An Ensemble LSTM for Rare Event Detection

S. N. Ramchand
866648

Submitted to Swansea University in partial fulfilment
of the requirements for the Degree of Master of Science

Swansea University
Prifysgol Abertawe

Department of Computer Science
Swansea University

30th September 2020

Declaration

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ……Suraj Ramchand…… (candidate)

Date ……23/11/2020……

Statement 1

This work is the result of my own independent study/investigations, except where otherwise stated. Other sources are clearly acknowledged by giving explicit references. I understand that failure to do this amounts to plagiarism and will be considered grounds for failure of this work and the degree examination as a whole.

Signed ……Suraj Ramchand…… (candidate)

Date ……23/11/2020……

Statement 2

I hereby give my consent for my work, if accepted, to be archived and available for reference use, and for the title and summary to be made available to outside organisations.

Signed ……Suraj Ramchand…… (candidate)

Date ……23/11/2020……

Abstract

In the complex study of rare diseases, it is crucial to find distinctive clinical features that allow for accurate or possibly early diagnosis. However, as suggested in the class of of such conditions, its rarity is a contributing factor of it being understudied. Fabry is a type of rare disease that has a broad range of phenotypes, making it increasingly difficult to detect and diagnose. The consequences of which are dire where Fabry patients could experience premature death almost 20 years in advance. Although the condition in focus was initially Fabry, the lack of data pushed this study towards the use of Sepsis data which is similar in nature. This study explores the use of clinically sound methods of data imputation for datasets with temporal nature and have large amounts of missing data. Additionally, this study reports the features selected by the machine learning algorithms. Gleaning from the knowledge from past studies, this study shows the differences in using data that was imputed with the best clinical judgement and compares it. The findings of this study is purposed as a foundation for future work on the use of artificial intelligence in rare event detection models and temporal datasets.

Acknowledgements

I would like to take this opportunity to credit Dr. Xie, Dr. Cole and Dr. Franks for their guidance and unwavering support for the development of this paper.

Contents

Chapter 1

Introduction

From robot health practitioners [1] to personalised and preventative medicine [2], the healthcare sector is edging towards an exponential increase, with health spending across the globe doubling in the next three years[3].Not only that, but the World Health Organization claims that 80% of premature coronary heart diseases(CHD's), strokes and type II diabetes can be prevented. Additionally, 40% of cancer could be prevented if risk factors, such as diet and levels of physical activity, were modified early on[4].

Healthcare has evolved rapidly from the days where patients with cystic fibrosis only had a life expectancy of 20 years, to now having an average expectancy of 40 years[5]. Although, as the focus tend towards diseases with a higher prevalence, a disregard for those with an rare condition builds. It was shown that the population pool of rare disease patients could be approximated to the size of the world´s third largest country[6].

Before the Orphan Drug Act of 1983, the lack of public awareness led to rare disease patients being known as health orphans, as they lacked a treatment plan or even a fixed diagnosis. Soon, after this policy was enacted the pharmaceutical agencies coined the term 'Orphan drugs' to refer to Rare disease drugs being developed in direct response to the change in the financial burden of researching, developing and testing.

To date, there are over 7000 different rare diseases which afflict millions of individuals in the World and are responsible of the deteriorating physical health, mental health and socioeconomic conditions. These conditions are mysterious due to their rarity, resulting in them being under studied and having neither formal diagnostic criteria nor cure. Around 80% of these conditions are attributed to genetic mutation and other arise from an exposure to toxins or infectious agents and, occasionally, from an adverse response to

therapeutic interventions[7]. Some of the rare disease patients face journeys that could be long and difficult, spending an average of 7.6 years and going through seven different doctors before obtaining a final diagnosis[8].

This is just the tip of the ice berg, as the individuals affected with any particular rare disease is relatively small and the pool of rare conditions is so large, a series of challenges complicates the development of safe and effective drugs and medical devices to prevent, diagnose, treat, or cure these conditions. Treatments for rare conditions are therefore increased. Take for example hereditary angioedema which is a life-threatening genetic condition which causes idiopathic swelling in several areas of the body and has a prevalence of 2 in 100,000 individuals. Its treatment can cost approximately $490,000 thousand for one patient in a single year[9]. Further, reviewing the annual median drug cost per patient in the US, it is found that the estimated cost of non-orphan drugs per patient is $28, 000 while orphan drugs could cost around $140,443 [10].

Fabry Disease (OMIM 301500) is one such disease. Fabry Disease (FD) is a progressive, life-threatening, multi-systemic, genetic, lysosomal storage disorders with an estimated prevalence rate of 1 in 40,000 to 1 in 60,000 live male births. The prevalence rate among females remains unknown to date[11, 12]. FD is characterised by deficiency in transcription or lack of a functional enzyme which leads to the accumulation of a fatty waste product called globotriaosylceramide (GL-3) and related glycosphingolipids in different organs.

The main challenge in diagnosing Fabry is that there are no fixed event marking its start or presence and that the prodromal symptoms are non-specific, thus making it challenging to identify even among healthcare professionals [13]. During childhood, symptoms could include neuropathic pain in the extremities, hypohidrosis, and gastrointestinal symptoms such as abdominal pain, diarrhea, and food intolerance.These clinical manifestations could be easily mistaken for other gastrointestinal conditions. Echoing similar sentiments, a study on the misdiagnosis of Fabry found that rheumatic fever, food intoxication or a petechiaea, in order of most misdiagnosed condition, are some of the incorrect diagnosis of Fabry [14].

The obvious traits of Fabry only start showing at adulthood by which time, significant organ damage would have already occurred. As with most rare diseases, it takes an average of 7-10 years and require up to seven different specialist for an accurate diagnosis [15]. This crippling late manifestation would not only seriously disrupt the lives of patients, if left untreated, FD patients could have their life expectancy reduced by up to 20 years [16].

It is, therefore, necessary for FD to be better understood through advanced methodologies. This study aims to understand the pathology of Fabry using Artificial Intelligence (AI) and discuss the effectiveness of a rare event detection methods. This paper would be helpful to direct future work in this area, for example, by pointing to a method to detect rare events in Electronic Health Care Records (EHR).

1.1 Fabry Disease: A rare genetic disorder

This study focuses on showcasing the importance of temporal deep learning techniques in identifying relevant biomarkers for rare conditions. The background information presents the pathology, biomarkers and treatment options of Fabry, as well as AI techniques applied on the detection of rare diseases.

1.1.1 Pathology

Fabry is a multi-system disorder affecting mainly the nervous system, skin, heart, kidneys and the eyes. As it affects major organs in the body, the average lifespan of FD patients have a significantly reduced lifespan. The average life expectancy of Fabry patients are 58 and 75 years for male and female, respectively [17].

It's inheritance pattern is X-linked [18], but current findings suggest that females with a mutation on one X chromosome can experience the same severity, although only evident at a later stage[19]. Atypical manifestations may occur in males (atypical variants), with manifestations more or less confined to 1 organ system (kidneys or heart) and later onset[20].

The condition is characterised by a deficiency or malfunction of the α-Gal A enzyme, localized on Xq22.1, which results in progressive storage of glycosphingolipids including globotriaosylceramide (Gb3) and other glycolipids. The accumulation occurs in many types of cell, mainly in the kidney, heart, liver and smooth muscle of the vascular system [21].

Classic manifestations of Fabry Patients including onset of acute and chronic neuropathic pain(Acroparaesthesia), hypohidrosis, angiokeratoma and gastrointestinal symptoms typically occur in the first two decades of their lives, which significantly reduces the quality of living [22]. As the condition progresses it leads kidney dysfunction and renal failure, cardiovascular disease, stroke, and premature death. Additional clinical features are elaborated further below beginning with Acroparaesthesia.

Acroparaesthesia refers to the condition of burning or tingling or numbing sensations affecting various parts of the body. It is caused by the excessive storage of Gb3 in nerve endings and dorsal root ganglia. Another explanation for the condition is that the exposure to cold causes the blood vessels to constrict and small fibre mal-perfusion due to accumulation of storage material in cutaneous vessels and vasa vasorum [23]. Luciano et al. found a duration-dependent dysfunction of a-delta and c-fibres that further impairs the cold than warm sensation of Fabry cases [24]. The decreased cold and warm senses is phenomenon more evident at adulthood than childhood. Identifying specific neuropathic pain pathways in patients would be be beneficial for a more precise treatment. A wide spectrum of other neurological and psychological manifestations include fatigue, tinnitus, recurrent vertigo, headache and depression [25].

Hypohidrosis is the reduced ability to perspire due to the accumulation of Gb3 in the eccrine sweat glands and their associated blood vessels [26]. It is more evident in male than in female patients. This condition is commonly associated with physical activities mainly in the summer where temperature changes in the body are triggers [23]. Hypohidrosis yields to an inability to cool the body during exercise. If left untreated, it could cause hypothermia, heat stroke, increased neuropathic pain and death [27].

Angiokeratoma are small, raised, dark-red spots on the skin. Although this condition is not specific to FD, it is the only known sign that presents early and would allow for early detection of FD. In the absence of angiokeratoma, early clinical diagnosis is often difficult to establish [28].

Gastrointestinal symptoms are common, presenting with diarrhoea, constipation, recurrent nausea or vomiting and abdominal pain. These symptoms often lead to hospitalisation and are sometimes mistaken for chronic inflammatory bowel disease. Altered intestinal motility and diarrhoea could be caused by the accumulation of fats in intestinal autonomic nerve ganglia while malabsorption could be caused by storage deposition in the small intestine [29].

Both Cardiac and renal deterioration are symptoms that are observed in adults and children, hence early diagnosis and careful monitoring is necessary [30]. A biomarker that is associated with detecting renal dysfunction is the level of a major plasma protein (albumin) in the urine. The condition known for excessive levels of the protein is known as albuminuria, usually the first indication of renal dysfunction [31]. Progressive

renal insufficiency occurs in older patients, although significant renal involvement has previously been reported [32].

