Prediction of ICU Readmissions using LSTM in Low and Middle Income Countries

A dissertation submitted for the Degree of Master of Business Analytics

N. F. Fazla

University of Colombo School of Computing

2024

Declaration

Name of the student: N. F. Fazla	
Registration number: 2020/BA/013	
Name of the Degree Programme: Master of Business Analytics	
Project/Thesis title: Prediction of ICU Readmissions using LSTM in Low and Middle Income Countries	

1. The project/thesis is my original work and has not been submitted previously for a degree at this or any other University/Institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.

2. I understand what plagiarism is, the various types of plagiarism, how to avoid it, what my resources are, who can help me if I am unsure about a research or plagiarism issue, as well as what the consequences are at University of Colombo School of Computing (UCSC) for plagiarism.

3. I understand that ignorance is not an excuse for plagiarism and that I am responsible for clarifying, asking questions and utilizing all available resources in order to educate myself and prevent myself from plagiarizing.

4. I am also aware of the dangers of using online plagiarism checkers and sites that offer essays for sale. I understand that if I use these resources, I am solely responsible for the consequences of my actions.

5. I assure that any work I submit with my name on it will reflect my own ideas and effort. I will properly cite all material that is not my own.

6. I understand that there is no acceptable excuse for committing plagiarism and that doing so is a violation of the Student Code of Conduct.

Signature of the Student	**Date (DD/MM/YYYY)**
N.F. Fazla	29/09/2024

Certified by Supervisor(s)

This is to certify that this project/thesis is based on the work of the above-mentioned student under my/our supervision. The thesis has been prepared according to the format stipulated and is of an acceptable standard.

	Supervisor 1	**Supervisor 2**	**Supervisor 3**
Name	Dr. K. A. T. K. Karunanayake		
Signature			
Date	30/09/2024		

I would like to dedicate this thesis to my parents for being my source of inspiration, support and guidance!

ACKNOWLEDGEMENTS

I would like to express a deep sense of gratitude and appreciation to all those who made it possible for me to do the project successfully. First and foremost I would like to express my heartiest gratitude towards my supervisors, Dr. K.A.K.T. Karunanayaka and Dr. Abigail Beane. Without their wise counsel and guidance, it would have been impossible to do this project successfully.

I extend my heartfelt thanks to my parents and friends for their invaluable support, which greatly contributed to the successful completion of the project.

ABSTRACT

This study addresses the persistent challenge of Intensive Care Unit (ICU) readmission, focusing on the unique context of Lower and Middle Income Countries (LMICs). Despite advancements in medical technology, ICU readmissions remain a critical issue, with implications for healthcare resources and patient outcomes. To address the challenge of ICU readmission, accurate prediction models are needed to identify patients at high risk of readmission because the prediction of readmission before the patient is discharged, will help physicians re-evaluate the discharge of the patient and reduce the immature discharges. The existing literature predominantly stems from high-income countries (HICs), and this study aims to fill the gap by developing a predictive model tailored to LMICs context. It utilizes the Long Short-Term Memory (LSTM), known for its ability to capture temporal dependencies in sequential patients' data to predict the early ICU readmission (readmission within 48 hours followed by index discharge) of the patients and feature ablation test to extract the important factors associated with ICU readmission. 2.85% (306) of discharges to the wards were later readmitted within 48 hours to the intensive care unit. The LSTM model with a cost-sensitive training had significantly better performance (area under the receiver operating curve, 0.68) compared to the baseline models with traditional machine learning approaches. It highlights that the deep learning models improve the accuracy of decision-making in predicting ICU readmission.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

ABSTRACT iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

LIST OF TABLES viiii

CHAPTER 1: INTRODUCTION 1

1.1 Motivation 2

1.2 Statement of the problem 2

1.3 Research Aim and Objectives 3

1.3.1 Aim 3

1.3.2 Objectives 3

1.4 Scope 4

1.5 Structure of the dissertation 4

CHAPTER 2: LITERATURE REVIEW 5

CHAPTER 3: METHODOLOGY 8

3.1 Dataset preparation 8

3.2 Features selection 9

3.3 Feature extraction for reason for admission 11

3.4 Statistical features for baseline models 11

3.5 Models 12

3.5.1 Baseline models 12

3.5.2 LSTM model 12

3.6 Model training 12

3.7 Feature ablation test for feature importance 13

3.8 External validation 14

3.9 Further exploration on Transformer-based model 15

3.10 Ethics 17

CHAPTER 4: EVALUATION AND RESULTS 18

4.1 Patient characteristics 18

4.2 Models 19

4.3 Ranking of features 21

4.4 External validation 21

4.5 Comparison with transformer-based model 22

CHAPTER 5: CONCLUSION AND FUTURE WORK 24

5.1 Conclusion 24

5.2 Future work 25

APPENDIX I

REFERENCES II

LIST OF FIGURES

Figure 1: Consort diagram of inclusion and exclusion of admissions to the study 9
Figure 2: Statistical features for baseline models 11
Figure 3: LSTM architecture 13
Figure 4: High level architecture of transformer-based model 16
Figure 5: AUC curve of LSTM model for LMIC dataset 19
Figure 6: Feature ranking chart 21
Figure 7: AUC curve of LSTM model for daily MIMIC III dataset 22
Figure 8: AUC curve of transformer-based model for LMIC dataset 23

LIST OF TABLES

Table 1: Variables selected for model training 10

Table 2: Comparison of case-mix, organ support and severity of illness between non-readmitted and readmitted admissions 15-16

Table 3: AUC score comparison between baseline models and LSTM 17

Table 4: AUC score comparison to previously proposed deep learning models developed for patients in HICs 17

Table 5: AUC score of the LSTM models on different datasets 22

Table 6: Comparison of AUC scores between the LSTM models and the transformer-based models on different datasets 23

CHAPTER 1

INTRODUCTION

While advancements in medical technology and the expertise of healthcare professionals have improved patient outcomes in Intensive Care Units (ICUs), the issue of ICU readmission remains a persistent challenge (Elliott, Worrall-Carter and Page, 2014). ICU readmission is defined as a second admission to the same ICU from which the patient was originally discharged during the same hospitalization (Brown, Ratcliffe and Halpern, 2013). Early ICU readmissions refer to instances in which patients return to the ICU within 48 hours of their initial discharge. Conversely, ICU readmissions, occurring between 3 and 14 days subsequent to discharge within the same hospital encounter, are characterized as late readmissions (Aryal et al., 2023).

Currently the incidence of ICU readmissions within the first 48 hours is widely used as a performance metric reflecting the quality of intensive care medicine (Aryal et al., 2023; Lin et al., 2018). High readmission rates not only strain healthcare resources but also have negative implications for patient outcomes, including increased morbidity, mortality, and healthcare costs (Aryal et al., 2023; Lin et al., 2018; Jo et al., 2015). Most of the available evidence regarding the occurrence, features, related risk factors, and resulting outcomes of ICU readmissions is derived from critical care populations and health systems in high-income countries (HICs) (Aryal et al., 2023).

Low and middle income countries (LMICs) face distinct healthcare challenges compared to HICs, including limited resources, inadequate infrastructure, higher mortality and a higher burden of infectious diseases and non-communicable diseases (Getzzg, 2023; Miranda et al., 2008). Consequently, the factors influencing ICU readmission in these regions may differ significantly from those observed in HICs and continue to be largely lacking (Aryal et al., 2023). Thus the development of prediction models tailored to these specific contexts is crucial to enhance patient care and resource allocation.

Deep learning techniques such as Long Short-Term Memory (LSTM) can be applied to develop such predictive models for ICU readmission because of their ability to effectively capture temporal dependencies and make predictions based on the history of observations by utilizing the sequential nature of the data (Lin et al., 2019). This capability makes deep learning approaches particularly suitable for analyzing and predicting outcomes in ICU patients over traditional statistical methods and machine learning techniques, where temporal dynamics play a significant role in understanding their health status and identifying potential risks including ICU readmissions (Cascarano et al., 2023).

1.1 Motivation

Due to the increased interest in the ICU readmission prediction, several predictor models have been proposed using classical machine learning techniques (Rojas et al., 2018; Pakbin et al., 2018; Veloso et al., 2014) and deep learning techniques (Lin et al., 2019; Barbieri et al., 2020; Ashfaq et al., 2019). But the predictive models proposed using the data collected from the patients of HICs, may not yield accurate predictions for ICU readmission in LMICs because of the models' inherent limitations in capturing factors that are contextually specific to readmission occurrences in LMICs. It provides a compelling basis for the development of a dedicated model for LMICs, highlighting the necessity to consider factors that are distinctly relevant to the healthcare settings in these regions.

