2023

Development of a Hospital Discharge Planning System Augmented with a Neural Clinical Decision Support Engine

David Mulqueen

Technological University Dublin, Ireland

Development of a Hospital Discharge Planning System augmented with a Neural Clinical Decision Support Engine

David Mulqueen

A dissertation submitted in partial fulfilment of the requirements of

Dublin Institute of Technology for the degree of

M.Sc. in Computing (Stream)

2/3/2023

Declaration

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Stream), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: David Mulqueen

Date:

Digitally signed by David Mulqueen
Date: 2023.03.01 14:28:33 Z

Abstract

The process of discharging patients from a tertiary care hospital, is one of the key activities to ensure the efficient and effective operation of a hospital. However, the decision to discharge a patient from a hospital is complex, as it requires multiple interactions with nurses, family, consultants, health information records and doctors, which can be very time consuming and prone to error.

This thesis descries how a neural network based Clinical Decision Support system can be developed, to help in the decision making process and dramatically reduce the time and effort in running the discharge process in a hospital.

A neural network model was developed using approximately one year of historical snapshots of inpatient information from a 180 bed hospital in Ireland. The model used a diverse set of inputs, which included a variety of patient metrics such as temperature, heart rate and length of stay etc, but also included more complex free text inputs such as patient progress notes and doctor's reports.

The model developed from these inputs was able to predict a patient's discharge with an F1 macro score of .71, but more importantly was able to quickly identify the vast majority of patient who have the potential to be discharged, which is typically a small subset of the inpatients. This decision support system may then present the small subset of potential discharges to the clinical discharge team, which can help guide the process. This is far more effective than having the discharge team analyse all inpatient's details or relying on human input to prompt the discharge process. This also allows for concentrated efforts on patients where the discharge decision may be questionable and higher levels of patient support are required.

The clinical decision support system developed can assess all inpatients (180) in

less than a minute versus the many hours that are required if competed manually by clinicians. This model is transferable and scaleable to larger institutions and can cater for all hospital specialities.

This model is currently being trialed in the selected institution with further feedback and model updates being fed back into the system to improve the model. Future enhancements may include more advanced text modeling, as hardware limitations limited the more advanced model capabilities.

Keywords: neural network, patient discharge, TensorFlow, Keras, F1 macro, model, hospital, patient

Acknowledgments

I would like to take this opportunity to thank my supervisor Dr. Robert Ross for his advice and guidance during the writing of my Thesis. His advice and encouragement at key stages of the development, helped immeasurably in the avoidance of some rabbit holes of thought and lengthy investigations.

I would also like to thank my wife. Her encouragement, support and inspiring guidance at every stage of my masters, was critical to achieving my outcomes.

Contents

List of Figures

List of Tables

List of Acronyms

HIS	Hospital Information System
Inpatient	A patient who stays overnight in a health care facility
GPU	Graphics Processing Unit
CRISPDM	CRoss Industry Standard Process for Data Mining
BERT	Bidirectional Encoder Representations from Transformers
TF-IDF	Term Frequency–Inverse Document Frequency
DPIA	Data Protection Impact Assessment,
ICD	International Statistical Classification of Diseases and Related Health Problems,
ATC	Anatomical Therapeutic Chemical,
HSE	Irish Health Care Executive,
EPR	Electronic Patient Record,
DAN	Deep Averaging Network
BoW	Bag of Words

Chapter 1

Introduction

1.1 Background

Hospital management and clinicians are in a constant struggle to balance the time treating each patient's condition appropriately and distributing healthcare resources to those most in need (Eigner & Bodendorf, 2020). Optimal patient flow is one of the key metrics hospitals use to ensure that patient safety goals, patient satisfaction and resources are being used effectively. If patients must wait a single extra day in a valuable hospital inpatient bed, because the resources were not in place to discharge the patient, it can dramatically affect the operational capacity of the healthcare institution.

The optimal flow is not only ensuring that patients are discharged promptly, but also ensuring that the patient does not get re-admitted due to a complication associated with the inpatient stay. There can be numerous factors which could affect a re-admission including co-morbidities (van Walraven, Jennings, & Forster, 2012), social factors (Al-Qudah, 2020), clinical issues and effective discharge management. (Al-Qudah, 2020).

Providing clinicians with appropriate electronic tools, has shown to improve the discharge decision-making process when tools are presented in an appropriate manner and when there is buy-in from all the teams involved (Graham, James, & Spertus, 2018).

This study has investigate the development of a supervised machine learning model using historical electronic patient data, to accurately predict future discharge of patients from a tertiary care hospital in Ireland.

This study is based around the broad scope of ACM 2012 of Health Information System and Probabilistic Computing, which has been studied extensively in a variety of publications which have proposed similar studies(Al-Qudah, 2020)(Benbassat & Taragin, 2013)(Goncalves-Bradley, Lannin, Clemson, Cameron, & Shepperd, 2016)(Graham et al., 2018)(Donzé, Aujesky, Williams, & Schnipper, 2013).

1.2 Research Project

There are many reasons why hospital administration might find discharge prediction attractive. Hospital overcrowding is a common problem, with adverse consequences for both the quality of patient care and for healthcare costs, with shorter lengths of stay have been associated with reductions in the total cost of a hospital admission (Clancy, 2009). Overcrowding has been associated with decreased patient satisfaction, as well as a higher risk of in-hospital complications and mortality (Clements et al., 2008). Overcrowding in healthcare is neither a short-term or a local Irish problem, there are more elderly people than ever before which puts higher demands on healthcare in every country (McDermid & Bagshaw, 2011).

Providing clinicians with a discharge tool that can help reduce the total time required to make an informed discharge decision that represents the minimum risk to patients, is the overall objective of this system. Deploying an Artificial Intelligence system used as a decision-support tool can also compensate for the different levels of knowledge and experience of the discharge personnel, by objectively and individually analysing patient data, and highlight clinically relevant correlations that the physicians may not recognise.

The discharge tool's machine learning model can also help the overall discharge decision-making process, by identifying high-risk patients and dramatically reducing the time required to make patient discharge risk assessments. The tool, would give the added benefit of providing high risk discharges with extra resources and interventions and discharge planning can be adapted accordingly.

1.3 Research Setting

This system development was completed in Beacon hospital in Ireland, which is a private tertiary care facility with over 180 inpatient beds, 8 operating theatres, 2 catheterisation labs, 4 endoscopy suites, over 1600 healthcare professionals and upwards of 300 consultants.

Beacon Hospital operates as a full-service acute hospital. Departments include, Cancer Care, Cardiology, Cardiothoracic Surgery, Emergency Medicine, Endocrinology, Endoscopy, ENT, General Medicine and Surgery, Intensive Care, Neurology & Neurosurgery, Orthopaedics, Physiotherapy, Paediatric Care (medical & surgical). While paediatric services are offered, the hospital in-patients would mainly consist of adults (aged 18 or older).

Beacon Hospital treated over 207,000 patients in 2022 across all areas of medicine and surgery. The hospital regularly runs near full capacity and is in a constant state of growth to cater for the need of patients.

Beacon Hospital uses technology at the core of all patient care, which provides a rich pool of data for the purposes of research. All the data for this study was gathered at this facility from the period 2021 to 2022.

Full ethics approval was obtained from the Beacon Hospital Research Ethics Committee on the 10th Sept 2022. Patient consent was not required a no patient identifiers were use in the development of this system.

1.4 Research Objectives

The purpose of the research is to develop a Hospital Discharge Planning System augmented with a neural network based Clinical Decision Support learning engine.

The high level major steps required to achieve this objective are outlined below:

- Explore the literature with respect to the development of a clinical decision support application and use of neural network models in the patient discharge process;
- Gather a dataset from historical patient discharges in the hospital;
- Perform cleaning, feature scaling and feature engineering on the dataset;
- Use the dataset to develop a neural network model, which accurately reflects the data;
- Analyse the results of the model in terms of the overall objective of deploying a clinical decision support system.

As the development and investigation of data progressed, it was found that the text features in the dataset would be pivotal in improving the performance of the models. To this end, more focus was given to researching methods and techniques involved in encoding unstructured clinical text features.

All the implementation steps will be performed in a combination of Python language and C# languages. Multiple extensive libraries were used in the development including:

- Tensorflow / Keras - open-source software library for neural network model implementation and deployment,
- Scikity-Learn -is an open machine learning library featuring various classification, regression and clustering algorithms,
- imblearn - package offering a number of re-sampling techniques commonly used in data sets showing strong between-class imbalance,
- Huggingface - a library of state of the art pre-trained models which were used in parallel with Tensorflow.

1.5 Research Methodologies

This research uses the CRISP-DM methodology as outlined in figure 1.1, to step through each of the critical milestones which need to be met, in developing an understanding of the data mining investigation. It should be noted that deployment and feedback on the system will not be in scope of this thesis.

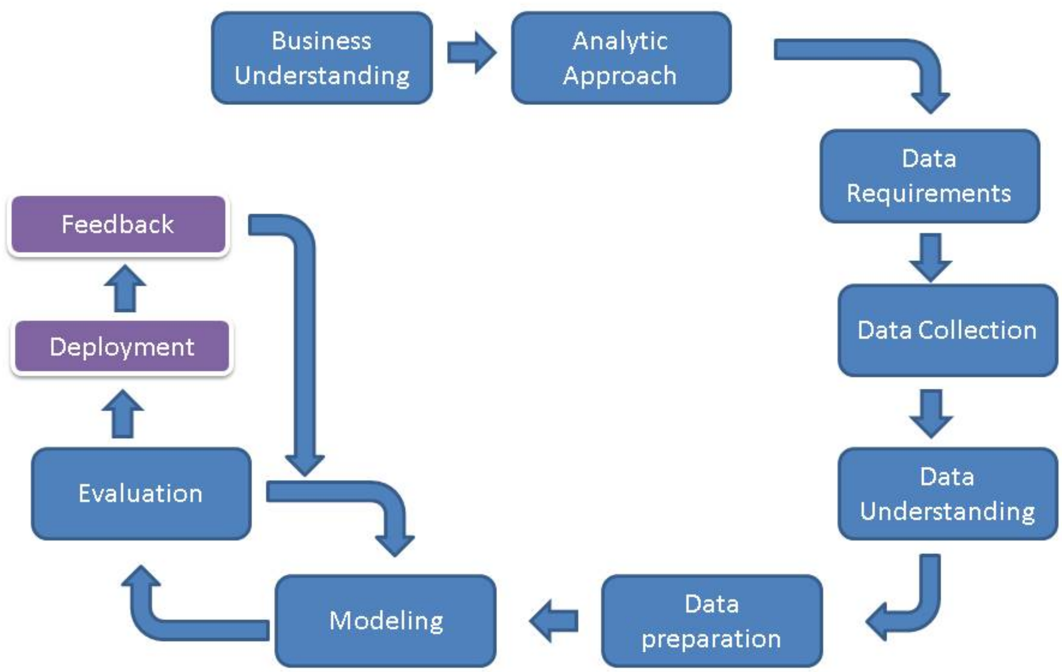


Figure 1.1: CRISPDM

A breakdown of each of these CRISP-DM steps is outlined as below:

1.5.1 Business Understanding & Problem Definition

The primary goal of the research is to create an AI augmented Discharge Planning system that can be deployed in the hospital. To deploy the system into an operational environment, a high level understanding of the current process, the people involved and technology in use must be achieved. This milestone will outline how this understanding was achieved.

1.5.2 Data Requirements

This milestone outlines the steps required to understand how data requirements are defined, which directly relate to the business understanding and the problem definition.

1.5.3 Data Collection

Once the requirements have been defined, the technical process outlines how this data was gathered for the various systems around the hospital.

1.5.4 Data understanding

Once this data is gathered, it is crucial that this data is analysed to ensure that:

- The correct data has been gathered
- Outliers have been identified
- Structure and variability of the data is understood
- The format and data types are identified.

1.5.5 Data Preparation

In this milestone the data preparation steps are outlined which may include steps such as:

- Removing outliers where appropriate
- Correcting spellings
- Standardisation of the data ranges
- Encoding and tokenization of data.

1.5.6 Modeling

Once the data is ready for modeling, a number of modeling techniques are used to evaluate whether a model can be developed which can accurately predict the discharge of a patient.

1.5.7 Evaluation

To assess if a model has the capability to delivery a system which is advantageous to the hospital, a number of metrics are investigated and detailed to determine the viability of the system.

1.5.8 Deployment

Typically once a viable system is found, it is deployed into the operating environment. To help limit the time for this development, this stage is considered to be out of scope

1.5.9 Feedback

Once the system is deployed, there should be feedback from both end users and operational data, which can help improve the system. This stage is also considered out of scope for this dissertation.

1.6 Scope and Limitations

Currently, there are a number of assumptions that will need to be made to ensure that that an effective model can be developed. The data collected, may include patients who have been discharged but may need to return to the hospital, this could be classed as a discharge in error, and not something that should be learned by the models. To this end, only discharged patients, who do not present to the hospital again within 72 hours, are deemed correct learnable discharges. Unfortunately patients may not return to the same hospital for readmission and this this data on readmission will not be accessible.

It should also be noted that the data will be limited to a single 180 inpatient bed hospital, but does provide a diverse set of clinical departments, which is representative of a typical tertiary care hospital. This diverse set of clinical departments also includes patient which are palliative care patients, this may cause outliers in acquired data and may need to be excluded. While the models developed may need to be limited to clinical speciality, in line with the study outcomes completed by Levin et al, this will be avoided to reduce the complexity of the developed system (Levin et al., 2020).

The size and complexity of models were limited to the hardware available during the study and development of this system in the hospital. While this limitation could have been avoided with the use of cloud based systems, this was not permitted due to confidentiality requirements of patient data.

The Deployment of this system into a live environment and the associated feedback which would be obtained from such a deployment, will be considered out of scope of this thesis. This scope limitation was introduced to help limit the amount of time required for development, but these steps will be implemented as part of a planned live deployment and future work.