Many attempts have been made over the years to find a disease-specific marker that would ideally serve as a rapid screening tool as well as indicator of therapy response.

1.1.1.1 Diagnostic Pathway and Biomarkers

Although Fabry mainly affects the heart, kidney and skin, no suitable suitable plasma or urine marker has been found [24]. In tissue biopsies, Gb3 which accumulates in lysosomes, is routinely used to diagnose the disease. However, there are debates on the usefulness of it. Young's study on hemizygotes, heterozygotes and healthy controls did not identify significant increase in Gb3 levels in plasma and urine in heterozygotes and hemizygotes with non-classical mutation [33]. Moreover, Schiffmann did not find any correlation between plasma or urine Gb3 and clinical response to Enzyme Replacement Therapy (ERT) [34]. On the use of detecting Gb3 deposits in peripheral blood mononuclear cells (PBMCs), it was found to have failed in detecting Gb3 deposits in atypical mutations.

While controversies surround the use of Gb3 as a biomarker for Fabry, the product of its degradation, Globotriaosylsphingosine (lysoGb3), is currently used in disease monitoring [29] and identifying pathogenicity of a mutation for homozygotes as well as heterozygotes. Fig. 1.1 shows the current diagnosis procedure. An evaluation/validation study of 124 patients found that higher lysoGb3 levels were correlated with the diagnosis of Fabry [35].

In a study [36], females with normal -Gal activity who were found with accentuated levels of lysoGb3 had a subsequent diagnosis of Fabry. Furthermore, lysoGb3 levels have been shown to decrease with ERT, especially among patients with classical forms of the disease [34]. These cogent findings result in lysoGb3 being widely accepted as the most accurate marker of the disease.

Several other biomarkers have been explored but are not yet widely in use due to the lack of data. Cammarata et al successfuly identified four micro-RNAs specific for Fabry patients in a pilot study regardless of mutation type, sex and age [37]. However, two of which were linked with endothelial dysfunction and the the size of the study population with 30 patients and 30 controls respectively, meant it could not be generalised. Other attempts on identifying Fabry-specific biomarkers include identification of new Gb3 isoforms using metabolomics [38] and using proteome analyses to measure abnormal urinary protein excretion; however, none were able to validate a clinically useful parameters.

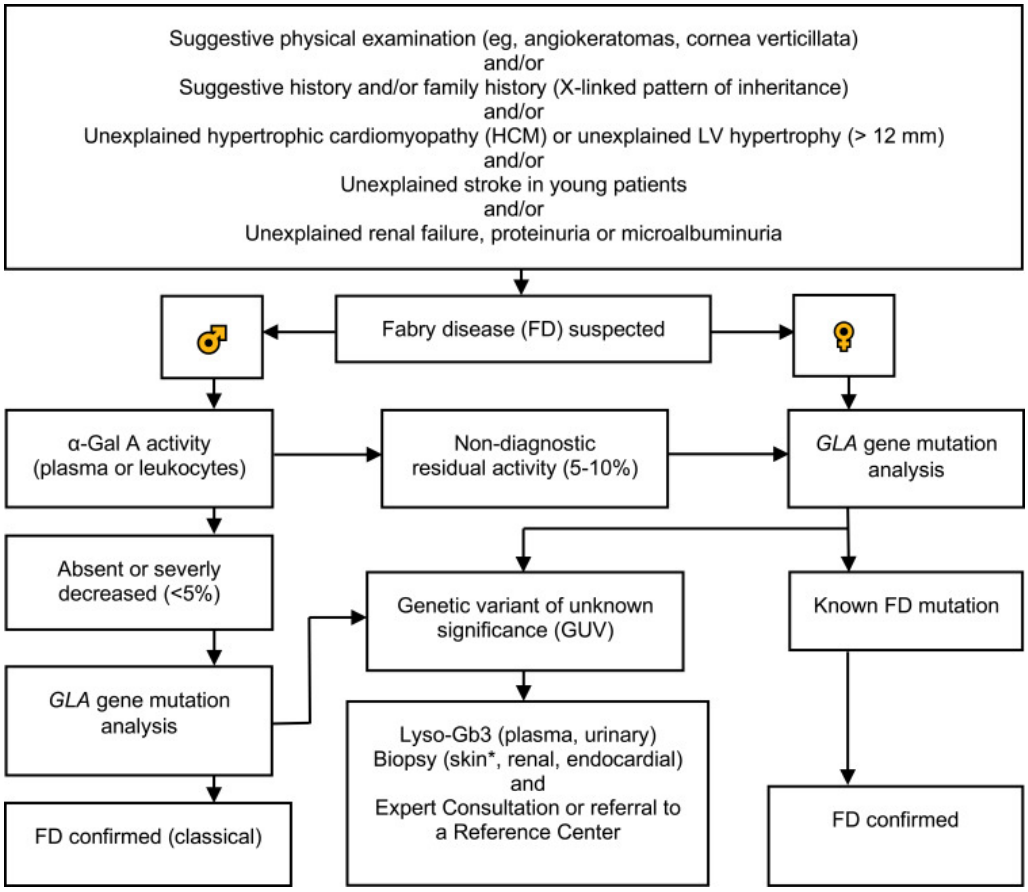


Figure 1.1: Diagnosis Pathway

1.1.2 Application of Artificial Intelligence

As noted above, existing studies using basic analysis such as correlation have found meaningful information between biomarkers, severity of the disease, outcomes of treatments and the clinical outcome, Fabry. Although, some studies lack confidence when applied to a different pool of patients, it is primarily due to the heterogeneity of the diseases and its manifestations.

While simple analytical methods can provide insight into correlations between biomarkers, they are not capable of modelling all the possible scenarios and might overlook some hidden patters such as time-trends on the granular patient level. On the other hand, advances in information technology have led to a more widespread application of artificial

intelligence (AI) and machine learning in the field of medicine and healthcare [39], where they are capable of performing nearly as well as an expert would on the subject area.

AI and machine learning typically use large, multivariate datasets to train algorithms, which are then used to make predictions and on the new or test data. Importantly, the computations by which these methods generate their output are not explicitly developed by a programmer, but instead are implicitly learned by the algorithm from training dataset (hence the term "machine learning").

Given the complexity in identifying diagnostic and response biomarkers in conditions, even more so for rare diseases, research on the subject matter would greatly benefit from the application of AI and machine learning techniques. Case in point, while it is virtually impossible for a physician to memorize information about thousands of rare diseases as well as pathologies in relation to common ailments, modern computers can easily "memorize" and handle huge quantities of digital information.

A scoping review [40] that spans from 2010 to 2019, further emphasises the lack of publications in this area. Fig. 1.2 showcases the number of published studies on the AI in the field of Rare disease, ranging from improving time to diagnosis with machine learning to assessing genetic mutation through the patients symptomology.

All these studies apply a series of methods with the most common being the application of ensembled techniques, which are characterised by the congregation of artificial learning models, that are capable of best differentiating between a case and a control. These models uniquely posses different characteristics that when combined yields to a better results than each single one alone[41]. Fig. 1.3 enhances the image of the algorithms applied to rare diseases and not only hints at the lack of literature but also the fact that Artificial neural networks are not the most common technique applied to rare disease.

Digging in a bit deeper, shows that the use of Neural networks through a series of application in healthcare, tends to lead to a higher classification and prediction accuracy, as noted in [42, 43]. The use of these networks extends to the domain of imbalanced datasets and temporal learning, which are the main attributes of our dataset. In [44], a Diagnostic decision support systems (DDSSs) that predicts onset of a rare condition was developed and could calculate the probability that a patient would develop a disease based on the patient symptoms, its top suggestion matched the confirmed diagnosis in 89.25% of cases. The underlying neural network understood the temporality and rarity of the conditions and obtained such promising results.

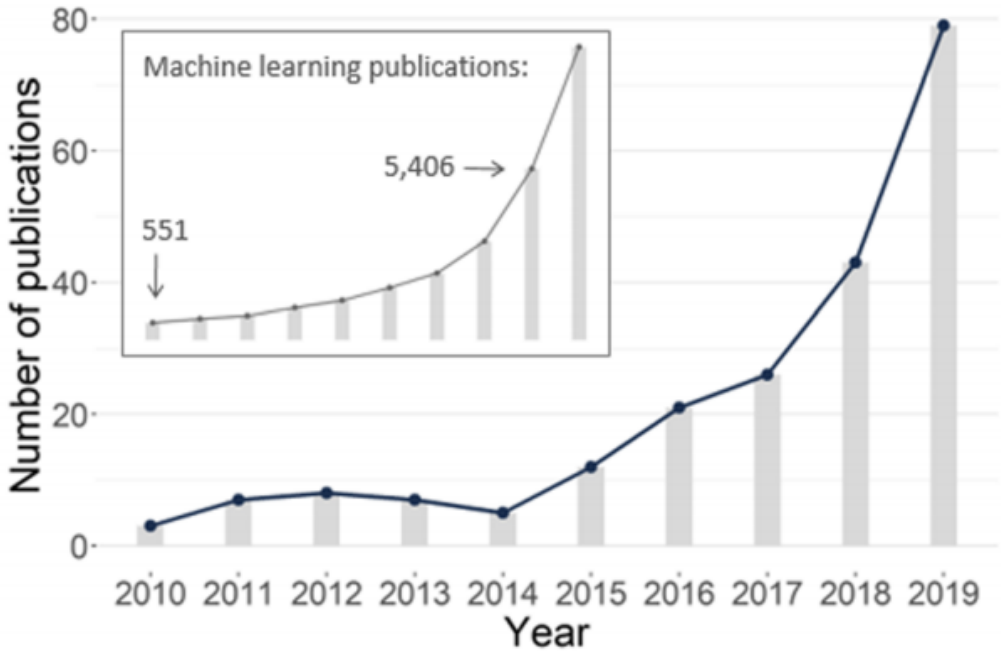


Figure 1.2: Publications on Applying AI to tackle the battle of Rare conditions Vs AI in healthcare (inside figure)[41]

Despite the evident potential of the application of AI in rare disease detection, the literature on its application in the study of rare diseases was scarce and even more so for Fabry. This gap that was identified is the motivation of this study, which aims to present a rare event detection algorithm which can provide for further a better understanding of the rare conditions, severity levels and responses in biomarkers with clinical soundness.

