 4, and chapter 3.

5.5.3.1 Machine Learning Algorithms

eXtreme Gradient Boosting (XGBoost): The eXtreme Gradient Boosting (XGBoost) algorithm [28] is an ensemble-based learning (boosting) model. The XGBoost implements gradient boosted decision trees [29] designed for performance and speed. A recent implementation of the gradient tree boosting machines involves combining the predictions of many "weak learners" of decision trees into a strong predictor. In addition, it uses a more regularised model formalisation to control the overfitting and give it better performance [28]. One of XGBoost significant advantages is that it is designed for scalable datasets. Learning rates (0.01, 0.1, 1), the number of estimators (5, 50, 250), and the maximum depth of 1, 3, 4, 5, 9 were used for the hyperparameters in the cross-validation stage. Table 17 describes the hyperparameters' values (All Dataset features, G, C, L, and M).

Random Forest: Random Forests (RF) is an ensemble learning method (bagging) for classification that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification). Each time, a model is built based on the decision tree trained on row/features sampling with replacement. Thus, every time the decision tree model is built, new rows are fed into the new decision tree learners (bootstrapping). The bootstrapping process occurs in parallel training until models achieve (n) of trained on the decision tree. Eventually, models are aggregated and generated from bootstrapping using the majority voting to give the final predictive output. In the context of our study, the voted majority of the RF is 0: short LOS or 1: long LOS. The RF classifier is input with a decision tree as the base learner, consisting of up to 250 trees with a Gini index and a maximum depth of 21. Table 17 donates all hyperparameters values (All Dataset features, G, C, L, and M) using the RF classifier.

Gradient Boosting Machines: Gradient Boosting (GB) is a powerful ensemble learning technique (boosting) for building predictive models [29]. It works by producing a prediction model in an ensemble of weak prediction models like decision trees. It creates new base learners to be maximally correlated with the negative gradient of the loss function and associated with the whole ensemble. Therefore, it builds the model (weak learner), and it improves models' errors over time. It achieves its best performance over a sequential process after training and learning, and eventually, an improved model with better predictive outcomes is obtained. Learning rates of 0.01, 0.0, 1, 10,100 were used with the number of trees of 5, 50, 250, and 500 as well as the max depth of 1, 3, 5, 7, 9. GirdSearch (cross 5-fold validation) was used and attained the GB's hyperparameters values and setups according to experimental sets (All Dataset features, G, C, L, and M) as described in Table 17.

Logistic regression (LR): Logistic Regression is a statistical method based on the use of a logistic function (sigmoid function) to model the output of binary values (0 or 1) [27]. The logistic regression model (L1, L2, and elasticnet) is used as the regularisation (penalty) or no regularisation input. In addition, the solver (newton-cg, lbfgs, liblinear) and Inverse of regularisation strength (C) were used as a positive float value. The GirdSearch with cross 5-fold validation obtained the LR's hyperparameters values and setups per the experimental sets (All Dataset features, G, C, L, and M), as described in Table 17.

Multi-layer perceptron neural network (MLP): The MLP [35] is a machine learning predictive model that mimics the neural networks stimulated by the biological neural networks and solves challenging computational tasks such as predictive modelling tasks. The feedforward, multi-layer perceptron neural network comprised three hidden layers with 10, 50 and 100 neurons. The activation functions are ReLU, Tanh, and Logistic. The network was trained on three learning rates (constant, invacaling, and adaptive). All MLP's hyperparameters values for the 5 experimental sets (All dataset features, G, C, L, and M) are described in Table 17.

6.4.3.2 The Hyperparameters of the predictive models.

Hyperparameters selection was implemented as model-based. Gridsearch strategy with cross 5-fold cross-validation was practised to find the hyperparameters used to get good predictive results in the binary approach. This step is essential in practice and experimental settings to allow tailing the behaviour of machine learning models, especially in the context of this study (the electronic medical records dataset "Al-Ain hospital"). Table 17 discusses each model with its hyperparameter values and explanation.

Table 17. Hyperparameters of the predictive models

Parameters	Value	Description	Experiment setup				
			All*	G	C	L	M
Logistic Regression (LR)							
C	1	To control penalty strength (Inverse of regularization strength), and it must be a positive value.	1000	C=1	C=0.001	C=0.01	C=1
Solver	liblinear or newton-cg	for regularization (penalty) and optimization problem.	N/A	liblinear	newton-cg	liblinear	liblinear
Multi-layer Perceptron (MLP)							
hidden_layer_sizes		Describes the ith element represents the number of neurons in the ith hidden layer.	10	50	50	N/A	10
activation		Refer to activation function for the hidden layer.	logistic	N/A	logistic	logistic	tanh
learning_rate		Learning rate schedule for weight updates.	0.01	0.01	0.01	0.01	0.01
Random Forest (RF)							
n_estimators		Describes the number of trees in the forest.	50	250	5	250	50
max_depth		Describes the maximum depth of the tree.	None	8	4	8	8
max_features		Describes the number of features to consider when looking for the best split.	None	sqrt	log2	log2	sqrt
Gradient Boosting (GB)							
n_estimators		Describes the number of boosting stages to perform.	500	50	5	50	500
max_depth		Refers to the maximum depth limits the number of nodes in the tree.	1	9	1	5	N/A
learning_rate		Learning rate shrinks the contribution of each tree	0.01	0.01	0.01	0.01	0.01
eXtreme Gradient Boosting (XGBoost)							
n_estimators		Describes the number of gradients boosted trees (equivalent to the number of boosting rounds)	100	5	5	5	250
max_depth		Describes the maximum tree depth for base learners	None	3	1	5	3

learning_rate		Describes the Boosting learning rate	None	1	0.01	0.1	0.01

*G+C+L+M

Algorithm 3 illustrates the steps involved in baselining stage 1 for the candidate models in the predictive LOS framework, whereas algorithm 4 presents the steps involved in the evaluation of the outperforming models from stage 2.

Algorithm 3-Prediction Pseudocode for baselining stage (1): Cross-validation with Hyperparameters

Input:
K // number of k-fold, $K = [k_1, k_2, .., k_5]$
Target label // *Short_LOS, Long_LOS*
D = transformed_d^{cEHR} // contains input features *X*, and output feature *y*
H // set of hyperparameters H_{sets} with different values (model-based)
M // *set of models = ['XGBoost','RF','GB','LR','MLP']*
for *i*=1 to k_n: // *k-fold = 5*
  Split *D* into D_i^{train}, D_i^{test} ,$D_i^{validate}$ for the i'th split
  for *j*= 1 to D_i^{train}, D_j^{test} ,$D_j^{validate}$ for the j'th split
  foreach *h* in H_{sets} *do:*
    Train ***M*** on D_i^{train} with hyperparameter set *h*
    Compute test error E_j^{test} for ***M*** with D_j^{test}
 Select optimal with hyperparameter set h^* from H_{sets}
 Train ***M*** with D_i^{train} using h^*
 Compute test error E_j^{test} for ***M*** with D_i^{test}
 Compute validation error $V_j^{validate}$ for ***M*** with $D_i^{validatet}$
Output:
Set of outperformed models matrix M_P, performance estimation on $V_j^{validate}$ ***vs.*** E_j^{tes}

Algorithm 4- Pseudocode for Outperforming Models Evaluation stage with Hyperparameters

Input:
Dataset // *EHR_Features,* $fc=[f_{c1}, f_{c2}, ..., f_{cn}]$
Target label // *Short_LOS, Long_LOS*
M_P // *outperformed models from stage 1 with hyperparamters*
Split *dataset into:*
  - *training matrix* D_{TR}
  - *testing matrix* D_{TS}
Training:
for *each* ***m*** *in* M_P *:*
  train ***m*** *using* D_{TR}
  trained M_{TR} ← *compute* ***m*** *performance with AUC*
Testing:
for *each* ***n*** *in* M_P *:*
  train ***n*** *using* D_{TS}
  tested M_{TS} ← *compute* ***n*** *performance with AUC*

Output*: winning best model* M_B ← *Evaluated models on* M_{TR} *and* M_{TS}

Figure 33. Pseudocode for algorithms 3 and 4

5.5.3.2 Models' Evaluation Metrics

A set of evaluation metrics was utilised to evaluate classifiers in the predictive LOS framework of ICU hospitalisations from real hospital datasets. In the first stage (models' benchmarking stage), cross-validation with k-fold =5 was implemented to estimate the skills of proposed classifiers on unseen data. The metrics such as Accuracy, Precision Sensitivity, Specificity and AUC are used to assess the performance of the classifiers in predicting the short and long LOS for actual admissions Al-Ain hospital dataset. Also, the statistical measure (confidence interval: CI) was used in this chapter. Therefore, the performance evaluation metrics are denoted as follows:

Accuracy: donates the ratio of the correct predictions to the total of a number of predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where TP: True Positive, FP: False Positive, FN: False Negative, TN: True Negative.

The accuracy does not describe the predictive story in imbalance-class datasets. Therefore, other metrics are used to evaluate imbalanced data, such as Precision, Sensitivity, G.mean and IBA.

Precision: refers to the number of positive classifications that are actually correct or called positive predicted value '**PPR'**.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

***Sensitivity (Recall)*:** measures the proportion of actual positives that is well classified or called the true positive rate '**TPR'** or Recall.

$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

***Specificity*:** measures the proportion of actual negatives that is well classified (true negative rate '**TNR'**)

F1-Score: It can be interpreted as the weighted average of precision and recall. F1-Score = 1 is the best possible value, and F1-Score close to 0 is the worst value.

$$F1 - Score = \frac{2*Preision*Recall}{Precision+Recall} \quad (4)$$

***Area Under the ROC Curve (AUC)*:** measures the quality of the model's predictions regardless of what classification threshold is chosen. It represents the area under the ROC curve plots (TPR vs. FPR)

$$\text{TRP} = \frac{TP}{TP+FN}, \qquad \text{FPR} = \frac{FP}{FP+TN} \quad (5)$$

where TPR is the true positive rate, and FPR is the false positive rate.

The ROC plot visualises the tradeoff between the classifier's sensitivity and specificity.

PR-AUC

The area under the precision-recall curve (PR-AUC) is a curve that combines precision and recall in one plot (single visualisation). Thus, once precision and recall are calculated in every threshold, the higher on y-axis curve is, the better the model performance. Therefore, the optimal operating point on PR curve is the upper right corner, and the values of PR-AUC range from 0 to 1, with a note that 1 describes a perfect classifier [196].

k-Fold Cross-Validation: is a statistical method used to estimate the skill of the machine learning model. The k-Fold cross-validation procedure (resampling) was used in this work. The CV (k-Fold Cross-Validation) is denoted with the following equation:

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L \quad (6)$$

where *i* is the observation by randomisation. $\hat{f}^{-k}(x)$ is the fitted function which is computed with the *k*th part of the data removed. K = N (leave-one-out) cross-validation, ad *k(i)* = *i* for the *i*th observation and the fit is computed using all data except the *i*th. Typically K choices are 5 or 10.

Confidence Interval: quantifies the uncertainty of an estimate for the predicted outcomes of the evaluated classifiers.

The classification error is calculated per the following formula:

$$\text{classification error} = \frac{incorrectpredictions}{totalpredictions} \quad (7)$$

Therefore, the confidence interval is achieved [197] by calculating the formula:

$$\text{confidence interval} = \text{error} +/- \text{const} * \text{sqrt}\left(\frac{(error*(1-error))}{n}\right) \quad (8)$$

Also, two measures were used in this chapter Log Loss [198] and the *Left-Curves (left)*

$$\text{Left} = \frac{PredictedRate}{AverageRate} \quad (10)$$

for the models' predictions explainability stage.

5.6 Results

This section reports the proposed ICU predictive framework results from a real-hospital dataset (Al-Ain hospital). The predictive framework was evaluated using cross-validation (stage 1) and with two portions (training and testing) to put the model into the practical aspect (stage 2). The LOS predicted results are explained within the LOS framework for the best performing model (stage 3).

5.6.1 Cross-Validation results (stage 1: Baselining)

In this work, a k-Fold (k=5) Cross-Validation was utilised (k groups approach). It is a practical procedure, especially to evaluate the classifier's performance in case of limited data. Hyperparameters optimisation is used to assess the predictive framework using k-Fold with the best model's tuned parameters. Experiment 1 in Table 18 reports the best results of Accuracy, Precision Sensitivity, Specificity and AUC amongst all participated five experiments. In experiment 1, the XGboost model showed relatively better results than other models (RF, GB, MLP, and LR) while particularly using the AUC model's measured outcomes (77.9% and 74.4%) for validation (V) and testing (T) scores. In comparison, MLP recorded AUC of 59.4, 57.5 for V and T, respectively.

Table 18. A comparison between features' selection sets based on the patient's information profile (cross-validation and testing reported results) LOS Predictive Framework on real hospital data (Al-Ain hospital, UAE)

Model	Accuracy %		Precision %		Sensitivity %		Specificity %		AUC %	
	V*	T**	V	T	V	T	V	T	V	T
Experiments 1: All dataset (G+C+L+M) ICU features										
XGBoost	83.7	78.9	87.3	83.2	90.8	86.7	64.9	62.1	77.9	74.4
RF	82.3	77	83.6	78.8	94.1	90.9	50.8	47	72.5	68.9
GB	80.9	74.6	84.6	81.2	90.1	81.8	56.1	59	73.1	70.5
MLP	73.7	70.3	77.1	72.1	90.8	92.3	28.1	22.7	59.4	57.5
LR	77.5	69.9	84.8	80.8	84.2	73.4	59.6	62.1	71.9	67.8
Experiments 2: General (G) ICU features										
XGBoost	72.2	68.9	73.3	69.5	97.4	97.2	5.2	7.5	51.3	52.4
RF	70.8	67	73.8	69.3	92.8	93	12	10.6	52.5	51.8
GB	70.8	67.9	72.9	69.2	95.4	95.8	5.2	7.5	50.3	51.7
MLP	72.7	69.4	73.4	69.7	98	97.9	5.2	7.5	51.6	52.7
LR	72.7	69.4	73.2	69.5	98.7	98.6	3.5	7.5	51.1	52.3
Experiments 3: Clinical (C) ICU features										
XGBoost	72.7	68.4	72.7	68.4	100	100	0	0	50	50
RF	73.2	68.4	73.3	68.8	99.3	98.6	3.5	3	51.4	50.8
GB	72.7	68.4	72.7	68.4	100	100	0	0	50	50
MLP	72.7	67.5	72.7	68.1	100	98.6	0	0	50	49.3
LR	72.7	67.5	72.7	68.4	100	100	0	0	50	50
Experiments 4: laboratory (L) ICU features										

XGBoost	71.8	67.5	72.9	68.5	97.4	97.2	3.5	3	50.4	50.1
RF	72.7	68.4	72.9	68.4	99.3	100	1.7	0	50.5	50
GB	73.2	67.9	73.3	68.1	99.3	98.6	3.5	0	51.4	49.3
MLP	72.2	67.9	73	68.3	98	99.3	3.5	0	50.8	49.7
LR	72.2	67.9	73	68.3	98	99.3	3.5	0	50.8	49.7
Experiments 5: Medication (M) ICU features										
XGBoost	79.9	77	85.3	79.9	87.5	88.8	59.6	51.5	73.6	70.2
RF	77	76.6	81	95.1	89.5	95.1	43.8	36.3	66.7	65.7
GB	82.3	75.1	86.6	89.5	89.5	89.5	63.1	43.9	76.3	66.7
MLP	80.9	78.5	85	89.5	89.5	89.5	57.8	54.5	73.7	72
LR	81.8	78	87	89.5	88.2	89.5	64.9	53	76.5	71.3
Models' average performance of all experiments										
XGBoost	76.06	72.14	78.3	73.9	94.62	93.98	26.64	24.82	60.64	59.42
RF	75.2	71.48	76.92	76.08	95	95.52	22.36	19.38	58.72	57.44
GB	75.98	70.78	78.02	75.28	94.86	93.14	25.58	22.08	60.22	57.64
MLP	74.44	70.72	76.24	73.54	95.26	95.52	18.92	16.94	57.1	56.24
LR	75.38	70.54	78.14	75.3	93.82	92.16	26.3	24.52	60.06	58.22

**V: Validation set, **T: Test set*

Hence, experiment 1 produced the best predictive LOS performance compared to the other experiments in Table 18. Experiment 1 was selected for a further look at the estimated skill of a classification method using the calculated Confidence Interval of 95% (95%CI). Table 19 indicates the classification error of each model with metrics (Accuracy, Precision Sensitivity, Specificity and AUC) with the CI of 95% or the true classification error of each model is likely to be within the range of the +/- CI 95% values.

Table 19. Baselining predictive models on Al-Ain dataset (Experiment 1, "G+C+L+M") with hyperparameters and cross-validation (k-Fold = 5) approaches & 95%CI

Model	**Accuracy % + 95% CI**		**Precision % + 95% CI**		**Sensitivity % + 95% CI**		**Specificity % + 95% CI**		**AUC % + 95% CI**	
	V*	**T****	**V**	**T**	**V**	**T**	**V**	**T**	**V**	**T**
XGBoost	83.7 (76.5-90.9)	78.9 (70.9-86.9)	87.3 (80.8-93.8)	83.2 (75.9 – 90.5)	90.8 (85.1 – 96.5)	86.7 (80-93.4)	64.9 (55.5 – 74.3)	62.12 (52.6-71.6)	77.9 (69.8 - 86)	74.4 (65.8 - 83)
RF	82.3 (74.8-89.8)	77 (68.8-85.2)	83.6 (76.3-90.9)	78.8 (70.8 – 86.8)	94.1 (89.5 – 98.7)	90.9 (85.3 – 96.5)	50.8 (41-60.6)	46.96 (37.2-56.7)	72.5 (63.7-81.3)	68.9 (59.8-78)
GB	80.9 (73.2-88.6)	74.6 (66.1-83.1)	84.6 (77.5-91.7)	81.2 (73.5-88.9)	90.1 (84.2-96)	81.8 (74.2-89.4)	56.1 (46.4-65.8)	59 (49.4-68.6)	73.1 (64.4-81.8)	70.5 (61.6-79.4)
MLP	73.7 (65.1-82.3)	70.3 (61.3-79.3)	77.1 (68.9-85.3)	72.1 (63.3-80.9)	90.8 (85.1-96.5)	92.3 (87.1-97.5)	28.07 (19.3-36.9)	22.72 (14.5-30.9)	59.4 (49.8-69)	57.5 (47.8-67.2)

LR	77.5 (69.3-85.7)	69.9 (60.9-78.9)	84.8 (77.8-91.8)	80.8 (73.1-88.5)	84.2 (77.1-91.3)	73.4 (64.7-82.1)	59.64 (50-69.3	62.1 (52.6-71.6)	71.9 (63.1-80.7)	67.8 (58.6-77)

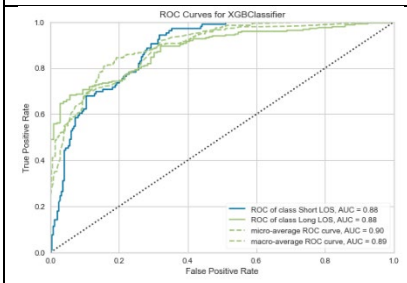
**V: Validation set, **T: Test set*

The calculated average overall models' performance of all experiments in Table 18 within scenario 1 revealed a preference towards XGboost Model. Therefore, XGboost is evidenced to be a robust classifier. Hence, XGBoost is our selection in the baselining stage (Figure 31).

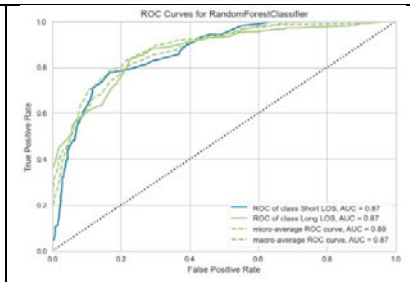
5.6.2 Outperforming Models' Evaluation (Stage 2: Practical Experimentation)

In this stage, the five main experiments (G+C+L+M, G, C, L, M) experimented with the same tuned models' parameters are used in the previous stage (baselining). For this purpose and based on the previous stage (Baselining), the most outperforming models were selected for a further performance evaluation on a practical aspect to attest candidate models' performance. The result of this experimental procedure is the best performing model. The performance measure is expressed through the ROC curves that display TPR on Y-axis and FPR on the X-axis. Two portions were used for the practical experimentation step: the training set (66%) and the testing set (34%). Figure 34 illustrates the ROC of the candidate's models in stage 2.

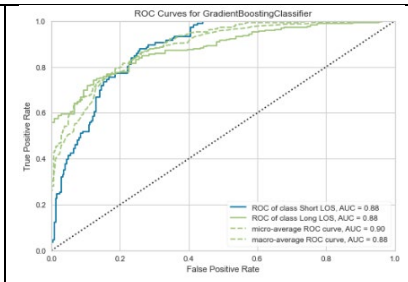
All dataset G+C+L+M features



a

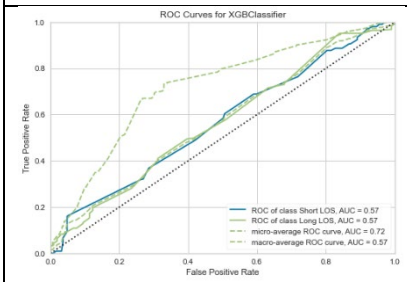


b

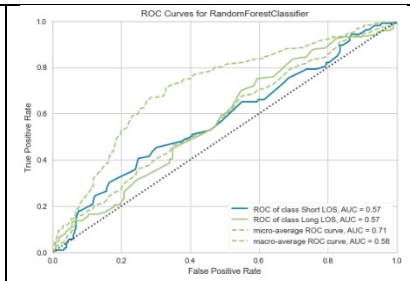


c

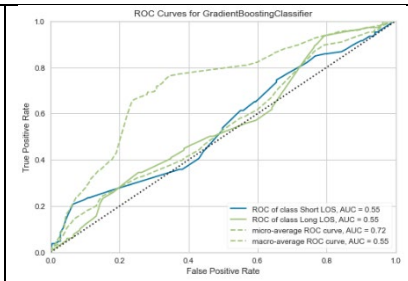
General (G) ICU features



d



e



f

Clinical (C) ICU features

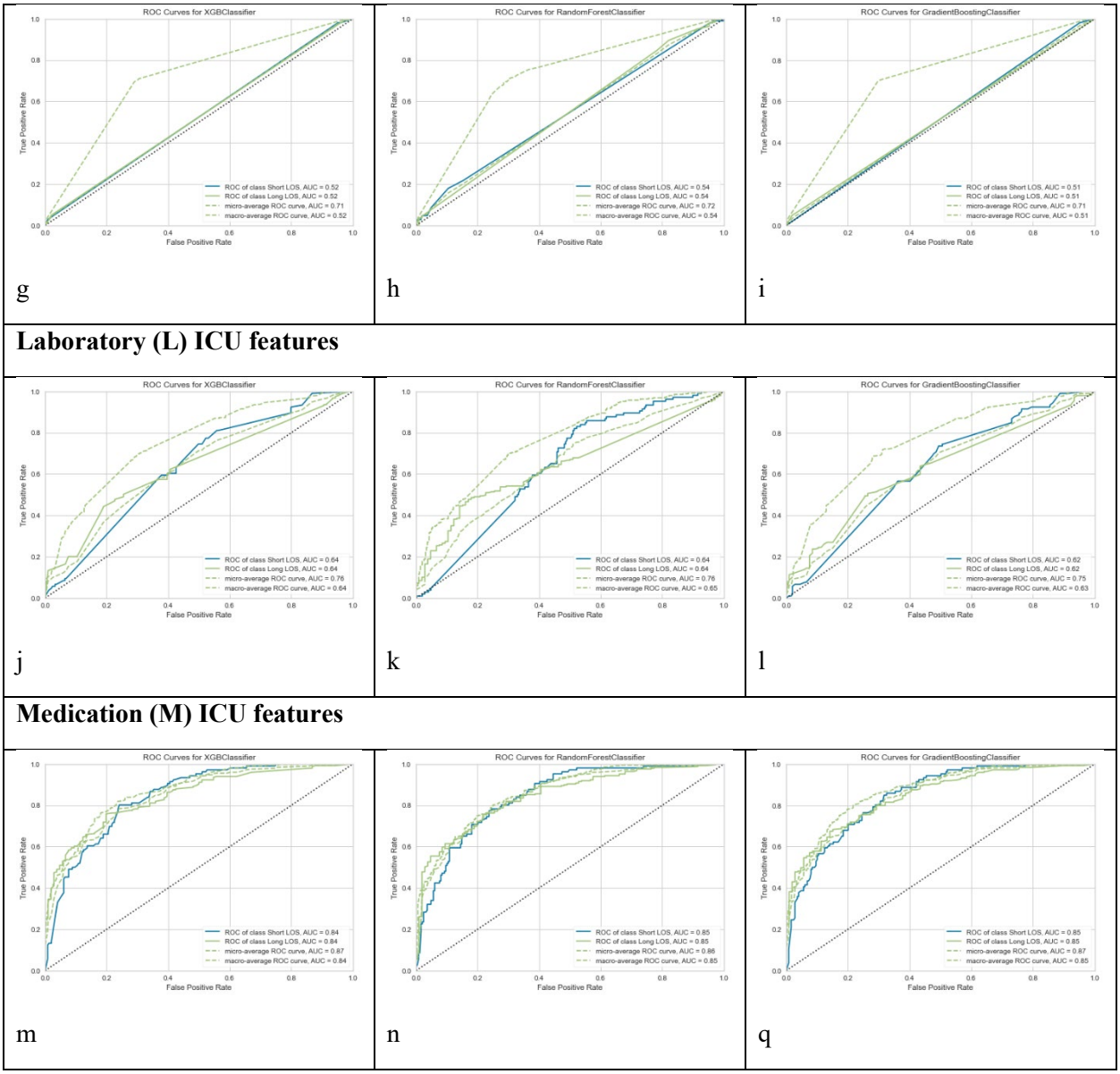


Figure 34. Practical experimentation stage (ROC Results for XGBoost, RF, and GB)

The performance of the three models in the practical experimentation stage was close to each other. For example, when evaluating the ROC for the XGBoost classifier, it was observed that the set of all of the experiments (G+C+L+M) achieved the best ROC (88%: CI%95 [81.6%-94.4%]) for short LOS and long LOS. The second apparent results are with the Medication (M) ICU features, where the XGBoost achieved ROC= 84%: [76.8%-91.2%] for short LOS and long LOS classes. This confirms the model's ability to commence fewer prediction errors in both classes (short and long LOS).

The Gradient Boosting achieved comparatively similar results to the XGboost in all features experiments (G+C+L+M), with ROC 88%: CI%95 [81.6%-94.4%] for short LOS and long LOS. However, Gradient Boosting obtained slightly better results with the experiment of Medication (M), achieving ROC 85%: CI%95[78%-92%] for short LOS and long LOS, respectively. At the same time,

the XGboost results were slightly better than GB in the general (G) and clinical (C) experiments. However, the General ROC results attained XGboost 57% and Gradient Boosting 55%, Clinical short and long LOS managed to obtain 52%: CI%95[42.2%- 61.8%], and 51% CI%95[41.2%- 60.8%] for XGBoost, and Gradient Boosting respectively as the least important ROC results in the experimentation stage. Moreover, XGboost attained ROC of 64%: CI%95[54.6%- 73.4%] and 62%: CI%95[525%- 71.5%] for GB in the laboratory experiment.

Finally, the Random Forest achieved steadily ROC results for both classes (short and long) LOS. In all features (G+C+L+M), the RF obtained the highest ROC (88%: CI%95 [81.6%-94.4%]) within the five experiments, then Medication experiments with ROC of 85%: CI%95[78%-92%], and 64%: CI%95[54.6%- 73.4%] for laboratory, 54%: CI%95[44.2%- 63.8%] 57%: CI%95[47.3%- 66.7%] for clinical and general experiments respectively.

5.6.3 Explaining the XGboost Predictive Results (stage 3)

This section explains the predictive results of the winning model in the proposed framework. The winning model is the most robust classifier based on the predicted outcomes and ability to attain stable and reliable results based on different experimentation setups from Al-Ain hospital data. The XGboost achieved the desired outcomes; therefore, it is our selection model for further results explanation. This chapter is expected to reveal the black box of the predictive classification model (XGboost) and make it more understandable and easy to explain for non-machine learning people. This may include healthcare workers in hospitals and healthcare givers such as hospital managers, clinicians, hospital nurses, and health insurance companies. Recent research papers [199] [200] discussed the advantages of the successful implementation of the xAI in healthcare and medical research.

The predictive outcomes are explained from two perspectives. The first approach is the explainability of the general classifier' predicted outcomes using the whole dataset. It is referred to as (predictive outcomes with the model's overall explainability) in the ICU dataset. In the second approach, the patients are considered in the perception of the explanation. Therefore, it is referred to it as the model's patient-centred prediction outcomes' explainability.

For this purpose, the ExplainerDashboard prediction explainer [201] was exploited as the explainable artificial (xAI) tool. The xAI tool builds explainable interactive dashboards to analyse the classification and predict the results. ExplainerDashboard xAI libraries are compatible with Python 3.6. The dashboard is running over the Local Server of the experimenting computing instance (http://localhost:8080). The XGboost classifier is used with hyperparameters' values per table (Table 17). The ExplainerDashboard was utilised in this research after examining similar methods such as

SHAP (chapter 2 and 4) or other features traditional of the specific models such as features' importance of RF, GB, etc. The ExplainerDashboard provides a comprehensive analysis of the predictive outcomes. Hence, we are working with UAE and Al-Ain hospital, and we considered providing them with xAI enabled facility.

5.6.3.1 Predictive outcomes of XGboost overall classifier explainability

a. XGboost performance metrics explainability

Based on the reported XGboost performance metrics for the short and long LOS labels (Table 20), it was noticed that the achieved accuracy for short and long LOS classes is 94.6%. Precision for both labels (short and long) LOS is 91.4% and 96%, respectively. Recall and F1-Score are important metrics where XGboost achieved relatively close results, as seen in Table 20. In addition, XGboost showed robustness to differentiate between the short LOS and long-predicted LOS classes with ROCAUC 98%. This confirms the ability of the model to achieve high predicted desired results with a negligible error percentage (2%).

Table 20. Model performance metrics

for short LOS vs. long LOS for XGBoost classifier: the xAI Tool

Metric	**Short LOS**	**Long LOS**
Accuracy	94.6%	94.6%
Precision	91.4%	96%
Recall	90.6%	96.4%
F1-Score	91%	96.2%
ROC-AUC_	98%	98%
PR-AUC_	95.6%	99.1%
Log Loss	0.211	0.211

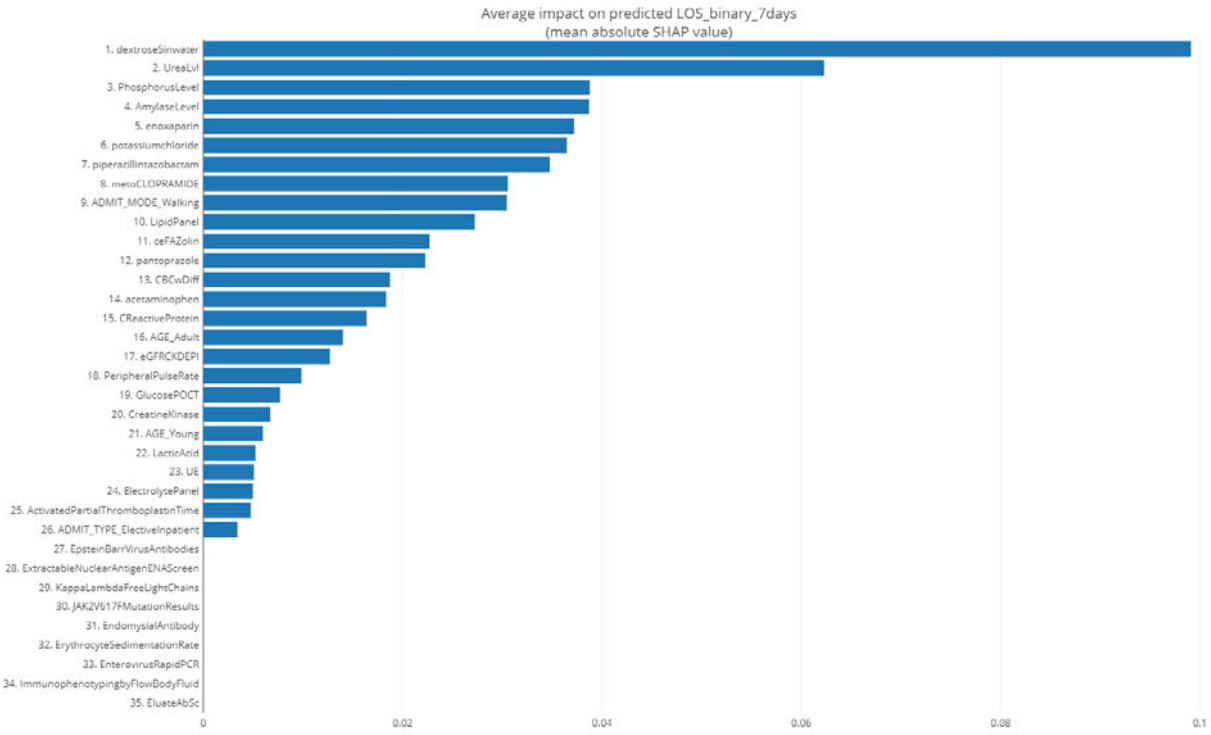
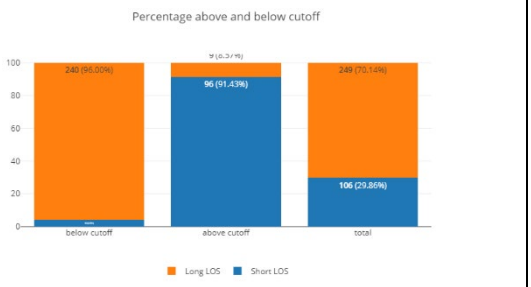
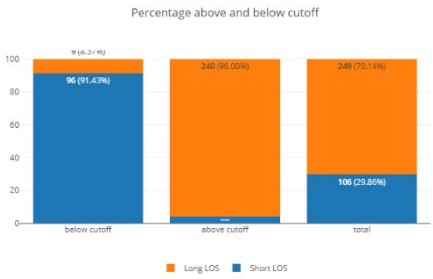


Figure 35. Features Importance by mean absolute SHAP value: XGboost

Figure 35 shows the average impact of features on the predicted LOS (short and long) labels using the mean absolute SHAP value. Both Medication and Laboratory information contributed to the most features impact on the XGboost model's decision. In addition, the general admission' features Admission Mode (walking), and Age (Adult, Young) made important weight (features importance) in the decision of the XGboost. It is noted that the clinical features did not appear in the features' importance plot (Figure 35). This can be interrupted as the SHAP method [67] measures the influence of a feature (global influence) by comparing model predictions with or without the feature. The SHAP's method computes the contribution of each feature to the prediction. Eventually, SHAP values provide information about each feature's contributions to individual prediction (Figure 35).