Deep learning models can effectively model the temporal dependencies and sequential patterns in time series data, which is crucial in ICU readmission prediction and it offers the advantage of transfer learning, where models pretrained on specific datasets can be fine-tuned to predict the ICU readmissions to a specific ICU in LMICs with limited data (Hosna et al., 2022). By considering the need for a prediction model for ICUs in LMICs and the above mentioned advantages of deep learning techniques, this study aims to develop a predictive model for ICU readmission using LSTM in the setting of LMICs to support healthcare providers in decision making.

1.2 Statement of the problem

To address the persistent challenge of ICU readmission, coupled with its impact on healthcare resources, patient outcomes, and costs, accurate prediction models are needed to identify patients at high risk of readmission because the prediction of readmission before the patient is discharged, will help physicians re-evaluate the discharge of the patient and reduce the immature discharges.

LMICs, confronting unique healthcare challenges, require tailored approaches to predict ICU readmission and understand the factors influencing it. Existing predictive models often lack specificity for LMICs compared to HICs as none of the studies have so far proposed a deep learning predictive model capturing time series nature of the patients’ data for ICU readmissions in LMICs.

This study will address this void by developing a predictive model that utilizes the LSTM technique to capture the temporal nature of the patients' data collected from ICUs in LMICs. Further, this study will extract the risk factors associated with readmissions to ICUs in LMICs using feature importance techniques. The ability of predicting the patients at high risk of readmitting with related risk factors, will empower healthcare providers with the ability for proactive interventions, resulting in improving patient outcomes, enhancing care quality, facilitating care coordination, mitigating financial burdens and resource utilization within the ICU setting and beyond.

1.3 Research Aims and Objectives

1.3.1 Aim

This study aims to develop a predictive model for ICU readmission for the regions of LMICs utilizing LSTM technique and extract the key factors associated with ICU readmissions in these regions.

1.3.2 Objectives

The primary objective of this study is to develop a robust and effective prediction model using LSTM technique to forecast the likelihood of patients being readmitted to the ICU within 48 hours following index discharge in the region of LMICs as longer intervals between index discharge and readmission are less likely to be related to ICU discharge circumstances (Brown, Ratcliffe and Halpern, 2013; Brown, Ratcliffe and Halpern, 2015).

Specific objectives are,

- To build an early ICU readmission (within 48 hours following the discharge) prediction model in the setting of LMICs by utilizing the LSTM technique
- To extract the key risk factors associated with ICU readmission in these region
- To compare the performance of the LSTM model to the baseline models and deep learning models developed for HICs

1.4 Scope

The scope of this study is to implement a predictive model using LSTM to predict early readmissions of adult patients (age greater than 18) to ICUs in the same hospitalization and extract the key features associated with early ICU readmission from the data collected from the ICUs in the LMICs.

1.5 Structure of the dissertation

The rest of the dissertation is organized as follows,

Chapter 2: Literature review summarizes the previous works carried out in the field, providing a comprehensive review of relevant literature to contextualize the current study. This section critically examines existing studies, highlights key findings, identifies gaps, and establishes the theoretical framework for the study.

Chapter 3: Methodology describes methodological approach adopted for the study, encompassing description and preprocessing of the data source. It outlines the criteria and rationale for feature selection, followed by the process of feature extraction. It further explains the steps involved in model development and model training. Finally, it explains the concept of feature ablation tests, providing insights into the methodology employed to assess the impact of individual features on the model's performance.

Chapter 4: Evaluation and Results presents the outcomes of the study, detailing the results obtained for patients characteristics of both readmitted and non-admitted groups, the developed models compared to previously proposed models and the feature ablation tests. It provides a comprehensive analysis of the model performances, highlighting key findings, and discusses the implications of the results.

Chapter 5: Conclusion and Future work discusses the key findings of the study, emphasizing the contributions made to the field. It provides a concise summary of the results, discussing their implications and relevance to the research question. The section also reflects on the limitations of the study and suggests avenues for future work

CHAPTER 2
LITERATURE REVIEW

The literature surrounding ICU readmissions prediction techniques encompasses a diverse array of studies, methodologies, and insights that collectively contribute to the advancement of healthcare decision-making. This section reviews the key findings and trends within this domain, highlighting both the progress achieved and the gaps that persist.

Existing predictive models for ICU readmission often rely on traditional statistical methods or classical machine learning techniques (Rojas et al., 2018; Pakbin et al., 2018; Veloso et al., 2014). When discussing classical machine learning methods, the study by Rojas et al. (2018) applied a gradient boosting technique and the study by Pakbin et al. (2018) applied the extreme gradient boosting (XGBoost) technique to predict the readmissions. They both utilized data from the last 24 hours prior to ICU discharge. Though these studies have achieved some success in predicting ICU readmissions, they have limitations. For instance, these studies often lack the ability to handle time-series data and capture temporal dependencies, which are important in ICU readmission prediction (Cascarano et al., 2023).

Deep learning models are well-suited for handling sequential data, such as time-series data as they have the ability to learn from the temporal patterns in the data and capture temporal dependencies, making them a powerful tool for predicting ICU readmissions accurately. Thus to overcome the above mentioned limitations in traditional methods and classical machine learning techniques, the use of deep learning models has been explored in ICU readmission prediction (Lin et al., 2019; Barbieri et al., 2020; Ashfaq et al., 2019) given deep learning techniques have shown to outperform shallow machine learning models and traditional data analysis approaches in many applications (Miranda et al., 2008). Ultimately they have shown better performance compared to existing state-of-the-art models in predicting ICU readmissions (Lin et al., 2019).

Studies introducing deep learning models for predicting ICU readmission have employed a Recurrent Neural Network (RNN) as in the work by Barbieri et al. (2020), or an RNN with LSTM as demonstrated by Lin et al. (2019) and Ashfaq et al. (2019). While studies by Lin et al. (2019) and Barbieri et al. (2020) predict the readmission within a 30-day period following ICU discharge, the study by Ashfaq et al. (2019) predicts the readmission within different time intervals; 24 hours, 72 hours, 7 days, 30 days, and bounceback readmissions within the same hospital admission. However longer intervals between index discharge and readmission are less likely to be related to ICU discharge circumstances (Brown, Ratcliffe and Halpern, 2013; Brown, Ratcliffe and Halpern, 2015).

Studies by Lin et al. (2019) and Barbieri et al. (2020) utilized longitudinal clinical data obtained from the publicly accessible MIMIC-III dataset. In contrast, the study by Ashfaq et al. (2019) relied on data gathered from more than 7500 congestive heart failure (CHF) patients hospitalized between 2012 and 2016 in Sweden. The primary distinction in the patient populations studied lies in the fact that the latter specifically concentrated on a population afflicted with a specific disease (CHF), thereby narrowing the target population, whereas the former two did not focus on any disease-specific population.

In the context of inclusion and exclusion criteria of patients, all three studies discussed above share the common practice of including only adult patients and excluding those who died at the time of discharge of index admission. However, a discrepancy arises in the inclusion of patients discharged and subsequently deceased within 30 days, with varying approaches in the studies conducted by Lin et al. (2019) and Barbieri et al. (2020). While Barbieri et al. (2020) excluded these patients, Lin et al. (2019) included them and assigned them into the readmitted class. Notably, the rationale behind the inclusion of these specific patients is not explicitly justified or supported by an argument by Lin et al. (2019).

In terms of features used for the deep learning model training, the studies extracted the features as follow,

1. Lin et al. (2019) extracted three distinct sets of features: chart events, ICD-9 embeddings, and patient demographic information.
2. Barbieri et al. (2020) incorporated the temporal dynamics of code embeddings computed by neural ordinary differential equations (ODEs).
3. Ashfaq et al. (2019) extracted both human and machine derived features.

Among all three studies discussed above, only the study by Ashfaq et al. (2019) addresses the class imbalance issue by introducing cost-sensitive classification.

All the aforementioned studies focused on predicting the readmissions through the utilization of supervised machine learning techniques. In contrast, the study by Veloso et al. (2014) adopted a clustering methodology to predict the readmissions. This study used the data collected at the time of discharge and applied different clustering algorithms that are k-means, k-medoids, x-means and DBSCAN.