1.7 Document Outline

The layout of the dissertation is outlined as shown in the list below, with six chapters in total, each divided into sections:

- Chapter 2 (Literature Review) gives an overview of the current state-of-the-art literature relating to discharge prediction, clinical decision support systems, neural network models and information extraction from unstructured clinical text. The gaps in the existing research will be highlighted to propose the research question for this project.

- Chapter 3 (Data Analysis) will elaborate on the process of data collection, cleaning and pre-processing. A complete understanding of the data to be used will also be outlined.

- Chapter 4 (Methodology) details the practical implementation of models including the context in which they were developed.

- Chapter 5 (Results & Discussion) will provide the results for the models developed and a detailed analysis of the model outputs. This chapter will also elaborate the strengths and weaknesses of the research.

- Chapter 6 (Conclusion) will summarize the findings of the research undertaken during the dissertation including problem definition, critical analysis of the design, developments completed, evaluation of the results and scope for further research.

Chapter 2

Literature Review

In an effort to ensure the research conducted uses the knowledge, experience and the wealth of research data surrounding this subject, a literature review has been completed to highlight the specific knowledge on which a system can be designed.

The literature review will be split into three distinct areas where research was appropriate, given the type of data involved in the overall project scope. These sections are patient discharge process, discharge support tools and data mining unstructured text. A summary will then be detailed highlighting the key learnings which can derived from this literature review, which can be used to make key design decisions.

2.1 Patient Discharge Process

The patient discharge process can be a highly faceted process, which may be specific to a patient's condition and procedures performed on the patient. Numerous clinical considerations must be analysed to make a discharge decision including drug regimen, conscious state of patient, wound state, recent lab results, current patient telemetry and recent nursing progress notes etc. But there are also numerous non-clinical considerations which also can skew any clinical decision such as patients home life (do they live alone?), time line on follow up appointments, patient's insurance level of cover, patients ability to medicate correctly.

Several studies have shown that a good understanding of the current discharge process, including the matrix of personal communications, is key to a redesign of the process. It should be noted that the discharge process may be specific to the institution or clinical speciality. For example, the discharge process completed by a cancer patient may be very different to a patient with a broken hip (Gonçalves-Bradley & S, 2022).

The complexity of the discharge process is demonstrated in the standard HSE (Irish Health Service Executive) document called "A practical guide to discharge and transfer from hospital" (HSE, 2017). This comprehensive forty six page document outlines the nine steps required to process a discharge for each patient. Each step of the process has multiple different stakeholders and step specific logic trees, all which would require extensive interdisciplinary team communications.

To help with this communication and workflow, a modern digital orientated hospital will typically use a EPR (Electronic Paper Record) to store all patient related information. The EPR, also known as an electronic medical record (EMR), is a digital version of a patient's medical history, which includes information on the patient's medical conditions, diagnoses, medications, allergies, lab results, and other medical information. The EPR can be just one system from one vendor or can be a loose amalgamation of products from multiple vendors. A discharge tool will need to access the EPR to obtain patient information to make an informed discharge decision.

2.2 Discharge Support Tools

A discharge tool will typical take for the form of a clinical Decision Support System (CDSS) in a hospital environment. A CDSS is an application that analyzes data to help healthcare providers make decisions and improve patient care. A CDSS focuses on using knowledge management to get clinical advice based on multiple factors of patient-related data. Clinical decision support systems enable integrated workflows, provide assistance at the time of care and offer care plan recommendations (in this case discharge recommendation).

CDSS in a clinical environment improve health care process measures, as shown in the systematic review "Effect of clinical decision-support systems" in 2012 (Bright et al., 2012). In the 15,176 abstracts and 128 studies reviewed, both commercially and locally developed CDSSs improved the health care process measures related to performing a healthcare service.

In more recent systematic study in 2021 by Olakotan et al, decision support systems have been shown to cause alert overload which can increases the mental workload of clinicians and decreases situation awareness in the workplace (Olakotan & Yusof, 2021). Ensuring the the system designed, minimises the issues with alert overload, will need to be given specific attention when developing the discharge support system. To avoid some of these overload issues, successful implementation were shown to be dependent on the reliability and adaptability of the decision support system and involvement of key users in the implementation process(Klarenbeek, Schuurbiers-Siebers, van den Heuvel, Prokop, & Tummers, 2021).

Along with the discharge specific system challenges, there are also some more generic challenges that apply to the development of clinical decision support implementations. There is a definite need for any system to be able to support a fast paced and mobile patient journey as expertly summarised by "speed is everything" (Bates et al., 2003). Any system that is implemented which further slows the ability of clinical staff to complete their clinical duties, will be difficult to deploy and harder yet to justify. This also extends to the time commitments for staff to fully and appropriately test any system which is to be implemented. There is also a concern that ill-conceived and poorly-designed alerts can distract health care providers from other important aspects of a patient's care (Nishimura et al., 2015;2016;). Having the user interface for this discharge system form a core part of of the existing workflow process, may help with both speed and minimise the level of alerting that needs to be performed.

Due to the number and complexity of discharge variables such as drug regimen, wound state, recent lab results, current patient telemetry and recent nursing progress notes, making a discharge decision can be a complex process and can sometimes be made on the "gut feel" of an experienced clinician. To help with this discharge process, several manual tools have been deployed in the past such as the simple carry forward method deployed in a New England Medical Centre (McCoy, Pellegrini, & Perlis, 2018). This carry forward model is based a mean value calculated comparing discharge numbers with yesterday, a week ago on the same day and a year ago. While this process is simple it was found to improve performance.

A large systematic review of 71 studies identified some of the main characteristics which were common with successful deployment of clinical decision support type systems. These characteristics included integration as part of routine clinical work flow and diagnostic or therapeutic recommendations integrated into the system. also all pertinent data should be available to the user at the time and location of decision. (Barnes, Hamrock, Toerper, Siddiqui, & Levin, 2016).

Measuring the discharge support tool's success, is important to show its value in the clinical setting. The patient re-admission rate is a metric which has proven to be successful in papers in the Israel Journal of Health Policy Research and a literature-based estimate in the Journal of Evaluation in Clinical Practice (Benbassat & Taragin, 2013) (van Walraven et al., 2012). Both papers have shown that a measurable readmission rate decrease can be achieved by deploying a discharge tool like this.

The value of the discharge tool, can be given a monetary value by the percentage decrease of re-admissions and lowering patient average length of stay (van Walraven et al., 2012). This monetary value could be used to sell this project to management within the hospital.

A study in 2019 of 3328 in-patient discharges from a teaching hospital demonstrated the need for discharge support tools. During the study, 838 patients were identified, which were not discharged on the predicted day. The breakdown of these patients included 38.5% who had no clinical reason identified, 33.8% had delayed follow-up on patient progress by the care team, 16.9% who had care transition issues to home or facility and 10.8% had a patient or family request to remain in the hospital (Safavi et al., 2019). These types of delays in discharge may be quite typical for many hospitals and demonstrate that a significant problem exists and the current discharge process may be inadequate.

From the patients perspective, spending time as an inpatient in a hospital when there is no clinical need, does come with risks. Inpatients run the risk of hospital acquired infections, which can have significant long term health consequences (example MRSA, cDIF and CLABSI) and medication errors, which cost 7000 lives per year in the U.S. (Flynn, Liang, Dickson, Xie, & Suh, 2012) (Gould, 2006).

2.2.1 Machine Learning tools in the Discharge Process

While manual tools have been effectively deployed in a small number of hospitals, there is very little evidence of actual deployment of artificial intelligence to an operational role. Nevertheless, using automated machine learning tools has been experimented with in a number of institutions in the papers from the Annals of data science (Gramaje, Thabtah, Abdelhamid, & Ray, 2019) and in the description of a comprehensive development in the highly rated journal from BJM innovations (Levin et al., 2020). In general, the developments have been successful and achieved AUC of 0.70 in predicting the patient discharge on the same day.

The available metrics which are used to build a machine learning model, are generally extremely critical in determining its overall prediction capabilities (McCoy et al., 2018). As hospitals vary in their digital maturity levels, some patient metrics may not be available, which may hamper the development of highly rated models.

The common trend in the research was that the following metrics are key in predicting the discharge of a patient (McCoy et al., 2018) (Levin et al., 2020)(Gramaje et al., 2019).

- Hospital length of stay
- A metric on the number of labs which have been ordered
- Medication which a patient has been prescribed
- The medical discipline under which the patient is being treated
- Gender of the patient
- Maximum and minimum hear rate
- Mobility of the patient

- Whether the patient was admitted to ICU
- Lab test ordered
- Progress Notes
- Clinical notes.

Any system design should contain these metrics at minimum. Gathering this temporal data into a single coherent data set, which can be spread across multiple subsystems in an EPR, will be the first challenge but once the data has been assembled there will be more data feature specific challenges. While metrics like hospital length of stay (example 10 days), is simple to understand and straight forward to input into a neural network, the unstructured free text in progress notes or clinical notes will be far more difficult to process, to extract knowledge and pass to a neural network.

2.3 Data Mining Free Text

As some of these key metrics may contain free text (for example; progress notes and clinical notes), it would be prudent to ensure the latest techniques and information extraction methods would be used to derive appropriate discharge related information from this type of feature.

Below is an example extract of a nursing progress note for a patient that was discharged.

"RECEIVD FROM NIGHT DUTY STAFF , STABLE, ALERT AND ORIENTED, ON FASTING FOR TANGO FOR TAVI WORKUP AND PUT VITALS, STABLE"

This nursing progress note contains manufacturer product terms ("TANGO" - Manufacturer of Stress Test Monitor), clinical terms (TAVI - "Transcatheter aortic valve implantation"), bad spelling ("RECEIVD") and abbreviated English, which will all add to the complexity of deciphering meaning and context, when processing as an input for neural network. Outlined below is some of the key literature reviewed in the domain of clinical free text and neural network model development.

While numerous deep learning neural network architectures have been developed for the goal of knowledge extraction from medical records, papers from Rebane et al and Wei have demonstrated that the use of LSTM (Bidirectional) and RNN demonstrate significant improvement in performance, due to short term / long term memory capabilities in the algorithms (Rebane, Samsten, & Papapetrou, 2020) (Wei et al., 2019). The dataset utilised in this study consisted of information about diagnoses and medications for 1,314,646 patients obtained from the research infrastructure Swedish Health Record Research Bank. This record bank complies with ICD-10 (International Statistical Classification of Diseases and Related Health Problems) coding standards for health problems and ATC (Anatomical Therapeutic Chemical) coding standard for medications.

Key insights into free text can also be derived by extracting medical codes from free text. These medical codes are typically based on international medical coding UMLS Metathesaurus(Sung, Hsieh, & Hu, 2020) (Rebane et al., 2020). Unfortunately due to the proprietary nature of Irish medical condition coding and only for statistical purposes post discharge, these models may not be particularly useful (Murphy, Wiley, Clifton, & McDonagh, 2004)(Information & Authority, 2021).

There has been numerous efforts to determine the clinical status of patients, based on progress notes due to the insightful nursing and doctor's knowledge they can contain and regularity on which they are completed (at least one note per clinician shift)(Kapinski et al., 2018)(Guan et al., 2019) (Chen, Dredze, Weiner, & Kharrazi, 2019)(Toyabe, 2012). There were no examples identified of patient progress notes being used to determine the probability of discharge.

While generative models like ChatGPT and BioGPT are typically used to generate text, they also can be refined to allow for classification. BioGPT follows the Transformer language model, and is pre-trained on 15 million PubMed abstracts (Luo et al., 2022). Experiments demonstrate that BioGPT achieves 6% better performance compared with BioBERT based model.

BioLinkBERT is a transformer encoder (BERT-like) model pre-trained on a large corpus of Medical documents including PubMed. It improves on BERT by capturing document links such as hyperlinks and citation links to include knowledge that spans across multiple documents into its pre-training. Specifically, it was pretrained by feeding linked documents into the same language model context, besides a single document. BioLinkBERT can be used as a drop-in replacement for BERT. It achieves better performance for general language understanding tasks (e.g. text classification). In the very recent review on natural language processing published by Hall et al, a BioLinkBERT based model has shown to improve performance by 5% over a standard BERT based model and 2% over the domain specific BioBert model (Hall, Chang, & Jayne, 2022).

2.4 Summary and Implications

Several key leanings can be extracted from this review of available literature which includes:

1. The decision support system models should be specific to the patient's procedure or clinical specialisation in which the patient was treated, to maximise the model's accuracy.

2. The user interface of the clinical support system should form an integral part of the workflow process.

3. Some key metrics are required as input to a machine model to ensure adequate performance.

4. Clinical input in the design of the system is key to its success.

5. Patient progress notes from nurse and doctors would contain a wealth of information on the current status of the patient and could be highly relevant in any discharge decision,

6. Bert and the more domain specific BioBERT text representation based models, can provide high levels of accuracy than more traditional LSTM/CNN based text encoding models.

The papers reviewed did not detail data cleaning techniques used or any feature engineering which could be used to improve the accuracy. Some of the more advanced models and data preparation techniques could be deployed to significantly improve the accuracy.

An area of research which did not seem to be covered in existing research was the concept of using this tool to identify the “sickest” patients needing the most care. The machine learning tools will output the percentage probability of being discharged – with patients with the lowest probability being the “sickest” patients. Using this high-level assessment of a patient may prove useful in day-to-day patient management, if it can be proved clinically relevant.

Decision support tools for aiding the discharge process, have not been extensively researched in Ireland and no journals or articles were found which were in any way tailored to an Irish context. This is probably due to the limited deployment of hospital wide information systems in Public Hospitals in Ireland, which would be required as the basis for this research.

Chapter 3

Data Analysis

3.1 Introduction

This chapter will elaborate on the data used to develop the neural network, which is the core component of the hospital discharge planning system. This chapter loosely adheres to the CRISP-DM framework as shown in fig 1.1 to ensure all the appropriate data was gathered, analysed and prepared for modeling in subsequent chapters.