1.1.3 Rare Disease and Fabry Data Registries

One of the most extensive knowledge bases for rare diseases is Orphanet [45]. It provides information including disease epidemiology, associated genes, inheritance types, disease onsets and useful references for the clinical terminologies. It is also linked to special care services, patient care registries amongst other resources.

EuroBioBank leverages on RDConnect for information sharing. RDConnect is a centralised data repository that contains a combination of registries, biobanks and genetic data for research on rare disease. It also has an in-house bioinformatics tool for data linkage and exploration which is useful for non-experienced users [45]. It follows an easy

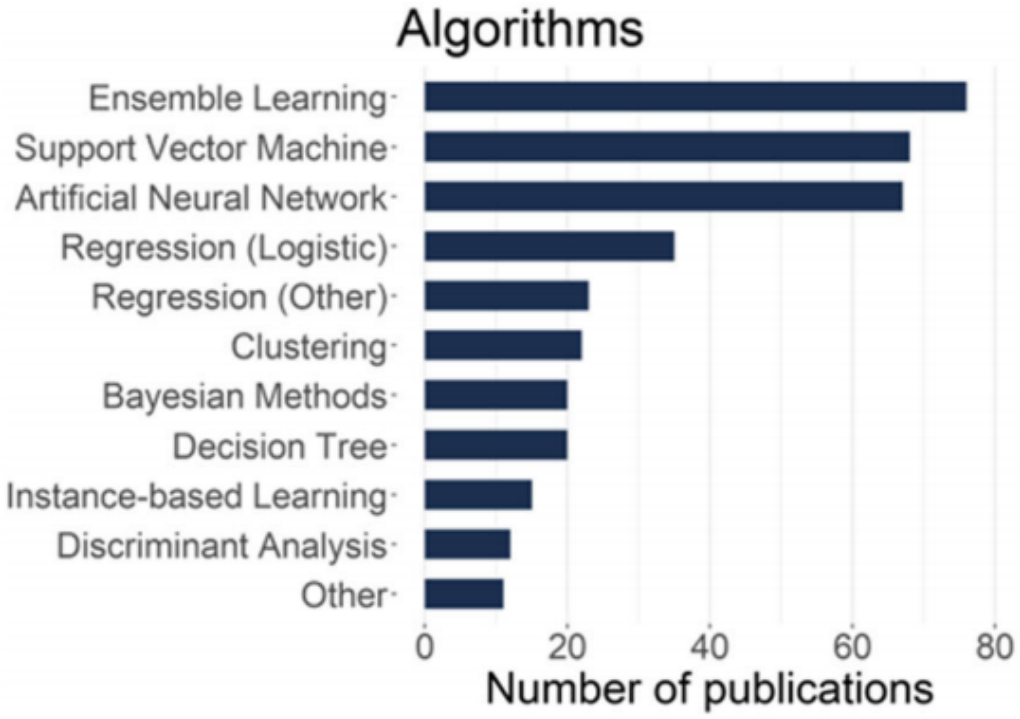


Figure 1.3: Artificial Intelligence algorithms used in Rare disease publications[41]

process for requesting access to the data through e-mail. Researchers seeking to have access should have basic information on the rare disease of interest prepared prior to the request.

Fabry Outcome Survey (FOS) [46] is a European based database for Fabry patients with who are receiving, or are suitable candidates for, enzyme replacement therapy with agalsidase alfa in various parts of Europe. The patients include both genders of full age range. It also contains historical clinical records, including the year of diagnosis, demographic details, family history and treatment for each patient that allows for a more comprehensive research on the disease. Additionally, signs and symptoms are recorded using an extensive checklist, which covers majority of the Fabry's pathology. Quantitative measurements from routinely collected data in primary care are also included in FOS. Although results from advanced scans and investigations such as cardiac and renal ultrasound and biopsy data are part of the dataset, these measures are not routinely collected thus lesser data of these measures are found. Essential biomarkers such as measurements of a-galactosidase A activity are recorded as categorical variables. It is evident that FOS would be the ideal data source for studies on Fabry. However, obtaining

the data requires applicants to file in a request with the scope and objective of the project and intended use of the data. It also requires an approval from an ethical approval board since it contains patient data.

The final registry for data is the Secure Anonymized Information Linkage (SAIL) databank [47]. It contains nearly 80% of care information of the Welsh population registered with Welsh healthcare services. It acts at a gateway to data from a range of healthcare providers and services. Some of the many datasets available in SAIL include primary care, outpatient, hospital, mortality, emergency and demographic data. The anonymized patient data allows for data linkage, similar to FOS. Another similar aspect is that it contains historical records and electronically recorded measures associated with patients. SAIL differs from FOS in that it contains data of patients registered with general practices in Wales instead of only Fabry patients. However, it was found that only 60 Fabry patients exist within the databank. The request for access to SAIL data is similar to that of FOS. Data users are also required to take a certification on ethical data usage before being granted the access.

An important note when applying for access in these registries is that they require payment for use. Although, the price would be insignificant when compared to the intangible cost of improving the diagnosis rate of medical conditions in patients and allowing better treatments or preventive measures to be found from the analyses of these data.

1.2 Sepsis

It can be noted that moving into the 21st Century there is a greater emphasis on data linkage in healthcare. Thus, the availability of such resources should foment a greater research into the pathology of rare diseases.

Although the premise of this study is on Fabry, at the time of the study data on Fabry was not yet made available. As such, a replacement of condition was called for. However this, should not shift the motivation of this study, which is to improve the clinical outcome of Fabry patients and potentially yield an algorithm which is capable of predicting a condition in a temporal highly imbalanced dataset.

Sepsis is similar to Fabry in that both conditions require analysis of biomarkers, usually assessed against normal ranges. Both are heterogeneous conditions and are temporal in nature. If not well managed or left undiagnosed, both conditions could deteriorate drastically, resulting in premature death. Within their respective datasets

which is highly imbalanced and contains a large amount of missing data, both conditions form the minority of the clinical outcomes.

In 1991, the American Society of Chest Physicians (ACCM), introduced the idea that sepsis was caused by the host's response to the infections rather than the invading organism alone. According to them sepsis was; "a life threatening organ dysfunction caused by a deregulated host response to an infection" [48]. This could lead to tissue damage, organ failure and eventually death, if not treated [49]. It is estimated that 30 million people develop sepsis globally and accounts for 6 million deaths annually [50].

Sepsis prediction is highly relevant though complicated, mostly due to a low specificity of usable physiological parameters [51]. The reliable and early identification of sepsis is often complicated by its broad range of clinical manifestations, which attributes to delays in treatment and diagnosis. Several studies adopted the use of machine learning techniques to identify clinical features useful for predicting the disease outcome. In 2016, Calvert et al. introduced a novel algorithm called InSight which is capable of extracting key diagnostic features from the MIMIC III sepsis database [52].

1.2.1 Clinical Scores

Another important aspect to training deep learning models on medical data, especially temporal data, is the ability to rely on previously assessed clinical scores to facilitate in picking up hidden trends.

1.2.1.1 Systemic Inflammatory Response (SIRS)

Systemic Inflammatory Response (SIRS) could be used to describe if the sepsis was present on patient or not. SIRS was and still is quite common (almost up to 90%) among the patients in the Intensive Care Unit (ICU)[53].

Although, by examining the SIRS criterion's, clinicians and doctors were able to figure out if the patient had a serious medical situation with high sensitivity, but its specificity was quite low to determine if the patient was developing sepsis or not. Therefore, same consensus got back together in 2016 in order to come up with a more specific explanation and description in order to detect sepsis. This lead to the development of the sequential organ failure clinical values.

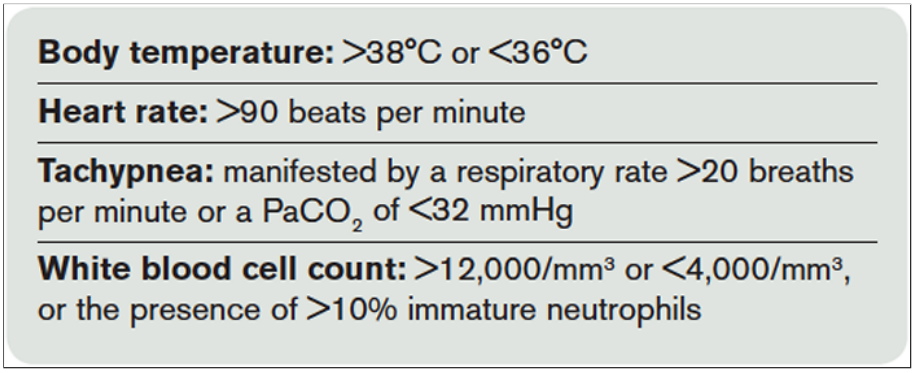


Figure 1.4: SIRS Table Scoring Criteria[54]

1.2.1.2 quick SOFA

Severe cases of Fabry and Sepsis result in multiple organ failure. The measure of extent of this measure and the rate of disease deterioration could be quantified using the quick Sequential Organ Failure Assessment(qSOFA)[55]. This method detects the propensity to multi-system organ failure and consists of attributing a point for each criteria listed in Fig. 1.5.

By assigning qSOFA scores at various time points in the temporal Sepsis data, it is possible for the LSTM model to identify trends in organ failure with time.