However, one of the main concerns of the deep learning models over the traditional statistical and machine learning models is the complexity and interpretability (Amorim et al., 2020). They are often considered "black boxes" meaning that it can be challenging to understand the underlying factors driving the predictions. Therefore very few studies have attempted to understand and interpret the deep learning predictive model. Feature interpretation of the machine learning models is crucial, and more imperative for clinical applications. This task is much more complex for deep learning techniques, which has made recent works short of explaining the decision making logic and model interpretation (Lin et al., 2019).

Summing up, the literature on the ICU readmissions prediction through classical machine learning and deep learning approaches contributes to enhancing patient outcomes and optimizing healthcare resources. However, while progress has been made, the following remains unaddressed,

- **Absence of LMICs oriented models**: A noticeable gap is the absence of deep learning models tailored for LMICs. These models would need to account for the unique healthcare landscape, resource constraints, and patient demographics prevalent in LMICs.

To address the lack of the representation of LMICs specific ICU readmission prediction models, this study aims to implement a LSTM model specifically designed to predict ICU readmissions among adult patients in the ICUs of LMICs and extract the key factors associated with ICU readmissions in these regions.

CHAPTER 3

METHODOLOGY

3.1 Dataset preparation

This multi-centre retrospective study uses the data submitted by a collaborating registry to the CCAA repository. Data elements including demographics, admission diagnosis, pre-existing comorbidity, route of admission, ICU outcome, physiological vital signs, laboratory measurements, readmission status, admission and discharge summaries, are routinely reported as part of the core dataset and daily dataset. Data is collected daily contemporaneous to patient care and all data collectors are trained in definitions and data acquisition.

ICUs in the registry that collected daily data were only included in the study. From data submitted between 2020-12-01 and 2023-11-30, patients were screened out based on the following criteria to predict readmissions within 48 hours as longer intervals between discharge and readmission are less likely to be related to ICU discharge circumstances (Brown, Ratcliffe and Halpern, 2013; Brown, Ratcliffe and Halpern, 2015),

1. Patients under 18 years at the time of admission (Lin et al., 2019; Barbieri et al., 2020)
2. Patients who died in the first ICU admission (Lin et al., 2019; Barbieri et al., 2020)
3. Patients discharged to ward and dead within 48 hours (Barbieri et al., 2020)
4. Patients who were discharged to home or transferred to another ICU/another hospital as their readmission status wasn't captured in the data (Rojas et al., 2018)

Then all the patients and their corresponding ICU stay records in the final dataset were labeled as positive (readmitted) or negative (not-readmitted) based on their post-discharge outcome as follow,

- Patients discharged to ward and not readmitted to ICU within 48 hours as negative
- Patients discharged to ward and readmitted to ICU within 48 hours as positive

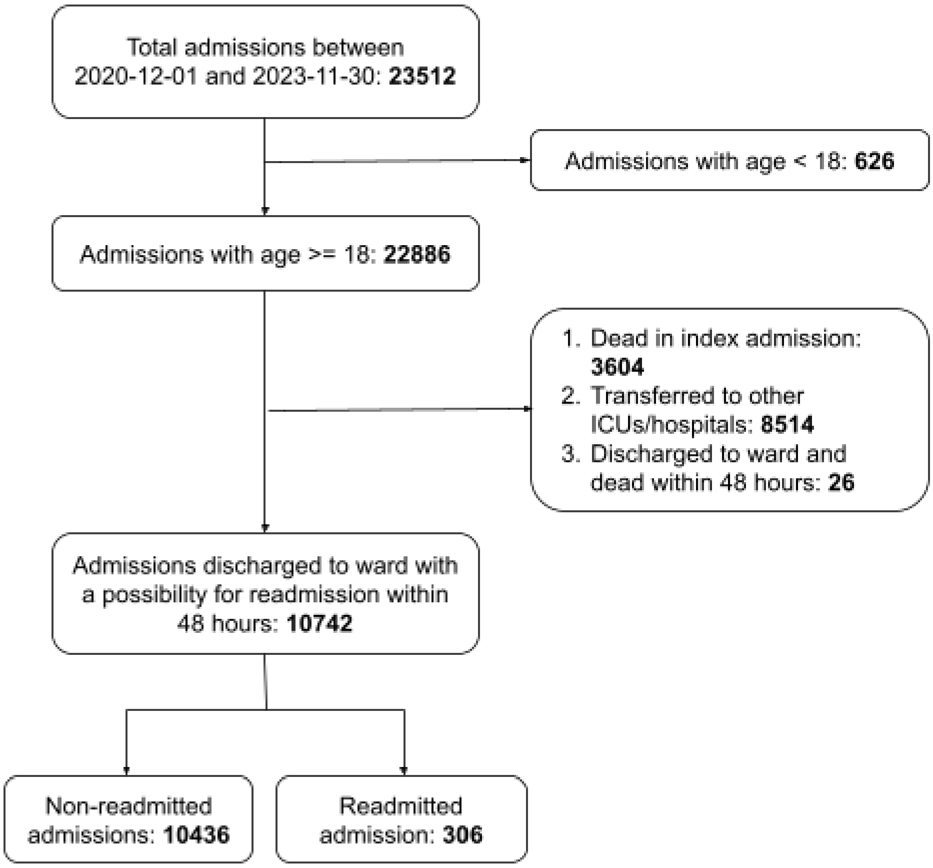

Figure 1: Consort diagram of inclusion and exclusion of admissions to the study

3.2 Features selection

Because of the availability of the variables in the dataset, various combinations of completeness thresholds and the corresponding number of features were systematically explored. Following the analysis, the completeness threshold for features selection was finally chosen to be 80% as basen on the data quality literature, threshold of 0.8 could be chosen to indicate good quality (Salati et al., 2011).

The selected set of features can be categorized based on their temporal nature into the following,

1. **Features that don't change over time**: From variables collected at admission and discharge
2. **Features that change over time**: From variables collected during each day of the entire ICU stay

All the features that were selected for model training are shown in Table 1 below. After selecting the features, to handle the missingness of temporal features in the dataset, filling the missing values based on the last known value known as backward filling was used (Lin et al., 2019). Then to cope with the different length of ICU stay records of patients, the data of each patient's last ICU stay was padded up to the maximum length of ICU records of patients in the dataset.

Table 1: Variables selected for model training

Feature	*Feature that doesn't change over time*	*Feature that changes over time*	*Feature type*
Age (years)	✔		Discrete
Gender	✔		Nominal
Source of admission	✔		Nominal
ICU length of stay (days)	✔		Discrete
Hospital length of stay prior to ICU admission (days)	✔		Discrete
Admission type	✔		Nominal
Emergency surgery	✔		Nominal
Reason for admission	✔		Nominal
Mechanical ventilation		✔	Nominal
Cardiovascular support		✔	Nominal
GCS score	✔		Discrete
APACHE II score	✔		Discrete
Systolic blood pressure (mmHg)	✔		Continuous
Diastolic blood pressure (mmHg)	✔		Continuous
Heart rate (b/min)	✔		Continuous
Antibiotic use		✔	Nominal
Renal replacement therapy		✔	Nominal
Temperature (C)		✔	Continuous
Respiratory rate (/min)		✔	Continuous
Fraction inspired oxygen (/)		✔	Continuous
After hours discharge	✔		Nominal

3.3 Feature extraction for reason for admission

Due to the high cardinality of the ***reason for admission*** variable, applying one-hot encoding would lead to high dimensional sparse data, and would lead to a computational inefficiency (Kassymzhomart Kunanbayev, Islambek Temirbek and Amin Zollanvari, 2021). Additionally, ordinal encoding would introduce an ordinal nature among unordered categorical options and it can affect the performance of the predictive model (Kassymzhomart Kunanbayev, Islambek Temirbek and Amin Zollanvari, 2021). To address these challenges, the approach proposed in the study by Guo and Berkhahn (2016) named entity embedding for categorical variables was employed for that variable, resulting in each unique option of the variable being represented as a 128-dimensional vector.

3.4 Statistical features for baseline models

To compare the LSTM model to the classical machine learning methods, the statistical features from the entire ICU stay were extracted. As linear regression approach has been widely used in ICU readmission prediction (Lin et al., 2019), slopes and intercepts of the linear regression line (m and c in $y = mx + c$) were extracted as additional features to characterize the linear variation of temporal continuous data such as temperature. For the categorical data such as mechanical ventilation, as in the study by Lin et al. (2019), the majority value and number of times the values change over the entire ICU stay were extracted. Figure 2 illustrates the extracted statistical features from both continuous and categorical features for the baseline models.