3.2 Business Understand of Data

It is important to understand the business process and all the characteristics of a hospital discharge decision, in order to help in specifying precisely what its current challenges are, how to evaluate these challenges and come up with a an appropriate solution.

In order to gain an understanding of the current hospital discharge process, the following information gathering activities were undertaken:

- Interview with Nursing Supervisor - person who has overall responsibility for managing the discharge process,
- Interview with admission / discharge management employees - the individuals work with nursing / administration / patients to manage the discharge process,

- Interview with inpatient nursing - this group are responsible for giving key insights on patients during the discharge process,
- Joined the "Bed Huddle" process on two occasions - the "Bed Huddle" process is a cross disciplinary team meeting, which is scheduled twice daily where discharges are discussed.

Through these information gathering processes a number of key information points were learned which included:

- The bed huddle process take place twice a day, where individual potential discharges are discussed. In light of this, it has been decided to snapshot a collection of data from the various HIS systems which should take place twice a day, to reflect this business process.
- Potential patient discharges are identified using inputs from Nursing / Doctors / Consultants and Family. This typically takes place over phone calls and personal visits. This can be a very time consuming process.
- Patients are often identified for discharge / non-discharge by physical and mental attitude of patients. These can be reflected in patient progress notes and doctor's reports. Patient progress notes should be included in the data snapshot process
- Pressure to discharge / not discharge an individual patient can come from many different and diverse sources such as: need of an incoming patient, family and social needs of a patient, known / unknown future inpatient requirements, number of theatre appointments or even day of the week.
- The current discharge process is very time consuming and may require multiple interactions with numerous individuals involved with patient care to make a decision on a single patient.

Based upon this process understanding, the available electronic information systems in the hospital and literature reviewed, the following data set elements outlined in table 3.1 were selected to analyses. All of these data items are available in a combination of HIS systems within the hospital, which would need to be gathered on a snapshot basis into an appropriate data warehouse

Attribute	**Description**	**Data Type**
Account Number	Unique ID for each Patient	string
SampleDT	Sample Date Time	Date Time
DaysSinceAdmitted	Number of days since patient admited	int
Age	Age of the patient when sample was taken	int
Sex	Sex of the patient	categorical
LOS	Current length of stay	int
Observations	Blood Pressure, Heart Rate, Respiratory Rate, Temperature and SPO2 of Patient	int
LabOrderCnt	Number of lab tests ordered split into high low and normal	int
MedOrderCnt	Medication Order count	int
NNText	Nursing progress notes in last 12 hrs	string
DoctorText	Doctors progress notes in last 12 hours	string
LabResults	White cell count Lactate Lab results	int
AdmitPriority	Method of patient admission	categorical
MaritalStatusID	Current martial Status	categorical
SepesisPositive	Sepsis Positive in last 12 hours	binary
EWS	National Early Warning System An metric based on HSE NEWS standard (Maguire et al., 2016)	int
Discharged	Patient Discharged (Target)	binary
Re-admitted	Patient Readmitted post discharge - used to filter data	binary

Table 3.1: List of Dataset Features

3.3 Data Warehousing

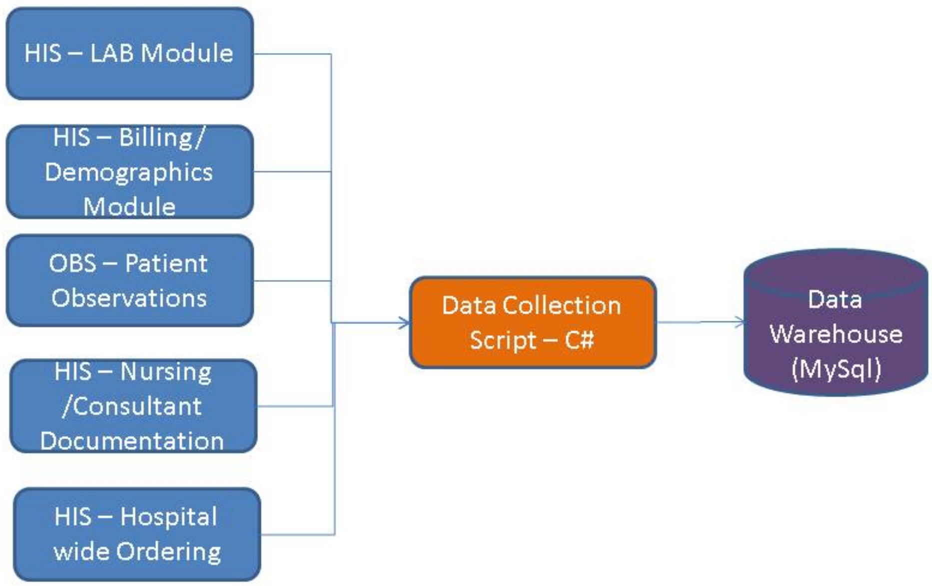


Figure 3.1: Data Collection Script

A c# Data Collection Script was developed, which started on a schedule twice a day, at the end of each nursing shift (8am to 8pm). At the scheduled time, the script would connect to each system pull a snapshot of each inpatients data via the appropriate APIs into a Data Warehouse. The Data Warehouse consisted of a single table with each row containing the full set of patient data captured at that time.

The schema for the data warehouse reflects the dataset outlined in table 3.1, with one row for each instance of data collected.

Health records are inherently multi-modal, containing a mix of provider's notes, order codes, laboratory data, images, vital signs, and increasingly also genomic sequencing and more. The partnered institution would have in the region of 100 systems and databases containing patient information, which is typical of a modern hospital. The multi-modality of Hospital Information System (HIS) is only going to grow, having jumped twenty fold from 2008 to 2015. No modality in isolation provides a complete picture of a person's health state. Analyzing pixel features of medical images frequently requires consulting structured records to interpret findings, so is important that any model developed should be multi-modal in nature and pull information from a variety of systems ("Survey: Deep Learning Concepts and Techniques for Electronic Health Record", 2018;2019;).

As typical in a hospital with multiple specialisms, in the partnered hospital, the multi modal patients data is split in multiple silos of information in various hospital information systems. The discharge dataset features needed to be pulled and amalgamated from all these systems into a single warehouse before being processed. The features selected for analysis from the various systems were derived from the business understanding of the data and literature reviews completed as part of this study.

There are three main groups of data which will require separate processing in table 3.1 patient metrics (temperature, BP etc), nursing notes and doctors notes.

The processing of patient metrics will follow the straight forward process of cleaning and normalisation before being processing by the model, but the progress notes will require extensive processing before it can be processed by any model. See table **??** for an outline of the processing used.

3.4 Data Analysis

Once this data was pushed into a single data warehouse, a data investigation was performed using Python code which included the following steps:

1. Statistical analysis: Basic statistics such as distribution, average, max, min, standard deviation, normalization, mode, the skewness of variables were computed for analysis,

2. Missing value analysis: Count and percentage count of the missing values of target and predictor variables were calculated,

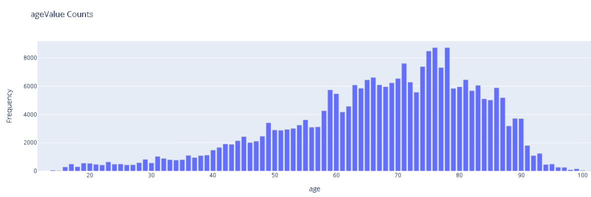
3. Outlier analysis: Outlier analysis were performed to find out the values lying

out of range e.g. temperature of patients,

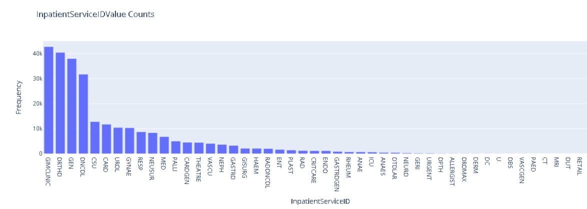
4. Exploratory data analysis: Frequency distribution plots of numeric and categorical variables with respect to the target variable,

5. Analysis of free text length, spelling, word counts, word clouds, use of language to get overall of the text content.

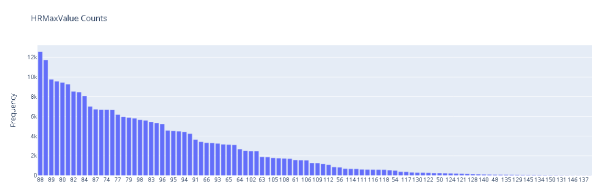
Once the data was collected, a data analysis was completed to ensure the quality of the data sent to the models would be adequate.



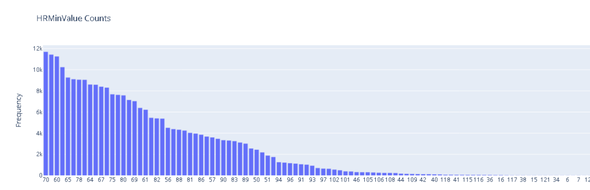
(a) Age



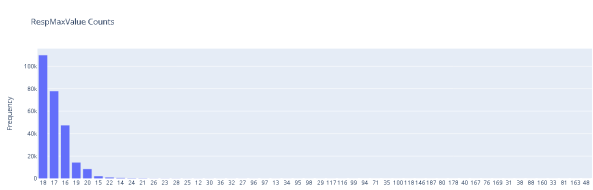
(b) Inpatient Service



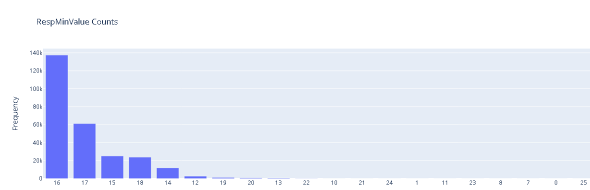
(c) Heart Rate Max



(d) Heart Rate Min

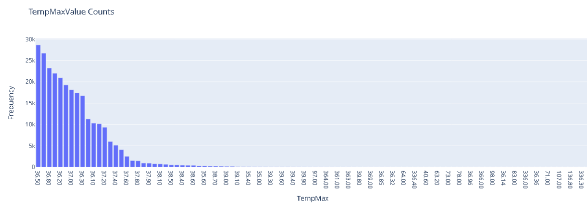


(e) Resp Rate Max

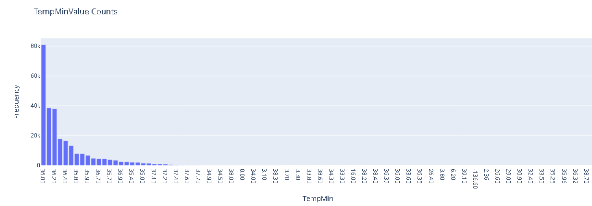


(f) Resp Rate Min

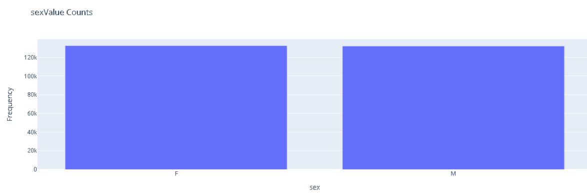
Figure 3.2: Data Distributions



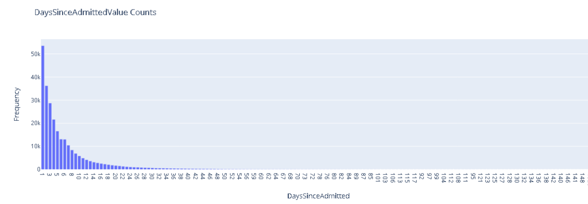
(a) Temp Max



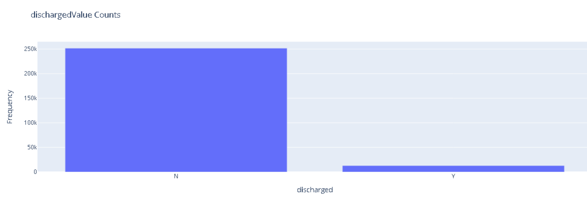
(b) Temp Min



(c) Sex



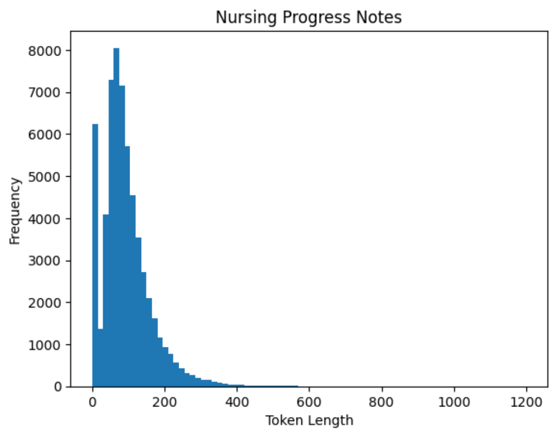
(d) Days Since Admitted



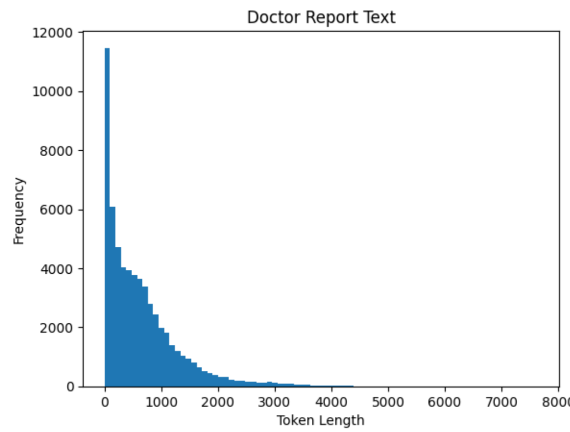
(e) Discharge Count

Figure 3.3: Data Distributions

A selection of the graphical analysis is outlined in figure 3.2 and 3.3 . Based on this understanding, a number of data cleaning activities were implemented. Special note should be given to the discharge count graph, which clearly shows that the data is highly imbalanced, which will need further consideration when performance metrics are used and re-balancing may be required. The Temperature Min graph also demonstrates that a level of data cleaning is required as temperature of 2.36 degree centigrade is not possible for living human. A good balance between male and female is maintained in the data set, which should minimise sex related biases being developed in the models.