Figure 1.5: quick SOFA Score Criteria

1.2.1.3 SOFA

The Sequential Organ Failure Assessment (SOFA)[56], a similar metric for assessing organ functions, requires a 24 hour window for the accurate score calculation. The formula used in its calculation to be calculated and is also more complex than in qSOFA. Yet, with the understanding that clinical scores are vital, an hourly SOFA score for each organ group calculated and stored as additional features in the dataset. The scoring criteria can be found in Fig. 1.6. It aims to understand the state at which an organ is, as compared to its optimal function

Variables/score	0	1	2	3	4
P_aO_2/FiO_2 (mmHg)	>400	≤400	≤300	≤200	≤100
Platelets ($\times 10^3$/uL)	>150	≤150	≤100	≤50	≤20
Bilirubin (mg/dL)	<1.2	1.2–1.9	2–5.9	6–11.9	>12
Cardiovascular (Hg/kg/min)	–	MAP<70	Dop≤5	Dop>5 (Epi≤0.1)	Epi>0.1
Glasgow Coma Scale	15	13–14	10–12	6–9	<6
Creatinine (mg/dL)	<1.2	1.2–1.9	2–3.4	3.5–4.9	>5

MAP = mean arterial pressure; Dop = dopamine; Epi = epinephrine.
doi:10.1371/journal.pone.0031256.t002

Figure 1.6: SOFA Table Scoring Criteria[57]

Chapter 2

Methodology

2.1 Data Origin

The Sepsis dataset used in this study is publicly available and was obtained from the website which hosted the PhysioNet/Computing in Sepsis Challenge held in 2019[58]. The focus of the competition on predicting sepsis with a 6 hour history of patients using anonymized Intensive Care Unit EHR before the onset time of sepsis according to Sepsis-3 clinical criteria[51].

The original data source was the Beth Israel Deaconess Medical Center. According to the National Institutes of Health, the challenge did not belong to human subject research and thus did not require an institutional ethics approval for the use of patient data.

The data consisted of a combination of averaged hourly vital signs, laboratory values, patients age and demographics. In particular, the data contained 40 clinical variables: 6 demographic features (discrete), 8 vital signs (continuous) and 26 laboratory measurements (continuous), listed in Table 2.1. As a whole,the data included over 1.3 million hourly time windows and 7.5 million data points.

2.2 Data Pre-Processing

Initial analysis led to the discovery that there were columns with a high number of missing values. This meant that filling this values would require a complex approach to maintain the clinical significance of the data and to not generate false patterns. Thus, Imputing the dataset initially was out of question.

Table 2.1: Feature table of Sepsis Dataset

Vital signs (columns 1-8)	
HR	Heart rate (beats per minute)
O2Sat	Pulse oximetry (%)
Temp	Temperature (Deg C)
SBP	Systolic BP (mm Hg)
MAP	Mean arterial pressure (mm Hg)
DBP	Diastolic BP (mm Hg)
Resp	Respiration rate (breaths per minute)
EtCO2	End tidal carbon dioxide (mm Hg)
Laboratory values (columns 9-34)	
BaseExcess	Measure of excess bicarbonate (mmol/L)
HCO3	Bicarbonate (mmol/L)
FiO2	Fraction of inspired oxygen (%)
pH	N/A
PaCO2	Partial pressure of carbon dioxide from arterial blood (mm Hg)
SaO2	Oxygen saturation from arterial blood (%)
AST	Aspartate transaminase (IU/L)
BUN	Blood urea nitrogen (mg/dL)
Alkalinephos	Alkaline phosphatase (IU/L)
Calcium	(mg/dL)
Chloride	(mmol/L)
Creatinine	(mg/dL)
Bilirubin_direct	Bilirubin direct (mg/dL)
Glucose	Serum glucose (mg/dL)
Lactate	Lactic acid (mg/dL)
Magnesium	(mmol/dL)
Phosphate	(mg/dL)
Potassium	(mmol/L)
Bilirubin_total	Total bilirubin (mg/dL)
TroponinI	Troponin I (ng/mL)
Hct	Hematocrit (%)
Hgb	Hemoglobin (g/dL)
PTT	partial thromboplastin time (seconds)
WBC	Leukocyte count (count*10^3/µL)
Fibrinogen	(mg/dL)
Platelets	(count*10^3/µL)
Demographics (columns 35-40)	
Age	Years (100 for patients 90 or above)
Gender	Female (0) or Male (1)
Unit1	Administrative identifier for ICU unit (MICU)
Unit2	Administrative identifier for ICU unit (SICU)
HospAdmTime	Hours between hospital admit and ICU admit
ICULOS	ICU length-of-stay (hours since ICU admit)
Outcome (column 41)	
SepsisLabel	For sepsis patients, SepsisLabel is 1 if ttsepsis6 and 0 if t<tsepsis6. For non-sepsis patients, SepsisLabel is 0.

Within the initial column in Fig. B.1, there is a clear pattern that shows most data missing came from laboratory measurements which were not routinely collected and were optional for collection. A closer look at the frequency of laboratory measurement over length of stay in Fig. 2.1 suggests that there is evidence to conclude that the missing laboratory measurements follows a Not Missing At Random (MAR) description. Supporting literature on imputation suggests that for datasets that are classed as Not Missing at Random (NMAR) should not be imputed[59]. This is consistent with a clinical environment where patients laboratory tests conducted at specific time intervals. Creatinine is a lab value which is measured both at around 6 hours and 24 hours. In [60] it was suggested time between tests period of 8 hours is required to note the difference in Kidney functionality instead of imputing values. This 8 hour period is further reviewed in a study assessing the "The Top 10 Things Nephrologists Wish Every Primary Care Physician Knew" [61], which concludes that depending on the physiology and age of the patient an increase of serum createnine within 8 hours could reflect a reduction is the Glomerular filtration rate (GFR), indicative of renal dysfunction.

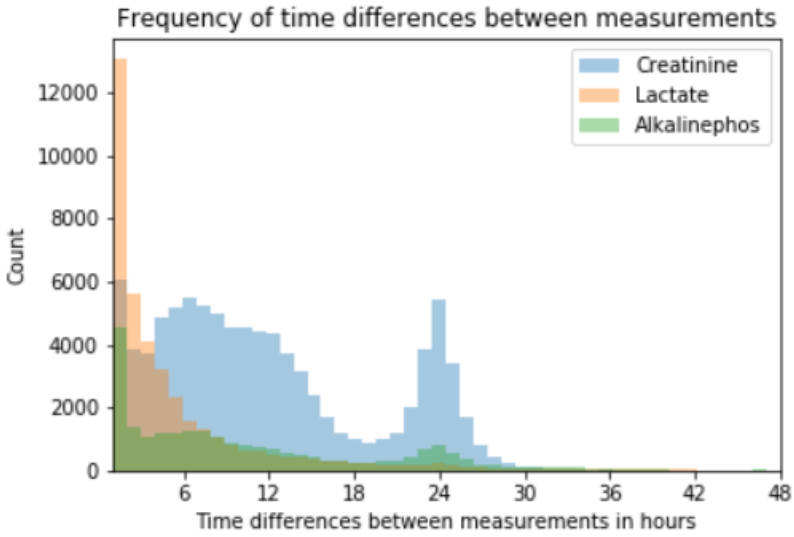


Figure 2.1: Frequency of time differences between measuremetns [62]

On the other hand, the vital and demographic data show a pattern of Missing At Random (MAR). The values excluded from the study was the location of the patient during the measurements (Unit1 and Unit2), where Unit1 is an ICU stay in the Medical Intensive Care Unit or Unit 2, the Surgical Intensive Care unit. These features excluded

as they contained above 50% of missing data and there is a lack of literature to suggest a correlation between the patients unit and the diagnosis of Sepsis.

It was noted that there was a substantial number of missing Diastolic Blood Pressure (DBS), which was easily calculated by using the Systolic Blood Pressure(SBP) and the Mean Arterial Pressure(MAP)[63], as seen below:

$$DBP = ((MAP * 3) - SBP)/2 \tag{2.1}$$

A custom function to implement this mathematical function was also created Listing 2.1.

```
############## MAP = (SBP + 2(DBP))/3  Now we can fill in NaN of DBP as all SBP
    missing values are alos missing DBP ###################

# Dyastolic Blood Pressure Missing Values
DBP = trainSetA_concat.copy()
DBP_missing = DBP[DBP['DBP'].isna() == True]
DBP_missing.isna().sum()

############## Partially filling In DBP missing values from MAP and SBP
    ###################

for row in range(1,len(trainSetA_concat)):

    if pd.notna(trainSetA_concat.loc[row,'SBP']) and pd.isna(trainSetA_concat.loc[row
        , 'DBP']):
        trainSetA_concat.at[row, 'DBP'] = (((trainSetA_concat.loc[row, 'MAP']*3) -
            trainSetA_concat.loc[row, 'SBP'])/2).round(0)
    else:
        pass
        #print('DBP or No SBD Value Present')
```

Listing 2.1: Calculation of Diastolic Blood Pressure

With all of the above in mind, a flow diagram was generated to describe the approach taken in both case scenarios, as seen in Fig. 2.2.

2.2.1 Random Under-Sampling

An added challenge of this dataset is that it is highly imbalanced where Sepsis patients were the minority class, representing 7.5% of all patients. In order to account for the heterogeneity of Sepsis while allowing the algorithm to be capable of detecting it, a resampling method is necessary in the training of the individual submodels.