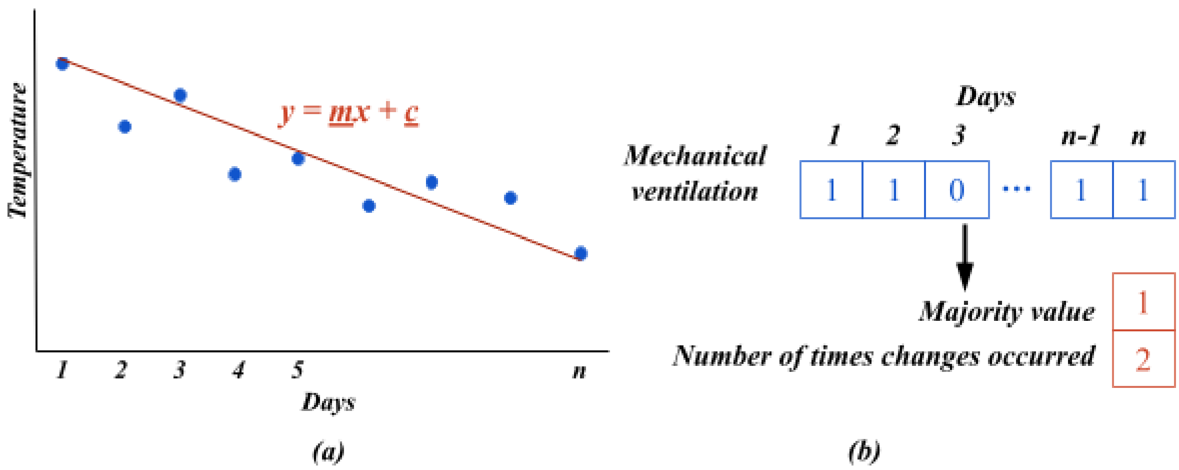

Figure 2: Statistical features for baseline models

Figure (a) shows the statistical features extraction for continuous variable and Figure (b) shows the statistical features extraction for categorical variable

3.5 Models

3.5.1 Baseline models

As developed in the study by Lin et al. (2019), Logistic regression with L2 regularization penalty and Random Forest were chosen and trained as baseline models of the study.

3.5.2 LSTM model

A recurrent neural network (RNN) with long-short term memory (LSTM), was chosen for its capability to uncover intricate patterns within sequential data and to make predictions based on them, especially for clinical measurements where there can be lags of unknown duration and missing values in a time series (Lin et al., 2019). Figure 3 shows the architecture of the implemented LSTM model. Two LSTM layers, followed by a dense decision layer with one output neuron activated by a sigmoid function were built.

3.6 Model training

The final dataset was divided into training (75%), validation (12.5%), and testing (12.5%) sets. Since the dataset is highly imbalanced (306 samples for positive class and 10436 samples for negative class), to address this, all the models were trained with and without class-weights (cost-sensitive training). Then the model was trained on the training set, validated on the validation set to fine-tune hyperparameters, and its performance was then assessed on the testing set.

To optimize the model's performance, hyperparameter tuning was conducted. Adjustments to the number of layers, units per layer, learning rates, and activation functions were explored using a hyperparameter search space to achieve the best possible area under the receiver operating characteristic curve (AUC-ROC) score. After training the model using the training set, its performance was evaluated on the test set using AUC-ROC score and compared to baseline models' performance.

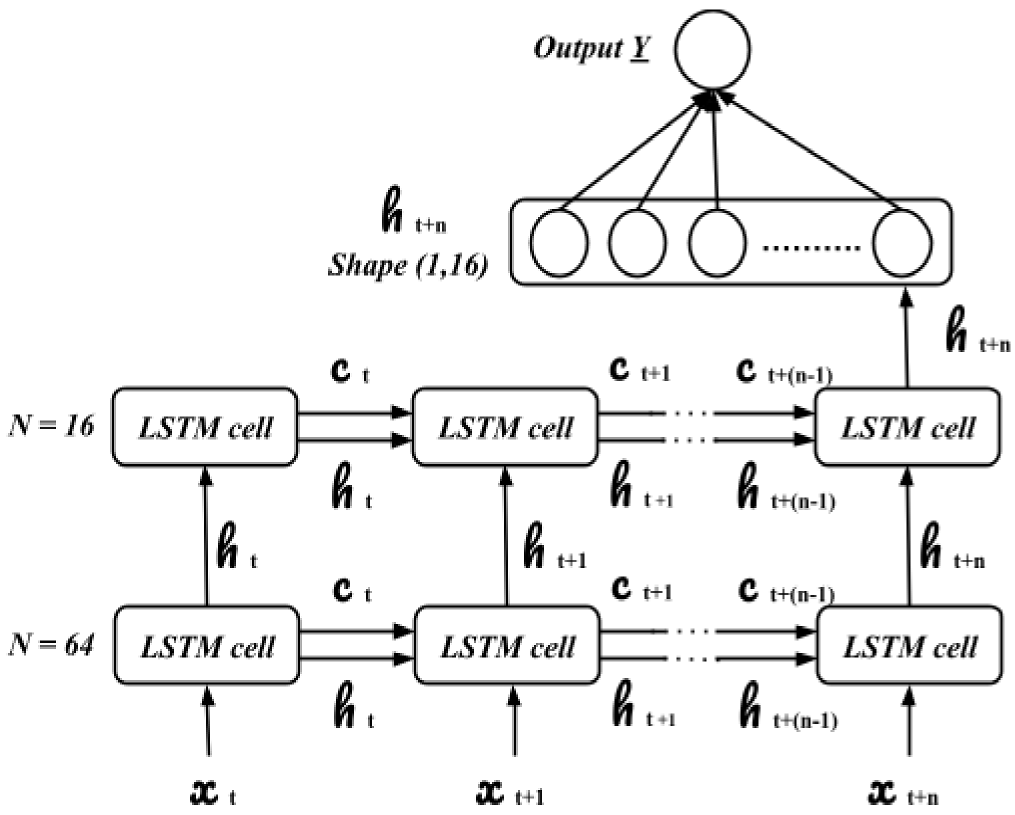

Figure 3: LSTM architecture

It Includes 2 LSTM layers, and a dense layer. *N* is the number of neurons in each LSTM layer. h_t and C_t are hidden activation states and cell memory states. x_t is the input time series, and $\underline{Y}$ is the predicted output.

3.7 Feature ablation test for feature importance

The feature ablation test on the features that change over time, was conducted to better understand the underlying logic of the proposed model. By following the approach in the study by Lin et al. (2019), each time by changing only one feature of features that change over time, in the test set to its normal value in humans, the AUC-ROC score of that test set was recorded.

The reduction in AUC-ROC score for each feature's normal imputation was calculated by comparing it with the AUC-ROC score of the original test data and if there was no reduction, a value of 0 was assigned. Then all the features that change over time were ranked according to the reduction in the AUC-ROC score. Figure 5 shows the results of the feature ablation test based on the reduction of AUC-ROC score of the prediction results after replacing the original feature with its normal value.

Since this normal imputation approach can't be applied to all the features that don't vary over time, the importance of them wasn't evaluated using the feature ablation test with normal imputation technique.

3.8 External validation

To evaluate the impact of the number of patient records and data points on model performance, an external validation was conducted using the MIMIC-III dataset (Johnson et al., 2016). The MIMIC-III dataset, which contains electronic health records (EHRs) from intensive care unit (ICU) patients, is a widely used, de-identified and publicly available collection of medical records.

The same patient inclusion and exclusion criteria described in section 3.1 were applied to the MIMIC-III dataset to prepare the final dataset. All patients and their corresponding ICU stay records in the final dataset were then labeled based on their post-ICU discharge outcomes. Patients who were discharged to the ward and not readmitted to the ICU within 48 hours were labeled as negative (not readmitted), while those readmitted to the ICU within 48 hours were labeled as positive (readmitted). In total, 26,179 patients were classified as negative, and 857 patients were classified as positive.