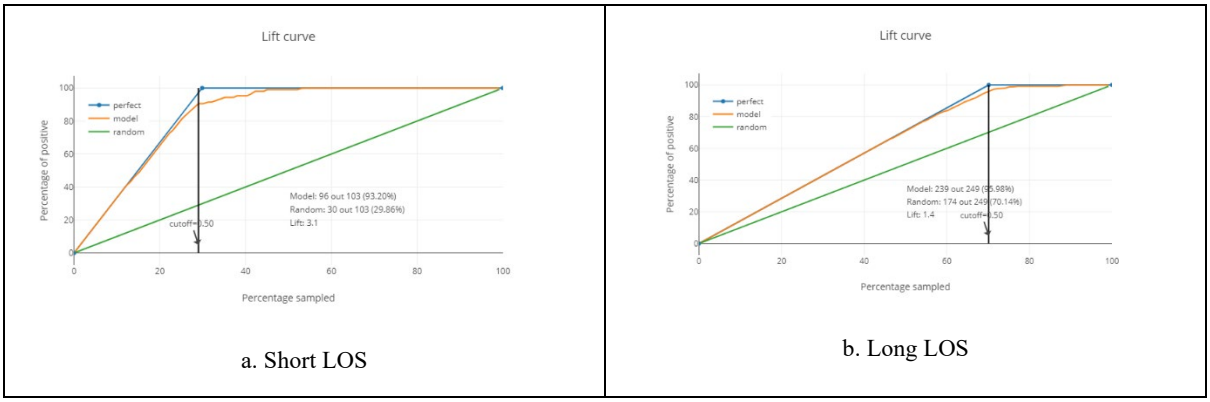
a. Short LOS



b. Long LOS

Figure 36. Percentage above and below cutoff: short LOS and long LOS

Figure 36 is the classification plot of the short LOS and long LOS labels. It describes the distribution of labels above and below the cutoff. The cutoff of the classification report is 80% per the xAI method. This donates that 91.43% of the XGBoost classified short LOS cases are above the cutoff (Figure 36-a). While (Figure 36-b) states that 96% of the XGboost long LOS cases are above the cutoff. XGBoost proves to be able to differentiate the two labels effectively. XGBoost succeeds to classify and predicting the above threshold value (cutoff=0.8) as positive and those below as negative.



a. Short LOS

b. Long LOS

Figure 37. Left curves for the XGBoost classification results: short LOS and long LOS

The left curve (Figure 37) is an important measure that helps with the predictive classification model's effectiveness (XGboost). Any given number of cases (percentage of samples: Figure 6) illustrates the expected number of positives we would predict if we did not have a model but simply selected random cases. The left curve provides a benchmark against which we can see the model's performance [202]. For example, based on equation (10), the XGBoost model gives us a left 3.1 predicting the short LOS class and a lift of 1.4 to predict the long LOS class. A good classifier will provide us with a high lift when we act on only a few cases, and as we include more cases, the lift will decrease. The left curve with the best classifier (that commences fewer errors) would overlap with the existing curve at the start, then continue with a slope of 1 until it is all success, then continue horizontally to the right [202]. This is clearly projected in Figure 37 (a & b). Thus, left curves aid beds managers or healthcare decision-makers to understand the decision made by the classification model and the way how it impacts the healthcare decisions and strategies for managing resources' utilisations and beds' availability.

The predictive outcomes overall classifier's explainability approach is clearly proven that the XGboost is robust with high and desired predictive outcomes (short and long LOS). Furthermore, the xAI tool explained the XGBoost classifier's inner workings with ease and made the decision predicted outcomes clear and understandable to a non-data machine learning specialist.

5.6.3.2 Patient-Centered Prediction Outcomes' Explainability

The ability and the inner workings are assessed to explain the decision of the XGBoost. For this purpose, a random patient (de-identified) was selected with the given data index (998). The index values are the numerical order in Pandas DataFrame by Python. Therefore, the number 998 represents the patient's case (patient profile) or admitted information (General, Clinical, Laboratory and Medication) in the dataset.

The xAI tool provides a range of explainable prediction components, including individual prediction's explainability. The individual prediction components were utilised, including the (pdp) plot of the feature, the contribution to prediction probability, the dependents' plot of the feature, and the prediction percentage of each class label. In addition, UreaLvl feature was used to evaluate the XGboost performance at a feature and an individual patient level. The selection of the UreaLvl is indiscriminately and only used to explain XGboost evaluation at the patient-centred level.

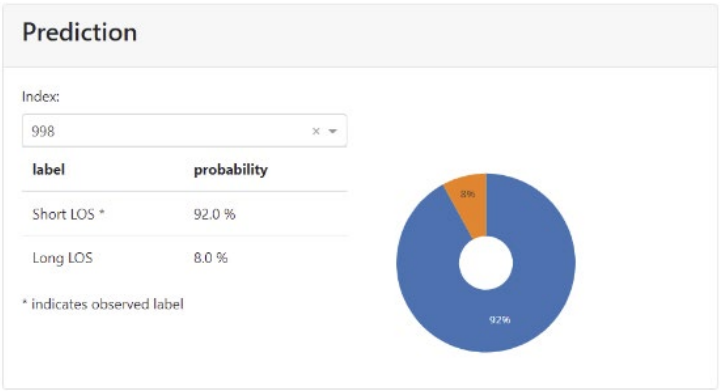
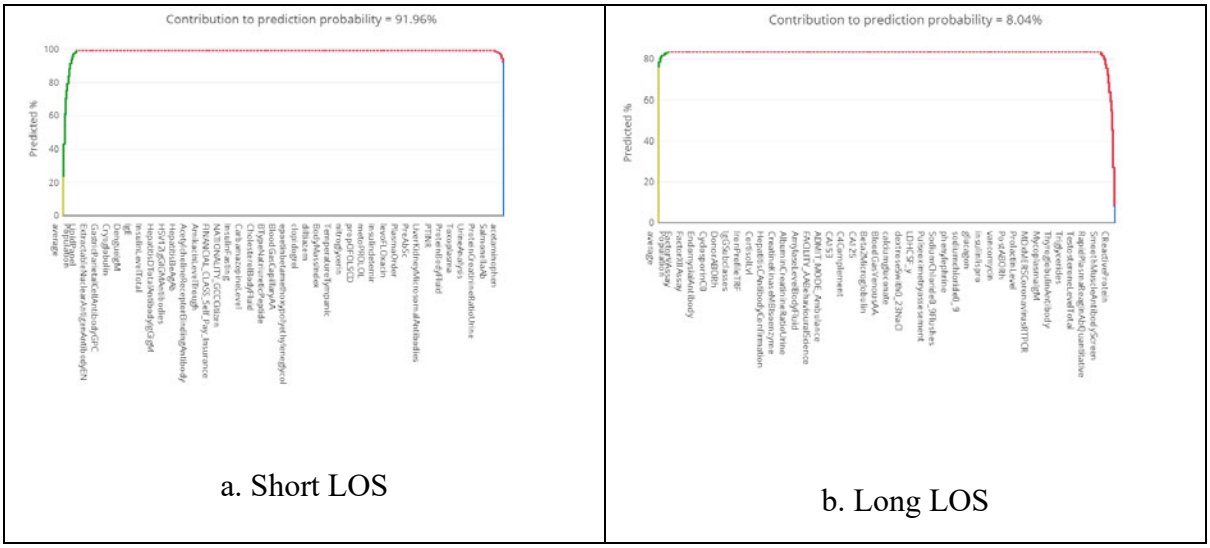


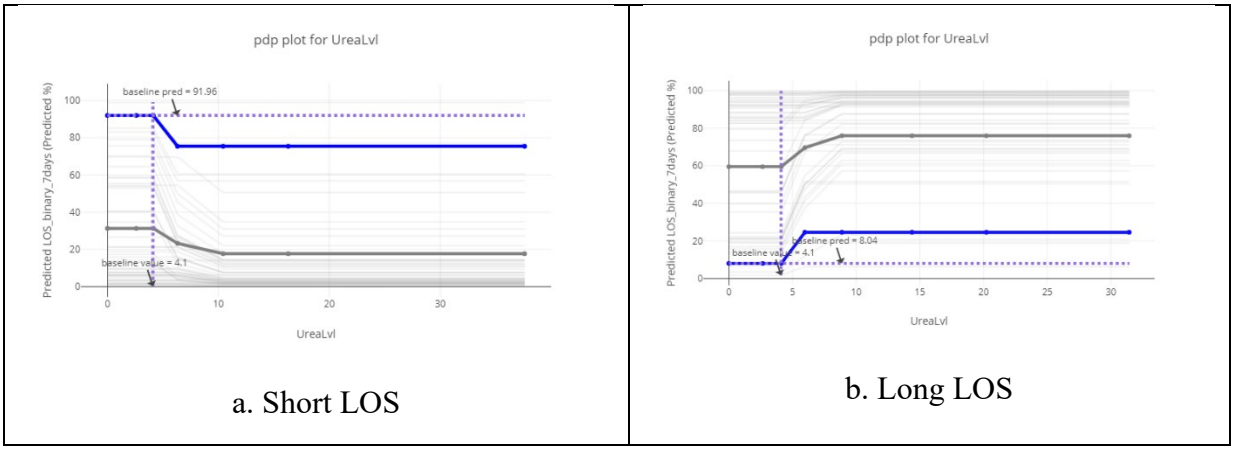
Figure 38. Prediction percentage by class labels (short LOS and long LOS)



a. Short LOS

b. Long LOS

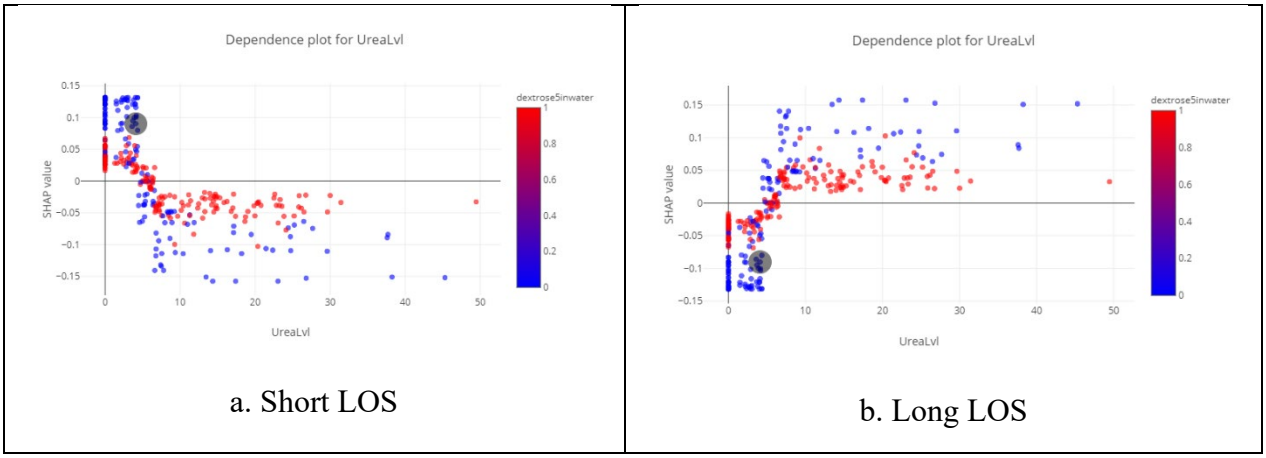
Figure 39. Contribution to the prediction probability (short and long) LOS classes



a. Short LOS

b. Long LOS

Figure 40. Partial dependent plot (pdp) for UreaLvL feature (short and long) LOS classes



a. Short LOS

b. Long LOS

Figure 41. Dependence plot for UreaLvl: (short and long) LOS classes

Every single feature was evaluated, which is shown in Figure 39, including how it affected the XGboost prediction. The contribution of features to the model's outcomes shows the breakdown of every single feature in the XGBoost and how it affected the final prediction for patient 998. The breakdown shows how the model thinks that patient 998 was predicted to be a short stay.

Figure 39 shows that the final contribution to the XGBoost prediction is 91.96%, and this is because the features that contributed have a high predicted probability in the patient's case. This is justified by the electronic health records (Al-Ain hospital) of the features present in the patient case. Therefore, the more features are inserted into the model (General, Clinical, Laboratory and Medication), the more able is the model to provide a reliable prediction. The features per Figure 39 (a-b) represent how each feature adds up to the final contribution of the short LOS prediction. The partial dependent plot (pdp) in Figure 40 illustrates how the prediction changes based on each feature input. For instance, Figure 40 (a) clearly shows the partial feature contribution to the prediction of short LOS outcomes. Still, if we look at the same feature from the perspective of the long LOS label, we can see a weak partial contribution to the predicted outcomes. Figure 41 (a-b) epitomises the dependence plot of UreaLvl according to SHAP values in relation to dextrose5inwater. The relationship (random selection)

between UreaLvl and dextrose5inwater features and their impact on the XGBoost prediction outcomes (short and long LOS) are depicted.

Finally, the xAI tool provides an overall picture of the XGBoost predicted outcomes or the prediction decision of each class label at the patient level (Figure 38). For example, the XGboost results designate that the patient is likely to stay (Short LOS) with a probability of 92% and the negligent likeliness to stay (long LOS) with 8%, eventually, and after explaining the inner workings of the XGboost from the features' perspective and their effect on the models' prediction outcomes. Also, the interaction between two features to the overall picture of the final prediction, the xAI tool provides the ability for us to understand the inner details of the chosen or the winning and outperforming model (XGBoost). Notably, these explanations may guide the concerned person of the AI model, such as the beds' managers, clinicians and healthcare insurance companies, to investigate each feature's or features' interactions on the model's outcomes or find un-explored relationships between interacted features and the way in with their interactions impact on the model's prediction classes.

5.7. Discussion

The evaluation of the proposed LOS framework using cross-validation and practical experimentation approaches with hyperparameters' optimisation is affirmative to achieving one of the foremost anticipated goals in this study. It was observed from all experiments (stage 1), that the ICU-LOS predictive framework obtained low variance and low bias with both approaches. This confirms the robustness of having validation set in a cross-validation approach to find the best parameters of each model (model tuning) and achieve optimised predictive results. Furthermore, this approach is helpful to reduce overfitting and to prevent underfitting as they are common issues in the machine-learning space. Eventually, this leads to an improved predictive machine-learning performance.

In the cross-validation (baselining), the statistical inference (Confidence Intervals "CI 95%") is used during the stage of the models' performance evaluation. The CI %95 gives the interval estimators for the prediction error, and it provides a more informative way to analyse and interpret predictive results. Therefore, the uncertainty estimate helps the hospital beds managers understand how good or bad the LOS classifiers' performance is. Moreover, it helps to look at which model is less complex and more interpretable. The experiments (stage 1) showed that the proposed LOS predictive framework could be a doable approach. For instance, the XGBoost validation set's AUC is 77.9% (69.8% – 86%) with CI%95, and the testing set's AUC is 74.4% (65.8%– 83%). This indicates that the XGboost can differentiate between the two labels (short LOS and long LOS) with an AUC of 77.9% on the validation set, and the true classification error of the XGboost is likely between 69.8% and 86%. Similarly,

XGboost Model's AUC is 74.4% on the testing set, and it is likely to predict the LOS labels truly within 65.8%– 83%. Thus, the cross-validation with hyperparameters' optimisation approach (stage 1) was beneficial and obtained desired outcomes to predict the short and long LOS with low variance and low bias.

The validation procedure provides an unbiased evaluation of a model fit on the training dataset while turning the model hyperparameters. The model evaluation with the stage 2 approach aimed to test the models' robustness during the practical implementation in real-world scenarios. For instance, if we have a new admitted case to the hospital (ICU), it is crucial to depend upon models unlikely to commence any significant error margin to differentiate short LOS from long LOS. Also, it is essential to attest the performance of the chosen model before putting the predictive model in practice in real settings. The predictive performance of the three candidate models was comparatively close to each other, with more favour towards the XGBoost, which achieved slight better-obtained results within the whole dataset features experiments (G+C+L+M). This is an important aspect that the more feature we have, the better predictive results we can achieve. It was apparent that the three candidates' models' performance improved with more features considered (G+C+L+M). This pattern is confirmed due to their common characteristics' methods to achieve reduced variance, such as bagging (Random Forest) and the ability to solve two-classes classification problems such as boosting (Gradient Boosting and XGBoost). The bagging and boosting methods are the most used methods in ensemble learning, and they are robust and accurate models.

Choosing the most suitable classifier is subjected to their performance in the proposed LOS framework. Therefore, we have decided to take XGboost to further explain to the non-specialised person in artificial intelligence, such as the hospital beds manager or generally people working on healthcare and for hospital CDSSs. This is called opening the black box in the predictive classification tasks (model's explainability). This is important for clinical decision support systems (CDSS), aiding healthcare professionals in their clinical decision making and predicting outcomes [203] which one is. The requirements CIS systems go beyond the model performance [204]. The CDSSs are established in clinical settings to exhibit proven safety [205].

The critical part of the predictive model safety is that the machine learning model is expected to understand and explain its predictions, including the situation when the ML models cannot explain what is intended. For this purpose, the xAI tool was used to unveil the XGboost classifier's predictions and explain the XGboost classifier's inner workings. As a result, it will benefit the hospital ICU beds manager and healthcare decision-makers to understand why such a chosen classifier (machine-learning

model) achieved specific predictions. Furthermore, it gives non-technical experts the ability to inspect the working performance of the model without dependence on artificial intelligence expertise.

Additionally, it is important to explain each patient's prediction, which will be the micro-level of the model's ability to explain the predictions at an individual level. This is crucial at a decision level for healthcare workers at hospitals, including hospital managers, beds managers and clinicians, to understand each prediction associated with a particular patient's case for further clinical and financial decisions. The utilised xAI tool explains the inner workings of the predictive models such as XGboost that tell people (humans in the loop) whether they understand what the model does and what does not. Therefore, this is an essential indication that they can gain intuition when the model is possibly missing important prediction information. As a result, the healthcare decision-makers will be able to overrule it.

Overall, the predictive outcomes classifier's explainability of the XGboost classifier revealed a high predictive ability of the chosen model. For example, the Accuracy, Precision, F1-score, Recall, ROCAUC and PR-AUC scores were above 90% for both labels (short and long LOS). Thus, this confirmed the undisputed prediction ability of the XGboost model. Finally, the XGBoost explainable approach helps clinicians, healthcare workers, and insurance companies understand which features are likely to contribute to a short or long LOS once a new case is admitted to ICU rooms. This is important since it will facilitate the mission to hospital workers and beds managers to allocate resources effectively. An important note is that relying on a robust classifier can explain its inner workings on the predicted outcomes, which help numerously reach safe, adaptable and reliable CDSS with the advancement of the machine-learning model. Also, it benefits from their ability to predict their decision from a data-driven approach.

5.7.1 Practical Implications

Hospital healthcare workers such as beds managers, ICU clinicians and nurses can benefit from this study more in practical implementation. The study provided a sophisticated and explainable predictive framework to predict the inpatient length of stay at the time of ICU admission or transfer to ICU. The hospital managers and specialists can decide on complex inpatient cases based on the reported results from the predictive xAI-LOS approach. Hence, the xAI provides the ability for beds managers or clinicians to compare patients with similar admission profiles. In addition, the framework offers advanced patient health monitoring capabilities using the advancement of artificial intelligence techniques. Eventually, it helps the hospital and ICU healthcare workers to make better decisions in

the dynamic and challenging health environment where the uncertainties, whether internal uncertainties within the hospitals or external factors such as pandemics, can occur at any time.

The introduced ICU-LOS framework can be a viable solution for hospital management, improving patients flow in the ICU, resources capacity, including hospital beds and ventilators, staffing, stock management, medical supplies, and other hospital resources allocation. Effective bed management still can accommodate newly admitted patients at times of uncertainty. This is the most important goal to achieve by hospital healthcare workers, caregivers, and healthcare stakeholders. In contrast, inefficient substandard hospital management leads to an overabundance of patients, more stress on healthcare workers, and consumption of hospital resources. As a result, this may lead to rejecting newly admitted patients if all beds are fully occupied, a lack of enough hospital resources, or hospital healthcare workers are busy.

Another viable benefit of the introduced framework is that it reduces health care expenditure, reduces wasted time, and increases the quality of service due to satisfactory bed management and effective hospital resources management. Integrating our framework into the hospital CIS systems can help hospitals and ICU rooms be more effective and productive.

5.7.2 Work Limitations

Although our study discussed many benefits of predicting LOS in an ICU setting, it also has some limitations. Firstly, the external validation was not performed on other datasets with similar characteristics to the Al-Ain dataset. This is due to the complication of getting real hospital data, especially the longer time (a total of one year and two months) that we waited until we received the AL-Ain dataset via UAEU and WSU. This involved visiting UAEU and Al-Ain hospital, meeting with involved stakeholders, and applying SEHA ethics via Al-Ain hospital. More importantly, the time is taken to get the ethics approved by Al-Ain hospital due to the COVID-19 pandemic, which stressed the hospital's operations. Also, until the data became available for UAEU. After that, we applied for the current WSU ethics amendment for this study. Eventually, we were able to access and commence working on the data at the end of March 2021. At the beginning of the study, we intended to evaluate the predictive LOS ICU framework on multiple real hospital datasets to assess the robustness of the proposed framework on more datasets. In a future study, we aim to request additional real hospital datasets with the same data characteristics (the Al-Ain dataset) and evaluate our framework via the external data validation approach. Secondly, the International Classification of Diseases (ICD) used for this work is the ICD-10cm. This is because Al-Ain hospital only uses this disease coding schema for their CIS (EHRs) hospital system. Due to this, the validation part of the predictive model could not

be achieved on the ICD-9 classification system according to our previous works [131] and [96]. Therefore, we will consider validating our framework within the same ICD schema (ICD-10cm) for future research.

Thirdly, the sample size in our study is relatively small due to the availability of the data collection. This disallowed us from evaluating the ICU-LOS framework on larger datasets. As a result, the advancement for deep neural network techniques could not be exploited to predict LOS at the time of hospital admission and get more data insight using the state-of-the-art deep learning models. We intend to obtain larger datasets in the future study, either expanding the data collection period from a large hospital or acquiring health linkages' datasets from the appropriate enormous datasets resources such as the Centre for Health Record Linkage (CHRL) [206]. Finally, the relative accuracy of the compared model was not studied to clinician estimates of LOS used in CIS. Still, we believe it can be used as a guide for quality improvement initiatives. For example, LOS indices, which compare expected to observed LOS, have been proposed as efficiency and hospital performance markers. Using the xAI tool for the patient-specific predicted LOS to measure expected LOS may improve the accuracy of such indices. Ultimately, this will allow hospitals to generate more representative quality metrics and, in reimbursement schemes that incentivize quality care, avoid punishment for taking on higher-risk patients.

5.8 Conclusion and Future work

This chapter presented a practical research predictive ICU framework to predict patients' length of stay at the time of admission to ICU rooms using real hospital data. While achieving desired predictive results during the three stages of the LOS predictive framework, it proved to be a practical and direct research implication for a genuine solution for beds management and resource utilisation in ICU settings. Furthermore, while ensemble learners showed robust and desired results within the three proposed stages in the framework, the XGboost model proved to be the best model due to its ability to explain the inner workings of non-AI people. Thus, our approach is the first among reviewed literature to provide a practical and AI explainable framework for predicting ICU patients' length of stay from a data-driven approach to the best of our knowledge. Further, the framework is not limited to any disease nor any health condition. Therefore, researchers can use it in clinical research and electronic health records for further important hospital predictive tasks such as improving patients at risk of mortality. The future work will develop research on user-centred clinical predictive information system working that puts the users of the proposed research work in the perception of the daily hospital workflow. A good in-depth future research study of the utilisation of xAI in hospital, ED and ICU

settings to codify the xAI for clinical information systems will establish authentic ML-XAL implementation and regulate their use in EHRs and for the healthcare system.

6. Chapter Six: Assessment of Hospital External Resources Demanding Factor: Simulating Public Health Measures during Pandemics

6.1 Chapter Background

Managing hospital resources effectively goes beyond the internal input variables from within the hospital settings. Practically, healthcare systems must consider factors that have a direct impact on hospital healthcare services. The health services are connected as one body within the healthcare system; however, suppose decision-makers in healthcare assessment settings fail to assess, examine, and predict external factors or be prepared for them before they happen. In that case, this will lead to the whole failure of the hospitals. Consequently, the healthcare system will fail accordingly since the hospitals are the most important component of the healthcare system. Also, unexpected scenarios can create pressure on hospital resources, such as pandemics. This research proposes to assess one of the main factors that will directly impact the healthcare system if no action or response is planned before from the healthcare ministerial or people at the decision level. Communicable infectious diseases are a public health concern. When an infectious disease passes the threshold and turns from an outbreak to a global spread, this alarms the healthcare systems and hospitals to action an immediate plan to prepare and manage their hospital and hospitalization occupancy beds. For example, the novel coronavirus (SARS-CoV-2) emerged in December 2019 in China and was declared a global infectious disease pandemic in March 2020 by WHO [207]. Responding to COVID-19, most world countries imposed lockdowns and strict measurements to control the spread of the disease within their borders. As a result, the healthcare system in certain countries or regional areas was overwhelmed with a high number of COVID-19 hospitalisations. Other hospitals were under strain even they were not dealing with an influx of COVID-19 patients[4]. Therefore, many hospitals achieved an overabundance of patients in the first wave.

Consequently, this created a persistent need to consider uncertainties from external factors when studying or providing a practical framework for hospital resources utilisation. In 2020, the research was motivated to consider the external public health factor to study and examine their effect on hospital resources' utilisation. While the pandemic progresses, the research inspired to simulate and measure

[4] https://www.abc.net.au/news/2021-09-03/hospital-influx-covid-patients-national-cabinet/100430286

the public health policies that were levied by different countries and report which public policies are likely to protect the healthcare system from inevitable failure. These policies are external factors that can help the healthcare system be prepared. Thus, this creates imperative research questions:

3.a How to measure hospitalised cases from non-pharmaceutical interventions (NPIs) based on different governmental policies, and how do the hospitalised cases demand hospital resources accordingly?

1.f How machine-learning models can predict LOS of hospitalised cases during pandemics?

Research question 3.a is connected to the integral research question 3 in chapter 1, while research question 1.f is linked back to the integral research question 1 in chapter 1.

For this purpose, this chapter addresses the above question 3.a by referring to three countries (two case studies). The first case study was designed to measure the different non-pharmaceutical intervention public policies and measured via simulated epidemiological models. The results of both case studies revealed the importance of applying NPIs measures to protect the healthcare system and, therefore, the hospital resources and prevent any major consequences in the quality delivery of healthcare. The second case study simulated the applied strict state measure (curfew) and how the NPIs (public health order/measure) can protect from a severe impact on the healthcare systems (case study in this chapter). The cases' choice of the countries in case studies 1 and 2 is rationalised based on the length of days in lockdown as well as the existence of the information about those countries being available on the web during the pandemic as good examples to reflect specific NPIs measures.

The third case study projected the hospitalised cases' purpose from case studies 1 and 2 and how they could impact hospital resources. The third case study provided a ground research attempt to evaluate the pandemic's impact on resources' utilisation at a scalable level. Case study 3 attempts the research question 1.f of this chapter. The LOS predictive methods from Chapters 4-5 supported the case study and showed promising results.

The epidemiological models (case studies 1 and 2) applied to three distinct countries help to determine and forecast the spread of infectious viruses and, therefore, help to determine and simulate the influx of patients and their potential impact on hospital resources and beds' occupancy, which ultimately is related to LOS estimation (case study 3). Case study 1 in this chapter uses Johns Hopkins, JHU CSSE (Figure 1 in chapter 1), while case study 2 uses the same data source and our world in data (Figure 1 in chapter 1). Finally, the third case study in this chapter uses the COVID-19 linkage dataset (Figure 1 in chapter 1).

6.2 Case Study (1) - A simulated measurement for COVID-19 Pandemic using the Effective Reproductive Number on an Empirical Portion of Population: Epidemiological Models

Case Study Summary

COVID-19, as a global pandemic, has had an unprecedented impact on the entire world. Projecting the future spread of the virus in relation to its characteristics for a specific suite of countries against a temporal trend can provide public health guidance to governments and organisations. Therefore, this case study presented an epidemiological comparison of the traditional SEIR model with an extended and modified version of the same model by splitting the infected compartment into asymptomatic-mild and symptomatic-severe. The NPIs were exposed and derived layered model into two distinct case studies with variations in mitigation strategies and benchmarking and comparison. This case study explores the United Arab Emirates (a small yet urban centre with a multi-cultural population) and Victoria, Australia (where clear sequential stages NPIs were implemented). Further, this case study concentrated on extending the models by utilising the effective reproductive number (R_t) estimated against time, which is more realistic than the static R_0. The aim was to assess the potential impact of NPIs within each case study. Compared to the traditional SEIR model, the results supported the modified model as being more sensitive in terms of peaks of simulated cases and flattening determinations.

6.2.1 Introduction and Related Work

Starting in 2020, humankind has increasingly suffered from the spread of a new pandemic characterized by acute respiratory and vascular symptoms produced by a novel coronavirus strain known as SARS-CoV-2 [208]. The virus, which initially emerged in Wuhan, China, in November 2019, later was considered a full-fledged outbreak before being declared by the World Health Organisation (WHO) as a public health emergency of international concern [209] around early 2020. Today, COVID-19 has no known approved vaccine, and no treatment is being considered effective. Meanwhile, governments and health institutions need assistance visualizing, simulating and assessing effective non-pharmaceutical interventions ((NPIs) to mitigate this virus' unpredictable behaviour and control its spread. Modelling techniques allow simulation and prediction of Covid-19 growth trends and guide pre-emption and preparation. However, it is important to properly introduce model parameters to understand the spread pattern of the infection under different mitigation strategies [210]. NPIs utilized to mitigate the spread of the COVID-19, and lockdown strategies have served as effective input to these simulations and allowed to present a range of multiple output scenarios. The literature so far observed a range of data mining as well as statistical and mathematical approaches. The

Susceptible Exposed Infectious Recovered (SEIR) model is a widely used mathematical technique to evaluate mitigation strategies and NPI measures [211]. The SEIR model relies on various disease outbreak parameters, which the scientific community understands much better now than at the earlier stage of the COVID-19 pandemic. Furthermore, the model allows represents various categories of symptomatic levels, providing a more accurate simulation of the pandemic. Prior work in mathematical modelling has shown the implementation of SEIR for specific regions as well as its ability to be modified to model specific research aims or scenarios such as in [212, 213]. In the work presented in [214], the authors discussed a range of parameters that can be introduced to model COVID-19 and improve the accuracy of SEIR models, as applied to eight countries. Also, the explicit application of SEIR modelling to specific countries has been noted. As an example, in [215], for the case of China, a layer of quarantined patients was incorporated, and those who had passed away ultimately allowed a prediction of peaks in various regions of China. In [216], the SEIR model was modified to include domestic passenger movement data to yet again predict the peak of the epidemic. In [217], the conventional SEIR model was applied to various social distancing mitigation strategies, where sustained application of NPIs was able to mitigate the spread of COVID-19 infection. SEIR modelling was also applied to project the health infrastructure needs, such as ICU beds and hospitalization needs, .in France [218]

The current research case study reports on how the effect of NPI measures can be investigated and compared according to the change of effective reproductive number (R_t) using simulation techniques. Two case studies were selected for simulation. The two simulations donate United Arab Emirates (UAE) and Victoria State in Australia (VIC) in this study. The former case represents a growing urban centre with a highly social and mobile society that has slowly exited its lockdown strategies after an initial outbreak of COVID-19 [219, 220]. It is also a multi-cultural nation with a diverse diaspora, two major air transport hubs and a high standard of living. The UAE deserves individual focus concerning the potential spread of COVID-19 since it can provide valuable insight to other similar countries. It is noted and acknowledged that previous work already discussed NPI measures undertaken by South Asian and Gulf countries to mitigate the spread of COVID-19. However, those lack meaningful modelling results [221]. The later case of VIC illustrates a unique perspective on a regional (second) outbreak of COVID-19 in a large country like Australia. Victoria has had to enforce a new round of lockdown measures. However, in contrast to the UAE, the state government in Victoria levied a clear and sequential staged list of interventions [222], with implementation date availability accurate to the nearest hour. In sum, the two case studies of the United Arab Emirates and Victoria were chosen due to the following apparent differential aspects. At first, Victoria was facing a second wave of COVID-19 and interventions were enforced quickly, in stages and stringently, whereas UAE had gradually

relaxed its lockdown strategies (such as reopening of international flights as soon as possible). Secondly, unlike Victoria, most of the interventions for UAE were not publicly available in a clear chronological form to the authors, for example via governmental websites. Recent research reported that the primary information source for health care workers in the UAE is social media [223] and not authentic governmental sources. The utilization of two distinct case studies allows us to judge the suitability and sensitivity of the proposed model in this study to capture different intervention settings and scenarios.

The application of these simulation models was further considered by the availability of information regarding mitigation strategies for each country. As stated prior, in the case of the UAE, no clear or segmented mitigation strategies were available to the authorship team to guide the model simulation inputs. For VIC, clear public health mitigation strategies were publicly available and were used as model inputs. As such, this provided further evidence of the impact of disclosing and inputting mitigation strategies on simulating COVID-19 spread within different populations. Accordingly, the simulation for VIC considered four main periods of mitigation strategies; that is, a no intervention period followed by three main NPI periods. This case study aims to provide future forecasting estimations about the spread of COVID-19 in the UAE and VIC, Australia, within different scenarios using the SEIR models. Specifically, two additional layers are introduced by splitting infectious into asymptomatic/symptomatic-mild and symptomatic-severe. This discretion is integral for the study of COVID-19 spread because it is established that many patients can go untested due to no visible symptoms. However, the virus can transmit from such patients and many patients diagnosed show mild symptoms [224].

6.2.2 Method

Two different case studies were simulated in the current research, both of which were at different points in their fight against COVID-19. The first one (the UAE case study) evaluated the effects of NPIs when clear policies are neither publicly communicated nor publicly available. The second case study evaluates the potential consequences of adequately disclosing and disseminating state policies (VIC/ Australia simulation).