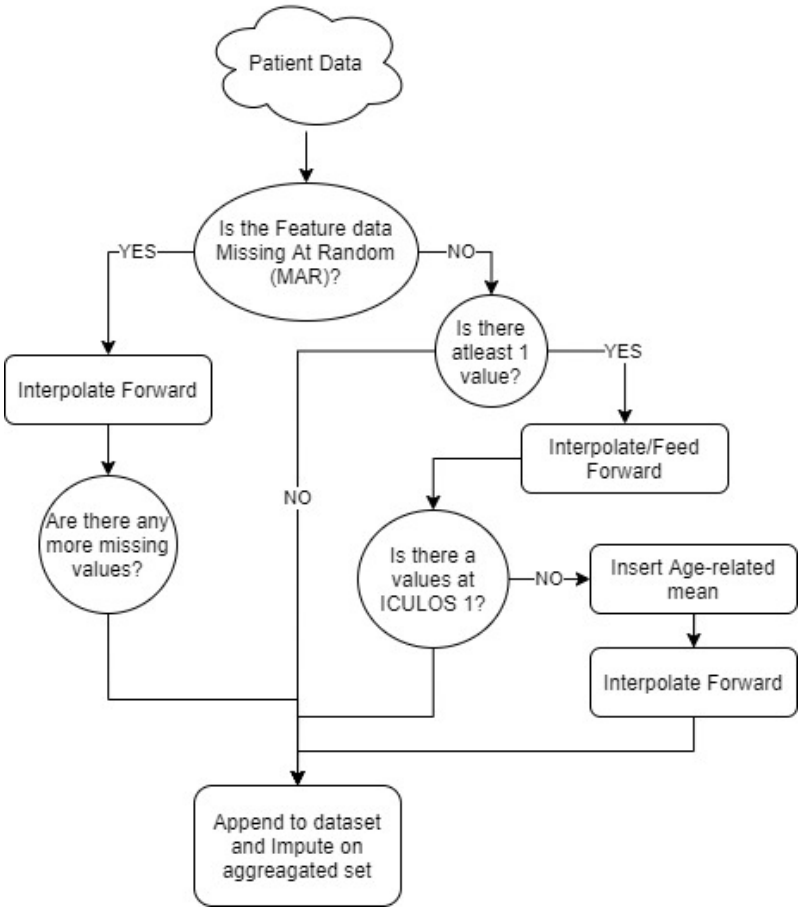


Figure 2.2: Imputation Criteria Flow Diagram

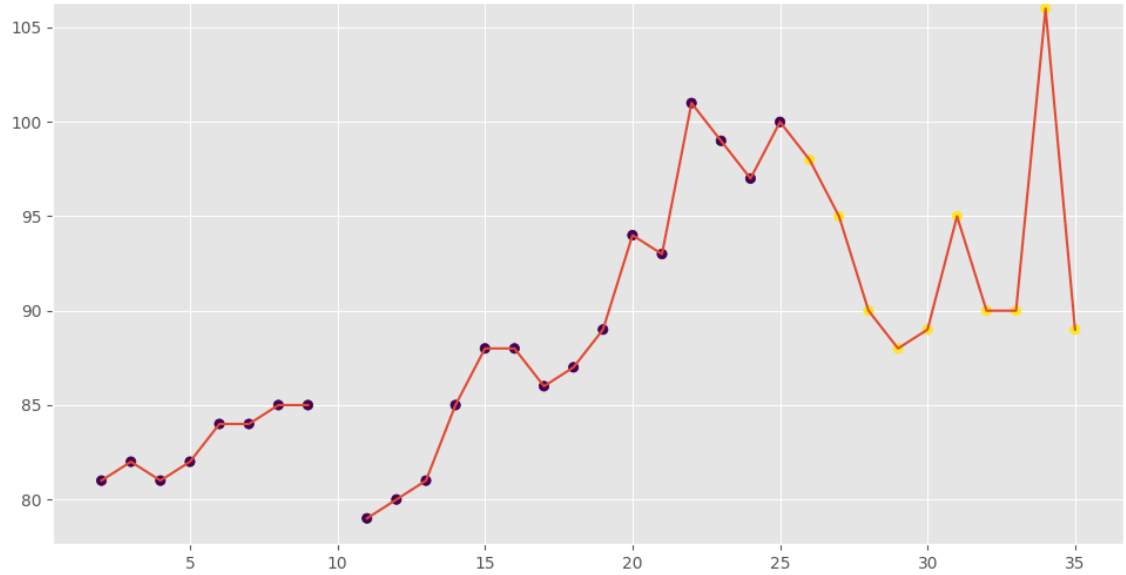
To avoid losing vital or hidden patterns within the data, only non-septic patients were randomly under-sampled. This resulted in a total of 4000 patients being excluded from the study. The high number of excluded patients aids the training of the submodels, which are Long Short Term Memory (LSTM) models, to better discriminate sepsis and non-sepsis patients. An added advantage is that the computational time and resource is reduced in the training of the individual models.

After the individual models were joined in the ensemble model, the full dataset was used in the training and prediction of the full model.

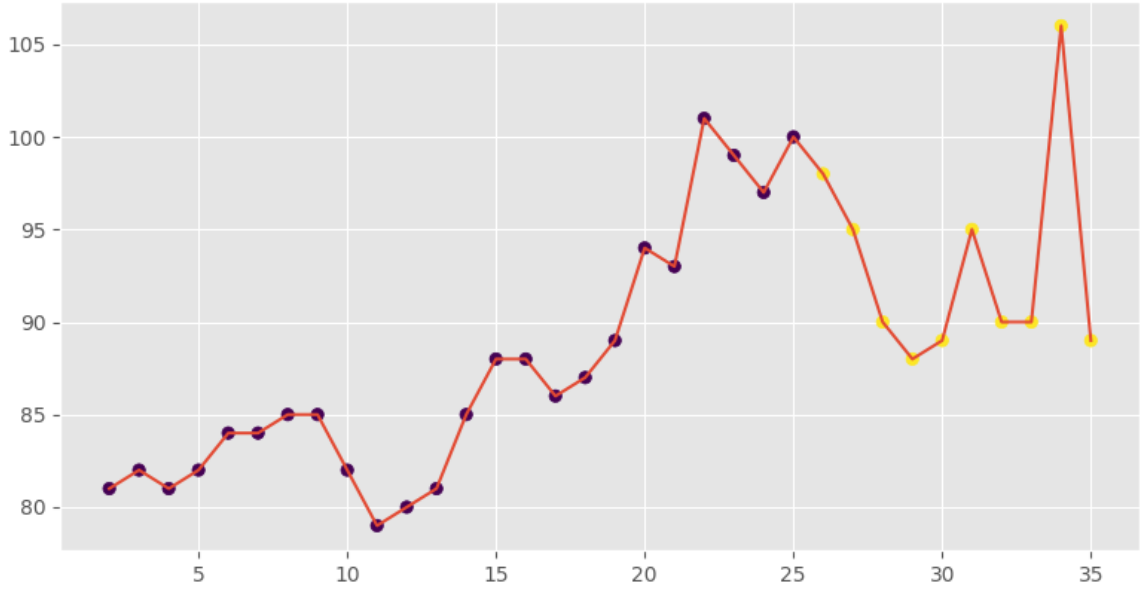
2.2.2 Clinical Interpolation

The following figure, Fig. 2.3, shows how the forward interpolation algorithm works by picking up the previous available data point which is not NA, Null or NAN and linearly interpolated the missing values until the next non-missing value. The linear function allows the model to identify trends such as a deterioration or improvement in a biomarker, as events that occur in a series of time-steps.

It is crucial for the first method of interpolation to be forward so as to not allow the interpolation algorithm to assign inferred values to measures which were not recorded at an earlier point in time. In sum, performing backwards fill first could result in a less clinically accurate measurement for missing especially for key clinical features such as the heart rate.



a) Sepsis Patient Heart Rate Per Hour in the ICU



b) Interpolated Sepsis Patient Heart Rate Per Hour in the ICU

Figure 2.3: A depiction of the Interpolation method for Continuous Data. Yellow dots sow the patient was diagnosed with Sepsis.

2.2.3 K-Nearest Neighbour Imputation

After filling some missing values through forward fill, Fig. B.1 shows that there are still columns containing over over 50% of missing values in both the sepsis and non-sepsis

patients. Since the number of missing data in both group remains 50% or above, imputing data for these columns would result in a large variation which is not clinically accurate. As such, these columns were excluded from the dataset.

The K-Nearest Neighbour (KNN) method was the second step taken for imputing missing data. This method was used in a study for imputing missing data in a genomics dataset [64]. Moreover, a study on the application of clustering methods for missing data [65] presented that KNN imputation method is applicable and is useful when handling medical data.

The KNN algorithm uses the euclidean distance as a similarity metrics. Taking into consideration the full set of features, similarity scores are calculated and totalled across each patient. Individuals with similar scores are considered as nearest nearest neighbours. After obtaining the feature vectors of the the nearest 10 members of an individual, it uses the distance of the feature space as weights to identify the best fitted value for the missing column for imputation. Higher K values is directly proportionate to the level of certainty in the value used for imputation.

To ensure clinical soundness of the imputed values, each biomarker was evaluated against their normal manifestation range after imputation.

2.3 Ensemble Sequential Deep Learning

The final step before presenting the data to LSTM models, is to Z-score all non- categorical variables, to reduce the computation power required to run the algorithm, and to aid the algorithm in detecting changes in continuous variables more effectively.

LSTM have proven to be great at modeling temporal sequences[66], as they manage to overcome the vanishing error problem. They can learn to connect small time lags with over 1000 discrete time steps by applying a constant error through the "constant error carrousels" (CECs) within the LSTM cells. This aids in retaining the general pattern of the dataset.

2.3.1 Architectures

Two LSTM Architectures were chosen to model Sepsis.

A simple Vanilla LSTM consists of a Padding and Masking Layer which allowed for the variable batch_size of patient data. This was followed by a Batch Normalization layer which aids the next layer in interpreting the data better by normalizing the feature vectors.

Following on, two 400 Neuron RELU activated LSTM layers are used to learn the temporal development of Sepsis in patients. These layers are constructed by consecutive neurons with a forget,input and output gate. The forget gate is responsible of forgetting a portion of the information learned from the precious neuron as to avoid over-fitting the model. The input gate is where information to each neuron is fed through, this will then be added on to the information from the previous neuron. A final output gate, will transmit the activated, transformed and aggregated neuron output to the next cell.