Given the differences in data collection between the original dataset (LMIC dataset that contains one vital sign measurement per day) and MIMIC-III dataset (which contains higher temporal resolution data), two distinct datasets were prepared from final MIMIC-III dataset,

1. **Daily MIMIC III (one vital sign per day)**: To replicate the data collection pattern in the LMIC dataset, the vital sign measurements in MIMIC-III were aggregated to one value per day by selecting the worst recorded value for each day. This dataset allows the impact of patient count and data resolution on model performance in a lower-resolution setting, mirroring the LMIC data collection environment, to be validated.

2. **Hourly MIMIC III (hourly vital signs per day)**: In contrast, the original hourly vital sign measurements from MIMIC-III were used to create a higher temporal resolution dataset. The effect of having more points in the temporal data on model performance was evaluated, particularly in comparison to the LMIC dataset with its lower frequency of vital sign data collection.

To ensure consistency in feature representation, the similar set of features included in the LMIC dataset was selected from the MIMIC-III dataset. The selected set of features can be categorized into the following,

1. Demographics (features that don't vary with time)
2. Chart events (features that vary with time)
3. ICD-9 embeddings (features that don't vary with time)

After selecting the features, the same approach outlined earlier in section 3.2 for handling missing values and varying ICU stay lengths was applied. Specifically, backward filling was used to address missing temporal data, and patient stays were padded to ensure uniformity in the length of ICU stays, consistent with the method previously described.

For each of the two datasets (daily MIMIC-III, and hourly MIMIC-III), a LSTM model was implemented, and then trained, validated and tested. By comparing model performance across these three datasets, the impact of number of features and data resolution on readmission prediction was quantified, and the robustness of the model across different settings was validated.

3.9 Further exploration on Transformer-based model

In addition to the primary focus on the LSTM model for ICU readmission prediction, the performance of a Transformer-based model was also explored on the same datasets to assess its suitability for this task. While LSTMs are well-suited for sequential data processing, recent advancements in deep learning have highlighted the efficacy of transformer models in capturing long-range dependencies and handling sequential data more effectively in some cases (Reza et al., 2022).

To ensure a comprehensive comparison between the two architectures, the transformer models were trained, validated and tested on the three datasets described earlier,

- LMIC dataset
- Daily MIMIC III
- Hourly MIMIC III

The Transformer model was adapted for time-series prediction by employing a sequence-to-sequence framework, with multi self-attention mechanisms utilized to capture the temporal patterns of vital signs across different time resolutions. The high-level architecture of the transformer-based model implemented is shown in Figure 4.

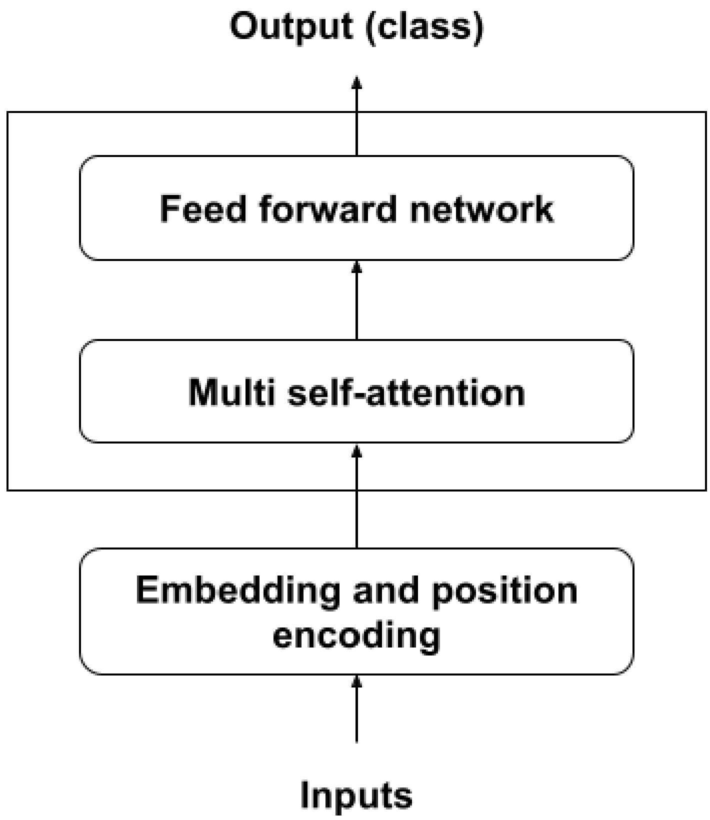

Figure 4: High-level architecture of transformer-based model

Transformer-based models were compared to LSTM models, in addition to the baseline models, to obtain insights into which model was more appropriate for readmission prediction under different data conditions. Additionally, it was evaluated whether modern transformer models offered any significant advantages over traditional recurrent neural networks, such as LSTMs, in this context.

3.10 Ethics

Since this study uses only the data previously collected (secondary data) de-identified data, and no individual can be traced back using this data, this study does not need ethical approval. Thus a waiver of ethical review was obtained from Oxford Tropical Research Ethics Committee (OxTREC) (Appendix A).

CHAPTER 4

EVALUATION AND RESULTS

4.1 Patient characteristics

During the study period, 10742 ICU admissions were eligible for readmission and early ICU readmission rate was 2.85% (n=306) (Figure 1). Patient characteristics of the whole eligible population are described in detail in Table 2. Readmitted admissions had older patients compared to non-readmitted admissions (68 IQR 41-72 vs. 58 IQR 52-78). Though there is a small percentage difference in the organ support during the first 24 hours between the readmitted and non-readmitted admissions, a higher difference can be observed in the organ support during the entire ICU stay. Also a higher APACHE II severity score was reported in readmitted admissions (13 IQR 8.25-19 vs. 10 IQR 6-15).

Table 2: Comparison of case-mix, organ support and severity of illness between non-readmitted and readmitted admissions

Variable	*Non-readmitted admissions*	*Readmitted admissions*
Patients N (%)	10436 (97.15)	306 (2.85)
Gender N (%)		
Male	5865 (56.2)	182 (59.48)
Female	4561 (43.7)	124 (40.52)
Intersex	10 (0.1)	0 (0)
Age (years) Median (IQR)	58 (41 - 72)	68 (52 - 78)
Admission type N (%)		
Non-operative	8232 (78.88)	244 (79.74)
Post-operative	2204 (21.12)	62 (20.26)
Elective surgery	1571 (71.28)	46 (74.19)
Emergency surgery	633 (28.72)	16 (25.81)
Source of admission N (%)		
Ward	1800 (17.25)	72 (23.53)
Emergency department	6197 (59.38)	171 (55.88)
ICU/HDU	634 (6.08)	19 (6.21)
Operating theater	1762 (16.88)	44 (14.38)
Missing	43 (0.41)	0 (0)
Reason for admission		
Pneumonia	1494 (14.32)	46 (15.03)
Emphysema/bronchitis	573 (5.49)	31 (10.13)
Sepsis, other	517 (4.95)	12 (3.92)
CVA, cerebrovascular accident/stroke	422 (4.04)	12 (3.92)
GI medical, other	391 (3.75)	15 (4.9)

Organ support during first 24 hours of admission		
Mechanical ventilation	1698 (16.27)	50 (16.34)
Cardiovascular support	959 (9.19)	34 (11.11)
Renal replacement therapy	418 (4.01)	16 (5.23)
Antibiotics	7932 (76.01)	245 (80.07)
Organ support during ICU stay		
Mechanical ventilation	2273 (21.78)	89 (29.08)
Cardiovascular support	1636 (15.68)	63 (20.59)
Renal replacement therapy	686 (6.57)	40 (13.07)
Antibiotics	8714 (83.5)	277 (90.52)
Apache ii score* Median (IQR)	10 (6 - 15)	13 (8.25 - 19)

* Normal imputation was done to handle the missingness

4.2 Models

The LSTM model was trained on 75% and validated on 12.5% of the LMIC dataset. The training process was performed with and without a cost-sensitive approach to address the class imbalance issue. After the model was trained, to evaluate the model's performance on the unseen data, the trained model was tested on the rest of the 12.5% of the data. Figure 5 shows the obtained AUC curve of the LSTM model trained with a cost-sensitive approach for the train and test sets. The same approach was applied for the baseline models as well.