6.*2*.2.1 Model Description and Parameters

A simulated compartmental model was implemented to measure the spread of COVID-19 using an empirical population sample (across both case studies). The simulated model is built upon an extended version of the SEIR Model [36]. In the first instance, a portion of the UAE population (2,998,325) was used, which is the total empirical population sample in GLEAMviz software in this

study into five compartmental states. The Susceptible (S), Exposed (E), Infected Asymptomatic-mild (I_a), Infected Symptomatic-severe (I_s) and Recovered (R). Similarly, for the second simulation, all inhabitants of the VIC were grouped with the total population of the study sample for VIC (infected areas with COVID-19 in VIC). The constant N ($N = S + E + I_a + I_s + R$) denotes the total population (N = 2,998,325 for UAE simulations and N=6,421,132 [225] for VIC simulation. The categories of the compartments are further described below.

- **Susceptible (S)**: All non-immune susceptible empirical population sample in this study.
- **Exposed (E)**: Latent but not yet infectious or "have no symptoms, and they cannot spread the virus yet".
- **Infected Asymptomatic- Symptomatic-mild (Ia)**: Refers to transmission of the virus from a person who does not develop symptoms or with mild symptoms [226] to another person (not yet latent but suspected).
- **Infected Symptomatic-severe (Is)**: The state of COVID-19 infection can progress to severe disease with dyspnoea and severe chest symptoms [227].
- **Recovered (R)**: Population showing immunity for COVID-19 after infection recovery.

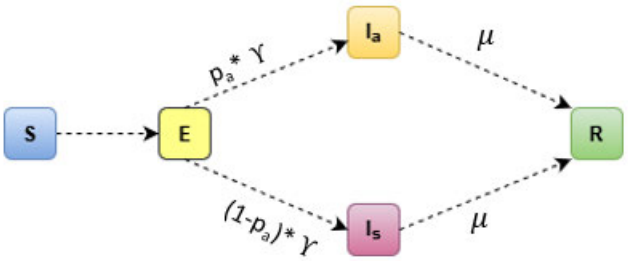


Figure 42. The architecture of the proposed SEI_aI_sR Model

In the course of many diseases, an unknown fraction of the infected hosts are still able to spread the disease while remaining symptoms-free (asymptomatic) [228]. In the proposed model, asymptomatic cases are combined with mild cases in the same fraction according to the WHO reported proportion of the infections [226]. Rationally, we split the asymptomatic and mild into I_a compartmental state and the symptomatic-severe into I_s compartmental state and the understanding of spreading growth for each compartment is a worthwhile attempt. Further, it is deemed as an important research task to evaluate the behaviour of each compartment in the pandemic event and for further compartments' evaluation in relation to the NPI mitigation strategies such as social distancing, lockdown, wearing masks, and more strategies that arise as the pandemic progresses (which reinforces the aim to apply the modelled simulation to two distinct case studies). Therefore, according to the WHO new

classification, the proposed epidemiological model espouses the asymptomatic-mild and severe states [226] for COVID-19 infected cases. The (beta) time-based (t) (β_t) describes the transmission rate varies according to social distancing, remote working, closing schools, wearing masks, etc. Alpha (α) indicates the reduction in the transmission rate of β in the isolated infectious symptomatic (severe), where patients are isolated [229]. The incubation period (Υ) is a period from the state of exposure to the disease to become infectious. This case study used the value of Υ = 1/ 5.2 'days' [230]. The recovery rate (μ) in the proposed model indicates the time until an infectious case is recovered. Previous research [231] tells us the recovery time for COVID-19 is 14 days. This value (recovery rate μ = 1/14 days) is used in the model of this case study. More information regarding the parameters used in this study is discussed in Table 1. The COVID-19 pandemic transmission in this model can be described by:

$$\dot{S} = -\beta t S(I_a + I_s) \quad (1)$$

$$\dot{E} = \beta t S(I_a + I_s) - \Upsilon E \quad (2)$$

$$\dot{I}_a = \Upsilon P_a E - \mu I_a \quad (3)$$

$$\dot{I}_s = \Upsilon(1 - P_a)E - \mu I_s \quad (4)$$

$$\dot{R} = \mu I_a + \mu I_s \quad (5)$$

where N = S + E + I_a+ I_s+ R. The beta$_{(t)}$ (β_t) is calculated according to equation (6) below:

$$\beta t = R_t \, \mu / P_a \, \alpha + (1 - P_a) \quad (6)$$

The model of this case study is compared to the SEIR model [211]. The aim is to investigate differences between the infected (I) simulated cases of the SEIR model against the proposed compartments (I_a and I_s). Furthermore, the effect of this variation on the recovered simulated cases in the comparison of both models is analysed.

The traditional SEIR model equation is formed as the following:

$$\dot{S} = -\beta t S I \quad (7)$$

$$\dot{E} = \beta t S I - \Upsilon E \quad (8)$$

$$\dot{I} = \Upsilon E - \mu I \quad (9)$$

$$\dot{R} = \mu I \quad (10)$$

where N = S + E +I+ R. Figure (2) represents the traditional compartments for the SEIR model. The beta$_{(t)}$ (β_t) is time-dependent. Therefore, β_t is denoted with the following equation:

$$\beta t = R_t \mu \quad (11)$$

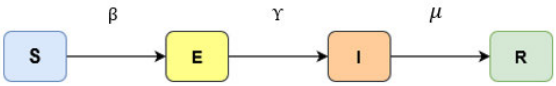


Figure 43. SEIR Model

Figure 42 and Figure 43 compare the dissimilarity between I_a and I_s compartments of SEI_aI_sR to the infected compartment I of SEIR. Equations (1-6) are the COVID-19 transmission equations of SEI_aI_sR in contrast to the SEIR equations 7-11.

Table 21. Model parameters and description

Parameter	Description	Value(s) & Ref
Υ	The incubation period from the state of exposure to the disease to become infectious in both SEIR and SEI_aI_sR	5.2 days [230]
P_{as}	Probability of being Asymptomatic - Symptomatic (Mild)	0.8 [226]
P_{sv}	The probability of being Severe Symptomatic requires hospitalization = $(1 - P_a)$	0.2 [226]
β_t	$Beta_{(t)}$ or transmission rate, which describes the spread of disease in the community.	Equation 6 for ~~S1~~ and Equation 11 for ~~S1~~ (given above)
α	Alpha: Reduction in transmission rate (severe)	0.5 [229]
μ	The recovery rate for (S1 and S2): Indicates the time until the infectious case becomes recovered	14 days [231] [232] μ (1/14 days = 0.07 "day $^{-1}$")

6.2.2.2 Estimating the Effective Reproductive Number

The R_t (effective reproductive number) measures the transmission potential of COVID-19, which is also referred to as the average number of people who will catch the disease from a single infected individual. When the pandemic occurs, the effective reproductive number R_t measures which will become infected per infectious person at a time (t). The most well-known version of R_t is the basic reproductive number R_0. However, the R_0 is a single measure that does not reflect changes in disease transmission, behaviours and restrictions in communities over time. Alternatively, as the pandemic progresses, mitigation strategies could be tightened, more restrictions imposed, or relaxed. This enables R_t to vary over time. Therefore, the R_t value is subject to variation after or before the introduction of NPIs. The real-time Bayesian estimation [233] was used to estimate the R_t and work implementation by [234]. Figures 3 and 9 in the results and discussion section provide the calculated R_t values for UAE and VIC, respectively.

The Modelling Software

The GLEAMviz was utilised with the client simulator [235], combining world data such as country populations and human mobility. The GLEAMviz simulator elaborates compartmental stochastic models [236] for disease transmission in a global epidemic event. To forecast the number of estimated future compartments for the COVID-19 epidemic in the UAE, a previous model, "Global Epidemic and Mobility GLEaM H1N1 schematic" [235], was exploited and depicted spread such as an epidemic disease. The model was modified to include the compartment of asymptomatic-mild and symptomatic-severe layers. It is worth noting that I_a represents the asymptomatic-mild cohort and the I_s represents the symptomatic-severe cohort in the study. Figure *42* represents the schematic for the proposed epidemiological model compartments in this case study.

6.2.3 Results and Discussion

Currently, there is no cure or effective vaccine for Covid-19 while the pandemic continues to spread, and there are more daily confirmed positive cases and deaths recorded worldwide. Aptly, it is necessary to maintain and measure NPIs effectiveness and figure out how to flatten the pandemic curve with long term interventions until successful vaccines are widely available or effective treatment is available. This section reflects the findings from both the UAE and VIC case studies. Particularly, the former, where detailed NPIs are not publicly available, and the latter strategy, where information is made available to the public by the Victoria government's official website [237]. The proposed model is sensitive to the contact rate s that determines the change of the R_t value, which is the essential entry to the simulation to reflect the policy outcomes in real-time R_t measurement.

The GLEAMviz simulations were run to initialize the spread of COVID-19 in the UAE and VIC starting on January 29 and January 20 2020, respectively. Since the GLEAMviz is limited to 365 days, the simulations end on 28/1/2021 for UAE and on 18/1/2021 for VIC, Australia. The proposed model SEI_aI_sR (an extended model from SEIR) was simulated for the situation of UAE and for the case of VIC. After that, the results of the three simulations were compared on an empirical portion of the population of UAE and VIC. The simulations considered the changes in the R_t according to the changes (tightening or easing) in policies for both the UAE, between 29/1/2020 and 2/8/2020, and VIC, between 20/1/2020 and 31/8/2020. The data used to calculate the R_t was fetched from the Github repository of "Our World in Data" [238] in the case of UAE, while VIC data was fetched from John Hopkins' official Github [239]. Furthermore, the data pre-processing step was applied to get the daily new cases from John Hopkins' official Github. The attributes of the data are the date, name of country and number of new daily cases (k) for UAE and for VIC. This step is essential in measuring R_t [233]

[234], and the decision was to change R_t for simulations inputs by 0.5 points of R_t each increase in the R_t or increment. This is assuming the 0.5 value has a noticeable impact on the simulation results.

6.2.3.1 COVID-19 Simulations in Undisclosed Public Health Strategy for the Public (UAE)

Figure *44* shows the real-time R_t for UAE. Since UAE went through different social distancing, restrictions, and easing of restrictions strategies, estimating the R_t is essential to measure and reflect the policy outcomes on UAE and VIC simulations using the empirical data over time. The simulations' parameters' values (Table 21) are the input in UAE and VIC simulations. The strategy in this case study was to update the simulations upon each 0.5 difference in R_t value to adopt the policy changes at a point in time. The model was fed with R_t values in the GLEAMviz exception layer, indicating the policy measured at a time (t). Because of the limitation of the new COVID-19 daily cases as obtained by the Github repository of "Our World in Data" [238], the R_t measure started on 23/3/2020. Therefore, the value of R_0 was kept constant from 29/1/2020 until 23/3/2020. The value of R_0 was set to 2.5 according to the WHO [240] report since there was no available data about R_t at the beginning of the pandemic.

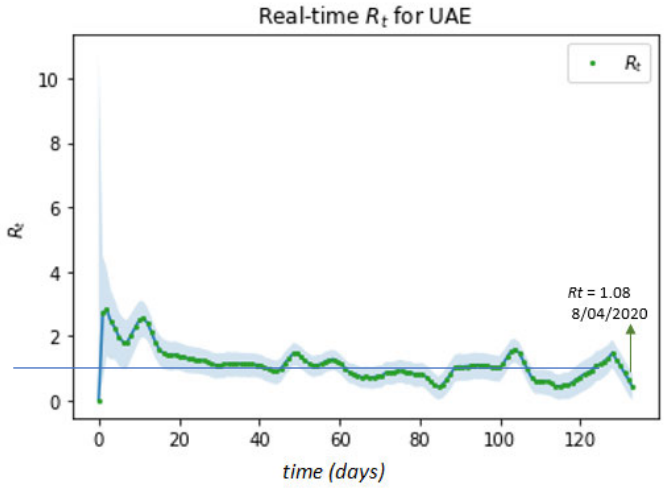


Figure 44. Real-time R_t for UAE from 23/3/2020 to 2/8/2020

R_t in the UAE fluctuated between zero at the beginning of the pandemic, and 3, during the first two weeks of the pandemic. After that, a decline in the R_t was noted, reaching a value of around 1 around the 40th day. Around the 50th day of the pandemic, R_t increased to about 2 and declined after that to less than 1 between the 60th and 90th day of the pandemic. The sharpest decrease in R_t was observed between days 110 and 120 of the pandemic. By the end of the simulation, R_t was noted as 1.08.

Figure 45 (a, b & c) presents the simulated median rate (along with 95% Confidence Interval (CI) of the median) of asymptomatic-mild, severe, and recovered COVID-19 cases per 1000 population. As of 23/05/2020, the median rate (95% CI) of asymptomatic-mild cases was 11.25 (5.62 –12.26) per

1000 population, with a cumulative median of 356.64 (173.07 - 540.17) per 1000 population (Figure 46 a). In the case of severe COVID-19, cases peaked by 20/5/2020, and the simulation predicted that there will be no severe cases after 16/12/2020. The median rate was simulated at 2.83 (1.35 – 3.02) per 1000 population and a cumulative median of 81.36 (39.02 – 126.22) per 1000 population (Figure *46* (b)). Simulated severe COVID-19 cases are an essential compartment to estimate the number of population that may require advanced health services, critical care services, or even hospitalization care. Simulation of severe cases will facilitate evaluating the needs for health services and identifying anticipated needs for patients with severe cases. After that, a simple comparison of estimated numbers and availability of health services will provide a valuable need assessment and identify potential gaps in medical services. However, the lack of healthcare indicators from the UAE limited such comparison and restricted the ability to anticipate the abovementioned gap.

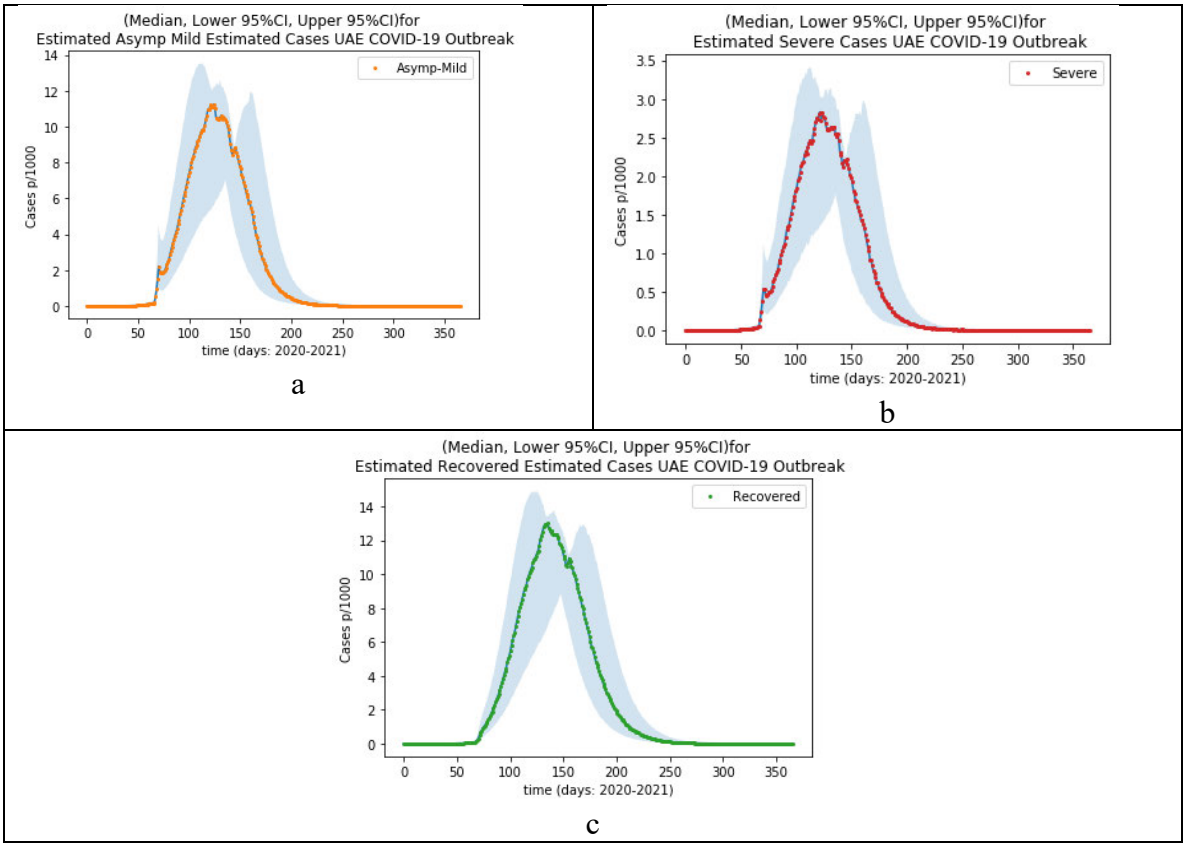


Figure 45 (a,b,c). Median, Lower 95%CI, Upper 95%CI for asymptomatic-mild, severe and recovered estimated cases (UAE)

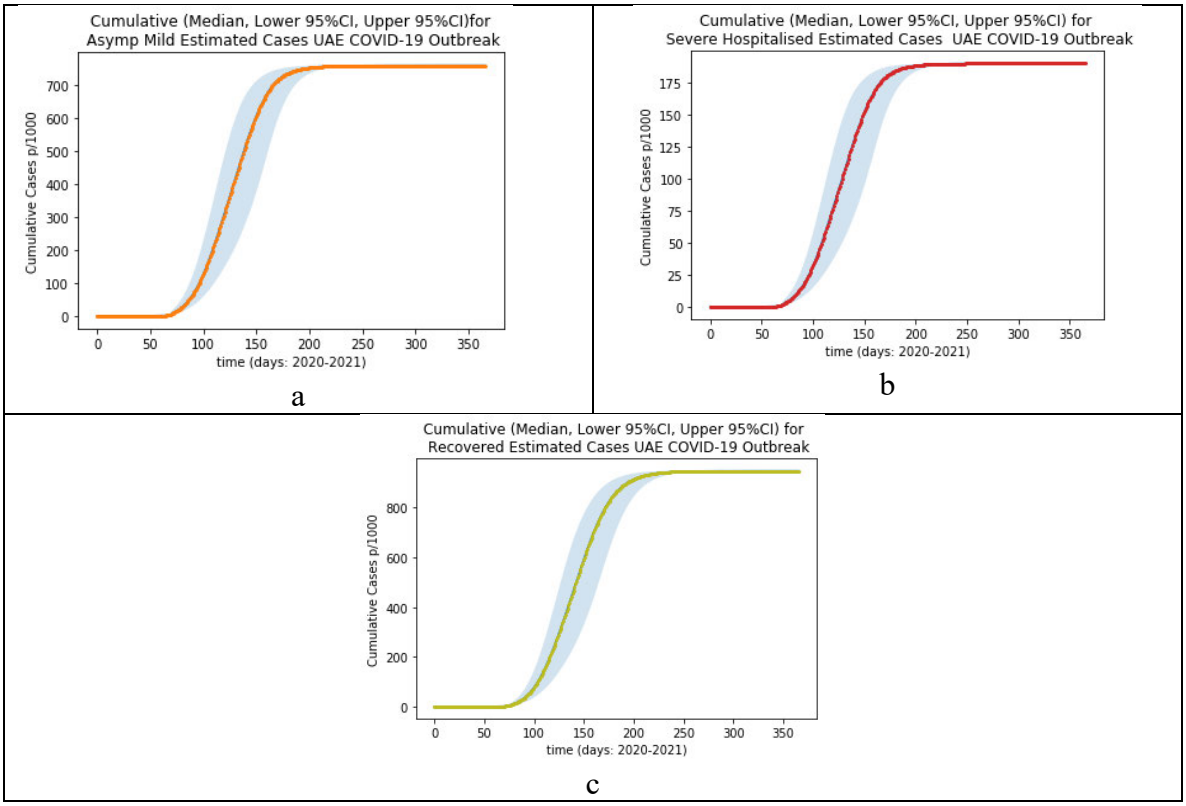


Figure 46 (a,b,c). Cumulative (Median, Lower 95%CI, Upper 95%CI) for asymptomatic-mild, severe and recovered estimated cases (UAE)

Recovered cases in UAE simulation followed the same trends of asymptomatic-mild and severe cases. Figure *45* (c) illustrated that the peak median recovered cases reached 12.98 per 1000 (6.92 – 13.41), and the median cumulative was 418.15 (206.55 - 622.50) per 1000 population (Figure 46 (c)). Thus, the simulation of Figure 45 estimated recovered cases to flatten by 27/1/2021. With reference to the flattening of the curve, in general, it was observed that the reported results are in line with prior literature [241], where it was shown that lockdown and stringency measures are required to be sustained for anywhere between 3 to 5 months to flatten the curve (albeit for the case of United Kingdom).

6.2.3.2 COVID-19 Simulations in Disclosed Public Health Strategies (Restrictions) for the Public (Victoria, Australia)

The first period of COVID-19 Outbreak in VIC to March 28/3/2020

The simulation was run by initialising the spread of COVID-19 in VIC starting on January 20 [242]. The proposed model takes into consideration international travel and mobility simulation. The non-intervention period is the period prior to March 2020 ($R_0 = 2.5$) [243]. The R_t calculations were initiated on 11/3/2020. The R_t measured values fluctuated between zero at the beginning of the pandemic and 3 (14/3/2020) (Figure 47). Therefore, the simulation model on VIC's population data (empirical) is adopted to these variations up to 28/3/2020 and prior March restrictions. The simulation predicted that

at the peak (28/03/2020) for asymptomatic- mild cases, the median cumulative was 0.058346 (0.006884 - 0.127451) per 1000 population. Figures 8 demonstrate the median rates per 1000 for asymptomatic-mild cases, recovered and severe cases in VIC outbreak.

Stage 3 Intervention: from 28/3/2020 and the period up to 31/5/2020)

The Victorian government announced stage 3 [244] to be effective at midnight on March 30 2020 [245]. Victoria's stage 3 indicates that there are four reasons to leave the house: 1) for food and supplies, 2) for exercise, 3) for medical care, 4) for work and education (if necessary), according to the Premier of Victoria [245]. The proposed model simulated stage 3 of Victoria, Australia, along with travel restrictions, to reduce the R_t to less than 1 (Figure 47) between 1/4/2020 and 30/4/2020. However, there was an increase in R_t values that reached $R_t = 1.5$ by 5/5/2020. The simulation model prediction uses the empirical data population of VIC by the end of the intervention of stage 3 (31/5/2020). The reported results from simulation VIC showed the median rate of 0.788229 per 1000 population for asymptomatic-mild, the median rate of 0.452507 per 1000 cases for recovered cohort, and a median rate of 0.19662 per 1000 population for severe cases. Figure 49 (a, b, and c) provide more information about the median rates during the first intervention period.

Stage 2 Intervention: 1/6/2020 30/6/2020

On May 24 2020, the Victorian government [237, 246] announced easing the restrictions to be effective on 1/6/2020, and moved to stage 2, further easing the restrictions [237] on Jun 22 2020. The R_t (Figure 47) was measured, which showed stable R_t values that were below 1.0 until 12/6, when the R_t started to increase again and reached 2 by 6/7/8. This occurred due to the nature of the interactions, such as loopholes in hotel quarantine, social events, ceremonies, community services, sport and exercise, cafes and restaurants, travel, and entertainment [237]. The model estimated the asymptomatic-mild cases to reach the peak median of 3.7 per 1000 population by 30/6/2020 (Figure 48) and the recovered and Severe cases to peak on 14/7/2020 with a median of 3.9 and 0.91 per 1000 population, respectively.

Back to Stage 3 Intervention: 1/7/2020 to 1/8/2020

The Victorian government has announced a number of Victorian postcodes to return to stage 3 (stay home restriction from Midnight 1/7/2020) [247]. On 8/7/2020, Melbourne city and Mitchell Shire turned into an outbreak centre in Australia with a dramatic increase in daily new cases. As a result, the Victorian state government isolated infected areas and performed a lockdown within those areas to contain the spread of COVID-19 cases. Also, strict social distancing (stage 3) was enforced according to the public health act [248], and all Melbourne city and Mitchell Shire residents were instructed to

wear a face mask. The R_t was calculated around 2 by 7/7/2020 and found it to drop again to 1.12 by 1/8/2020.

Stage 4 Intervention (curfew): 2/8-2020 to 31/8/2020

Victorian state government applied further restrictions (stage 4) effective 2/8/2020 for people living in metropolitan Melbourne and stage 3 for the rest of regional Victoria [249]. In this stage, the curfew was applied to ensure people remained at their homes from 8 pm until 5 am, and the only reason to leave home is for work, medical care or caregiving [250]. The curfew measured R_t reached 0.5 by 20/8/2020 (Figure 47). The simulation was kept running until 18/1/2021 (the maximum running time for simulation is 365 days). It was noticed that there would be a steady increase of simulation asymptomatic-mild, severe and recovered cases up to the end of the year (mid of October 2020). However, a sharp and exponential trend for asymptomatic-mild and severe predicted cases will reach the second peak by late December, considering that the restrictions are lifted in mid of September. Therefore, the second (or, in reality, third) wave is predicted to reach its peak by the end of December 2020 with thousands of daily reported numbers in VIC. This estimation considering the value of R_t will be over 2.0 after the end of October 2020. However, the bottom of the second wave was not observed, or the curve flattening was not detected due to the limitation of the modelling software (GLEAMviz), which is limited to 365 days as a running simulation time in days. Figures 10 (a), (b), and (c) present the median rates for cumulative, asymptomatic-mild, and severe COVID-19 cases.

With the increased global concerns about COVID-19, strict NPI measures have become necessary to mitigate the risks associated with COVID-19. Citizens' commitment is critical to controlling the epidemic. When citizens adapt to the NPI measures, a reduction in the spread of the epidemic is expected. The combined efforts from both governments and citizens are, then, critical for designing and adapting effective NPI measures. This is simply reflected in the epidemic curve of the pandemic.

In the current study, the effect of NPIs was assessed utilising R_t using advanced simulation models for UAE and VIC. In both cases, the model established potential evidence of effective NPIs to control the spread of COVID-19, especially when the model's modifications were introduced to meet the characteristics of the pandemic. Adopted NPIs in both geographic locations effectively reduced the adequate reproduction number below one. Further, the results indicated that the rapid introduction of NPIs has a more effective reduction of the spread of the epidemic.

Multiple models evaluated the effectiveness of state measures to control COVID-19 spread. A direct link was established between the effectiveness of NPIs and reducing the reproduction rate (R_t) [251]. As well, strict and stringent state measures were found to have a swift effect in mitigating the spread of COVID-19 [252]. The results are in line with the literature and indicate that effective implementation of NPI measures has, potentially, profound consequences on the epidemic curve of

COVID-19 not only by reducing the number of newly reported or simulated cases but also by reducing the effective reproduction number. The message behind such results is a cornerstone for communicating public health policies and the implementation of NPIs. Of interest when comparing the models from UAE and VIC, the feature that calculated the reproduction rate could be facilitated by the number of reported cases without the need for details related to NPIs (as in the case of the UAE). While this note was not introduced in the literature, it becomes interesting to compare results from different population subgroups from within different cultures. This note is related to the sensitivity of R_t to the contact rate, which is critical in the spreading or containing the spread of COVID-19. Contact rates are, in turn, dependent on NPIs, which are essential for mitigating the disease. Regardless, the study provides evidence that the effect of NPI measures could be evaluated and discussed using the effective reproductive rate. This is an added value to public health professionals and could be used when designing and implementing mitigation strategies, such as a discussion on whether suppression or control is more appropriate. The results showed how a sudden relaxation of interventions in Victoria could lead to another swift outbreak.

On the other hand, traditional SEIR models seem to be limited in assessing the effect of NPI measures on the epidemic curves of COVID-19. This note is directly related to the need to consider the characteristics of the disease, with COVID-19 representing itself uniquely as asymptomatic cases that needed to be fine-tuned when designing the model compartments. Disease severity, therefore, is deemed critical for modelling and simulating the transmission of COVID-19 within populations. Within the uncertainties associated with COVID-19, time will tell if these asymptomatic-mild cases are of more significant concern for disease transmission. Regardless, modelling and simulation techniques should consider modifying the traditional SEIR to present the epidemic curve better.

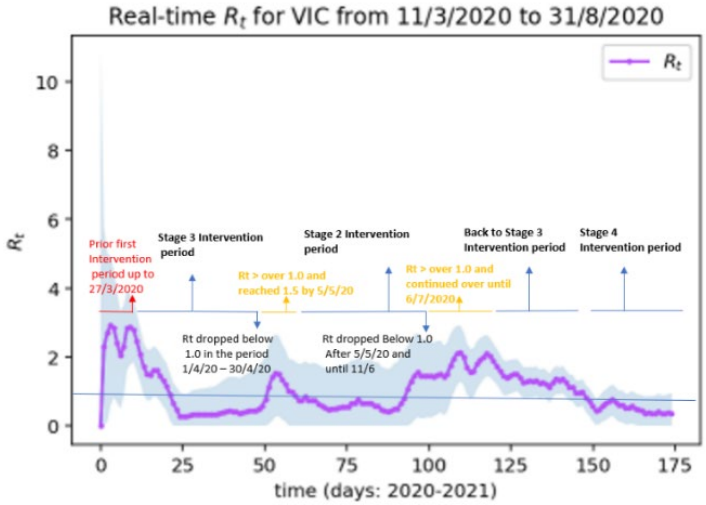


Figure 47. Real-time R_t for VIC in COVID-19 outbreak

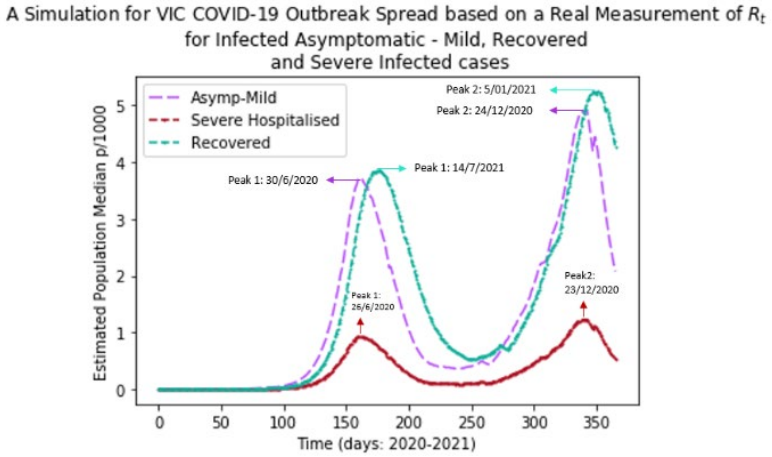


Figure 48. Asymptomatic-mild, recovered and severe cases in VIC, COVID-19 outbreak

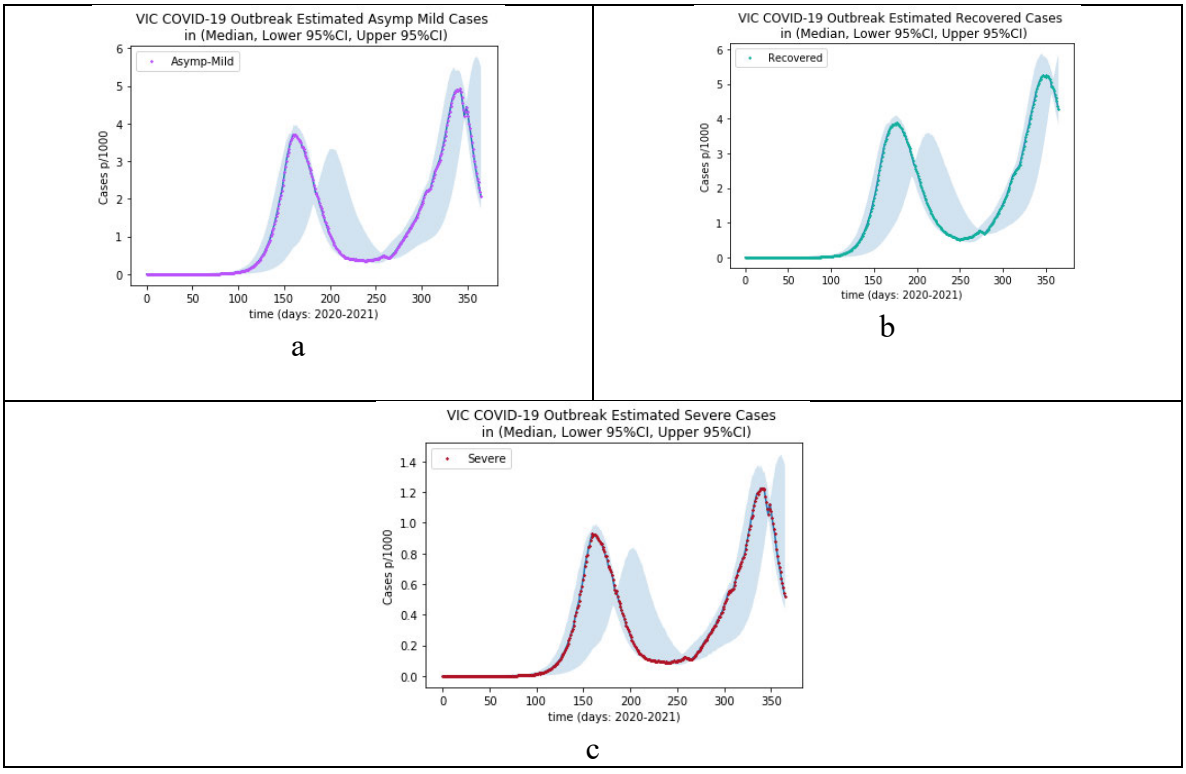


Figure 49. Median for asymptomatic-mild cases, recovered and severe cases in VIC, COVID-19 outbreak

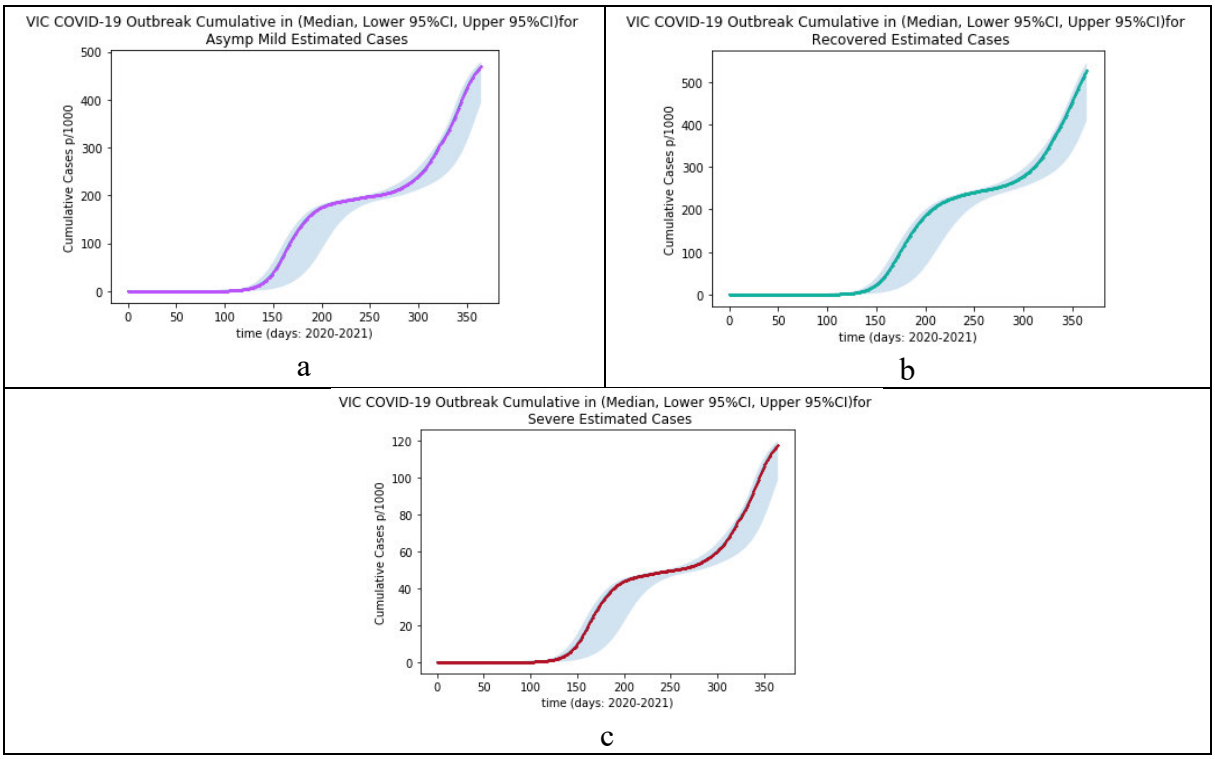


Figure 50. Cumulative (Median, Lower 95%CI, Upper 95%CI) for asymptomatic-mild, recovered and severe estimated cases

6.2.4 Study Limitations

The study provides an interesting outlook on the computation of R_t with respect to stated interventions; however, there are some limitations also associated with the simulations. First, the model is evaluated on empirical population data; it was not examined on real confirmed cases due to the lack of many variables necessary for stochastic compartment models. Second, transmission data may simply not be available or is made private by the authorities, which has ultimately limited the potential to run the model on real-world data and evaluate the predictions of the simulated model against asymptomatic-mild and real severe cases. It was presumed that the entire population of the sample country (in the case of UAE and VIC) is susceptible. Prior work has utilized other ranges, such as 70% [32].