(a) Nursing Progress Notes



(b) Doctors Report

Figure 3.4: Text Token Length

Figure 3.4 demonstrates there is a significant difference in the length of the text between nursing progress notes and doctor's notes. This is understandable as both serve a different purpose, whereby nursing notes are used to document the continuous progress of an in-patient and doctor's reports document a specific investigation and may contain full background history of a patient and diagnosis.

This dataset for each inpatient was captured every 12 hours from multiple hospital systems and sent to a data warehouse system. This data was captured for a period of 17 months from 2nd July 2021 to 1st of Nov 2022. This resulted in 264,558 data items collected. An overview of data gathering process is outlined below in Figure 3.1

3.5 Data Preparation

The below sections outline the data preparation that was completed to ensure data sent to the machine learning algorithms excluded outliers and errant data, which could effect performance of the models.

3.5.1 Data Cleaning

In total, there was 264,558 snapshots of raw data collected, but all this data was raw inputs from various patient information systems. These systems offer relatively poor user interface checking of data and allow end users (Nurses) to submit obviously errant data. For example, a maximum body temperature of a patient can be submitted as 375 degrees centigrade, which is obviously missing a decimal separator and should read 37.5.

To ensure only correct data is passed to the machine learning engines, a number of filters were applied to correct the data. These filters included:

- Max / Min temperature of patient restricted to 33 and 45 degrees respectively
- Max / Min Blood pressure was restricted to 20 and 250 bpm respectively
- Max / Min Hear Rate was restricted to 39 and 117 beat per minute respectively
- Only adults patients were included (to avoid any DPIA issues)
- Length of stay restricted to 90 and lower - outlier patients were excluded
- Removal of deceased patients as they will be marked as discharged
- Removal of any patients that were discharged against medical advice
- Removal of any patient that did not wait to be formally discharged and self discharged
- Removal of any patients that were re-admitted into the hospital under the same consultant within 7 days. These were considered to be bad discharge decisions.

In total 10,770 outlier/errant data elements were removed using the above filters. This represents 4% of the overall data collected, which is statically insignificant.

3.5.2 Data Imbalance

One of the largest problems with the dataset is that there is a severe imbalance of patients who are discharged, compared to those who are not discharged. On a daily basis, only in the region of 5% of patients get discharged. So out of 180 inpatient beds, only 9 patients will get discharged on a daily basis on average.

As the overall objective is to identify patients who should / should not be discharged and given that there equal negative consequences for miss classification in each category, (false negative/ false positives), there is no bias to favour false negatives over false positives and vice versa.

As the system will be tuned to reduce all miss classifications, the F1 score will be used to assess the performance of the data imbalance, rather than a number of false negatives or false positives. To evaluate the sampling techniques, the TFIDF model was used as it was fast to execute and required minimal changes to the model for each variation of technique.

To assess if any sampling techniques could be used to reduce the significance of the data imbalance, the following techniques outlined in table 3.2 were applied using the TFIDF model to see if performance could be improved.

Sampling Technique	**Max F1 Score after 10 Epocs**
Custom Weight Loss Function (Weighted Binary Cross Entropy)	.7089
No balancing	.7054
Using SMOTE	.6821
Over sample Minority Class	.6648
Under-sample majority class	.5651
Tensorflow Model Class Weight setting	.6866

Table 3.2: Data Imbalance Investigation

In line with papers from Attiga et al and Arya et al, we can see that re-sampling does not improve performance of the TFIDF model, and can have very significant decreases in performance in several instances (Attiga et al., 2018)(Arya & Sastry G, 2020). As the custom loss function (Weighted Binary Cross Entropy) demonstrated a minor improvement in performance, this will be used in all further model development.

As the data set is large (237,324), the 5.3% of samples that were discharged still provided a significant amount of samples of patients who were discharged (12,036). So even though we have a highly imbalanced data set, we do have have significant amount of samples of each class.

3.5.3 Correction of English / Spellings

During the data analysis, it was noted that there was a wide variation in the quality of the English which was used in the the Nursing Progress notes and Doctor's notes. The Doctor's progress notes were found to have 41% non-English words and the nursing notes were found to have 35% non-English.

Text encoding techniques, where the input words are matched against a learned dictionary for words (example BERT, Universal Sentence Encoder etc) would perform poorly due to English quality. It was therefore decided that the English would be corrected where possible, to help with performance of these models.

There were various reasons for this high degree of non-English words in reports including:

- Poor spelling, as the user interface for the hospital system did not have spell checker capability on the input screens
- Extensive use of abbreviations and jargon specific to the medical field and to the specific hospital
- Use of Latin medical terminology
- Use of Irish drug names

Spelling corrections, abbreviations replacement and Latin replacement script was written in c# to implement all the corrections. The script pulled text directly from the data warehouse database, the text was then corrected by the script and then inserted back into the same location. As part of this development, a customised dictionary was developed to help in the replacement of localised jargon. In total, 98,111 words were added to the default general English dictionary and were harvested from drug dictionaries, HSE standard documentation (abbreviations), local hospital policies and procedures and local place names (example 3rd floor ward).

This script was written in c# due the performance limitations of Python in analysing large sets of text and correcting spelling. A multi threaded c# application was able to perform the cleanup in hours versus days required for python.

The final deployment of this system into a live environment, will require that all notes will need pre-processing to correct spelling before a model is run. In a live environment, where processes are time sensitive, it is prudent that this process is efficient.

3.5.4 One-Hot Encoding: Categorical Variables

The warehouse dataset set has 4 categorical variables (sex, InpatientServiceId, Admit-ProirityID, Marrital StatusID) out of 47 feature variables. In order to build a neural network model, categorical variables need to be converted into numeric variables, as neural network models machine only work on binary or numeric variables. To overcome this problem, dummy variables were constructed out of categorical variables using the Pandas library in Python.

3.5.5 Normalisation and Scaling

All learning algorithms are known to provide dramatic predictions on unscaled or unstandardized feature inputs if they are not scaled. To ensure that all the feature values are on the same scale, normalization or standardization is a mandatory step to be carried out before proceeding to model building. Where appropriate all training / test data was normalised using the standard scalier algorithm to help avoid an scaling issues.

3.6 Data preparation

The data preparation phase covers all the activities involved in the transformation and cleaning of the data, to make it fit for use in the modeling phase. The missing values, noise and outliers present in the data identified during the data understanding phase will be removed where appropriate in data pre-processing.

As some of the major inputs to the learning engine was text, a series of prepossessing was applied to the text to help with the improving accuracy of the models. These prepossessing techniques included removing stop words, making all text lower case, tokenization of text depending on modeling technique, correcting spellings, applying word translations (Latin) and extending standard medical abbreviations.

3.7 Data Hypothesis

It is hypothesised that free text note features in the dataset will be a significant component in determining if a patient will be discharged. To this end, the following hypothesis will be explored as part of the data investigation:

"Can adding unstructured free text data to a set of patient observation metrics, help to improve a machine learning model's performance, when predicting probability of hospital in-patient discharge?"

In order to determine the hypothesis, if adding unstructured text data to a set of patient observation metrics helps to improve a machine learning model's performance when predicting probability of hospital in-patient discharge, the following tests will be completed:

- Tensorflow model will be built which only contains patient observations and results will be recorded
- The same model will be extended by adding encoded text from the patient progress notes. The TFIDF text encoding technique will be used due to its simplicity and speed of calculation.

- This model will then be extended again by adding encoded text from the Doctor's notes also using the TFIDF encoding technique.

The TFIDF based model will be used to assess the hypothesis, due its high F1 scores and its speed of convergence.

The performance of each model will be measured using the F1 macro metric, as the data is highly imbalanced. If performance improvements are seen with the addition of each text feature, the hypothesis will be considered to be proven.

3.8 Strength and Limitations

This section outlines the strength and limitations of the dataset collected. One major strength of this dataset is that all data will come from a current operational environment, which will reflect the typical problems associated with running a hospital information system in the 21st century.

The gathered data is also extensive with over a quarter of a million distinct snapshots of a patients status. This dataset represents the operation of the entire inpatient environment for the hospital for over an entire year and like all hospitals the data would also reflect the pressures on an inpatient environment with the change in season changes (e.g. flue, COVID, winter vomiting etc).

3.9 Summary

The exploratory data analysis is done on the patient discharge dataset to understand the distribution of predictor variables and their distribution. Significant data cleaning issues were found, which would affect performance of the models.

The dataset was found to have significant data balance issues with respect to the target variable "discharge". Investigations were completed to determine if this data balance could be corrected. A small performance increase was found when using a custom developed loss function.

Significant data quality was also found with the text inputs. Significant text corrections were required to ensure that text matching in pre-trained models would be effective.

Chapter 4

Methodology

This chapter outlines the details of the models developed, which is the core component of the neural decision support engine. The models use the output data which was prepared in chapter three, but each model will use this source data in a different manner depending on he requirement and unique attributes of each model.

4.1 Model Evaluation Metrics

Choosing an appropriate metric is challenging generally in applied machine learning, but is particularly difficult for this highly imbalanced classification. Most of the standard metrics that are widely used (Accuracy) assume a balanced class distribution and if applied to this classification problem could be misleading.

A metric which is independent of the number of training instances overall, would be a far more representative metric. The macro-F1 score, gives equal weights to each class. The macro-F1 metric evaluates the algorithm from a class standpoint: high macro-F1 values indicate that the algorithm has good performance on all the classes, whereas low macro-F1 values refers to poorly predicted classes.

All models developed in this section will use the macro-F1 score as the primary model performance indicator.

As there will be a tendency to over-fit, early stopping will be applied to validation macro-f1 score, as this is what we are trying to maximise (power of model to predict discharge). While it would also be prudent to monitor validation loss for early stopping, there may be situations where loss may be increasing along with macro-f1 score, but there may be some small level of over fit which may be helpful in the model, given the imbalance in the training data.

Although the F1 metric will provide a good insight into the performance of the model, it is not a perfect metric in terms of overall objective of providing a clinical decision support system which quickly identifies patients for discharge. The F1 metric does not give any insight into how the probability of discharge is distributed across all predictions. As the vast majority (95%) of patients are not discharged, the vast majority should have a 0% probability of discharge. If this is not the case, the system would be ineffective in quickly identifying discharge patients, as the user of the system would need to search through a large number of patients.

A new custom metric was developed which can be used along side the F1 metric, which gives insight into the number of errors which are in the 0 to .10 percent probability bin. This metric is call the Discharge Capability Factor (DCF). Appendix A.1 outlines the discharge capability factor's performance.

4.2 Model Data & Environment

In general for all models the data set was split between a training data set (80%) and a test data set (20%) to allow for comparison between models. This training / test split allowed for 189,859 items in training and 47,465 in the test data set, post data cleanup processing. Stratification and a static random seed was introduced to ensure there were sufficient classes of the target feature (discharge) in each data set.

All models were optimised using the Adam optimiser with the learning rate adjusted where required to allow for faster convergence. For example, a small learning rate (1e-5) was required in BERT / BioBERT models in line with the model's development guide.

It should be noted that specific limitations had to be applied to protect confidentiality of information including:

- Patient information could only be stored on hospital systems. Cloud storage and processing was not permitted,
- Due to the complexity of the models and and large size of data all models were run on GPU (Nvidia RTX 3060 Ti) to allow for more reasonable times to convergence. This hardware was limited to 8GB of RAM, which limited the size and complexity for the models developed,
- Models which incorporated multiple state-of-the-art modeling techniques were not possible (Logformer mixed with BioBERT) due to limited hardware.

4.3 Modeling

The main aim of this study is to develop a Hospital Discharge Planning system, augmented with a neural network based Clinical Decision Support learning engine. Based upon the research and current state of development, the Keras, Tensorflow and Huggingface PyTorch libraries will be used to provide the underlying neural network modeling capabilities.

As hypothesised the free text elements will be a major factors in the capability of the neural network to learn how to discharge patients discharge, so specific emphasis will be given to a number of advanced text encoding techniques to help with improving the performance.

From the literature review, it was found that there are numerous techniques and methods which can be used to extract useful information from the text using text representations. A text representation aims to numerically represent the unstructured text to make it computable by a mathematical model such as a neural network. It is hypothesised that the free text features in the data set will contain significant knowledge on the discharge of a patient, and these representations will be fundamental to improving the performance of the models.

There are numerous different techniques for representing text to a model with each having is own advantages and shortcomings. For example a simple technique of counting the number of times a word is used in a free text note (Bag of words) is simple to implemented, but ignores the context of a word in relation the other words in a sentence, which may be fundamental to the meaning.

As this development would take place in a clinical setting, a number of text encoding techniques were chosen for analysis basis on their bases for use in a clinical setting and experience and knowledge obtain from the literature review.

These encoding techniques will include:

- TFIDF Text Representations
- BERT
- BioBERT
- Keras FNet
- Google Universal Scentence Encoder
- Ensemble of BioBERT word encoder mixed with BioBERT scentence encoder

Each of these text encoding techniques along with the associated architecture is discussed below.

4.3.1 TF-IDF Text Representation Model

Term Frequency Inverse Document Frequency is a statistic that aims to better define how important a word is for a document, while also taking into account the relation to other documents from the same corpus. This technique was chosen as a baseline on which other methods could be assessed.

TF-IDF text encoding technique can be broken down into two parts TF (term frequency) and IDF (inverse document frequency).

Term frequency works by looking at the frequency of a particular term you are concerned with relative to the document. While there are multiple ways, of defining frequency of a word this work will simply use the number of times the word appears in a document (raw count).