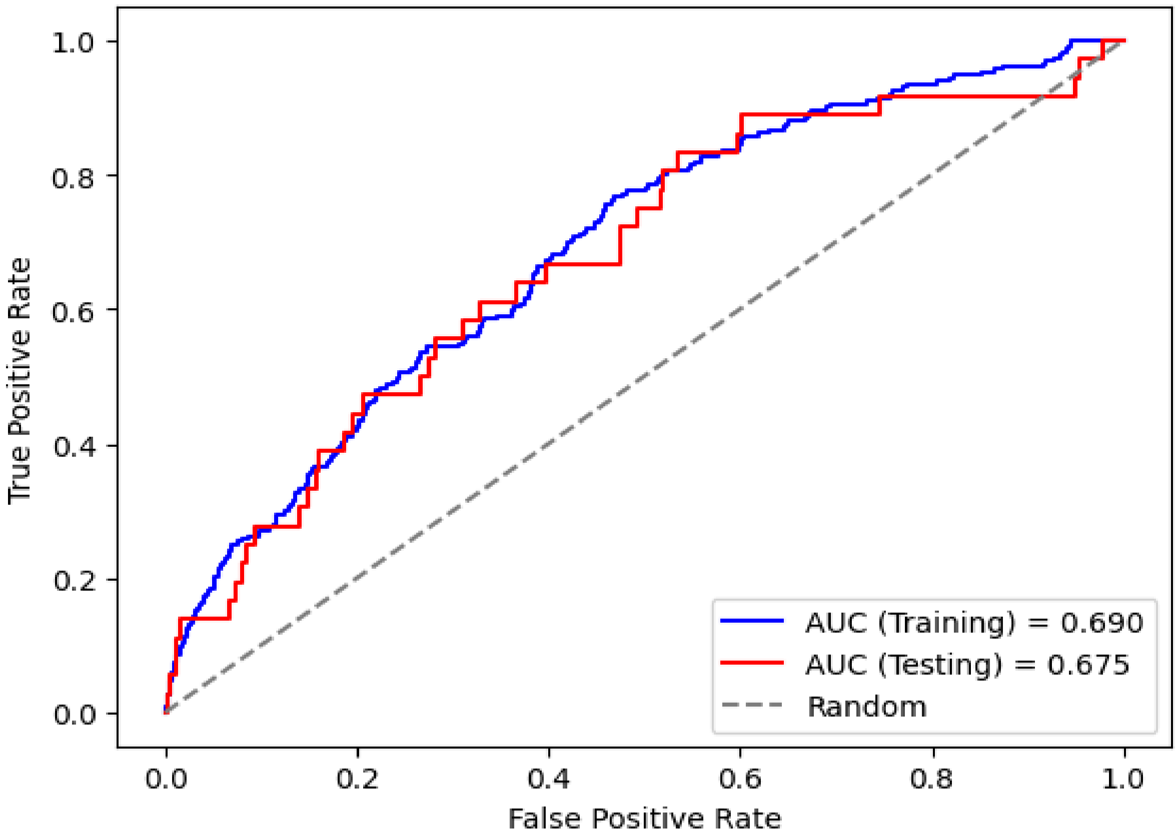

Figure 5: AUC curve of LSTM model for LMIC dataset

Both LSTM model and baseline models for the LMIC dataset achieved a slightly higher AUC score with cost-sensitive training (Table 3). Among baseline models, LR model with cost-sensitive training obtained a higher AUC score of 0.635 and LSTM with cost-sensitive training obtained the highest AUC score of 0.675 compared to the baseline models.

When evaluating the performance of this model by comparing to other deep learning models proposed based on the regions of HICs, this LSTM model has a lesser AUC score. Table 4 compares the AUC score of the LSTM model to previously proposed deep learning models.

Table 3: AUC score comparison between baseline models and LSTM

Model	*Cost-sensitive training*	*AUC score*
	Baseline models	
LR	No	0.612
LR	Yes	0.635
RF	No	0.568
RF	Yes	0.573
	Deep learning model	
LSTM	No	0.649
LSTM	Yes	**0.675**

Table 4: Results comparison to previous deep learning models developed for patients in HICs

Study by	*Model*	*Dataset*	*AUC score*
Lin et al., 2019	LSTM	MIMIC-III	0.791
Barbieri et al., 2020	RNN	MIMIC-III	0.739
Ashfaq et al., 2019	LSTM	CHF patients hospitalized in sweden	0.770
This study	LSTM	A registry of CCAA data repository	0.675

4.3 Ranking of features

Figure 6 visualizes the ranking of features that vary with time resulting in the feature ablation test. Y-axis shows the reduction in ROC score after imputing normal values for each feature that varies with time and X-axis shows the corresponding feature.

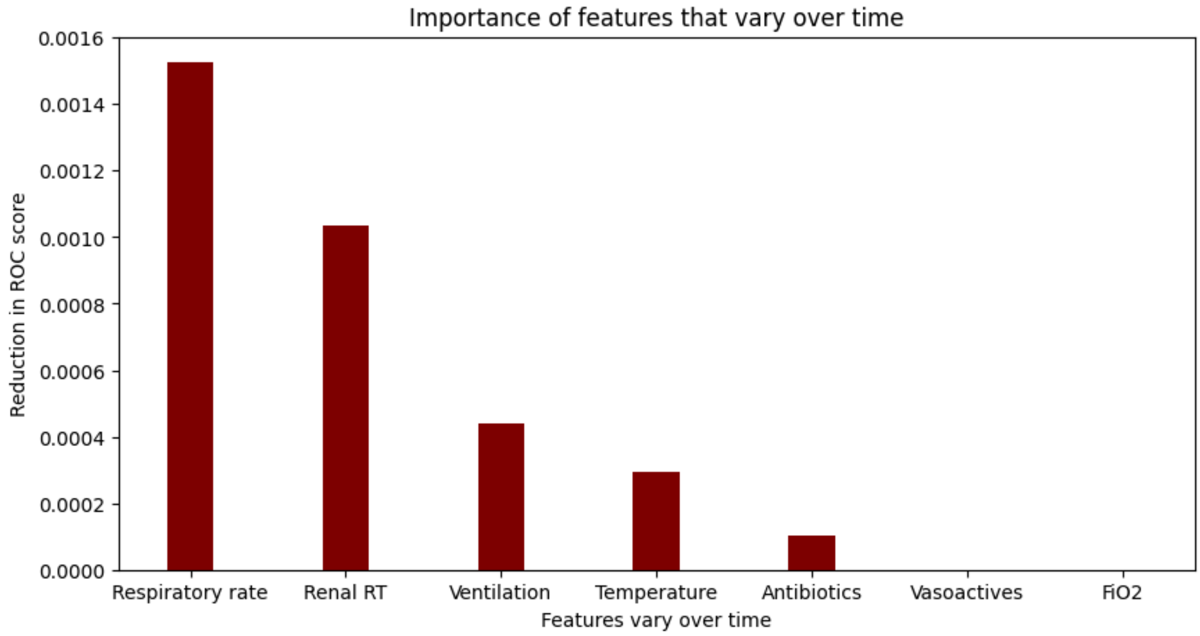

Figure 6: Feature ranking chart

4.4 External validation

For the external validation of high resolution of temporal data on the performance of the model, since the LSTM model and the baseline models trained on the LMIC dataset achieved higher accuracy with cost-sensitive training, the LSTM models for both daily MIMIC III and hourly MIMIC III datasets were trained with a cost-sensitive approach as well.

The LSTM models for the LMIC dataset and daily MIMIC III dataset (Figure 7) obtained approximately similar AUC scores, while the LSTM model for hourly MIMIC III dataset achieved higher AUC score than the models for the other two datasets (Table 5).