Further, complementary logistic modelling was executed on the scenarios for the UAE. Furthermore, the severe cases were not studied and reported that require hospitalizations. Therefore, a future study should include forecasting severe cases that may require hospitalizations in the model. Glemaviz software application does not allow accessing the mathematical equations used to run the model. This limits the ability to adjust the disease's characteristics within the equations. This may be a reason behind discrepancies in UAE and VIC simulations. Still, Glemaviz is a user-friendly application that

allows public health professionals to run simulation models without an in-depth understanding of advanced mathematical equations.

6.3 Case Study (2) - The Effect of Strict State Measures on the Epidemiologic Curve of COVID-19 Infection in the Context of a Developing Country: A Simulation

Case Study Summary

COVID-19 has posed an extraordinary global public health threat and caused a substantial number of severe cases, which imposed extended hospitalisation and stressed healthcare services in most affected countries. In response to the pandemic, governments introduced a series of non-pharmaceutical interventions (NPIs) which led to severe impacts on the economic and social aspects of society. The impact of these interventions and measures on the spread of the COVID-19 pandemic is not well studied within developing countries. This case study simulated the COVID-19 pandemic curve trajectories in Jordan between February and May and assessed the effect of Jordan's strict NPI measures on the spread of COVID-19. A modified susceptible, exposed, infected, and recovered (SEIR) epidemic model was utilised. The compartments in the proposed model categorised the Jordanian population into six deterministic compartments: suspected, exposed, infectious pre-symptomatic, infectious with mild symptoms, infectious with moderate to severe symptoms, and recovered. A GLEAMviz client simulator was used to run the simulation model. Epidemic curves were plotted for estimated COVID-19 cases in the simulation model and compared against the reported cases. The simulation model estimated the highest number of daily new COVID-19 cases in the pre-symptomatic compartmental state to be 65 cases, with an epidemic curve growing to its peak in 49 days and terminating in 83 days; a total simulated cumulative case count was 1048 cases. The curve representing the number of actual reported cases in Jordan showed good pattern compatibility to that in the mild and moderate to severe compartmental states. The reproduction number under the NPIs was reduced from 5.6 to less than 1. Thus, NPIs in Jordan seem to control the COVID-19 epidemic and reduce the reproduction rate effectively. Early strict intervention measures showed evidence of containing and suppressing the disease. Consequently, this approach helps hospitals to manage their resources, including ICU units, COVID-19 admitted patients, and staffing efficiently.

6.3.1 Introduction and Related Work

Newly evolved coronaviruses (CoVs), such as the severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV), have posed global public health threats, including the 2003 outbreak in Guangdong, China, and the 2012 outbreak in the Middle East, respectively [253]. Similarly, SARS-CoV-2, an enveloped positive-sense RNA virus that

infects humans [254], was initially reported as a localised pneumonia epidemic around December 2019, in China, before being declared a pandemic by WHO in early 2020 [255, 256]. COVID-19, the disease caused by SARS-CoV-2, is today a pandemic and of high priority. As of June 21, 2020, more than 8.5 million confirmed COVID-19 cases and about 500 thousand deaths had been recorded worldwide [257].

While COVID-19 can cause severe illness and death, many uncertainties exist. The full extent of the pandemic, especially in developing countries, the full clinical spectrum of illness, including the prevalence of mildly symptomatic cases [256], and the true case fatality rates[258], are not indeed known. With 81% of infected cases developing only mild symptoms of COVID-19, it was suggested that many infected individuals with mild symptoms might not seek testing [259].

This adds to the uncertainty of COVID-19 [260], especially in developing countries with limited testing and treating capabilities, and may make the true case count as much as 10 times higher than reported [261]. Therefore, projecting a case count is essential for public health response measures and health system management.

Globally, two vital non-pharmaceutical intervention (NPI) strategies have been identified to control the spread of an epidemic: mitigation and suppression. The former focuses on slowing the spread of the disease, but not necessarily stopping it, by reducing the healthcare demand peak and by protecting at-risk groups. On the other hand, suppression focuses on reversing the epidemic's growth, reducing case numbers to low levels and maintaining that situation indefinitely [262, 263].

In developed countries, these measures have effectively controlled the spread of COVID-19 [262, 264, 265]. This effect has been assessed using mathematical modelling that simulated the spread of SARS-CoV-2 infection across the population and shaped control measures that might mitigate future transmissions [262, 264, 266]. One outcome of such a simulation is the predicted epidemic curve, representing the number of infections caused by the virus over time. Using a set of parameters, such simulation measures the impact of different interventions that can directly affect the predicted epidemic curve [266]. Mathematical modelling, therefore, presented itself as a powerful tool for understanding the transmission of COVID-19 and exploring different scenarios. However, using such modelling in developing countries, where healthcare systems are relatively weak, protective equipment is scarce, and testing and treatment capacities are poor [259, 267, 268].

Jordan (the Middle-East region) activated its initial national response to COVID-19 on February 27 by banning non-Jordanian travellers from high-risk countries from entering Jordan. On March 2, the first COVID-19 case was reported for a national arriving from Italy. In the same week, Jordan initiated

a quarantine for arrivals from selected European countries. On March 15, a total of 12 new cases were reported. In response, more restrictions were imposed, where all educational institutions, tourism sites, cafes, and restaurants were ordered to close. All arriving passengers were then handled as suspected cases and immediately quarantined. By March 18, Jordan prohibited travel between governorates, suspended all flights, closed borders, suspended public transportation, closed commercial complexes, suspended non-emergency medical services, closed public and private sectors, implemented a stay-at-home policy, and prohibited public, social, and religious events. Jordan then declared a national lockdown, a state of emergency, and imposed a curfew and mandated wearing face masks in public places, including cars. During the early couple of days of the curfew, a complete nationwide lockdown banned people from leaving their households. Citizens were then allowed five specific days to move around and walk locally, and neighbourhood grocery stores were allowed to open between 10 am and 6 am. Driving was not allowed, and moving between administrative and geographic boundaries, was permitted under emergency circumstances.

The number of newly daily reported COVID-19 cases in Jordan fluctuated between 3 and 42 cases (the mean number of daily reported cases was 15 cases). As of May 1, the number of reported cases was 459 cases, including eight deaths. The reported cases seem to have clustered among persons within the same family, and a limited number of cases have been identified to be of unknown origin. Testing, taking place during the time at which this manuscript was prepared, has been conducted randomly, regardless of symptoms, within each of the 12 Jordanian governorates, and a limited number of cases has been identified using this approach. In early May, the number of local cases reached zero for about 10 days.

At the time of this case study (as discussed above), it was necessary to simulate the COVID-19 epidemic curve in Jordan, especially for those with mild symptoms and severe cases, as this will indicate the actual national situation. Without proper simulation of cases by clinical manifestation, decisions to reopen businesses will be arbitrary and not data-driven. From this perspective, the current research attempted to simulate the ongoing trajectory of the COVID-19 outbreak in Jordan and to model the effect of national interventions utilising real-time scenarios. Furthermore, to measure the effectiveness of NPIs on controlling or preventing any potential patients' stream that is expected to visit hospitals in NPIs is not applied. The simulation of the COVID-19 outbreak could be of added value for public health response planning and future expectations, as well as to assess the external factor that may stress hospitals if no NPIs are applied while the pandemic is progressing. The current research will also advance the knowledge about COVID-19 in developing countries and the effect of publicised responses implemented with widespread adherence and support in Jordan.

6.3.2 Method

6.3.2.1 Description of Data

The simulated daily new mild COVID-19 cases under S1 peaked on March 21, were 36 cases and a total duration of 49 days. After this, the simulated daily new mild case count started to decrease and reached, on Apr 27, zero daily new cases (total duration of the epidemic curve was 87 days). Thus, the estimated cumulative mild case count has reached its maximum at 794 cases around April 27. Under S2, the curve peaked at 174,082 cases around July 1 (a total of 151 days). The simulated daily new moderate to severe cases (S1) reached a maximum number on March 24 with a total of 46 cases (a total of 53 days). The number of simulated daily new moderate to severe cases, then, decreased to zero on April 27 (the total number of days for the epidemic was 87 days). Under S2, the curve peaked at a simulated daily new cases of 150,523 on July 3 (a total of 153 days).

6.3.2.2 Model Description

A modified susceptible, exposed, infected, and recovered (SEIR) epidemic model [269] (case study 1) to simulate the spread of COVID-19 in Jordan was utilised. The SEIR model simulates the spread of infectious disease, assuming that no births or deaths occur and no new individuals are introduced. As such, each individual is initially assigned to each of the following disease states (deterministic compartments): susceptible (S), exposed (E), infectious (I) or recovered (R). The deterministic compartments in the SEIR model are fairly sophisticated quantitative mathematical models yet are efficiently run utilising public data and known disease characteristics [269]. The standard SEIR model was modified by adding compartmental states that reflect the compartmental population and research needs.

The modified model categorised the Jordanian population into six deterministic compartments: susceptible, exposed, infectious pre-symptomatic (representing the total number of infections in Jordan), infectious with mild symptoms (i.e., not needing hospitalisation), infectious with moderate to severe symptoms (i.e., requiring hospitalisation), and recovered. In designing the modified simulation model, it was assumed that an exposed individual might become infectious, pre-symptomatic, and then may progress to recover, or progress to become either a mild or moderate to severe symptomatic individual, both of whom may then progress to recovered. The following brief shows the compartmental states applied in this study:

- **Susceptible**: all of the non-immune population in the case study (the entire Jordanian population).

- **Pre-symptomatic**: population producing or showing no COVID-19 symptoms yet, albeit infectious [263].
- **Symptomatic** (mild or moderate to severe): population showing COVID-19 symptoms.
- **Recovered**: population recovered from COVID-19 infection.

The modified model predicts the number of simulated COVID-19 cases by each compartmental state in Jordan. It also has the potential to distinguish hidden (asymptomatic or mild, not seeking hospital care) from identified infected cases needing hospitalisation (moderate to severe cases). Indeed, standard SEIR models are estimated by assuming that all infected people are reported. Such an assumption for the novel COVID-19 pandemic is unreasonable mainly, as many infected people show no symptoms or mild symptoms and, as the testing procedure is not available in mass, many remain undetected [270]. The model also accounts for the hospitalisation of moderate to severe cases by adjusting the contact rate. It is assumed that such cases will be detected and quarantined within a healthcare setting as they will be seeking medical services. Hence, their contact rates will decrease tremendously.

6.3.2.3 Simulated Model and Modelling Software

The GLEAMviz was utilised using client desktop application version 7 simulator [271] that combines world data such as country population and human mobility. The GLEAMviz elaborates compartmental stochastic models [272] for disease transmission in a pandemic event. The analysis assumes that the first case entered Jordan on February 1, and the initial simulation started as such. A population size of 10.2 million was built into the client simulator. Moreover, the model allows for the limitation of mobility within the population and the restriction of travel as built-in functions within the designed models. The simulator provides rates within each compartment which were converted into numbers based on population size.

6.3.2.4 Model Parameterisation

A series of parameters were utilised to run the simulation model, as indicated by the simulator (Table 22) [272].

Table 22. Model parameter descriptions and values used for simulating the number of COVID-19 cases in Jordan.

Parameter and Symbols	Description	Scenario 1 Values
β (beta)	Describes the transmission rate	February 1 to March 17 = 0.37 March 18 to April 24 = 0.06 April 25 to May 15 = 0.20

		After May 15 = 0.37
α (alpha)	Reduction in transmission rate. (Moderate to Severe)	0.5
ε (epsilon)	The incubation period from the state of exposure to the disease to become infectious	1/5.2
Ps	Probability of developing severe SAR-CoV-2 symptoms	0.01
μ (mu)	Recovery rate	1/14 days
R_0	Basic Reproduction number	5.6

These parameters are as follows: Beta (β) describes the transmission rate and the spread of disease in the community. The β varies according to public health policies enforced or applied in communities, such as pandemic containment, social distancing, remote working, closing schools, etc. Since Jordan's culture is homogeneous, and people follow traditional forms for greeting, the standard contact rate (β) was set to 0.37 [231, 273, 274]. To reflect the status of measures in Jordan, an extra layer (exception) was added to designate the non-pharmaceutical interventions (NPIs) that took place on March 17. The contact rate value (β) was reduced from 0.37 to 0.06 [275] between March 17 and April 24. The contact rate value (β) was set to 0.2 between April 25 and May 15, reflecting the curfew's partial lifting and reopening of selected businesses. After that, the contact rate value (β) was set to its original value of 0.37. Alpha (α) denotes the reduction in the transmission rate of hospitalised (moderate to severe) cases. The value of α = 0.5 was used to reflect the negligible transmission rate of hospitalised patients. Epsilon (ε): the incubation period from the point of exposure to the disease becoming infectious. It is set to 5.2 days [261, 262, 276]. Ps: the probability of developing severe COVID-19 symptoms. This value was set at 0.01 [277]. Recovery rate (mu or μ) indicates the time until an infectious case becomes recovered. Previous research [278] reports that the recovery time for COVID-19 is 14 days (μ = 1/14 days). Hence, this value was used as the proposed model's recovery rate (μ = 0.07). R_0: the reproduction number for COVID-19. Based on the above values, R_0 was calculated as 5.6 (see Supplementary Table S1 for the formula). The basic reproduction number (R_0) measures the transmission (contagious) potential of COVID-19 and describes the average number of secondary infections caused by a typical primary infection in a completely susceptible population. An R_0 value of 5.6 was reported in other similar global simulations [279]. The literature reported that R_0 ranged between 2.3 and 6.5 [230, 231, 280, 281] and a re-analysis of Chinese data provided an updated

estimate of 5.7 (95% CI 3.8–8.9) [281]. Other published studies reported that for social gathering events such as wedding parties in Jordan.

Table 23. Model parameters' description and values are used for simulating the number of COVID-19 cases in Jordan under the hypothetical scenario of no-action (S2).

Parameter and Symbols	Description	Scenario 2 Values
β (beta)	Describes the transmission rate.	0.37
A (alpha)	Reduction in transmission rate (moderate to severe).	0.5
ε	The incubation period from the state of exposure to the disease to become infectious	1/5.2 days
P_s	Probability of developing severe SAR-CoV-2 symptoms	0.01
μ	Recovery rate	1/14 days
$R_0^{\#}$	Basic Reproduction number	5.6

The R_0 formula that was used in this case study is:

$$\beta = R_0 / \mu + \varepsilon$$

where μ is the recovery rate, and ε is the rate of the incubation period.

The simulation in this case study does not provide estimates for the proportion requiring intensive care units (ICU) within hospitals nor the estimated number of COVID-19-related deaths. Providing these estimates requires details of the clinical fraction of infected people, the likelihood of clinical cases being severely ill, and a detailed understanding of the capacity of the health services in Jordan.

Two basic models were run to simulate the estimated numbers of COVID-19 cases by clinical manifestation, assuming two separate scenarios: the NPI scenario (S1), which was implemented in Jordan, and the no-action scenario (S2). The former considered NPI implementation dates (starting March 17 and ending May 15), while the later assumed no NPIs took place (Table 23). For each compartmental state, the number of simulated daily new COVID-19 cases was plotted. Accordingly, the epidemic curves are presented along with the duration of the epidemic (in days) and the time to the peak (in days). Each S1 curve was also fitted against the reported daily number of cases.

6.3.3 Results

Figure 51 presents the number of daily new COVID-19 cases in the pre-symptomatic compartmental state, simulated under the S1 and S2 curves using the same scale. The S1 curve is demonstrated as a "baby" curve under the S2 curve that started after February 1 and ended before April 20. The simulation model, under S1, predicted that on March 20, the highest number of daily new cases in the pre-symptomatic compartmental state would be 65 cases, after which the number of simulated daily new cases would start to decrease. By April 24, the predicted daily new cases had levelled out to zero. Considering that the simulation was set to start on February 1, and the NPIs commenced on March 17, it took the epidemic curve 49 days to grow to its peak, and the total duration of the epidemic curve was predicted at 83 days. The cumulative number of cases was predicted at 1048. For the hypothetical scenario of no action (S2), the epidemic took a total of 147 days to reach its peak of 238,142 daily new cases by June 27, and the cumulative number of cases reached about 9.5 million around December 1.

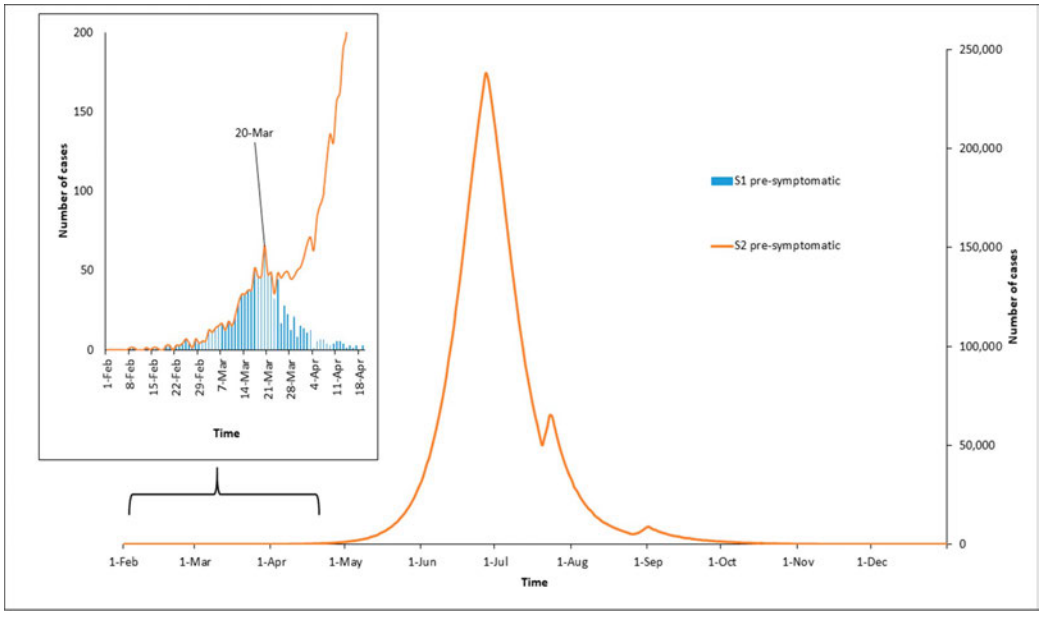


Figure 51. Simulated COVID-19 epidemic curves in Jordan under scenarios 1 and 2 (S1 and S2), utilizing the pre-symptomatic compartmental state

The simulated daily new mild COVID-19 cases under S1 reached their peak on March 21 with 36 cases and a total duration of 49 days (Figure 52), after which the simulated daily new mild case count started to decrease and reached, on April 27, zero daily new cases (the total duration of the epidemic curve was 87 days).

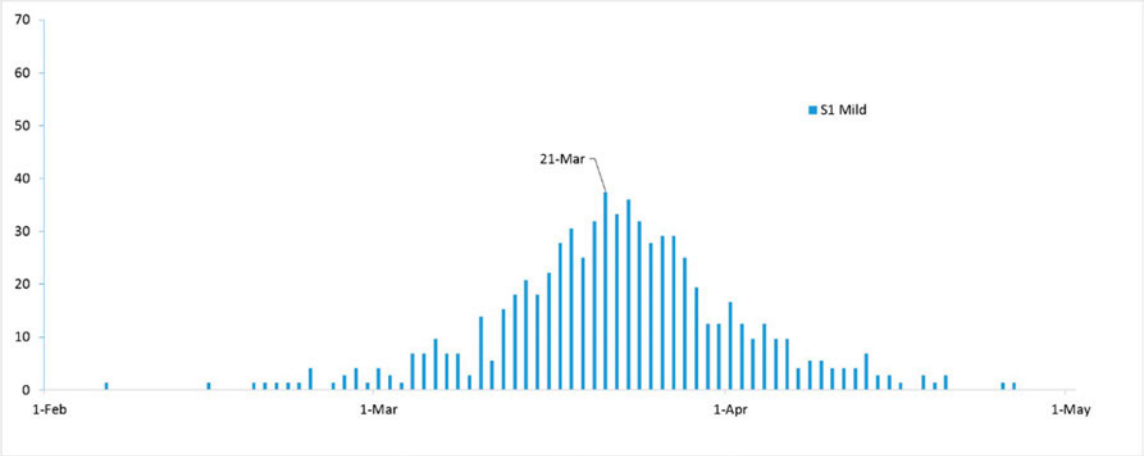


Figure 52. Simulated number of daily new COVID-19 cases in the mild compartmental state under scenario 1 (S1)

As seen in Figure 53, the simulated daily new moderate to severe cases under S1, reached a maximum number on March 24 with 46 cases (a total of 53 days). The number then decreased to zero cases on April 27 (the total number of days for the epidemic was 87 days).

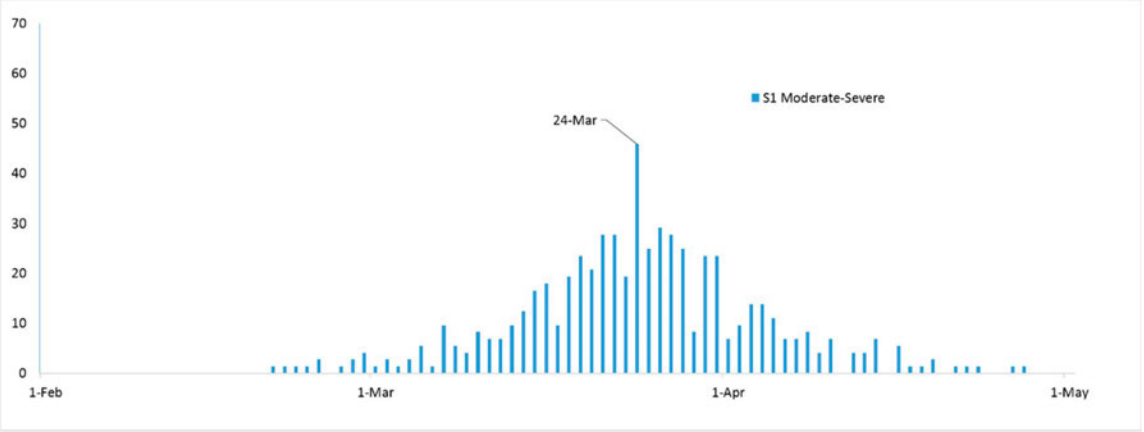


Figure 53. Simulated number of daily new COVID-19 cases in the moderate to severe compartmental state under scenario 1 (S1)

In Figure 54, the actual reported daily new cases were plotted in Jordan against the simulated cases in model S1. The curves representing the simulated number of daily new COVID-19 cases in both the mild and moderate to severe compartmental states had good pattern compatibility with those depicting the number of reported cases in Jordan, with a peak of new cases on March 24.

Under S1, the simulated cumulative recovery was 1044 cases by Jun 30. Out of the total cumulative cases, 695 cases were in the moderate to severe compartmental state, i.e., needing hospital care. At the same time, 795 were in the mild compartmental state, i.e., mostly hidden cases within the community.

Moreover, based on the S1 model, the simulated reproduction number (R_0) for COVID-19 after implementing NPIs in Jordan was estimated at 0.9.

Further comparisons between the S1 and S2 simulated models are presented in Figure 55.

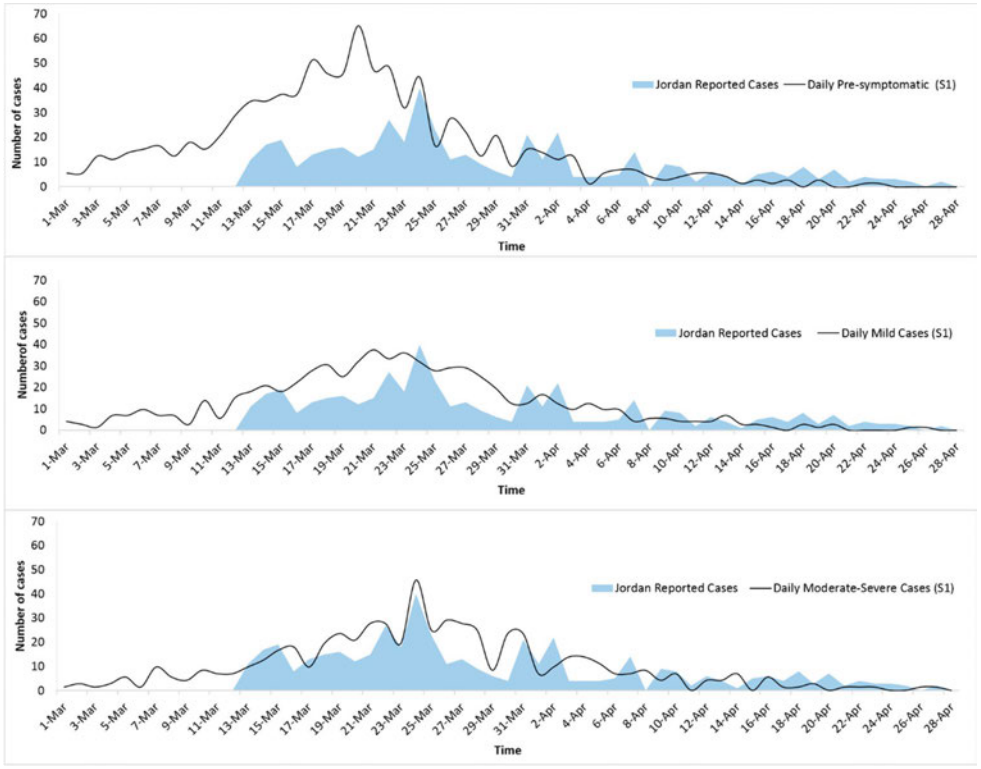


Figure 54. Number of daily new reported COVID-19 cases compared to S1-simulated numbers in the three compartmental states

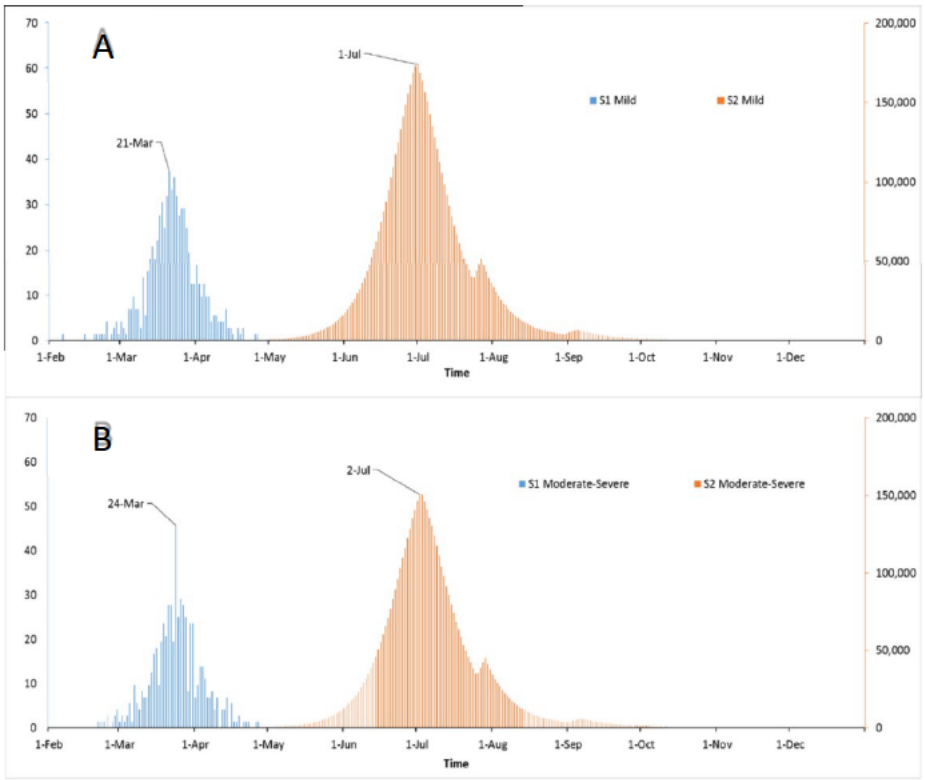


Figure 55. Simulated COVID-19 epidemic curves in Jordan under scenarios 1 and 2 (S1 and S2), utilising the (A) mild and (B) moderate to severe compartmental states.

6.3.4 Discussion

With COVID-19 imposing global public health and socioeconomic uncertainties, governments are counting on their people to adapt to NPIs to reduce the epidemic's impact. The combined efforts of both the government and the people are necessary to control the outbreak and eventually curb the possibility of having a large stream of hospitalised patients. Therefore, how people react and respond to the implemented NPI measures is critical to the epidemiological presentation of the epidemic. In this context, the current study assessed the effect of NPIs implemented in Jordan on the COVID-19 outbreak, utilising simulation techniques. The simulated epidemic curves for COVID-19 provided evidence that Jordan may have successfully implemented NPI measures. They kept the hospital in a natural operation state while facilitating suppressing (containing) the spread of the epidemic by reducing the number of daily new reported cases and the total duration of the epidemic. The effects of the adopted NPIs in Jordan on the number of daily new cases and the duration of the epidemic are even more appreciated when compared to the catastrophic impact of the hypothetical scenario of no action (Figure 55). The results suggest that swift, intensive, and targeted lockdowns in Jordan may have caused new COVID-19 cases to plummet and the health system to be protected. Therefore, this case study suggests that a strong containment policy implemented early on can combat the spread of a COVID-19 epidemic and effectively protect hospital resources from the overabundance of patients.

A recent study, which utilised statistical modelling based on Google reports on social distancing, assessed lockdown efficiency for 13 countries. Jordan, Italy, and Indonesia were categorised as countries with very high-level lockdowns. When correlating lockdown procedures and the infection rates to assess the impact of lockdown policies on R_0, Jordan was reported as a country with high lockdown efficiency for the period between February 15 and April 11. However, Italy and Indonesia were categorised as medium lockdown efficiency countries for the same period. Similarly, Germany and Spain were reported as "not gaining any productive results out of the lockdown procedures" for the same period, yet their efficiency levels improved between May and July.

On the other hand, India was reported to have a very strict lockdown policy yet was categorised, initially, to have a low lockdown efficiency. Later on, India was categorised to have a medium efficiency (between May and July). Late lockdown procedures detected in Brazil and the United States were reported to have a major impact on large outbreaks and to inversely contribute to elevated infection rates [251]. These results are in line with the simulated proposed model for Jordan. They suggest that the country has presented a successful strategy that allowed for the "snuffing" out of the COVID-19 pandemic early. Such success may be attributed to early adaptation to a complete national lockdown, early isolation of all arrivals and travellers for two weeks, and effective contact tracing through the already established crises management centre, which facilitated centralised decision making.

The Jordan Ministry of Health conducted national seropositivity (immunoglobulins M (IgM) and immunoglobulins G (IgG)) study to assess the effect of its measures in combating the spread of COVID-19. This comes as a continuation of the Ministry's random PCR testing after the simulated first wave. The positivity rate of the PCR test for about 700,000 randomly collected samples was less than 0.03%, while the positivity rate for SARS-CoV-2 antibodies is less than 1% for about 500,000 tests performed so far (Jordan's population is estimated at 10.2 million). The later results are still being updated, but the positivity rate aligns with the reported national numbers and the random PCR testing results. Thus, both results seem to point to the effectiveness of state measures in combating COVID-19 and support the case study findings.