Inverse document frequency looks at how common (or uncommon) a word is amongst the corpus. The reason we need IDF is to help correct for words like “of”, ”it”,”as”, “drug”, “patient”, etc. since they appear frequently in an English corpus. Thus by taking inverse document frequency, we can minimize the weighting of frequent terms while making infrequent terms have a higher impact.

The higher the TF-IDF score the more important or relevant the term is; as a term gets less relevant, its TF-IDF score will approach 0. TF-IDF is based on the bag-of-words (BoW) model, therefore it does not capture position in text, semantics, co-occurrences in different free text notes, but should serve as a good baseline on which to compare other text representations.

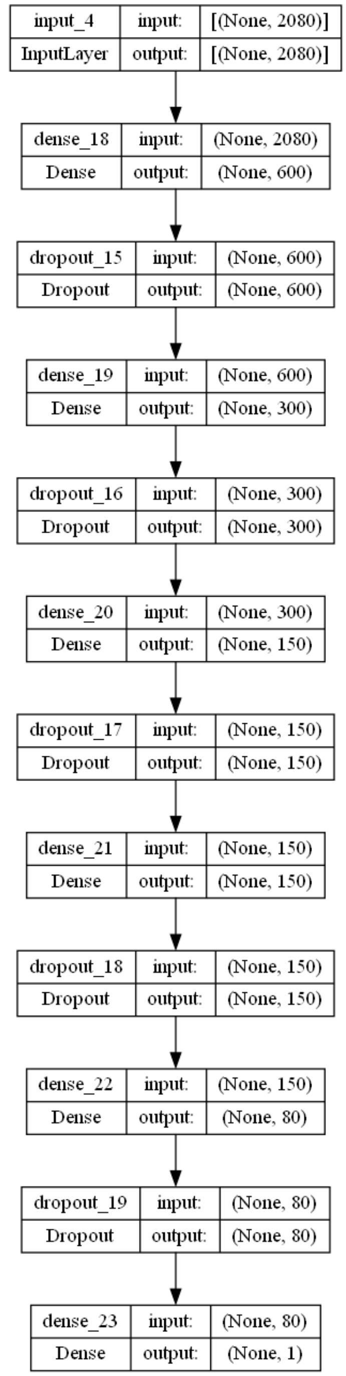


Figure 4.1: TFIDF Model

The architecture of the model that builds on the TF-IDF representation is quite simple, with all the word count frequencies and variables added as a single input into feed forward fully connected layers. In total the top 1000 words for each text document were used, as any more words did not improve performance. As the TFIDF representation produces a array of word counts, the observations data and counts were appended to produce a single input for the model. A high level of dropout and regularisation was implemented, to help stop the over-fitting on the training data. The output layer, as with all the models, will use a dense one node layer with sigmoid activation to represent the binary discharge decision of the patient.

4.3.2 BERT Text Representation Model

BERT (Bidirectional Encoder Representations from Transformers) is a recent (2018) paper published by researchers at Google AI Language. BERT's key technical innovation is applying the bidirectional training of Transformers, a popular attention model, to language modelling. A language model which is bidirectionally trained can have a deeper sense of language context and flow than single-direction language models.

This language model allows the encoding of words depending on their context, by using a huge corpus (Wikipedia and books) for training. The training of BERT is self-supervised i.e., it is done by masking random words from existing sentences, and having the model lean to guess the missing words. The architecture of BERT consists of multiple layers of transformers. The BERT model has 110 million parameters; the large BERT model has 340 million parameters. To limit the hardware requirements, a smaller model (distilroBERTa-base) was used, which has 82 million parameters. Devlin at al. made the implementation of the BERT model open source. But we will use the Python library transformers, which were made available by company HuggingFace (Devlin, Chang, Lee, & Toutanova, 2018). BERT is not domain specific as it was trained on generic Wikipedia and books, so may limited ability in encoding the characteristics of medical texts. The input of the BERT model is limited to 512 tokens so longer text will need to be truncated before being used. If critical information on discharge of the patient is truncated is may reduce the ability of the model to predict discharge.

Since the BERT model is pre-trained, it comes with a tokenizer, which we will use to tokenize our own input data. Fine tuning training of the BERT model were performed to allow for custom words and sequences in the data set to readjust the BERT model to help improve performance.

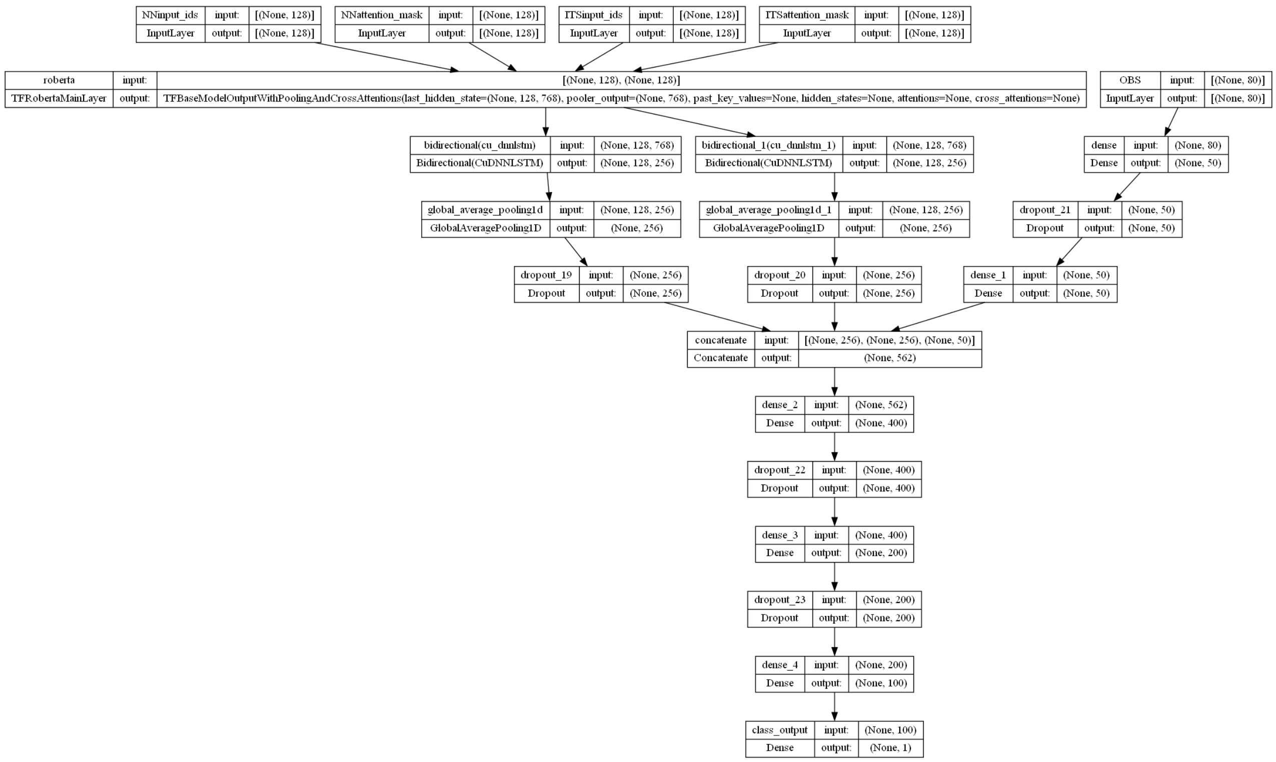


Figure 4.2: BERT Model

The tokenised text data (input_id and masks) was fed into the BERT encoder as two separate inputs along with a separate input for the non textual data. The tokenisation process was limited to 512 tokens, so text was truncated where appropriate to fit into this limitation. A bi-directional LSTM was used with the outputs of the BERT encoder and merged with the normalised data from the non textual data to derive the output. The output and layer is exactly the same as the TF-IDF output layer but the hidden layers at the head were customised to help improve performance of the model.

4.3.3 BioBERT Text Representation Model

BioBERT (Bi-directional Encoder Representations from Transformers for Biomedical Text Mining), is a domain-specific language representation model pre-trained on large-scale biomedical corpora. The BioBERT text transformer is based upon BERT, but has been trained on biomedical text including from PubMed, PMC and publicly available discharge summaries. This language domain pre-trained model, should contain the medical terms and phrases that are specific to the input text in the data warehouse and should lead to better outcomes.

According to the authors, BioBERT largely outperforms BERT and previous state-of-the-art language representation models in a variety of biomedical text-mining tasks, such as:

- biomedical named entity recognition
- relation extraction (such as identifying the relationship between a gene and a disease, or a protein and a small molecule)
- question answering (such as identifying what acronyms stand for and, therefore, their meaning and relevance)

Although discharge identification is not named as a possible use case, the model should perform well with this type of classification.

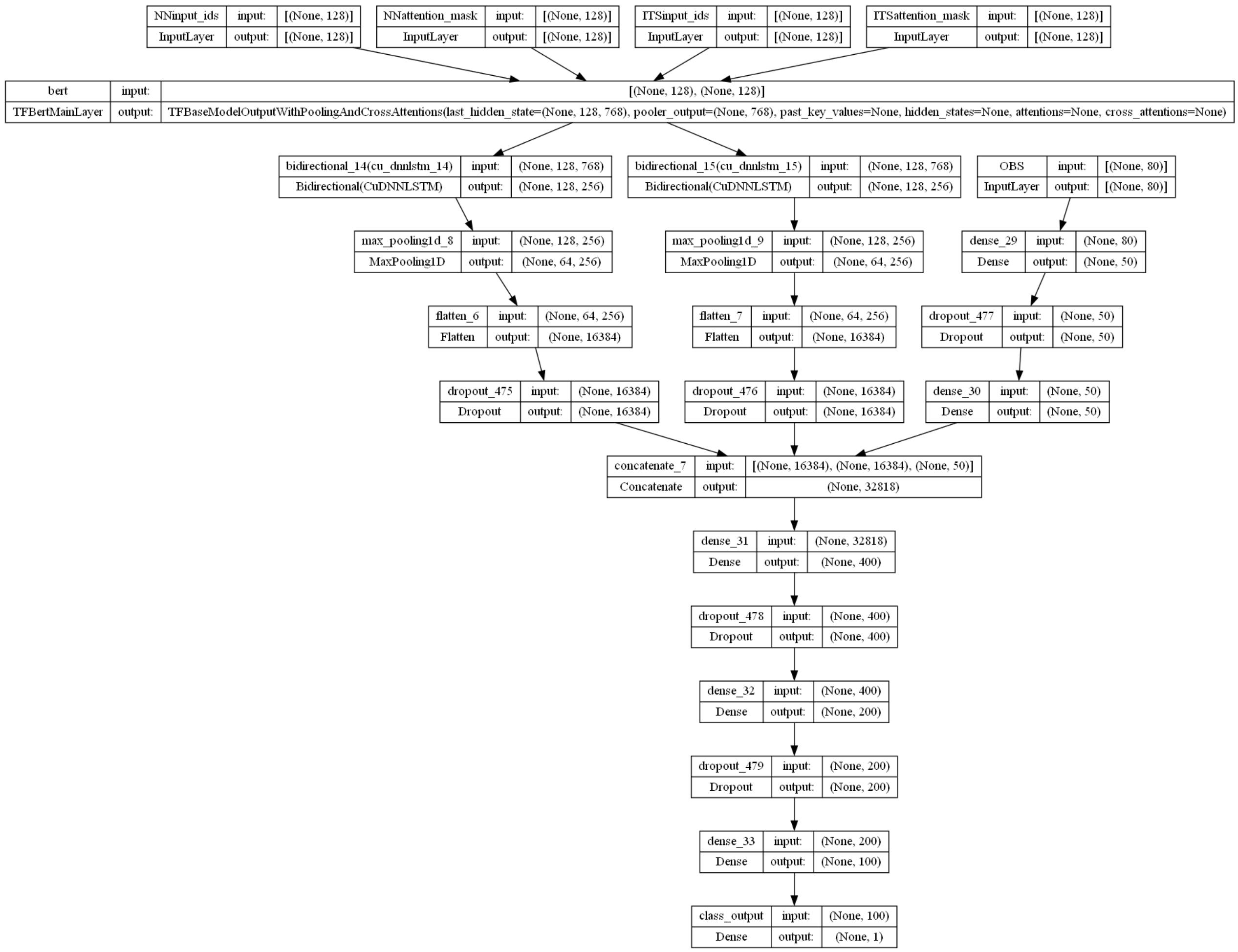


Figure 4.3: BioBERT Model

The tokenised text data (input_ids and masks) was fed into the BioBERT encoder as two separate inputs, along with a separate input for the non-textual data. As with BERT the tokenisation process was limited to 512 tokens, so text was truncated where appropriate to fit into this limitation. A bi-directional LSTM was used with the outputs of the BERT encoder and merged with the normalised data from the non-textual data to derive the output. The head for the BioBert and BERT were kept exactly the same to allow for direct comparison.

4.3.4 Keras FNet Text Representation Model

Due to the large dataset and limited hardware a model was chosen which would perform more efficiently than BERT/BioBERT, but have a similar performance. FNet was chosen, as according to the paper in 2020 titled "Mixing Tokens with Fourier Transform", the author claims this architecture is 80% faster then BERT while achieving 92-97% of the performance (Lee-Thorp, Ainslie, Eckstein, & Ontanon, 2022). This is achieved by replacing the self-attention mechanism in BERT with a Fourier Transform Layer, which helps to dramatically improve the speed of model convergance.