The final portion of this network consists of a series of TimeDistributed Dense Layers, where a Dense (fully-connected) operation to every time-step of a 3D tensor is applied. This leads to obtaining a probabilistic output the patient is suffering from sepsis at a given time-step.

A Second Stacked LSTM architecture was also used, as viewed in Fig. 2.4, this structure was obtained from a previously successful LSTM Sepsis classification model[67]. It is quite similar to the Vanilla LSTM, with the single addition of an extra LSTM layers, which allows the algorithm to model a second hidden state of the classification.

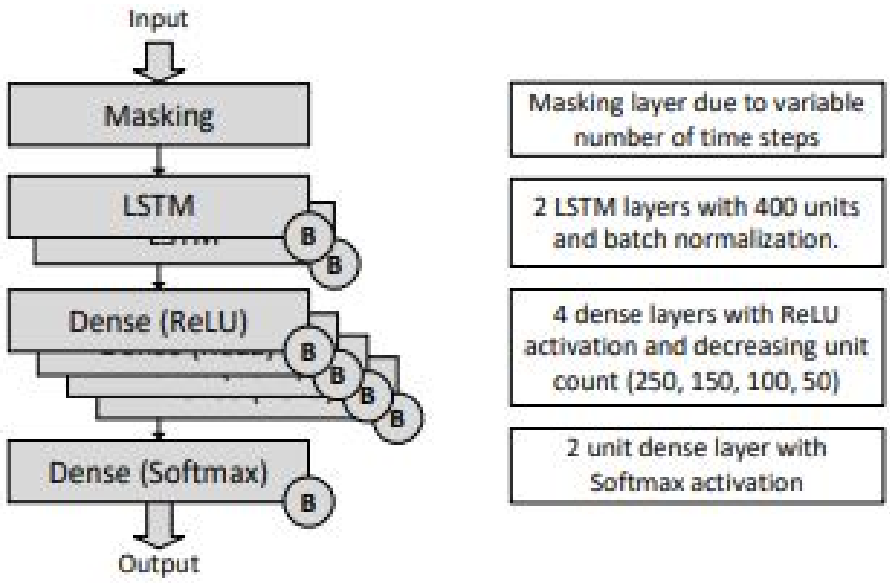


Figure 2.4: Stacked LSTM Architecture

2.3.2 Loss Functions

A novelty of this study is the comparison of four different loss functions, which are then stacked into an ensemble LSTM and used to yield a higher classification rate, by benefiting from the strengths of each loss function.

The function were as follows:

1. Sparse categorical Cross Entropy: the loss is a measure of the dissimilarity between the distribution of observed class labels and the predicted probabilities of class membership.

2. Binary Cross Entropy False Positive weighted[68]: the loss is a measure of the dissimilarity, but the adapted function is additionally weighted with the factor w = 0.7 to emphasize that a false negative prediction is worse than a false positive prediction

3. Weighted Mean Squared Error: As the name suggests, it is the squared difference between the target and the prediction values.the probability for each class is the multiplied by the according weights and summed over each column, to obtain the loss.

4. Sparse Categorical Focal Loss: is a Cross-Entropy Loss that weighs the contribution of each sample to the loss based on the classification error. This allows the function to output a low loss for correctly classified example but a high one for those harder to classify. It is a recent technique that aims to tackle imbalanced datasets.

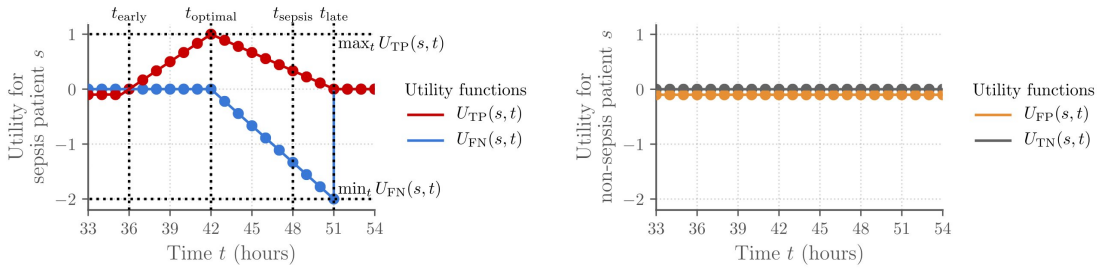
```
def weighted_mse(y_true, y_pred):# (batch_size, 2)

    # calculating squared difference between target and predicted values
    loss = K.square(y_pred - y_true)  # (batch_size, 2)

    # multiplying the values with weights along batch dimension
    loss = loss * [0.3 , 0.7]          # (batch_size, 2)

    # summing both loss values along batch dimension
    loss = K.sum(loss, axis=1)        # (batch_size,)

    return loss
```

Listing 2.2: Weighted mean squared error costum loss function

The code listing above (Listing 2.2) shows an the implementation of the weighted mean squared error function, with class weights attributed to each batch and prediction.

2.3.3 Evaluation Metrics

Different studies often employ different evaluation metrics, and such metrics do not necessarily reflect the clinical utility of sepsis detection and treatment. Traditional scoring metrics, such as area under the curve (AUC) metrics, do not explicitly reward early detection or penalize false alarms or overtreatment. For the Challenge, we devised a novel evaluation metric that addresses these issues and could be generally applicable to predicting infrequent events in time series data.



a) Sepsis Utility Scoring Function

b) Non-Sepsis Utility Scoring Function

Figure 2.5: A depiction of the utility score grading function in Sepsis and Non-Sepsis Patients [58].

Chapter 3

Results and Discussion

3.1 Results

This section aims to present the results compiled from the application of rare event detection methods in the early diagnosis of Sepsis.

All models were trained for 5 epoch, with the patience parameter set to 2. They were set to continuously monitor the loss of the model and where reset at every run, to lead to an effective classification the first time round.

3.1.1 Correlation Matrix and Feature Importance

Fig. 3.1 show a number of positive correlation between, mainly arising from the vital signs and the laboratory values. This further reiterates that the imputation of missing values in this study was quite thorough.

What is astounding, is the fact that SOFA and qSOFA features show a mild correlation with other features. Within a clinical environment the ideal metric would be the clinical scores and the biomarkers would serve as a more in depth assessment of the patient.

Another surprising feature and the only one that lacks a correlation with any of the variables is the Hospital Admission time. This could be due to this study mainly focusing on vital and laboratory events and excluding administrative transitions. It could also show a lack of correlation between the time that patients where admitted to ICU and their prognosis.

Fig. 3.2 is a depiction of a Decision tree classifier, which yields the top six most import features to be the Length of Stay in the ICU, The fraction of inspired oxygen, which is

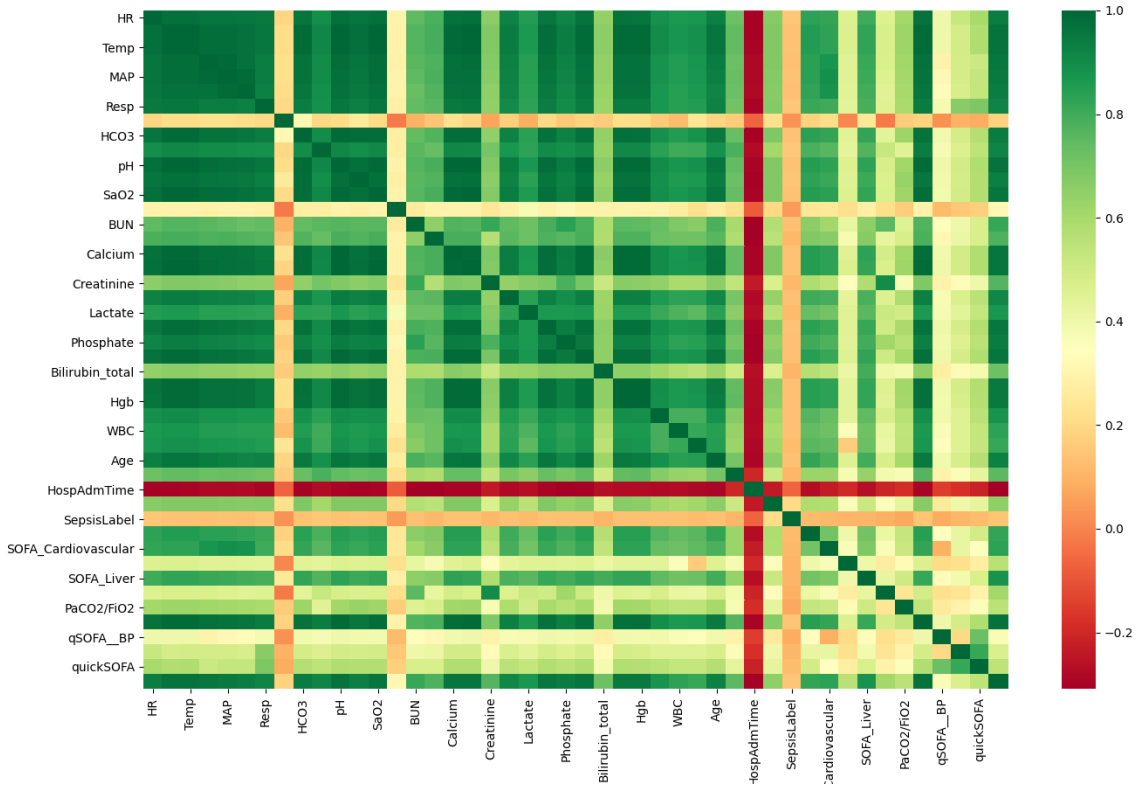


Figure 3.1: Correlation matrix of all model features

followed right after by th ration of temperature and then the PaCO2/FiO2. The last two are Bilirubin total and Lactate.