Strict NPI measures implemented in Jordan, which lasted for more than six weeks, appear to have reduced COVID-19 transmission and likely reduced the reproduction number to less than 1. A similar discussion was presented for the UK, for example [265], where, in the absence of control measures, the epidemic would quickly overwhelm the healthcare system. A combination of moderate interventions (school closures, shielding of older groups and self-isolation) was predicted to be

unlikely to prevent an epidemic that would far exceed the available ICU capacity in the UK. More intensive lockdown-type measures, however, indicated adequate protection of the healthcare system and hospitals from being overwhelmed. Importantly, the lockdown scenario for the UK effectively reduced R_0 to near or below one [265].

The study results are significant for public health decision-makers and risk communication and lessons learned. In case a new wave of the epidemic hits, the notion to initiate strict measures is supported by this model's outcomes and would strengthen public messages to enhance the proper implementation of strict measures. In addition, this data-driven approach is vital to ensure population commitment and to, perhaps, aid the ongoing efforts of other countries with similar resources and cultures.

6.3.5 Study Limitations

In infectious disease epidemiology, sensitivity analysis provides an insight into how the uncertainty of the model inputs affects the model output and which input tends to lead to variation in the output. Unfortunately, the GLEAMviz simulation software application does not provide compartment modelling in the form of accessible algorithms. Therefore, the inputs of the compartments are the only parameters that the end-user can control. This limited the research in this study's abilities to examine the algorithm of GLEAMviz and conduct sensitivity analyses. In a future study, the simulation could be further improved by introducing epidemiological compartmental models in computational algorithms to be evaluated with a suitable sensitivity analysis. However, the Susceptible, Infected and Recovered (SIR) original model is a standardised one that has been in use for several years in the epidemic investigation. Significantly, optimising the model parameter values to facilitate a proper agreement between the simulated and reported COVID-19 cases (as presented in Figure 54) improves the model's validity.

A combination of NPIs, isolation and contact tracing has been reported to present a synergistic effect that increased the prospect of containment of COVID-19 [282]. Knowing that Jordan has implemented strict contact tracing and isolation of contacts limits the ability to clearly compare the actual reported numbers to those presented under S1. Until detailed information about cases identified via contact tracing and isolation is made available, the presented model (S1) is the only available method to meet the objective of the current study. Moreover, the numbers presented under S2 seemed to be of high values, as the scenario assumed that no prevention and control measures were implemented. Their interpretation, therefore, should be limited to a comparison with S1 and should be seen as mostly hypothetical.

The simulation presented in the current study has limitations. It was designed to monitor the evolution of the COVID-19 epidemic spread in Jordan, utilising parameters presented about the disease from experience within developed countries. However, at this stage of the epidemic, country-specific parameters are not available. Furthermore, the contact rates used in the current simulation were generalised for the whole population and did not consider variability within households or local communities. However, the assumption of a universal contact rate used in the proposed model was adjusted for all cases with moderate to severe clinical manifestations, considering that these cases are most likely to be detected within healthcare settings and be hospitalised, where their contact rate was reduced to its minimum to overcome this limitation.

Furthermore, recognising co-morbidities within the population structure of Jordan and incorporating them within the compartmental states is assured of added value in this simulation. However, the reports from Jordan did not specify co-morbidities and only stated the number of cases. As stated before, this is a limitation to the proposed model and limits the research abilities to assess and compare. However, the aim was to evaluate state measures and compare simulated numbers to reported ones. Also, the case study aimed to look at the number of cases that require hospitalisation and measure its effect on the healthcare system and hospitals (external factor). Therefore, future research should consider this population structure of co-morbidities and fine-tune the results to reflect such factors within simulation models.

6.4 Case Study (3) - Predicting the LOS for COVID-19 Hospitalisation from Linkage Data Sources

Case Study Summary

This case provided an investigation of the role of length of stay predictive models used in this study to predict COVID-19 LOS from admitted inpatient hospitalisations (de-identified). This case study aims to provide a fundamental and initiate research approach to predict LOS for the communicable infectious disease that acquires hospital admission for certain cases in the pandemic. The case study used an open-access dataset for hospital admissions [283] to predict COVID-19 hospitalisations based on various public health measures. This case study supports healthcare professionals in ICU and hospital settings to look at external factors (public health NPIs) studied (cases 1 and 2) and how un-controlled NPIs may lead to increased hospital occupancy. Furthermore, the case study is the first in the literature to baseline the predictive models against deep neural networks in predicting inpatient LOS to provide a practical research framework to improve the workflow in hospital resources utilisation, to the best of our knowledge.

6.4.1 Introduction and Related Work

Predicting LOS for COVID-19 hospitalisations requires understanding factors that demand occupancy of hospital beds, such as internal and external factors. At the same time, external factors are studied in case studies 1 and 2 of this chapter in this thesis. Therefore, managing beds occupancy efficiently in hospital settings requires understanding LOS for COVID-19 inpatients, which is a crucial task to achieve. However, the literature reported trivial attempts [284-286]. While these attempts are limited to private datasets that each belongs to one hospital choice, the literature did not report nor attempt to examine the scalability of the predictive modelling using a larger dataset. This motivates this case study to explore the LOS COVID-19 as the first research attempt, which provides a methodological approach to deal with LOS prediction from multiple hospital resources to measure the COVID-19 pandemic impact on hospital beds occupancy from the LOS prediction approach.

This case study supports the first case study and evaluates the importance of the existence of thorough research attempts to assess epidemics/pandemics' impact on hospital resources' utilisation based on external measures (e.g. NPIs) and internal factors such as the ability to

predict COVID-19 inpatients during the progress of the pandemic. Until the time of the thesis submission, COVID-19 is progressively and rapidly spreading around the world. Therefore, this case study is ground research that contributes to predicting infectious diseases inpatients into hospitals (hospitalised cases) and assists hospital managers, public health and health informatics researchers in understanding the COVID-19 LOS further.

6.4.2 Method, Reported Results and Discussion

This section describes the dataset choice, the steps toward the predictive algorithms benchmarking, and the outperforming model for the COVID-19 LOS prediction. The chapter provided a pilot study that compares the prediction capabilities of the best outperforming ensembles of machine-learning in this thesis against the performance of the deep neural networks in multi-class classification. To this view, the case study links the external hospital factors such as the different NPIs measures to demand for hospital and ICU resources, then guide hospital managers and healthcare workers to be prepared on this ground. The proposed COVID-19 LOS in this study is a robust predictive architecture that can be a valuable tool to baseline the COVID-19 LOS prediction models in hospitals and ICUs based on multiple admissions types. The following sections describe the study design, while this section refers to the methods followed in previous chapters to carry out this case study.

6.4.2.1 Data Description

The dataset [283] used to achieve the aims of this pilot compromised admission features related to patients' hospitals and length of stay on a case basis. The dataset is de-identified and available via google dataset (open-source). The dataset (linkage dataset) combined multiple hospitals' COVID-19 hospitalisations and joined them in one dataset, as described in Table 24. To utilise the dataset in this research, it was coordinated and added to the existing ethical approval number (H13511) by the Western Sydney University ethics committee. The ethical approval is in the Appendix of Chapter 1. The dataset attributes are described in Table 24.

Table 24. Dataset Features description

Feature	**Value/description**
case_id	Admitted case id
Hospital	Name of the hospital
Hospital_type	Hospital type
Hospital_city	City of the hospital
Hospital_region	Region of the hospital

Available-Extra-Rooms-in-Hospital	Number of Extra rooms available in the hospital
Department | Department overlooking the case	Department Name | ['radiotherapy' 'anesthesia' 'gynecology' 'TB & Chest disease' 'surgery']
Ward_Type	Type of ward: ['R' 'S' 'Q' 'P' 'T' 'U']
Ward_Facility	Ward Facility: ['F' 'E' 'D' 'B' 'A' 'C']
Bed_Grade	Condition of Bed in the Ward
patientid	Patient id
City*Code*Patient	City Code for the patient
Type of Admission	Admission Type registered by the Hospital: ['Emergency' 'Trauma' 'Urgent']
Illness_Severity	Severity of the illness recorded at the time of admission: ['Extreme' 'Moderate' 'Minor']
Patient_Visitors	Patient Visitors
Age	Age category: ['51-60' '71-80' '31-40' '41-50' '81-90' '61-70' '21-30' '11-20' '0-10' '91-100']
Admission_Deposit	Deposit at the Admission Time
Stay_Days	Stay Days by the patient (LOS): ['0-10' '41-50' '31-40' '11-20' '51-60' '21-30' '71-80' 'More than 100 Days' '81-90' '61-70' '91-100']

6.*4.2*.2 Exploratory Data Analysis (EDA)

The data contain a large number of records (case-id 318438 admitted cases) to multiple hospitals in the dataset. Therefore, investigating the dataset is deemed an essential task before going to the predictive stage. The EDA helps to evaluate the attributes before carrying out predictive tasks. Furthermore, EDA aims to define whether a relationship exists between data attributes and the target (LOS = Stay_Days). EDA approach was followed to deal with the large nature and discover data insights for decision-making in a hospital or healthcare assessment system, perhaps, predictive insights if the data can guide the prediction stage further. The EDA used state-of-the-art EDA methods ([287], [288]) to achieve this task, which helped analyse data features further and get data insights.

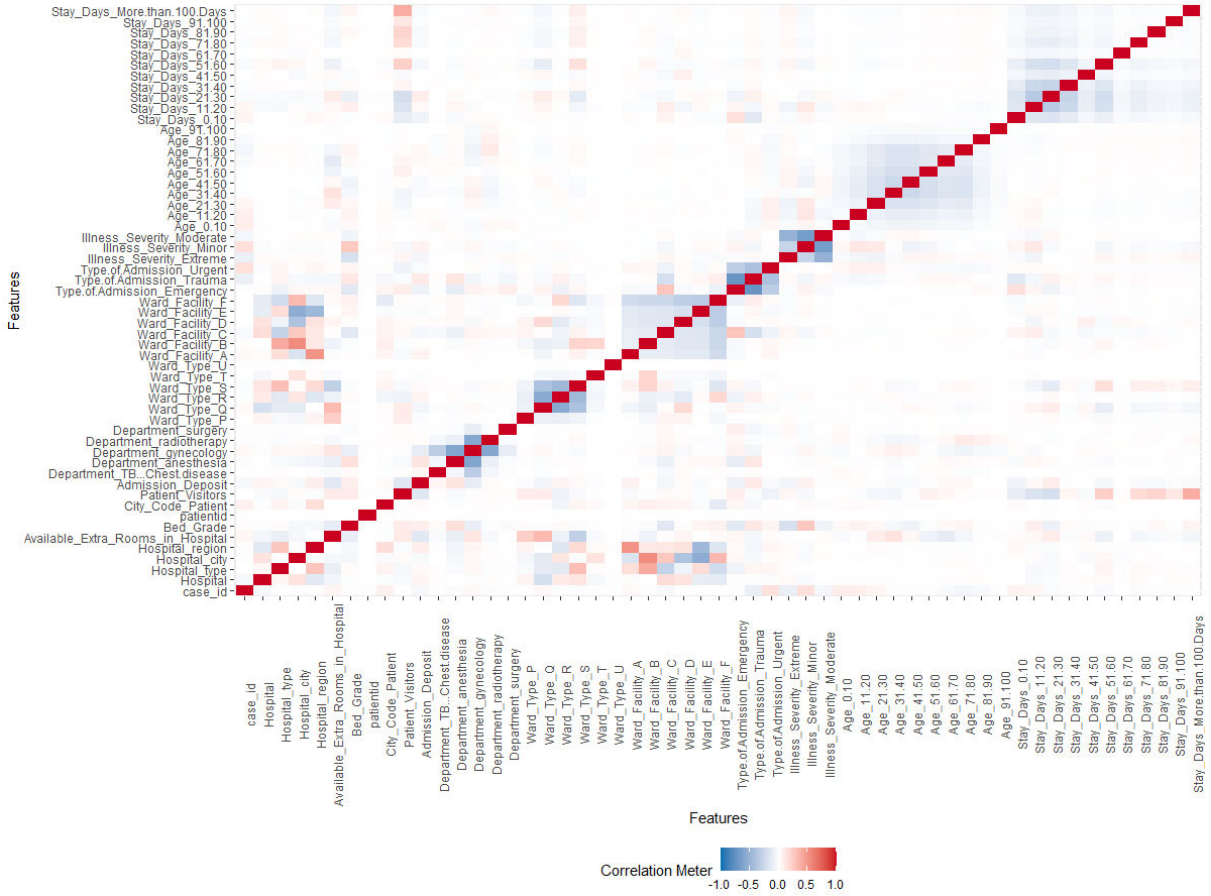


Figure 56. Correlation matrix for the dataset attributes and values of attributes

The first step is to look at the correlation (relationships) between features and their values. Therefore, a correlation was performed in Figure 56 to assess the features' correlations to the target variables LOS (Stay_Days). It was noticed that more hospital stays mean better hospital wards. Illness severity was associated with more extended stays, which is justified by the acute care requirements for the admitted severe cases. Patient visitors were associated with more extended hospital stays, and this is naturally described because of the families or people concerned about patients' health outcomes during the pandemic. As COVID-19 is a new disease, the health outcomes of patients who stayed for prolonged LOS were not clear to their families and beloved ones. Most of the other attributes were less statistically significant to the LOS (Stay_Days).

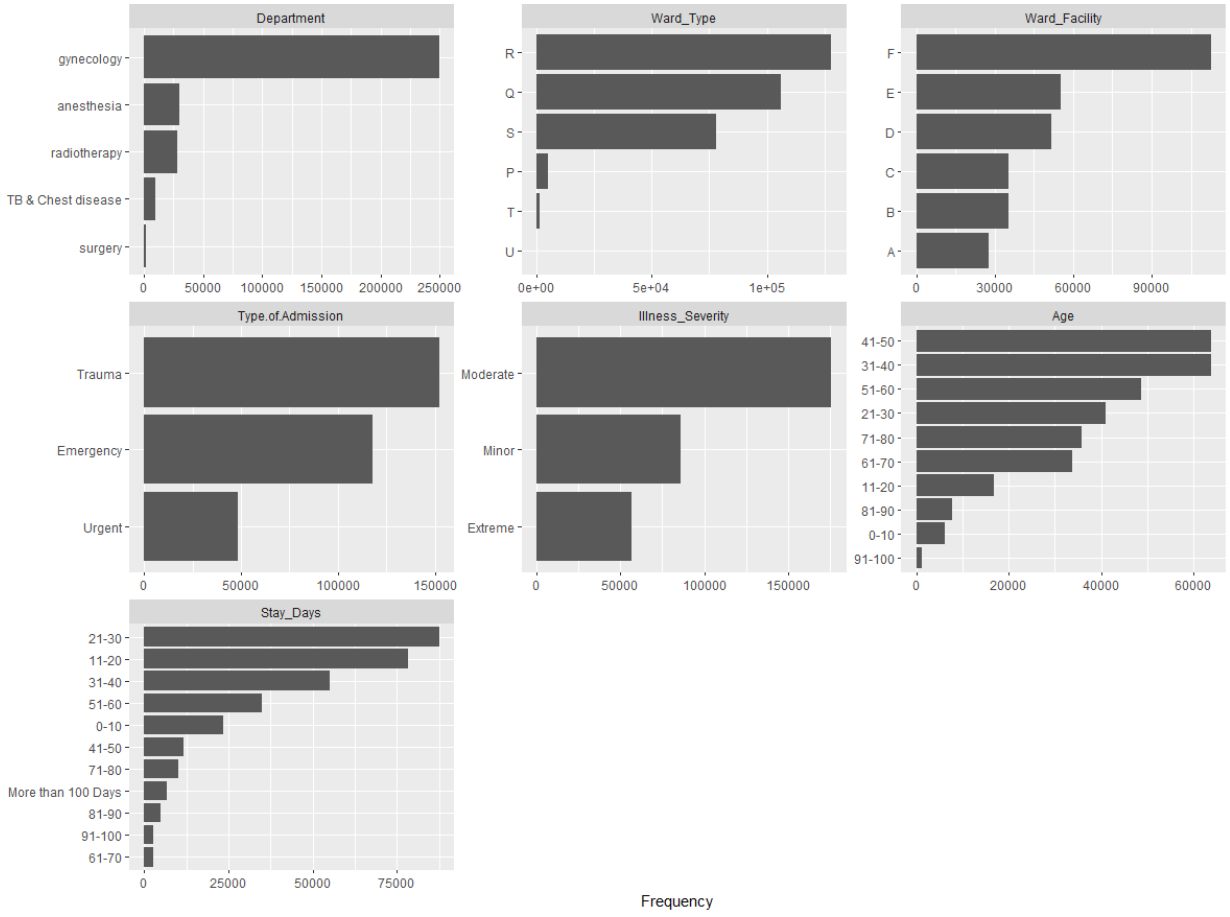


Figure 57. Distribution of categorical variables

Some variables were of the categorical variables EDA interest, as shown in Figure 57. Based on the EDA for the distribution of categorical variables in the dataset (linkage dataset). Majority of COVID-19 admission transferred to the gynecology department. This is an interesting finding as the dataset did not report the gender of admitted patients. The second most important department by admission is anesthesia. However, since there are no clinical inputs about the patients, it was not possible to correlate the department admission to the clinical procedure followed for each admitted case. Wards type values R, Q, and S were the most dominant observed values in the dataset. Similarly, the ward facility's F, E, and D values were the most frequent observed values. Looking at the type of admission, it was perceived that “Trauma” is the most known type of admission among all hospitals participating in this dataset. “Emergent” mode of admission was also a noticeably frequent type of admission. Most of the admitted cases by “illness severity” are moderate cases. However, extreme (severe) admitted cases attributed to about 17% of all admitted cases in the dataset. This is a clear indication that COVID-19 stressed hospitals’ resources and caused significant stress on the staff workflow. Also, this analysis confirmed the importance of studying the epidemiological

curves during the non-pharmaceutical interventions per the case study (1) of this chapter. Age groups of 41-50 and 31-40 were equally most admitted cases. Since there were no patients' profiles and demographic attributes recorded, these figures are not justified and may require further investigation with more clinical information and a clear picture of the patients' profiles. Finally, Figure 57 revealed that most hospitals' LOS is in the range of 21-30 days and 11-20. At the same time, prolonged stays (31-40, 51-60, and 71-80 days) showed substantial reported numbers. These numbers indicated extended longer stay more subjected to extensive treatment course and, therefore, are likely to demand hospital resources. It is noted that the proper values of variables (Ward Type, Ward Facility, Hospital Region, Hospital City) are not available explicitly once. Figure 59 compares important common categorical features in the dataset against the Stay_Days variables. Figure 58 represents a projection of prolonged LOS in the dataset, whereas the labels 0-10 are per the description of the Stays_Days column in Table 25.

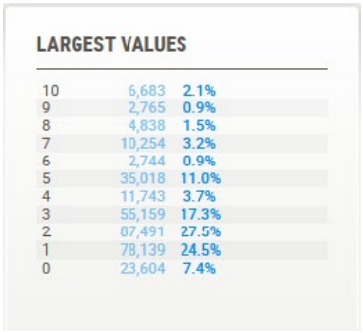


Figure 58. Prolonged LOS percentages in dataset per label (the larger the label, the more extended LOS)

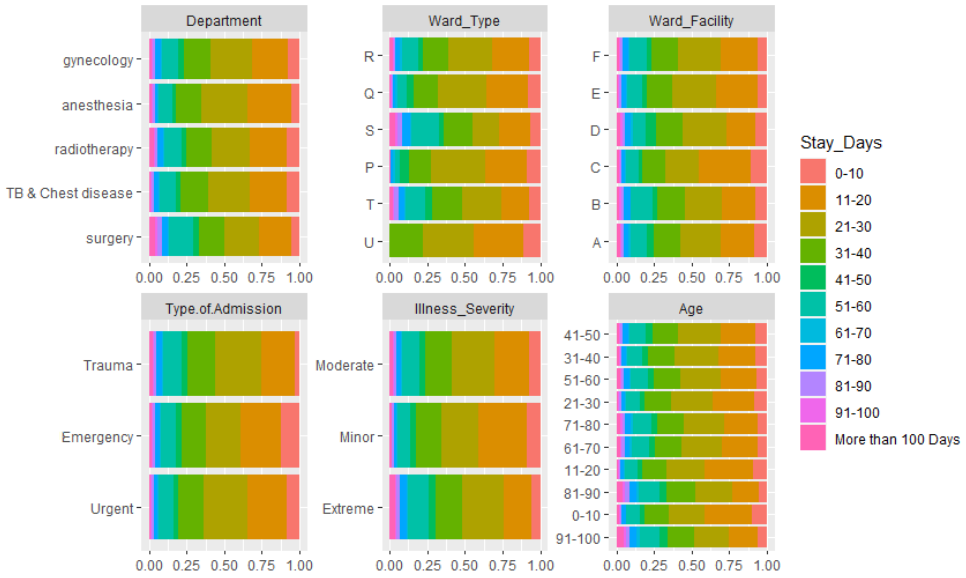


Figure 59. Categorical variables against Stay_Days (LOS)

More EDA analysis in relation to the dataset is summarised in the appendix of this chapter.

6.4.2.3 Data Pre-processing

The dataset contained missing values from Bed_Grade (n=113), and CityCodePatient (n=4532). Therefore, the missing value technique imputes most frequent from sklearn library (SimpleImputer) [289] was used to treat missing value.

6.4.2.4 Models Description, Predictive Results, and Discussion

The model in the case study was chosen according to the desired performance of ensemble learning predictive models in the thesis. Therefore, the choice benchmarks three models (Random Forest, Gradient Boosting, and eXtreme Gradient Boosting), as described in Chapters 4-5, and evaluates their performance against each other. Therefore, the methodology in this case study is adopted from the previous research framework utilised in the previous LOS chapters. The utilisation of the methods (Random Forest, Gradient Boosting, and eXtreme Gradient Boosting) follows the algorithmic structure as their utilisations in Chapters 2, 4-5.

The benchmarking strategy is achieved in two approaches. The first approach is benchmarking with hyperparameters for each model (per chapter 5), and without the class balancing approach, and the second approach follows the class balancing SMOTE (chapter 4) due to its desired outcomes. The confusion matrix was utilised to compare the prediction outcomes of the winning model in approach 1 vs approach 2.

6.4.2.4.1 LOS Benchmarking with Hyperparameters Results (Prediction without Class Balancing)

Table 25. Comparison between RF, GB, and XGB approach (Prediction without class balancing)

Stays _Days	Precision			Recall			F1-Score		
	RF	GB	XGB	RF	GB	XGB	RF	GB	XGB
Label 0: 0-10 days	0.46	0.46	0.35	0.07	0.01	0.12	0.13	0.02	0.17
Label 1: 11-20 days	0.4	0.38	0.39	0.54	0.46	0.55	0.46	0.42	0.46
Label 2: 21-30 days	0.42	0.4	0.42	0.62	0.68	0.58	0.5	0.5	0.49
Label 3: 31-40 days	0.41	0.36	0.38	0.24	0.31	0.23	0.3	0.33	0.28
Label 4: 41-50 days	0	0	0	0	0	0	0	0	0
Label 5: 51-60 days	0.39	0.42	0.39	0.49	0.25	0.41	0.43	0.32	0.4
Label 6: 61-70 days	0	0	0	0	0	0	0	0	0
Label 7: 71-80 days	0	0.39	0.11	0	0.01	0.01	0	0.02	0.01
Label 8: 81-90 days	0	0	0.26	0	0	0.12	0	0	0.17
Label 9: 91-100 days	0	0.2	0	0	0	0	0	0	0
Label 10: (100 +) days	0.57	0.61	0.56	0.34	0.3	0.36	0.42	0.4	0.44
AVG Accuracy	RF: 0.41, GB: 0.39, XGB: 0.4								
Weighted Prediction	0.37	0.37	0.36	0.41	0.39	0.4	0.37	0.34	0.36

The benchmarking results from approach 1 reported weak performance for all models to predict LOS for desired outcomes for all labels (Table 27). Also, it was noted that the models suffered from unreliable results due to the imbalance class issue (as described in chapter 4). However, chapter 4 provided a doable methodological approach to deal with it for LOS hospitalisation prediction. Thus, the same approach is adopted in approach 2 of this case study. RF is chosen to be further evaluated with class balancing.

6.4.2.4.2 LOS Benchmarking with Class Balancing Prediction Results

Table 26. Class balancing prediction report for random forest

Stays_ Days	Precision	Recall	Specificity	F1-Score	G.mean	IBA
0-10	0.61	0.67	0.96	0.64	0.80	0.63
11-20	0.41	0.42	0.94	0.42	0.63	0.38
21-30	0.41	0.57	0.92	0.49	0.72	0.51
31-40	0.43	0.21	0.97	0.28	0.45	0.19
41-50	0.44	0.71	0.97	0.69	0.83	0.67
51-60	0.68	0.49	0.96	0.52	0.68	0.44
61-70	0.55	0.93	0.99	0.93	0.96	0.92
71-80	0.92	0.73	0.97	0.73	0.84	0.70
81-90	0.73	0.93	0.98	0.87	0.95	0.90
91-100	0.82	0.94	0.99	0.94	0.96	0.93
100 days+	0.94	0.90	0.99	0.89	0.93	0.86
Weighted Prediction	0.68	0.68	0.97	0.67	0.80	0.65

Table 26 reports the class balancing results for RF. It is noted that the results are improved significantly for all LOS predicted labels. Class labels 61-70, 81-90, 0.93 days showed the most predicted desired performance (IBA, G.mean), where RF is able to predict inpatient LOS cases with similar characteristics successfully. On the other hand, class labels 31-40, 11-20 days revealed in the unreliable prediction, which is justified from a data prediction perspective due to the limited number of inpatients within those labels. The confusion matrix in Figure 60 shows the lack of ability of RF to predict LOS in multi-label prediction effectively. In contrast, the confusion matrix (Figure 61) for the same model (RF) shows an improved multi-label prediction of LOS outcomes with the class-balancing method (SMOTE).

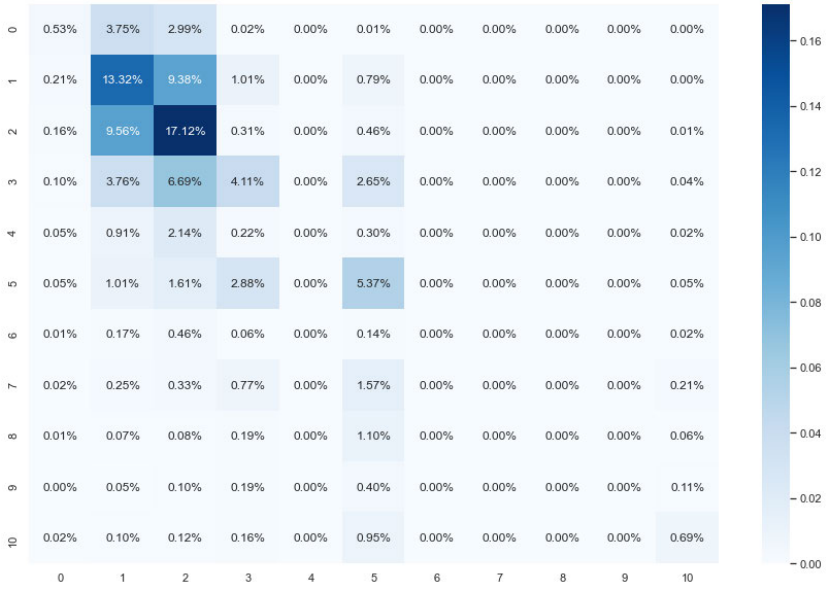


Figure 60. Confusion matrix for RF: prediction multi-label without class balancing

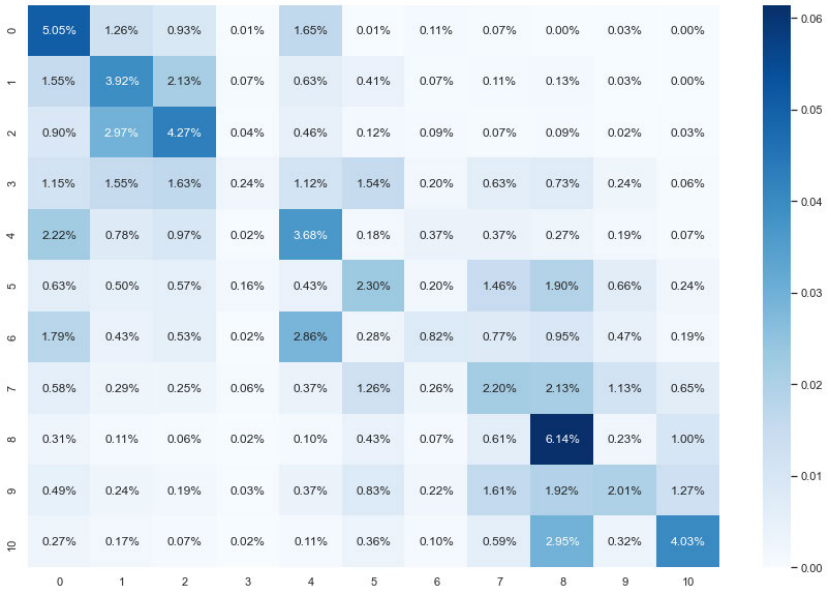


Figure 61. Confusion matrix for RF: prediction multi-label with class balancing

Class balancing methods provide a powerful evaluation to improve prediction results with the case of imbalanced class labels as they appear like the prediction problem in this research. The RF with SMOTE showed improved LOS predictive outcomes, which can help hospital decision-makers decide the suitability of the prediction outcomes based on IBA and G.mean evaluation metrics for class balancing outcomes. Furthermore, the confusion matrix helps visualise the instances where FN vs FP and TP vs TN of the predicted LOS observation. It was noticed that shorter LOS (0, 1, 2) and higher LOS (8, 9) labels mean higher LOS predicted LOS outcomes. In contrast, other predicted labels (3, 6, 9) reported fewer means of predicted IBA.

The choice of the dataset in this case study is in accordance with the COVID-19 open access hospital dataset availability. However, since the disease emerged in December 2019 and became a global pandemic in 2020, it was difficult to obtain COVID-19 real hospitalisation despite our attempt with Al-Ain hospital to validate the proposed real hospital LOS predictive framework as studied in chapter 5. The decision was made to find a suitable open-access dataset with the minimum useful admission attributes that can run a predictive LOS model from this thesis and discover insights from the data. For this purpose, after multiple searches for an open-access COVID-19 dataset, the dataset evaluated in this case study was deemed suitable for this research even though there is missing information on the patients' profiles, such as clinical, laboratory, and medication information. Regardless of these limitations, it was worthy of providing a research attempt for researchers in the field of health analytics to utilise the ability of ML predictive models. Further, the attempt aimed to advance the research in LOS prediction during the pandemic and understand further the nature of the infectious disease and its impact on ICU and hospital resources' utilisation as reported in the results of this case study.

6.4.2.4.3 Research Implication of COVID-19 LOS Case Study

One of the direct research implications of the proposed predictive COVID-LOS architecture is the doability of the predictive architecture to differentiate the duration of stay for predicted cases based on LOS various stay categorised as provided from the COVID dataset used in this study. Hence, the prediction nature of the COVID-19 linkage dataset is multi-class prediction; the predicted results showed efficient predicted outcomes, especially with the issue of class imbalance problems. The class balancing approach with the ensemble learners can help healthcare workers determine factors that can influence the influx of patients during outbreaks and plan hospital beds and wards facilities more efficiently and proactively. On the other hand, the performance evaluation metrics (confusion matrix) can guide beds managers with the assistance of clinical information systems developers to provide healthcare guidelines to healthcare workers, including clinical doctors and nurses, to prioritise cases based on LOS and illness severity. Finally, a vital research implication of COVID-LOS predictive architecture is the fact that the predicted outcomes provide insights for healthcare insurance companies and governments' healthcare bulking mechanisms to make finical decisions based on the LOS categories and how likely it will demand hospitals' resources and, therefore, the finical cover decision factor.

6.5 Chapter Conclusion and Future Direction

The first case study (curfew in Jordan) demonstrated that NPIs during the first wave of COVID-19 in Jordan seemed to effectively control the COVID-19 epidemic, reduce the reproduction rate, and protect hospitals from being overwhelmed. In addition, early strict intervention measures showed evidence of containing and suppressing the disease. The case study showed that simulating the spread of infectious disease in the early stage of the outbreak revealed great benefits to protect the healthcare system from failure, protecting vulnerable people from being hospitalised, and reducing the risk of mortality during the outbreak. Studying and simulating the epidemiological model certainly can help hospital managers and healthcare assessment systems look at the probability of a high number of patients admitted to the hospital and eventually stress hospital resources in scenarios in that healthcare authorities do not apply NPIs. Bed and hospital managers support public health policies and NPIs measures and discuss how epidemiologic simulation curves can guide them to manage hospital resources, beds, and staffing and predict ICU inpatients' length of stay efficiently.

The second case study introduced an extended version of the SEIR model by forking the infectious compartment into two categories: asymptomatic-mild or symptomatic-severe. Despite the lack of real data, the case study illustrated how the effective reproductive number (and its change over time) could be computed using available parameters. This computation has allowed us to forecast and predict the outlook of COVID-19 in the UAE as the sample country of the investigation by using the two variations of the SEIR model. The modelling techniques were applied to VIC, which has a clearer and more documented list of interventions. The results show that the modified SEIR model is more sensitive and can determine when the diffusion will flatten. The case study reported certain limitations of the proposed method. Most concerning is the lack of real empirical data. The modified SEIR simulation can guide healthcare managers to prepare bed occupancy for the worst-case scenarios. Further, it helps to understand the possibility of severe hospitalisations and how their distribution can demand hospital resources' utilisations.

The third case study evaluated the impact of the COVID-19 pandemic and the stress caused by the disease on hospital resource' utilisation. The case study is the first in the literature to predict COVID-19 LOS at a scalable level using a de-identified dataset with multiple hospitals combined within the same dataset linkage data. Based on the first two case studies' motivation to examine the external factor that could impact hospital resource and beds' occupancy, the

third case study successively provided a doable pilot study to predict COVID-19 LOS based on different LOS categories. In addition, the case study provided a detailed EDA analysis that could potentially assist hospitals' healthcare systems in evaluating and managing hospital resources more efficiently based on the reported data insights. It is noted that the LOS predictive approaches in Chapters 4-5 showed successful implementation to predict new emerged diseases such as COVID-19, and the fact that those modes (ensemble learners: RF, XGB, GB) are viable choices in clinical information systems in predictive LOS tasks. In particular, the Random Forest is the desired model for LOS predictive tasks in clinical information systems with binary and multi-class prediction problems; therefore, the RF should be considered in the practical implementation of machine-learning for hospital resources' utilisation.