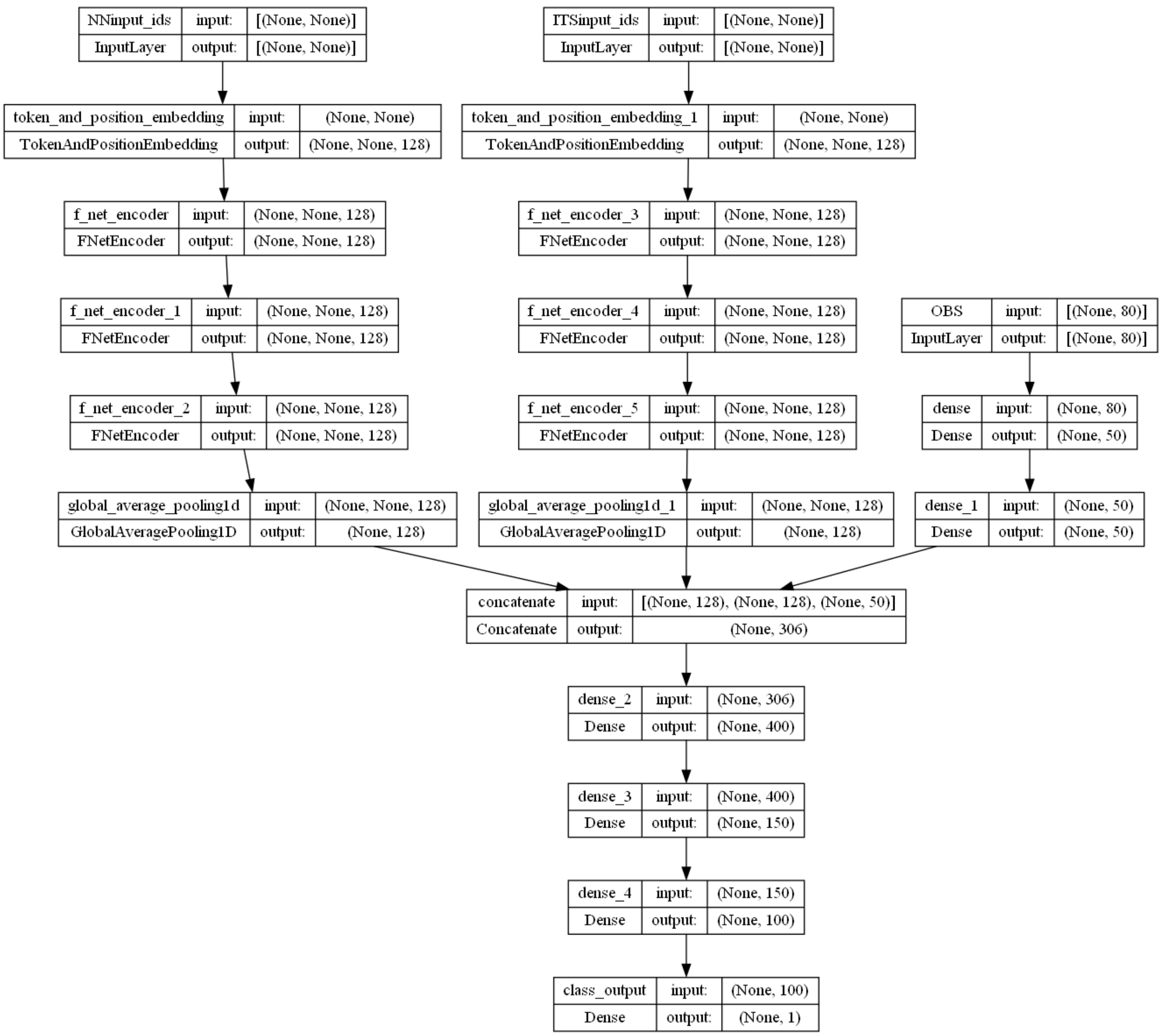


Figure 4.4: FNet Model

The developed model diagram in fig 4.4 outlines how the two textual inputs are processed through the FNet layers, which are then merged with the non-textural features into a simple feed forward dense network to produce an output. Once again the inputs are limited to 512 tokens so long text will be truncated.

4.3.5 Google Universal Sentence Encoder Text Representation Model

The Google Universal Sentence Encoder encodes text into high dimensional vectors that can be used for text classification, clustering, and other natural language tasks. It is trained on a variety of data sources to learn for a wide variety of tasks. The sources are Wikipedia, web news, web question-answer pages, and discussion forums. The pre-trained Universal Sentence Encoder is publicly available in Tensorflow-hub. It comes with two variations i.e. one trained with a Transformer encoder and the other trained with a Deep Averaging Network (DAN). The DAN encoder was used in this development, as it is computationally less expensive and with slightly lower accuracy than a Transformer encoder variation.

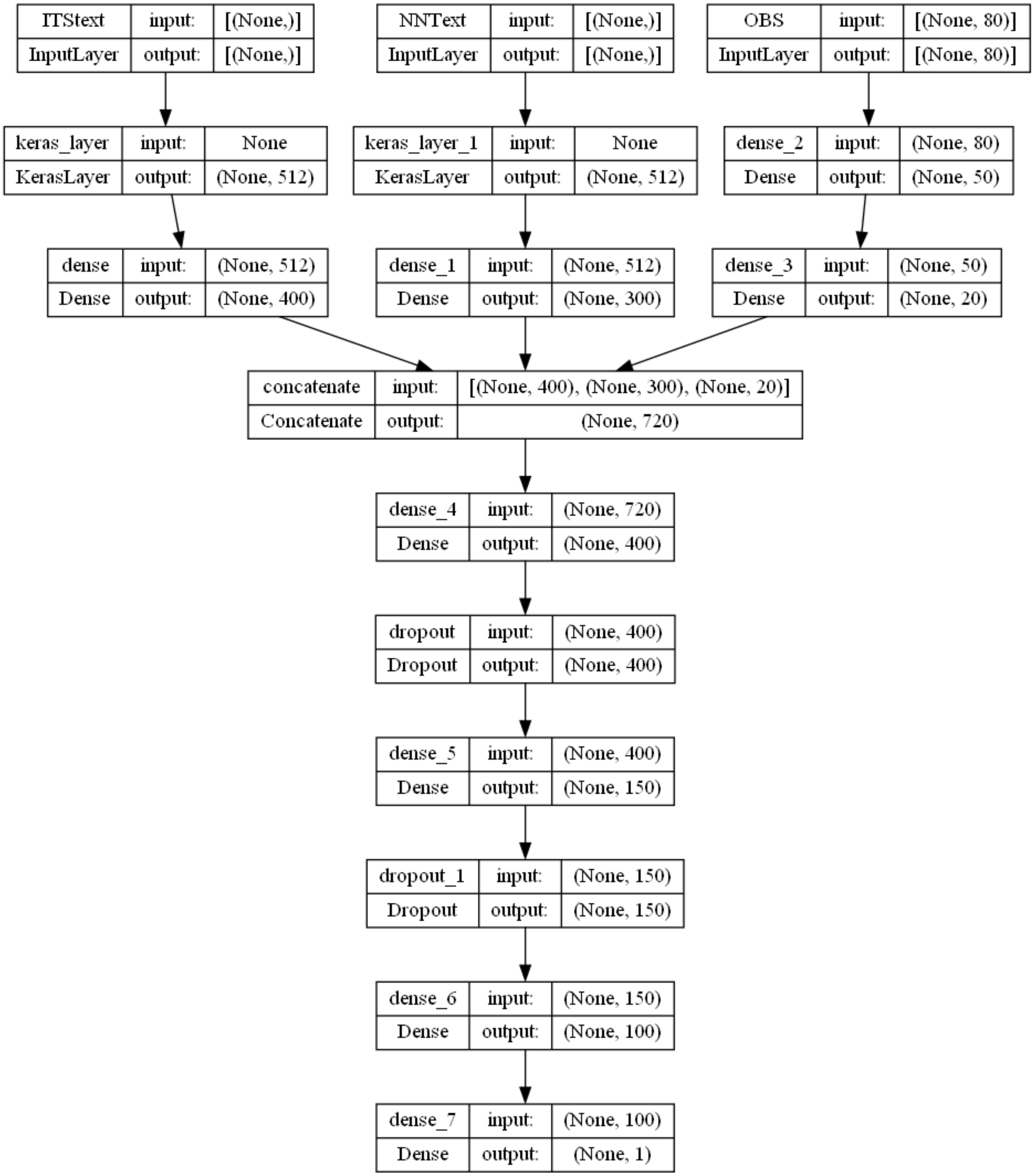


Figure 4.5: Google Universal Sentence Encoder

The Google Universal Encoder model in fig 4.5 is fairly simplistic, with the Nursing notes and Doctor's reports being encoded by the Universal Encoder directly and then merged with the non-textual features into an output dense network. The large size of the Doctor's text were not affected by any input length limitations. In order to adjust this model for better performance, the head of this model was customised in comparison with Bert/BioBert.

4.3.6 Ensemble of BioBERT Word and BioBERT Sentence Text Representation Models

To help bypass the oversize input problem with the doctors notes, an Ensemble of encoders were used on the text inputs. BioBERT was applied to the Nursing Notes as before, but a separate Transformer model based on BioBERT Sentence encoding was applied to the Doctor's notes.Sentence text representations can identify the order of words within a sentence and hence capture more context. As sentence text representations work at the sentence level rather that the word level, it is is not bound by input limitations of when dealing with long report like text. The model diagram is outlined in fig 4.6

Figure 4.6: Ensemble of techniques

4.4 Summary

Six models were developed which used varying techniques of text encoding to help achieve a significant F1 macro score.

Chapter 5

Results & Discussion

This chapter details the results of the models developed, along with a discussion for the performance of each model.

The performance of all the models was evaluated based on the F1 macro scores obtained after training each model on the training data set. The test data set was split 80:20 and used stratification and random seeds to ensure there was appropriate mix of data. Early stopping was applied to all models, with validation loss being monitored for five sequential reductions in performance to trigger a stop in convergence calculations.

5.1 Modeling Results

Table 5.1 outlines the results of the modeling completed. Early stopping was implemented to limit the time required for development.

Model	Max F1 Score
TFIDF	.7089
BERT with LSTM	.6642
BioBERT with LSTM	.6711
Kers FNet	.6665
Google Universal Sentence Encoder	.7102
Ensemble BioBERT and BioBERT sentence encoder	.6102

Table 5.1: Modeling Results

Based on on the highest F1 macro score achieved the Google Universal Sentence Encoder was chosen as the core technique for text encoding in the neural network.

5.2 Hypothesis Results

To test the hypothesis that the free text note features of the dataset (Nursing Notes and Doctors Notes) will be a significant component in determining if a patient will be discharged, the following table of tests were completed. While a separate model was used for each phase, the model was architecture was exactly the same with the exception of the number of inputs expanding at each phase to include each free text feature.

Model	Max F1 Score
TFIDF with no text	.3544
TFIDF with Nursing Notes	.5233
TFIDF with Nursing Notes and Doctors Notes	.7089

Table 5.2: Hypothesis Results

Based upon these results, the hypotheses has been accepted. Text note features are significant components in determining the probability of a patient discharge.

5.3 Discussion

The goal of this research was to develop a neural network based clinical decision support system, which could be used to help provide improved discharge decisions and dramatically reduce the time and effort in running the patient discharge process in a hospital.

The major inputs to the decision support system were patient telemetry, Nursing progress notes and Doctor's reports. While a small degree of cleaning was required with patient telemetry, extensive English corrections were required for the Nursing and Doctors' reports. The source systems for the data input (Hospital Information System) provide little in the way of spelling correction or auto correction, so it was no surprise that the quality of English was poor.

5.4 BERT Performance

The self-attention mechanism is one of the most important components that lead to the successes of transformer based models, which allows each token in the input sequence to independently interact with every other token in the sequence in parallel. This allow words to be assessed in the context of every word in the input sequence. However, the memory consumption of self-attention dramatically enlarges with sequence length, resulting in impracticable training time (44 days with a basic BERT model with no GPU) and easily reaching the memory limits of GPUs. Consequently, transformer-based models that leverage a full self-attention mechanism, such as BERT and BioBERT, always have an input sequence length limit of 512 tokens.

The average input token length for the Nursing progress and Doctors' reports notes was 93 tokens and 640 token respectively. This highlights the significant variation in the length of the two inputs and the GPU memory requirements when processing the doctors reports.

It can been seen from the histogram plots in fig 3.4 there is a significant difference in the token length. The Nursing notes token length would rarely exceed the 512 input limit of BERT, but the Doctors reports' would regularly exceed this limit. There are many ways to circumvent this 512 token limitation as outlined in the tech report from Li et al (Li, Wehbe, Ahmad, Wang, & Luo, 2022a), but due to limited hardware it was decided to shorten the text to the last 512 tokens in the text. The last 512 tokens were chosen, as it was assumed that reports are generally temporal in nature and the discharge of the patient would be discussed towards the end of the document.Text that was removed may have a negative impact on performance.

This model was restricted to a batch size of 8, due to the restrictions in hardware available at the time of development. BERT type models typically required batch size in the region of 64 - 4096 to achieve the best results (Popel & Bojar, 2018). This small batch size may also have had a detrimental impact on performance.

5.5 BioBERT Performance

The BioBERT model is based upon BERT and also has the limitation of the 512 input, but has been trained on Medical text which is more closely aligned to this domain of study. The specific HuggingFace model that was using is called "emilyalsentzer-Bio Discharge Summary BERT" and was trained on Publicly Available Clinical documentations including PubMed, PMC, and all MIMIC notes which include discharge summaries. The HuggingFace models are compatible with Keras libraries after translation.

While a performance increase was seen in comparison with standard BERT the difference was a minimal (less than 1%) improvement. The could be explained by the lack of international standards in clinical documentation and the extensive use of local Irish clinical terms and terminology.

As this model was also limited to a batch size of 8, the models performance would also have suffered for this reason.

5.6 TFIDF Performance

While the BERT and BioBERT models were based on encoding of words based on learned data from both the learned weights from the corpus and the trained BERT/BioBERT model, the TFIDF algorithm purely uses its own corpus to create the word representations. Jargon, vernacular and colloquialisms which were specific to the hospital, would get represented in the model and could have significant impact on improving the performance.

An example would be peoples names within the hospital - "Please ask Mary to complete a review" - this sentence may signify that the patient is having significant issues and may not be ready for discharge. This understanding may be picked up by TFIDF encoding due to corpus rapidity, but would not be picked up by a BERT based model due to its vague local nature.