Table 5: AUC score of the LSTM models on different datasets

Dataset	*AUC score*
LMIC dataset	0.675
Daily MIMIC III	0.681
Hourly MIMIC III	**0.712**

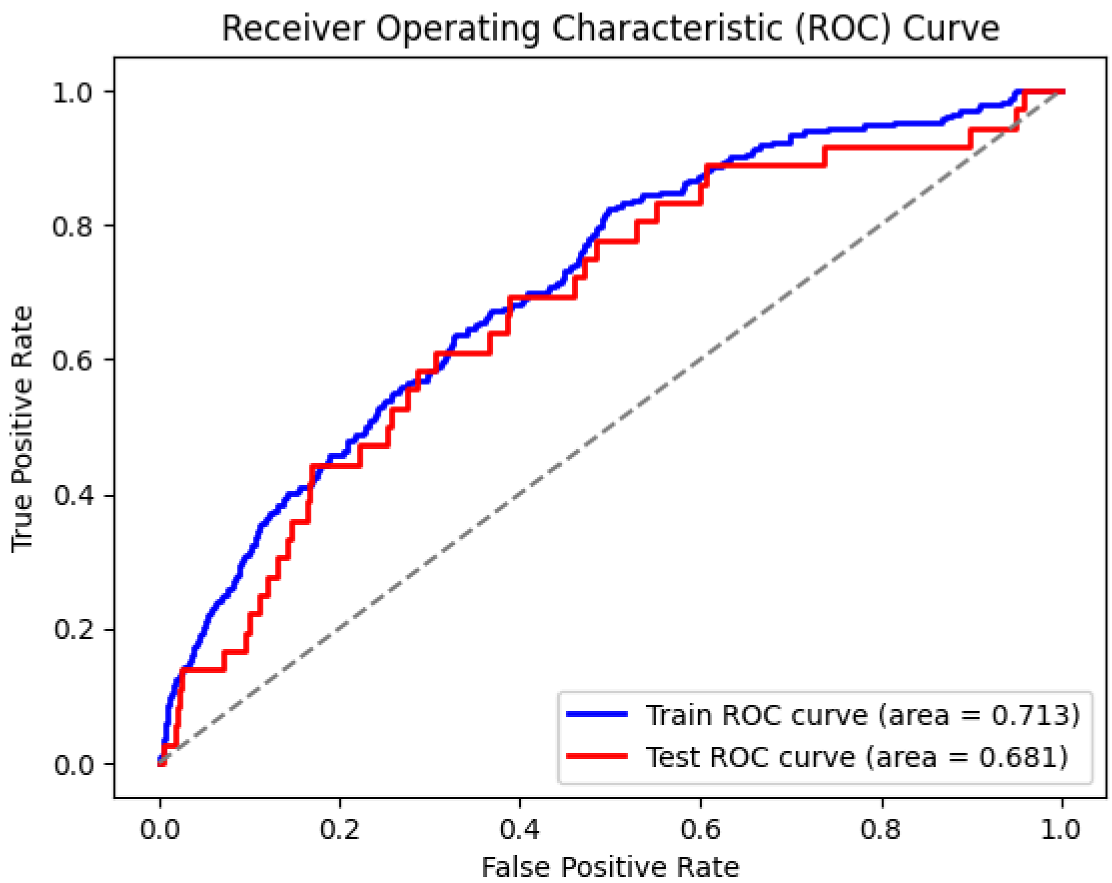

Figure 7: AUC curve of LSTM model for daily MIMIC III dataset

4.5 Comparison with transformer-based model

Figure 8 shows the AUC score obtained for training and test data of LMIC dataset using the transformer-based model. For the LMIC dataset, compared to the transformer based model, the LSTM model obtained higher accuracy than it. Additionally, though the result is the same for daily MIMIC III dataset, it achieved higher accuracy for hourly MIMIC III dataset (Table 6). The reason could be that transformer-based models are designed to capture the dependency of longer sequence, the number of datapoints in both the LMIC dataset and daily MIMIC III might not be enough to train the transformer model.

Table 6: Comparison of AUC scores between the LSTM models and the transformer-based models on different datasets

Dataset	*AUC score*	
	LSTM model	***Transformer-based model***
LMIC dataset	0.675	0.659
Daily MIMIC III	0.681	0.662
Hourly MIMIC III	0.712	0.709

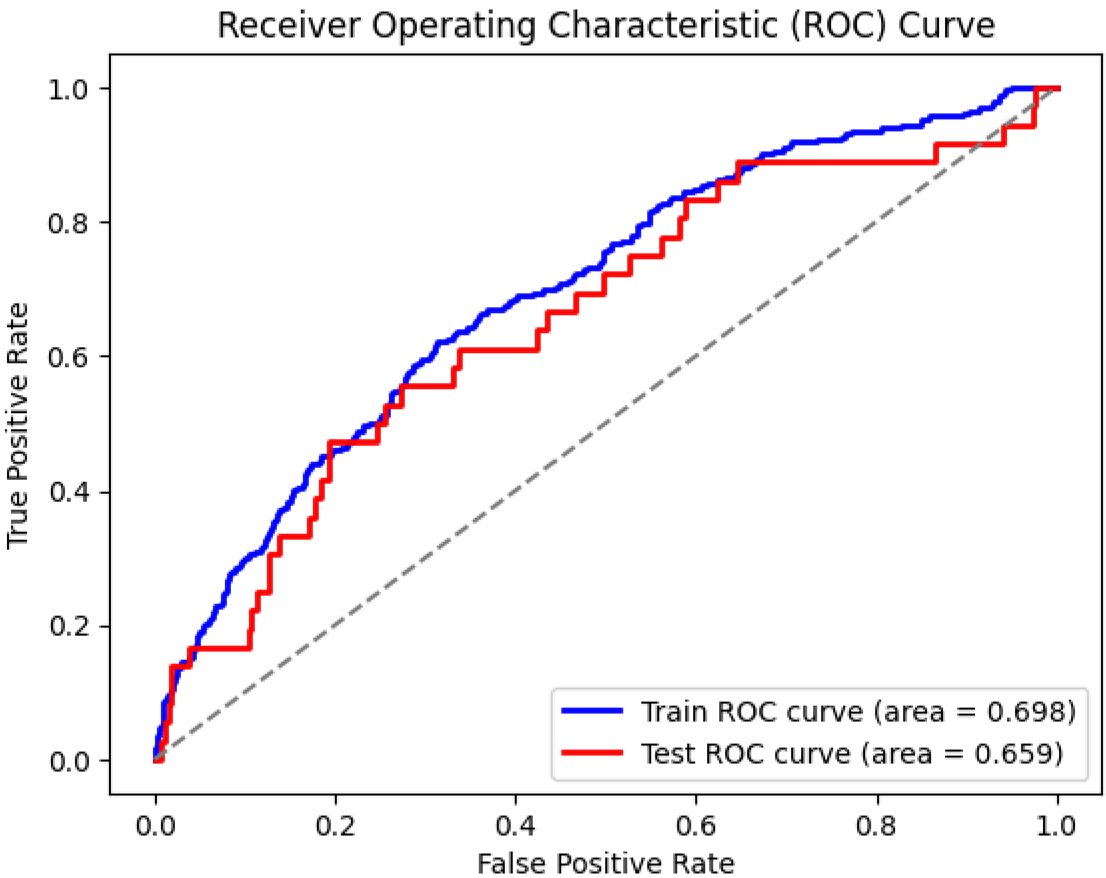

Figure 8: AUC curve of transformer-based model for LMIC dataset

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this study the LSTM model outperformed the chosen baseline models (traditional machine learning approaches) for this dataset (LMIC dataset) from the ICUs of LMICs as similar to the study by Lin et al. (2019) based on data from ICUs of HICs. It further supports the statement that capturing the temporal data from patients' ICU stay would improve the predictive ability of the model. From the feature ablation test, it can be seen that respiratory rate is the important factor learned by the model followed by renal replacement therapy, ventilation, temperature and antibiotics, while vasoactive and fraction inspired oxygen don't have significant influence on the model's decision. However, using this feature ablation test, the important features that influenced the prediction of one given sample can not be explained.

When compared to the proposed deep learning models for HICs (Lin et al., 2019; Barbieri et al., 2020; Ashfaq et al., 2019), the AUC-ROC score of the proposed model is comparatively lower. However, external validation on the daily MIMIC III dataset, with a similar feature representation, suggests that the limited number of selected features due to the constraints of the LMIC setting, negatively impacts model performance. Furthermore, the external validation on the hourly MIMIC III dataset highlights that having data with higher temporal resolution improves the model's ability to capture nuanced patient trends, which, in turn, could lead to better predictive performance in settings with more granular data collection.

Additionally, the results of the exploration on the transformer-based models indicate that both the transformer and LSTM models performed similarly when more temporal data was available, likely due to the transformer's ability to capture long-range dependencies effectively through its self-attention mechanism (Wahid et al., 2023). However, when less temporal data was available, the LSTM outperformed the transformer. This suggests that LSTMs, which are designed to process sequential data in a stepwise manner, may be better suited to learning from smaller, more sparse time series datasets (Reza et al., 2022).

In addition to the limited number of features and temporal data points, another limitation of the study is generalizing the model for LMICs. As the model was built on the subset of the data repository of LMICs, the model can not be generalized for all LMICs. This will be addressed in the future work.