Appendix

EDA Analysis

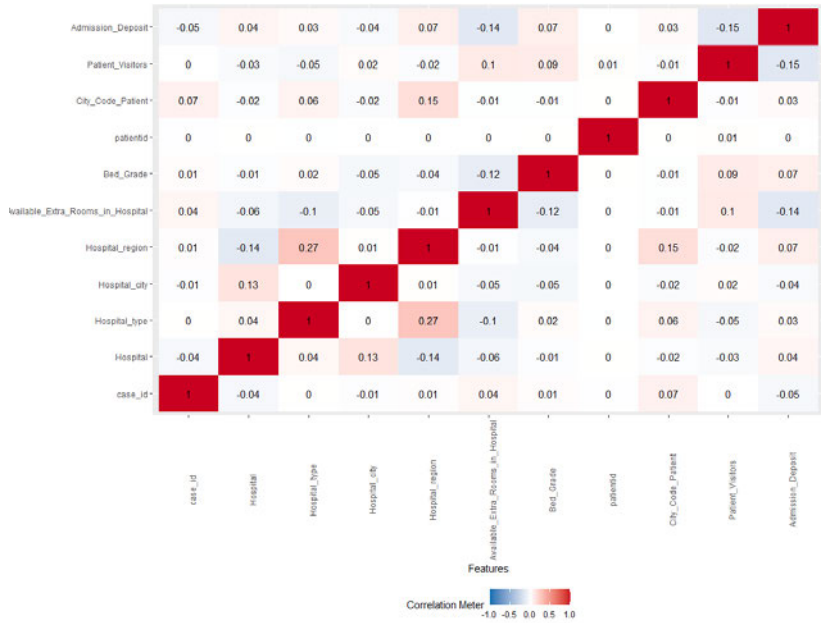


Figure 62. Correlation meter for continuous valuables

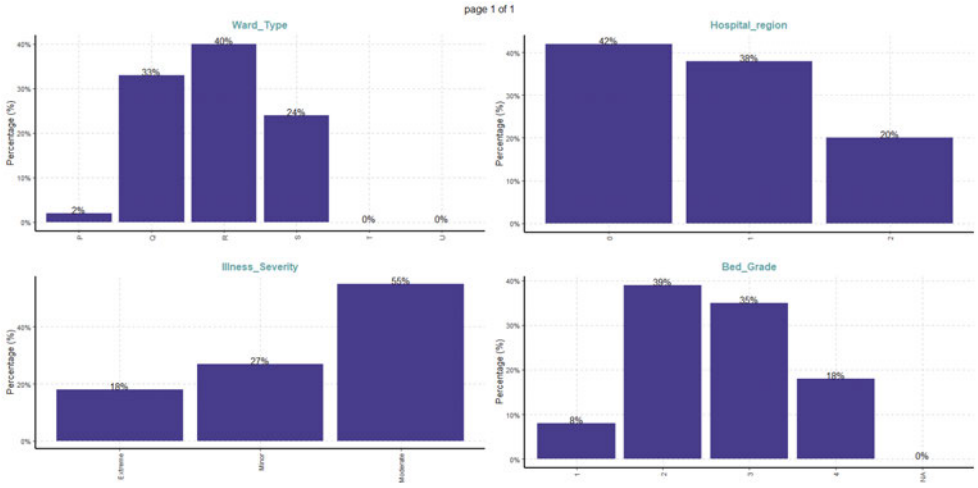


Figure 63. Distributions of categorical variables in the dataset

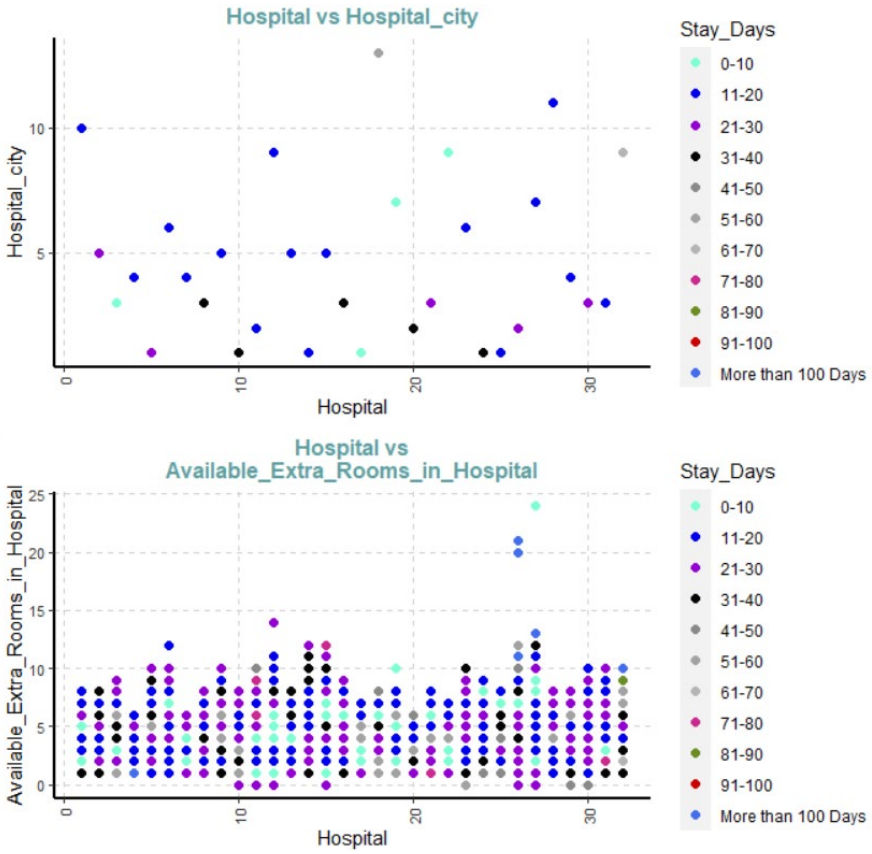


Figure 64. Scatter plot -bivariate- for numerical variables (hospital vs hospital city, hospital vs available extra rooms in hospital) against Stay_Days

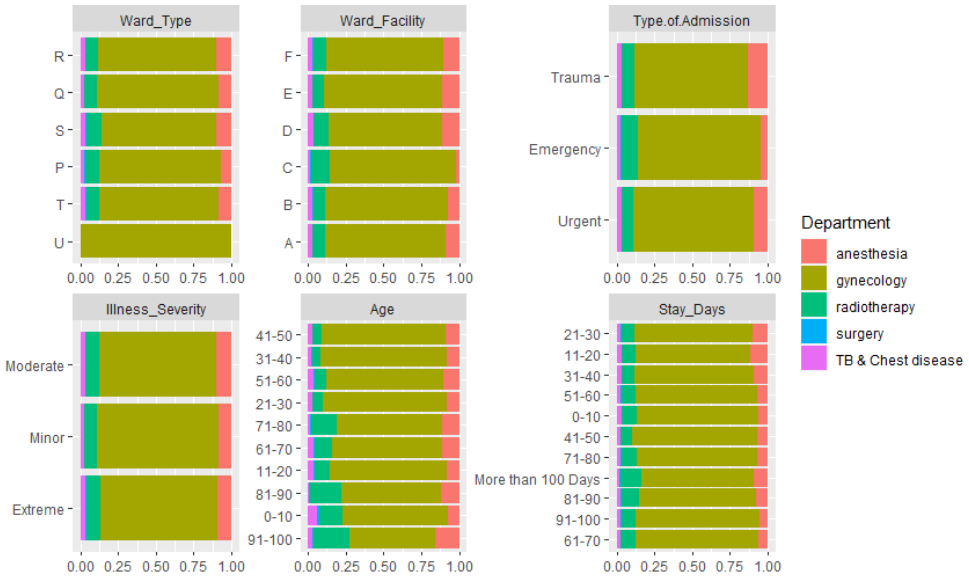


Figure 65. Categorical variables against hospital department variable

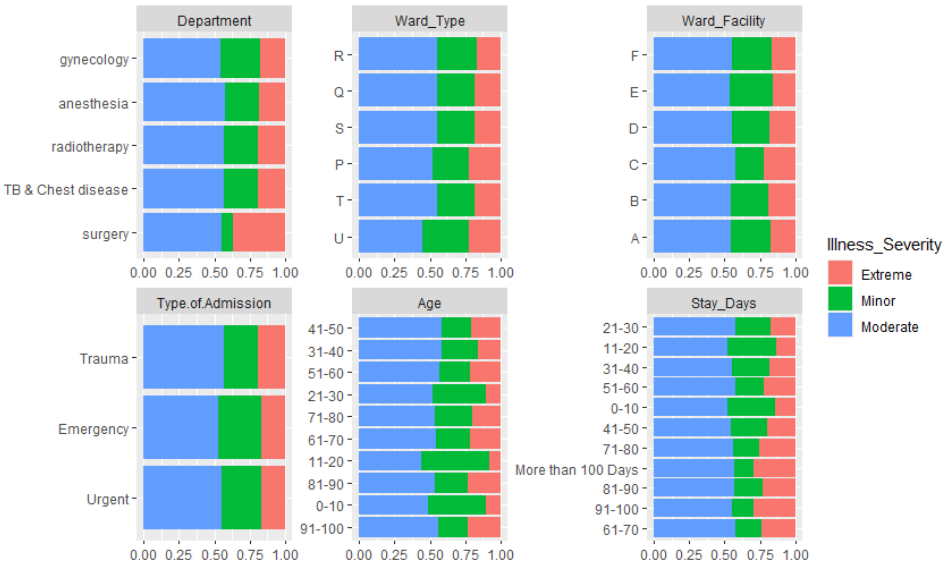


Figure 66. Categorical variables against IIilness severity variable

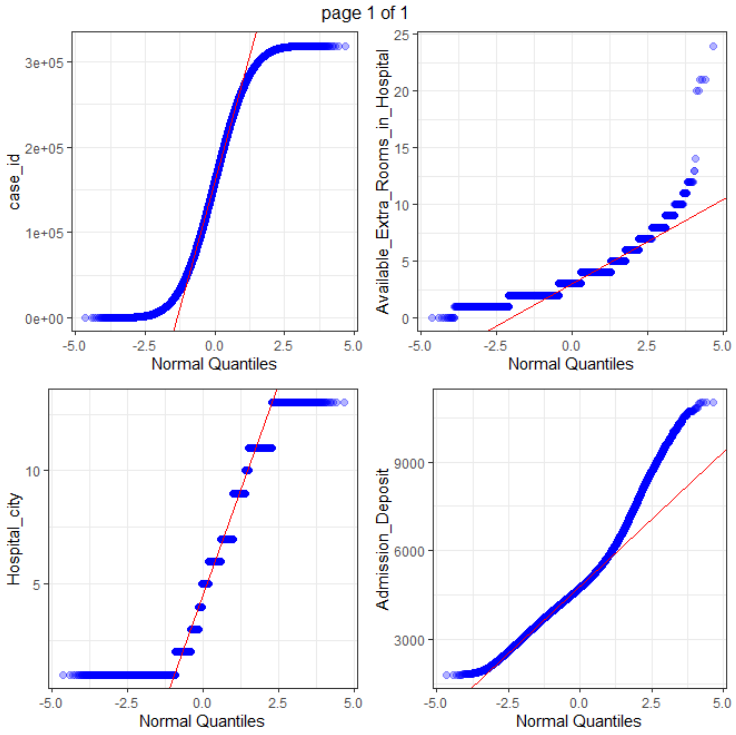


Figure 67. Quantile-quantile plot for numeric variables

7. Chapter Seven: Conclusion and Future Work

7.1 Conclusion

The primary aim of this thesis was to investigate the feasibility and robustness of predictive machine-learning models in the context of improving hospital resources' utilisation with data-driven approaches and predicting hospitalisation with hospital quality assessment metrics such as length of stay. The length of stay predictions includes the validity of the proposed methodological predictive framework on each hospital's electronic health records data source. In this thesis, we relied on EHRs to drive a data-driven predictive LOS research framework that suits the most demanding hospital facilities for hospital resources' utilisation context. The thesis focused on the viability of the methodological predictive length of stay approaches on dynamic and demanding healthcare facilities and hospital settings such as the intensive care units and the emergency departments. While the hospital length of stay predictions are (internal) healthcare inpatients outcomes assessment at the time of admission to discharge, the thesis also considered (external) factors outside hospital control, such as forecasting future hospitalisations from the spread of infectious communicable disease during pandemics. The internal and external splits are the thesis' main contributions. Therefore, the thesis evaluated the public health measures during events of uncertainty (e.g. pandemics) and measured the effect of non-pharmaceutical intervention during outbreaks on future hospitalised cases. This approach is the first contribution in the literature to examine the epidemiological curves' effect using simulation models to project the future hospitalisations on their strong potential to impact hospital beds' availability and stress hospital workflow and workers, to the best of our knowledge.

The main research commonalities between chapters are the usefulness of ensembles learning models in the context of LOS for hospital resources utilisation. The ensembles learning models anticipate better predictive performance by combining several base models to produce an optimal predictive model. These predictive models explored the internal LOS for various chronic and acute conditions using data-driven approaches to determine the most accurate and powerful predicted outcomes. This eventually helps to achieve desired outcomes for hospital professionals who are working in hospital settings.

7.2 Research Implications

The thesis offers manifold practical research implications and research guidelines for hospital beds managers, clinicians, nurses, hospital management, and researchers in hospitals' electronic health records and public health settings. Each chapter discussed the research implication of the proposed methodological predictive approaches.

The thesis implications are split across recommended data-driven methods and implications for clinicians, where both are important in real setting applications. This is ultimately the framework (or a checklist) to focus on for the next project. Therefore, a clinician in the ideal world will have a checklist in a web-based app that is running the predictive and simulated methods in the background. The xAI techniques will allow the clinician to choose the options appropriately. Therefore, here we emphasise and summarise the overall research implications as follows:

LOS Predictive Models in Emergency Department Settings: chapter two provides research methods to predict inpatients' length of stay at the time of emergency hospitalisation. The chapter utilised an open-source emergency room dataset to fulfil the thesis and chapters' expected aims. The results summarised the state of ED-LOS hospitalisation, considering the ED factors in emergency department hospitalisations are important to predict the LOS of ED inpatients. Furthermore, the study reported that ED-LOS predicators (triage by a physician, weekend day of ED admission, X-ray, transfer, and age group) are deemed a vital decision boundary in the prediction output of inpatients for ED-LOS discharge. The direct and implied research benefits from chapter 2 are:

- Assessment of the effectiveness of predictive models provides profound predictive information for ED beds managers and ED doctors and nurses to better anticipate inpatients' decision guidance for different timing and under different ED situations and inpatients' circumstances.
- The ensembles learners are robust where their performance is vigorous under different ED admitted circumstances. Hence, the reliability of prediction models is a must-task to achieve in dynamic and fast decision-making environments such as ED. Therefore, the ensembles predictive models provide feasible opportunities for ED researchers to exploit them in other ED prediction tasks. The explainable AI approach confirmed the robustness of ensembles learners in the context of predicting ED-LOS.

LOS Predictive Models in Intensive Care Units: chapter 3 compared regression and classification predictive methods to derive predictive LOS findings for which approach is more healthcare decision-based in ICU settings. The chapter utilised two different disease-based case studies to fulfil this objective. The findings guide to:

- Binary prediction problems make a clearer decision-making sense than regression predictive models. For example, the classification approach in the sepsis case successfully differentiated two (short vs long) LOS labels. This helps hospital managers and clinicians to assess which clinical variables are more likely to lead to which LOS prediction (LOS category). Thus, binary prediction tasks are suitable for clinical information systems where the healthcare decisions are carried out by non-machine learning experts.
- Regression prediction models, especially ensemble models and the stacking ensembles, provided a powerful ability to investigate the correlation between the LOS and the associated clinical predictors from inpatients' hospitalisation.

The findings in chapter 4 support the adoption of predictive machine-learning methods in imbalanced class LOS distribution.

- The chapter provided a viable approach to treating imbalance class methods regardless of the feature selection methods. Lung cancer is a case study of verifying the robustness of the doable class balancing LOS predictive framework from a data-driven approach. The class balancing method SMOTE with Random Forest (ensemble) model enables the clinician to apply procedure and actions, including evaluating the severity of inpatients' admission at the time of hospitalisation as a clinical information systems assistive tool.
- The findings support clinicians to detect early clinical problems with patients at the time of admission.
- The findings assist junior doctors in ICU settings to manage inpatient hospitalisation based on the desired clinical performance of the RF-SMOTE approach.
- The explainable AI predictive approach helps hospital and beds managers allocate resources efficiently, reduce medical resource waste, and potentially improve patient outcomes.

Validation of the LOS Predictive Model from Real Hospital Settings: chapter 5 studied the predictive LOS models in real hospital settings. Therefore, the importance of the reported findings is as follows:

- Validation of the LOS predictive model was done using real hospital data with the proposed methodological framework.
- The proposed LOS predictive framework can be a doable solution since it considers the safety aspect of machine-learning and its ability to explain the inner working and, therefore, make the LOS predictive decision explainable to beds managers, nurses and clinicians.
- The chapter provided an approach to explain the prediction results of machine-learning ensemble models for all hospitalisations and at the level of individual patients (patient-centred).
- The explainable predictive LOS approach helps to understand the hospital variables that are more significant to the health-decision making at the patient level or beds' occupancy level and resources' management level.

Assessment of Public Health Measures (External Factors) during Pandemic and Measurement of their Impact on Predicting Hospital LOS chapter 6 studied the effect of public health measures during pandemics, besides evaluating the LOS from linkage hospital data sources. The main findings are:

- Measured and simulated the effect of non-pharmaceutical intervention during COVID-19 pandemics and studied its potential impact on hospital resource utilisation by forecasting future hospitalised cases.
- The projection of future hospitalised cases is guidance and indication for hospital management systems and hospital managers to be proactively prepared in different circumstances.
- The chapter supported the findings of future hospitalised cases by predicting COVID-19 from real and de-identified COVID-19 hospitalisations using a dataset (hospitals linkage data source). The comprehensive EDA data-driven approach and LOS benchmarking of the ensembles predictive models showed the importance of simulating the future hospitalised cases using the derived SEIR models.

- The COVID-19 LOS multi-label prediction task is feasible for hospital resources' utilisation based on the results from the pilot study via the large size dataset used to perform the predictions.

Overall, the utilised predictive LOS approaches are guidelines for the development of the proposed LOS predictive frameworks of each chapter in the thesis. Furthermore, the findings assist hospital management in applying them in the context of the clinical information systems for improving resources' utilisation in a doable predictive LOS stay system that includes internal and external factors in the perception of the system implementation.

7.3 Future work

According to the findings and conducted results in the thesis, this section reports research directions:

Scalable prediction. Prediction on multiple data sources with similar attributes is research motivating, which is slightly touched on in the last chapter with the COVID-1 LOS dataset. However, it is worth investigating the scalability of the LOS prediction from multiple data sources (hospitals) that have similar data characteristics, including Admission, Laboratory Clinical, and Medication Information. This confirms the patient's profile predictions micro-levelled to the patient's level (patient-centred). Moreover, this will allow the discovery of more data patterns from large and scalable data and, therefore, support the healthcare assessment systems and health decision-makers for LOS predictive tasks.

Hospital Big Data Analytics is currently emerging in the healthcare sector worldwide. However, one of the major obstacles facing advancing this research is the lack of health standardisation and dataset sharing and collaboration between healthcare institutional data owners globally. Besides hospitals, governmental healthcare sectors, global healthcare organisations, and international healthcare policymakers can play a vital role in international efforts by establishing global health analytical data hubs worldwide with global standards. Furthermore, hospitals and medical, educational institutes can be central in guiding the research and finding big data health centres in various world regions connected with similar goals. Thus, this requires a collaboration between world governments and world health organisations to build health infrastructure that can accommodate the need of the collaborative parties, from data housing and distributed data processing to many more technical and logistic requirements. In a future study, the ultimate goal is to apply big data analytics and utilise suitable techniques

on big EHRs data. A web application will be built to implement the algorithmic predictive models based on the chapter 5 xAI approach.

This thesis provides a research opportunity for other predictive and decision-making tasks in healthcare assessment systems and clinical information systems that are also vital for patient care and hospital clinical decision making, such as the prediction of the risk of mortality at the time of hospital admission, prediction of hospital readmission, prediction of hospital capacity and hospital crowding, and the prediction of diseases severity for hospitalised medical cases.

Furthermore, the thesis highlighted the importance of employing explainable machine learning approaches to explain prediction for non-machine learning people and reveal the black box of the machine learning models so that they are more understandable, explainable and provide clear insights from the data for decision making in medical and healthcare studies and other domains. The findings on xAI support the decision making for healthcare workers and provide more understanding of the patient's status to third parties such as healthcare insurances and government healthcare assessment systems. The thesis encourages researchers in the field of health analytics to examine LOS with other studies that rely on medical imaging to diagnose, support clinical decision and hospital resources utilisation with advancements in medical imaging and image analysis where the various aspects of this domain could potentially enhance the adoption of machine learning and deep neural networks in clinical and healthcare studies via the explainable Artificial intelligence approaches.

References

[1] H. Baek, M. Cho, S. Kim, H. Hwang, M. Song, and S. Yoo, "Analysis of length of hospital stay using electronic health records: A statistical and data mining approach," *PloS one,* vol. 13, no. 4, p. e0195901, 2018.

[2] S. K. Dixit and M. Sambasivan, "A review of the Australian healthcare system: A policy perspective," *SAGE open medicine,* vol. 6, p. 2050312118769211, 2018.

[3] P. Santibáñez, V. S. Chow, J. French, M. L. Puterman, and S. Tyldesley, "Reducing patient wait times and improving resource utilization at British Columbia Cancer Agency's ambulatory care unit through simulation," *Health care management science,* vol. 12, no. 4, p. 392, 2009.

[4] E. M. Carter and H. W. Potts, "Predicting length of stay from an electronic patient record system: a primary total knee replacement example," *BMC medical informatics and decision making,* vol. 14, no. 1, p. 26, 2014.

[5] A. I. o. H. a. W. (AIHW), "Australia's hospitals at a glance 2016–17," 27 Jun 2018 2018. Accessed: 15/09/2018. [Online]. Available: https://www.aihw.gov.au/reports/hospitals/ahs-2016-17-at-a-glance/contents/table-of-contents

[6] L. Siciliani, P. M. Sivey, and A. Street, "Differences in Length of Stay Between Public Hospitals, Treatment Centres and Private Providers: Selection Or Efficiency?," 2011.

[7] S. Siddiqui, S. Ahmed, and R. Manasia, "Apache II score as a predictor of length of stay and outcome in our ICUs," *JPMA. The Journal of the Pakistan Medical Association,* vol. 55, no. 6, pp. 253-254, 2005.

[8] W. A. Knaus *et al.*, "The APACHE III prognostic system: risk prediction of hospital mortality for critically III hospitalized adults," *Chest,* vol. 100, no. 6, pp. 1619-1636, 1991.

[9] F. Rahmani, H. E. Bakhtavar, S. S. Vahdati, M. Hosseini, and R. M. Esfanjani, "Evaluation of MGAP and GAP Trauma Scores to Predict Prognosis of Multiple-trauma Patients," *Trauma Monthly,* vol. 22, no. 3, 2017.

[10] M. T. Keegan, O. Gajic, and B. Afessa, "Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and influence of resuscitation status on model performance," *Chest,* vol. 142, no. 4, pp. 851-858, 2012.

[11] C. Li *et al.*, "Prediction of length of stay on the intensive care unit based on least absolute shrinkage and selection operator," *IEEE Access,* vol. 7, pp. 110710-110721, 2019.

[12] F. Tang, C. Xiao, F. Wang, and J. Zhou, "Predictive modeling in urgent care: a comparative study of machine learning approaches," *JAMIA Open,* 2018.

[13] HealthIT. "What is an electronic health record (EHR)?" https://www.healthit.gov/faq/what-electronic-health-record-ehr (accessed 30/07/2018, 2018).

[14] A. J. Steele, S. C. Denaxas, A. D. Shah, H. Hemingway, and N. M. Luscombe, "Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease," *PloS one,* vol. 13, no. 8, p. e0202344, 2018.

[15] J. Zhao, "Learning Predictive Models from Electronic Health Records," Department of Computer and Systems Sciences, Stockholm University, 2017.

[16] S. A. Al-Sayouri, E. Gujral, D. Koutra, E. E. Papalexakis, and S. S. Lam, "t-PINE: Tensor-based Predictable and Interpretable Node Embeddings," *arXiv preprint arXiv:1805.01889,* 2018.

[17] S. Gunasekar *et al.*, "Phenotyping using Structured Collective Matrix Factorization of Multi--source EHR Data," *arXiv preprint arXiv:1609.04466,* 2016.

[18] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE journal of biomedical and health informatics,* vol. 22, no. 5, pp. 1589-1604, 2018.

[19] N. Beeknoo and R. P. Jones, "The demography myth: how demographic forecasting underestimates hospital admissions, and creates the illusion that fewer hospital beds and community-based bed equivalents, will be required in the future," *Journal of Advances in Medicine and Medical Research,* pp. 1-27, 2017.

[20] H. Salah, "Predicting inpatient length of stay in Western New York health service area using machine learning algorithms," State University of New York at Binghamton, 2017.

[21] E. Condes and J. R. Arribas, "Impact of COVID-19 on Madrid hospital system," *Enfermedades Infecciosas y Microbiología Clínica,* 2020.

[22] J.-D. Van Wees *et al.*, "Forecasting hospitalization and ICU rates of the COVID-19 outbreak: An efficient SEIR model," *Bull World Health Organ,* 2020.

[23] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news,* vol. 2, no. 3, pp. 18-22, 2002.

[24] scikit-learn. "Gradient Boosting Regressor." https://scikit-learn.org (accessed 15/01/2020, 2020).

[25] L. Breiman, "Stacked regressions," *Machine Learning,* journal article vol. 24, no. 1, pp. 49-64, July 01 1996, doi: 10.1007/bf00117832.

[26] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks,* vol. 61, pp. 85-117, 2015.

[27] R. E. Wright, "Logistic regression," 1995.

[28] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.

[29] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in neurorobotics,* vol. 7, p. 21, 2013.

[30] L. Bottou, "Stochastic gradient descent tricks," in *Neural networks: Tricks of the trade*: Springer, 2012, pp. 421-436.

[31] J. Laaksonen and E. Oja, "Classification with learning k-nearest neighbors," in *Proceedings of International Conference on Neural Networks (ICNN'96)*, 1996, vol. 3: IEEE, pp. 1480-1483.

[32] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics,* vol. 21, no. 3, pp. 660-674, 1991.

[33] W. Nick, J. Shelton, K. Asamene, and A. C. Esterline, "A Study of Supervised Machine Learning Techniques for Structural Health Monitoring," *MAICS,* vol. 1353, p. 36, 2015.

[34] W. S. Noble, "What is a support vector machine?," *Nature biotechnology,* vol. 24, no. 12, pp. 1565-1567, 2006.

[35] G. K. Jha, "Artificial neural networks and its applications," *IARI, New Delhi, girish_iasri@ rediffmail. com,* 2007.

[36] M. Y. Li and J. S. Muldowney, "Global stability for the SEIR model in epidemiology," *Mathematical biosciences,* vol. 125, no. 2, pp. 155-164, 1995.

[37] C. Morley, M. Unwin, G. M. Peterson, J. Stankovich, and L. Kinsman, "Emergency department crowding: A systematic review of causes, consequences and solutions," *PloS one,* vol. 13, no. 8, p. e0203316, 2018.

[38] R. Forero, S. McCarthy, and K. Hillman, "Access block and emergency department overcrowding," *Annual Update in Intensive Care and Emergency Medicine 2011,* pp. 720-728, 2011.

[39] A.-A. I. o. H. a. Welfare, "Emergency department care 2017–18: Australian hospital statistics," 06 Dec 2018 2018, vol. HSE 216. Accessed: 15/12/2018. [Online]. Available: https://www.aihw.gov.au/reports/hospitals/emergency-department-care-2017-18/report-editions

[40] Statista, "Rates of emergency department (ED) visits resulting in admission in the U.S. in 2006 and 2014, by first-listed diagnosis (per 1,000 population)," 2021. [Online]. Available: https://www.statista.com/statistics/749689/ed-visit-rates-resulting-in-admission-in-us-by-diagnosis-type/

[41] A. Boyle, K. Beniuk, I. Higginson, and P. Atkinson, "Emergency department crowding: time for interventions and policy evaluations," *Emergency medicine international,* vol. 2012, 2012.

[42] B. E. Driesen, B. H. Van Riet, L. Verkerk, H. J. Bonjer, H. Merten, and P. W. Nanayakkara, "Long length of stay at the emergency department is mostly caused by organisational factors outside the influence of the emergency department: A root cause analysis," *PloS one,* vol. 13, no. 9, p. e0202751, 2018.

[43] Harutyunyan, H., Khachatrian, H., Kale, D.C. et al. Multitask learning and benchmarking with clinical time series data. Sci Data 6, 96 (2019).

[44] V. Liu, P. Kipnis, M. K. Gould, and G. J. Escobar, "Length of stay predictions: improvements through the use of automated laboratory and comorbidity variables," *Medical care,* pp. 739-744, 2010.

[45] S. M. Hosseininejad *et al.*, "Determinants of Prolonged Length of Stay in the Emergency Department; a Cross-sectional Study," *Emergency,* vol. 5, no. 1, 2017.

[46] P. R. Hachesu, M. Ahmadi, S. Alizadeh, and F. Sadoughi, "Use of data mining techniques to determine and predict length of stay of cardiac patients," *Healthcare informatics research,* vol. 19, no. 2, pp. 121-129, 2013.

[47] H. Maharlou, S. R. Niakan Kalhori, S. Shahbazi, and R. Ravangard, "Predicting Length of Stay in Intensive Care Units after Cardiac Surgery: Comparison of Artificial Neural Networks and Adaptive Neuro-fuzzy System," *Healthcare informatics research,* vol. 24, no. 2, pp. 109-117, 2018.

[48] M. Elbattah and O. Molloy, "Using Machine Learning to Predict Length of Stay and Discharge Destination for Hip-Fracture Patients," in *Proceedings of SAI Intelligent Systems Conference*, 2016: Springer, pp. 207-217.

[49] S. Barnes, E. Hamrock, M. Toerper, S. Siddiqui, and S. Levin, "Real-time prediction of inpatient length of stay for discharge prioritization," *Journal of the American Medical Informatics Association,* vol. 23, no. e1, pp. e2-e10, 2015.

[50] S. Levin *et al.*, "Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index," *Annals of emergency medicine,* vol. 71, no. 5, pp. 565-574. e2, 2018.

[51] X. Zhang, J. Kim, R. E. Patzer, S. R. Pitts, A. Patzer, and J. D. Schrager, "Prediction of emergency department hospital admission based on natural language processing and neural networks," *Methods of information in medicine,* vol. 56, no. 05, pp. 377-389, 2017.

[52] M. Gül and A. F. Güneri, "Forecasting patient length of stay in an emergency department by artificial neural networks," *Journal of Aeronautics and Space Technologies,* vol. 8, no. 2, pp. 43-48, 2015.

[53] R. A. Taylor *et al.*, "Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data–Driven, Machine Learning Approach," *Academic emergency medicine,* vol. 23, no. 3, pp. 269-278, 2016.

[54] A. McCoy and R. Das, "Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units," *BMJ Open Qual,* vol. 6, no. 2, p. e000158, 2017.

[55] Q. Mao *et al.*, "Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU," *BMJ open,* vol. 8, no. 1, p. e017833, 2018.

[56] J. Sundén-Cullberg, R. Rylance, J. Svefors, A. Norrby-Teglund, J. Björk, and M. Inghammar, "Fever in the emergency department predicts survival of patients with severe sepsis and septic shock admitted to the ICU," *Critical care medicine,* vol. 45, no. 4, pp. 591-599, 2017.

[57] M. J. Vermeulen *et al.*, "Are reductions in emergency department length of stay associated with improvements in quality of care? A difference-in-differences analysis," *BMJ Qual Saf,* vol. 25, no. 7, pp. 489-498, 2016.

[58] A. Azari, V. P. Janeja, and S. Levin, "Imbalanced learning to predict long stay Emergency Department patients," in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*, 2015: IEEE, pp. 807-814.

[59] W. S. Hong, A. D. Haimovich, and R. A. Taylor, "Predicting hospital admission at emergency department triage using machine learning," *PloS one,* vol. 13, no. 7, p. e0201016, 2018.

[60] C. Combes, F. Kadri, and S. Chaabane, "Predicting Hospital length of stay using regression models: application to Emergency Department," in *10ème Conférence Francophone de Modélisation, Optimisation et Simulation-MOSIM'14*, 2014.

[61] Y. Barak-Corren, A. M. Fine, and B. Y. Reis, "Early Prediction Model of Patient Hospitalization From the Pediatric Emergency Department," *Pediatrics,* vol. 139, no. 5, p. e20162785, 2017.

[62] C.-H. Chaou *et al.*, "Predicting length of stay among patients discharged from the emergency department—using an accelerated failure time model," *PloS one,* vol. 12, no. 1, p. e0165756, 2017.

[63] B. Graham, R. Bond, M. Quinn, and M. Mulvenna, "Using Data Mining to Predict Hospital Admissions From the Emergency Department," *IEEE Access,* vol. 6, pp. 10458-10469, 2018.

[64] D. Gligorijevic *et al.*, "Deep Attention Model for Triage of Emergency Department Patients," in *Proceedings of the 2018 SIAM International Conference on Data Mining*, 2018: SIAM, pp. 297-305.

[65] D. Golmohammadi, "Predicting hospital admissions to reduce emergency department boarding," *International Journal of Production Economics,* vol. 182, pp. 535-544, 2016.

[66] K. D. Mandl, D. Markwell, R. MacDonald, P. Szolovits, and I. S. Kohane, "Public standards and patients' control: how to keep electronic medical records accessible but privateMedical information: access and privacyDoctrines for developing electronic medical recordsDesirable characteristics of electronic medical recordsChallenges and limitations for electronic medical recordsConclusionsCommentary: Open approaches to electronic patient recordsCommentary: A patient's viewpoint," *Bmj,* vol. 322, no. 7281, pp. 283-287, 2001.

[67] SHAP. "TreeExplainer." https://shap-lrjball.readthedocs.io/en/docs_update/generated/shap.TreeExplainer.html (accessed 15/1/2021, 2021).

[68] J. Andersson, L. Nordgren, I. Cheng, U. Nilsson, and L. Kurland, "Long emergency department length of stay: A concept analysis," *International Emergency Nursing,* vol. 53, p. 100930, 2020.