The TFIDF model also does not have any input restrictions, so all text in reports can be considered available to recalculate weights.

5.7 Google Sentence Encoder

The Google Universal Sentence Encoder encodes text into high-dimensional vectors that can be used for text classification and other natural language tasks. The model is trained and optimized for sentences, phrases or short paragraphs. It is trained on a variety of data sources and a variety of tasks, with the aim of dynamically accommodating a wide variety of natural language understanding tasks. The input is variable length English text and does not have the BERT 512 input limitation, so the Doctor's Report did not need to be shortened.

5.8 FNet

The Keras FNet model was chosen as it uses Fourier Transforms to limit the number of parameters which help speed up the algorithm convergence time and is more memory efficient.

As FNet was based on BERT, it also has the 512 token input limitation. FNet regularly achieves 92-97% of the accuracy of BERT counterparts on the GLUE benchmark, but trains 80% faster on GPUs. The results obtained in this study demonstrated a marginal increase in performance on the pure BERT based algorithm.

5.9 Strength and Limitations of Results

While the highest score achieved was an F1 macro of .71, this still leaves significant room for performance improvement. The confusion matrix for the Google Universal Sentence encoder outlined in figure 5.1 was based upon a the test subset of data and a threshold of probability decision making of .50.

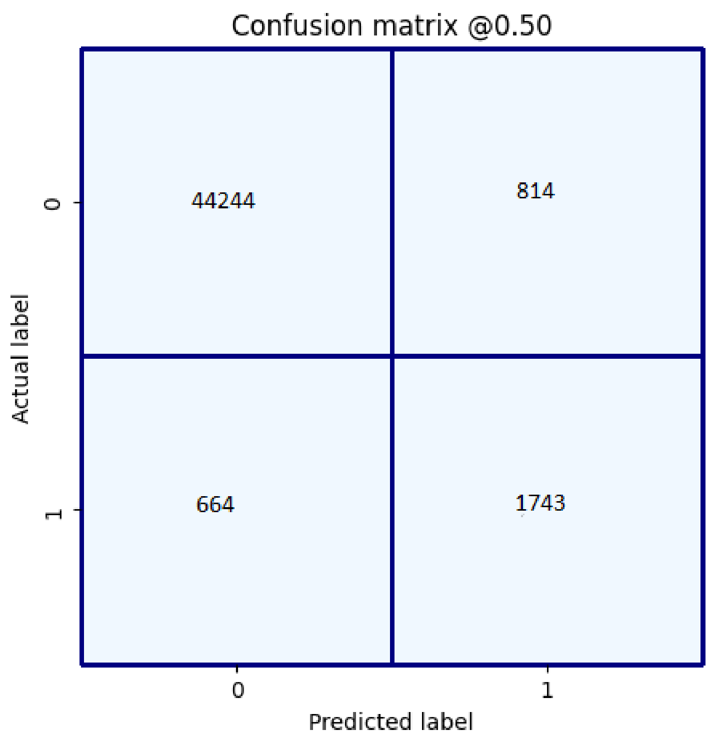


Figure 5.1: Confusion Matrix

This matrix outlines what the model achieved on this test set of data with 47,465 total records :

- Legitimate non-discharges detected(true negatives): 44244 (93.2% of total)

- Discharges incorrectly detected (false positives): 814 (1.8% of total)
- Non-discharges incorrectly detected (false negatives): 664 (1.4% of total)
- Discharges accurately detected: 1743 (3.6% of total)

While the model performance is reasonable, it does have some significant features when deployed in the context of a Clinical Decision Support engine used in a live hospital environment, which may prove extremely useful.

Typically, during each shift, a distributed (nurses, doctors, consultants , administration) decision will be made on all the inpatients to determine whether they are eligible to be discharged. On average, 95% of the patients will not be discharged and about 5% will be eligible for discharge. In this institution, there are 180 patient beds, so the objective would be to identify the 9 patients on each shift which will be discharged.

The below histogram outlined in fig 5.2 displays the output of the probability of discharge, for each patient from the model for all 180 patients from a single shift. The probability of discharge have been binned to give an overview of all patients.

As you can see, the 164 patients have been assigned a probability near zero for discharge and counting the other probability bins give a total of 16 other patients (2+2+0+1+2+3+4+1+1). All the other patients have been assigned probabilities of discharge in the range of 0.10 to 1.00 and have the potential to be a valid discharge.

But also more importantly, the model was able to quickly identify the vast majority patients who have the potential of discharge, which is typically a small subset of the inpatients on each shift.

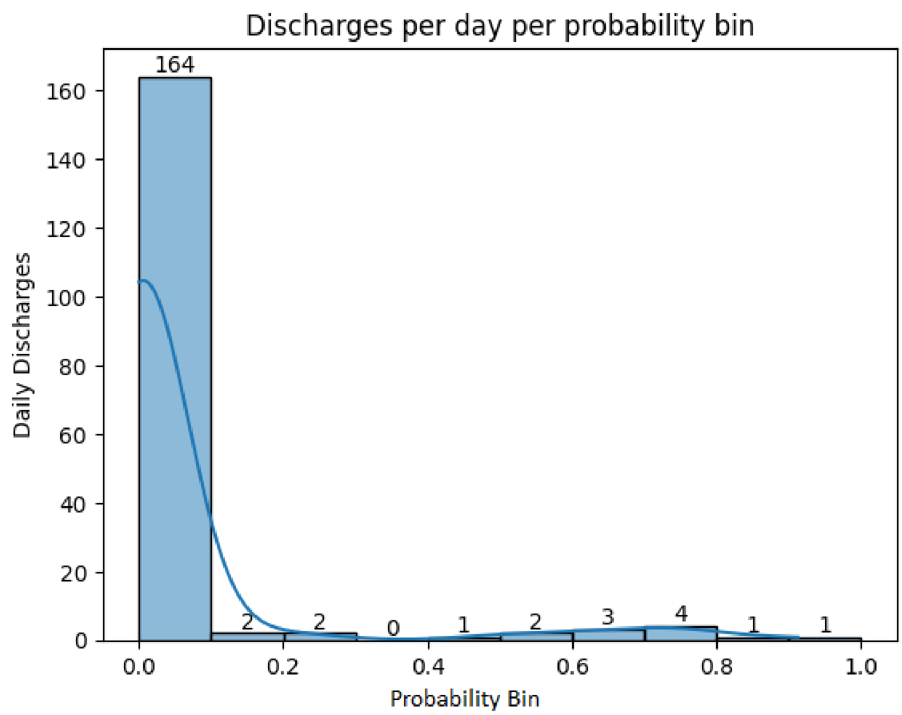


Figure 5.2: Probability of Discharges in a Shift

Simply concentrating the discharge team's efforts on processing discharges on the 16 patients with high probabilities, rather than on all 180 patients would save significant amounts of time.

Implementing this operational strategy, does risk that a patient would be missed when the model classifies a patient with a discharge probability less than 0.05.

Looking at all the test data (47,465) and errors when there was a prediction in the 0-.05 probability bin, shows that there was only 142 errors in this bin, which is 0.3 percent of the total predictions. Therefore the probability of a mis-classification in this classification bin (0-.05 probability) is 0.3 percent or 99.7 percent accurate. This risk can be further reduced by lowering the threshold for classification for a discharge, but his will increase the number of patients that would need to be reviewed.

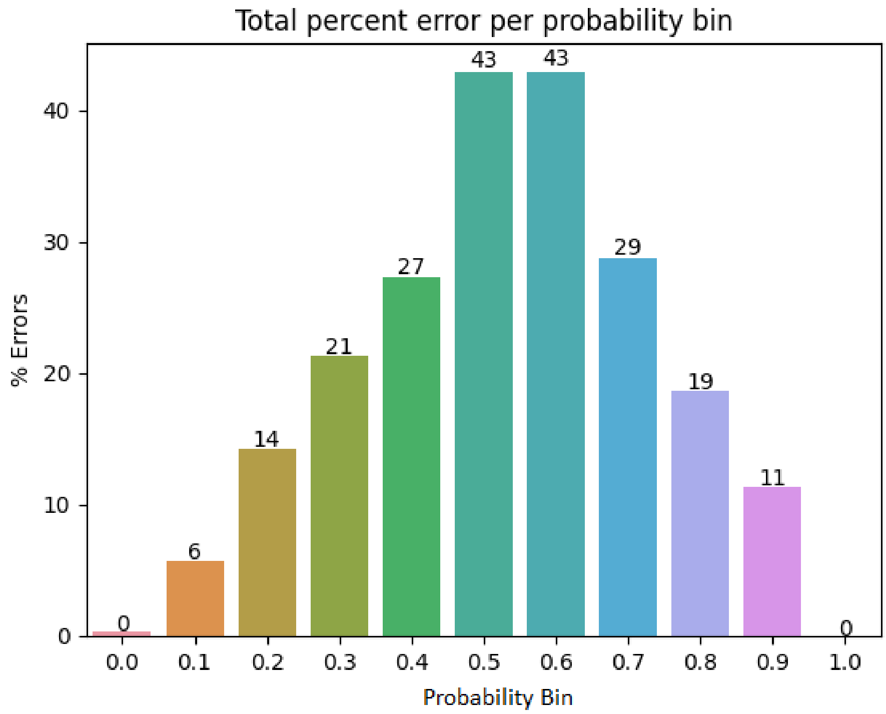


Figure 5.3: Total Percent Error per Probability bin

Outlined in figure 5.3 is histogram of the percentage errors in each probability bin for all the test data. The percentage error around the 0.4 to 0.6 probability is high (50%) as expected, as it is difficult to classify this patient with a threshold of 0.5. But the probability around the 0 - .1 probability of discharge is very low, and this is where the vast majority of patents will be classified.

With more time spent on modeling using faster hardware and more patient data being fed into the development, the percentage errors should be reduced making the system more effective over time and further reducing risk of a mis-classification.

It should be noted that there is a limit to how accurate the system can be. Historical patient discharge data used by the model, is assumed to be low risk discharge decisions. Even with filtering and data cleaning to remove any outliers, these low risk decisions may not have been low risk. Repetitive Human error in the historical data, could led to poor discharge output probabilities in the model. But the system will be deployed as a clinical decision support engine, where the final decision will always be made by the clinician.

5.10 Summary

The breakdown and evaluation of the overall system development is discussed in this chapter.

Five neural network models, which are the core of the Hospital Discharge Planning system Clinical Decision Support were developed and teased using a subset of the live data.

The model based upon Google's Universal sentence encoder demonstrated the best F1 macro score achieving a maximum of .71. While this score leaves room for improvement, it does not limit the capability of the system in deploying a clinical decision support engine. A decision support engine based on the probability outputs of the model, can help reduce the time and risk associated with discharging patients in an operational environment.

The next chapter will give the detailed summary of the research, contribution and impact of the research and the areas of future research.

Chapter 6

Conclusion

6.1 Conclusions

The purpose of the research has been to develop a Hospital Discharge Planning system augmented with a neural network based Clinical Decision Support learning engine.

At the initial stage of this research, a literature review was conducted, summarizing existing state-of-the-art clinical decision support systems, patient discharge prediction and text encoding techniques.

A neural network was developed using over a quarter of a million historical patient records gathered from a hospital in Ireland. This historical information included two free text fields, which were hypothesised to contain significant information on the discharge status of a patient. This hypothesised was then accepted through he sequential testing a standard model with and without text inputs.

One of the models developed using Google Universal Sentence Encoder, achieved an F1 macro score of .71 for predicting the patient discharge from the hospital. While this score left significant room for improvement, the model output does have some significant features when deployed in the context of Clinical Decision Support engine, which may prove extremely useful.

On average the number of patients that would typically be discharged from an 180 inpatient ward would typical be 9 patients. This model would on average predict 12 patient to be discharged with 3 false positives / false negatives. The vast majority of patients receive a low discharge probability below 5%, and if a low probability is predicted, it will be correct 99.7% of the time. This could allow the discharge teams, to concentrate the efforts on patients with high discharge probability. Patients with low discharge probability which have been marked for discharge by other processes may also need further scrutiny.

6.2 Problem Definition

The process of discharging patients from a tertiary care hospital is one of the key processes to ensure an efficient operation of a hospital. The discharge decision is complex and requires multiple interactions with nurses, family, consultants and doctors, which can be very time consuming.

This thesis descries how a neural network based engine can be used to provide a Clinical Decision Support system to help provide rigorous decisions and dramatically reduce the time and effort in running the discharge process in a hospital.

6.3 Summary of Findings

A neural network model was developed using approx. a quarter of a million snapshots of inpatient features from a medium size hospital in Ireland. The model used a diverse set of inputs, which included a variety of patient metrics such as temperature, heart rate and length of stay etc, but also included more complex free text inputs such as patient progress notes and doctor's reports.

The model developed from these inputs was capable of predicting patient discharge with an f1 macro score of .71, but also more importantly was able to quickly identify 99.7% of patients who have the potential to be discharged. A clinical decision support system based on this neural network could dramatically reduce the time and effort in running the discharge process in a hospital.

6.4 Contributions and Impact

This research explores the development of machine learning in the patient discharge process in an operational environment.

With the results obtained, there could be a number of positive operational impacts in deploying a system like this in an operational environment including:

- Help reduce risk of discharge process for both the patient and the hospital. Poor discharge decisions can affect the medical risk to the patient and multiple operational costs to the hospital.
- Optimise faster decisions for the patient and the hospital, to help save the patient stress and help staff to be more effectively deployed in managing patient care.
- Having medical staff spend more time dealing with patient concerns and questions before discharge, may help reduce work with re-admissions and questions when patients have returned home.
- The deployed system could also help identify questionable discharges, which may be highlighted for re-assessment.

All of these impacts together could make a very significant difference to the throughput and operational efficiency. This development process and model is also transferable and scale-able to larger institutions.