These are evidently features that appear in literature supporting the early diagnosis of sepsis, which asserts that the data preparation model has successfully modelled the original pattern of the data.

3.1.2 Model Evaluation

Even after tedious hours, Fig. 3.3 showcases that the results of the classifier are much to low to be industrially competitive. This is mainly evident in the low precision, which signals a high number of False Positive and False Negatives.

One variant of the model managed to achieve a higher precision of 30% but this was at the cost of misclassifying over 3% of sepsis rows.

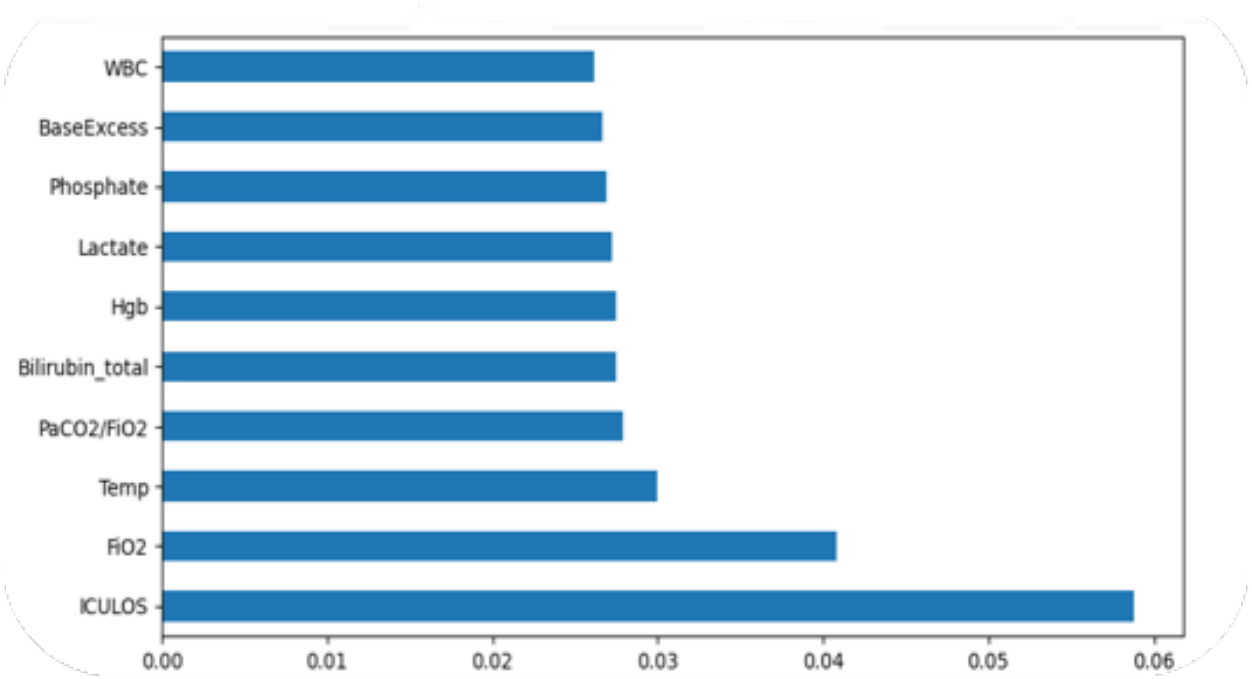


Figure 3.2: Feature Importance for of all model features

3.2 Discussion

From the analysis, we can conclude that there is evidence to believe that the data was imputed well and that there are correlations between variables. Although, it is true that the decision tree generated very low importance scores for each of the features.

So, breaking this down, the initial variables Length of Stay in the ICU. This is because during analysis of the dataset, evidence arose that patients suffering from a septic shock or septicemia where monitored for longer hours in the ICU. As sepsis is a multi systemic disorder, patients who have a septic shock are prone to sequential organ failure, which gives reason to believe that ICU length of stay is strongly correlated with being diagnosed with Sepsis[69].

The second most important variable is temperature and it has a strong correlation with most variables. This also has a strong correlation with Sepsis diagnosis. As literature suggest that a high fever is one of the first indicators of sepsis[33].

The next variable that shows a high importance and a correlation with other features is the fraction of inspired oxygen (FiO2) which combined with the PaO2 forms the third

		Minority Class			
Model Structure	Loss Function	Precision	Recall	Utility Score	Time to Train per epoch
Vanilla	Sparse Categorical Cross Entropy	23%	21%	0.18	8 minutes
	Weighted Mean Squared Error	20%	23%	0.17	5 minutes
	Focal Loss	30%	18%	0.18	26 minutes
Stacked	Sparse Categorical Cross Entropy	23%	21%	0.21	12 minutes
	Weighted Mean Squared Error	20%	23%	0.19	10 minutes
	Focal Loss	30%	18%	0.21	30 minutes
Ensembled	Sparse Categorical Cross Entropy	23%	23%	0.25	6 hours
	Weighted Mean Squared Error	23%	23%	0.25	3 hours
	Focal Loss	23%	21%	0.25	10 hours

Figure 3.3: Results for all the models

most important feature,according to the decision tree classifier.This ratio of PaO2 to FiO2 is known as the SOFA score for the respiratory systems, which allows to deduce if the patients lungs are functioning well[70] and highly correlated with sepsis diagnosis.

Again the final two values are Bilirubin Total and Lactate, both of these have an average importance and both form the basis of the SOFA score for kidney[71] function, with lactate, recently being recognises as a valuable biomarker for sepsis.

As to conclude the discussion, it is evident that the data exhibits some of literature related patterns and that if the feature importance yielded a higher importance for the confounding variables, then the prediction model would be more accurate.

The study also faced some limitations such as long training times, low recall and the structure of the data presented a challenge when embedding sequential models. Although with the resources available, a utility score of 0.25 was achieved.

Chapter 4

Conclusions and Future Work

In conclusion of this study, it was found that rare event detection method have the potential to tackle highly imbalanced datasets, but that a through data preparation stage is required. This study focused on modelling the hourly sequential nature of, which is quite different to the windowing time approach that most other studies attempt at early diagnosis of sepsis. This goes to show that the algorithm was able to correctly fit a sepsis model but has easily skewed of the normal and tends to over-fit or under-fit.

The new framework to imputing missing values, puts a great deal of importance on clinical accuracy, to retain the original significance of the data, which showcases responsible innovation in Artificial intelligence, where the human is given a higher importance. This is viewed to generate the most clinically competent version of this dataset studied, as it ranks the correct features according to their clinical importance.Added on to the inclusion of clinical scores, used currently in intensive care units, such as the SIRS, qSOfa and SOFA, which although not at the top did influence the diagnosis.

On the other hand the effectiveness of the algorithm may not imply industry compliance,this study aims to present a base line model for hourly classification of sepsis in sequential patient records, within a highly imbalanced dataset.

4.1 Future Work

Future Work, will entail revisiting the data preparations steps and applying a 6,12,24 hour windowing function. This should yield a higher clinical utility score, due to the change in the patients condition being more evident over time. An addition of other clinical scores

for sepsis detection, such as SIRS, which is defined under the new guidelines for quick sepsis diagnosis and could aid in improving the prediction[55].

Another area to focus future work on, is in testing customising our own focus loss function to be capable of tuning the cross entropy loss contribution parameter and adjust it to resemble a healthcare scenario, where false negatives are more heavily penalised than false positives. As well as defining a modifiable parameter for the class weight given to the probabilities.

Additionally, the need to experiment with more complex LSTM structures, may yield a better retention of information. Advances in Representation Learning also bring a new wave of rare event detection method, which may offer a visual boundary of the conditions[72]. This combined with deep learning interpretation techniques such as LIME and Shapely Values, could stand as the perfect representation of an explainable temporal neural network.

When access to the Fabry patient dataset is available, future work will look at analysing ECG data, specifically the left ventricular hypertrophy (LVH) and the recently emerging biomarker T1 mapping[73].This will involve real-time signal processing of ECG data combined with rare event detection, which may prove to yield novel results.

Overall, rare event detection is still a relatively new field in the area of deep learning and still has a vast pool of opportunities to explore, with this study serving as the baseline to evaluate against.

Bibliography

[1] G.-Z. Yang, B. J. Nelson, R. R. Murphy, H. Choset, H. Christensen, S. H. Collins, P. Dario, K. Goldberg, K. Ikuta, N. Jacobstein, D. Kragic, R. H. Taylor, and M. McNutt, "Combating covid-19—the role of robotics in managing public health and infectious diseases," *Science Robotics*, vol. 5, no. 40, 2020. [Online]. Available: https://robotics.sciencemag.org/content/5/40/eabb5589

[2] D. Cirillo and A. Valencia, "Big data analytics for personalized medicine," *Current opinion in biotechnology*, vol. 58, pp. 161–167, 2019.

[3] D. S. Allen, "2020 global health care outlook," vol. 2020, 2020. [Online]. Available: https://www2.deloitte.com/global/en/pages/life-sciences-and-healthcare/articles/global-health-care-sector-outlook.html

[4] W. H. Organization *et al.*, *Global recommendations on physical activity for health*. World Health Organization, 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK305057/

[5] J. A. Dodge, P. A. Lewis, M. Stanton, and J. Wilsher, "Cystic fibrosis mortality and survival in the uk: 1947–2003," *European Respiratory Journal*, vol. 29, no. 3, pp. 522–526, 2007. [Online]. Available: https://erj.ersjournals.com/content/29/3/522

[6] S. Nguengang Wakap, D. M. Lambert, A. Olry, C. Rodwell, C. Gueydan, V. Lanneau, D. Murphy, Y. Le Cam, and A. Rath, "Estimating cumulative point prevalence of rare diseases: analysis of the orphanet database," *European Journal of Human Genetics*, vol. 28, no. 2, pp. 165–173, Feb 2020. [Online]. Available: https://doi.org/10.1038/s41431-019-0508-0

[7] I. of Medicine, *Rare Diseases and Orphan Products: Accelerating Research and Development*, M. J. Field and T. F. Boat, Eds. Washington, DC: The National

Academies Press, 2010. [Online]. Available: https://www.nap.edu/catalog/12953/rare-diseases-and-orphan-products-accelerating-research-and-development

[8] C. Hendriksz, "Rare disease impact report: Insights from patients and the medical community," 06 2013.