5.2 Future work

Because of the time limitation, the importance of features that don't vary over time wasn't explored and this will be addressed in the future work. The importance of all the features in the model's prediction, irrespective of their temporal nature i.e. vary over time or not, would be explored using appropriate techniques and the features that influence on the prediction of a given sample will also be investigated by applying relevant techniques.

In addition to that, to generalize the model for LMICs, the representative ICU patients' data for LMICs from CCAA data repository will be extracted and its representativeness will be validated. Then an ICU readmission prediction model will be developed on the representative ICU patients' data to generalize the model for LMICs.

APPENDIX

Appendix A: Ethical waiver letter

REFERENCES

1. Elliott, M., Worrall-Carter, L. and Page, K. (2014). Intensive care readmission: A contemporary review of the literature. *Intensive and Critical Care Nursing*, [online] 30(3), pp.121–137. doi:https://doi.org/10.1016/j.iccn.2013.10.005.

2. Brown, S.E.S., Ratcliffe, S.J. and Halpern, S.D. (2013). An Empirical Derivation of the Optimal Time Interval for Defining ICU Readmissions. *Medical Care*, 51(8), pp.706–714. doi:https://doi.org/10.1097/mlr.0b013e318293c2fa.

3. Aryal, D., Hem Raj Paneru, Sabin Koirala, Sushil Khanal, Subhash Prasad Acharya, Karki, A., Dilanthi Gamage Done, Rashan Haniffa, Beane, A. and Salluh, J.I. (2023). Incidence, risk and impact of ICU readmission on patient outcomes and resource utilisation in tertiary level ICUs in Nepal: A cohort study. *Wellcome open research*, 7, pp.272–272. doi:https://doi.org/10.12688/wellcomeopenres.18381.2.

4. Lin, W.-T., Chen, W.-L., Chao, C.-M. and Lai, C.-C. (2018). The outcomes and prognostic factors of the patients with unplanned intensive care unit readmissions. *Medicine*, 97(26), p.e11124. doi:https://doi.org/10.1097/md.0000000000011124.

5. Jo, Y.S., Lee, Y.J., Park, J.S., Yoon, H.I., Lee, J.H., Lee, C.-T. and Cho, Y.-J. (2015). Readmission to Medical Intensive Care Units: Risk Factors and Prediction. *Yonsei Medical Journal*, 56(2), p.543. doi:https://doi.org/10.3349/ymj.2015.56.2.543.

6. Miranda, J.J., Kinra, S., Casas, J.P., Davey Smith, G. and Ebrahim, S. (2008). Non-communicable diseases in low- and middle-income countries: context, determinants and health policy. *Tropical Medicine & International Health*, [online] 13(10), pp.1225–1234. doi:https://doi.org/10.1111/j.1365-3156.2008.02116.x.

7. Lin, Y.-W., Zhou, Y., Faghri, F., Shaw, M.J. and Campbell, R.H. (2019). Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLOS ONE*, 14(7), p.e0218942. doi: https://doi.org/10.1371/journal.pone.0218942.

8. Cascarano, A., Jordi Mur-Petit, Jerónimo Hernández-González, Camacho, M., Nina, Polyxeni Gkontra, Chadeau‑Hyam, M., Jordi Vitrià and Karim Lekadir (2023). Machine and deep learning for longitudinal biomedical data: a review of methods and applications. *Artificial Intelligence Review*. doi:https://doi.org/10.1007/s10462-023-10561-w.

9. Rojas, J.C., Carey, K.A., Edelson, D.P., Venable, L.R., Howell, M.D. and Churpek, M.M. (2018). Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data. *Annals of the American Thoracic Society*, 15(7), pp.846–853. doi:https://doi.org/10.1513/annalsats.201710-787oc.

10. Pakbin, A., Rafi, P., Hurley, N., Schulz, W., Harlan Krumholz, M. and Bobak Mortazavi, J. (2018). *Prediction of ICU Readmissions Using Data at Patient Discharge*. [online] IEEE Xplore. doi:https://doi.org/10.1109/EMBC.2018.8513181.

11. Veloso, R., Portela, F., Santos, M.F., Silva, Á., Rua, F., Abelha, A. and Machado, J. (2014). A Clustering Approach for Predicting Readmissions in Intensive Medicine. *Procedia Technology*, 16, pp.1307–1316. doi:https://doi.org/10.1016/j.protcy.2014.10.147.

12. Barbieri, S., Kemp, J., Perez-Concha, O., Kotwal, S., Gallagher, M., Ritchie, A. and Jorm, L. (2020). Benchmarking Deep Learning Architectures for Predicting Readmission to the ICU and Describing Patients-at-Risk. *Scientific Reports*, [online] 10(1). doi:https://doi.org/10.1038/s41598-020-58053-z.

13. Ashfaq, A., Sant’Anna, A., Lingman, M. and Nowaczyk, S. (2019). Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics*, 97, p.103256. doi:https://doi.org/10.1016/j.jbi.2019.103256.

14. Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z. and Azim, M.A. (2022). Transfer learning: a friendly introduction. Journal of Big Data, 9(1). doi:https://doi.org/10.1186/s40537-022-00652-w.

15. Amorim, J.P., Abreu, P.H., Reyes, M. and Santos, J. (2020). Interpretability vs. Complexity: The Friction in Deep Neural Networks. doi:https://doi.org/10.1109/ijcnn48605.2020.9206800.

16. Brown, S.E.S., Ratcliffe, S.J. and Halpern, S.D. (2015). Assessing the Utility of ICU Readmissions as a Quality Metric. *Chest*, 147(3), pp.626–636. doi:https://doi.org/10.1378/chest.14-1060.

17. Salati, M., Brunelli, A., Dahan, M., Rocco, G., Dirk Van Raemdonck and Varela, G. (2011). Task-independent metrics to assess the data quality of medical registries using the European Society of Thoracic Surgeons (ESTS) Database. *European Journal of Cardio-Thoracic Surgery*, 40(1), pp.91–98. doi:https://doi.org/10.1016/j.ejcts.2010.11.004.

18. Kassymzhomart Kunanbayev, Islambek Temirbek and Amin Zollanvari (2021). Complex Encoding. doi:https://doi.org/10.1109/ijcnn52387.2021.9534094.

19. Guo, C. and Berkhahn, F. (2016). Entity Embeddings of Categorical Variables. *ArXiv*. [online] Available at: https://www.semanticscholar.org/paper/Entity-Embeddings-of-Categorical-Variables-Guo-Berkhahn/f9408afe434ab6ea5e852d92d5454063815a8685 [Accessed 3 Feb. 2024].

20. Johnson, A., Pollard, T., Shen, L., Lehman, L.-W., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. and Mark, R. (2016). OPEN SUBJECT CATEGORIES Background & Summary. *MIMIC-III, a Freely Accessible Critical Care Database*. [online] doi:https://doi.org/10.1038/sdata.2016.35.

21. Reza, S., Ferreira, M.C., Machado, J.J.M. and Tavares, J.M.R.S. (2022). A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert Systems with Applications*, 202, p.117275. doi:https://doi.org/10.1016/j.eswa.2022.117275.

22. Wahid, A., Yahya, M., Breslin, J.G. and Muhammad Ali Intizar (2023). Self-Attention Transformer-Based Architecture for Remaining Useful Life Estimation of Complex Machines. *Procedia Computer Science*, 217, pp.456–464. doi:https://doi.org/10.1016/j.procs.2022.12.241.

Oxford Tropical Research Ethics Committee

University of Oxford
Research Services, Research Governance Ethics & Assurance
Boundary Brook House, Churchill Drive, Oxford OX3 7GB
Tel. +44 (0)1865 (2)82106
E-mail: oxtrec@admin.ox.ac.uk

UNIVERSITY OF OXFORD

Ms Fathima Fazla
Data Scientist, NICS-MORU
fazla@nicslk.com

18 December 2023

By email

Dear Ms Fazla

Study title: Prediction of ICU readmissions using LSTM in low and middle income countries (LMICs)

PI: Fathima Fazla

Thank you for your email of 14 December 2023.

I can confirm that the above study is exempt from ethical review by the Oxford Tropical Research Ethics Committee (OxTREC) because analysis will be entirely of anonymised, previously collected data.

Yours sincerely

DocuSigned by:

BA168DF4624B463...

Karen Melham
Sponsorship and Ethics Lead

for
Research Ethics Manager, OxTREC