[69] A. G. D. o. Health. "Recommended triage method." http://www.health.gov.au/internet/publications/publishing.nsf/Content/triageqrg~triageqrg-method (accessed 25th, Novemeber, 2018).

[70] K.-H. H. Chip-Jin Ng, Jen-Tze Kuan,Te-Fa Chiu, Wei-Kong Chen, "Comparison Between Canadian Triage and," *Journal of the Formosan Medical Association,* vol. 109, pp. 828-837, 11 2010.

[71] N. Health. "Emergency Department Care " NSW Health. https://www.health.nsw.gov.au/Performance/Pages/emergency.aspx (accessed 2021)

[72] D. B. Richardson, "Reducing patient time in the emergency department," *MJA,* vol. 179, pp. 516-517, 2003.

[73] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. Ver Steeg, and A. Galstyan, "Multitask learning and benchmarking with clinical time series data," *Scientific data,* vol. 6, no. 1, p. 96, 2019.

[74] W. H. Organization, "WHO Sepsis Technical Expert Meeting - Meeting report," Geneva, 2018. [Online]. Available: https://www.who.int/servicedeliverysafety/areas/sepsis_meeting-report-2018.pdf

[75] (2018). *HSE 201, Admitted patient care 2016–17: Australian hospital statistics*. [Online] Available: https://www.aihw.gov.au/reports/hospitals/ahs-2016-17-admitted-patient-care/contents/summary

[76] A. S. Network, "Stopping Sepsis: A National Action Plan," in "A health policy report December 2017," The George Institue, Sydney, Australia, 2017. Accessed: 15/07/2019. [Online]. Available: https://www.georgeinstitute.org.au/sites/default/files/documents/stopping-sepsis-national-action-plan.pdf

[77] C. J. Paoli, M. A. Reynolds, M. Sinha, M. Gitlin, and E. Crouser, "Epidemiology and costs of sepsis in the United States—An analysis based on timing of diagnosis and severity Level," *Critical care medicine,* vol. 46, no. 12, p. 1889, 2018.

[78] M. J. Hall, S. N. Williams, C. J. DeFrances, and A. Golosinskiy, "Inpatient care for septicemia or sepsis: a challenge for patients and hospitals," 2011.

[79] N. O. Ghanem-Zoubi, M. Vardi, A. Laor, G. Weber, and H. Bitterman, "Assessment of disease-severity scoring systems for patients with sepsis in general internal medicine departments," *Critical Care,* vol. 15, no. 2, p. R95, 2011.

[80] B. Alsinglawi and O. Mubin, "Predictive Analytics and Deep Learning Techniques in Electronic Medical Records: Recent Advancements and Future Direction," in *Workshops of the International Conference on Advanced Information Networking and Applications*, 2019: Springer, pp. 907-914.

[81] B. Farzanegan and M. Zangi, "Predictor factors for sepsis diagnosis, length of ICU stay and mortality in ICU," *Journal of Cellular & Molecular Anesthesia,* vol. 2, no. 2, pp. 55-62, 2017.

[82] P. A. Hampshire, A. Guha, A. Strong, D. Parsons, and P. Rowan, "An evaluation of the Charlson co-morbidity score for predicting sepsis after elective major surgery," *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine,* vol. 15, no. 1, p. 30, 2011.

[83] D. W. Shimabukuro, C. W. Barton, M. D. Feldman, S. J. Mataraso, and R. Das, "Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial," *BMJ open respiratory research,* vol. 4, no. 1, p. e000234, 2017.

[84] Burdick H, Pino E, Gabel-Comeau D, McCoy A, Gu C, Roberts J, Le S, Slote J, Pellegrini E, Green-Saxena A, Hoffman J, Das R. Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. BMJ Health Care Inform. 2020 Apr;27(1):e100109. doi: 10.1136/bmjhci-2019-100109. PMID: 32354696; PMCID: PMC7245419.

[85] A. E. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific data,* vol. 3, p. 160035, 2016.

[86] S. Wang, M. B. McDermott, G. Chauhan, M. Ghassemi, M. C. Hughes, and T. Naumann, "Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii," in *Proceedings of the ACM Conference on Health, Inference, and Learning*, 2020, pp. 222-235.

[87] T. Zebin, S. Rezvy, and T. J. Chaussalet, "A deep learning approach for length of stay prediction in clinical settings from medical records," in *2019 IEEE Conference on*

Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2019: IEEE, pp. 1-5.

[88] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian journal of internal medicine,* vol. 4, no. 2, p. 627, 2013.

[89] J.-S. Li, Y. Tian, Y.-F. Liu, T. Shu, and M.-H. Liang, "Applying a BP neural network model to predict the length of hospital stay," in *International Conference on Health Information Science*, 2013: Springer, pp. 18-29.

[90] P. Ponikowski *et al.*, "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC," *European journal of heart failure,* vol. 18, no. 8, pp. 891-975, 2016.

[91] G. Savarese and L. H. Lund, "Global public health burden of heart failure," *Cardiac failure review,* vol. 3, no. 1, p. 7, 2017.

[92] W. Lesyuk, C. Kriza, and P. Kolominsky-Rabas, "Cost-of-illness studies in heart failure: A systematic review 2004–2016," *BMC cardiovascular disorders,* vol. 18, no. 1, p. 74, 2018.

[93] Y.-K. Chan *et al.*, "Current and projected burden of heart failure in the Australian adult population: a substantive but still ill-defined major health issue," *BMC health services research,* vol. 16, no. 1, p. 501, 2016.

[94] S. L. Jackson, X. Tong, R. J. King, F. Loustalot, Y. Hong, and M. D. Ritchey, "National burden of heart failure events in the United States, 2006 to 2014," *Circulation: Heart Failure,* vol. 11, no. 12, p. e004873, 2018.

[95] A. Almashrafi, H. Alsabti, M. Mukaddirov, B. Balan, and P. Aylin, "Factors associated with prolonged length of stay following cardiac surgery in a major referral hospital in Oman: a retrospective observational study," *BMJ open,* vol. 6, no. 6, p. e010764, 2016.

[96] B. Alsinglawi, F. Alnajjar, O. Mubin, M. Novoa, O. Karajeh, and O. Darwish, "Benchmarking Predictive Models in Electronic Health Records: Sepsis Length of Stay Prediction," in *International Conference on Advanced Information Networking and Applications*, 2020: Springer, pp. 258-267.

[97] H. Omar and M. Guglin, "Longer-than-average length of stay in acute heart failure," *Herz,* vol. 43, no. 2, pp. 131-139, 2018.

[98] M. S. Durstenfeld, M. D. Saybolt, A. Praestgaard, and S. E. Kimmel, "Physician predictions of length of stay of patients admitted with heart failure," *Journal of hospital medicine,* vol. 11, no. 9, pp. 642-645, 2016.

[99] P.-F. J. Tsai *et al.*, "Length of hospital stay prediction at the admission stage for cardiology patients using artificial neural network," *Journal of healthcare engineering,* vol. 2016, 2016.

[100] L. Turgeman, J. H. May, and R. Sciulli, "Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission," *Expert Systems with Applications,* vol. 78, pp. 376-385, 2017.

[101] N. binti Omar, E. Supriyanto, and R. H. Al-Ashwal, "Personalized Clinical Pathway for Heart Failure Management," in *2018 International Conference on Applied Engineering (ICAE)*, 2018: IEEE, pp. 1-5.

[102] J. A. Sterne *et al.*, "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls," *Bmj,* vol. 338, p. b2393, 2009.

[103] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research,* vol. 13, no. Feb, pp. 281-305, 2012.

[104] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of machine learning research,* vol. 12, no. Oct, pp. 2825-2830, 2011.

[105] F. Chollet, "Keras," ed, 2015.

[106] J. M. Valderas, B. Starfield, B. Sibbald, C. Salisbury, and M. Roland, "Defining comorbidity: implications for understanding health and health services," *The Annals of Family Medicine,* vol. 7, no. 4, pp. 357-363, 2009.

[107] S. Kenya *et al.*, "Using community health workers to improve clinical outcomes among people living with HIV: a randomized controlled trial," *AIDS and Behavior,* vol. 17, no. 9, pp. 2927-2934, 2013.

[108] D. R. Anderson, "The impact of resource management on hospital efficiency and quality of care," University of Maryland, College Park, 2013.

[109] J. D. De Jong, G. P. Westert, R. Lagoe, and P. P. Groenewegen, "Variation in hospital length of stay: do physicians adapt their length of stay decisions to what is usual in the hospital where they work?," *Health Services Research,* vol. 41, no. 2, pp. 374-394, 2006.

[110] F. Pecoraro, F. Clemente, and D. Luzi, "The efficiency in the ordinary hospital bed management in Italy: An in-depth analysis of intensive care unit in the areas affected by COVID-19 before the outbreak," *Plos one,* vol. 15, no. 9, p. e0239249, 2020.

[111] M. Hassan, H. P. Tuckman, R. H. Patrick, D. S. Kountz, and J. L. Kohn, "Hospital length of stay and probability of acquiring infection," *International Journal of pharmaceutical and healthcare marketing,* 2010.

[112] M. C. Blom, K. Erwander, L. Gustafsson, M. Landin-Olsson, F. Jonsson, and K. Ivarsson, "The probability of readmission within 30 days of hospital discharge is positively associated with inpatient bed occupancy at discharge–a retrospective cohort study," *BMC emergency medicine,* vol. 15, no. 1, pp. 1-6, 2015.

[113] E. Rocheteau, P. Liò, and S. Hyland, "Temporal Pointwise Convolutional Networks for Length of Stay Prediction in the Intensive Care Unit," *arXiv preprint arXiv:2007.09483,* 2020.

[114] C. W. Hanson *et al.*, "Effects of an organized critical care service on outcomes and resource utilization: a cohort study," *Critical care medicine,* vol. 27, no. 2, pp. 270-274, 1999.

[115] W. A. Knaus, J. E. Zimmerman, D. P. Wagner, E. A. Draper, and D. E. Lawrence, "APACHE-acute physiology and chronic health evaluation: a physiologically based classification system," *Critical care medicine,* vol. 9, no. 8, pp. 591-597, 1981.

[116] C.-C. Yeh *et al.*, "Quick-SOFA score≥ 2 predicts prolonged hospital stay in geriatric patients with influenza infection," *The American journal of emergency medicine,* 2019.

[117] G. S. Birkhead, M. Klompas, and N. R. Shah, "Uses of electronic health records for public health surveillance to advance public health," *Annual review of public health,* vol. 36, pp. 345-359, 2015.

[118] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE journal of biomedical and health informatics,* vol. 22, no. 5, pp. 1589-1604, 2017.

[119] E. S. Berner and T. J. La Lande, "Overview of clinical decision support systems," in *Clinical decision support systems*: Springer, 2007, pp. 3-22.

[120] Y.-Y. Jo *et al.*, "Prediction of Prolonged Length of Hospital Stay After Cancer Surgery Using Machine Learning on Electronic Health Records: Retrospective Cross-sectional Study," *JMIR medical informatics,* vol. 9, no. 2, p. e23147, 2021.

[121] W. E. Muhlestein, D. S. Akagi, J. M. Davies, and L. B. Chambless, "Predicting Inpatient Length of Stay After Brain Tumor Surgery: Developing Machine Learning Ensembles to Improve Predictive Performance," *Neurosurgery,* 2018.

[122] J.-E. Bibault, P. Giraud, and A. Burgun, "Big data and machine learning in radiation oncology: state of the art and future prospects," *Cancer letters,* vol. 382, no. 1, pp. 110-117, 2016.

[123] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *The New England journal of medicine,* vol. 375, no. 13, p. 1216, 2016.

[124] A. G. Singal *et al.*, "Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma," *The American journal of gastroenterology,* vol. 108, no. 11, p. 1723, 2013.

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

[126] X. Ma, Y. Si, Z. Wang, and Y. Wang, "Length of stay prediction for ICU patients using individualized single classification algorithm," *Computer methods and programs in biomedicine,* vol. 186, p. 105224, 2020.

[127] J. Fang, J. Zhu, and X. Zhang, "Prediction of Length of Stay on the Intensive Care Unit Based on Bayesian Neural Network," in *Journal of Physics: Conference Series*, 2020, vol. 1631, no. 1: IOP Publishing, p. 012089.

[128] C. Dominici *et al.*, "A nomogram for predicting long length of stay in the intensive care unit in patients undergoing CABG: Results from the multicenter E-CABG registry," *Journal of Cardiothoracic and Vascular Anesthesia,* vol. 34, no. 11, pp. 2951-2961, 2020.

[129] L. Y. Sun, A. Bader Eddeen, M. Ruel, E. MacPhee, and T. G. Mesana, "Derivation and Validation of a Clinical Model to Predict Intensive Care Unit Length of Stay After Cardiac Surgery," *Journal of the American Heart Association,* vol. 9, no. 21, p. e017847, 2020.

[130] I. W. Verburg, N. F. de Keizer, E. de Jonge, and N. Peek, "Comparison of regression methods for modeling intensive care length of stay," *PloS one,* vol. 9, no. 10, p. e109684, 2014.

[131] B. Alsinglawi *et al.*, "Predicting Length of Stay for Cardiovascular Hospitalizations in the Intensive Care Unit: Machine Learning Approach," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020: IEEE, pp. 5442-5445.

[132] K. Meadows, R. Gibbens, C. Gerrard, and A. Vuylsteke, "Prediction of patient length of stay on the intensive care unit following cardiac surgery: a logistic regression analysis based on the cardiac operative mortality risk calculator, EuroSCORE," *Journal of cardiothoracic and vascular anesthesia,* 2018.

[133] T. D. Best *et al.*, "Multilevel Body Composition Analysis on Chest Computed Tomography Predicts Hospital Length of Stay and Complications After Lobectomy for Lung Cancer: A Multicenter Study," *Annals of Surgery,* 2020.

[134] C. Pompili *et al.*, "Poor preoperative quality of life predicts prolonged hospital stay after VATS lobectomy for lung cancer," *European Journal of Cardio-Thoracic Surgery,* vol. 59, no. 1, pp. 116-121, 2021.

[135] J. Dong, Y. Mao, J. Li, and J. He, "Stair-climbing test predicts postoperative cardiopulmonary complications and hospital stay in patients with non-small cell lung cancer," *Medical science monitor: international medical journal of experimental and clinical research,* vol. 23, p. 1436, 2017.

[136] S. Li, K. Zhou, M. Wang, R. Lin, J. Fan, and G. Che, "Degree of pulmonary fissure completeness can predict postoperative cardiopulmonary complications and length of hospital stay in patients undergoing video-assisted thoracoscopic lobectomy for early-stage lung cancer," *Interactive cardiovascular and thoracic surgery,* vol. 26, no. 1, pp. 25-33, 2018.

[137] M. Soares *et al.*, "Characteristics and outcomes of patients with cancer requiring admission to intensive care units: a prospective multicenter study," *Critical care medicine,* vol. 38, no. 1, pp. 9-15, 2010.

[138] E. Azoulay *et al.*, "Predictors of short-term mortality in critically ill patients with solid malignancies," *Intensive care medicine,* vol. 26, no. 12, pp. 1817-1823, 2000.

[139] C. Andréjak *et al.*, "Admission of advanced lung cancer patients to intensive care unit: a retrospective study of 76 patients," *BMC cancer,* vol. 11, no. 1, pp. 1-9, 2011.

[140] D. Vicendese *et al.*, "Hospital characteristics, rather than surgical volume, predict length of stay following colorectal cancer surgery," *Australian and New Zealand journal of public health,* vol. 44, no. 1, pp. 73-82, 2020.

[141] M. Soares *et al.*, "Intensive care in patients with lung cancer: a multinational study," *Annals of oncology,* vol. 25, no. 9, pp. 1829-1835, 2014.

[142] X.-C. Zhang, Z.-D. Zhang, and D.-S. Huang, "Prediction of length of ICU stay using data-mining techniques: an example of old critically Ill postoperative gastric cancer patients," *Asian Pacific Journal of Cancer Prevention,* vol. 13, no. 1, pp. 97-101, 2012.

[143] K. T. Baghaei and S. Rahimi, "Sepsis Prediction: An Attention-Based Interpretable Approach," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2019: IEEE, pp. 1-6.

[144] S. García, J. Luengo, and F. Herrera, *Data preprocessing in data mining*. Springer, 2015.

[145] W. McKinney, *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012.

[146] K. Ho and P. Scott, "Zeta: a global method for discretization of cotitinuous variables," in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, 1997, pp. 191-194.

[147] E. J. Clarke and B. A. Barton, "Entropy and MDL discretization of continuous variables for Bayesian belief networks," *International Journal of Intelligent Systems,* vol. 15, no. 1, pp. 61-92, 2000.

[148] S. Jung, Y. Bi, and R. V. Davuluri, "Evaluation of data discretization methods to derive platform independent isoform expression signatures for multi-class tumor subtyping," *BMC genomics,* vol. 16, no. 11, pp. 1-10, 2015.

[149] A. Gupta, T. Liu, and S. Shepherd, "Clinical decision support system to assess the risk of sepsis using tree augmented Bayesian networks and electronic medical record data," *Health informatics journal,* vol. 26, no. 2, pp. 841-861, 2020.

[150] A. Tabaie, F. H. Chokshi, A. L. Holder, and S. N. Nemati, "Doubly-robust estimation of effect of imaging resource utilization on discharge decisions in emergency departments," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018: IEEE, pp. 3256-3259.

[151] Y.-S. Chen, C.-H. Cheng, C.-J. Lai, C.-Y. Hsu, and H.-J. Syu, "Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment," *Computers in Biology and Medicine,* vol. 42, no. 2, pp. 213-221, 2012.

[152] J. P. Allard *et al.*, "Decline in nutritional status is associated with prolonged length of stay in hospitalized patients admitted for 7 days or more: A prospective cohort study," *Clinical nutrition,* vol. 35, no. 1, pp. 144-152, 2016.

[153] Z. Rustam and N. Maghfirah, "Correlated based SVM-RFE as feature selection for cancer classification using microarray databases," in *AIP Conference Proceedings*, 2018, vol. 2023, no. 1: AIP Publishing LLC, p. 020235.

[154] X. Min *et al.*, "Multi-parametric MRI-based radiomics signature for discriminating between clinically significant and insignificant prostate cancer: Cross-validation of a machine learning method," *European journal of radiology,* vol. 115, pp. 16-21, 2019.

[155] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometrics and intelligent laboratory systems,* vol. 83, no. 2, pp. 83-90, 2006.

[156] A. Bustamam, A. Bachtiar, and D. Sarwinda, "Selecting features subsets based on support vector machine-recursive features elimination and One Dimensional-Naïve Bayes classifier using support vector machines for classification of prostate and breast cancer," *Procedia Computer Science,* vol. 157, pp. 450-458, 2019.

[157] S. Karthik, R. S. Perumal, and P. C. Mouli, "Breast cancer classification using deep neural networks," in *Knowledge computing and its applications*: Springer, 2018, pp. 227-241.

[158] S. F. Abdoh, M. A. Rizka, and F. A. Maghraby, "Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques," *IEEE Access,* vol. 6, pp. 59475-59485, 2018.

[159] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, 2008: IEEE, pp. 1322-1328.

[160] D. L. Wilson, "Asymptotic properties of nearest neighbor rules using edited data," *IEEE Transactions on Systems, Man, and Cybernetics,* no. 3, pp. 408-421, 1972.

[161] I. Tomek, "Two modifications of CNN," 1976.

[162] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter,* vol. 6, no. 1, pp. 20-29, 2004.

[163] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araujo, and J. Santos, "Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]," *ieee ComputatioNal iNtelligeNCe magaziNe,* vol. 13, no. 4, pp. 59-76, 2018.

[164] V. Nagarajan, "A critical analysis of Sampling Techniques for imbalanced data classification: An application to Social Media," Dublin, National College of Ireland, 2017.

[165] L. Lusa, "Improved shrunken centroid classifiers for high-dimensional class-imbalanced data," *BMC bioinformatics,* vol. 14, no. 1, pp. 1-13, 2013.

[166] T. Le, M. T. Vo, B. Vo, M. Y. Lee, and S. W. Baik, "A hybrid approach using oversampling technique and cost-sensitive learning for bankruptcy prediction," *Complexity,* vol. 2019, 2019.

[167] A. Amin *et al.*, "Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study," *IEEE Access,* vol. 4, pp. 7940-7957, 2016.

[168] M. Pal, "Random forest classifier for remote sensing classification," *International journal of remote sensing,* vol. 26, no. 1, pp. 217-222, 2005.

[169] S. Li and X. Zhang, "Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm," *Neural Computing and Applications,* pp. 1-9, 2019.

[170] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of biomedical informatics,* vol. 35, no. 5-6, pp. 352-359, 2002.

[171] W. E. Muhlestein, D. S. Akagi, J. M. Davies, and L. B. Chambless, "Predicting inpatient length of stay after brain tumor surgery: Developing machine learning ensembles to improve predictive performance," *Neurosurgery,* vol. 85, no. 3, pp. 384-393, 2019.

[172] V. García, R. A. Mollineda, and J. S. Sánchez, "Theoretical analysis of a performance measure for imbalanced data," in *2010 20th International Conference on Pattern Recognition*, 2010: IEEE, pp. 617-620.

[173] V. García, J. S. Sánchez, and R. A. Mollineda, "On the effectiveness of preprocessing methods when dealing with different levels of class imbalance," *Knowledge-Based Systems,* vol. 25, no. 1, pp. 13-21, 2012.

[174] B. Kim, O. Koyejo, and R. Khanna, "Examples are not enough, learn to criticize! Criticism for Interpretability," in *NIPS*, 2016, pp. 2280-2288.

[175] M. A. Ahmad, C. Eckert, and A. Teredesai, "Interpretable machine learning in healthcare," in *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, 2018, pp. 559-560.

[176] H. Zhang and M. Wang, "Search for the smallest random forest," *Statistics and its Interface,* vol. 2, no. 3, p. 381, 2009.

[177] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874,* 2017.

[178] A. Awad, M. Bader–El–Den, and J. McNicholas, "Patient length of stay and mortality prediction: A survey," *Health services management research,* vol. 30, no. 2, pp. 105-120, 2017.

[179] O. f. E. C.-o. a. D. (OECD), "OECD (2021), Length of hospital stay (indicator)," 2020. Accessed: Accessed on 21 July 2021. [Online]. Available: https://data.oecd.org/healthcare/length-of-hospital-stay.htm

[180] A. I. o. Health and Welfare, "Hospital Performance: Length of stay in public hospitals in 2011–12," AIHW, Canberra, 2013. [Online]. Available: https://www.aihw.gov.au/reports/hospitals/hospital-performance-length-of-stay-in-2011-12

[181] M. M. Islam, T. N. Poly, and Y.-C. J. Li, "Recent advancement of clinical information systems: Opportunities and challenges," *Yearbook of medical informatics,* vol. 27, no. 01, pp. 083-090, 2018.

[182] E. Levesque, E. Hoti, D. Azoulay, P. Ichai, D. Samuel, and F. Saliba, "The implementation of an Intensive Care Information System allows shortening the ICU length of stay," *Journal of clinical monitoring and computing,* vol. 29, no. 2, pp. 263-269, 2015.

[183] S. Calloway, H. A. Akilo, and K. Bierman, "Impact of a clinical decision support system on pharmacy clinical interventions, documentation efforts, and costs," *Hospital pharmacy,* vol. 48, no. 9, pp. 744-752, 2013.

[184] S. T. McMullin, T. P. Lonergan, C. S. Rynearson, T. D. Doerr, P. A. Veregge, and E. S. Scanlan, "Impact of an evidence-based computerized decision support system on primary care prescription costs," *The Annals of Family Medicine,* vol. 2, no. 5, pp. 494-498, 2004.

[185] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," *NPJ digital medicine,* vol. 3, no. 1, pp. 1-10, 2020.

[186] D. Romanow, A. Rai, M. Keil, and S. Luxenberg, "Does extended CPOE use reduce patient length of stay?," *International journal of medical informatics,* vol. 97, pp. 128-138, 2017.

[187] L. Su *et al.*, "Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis 3.0 by Machine Learning Models," *Frontiers in Medicine,* vol. 8, p. 883, 2021.

[188] P. V. Staziaki *et al.*, "Machine learning combining CT findings and clinical parameters improves prediction of length of stay and ICU admission in torso trauma," *European Radiology,* pp. 1-8, 2021.

[189] K. Alghatani, N. Ammar, A. Rezgui, and A. Shaban-Nejad, "Predicting Intensive Care Unit Length of Stay and Mortality Using Patient Vital Signs: Machine Learning Model Development and Validation," *JMIR Medical Informatics,* vol. 9, no. 5, p. e21347, 2021.

[190] T. Gentimis, A. Ala'J, A. Durante, K. Cook, and R. Steele, "Predicting hospital length of stay using neural networks on mimic iii data," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, 2017: IEEE, pp. 1194-1201.

[191] R. J. Steele and B. Thompson, "Data mining for generalizable pre-admission prediction of elective length of stay," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, 2019: IEEE, pp. 0127-0133.

[192] C. Combi and G. Pozzi, "Clinical information systems and artificial intelligence: recent research trends," *Yearbook of medical informatics,* vol. 28, no. 01, pp. 083-094, 2019.

[193] E. S. Berner, *Clinical decision support systems*. Springer, 2007.

[194] H. Quan *et al.*, "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data," *Medical care,* pp. 1130-1139, 2005.

[195] K. Potdar, T. S. Pardawala, and C. D. Pai, "A comparative study of categorical variable encoding techniques for neural network classifiers," *International journal of computer applications,* vol. 175, no. 4, pp. 7-9, 2017.

[196] S. E. Gerard, T. J. Patton, G. E. Christensen, J. E. Bayouth, and J. M. Reinhardt, "FissureNet: a deep learning approach for pulmonary fissure detection in CT images," *IEEE transactions on medical imaging,* vol. 38, no. 1, pp. 156-166, 2018.

[197] J. Brownlee. "How to Report Classifier Performance with Confidence Intervals." https://machinelearningmastery.com/report-classifier-performance-confidence-intervals/ (accessed 2021).

[198] A. Rezaei, R. Fathony, O. Memarrast, and B. Ziebart, "Fairness for robust log loss classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 04, pp. 5511-5518.

[199] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (xai): Toward medical xai," *IEEE Transactions on Neural Networks and Learning Systems,* 2020.

[200] A. M. Antoniadi *et al.*, "Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review," *Applied Sciences,* vol. 11, no. 11, p. 5088, 2021.

[201] "ExplainerDashboard." https://explainerdashboard.readthedocs.io/en/latest/index.html (accessed 2021).

[202] U. o. N. Dame. "Lift Charts." https://www3.nd.edu/~busiforc/handouts/DataMining/Lift%20Charts.html (accessed 2021).

[203] E. Zihni *et al.*, "Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome," *Plos one,* vol. 15, no. 4, p. e0231166, 2020.

[204] K.-H. Yu and I. S. Kohane, "Framing the challenges of artificial intelligence in medicine," *BMJ quality & safety,* vol. 28, no. 3, pp. 238-241, 2019.

[205] E. H. Shortliffe and M. J. Sepúlveda, "Clinical decision support in the era of artificial intelligence," *Jama,* vol. 320, no. 21, pp. 2199-2200, 2018.

[206] CHeReL. "The Centre for Health Record Linkage " https://www.cherel.org.au/ (accessed.

[207] WHO. "WHO Timeline - COVID-19." https://www.who.int/news/item/27-04-2020-who-timeline---covid-19 (accessed 2021).

[208] M. Ackermann *et al.*, "Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19," *New England Journal of Medicine,* 2020.

[209] D. Brahma, S. Chakraborty, and A. Menokee, "The early days of a global pandemic: A timeline of COVID-19 spread and government interventions," *Brookings, USA,* 2020.

[210] S. Lai *et al.*, "Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak: an observational and modelling study," *medRxiv,* 2020.

[211] M. Y. Li, J. R. Graef, L. Wang, and J. Karsai, "Global dynamics of a SEIR model with varying total population size," *Mathematical biosciences,* vol. 160, no. 2, pp. 191-213, 1999.

[212] J. Farooq and M. A. Bazaz, "A Novel Adaptive Deep Learning Model of Covid-19 with focus on mortality reduction strategies," *Chaos, Solitons & Fractals,* p. 110148, 2020.

[213] S. Dil, N. Dil, and Z. H. Maken, "COVID-19 Trends and Forecast in the Eastern Mediterranean Region With a Particular Focus on Pakistan," *Cureus,* vol. 12, no. 6, 2020.

[214] X. Zhou *et al.*, "Forecasting the worldwide spread of COVID-19 based on logistic model and SEIR model," *medRxiv,* 2020.

[215] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling," *arXiv preprint arXiv:2002.06563,* 2020.

[216] Z. Yang *et al.*, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *Journal of Thoracic Disease,* vol. 12, no. 3, p. 165, 2020.

[217] K. Prem *et al.*, "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study," *The Lancet Public Health,* 2020.

[218] C. Massonnaud, J. Roux, and P. Crépey, "COVID-19: Forecasting short term hospital needs in France," *medRxiv,* 2020.

[219] A. Sheikh, A. Sheikh, Z. Sheikh, S. Dhami, and D. Sridhar, "What's the way out? Potential exit strategies from the COVID-19 lockdown," *Journal of global health,* vol. 10, no. 1, 2020.

[220] R. AlQutob, I. A. Moonesar, M. R. Tarawneh, M. Al Nsour, and Y. Khader, "Public Health Strategies for the Gradual Lifting of the Public Sector Lockdown in Jordan and the United Arab Emirates During the COVID-19 Crisis," *JMIR Public Health and Surveillance,* vol. 6, no. 3, p. e20478, 2020.

[221] S. H. Bloukh, A. Shaikh, H. M. Pathan, and Z. Edis, "Prevalence of COVID-19: A Look behind the Scenes from the UAE and India," 2020.

[222] A. Saul, N. Scott, B. S. Crabb, S. S. Majundar, B. Coghlan, and M. E. Hellard, "Victoria's response to a resurgence of COVID-19 has averted 9,000-37,000 cases in July 2020," *The Medical Journal of Australia,* p. 1, 2020.

[223] A. S. Bhagavathula, W. A. Aldhaleei, J. Rahmani, M. A. Mahabadi, and D. K. Bandari, "Novel coronavirus (COVID-19) knowledge and perceptions: a survey on healthcare workers," *MedRxiv,* 2020.

[224] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, "Features, evaluation and treatment coronavirus (COVID-19)," in *Statpearls [internet]*: StatPearls Publishing, 2020.

[225] (2020). *Coronavirus update for Victoria - 25 July 2020*. [Online] Available: https://www.dhhs.vic.gov.au/coronavirus-update-victoria-25-july-2020

[226] W. H. Organization, "Coronavirus disease 2019 (COVID-19)," in "Situation Report – 73," 2 April 2020 2020. Accessed: 1/6/2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_4

[227] T. P. Velavan and C. G. Meyer, "The COVID-19 epidemic," *Trop Med Int Health,* vol. 25, no. 3, pp. 278-280, 2020.

[228] R. H. Chisholm, P. T. Campbell, Y. Wu, S. Y. Tong, J. McVernon, and N. Geard, "Implications of asymptomatic carriers for infectious disease transmission and control," *Royal Society open science,* vol. 5, no. 2, p. 172341, 2018.

[229] R. Verity *et al.*, "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *The Lancet Infectious Diseases,* 2020.

[230] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," *Journal of travel medicine,* 2020.

[231] B. Tang, N. L. Bragazzi, Q. Li, S. Tang, Y. Xiao, and J. Wu, "An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov)," *Infectious disease modelling,* vol. 5, pp. 248-255, 2020.

[232] R. Wölfel *et al.*, "Virological assessment of hospitalized patients with COVID-2019," *Nature,* vol. 581, no. 7809, pp. 465-469, 2020.

[233] L. M. Bettencourt and R. M. Ribeiro, "Real time bayesian estimation of the epidemic potential of emerging infectious diseases," *PLoS One,* vol. 3, no. 5, p. e2185, 2008.

[234] R. Vaidyanathan. "Estimating COVID-19's " https://www.datacamp.com/community/tutorials/replicating-in-r-covid19 (accessed 1/8/2020.

[235] D. Balcan *et al.*, "Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility," *BMC medicine,* vol. 7, no. 1, p. 45, 2009.

[236] P. E. Greenwood and L. F. Gordillo, "Stochastic epidemic modeling," in *Mathematical and statistical estimation approaches in epidemiology*: Springer, 2009, pp. 31-52.

[237] (2020). *Victoria's restriction levels*. [Online] Available: https://www.dhhs.vic.gov.au/victorias-restriction-levels-covid-19

[238] O. W. i. Data. "Research and data to make progress against the world's largest problems." https://github.com/owid/covid-19-data/tree/master/public/data (accessed covid-19-data).

[239] J. H. U. (JHU), "The Center for Systems Science and Engineering (CSSE) at JHU," 2020. [Online]. Available: https://github.com/CSSEGISandData.

[240] WHO. "Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)." WHO. https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) (accessed 2020).

[241] N. Ferguson *et al.*, "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," 2020.

[242] *Coronavirus (COVID-19) current situation and case numbers*. [Online] Available: https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/coronavirus-covid-19-current-situation-and-case-numbers

[243] (2020). *Modelling the current impact of COVID-19 in Australia* [Online] Available: https://www.health.gov.au/resources/publications/modelling-the-current-impact-of-covid-19-in-australia

[244] V. S. G. Health and Human Services. "Premier announces the commencement of stage 3 restrictions." 28/3/2020. https://www.dhhs.vic.gov.au/coronavirus/updates (accessed 1/7/2020, 2020).

[245] A. news. "Victoria in stage 3 coronavirus shutdown restrictions as cases climb to 821." ABC news, Australia. https://www.abc.net.au/news/2020-03-30/victoria-stage-3-coronavirus-restrictions-as-cases-rise/12101632 (accessed 1/7/2020, 2020).

[246] (2020). *Updates about the outbreak of the coronavirus disease (COVID-19): May 2020*. [Online] Available: https://www.dhhs.vic.gov.au/coronavirus/updates/202005

[247] V. S. Government. "Coronavirus (COVID-19) restrictions Victoria." https://www.vic.gov.au/coronavirus-covid-19-restrictions-victoria (accessed 2/7/2020, 2020).

[248] (2008). *Public Health and Wellbeing Act 2008*. [Online] Available: https://www2.health.vic.gov.au/public-health/infectious-diseases/public-health-wellbeing-act-regulations#:~:text=The%20Public%20Health%20and%20Wellbeing%20Act%202008%20and%20Public%20Health,of%20public%20health%20and%20wellbeing.