This work also highlights some of the limitations of clinical domain transformer models, when applied to localised free text clinical data. Models such as BioBERT did not perform as expected in comparison with the Google Universal Sentence encoder, due to a number of limitation such as; local vocabulary and input size limitations.

6.5 Future Work & Recommendations

This development work focused primarily on testing with test data and did not incorporate any real world feedback from clinicians and patients, which was out of scope.

The obvious next step is to deploy the system and seek feedback from all stakeholders, which may help tweak the system.

There is a wealth of text encoding techniques which could be used to extract useful knowledge form the Doctor's / Nurse's notes. Further analysis of the text and domain relevant encoders may help improve the models performance. Deployment of the system into other hospitals in Ireland, may also prove useful to allow a more localised version of the model to be trained, which has been exposed to notes from several different clinical institutions.

Using several separate models, which is specific to the clinical speciality of the patient may also be useful. While the developed model did take the clinical speciality into account as a feature in the data set, developing separate models may help improve the F1 score but this may add to the complexity of an operational discharge system.

Due to the 512 token input limitation of the BERT based transformers, consideration should be given to splitting long text notes into multiple smaller paragraphs, which could be assigned to different inputs in the model architecture. Truncating the text which is what was developed, may loose knowledge in the text that was critical to making a discharge decision. Alternatively a Longformer based transformer could be used to help encode the longer text inputs (Li, Wehbe, Ahmad, Wang, & Luo, 2022b).

A limitation of batch size of 8 was encountered due to the restrictions in hardware available at the time of development. BERT type models typically require a batch size in the region of 64 - 4096 to achieve the best results (Popel & Bojar, 2018). Redevelopment of the BERT based models on hardware which is not limited to an 8 batch size, may allow for immediate significant improvement.

References

Al-Qudah, H. S. S. (2020). The interventional models that aim to reduce the rate of hospital readmission [Journal Article]. *Researchers World*, *11*(2), 1-6. doi: http://dx.doi.org/10.18843/rwjasc/v11i2/01

Arya, M., & Sastry G, H. (2020). Deal–'deep ensemble algorithm' framework for credit card fraud detection in real-time data stream with google tensorflow. *Smart science*, *8*(2), 71-83.

Attiga, Y., Chen, S.-Y., LaGue, J., Ovalle, A., Stott, N., Brander, T., ... Francis-Lyon, P. (2018). Applying deep learning to public health: Using unbalanced demographic data to predict thyroid disorder. In *2018 ieee 9th annual information technology, electronics and mobile communication conference (iemcon)* (p. 851-858). doi: 10.1109/IEMCON.2018.8614888

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

Bates, D. W., Kuperman, G. J., Wang, S., Gandhi, T., Kittler, A., Volk, L., ... Middleton, B. (2003). Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *Journal of the American Medical Informatics Association : JAMIA*, *10*(6), 523-530.

Benbassat, J., & Taragin, M. I. (2013). The effect of clinical interventions on hospital readmissions: a meta-review of published meta-analyses [Journal Article]. *Israel Journal of Health Policy Research*, *2*(1), 1. doi: 10.1186/2045-4015-2-1

Bright, T. J., Wong, A., Dhurjati, R., Bristow, E., Bastian, L., Coeytaux, R. R., ... Lobach, D. (2012). Effect of clinical decision-support systems: a systematic review. *Annals of internal medicine*, *157*(1), 29.

Chen, T., Dredze, M., Weiner, J. P., & Kharrazi, H. (2019, 07). Identifying vulnerable older adult populations by contextualizing geriatric syndrome information in clinical notes of electronic health records. *Journal of the American Medical Informatics Association*, *26*(8-9), 787-795. Retrieved from https://doi.org/10.1093/jamia/ocz093 doi: 10.1093/jamia/ocz093

Clancy, C. M. (2009). Reengineering hospital discharge: a protocol to improve patient safety, reduce costs, and boost patient satisfaction [Journal Article]. *Am J Med Qual*, *24*(4), 344-6. doi: 10.1177/1062860609338131

Clements, A., Halton, K., Graves, N., Pettitt, A., Morton, A., Looke, D., & Whitby, M. (2008). Overcrowding and understaffing in modern health-care systems: key determinants in meticillin-resistant staphylococcus aureus transmission [Journal Article]. *Lancet Infect Dis*, *8*(7), 427-34. doi: 10.1016/s1473-3099(08)70151-8

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, *abs/1810.04805*. Retrieved from http://arxiv.org/abs/1810.04805

Donzé, J., Aujesky, D., Williams, D., & Schnipper, J. L. (2013). Potentially avoidable 30-day hospital readmissions in medical patients: Derivation and validation of a prediction model [Journal Article]. *JAMA internal medicine*, *173*(8), 1-7. doi: 10.1001/jamainternmed.2013.3023

Eigner, I., & Bodendorf, F. (2020). Decision support for patient discharge in hospitals – analyzing the relationship between length of stay and readmission risk, cost, and profit [Book Section]. In (p. 77-84). Cham: Springer International Publishing. doi: 10.1007/978-3-030-59595-1_6

Flynn, L., Liang, Y., Dickson, G. L., Xie, M., & Suh, D.-C. (2012). Nurses' practice environments, error interception practices, and inpatient medication errors. *Journal of nursing scholarship*, *44*(2), 180-186.

Goncalves-Bradley, D. C., Lannin, N. A., Clemson, L. M., Cameron, I. D., & Shepperd, S. (2016). Discharge planning from hospital [Journal Article]. *Cochrane database of systematic reviews U6*(1), CD000313.

Gonçalves-Bradley, C. C., DC Lannin, & S, S. (2022). Discharge planning from hospital. *Cochrane Database of Systematic Reviews*(2). Retrieved from `https://doi.org//10.1002/14651858.CD000313.pub6` doi: 10.1002/14651858.CD000313.pub6

Gould, I. (2006). Costs of hospital-acquired methicillin-resistant staphylococcus aureus (mrsa) and its control. *International Journal of Antimicrobial Agents*, *28*(5), 379-384. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0924857906003499` doi: https://doi.org/10.1016/j.ijantimicag.2006.09.001

Graham, M. M., James, M. T., & Spertus, J. A. (2018). Decision support tools: Realizing the potential to improve quality of care [Journal Article]. *Canadian journal of cardiology*, *34*(7), 821-826. doi: 10.1016/j.cjca.2018.02.029

Gramaje, A., Thabtah, F., Abdelhamid, N., & Ray, S. K. (2019). Patient discharge classification using machine learning techniques [Journal Article]. *Annals of data science*. doi: 10.1007/s40745-019-00223-6

Guan, M., Cho, S., Petro, R., Zhang, W., Pasche, B., & Topaloglu, U. (2019, 01). Natural language processing and recurrent network models for identifying genomic mutation-associated cancer treatment change from patient progress notes. *JAMIA Open*, *2*(1), 139-149. Retrieved from `https://doi.org/10.1093/jamiaopen/ooy061` doi: 10.1093/jamiaopen/ooy061

Hall, K., Chang, V., & Jayne, C. (2022). A review on natural language processing models for covid-19 research. *Healthcare Analytics*, *2*, 100078. Retrieved from https://www.sciencedirect.com/science/article/pii/S2772442522000326 doi: https://doi.org/10.1016/j.health.2022.100078

HSE. (2017). A practical guide to discharge and transfer from hospital [HSE standards documentation]. *HSE Standards documentation*, *1*(1), 47. Retrieved from https://www.hse.ie/eng/about/who/qid/resourcespublications/nationalintegratedcareguidance.pdf

Information, H., & Authority, Q. (2021). The need to reform ireland's national health information system.

Kapinski, N., Zielinski, J., Borucki, B. A., Trzcinski, T., Ciszkowska-Lyson, B., & Nowinski, K. S. (2018, Jun 19). Estimating achilles tendon healing progress with convolutional neural networks. *arXiv.org*. Retrieved from https://tudublin.idm.oclc.org/login?url=https://www.proquest.com/working-papers/estimating-achilles-tendon-healing-progress-with/docview/2073559470/se-2 (Copyright - © 2018. This work is published under http://arxiv.org/licenses/nonexclusive-distrib/1.0/ (the "License"). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License; Last updated - 2019-09-13)

Klarenbeek, S. E., Schuurbiers-Siebers, O. C. J., van den Heuvel, M. M., Prokop, M., & Tummers, M. (2021). Barriers and facilitators for implementation of a computerized clinical decision support system in lung cancer multidisciplinary team meetings—a qualitative assessment. *Biology*, *10*(1). Retrieved from https://www.mdpi.com/2079-7737/10/1/9 doi: 10.3390/biology10010009

Lee-Thorp, J., Ainslie, J., Eckstein, I., & Ontanon, S. (2022). Fnet: Mixing tokens with fourier transforms.

Levin, S., Barnes, S., Toerper, M., Debraine, A., DeAngelo, A., Hamrock, E., … Howell, E. (2020). Machine-learning-based hospital discharge predictions can support multidisciplinary rounds and decrease hospital length-of-stay [Journal Article]. *BMJ innovations*. doi: 10.1136/bmjinnov-2020-000420

Li, Y., Wehbe, R. M., Ahmad, F. S., Wang, H., & Luo, Y. (2022a). *Clinical-longformer and clinical-bigbird: Transformers for long clinical sequences* (Tech. Rep.). Cornell University Library, arXiv.org.

Li, Y., Wehbe, R. M., Ahmad, F. S., Wang, H., & Luo, Y. (2022b, 11). A comparative study of pretrained language models for long clinical text. *Journal of the American Medical Informatics Association*. Retrieved from `https://doi.org/10.1093/jamia/ocac225` (ocac225) doi: 10.1093/jamia/ocac225

Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T.-Y. (2022, 09). BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics*, *23*(6). Retrieved from `https://doi.org/10.1093/bib/bbac409` (bbac409) doi: 10.1093/bib/bbac409

Maguire, P. J., Power, K. A., Daly, N., Farren, M., McKeating, A., & Turner, M. J. (2016). High dependency unit admissions during the first year of a national obstetric early warning system. *International journal of gynecology and obstetrics*, *133*(1), 121-122.

McCoy, J. T. H., Pellegrini, A. M., & Perlis, R. H. (2018). Assessment of time-series machine learning methods for forecasting hospital discharge volume [Journal Article]. *JAMA network open*, *1*(7), e184087-e184087. doi: 10.1001/jamanetworkopen.2018.4087

McDermid, R. C., & Bagshaw, S. M. (2011). Icu and critical care outreach for the elderly. *Best Practice & Research Clinical Anaesthesiology*, *25*(3), 439–449.

Murphy, D., Wiley, M. M., Clifton, A. C., & McDonagh, D. (2004). Updating clinical coding in ireland: Options and opportunities..

Nishimura, A. A., Shirts, B. H., Salama, J., Smith, J. W., Devine, B., & Tarczy-Hornoch, P. (2015;2016;). Physician perspectives of cyp2c19 and clopidogrel drug-gene interaction active clinical decision support alerts. *International journal of medical informatics (Shannon, Ireland)*, *86*, 117-125.

Olakotan, O. O., & Yusof, M. M. (2021). The appropriateness of clinical decision support systems alerts in supporting clinical workflows: A systematic review. *Health Informatics Journal*, *27*(2), 14604582211007536. Retrieved from https://doi.org/10.1177/14604582211007536 (PMID: 33853395) doi: 10.1177/14604582211007536

Popel, M., & Bojar, O. (2018). Training tips for the transformer model. *CoRR*, *abs/1804.00247*. Retrieved from http://arxiv.org/abs/1804.00247

Rebane, J., Samsten, I., & Papapetrou, P. (2020). Exploiting complex medical data with interpretable deep learning for adverse drug event prediction. *Artificial Intelligence in Medicine*, *109*, 101942. Retrieved from https://www.sciencedirect.com/science/article/pii/S0933365719311546 doi: https://doi.org/10.1016/j.artmed.2020.101942

Safavi, K. C., Khaniyev, T., Copenhaver, M., Seelen, M., Zenteno Langle, A. C., Zanger, J., ... Dunn, P. (2019, 12). Development and Validation of a Machine Learning Model to Aid Discharge Processes for Inpatient Surgical Care. *JAMA Network Open*, *2*(12), e1917221-e1917221. Retrieved from https://doi.org/10.1001/jamanetworkopen.2019.17221 doi: 10.1001/jamanetworkopen.2019.17221

Sung, S.-F., Hsieh, C.-Y., & Hu, Y.-H. (2020, 02). Two decades of research using taiwan's national health insurance claims data: A bibliometric and text mining analysis on pubmed (preprint). *Journal of Medical Internet Research*, *22*. doi: 10.2196/18457

Survey: Deep learning concepts and techniques for electronic health record. (2018;2019;). , *2018-*, 1-5.

Toyabe, S.-i. (2012). Detecting inpatient falls by using natural language processing of electronic medical records [Journal Article]. *BMC Health Services Research*, *12*(1), 448. Retrieved from https://doi.org/10.1186/1472-6963-12-448 doi: 10.1186/1472-6963-12-448 van Walraven, C., Jennings, A., & Forster, A. J. (2012). A meta-analysis of hospital 30-day avoidable readmission rates [Journal Article]. *Journal of Evaluation in Clinical Practice*, *18*(6), 1211-1218. doi: https://doi.org/10.1111/j.1365-2753.2011.01773.x

Wei, H., Gao, M., Zhou, A., Chen, F., Qu, W., Wang, C., & Lu, M. (2019). Named entity recognition from biomedical texts using a fusion attention-based bilstm-crf. *IEEE Access*, *7*, 73627-73636. doi: 10.1109/ACCESS.2019.2920734

Appendix A

Additional content

A.1 Appendix 1 - Discharge Capability Factor

The Discharge Capability Factor (DCF) is a metric determines the number of errors the model will produce, when a patient is assess to having a probability of discharge of less than .05. If the model assess that a patient has a low probability of discharge and the patient was discharged this is assumed to be an error in the model and is something that need to be minimised. The objective would be to create a model where these errors are minimal.