[249] A. ABC news. "Melbourne placed under stage 4 coronavirus lockdown, stage 3 for rest of Victoria as state of disaster declared." ABC news, Australia. https://www.abc.net.au/news/2020-08-02/victoria-coronavirus-restrictions-imposed-death-toll-cases-rise/12515914 (accessed 2/8/2020, 2020).

[250] V. S. Gove. "Melbourne, Stage 4 restrictions." https://www.dhhs.vic.gov.au/stage-4-restrictions-covid-19 (accessed 20/8/2020, 2020).

[251] M. Al Zobbi, B. Alsinglawi, O. Mubin, and F. Alnajjar, "Measurement Method for Evaluating the Lockdown Policies during the COVID-19 Pandemic," *International Journal of Environmental Research and Public Health,* vol. 17, no. 15, p. 5574, 2020.

[252] K. A. Kheirallah *et al.*, "The Effect of Strict State Measures on the Epidemiologic Curve of COVID-19 Infection in the Context of a Developing Country: A Simulation from Jordan," *International Journal of Environmental Research and Public Health,* vol. 17, no. 18, p. 6530, 2020.

[253] A. C. Wong, X. Li, S. K. Lau, and P. C. Woo, "Global epidemiology of bat coronaviruses," *Viruses,* vol. 11, no. 2, p. 174, 2019.

[254] T. Phan, "Novel coronavirus: From discovery to clinical diagnostics," *Infection, Genetics and Evolution,* vol. 79, p. 104211, 2020.

[255] F. Yu, L. Du, D. M. Ojcius, C. Pan, and S. Jiang, "Measures for diagnosing and treating infections by a novel coronavirus responsible for a pneumonia outbreak originating in Wuhan, China," *Microbes and infection,* vol. 22, no. 2, pp. 74-79, 2020.

[256] WHO, "2019 Novel Coronavirus (2019-nCoV): Strategic preparedness and response plan," World Health Organization, 2020. [Online]. Available: https://www.who.int/publications/i/item/strategic-preparedness-and-response-plan-for-the-new-coronavirus

[257] W. H. Organization, "Coronavirus disease 2019 (COVID-19): situation report, 73," 2020.

[258] M. Battegay, R. Kuehl, S. Tschudin-Sutter, H. H. Hirsch, A. F. Widmer, and R. A. Neher, "2019-novel Coronavirus (2019-nCoV): estimating the case fatality rate–a word of caution," *Swiss medical weekly,* vol. 150, no. 0506, 2020.

[259] A. R. Tuite, I. I. Bogoch, R. Sherbo, A. Watts, D. Fisman, and K. Khan, "Estimation of coronavirus disease 2019 (COVID-19) burden and potential for international dissemination of infection from Iran," *Annals of Internal Medicine,* vol. 172, no. 10, pp. 699-701, 2020.

[260] S. Weston and M. B. Frieman, "COVID-19: knowns, unknowns, and questions," *Msphere,* vol. 5, no. 2, pp. e00203-20, 2020.

[261] N. Imai, I. Dorigatti, A. Cori, S. Riley, and N. Ferguson, "Estimating the potential total number of novel coronavirus cases in Wuhan City, China [cited 2020 Feb 2]," ed.

[262] N. M. Ferguson *et al.*, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team," *Imperial College COVID-19 Response Team,* p. 20, 2020.

[263] X. He *et al.*, "Temporal dynamics in viral shedding and transmissibility of COVID-19," *Nature medicine,* vol. 26, no. 5, pp. 672-675, 2020.

[264] S. Choi and M. Ki, "Estimating the reproductive number and the outbreak size of COVID-19 in Korea," *Epidemiology and health,* vol. 42, 2020.

[265] N. G. Davies *et al.*, "Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study," *The Lancet Public Health,* vol. 5, no. 7, pp. e375-e385, 2020.

[266] J. Panovska-Griffiths, "Can mathematical modelling solve the current Covid-19 crisis?," ed: BioMed Central, 2020.

[267] S. P. Brand *et al.*, "Forecasting the scale of the COVID-19 epidemic in Kenya," *MedRxiv,* 2020.

[268] K. Zia and U. Farooq, "COVID-19 outbreak in Oman: Model-driven impact analysis and challenges," *medRxiv,* 2020.

[269] J. Stehlé *et al.*, "Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees," *BMC medicine,* vol. 9, no. 1, pp. 1-15, 2011.

[270] V. Surveillances, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020," *China CDC weekly,* vol. 2, no. 8, pp. 113-122, 2020.

[271] D. Balcan *et al.*, "Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility," *BMC medicine,* vol. 7, no. 1, pp. 1-12, 2009.

[272] G. Chowell, J. M. Hyman, L. M. Bettencourt, C. Castillo-Chavez, and H. Nishiura, *Mathematical and statistical estimation approaches in epidemiology*. Springer, 2009.

[273] K. Prem, A. R. Cook, and M. Jit, "Projecting social contact matrices in 152 countries using contact surveys and demographic data," *PLoS computational biology,* vol. 13, no. 9, p. e1005697, 2017.

[274] J. Hilton and M. J. Keeling, "Estimation of country-level basic reproductive ratios for novel Coronavirus (SARS-CoV-2/COVID-19) using synthetic contact matrices," *PLoS computational biology,* vol. 16, no. 7, p. e1008031, 2020.

[275] P. Castorina, A. Iorio, and D. Lanteri, "Data analysis on Coronavirus spreading by macroscopic growth laws," *International Journal of Modern Physics C,* vol. 31, no. 07, p. 2050103, 2020.

[276] N. M. Linton *et al.*, "Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data," *Journal of clinical medicine,* vol. 9, no. 2, p. 538, 2020.

[277] M. M. Siwiak, P. Szczesny, and M. P. Siwiak, "From a single host to global spread. The global mobility based modelling of the COVID-19 pandemic implies higher infection and lower detection rates than current estimates," *The Global Mobility Based Modelling of the COVID-19 Pandemic Implies Higher Infection and Lower Detection Rates than Current Estimates (3/23/2020),* 2020.

[278] F. Pan *et al.*, "Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia," *Radiology,* 2020.

[279] G. Patrick, C. Walker, and W. Oliver, "The global impact of COVID-19 and strategies for mitigation and suppression," *WHO Collaborating Centre for Infectious Disease Modelling, MRC Centre for Global Infectious Disease Analysis: Abdul Latif Jameel Institute for Disease and Emergency Analytics, Imperial College London,* 2020.

[280] B. Tang *et al.*, "Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions," *Journal of clinical medicine,* vol. 9, no. 2, p. 462, 2020.

[281] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2," *Emerging infectious diseases,* vol. 26, no. 7, p. 1470, 2020.

[282] M. E. Kretzschmar, G. Rozhnova, and M. van Boven, "Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing," *Frontiers in Physics,* vol. 8, p. 677, 2021.

[283] Möbius. "COVID-19 Hospitals Treatment Plan Help hospitals to identify patients of high LOS risk." https://datasetsearch.research.google.com/search?query=COVID-19%20Hospitals%20Treatment%20Plan&docid=L2cvMTFtX2pzMTg4OQ%3D%3D (accessed.

[284] Q. J. Leclerc *et al.*, "Importance of patient bed pathways and length of stay differences in predicting COVID-19 hospital bed occupancy in England," *BMC health services research,* vol. 21, no. 1, pp. 1-15, 2021.

[285] B. Mahboub, M. T. Al Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 hospital length of stay and risk of death using artificial intelligence-based modeling," *Frontiers in medicine,* vol. 8, 2021.

[286] J. Ebinger *et al.*, "A Machine Learning Algorithm Predicts Duration of hospitalization in COVID-19 patients," *Intelligence-based medicine,* p. 100035, 2021.

[287] D. U. K. R. R. P. K. S. Putatunda. "Summarize and Explore the Data- SmartEDA " https://cran.r-project.org/web/packages/SmartEDA/vignettes/SmartEDA.html (accessed.

[288] B. Cui, "Introduction to DataExplorer," 2020. [Online]. Available: https://cran.r-project.org/web/packages/DataExplorer/vignettes/dataexplorer-intro.html.

[289] sklearn. "SimpleImputer." https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html (accessed 2020)

Thesis Appendix

Table 27. Al-Ain hospital data description (Abstract overview)

Columns	Description
Table: General Admission Information	
MRN	Medical Record Number (Unique for one patient)
FIN	- Unique ID for a patient visit – used to merge data from other sheets/tables such as (lab, clinical events, and Medication) - Visit ID (One MRN can have one or many FINs)
ENCNTR_TYPE	Type of the visit (ex; Inpatient, Emergency, Outpatient…)
ADMIT_NURSE_UNIT_DISP	The first unit the patient was admitted to
DISCHARGE_NURSE_UNIT_DISP	The last unit the patient discharge from
FINANCIAL_CLASS	Insurance class
TRANSFER_NURSE_DISP	Transfer to disposition by nurse recorded name
TRANSFER_DATE_TIME	Transfer to disposition (date and time)
PRIMARY_DIAG_CODE	Primary ICD 10 diagnosis code
PRIMARY_DIAG_DESC	Primary ICD 10 diagnosis description
SECONDARY_DIAG_CODE	Other/Secondary ICD 10 codes
SECONDARY_DIAG_DESC	Other/Secondary ICD 10 description
ENCNTR_ID	Interaction ID between patient and Al-Ain-Hospital
GENDER	Patient Gender: Female, Male
AGE_VISIT	Age in years at the time of visit
FACILITY	Facility name where services were performed.
ADMISSION_DATE_TIME	Patient's admission time in Day/Month/Year, Hour: minute
RACE	Patient race: National, Non-National, GCC National, Unknown
NATIONALITY	The nationality of the patient.
ADMIT_MODE	The mechanism by which a person begins the ICU visit care: Ambulance, Walking, Wheelchair, Stretcher, Carried, Private Vehicle, Police Vehicle.
ADMIT_SOURCE	The source from which the person was transferred/referred directly prior to the beginning of the healthcare visit care.
ADMIT_TYPE	A description of the manner in which the patient was admitted to the healthcare facility: Elective Inpatient, Emergent, Urgent.
LOS_days	The total number of patient days at the Hospital visit.
DISCH_TO_LOC	The patient's destination/Location or status upon discharge
DISCH_DISPOSITION	Discharge disposition or Discharge Status is the patient's anticipated location or status following the encounter.
Year	Year of admission
Table: Medication	
FINNO	Same as FIN

ENCNTR_ID	The interaction between a patient and healthcare provider(s)
CATDESC	Medication name
Cat	Same as CATDESC
Comment	Availability of Medication at Healthcare facility
Table: laboratory	
FIN	- Unique ID for a patient visit – used to merge data from other sheets/tables such as (lab, clinical events, and Medication)
O_CATALOG_DISP	Test Name (ORDER_MNEMONIC can be in abbreviation form)
R_TASK_ASSAY	The specific test name (it can be the same as ORDER_MNEMONIC, and sometimes it is the specific test in a panel like CBC w/Diff)
R_RESULT_STATUS	Status of requested result
PERFORM_DT_TM	Performed results (day, time)
RESULT_VALUE_ALPHA	Result if it is in Alphabet
RESULT_VALUE_NUMERIC	Result if it is in Numeric
PR_UNITS	Result Unit if Numeric
Table: Clinical	
FIN	- Unique ID for a patient visit – used to merge data from other sheets/tables such as (lab, clinical events, and Medication)
RESULT	The result value
EVENT_CD	Clinical event recorded value
EXP	The name of the requested data (ex; weight, BMI, temperature ...., etc.)
PERFORMED_DT_TM	Clinical event recorded value (day, time)

Table 28. Feature type per each table vs correlation p-Value for each feature (Pearson Correlation)

Feature	Correlation	Feature type	Feature	Correlation	Feature type
General Admission Information					
GENDER_0	-0.033	Nominal	FINANCIAL_CLASS_Self_Pay_Insurance	0.046	Nominal
GENDER_1	0.033	Nominal	NATIONALITY_GCCCitizen	0.038	Nominal
AGE_Adult	-0.008	Nominal	NATIONALITY_NonUAE_nonGCC	0.032	Nominal
AGE_Childern	-.117**	Nominal	NATIONALITY_UAECitizen	-0.052	Nominal
AGE_Senior	.069*	Nominal	NATIONALITY_UNKNOWN	-0.040	Nominal

AGE_Young	-0.021	Nominal	ADMIT_MODE_Ambulance	-0.001	Nominal
FACILITY_AAAlAinHospital	-0.014	Nominal	ADMIT_MODE_Carried	.074*	Nominal
FACILITY_AABehaviouralScience	0.014	Nominal	ADMIT_MODE_Stretcher	0.013	Nominal
ENCNTR_TYPE_InpatientAcuteCare	-.106**	Nominal	ADMIT_MODE_Walking	-0.037	Nominal
ENCNTR_TYPE_InpatientLongTermCare	.106**	Nominal	ADMIT_MODE_Wheelchair	.065*	Nominal
FINANCIAL_CLASS_0	0.020	Nominal	ADMIT_MODE_byVehicle	-0.011	Nominal
FINANCIAL_CLASS_Commercial_or_non Goverment_Insurnace	0.002	Nominal	ADMIT_TYPE_ElectiveInpatient	.088**	Nominal
FINANCIAL_CLASS_Goverment_Insurnace	-0.005	Nominal	ADMIT_TYPE_Emergent	-.081**	Nominal
FINANCIAL_CLASS_MANDATE_Insurance	-0.023	Nominal	ADMIT_TYPE_Urgent	0.022	Nominal
laboratory Information					
InsulinFasting	0.020	Scale	HepatitisBeAgAb	0.020	Scale
@24HourUrineVanillylmandelicAcid	0.020	Scale	HepatitisBeAntibody	0.020	Scale
@5HIAA24HourUrine	0.020	Scale	HepatitisBeAntigen	0.020	Scale
ACTH	0.051	Scale	HepatitisCAntibody	0.043	Scale
ALT	-.066*	Scale	HepatitisCAntibodyConfirmation	.c	Nominal
APCR	0.041	Scale	HepatitisDTotalAntibodyIgGIgM	.c	Nominal
ARRsupinePathologistcomment	.c	Nominal	HepatitisEIgGIgMAb	0.027	Scale
AST	-.075*	Scale	HerpesSimplexVirusGenotypePCRResult	.c	Nominal
AcetaminophenLevel	-0.047	Nominal	HomocystineTotal	0.042	Nominal

AcetylcholineReceptorBindingAntibody	.[c]	Nominal	IgA	0.054	Scale
ActivatedPartialThromboplastinTime	0.007	Scale	IgE	0.049	Scale
AdenovirusResults	.[c]		IgGCSF	0.042	Scale
AlbuminLevel	.138**	Scale	IgGLvl	0.013	Scale
AlbuminLevelBodyFluid	.092**	Scale	IgGSubclasses	0.020	Scale
AlbuminCreatinineRatioUrine	.061*	Scale	IgM	0.024	Scale
AldosteroneSupine	0.035	Scale	ImmunofixationElectrophoresisSerum	.[c]	Nominal
AldosteroneReninRatiocalcStanding	.[c]	Nominal	ImmunophenotypingbyFlowBlood	.[c]	Nominal
AldosteroneReninRatiocalcSupine	0.023	Scale	ImmunophenotypingbyFlowBodyFluid	.[c]	Nominal
AlkalinePhosphotase	.120**	Scale	InfluenzaH1Results	.[c]	Nominal
AlkalinephosphataseBodyFluid	0.020	Nominal	InfluenzaVirusAntigen	.[c]	Nominal
Alpha1Antitrypsin	0.017	Scale	InfluenzaeRapidPCR	.[c]	Nominal
AlphaFetoproteinTumorMarker	0.020	Scale	InsulinAntibodies	.[c]	Nominal
AmikacinLevelTrough	.076*	Scale	InsulinLevelTotal	0.021	Scale
AmmoniaLevel	-0.056	Scale	InsulinLikeGrowthFactor1_GF1	0.023	Scale
AmoebaAntibodies	.[c]	Nominal	IronLevel	.086**	Scale
AmylaseCSF	-0.020	Nominal	IronProfileTRF	0.048	Scale
AmylaseLevel	-0.029	Scale	JAK2V617FMutationResults	.[c]	Nominal
AmylaseLevelBodyFluid	0.040	Scale	KappaLambdaFreeLightChains	0.020	Scale
AngiotensinConvertingEnzyme	.[c]	Nominal	KleihauerABetketest	.[c]	
AntiStreptolysinOQuant	0.031	Nominal	LDHCSF_x	0.034	Scale
AntiIA2	0.018	Nominal	LMWHeparinLevel	0.035	Scale

AntibodyIdentification	.[c]	Nominal	LactateDehydrogenase	0.019	Scale
AntibodyScreen	.[c]	Nominal	LactateDehydrogenaseBodyFluid	.081**	Scale
AntibodyTiter	.[c]	Nominal	LacticAcid	-.071*	Scale
AntigenTyping	.[c]	Nominal	LacticAcidCSF	0.030	Scale
AntimitochondrialAntibodyScreen	.[c]	Nominal	LegionellaAntigenUrine	.[c]	Nominal
AntinuclearAntibodyScreen	.[c]	Nominal	LipaseLevel	0.032	Scale
AntinuclearAntibodybyIIFPatternand	.[c]	Nominal	LipidPanel	.073*	Scale
AntithrombinIIIAssay	.067*	Scale	LithiumLevel	-0.048	Scale
BTypeNatriureticPeptide	0.038	Scale	LiverKidneyMicrosomalAntibodies	.[c]	Nominal
BCRABL1Results	.[c]	Nominal	LupusAntiicoagulantTW	0.036	Scale
BenceJonesProteinUrine	.[c]	Nominal	LuteinizingHormone	0.020	Nominal
Beta2GlycoproteinsIgGIgM	-0.013	Scale	LymphocytesSubsetsCD4CD8byFlowCy	0.028	Scale
Beta2Microglobulin	0.020	Scale	MDxMERSCoronavirusRTPCR	.[c]	Nominal
BetaHydroxybutyratelevel	0.020	Scale	MagnesiumLevel	0.060	Scale
BetahCGQuantitative	-0.002	Scale	MagnesiumLevelUrine	0.020	Scale
BilirubinBodyFluid	0.028	Scale	MalariaScreen	.[c]	Nominal
BilirubinDirect	.065*	Scale	MeaslesIgG	.[c]	Nominal
BilirubinTotal	0.051	Scale	Metanephrines24HourUrine	0.048	Scale
BloodGasArterialAA	.087**	Scale	MiscellaneousSendOutLab	.[c]	
BloodGasArterialPOCT	0.008	Scale	MixingStudyAPTT	0.044	Scale
BloodGasCapillaryAA	-0.047	Scale	MixingStudyPT	0.040	Scale
BloodGasCapillaryPOCT	-0.005	Scale	MumpsIgG	.[c]	Nominal
BloodGasMixVenousPOCT	-0.019	Scale	MycobacteriumResults	.[c]	Nominal

BloodGasPleuralFluid	0.031	Scale	MycobacteriumtuberculosisRapidPCR	.[c]	Nominal
BloodGasVenousAA	0.012	Scale	MycoplasmaIgM	0.025	Scale
BloodGasVenousPOCT	0.007	Scale	MyoglobinUrine	-0.005	Scale
BloodGrouping	.[c]	Nominal	NTProBTypeNatriureticPeptide	0.026	Scale
BrucellaAntibodyTiterMelitensisAbor	.[c]	Nominal	NeutrophilCytoplasmicAntibodycANCA	0.044	Scale
CPeptide	0.010	Scale	NeutrophilCytoplasmicAntibodypANCAT	0.051	Nominal
CReactiveProtein	.067*	Scale	OligoclonalBandsCSF	.[c]	Nominal
C3Complement	.078*	Scale	OsmolalitySerum	0.049	Scale
C4Complement	.075*	Scale	OsmolalityUrine	.105**	Scale
CA125	0.040	Scale	PTINR	-0.003	Scale
CA153	0.018	Scale	PTHIntact	.070*	Scale
CA199	0.033	Scale	ParvovirusB19IgMAntibody	.[c]	Nominal
CBCwDiff	0.025	Scale	PathComments	.[c]	Nominal
CEA	0.034	Scale	PathologistCommentGL	.[c]	Nominal
CMVAbIgGIgM	0.015	Scale	PathologistReview	.[c]	Nominal
CalcitoninLevel	.[c]	Nominal	PathologistReviewCoagulation	.[c]	Nominal
CalciumLevel	.142**	Scale	PathologySendoutRequest	.[c]	Nominal
CalciumLevelBodyFluid	0.020	Scale	PhenytoinLevelTotal	0.015	Scale
CalciumLevelCorr	.105**	Scale	PhosphorusLevel	-0.052	Scale
CarbamazepineLevel	0.020		PlasmaOrder	.[c]	Nominal
CardiolipinAntibodies	0.025	Scale	PlateletAggregation	.[c]	Nominal
CardiolipinAntibodyIgG	-0.014	Scale	PlateletFunctionScreen	0.060	Scale
CardiolipinAntibodyIgM	0.029	Scale	PlateletOrder	.[c]	Nominal
CatecholaminesFractionated24HourUrin	-0.007	Scale	PostABORh	.[c]	Nominal
CeliacDisease	0.020	Scale	PostAbSc	.[c]	Nominal
CellCountwDiffCSF	-0.048	Scale	PostTxRxDATPoly	.[c]	Nominal

CellCountwDiffBody Fluid	0.038	Scale	PostXM	.[c]	Nominal
Ceruloplasmin	0.022	Scale	PotassiumLevel	.171**	Scale
ChlamydiaAntigen	.[c]	Nominal	PotassiumLevelBodyFluid	0.028	Scale
ChlorideLevel	.175**	Scale	PotassiumLevelUrine	-0.018	Nominal
ChlorideLevelCSF	0.027	Nominal	PreABORh	.[c]	Nominal
CholesterolBodyFluid	0.020	Scale	PreAbSc	.[c]	Nominal
CholesterolHDL	0.021	Scale	PreXM	.[c]	Nominal
CholesterolTotal	.119**	Scale	Prealbumin	0.020	Scale
ChromosomeAnalysis SolidTissueWorkup	.[c]	Nominal	Procalcitonin	-0.027	Scale
CopperLevel	.[c]	Nominal	ProlactinLevel	0.041	Scale
Cortisol24HourUrine	0.027	Scale	ProstateSpecificAntigen	.068*	Scale
CortisolLvl	0.026	Scale	ProstateSpecificAntigenProfile	0.027	Scale
CreatineKinase	0.021	Scale	Protein24HourUrine	0.055	Scale
CreatineKinaseMBIsoenzyme	-0.058	Scale	ProteinBodyFluid	.102**	Scale
CreatineKinaseMBMass	-0.043	Scale	ProteinCActivity	0.048	Scale
CreatinineBodyFluid	0.020		ProteinCSF	-0.002	Scale
CreatinineLvl	0.023	Scale	ProteinElecSerum	.[c]	Nominal
CrimeanCongoFeverIgG	.[c]	Nominal	ProteinElectrophoresisSerum	0.028	Scale
CrimeanCongoFeverIgM	.[c]	Nominal	ProteinSFreeAg	0.050	Scale
CrimeanCongoFever RTPCR	.[c]	Nominal	ProteinTotal	.186**	Scale
Crossmatch	.[c]	Nominal	ProteinUrine	0.021	Nominal
Cryoglobulin	.[c]	Nominal	ProteinCreatinineRatioUrine	0.054	Scale
CryoprecipitateOrder	.[c]	Nominal	QuantiFeronTB	.[c]	Nominal

CyclicCitrullinatedPeptideAntibodyIg	0.007	Scale	RBCOrder	.[c]	Nominal
CyclosporinC0	0.020	Nominal	RapidPlasmaReaginAbQuantitative	.[c]	Nominal
CytomegalovirusQuantitativeResults	0.020	Scale	ReninActivityStanding	0.020	Scale
DDimerTW	-.077[*]	Scale	ReninActivitySupine	0.028	Scale
DATInterp	.[c]	Nominal	RespiratoryPathogenIDPCRResults	.[c]	Nominal
DATMono	.[c]	Nominal	RespiratorySyncytialVirusRSVAntige	.[c]	Nominal
DATPoly	.[c]	Nominal	ReticCountAuto	0.022	Scale
DNAAntibodyDoublestranded	0.034	Scale	RheumatoidFactorQuantitative	-0.008	Nominal
DehydroepiandrosteroneSulphateDHEAS	0.028	Scale	RubellaIgG	.[c]	Nominal
DengueIgG	.[c]	Nominal	SalicylateLevel	0.036	Nominal
DengueIgM	.[c]	Nominal	SalmonellaAb	.[c]	Nominal
DenguefeverPCR	.[c]	Nominal	SendouttestMAYO	.[c]	Nominal
DigoxinLevel	0.058	Scale	SickleCellSolubilityTest	.[c]	Nominal
DonorABORh	.[c]	Nominal	SmoothMuscleAntibodyScreen	.[c]	Nominal
EchinococcusAntibody	.[c]	Nominal	SodiumLevel	.182[**]	Scale
ElectrolytePanel	.078[*]	Scale	SodiumLevelBodyFluid	0.044	Scale
ElectrolytePanelUrine	0.037	Scale	SodiumLevelUrine	.093[**]	Scale
EluateAbSc	.[c]	Nominal	SynovialFluidCrystals	.[c]	
EndomysialAntibody	.[c]	Nominal	Syphilis	0.021	Scale
EnterovirusRapidPCR	.[c]	Nominal	T3Free	.098[**]	Scale
EpsteinBarrVirusAntibodies	0.032	Scale	TestosteroneLevelTotal	0.020	Scale
ErythrocyteSedimentationRate	.099[**]	Scale	ThiopurineMethylTransferaseTpmtRbs	.[c]	Scale
EthanolLevel	-0.030	Scale	ThyroglobulinAntibody	0.033	Scale
ExtractableNuclearAntigenENAScreen	0.023	Scale	ThyroidFunctionTest	.085[**]	Scale

ExtractableNuclearAntigenAntibodyEN	.[c]	Nominal	ThyroidPanel	0.057	Scale
FSHLevel	0.020	Scale	ThyroidPeroxidaseAntibodyTPO	-0.003	Scale
FactorIXAssay	0.028	Scale	ThyroidStimulatingHormone	0.034	Scale
FactorVAssay	0.020	Scale	ThyrotropinReceptorAntibody	0.025	Scale
FactorVIIAssay	0.020	Nominal	TissueTransglutaminaseIgA	0.005	Scale
FactorVIIIAssay	0.033	Nominal	TotalThyroxin	-0.006	Scale
FactorXAssay	-0.027	Scale	Toxoplasma	0.029	Scale
FactorXIAssay	0.028	Scale	Transferrin	.066*	Scale
FactorXIIAssay	0.028	Scale	TreponemaPallidumHemagglutinationTPH	.[c]	Nominal
FecalCalprotectin	0.020		TriglycerideBodyFluid	0.030	Scale
FerritinLvl	0.013	Scale	Triglycerides	.080**	Scale
FibrinogenLvl	.079*	Scale	TroponinI	-0.055	Scale
FolateLevel	.091**	Scale	TroponinT	-.072*	Scale
FreeT4	.092**	Scale	UE	.065*	Scale
GGT	.105**	Scale	UnitTypingCount	.[c]	
GastricParietalCellAntibodyGPC	.[c]	Nominal	UreaLvl	.105**	Scale
GentamicinLevel	0.049	Scale	UricAcid	.118**	Scale
GentamicinLevelPeak	0.020	Scale	UricAcidBodyFluid	0.020	Scale
GentamicinLevelTrough	0.054	Scale	UricAcidUrine	-0.048	Nominal
GlucoseBodyFluid	.093**	Scale	UrinalysisPOCT	0.005	Scale
GlucoseCSF	0.028	Scale	UrineAnalysis	.132**	Scale
GlucoseFasting	0.045	Scale	UrineToxodrugScreen	.[c]	Nominal
GlucosePOCT	0.005	Scale	ValproicAcidLevel	-0.017	Scale
GlucoseRandom	-0.018	Scale	VancomycinLevelPeak	0.026	Scale
Glucose6PDQuantitative	0.051	Scale	VancomycinLevelTrough	.170**	Scale
GlutamicAcidDecarboxilaseAntibodies	0.007	Scale	VaricellaZosterAntibodyIgG	.[c]	Nominal

HBVQuantitativeResults	.c	Nominal	VaricellaZosterAntibodyIgM	0.028	Scale
HIV1and2WesternBlot	.c	Nominal	VaricellaZosterVirusResults	.c	Nominal
HIVAgAbCombination	0.037	Scale	VitaminB12Level	.080**	Scale
HIV1QuantitativeResults	0.020	Scale	VitaminD25HydroxyLevel	.071*	Scale
HLAClassIDiseaseAssociation	.c	Nominal	VonWillebrandActivity	0.020	Scale
HSV12IgGIGMAntibodies	0.035	Scale	ZincLevel	.c	
Haptoglobin	.068*	Scale	eGFRCKDEPI	-0.020	Scale
HbElecHPLC	0.036		iSTATChemistryPOCT	.146**	Scale
HemoglobinA1c	.092**	Scale	vWFAntigen	0.020	Scale
HeparinInducedThrompocytopeniabyFlow	0.039	Scale	HepatitisBCoreAntibodyIgM	0.020	Scale
HepaticFunctionPanel	-0.049	Scale	HepatitisBSurfaceAntibody	0.047	Scale
HepatitisAAntibodyIgG	.c	Nominal	HepatitisBSurfaceAntigen	0.025	Scale
HepatitisAAntibodyIgM	0.013	Scale	HepatitisBCoreAntibody	0.011	Scale
Clinical Information					
BodyMassIndex	-0.004	Scale	Pulseoximetryassesement	0.040	Nominal
DiastolicBloodPressure	-0.025	Scale	RespiratoryRate	-0.044	Scale
ESR	0.047	Scale	SystolicBloodPressure	-0.019	Scale
LDH	0.054	Scale	TemperatureAxillary	.066*	Scale
LDHBF	0.034	Scale	TemperatureOral	0.050	Scale
LDHCSF_y	0.028	Scale	TemperatureSkin	0.020	Scale
PeripheralPulseRate	-0.052	Scale	TemperatureTympanic	-0.032	Scale
Weight	-0.023	Scale			
Medication					
Aspirin	.108**	Nominal	insulindetemir	0.040	Nominal
Dextrose10with0_9NaCl	0.034	Nominal	insulinglargine	.066*	Nominal
InsulindegluDEC	-.078*	Nominal	insulinglulisine	-0.044	Nominal

SodiumChloride0_9Flushes	0.034	Nominal	insulinisophane	-0.005	Nominal
acetaminophen	.234**	Nominal	insulinisophaneinsulinregular	-0.024	Nominal
albuterol	.188**	Nominal	insulinlispro	0.045	Nominal
amiodarone	.091**	Nominal	insulinlisproinsulinlisproprotamine	0.020	Nominal
amoxicillin	-0.024	Nominal	insulinregular	-0.003	Nominal
amoxicillinclavulanate	0.026	Nominal	ipratropium	.263**	Nominal
ampicillin	0.006	Nominal	ketorolac	.099**	Nominal
atenolol	0.028	Nominal	labetalol	.123**	Nominal
atorvastatin	.139**	Nominal	levETIRAcetam	.109**	Nominal
atropine	0.026	Nominal	levoFLOxacin	.069*	Nominal
calciumgluconate	0.041	Nominal	levoTHYROxine	0.047	Nominal
captopril	-0.029	Nominal	lisinopril	.070*	Nominal
carvedilol	0.045	Nominal	magnesiumsulfate	-0.003	Nominal
ceFAZolin	.139**	Nominal	meropenem	.255**	Nominal
cefTRIAXone	.119**	Nominal	metoCLOPRAMIDE	.193**	Nominal
cefepime	.079*	Nominal	metoPROLOL	0.056	Nominal
ciprofloxacin	.102**	Nominal	metroNIDAZOLE	.116**	Nominal
clonazePAMCD	0.015	Nominal	midazolamCD	.188**	Nominal
clopidogrel	.106**	Nominal	neostigmine	.093**	Nominal
dextrose10inwater	0.032	Nominal	nitroglycerin	0.027	Nominal
dextrose20inwater	0.028	Nominal	norepinephrine	.093**	Nominal
dextrose25inwater	0.020	Nominal	ondansetron	.100**	Nominal
dextrose4sodiumchloride0_18	0.040	Nominal	oxyCODONEN	0.024	Nominal
dextrose5inwater	.407**	Nominal	pantoprazole	.216**	Nominal
dextrose5with0_23NaCl	0.032	Nominal	penicillinGsodium	0.049	Nominal
dextrose5with0_45NaCl	.183**	Nominal	phenylephrine	.110**	Nominal
dextrose5with0_9NaCl	.305**	Nominal	phenytoin	0.052	Nominal
dextrose50inWater	.089**	Nominal	phytonadione	.098**	Nominal
dextrose70inWater	0.056	Nominal	piperacillintazobactam	.300**	Nominal
diazepamCD	0.021	Nominal	potassiumchloride	.326**	Nominal
digoxin	.074*	Nominal	potassiumphosphate	0.059	Nominal
diltiazem	0.034	Nominal	potassiumsodiumhydrogencitrate	-0.016	Nominal
diphenhydramine	0.020	Nominal	pravastatin	0.028	Nominal
enoxaparin	.294**	Nominal	predniSONE	0.020	Nominal
epoetinalfa	0.020	Nominal	propOFOLSCD	.263**	Nominal
epoetinbetamethoxypolyethyleneglycol	.065*	Nominal	ranitidine	.062*	Nominal
fentaNYLN	.278**	Nominal	rosuvastatin	.062*	Nominal
fluconazole	.130**	Nominal	simvastatin	-0.044	Nominal
fondaparinux	0.044	Nominal	sodiumchloride0_45	.198**	Nominal
furosemide	.339**	Nominal	sodiumchloride0_9	.176**	Nominal

glucagon	0.014	Nominal	sodiumchloride3	.098**	Nominal
haloperidolCD	.121**	Nominal	sodiumchloride7	0.020	Nominal
heparin	0.043	Nominal	sulfamethoxazoletrimethoprim	0.004	Nominal
hydrALAZINE	.139**	Nominal	tacrolimus	0.020	Nominal
hydrochloroTHIAZIDE	.066*	Nominal	terlipressin	-.069*	Nominal
insulinaspart	.070*	Nominal	vancomycin	.148**	Nominal
insulinaspartinsulinaspartprotamine	-0.007	Nominal	warfarin	.092**	Nominal

* Significant: p-Value (1-tailed), ** Significant: p-Value (2-tailed)