[9] W. R. Lumry, "Hereditary angioedema: The economics of treatment of an orphan disease," *Frontiers in medicine*, vol. 5, pp. 22–22, Feb 2018, pMC5820358[pmcid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29503818

[10] S. A. Tilles, L. Borish, and J. P. Cohen, "Management of hereditary angioedema in 2012: scientific and pharmacoeconomic perspectives," *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology*, vol. 110, no. 2, pp. 70–74, Feb 2013, s1081-1206(12)00939-8[PII]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23352523

[11] R. Desnick, Y. Ioannou, C. Eng, C. Scriver, A. Beaudet, and W. Sly, "The metabolic and molecular bases of inherited disease," *by Scriver CR, Beaude AL, Sly WS, Valle D., McGraw-Hill, New York*, pp. 3733–3774, 2001.

[12] P. J. Meikle, J. J. Hopwood, A. E. Clague, and W. F. Carey, "Prevalence of lysosomal storage disorders," *Jama*, vol. 281, no. 3, pp. 249–254, 1999.

[13] G. Genoni, I. Demarchi, S. Bellone, A. Petri, F. Settanni, E. Dondi, M. Negro, L. Cortese, F. Prodam, and G. Bona, "Early diagnosis of fabry disease in children," *Minerva pediatrica*, vol. 63, pp. 425–30, 10 2011.

[14] C. Marchesoni, N. Roa, A. Pardal, P. Neumann, G. Cáceres, P. Martínez, I. Kisinovsky, S. Bianchi, A. Tarabuso, and R. Reisin, "Misdiagnosis in fabry disease," *The Journal of pediatrics*, vol. 156, pp. 828–31, 05 2010.

[15] P. Engel, S. Bagal, M. Broback, and N. Boice, "Physician and patient perceptions regarding physician training in rare diseases: The need for stronger educational initiatives for physicians," *J Rare Dis*, vol. 1, pp. 1–15, 01 2013.

[16] L. B. N. d. Silva, T. C. d. M. T. Badiz, M. M. S. e. S. Enokihara, and A. M. Porro, "Fabry disease: clinical and genotypic aspects of three cases in first degree relatives," *Anais brasileiros de dermatologia*, vol. 89, no. 1, pp. 141–143, 2014, 24626659[pmid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24626659

[17] S. Bryson, "Is fabry disease going to affect my life expectancy?" 2019. [Online]. Available: https://fabrydiseasenews.com/2019/10/21/is-fabry-disease-going-to-affect-my-life-expectancy/

[18] K. MacDermot, A. Holmes, and A. Miners, "Anderson-fabry disease: clinical manifestations and impact of disease in a cohort of 98 hemizygous males," *Journal of medical genetics*, vol. 38, no. 11, pp. 750–760, 2001.

[19] ——, "Anderson-fabry disease: clinical manifestations and impact of disease in a cohort of 60 obligate carrier females," *Journal of medical genetics*, vol. 38, no. 11, pp. 769–775, 2001.

[20] R. El-Abassi, D. Singhal, and J. D. England, "Fabry's disease," *Journal of the Neurological Sciences*, vol. 344, no. 1, pp. 5–19, Sep 2014. [Online]. Available: https://doi.org/10.1016/j.jns.2014.06.029

[21] R. Desnick, M. Banikazemi, and M. Wasserstein, "Enzyme replacement therapy for fabry disease, an inherited nephropathy," *Clinical nephrology*, vol. 57, pp. 1–8, 02 2002.

[22] R. J. Desnick and R. O. Brady, "Fabry disease in childhood," *The Journal of Pediatrics*, vol. 144, no. 5, pp. S20–S26, May 2004. [Online]. Available: https://doi.org/10.1016/j.jpeds.2004.01.051

[23] M. Ries, U. Ramaswami, R. Parini, B. Lindblad, C. Whybra, I. Willers, A. Gal, and M. Beck, "The early clinical phenotype of fabry disease: A study on 35 european children and adolescents," *European journal of pediatrics*, vol. 162, pp. 767–72, 12 2003.

[24] C. A. Luciano, J. W. Russell, T. K. Banerjee, J. M. Quirk, L. J. Scott, J. M. Dambrosia, N. W. Barton, and R. Schiffmann, "Physiological characterization of neuropathy in fabry's disease," *Muscle & Nerve*, vol. 26, no. 5, pp. 622–629, 2002. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mus.10236

[25] F. Birklein, "[mechanism-based treatment principles of neuropathic pain]." *Fortschritte der Neurologie-Psychiatrie*, vol. 70, pp. 88–94, 03 2002.

[26] L. Lao, M. Kumakiri, H. Mima, H. Kuwahara, H. Ishida, K. Ishiguro, T. Fujita, and K. Ueda, "The ultrastructural characteristics of eccrine sweat glands in a fabry disease patient with hypohidrosis," *Journal of dermatological science*, vol. 18, pp. 109–17, 12 1998.

[27] B. Chia and H. L. Tey, "Approach to hypohidrosis," *Journal of the European Academy of Dermatology and Venereology : JEADV*, vol. 27, 10 2012.

[28] E. Šabović, M. Tansek, U. Groselj, and V. Dragoš, "Angiokeratomas and treatment with enzyme replacement therapy in a patient with fabry disease," *Acta Dermatovenerologica Alpina Pannonica et Adriatica*, vol. 29, 06 2020.

[29] B. O'Brien, T. Shnitka, R. McDougall, K. Walker, L. Costopoulos, B. Lentle, L. Anholt, H. Freeman, and A. Thomson, "Pathophysiologic and ultrastructural basis for intestinal symptoms in fabry's disease," *Gastroenterology*, vol. 82, pp. 957–62, 06 1982.

[30] C. Kampmann, F. Baehner, C. Whybra, C. Martin, C. Wiethoff, M. Ries, A. Gal, and M. Beck, "Cardiac manifestation of anderson-fabry disease in heterozygous females," *Journal of the American College of Cardiology*, vol. 40, pp. 1668–74, 12 2002.

[31] N. Yuen, C.-W. Lam, T.-C. Chow, and M.-C. Chiu, "A characteristic dissection microscopy appearance of a renal biopsy of a fabry heterozygote," *Nephron*, vol. 77, pp. 354–6, 02 1997.

[32] M. El-Shahawy, C. Mesa, M. Koss, and V. Campese, "A 19yearold female with fever, acroparesthesia, and progressive deterioration of renal function," *American Journal of Nephrology - AMER J NEPHROL*, vol. 16, pp. 417–424, 09 1996.

[33] K. MacDermot, A. Holmes, and A. Miners, "Anderson-fabry disease: clinical manifestations and impact of disease in a cohort of 60 obligate carrier females," *Journal of medical genetics*, vol. 38, pp. 769–75, 12 2001.

[34] R. Schiffmann, J. Kopp, H. Austin, S. Sabnis, D. Moore, T. Weibel, J. Balow, and R. Brady, "Enzyme replacement therapy in fabry disease: A randomized controlled trial," *JAMA : the journal of the American Medical Association*, vol. 285, pp. 2743–9, 07 2001.

[35] M. Ries, K. Wendrich, C. Whybra, C. Kampmann, A. Gal, and M. Beck, "Angiokeratoma and pain, but not fabry's disease: Considerations for differential diagnosis," *Contributions To Nephrology - CONTRIB NEPHROL*, vol. 136, pp. 256–259, 10 2001.

[36] M. Ries, E. Mengel, G. Kutschke, K. Kim, F. Birklein, F. Krummenauer, and M. Beck, "Use of gabapentin to reduce chronic neuropathic pain in fabry disease," *Journal of inherited metabolic disease*, vol. 26, pp. 413–4, 02 2003.

[37] C. Whybra, C. Kampmann, I. Willers, J. Davies, B. Winchester, J. Kriegsmann, K. Brühl, A. Gal, S. Bunge, and M. Beck, "Anderson-fabry disease: Clinical manifestations of disease in female heterozygotes," *Journal of inherited metabolic disease*, vol. 24, pp. 715–24, 01 2002.

[38] N. Yuen, C.-W. Lam, T.-C. Chow, and M.-C. Chiu, "A characteristic dissection microscopy appearance of a renal biopsy of a fabry heterozygote," *Nephron*, vol. 77, pp. 354–6, 02 1997.

[39] S. Cakir and A. Otunctemur, "Artificial intelligence in medicine," *European Archives of Medical Research*, vol. 34, pp. 1–3, 12 2018.

[40] J. Schaefer, M. Lehne, J. Schepers, F. Prasser, and S. Thun, "The use of machine learning in rare diseases: a scoping review," *Orphanet Journal of Rare Diseases*, vol. 15, no. 1, p. 145, Jun 2020. [Online]. Available: https://doi.org/10.1186/s13023-020-01424-6

[41] S. Kaushik, A. Choudhury, P. K. Sheron, N. Dasgupta, S. Natarajan, L. A. Pickett, and V. Dutt, "Ai in healthcare: Time-series forecasting using statistical, neural, and ensemble architectures," *Frontiers in Big Data*, vol. 3, p. 4, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fdata.2020.00004

[42] X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J. R. Ledsam, M. K. Schmid, K. Balaskas, E. J. Topol, L. M. Bachmann, P. A. Keane, and A. K. Denniston, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The Lancet Digital Health*, vol. 1, no. 6, pp. e271 – e297, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2589750019301232

[43] I. Tobore, J. Li, L. Yuhang, Y. Al-Handarish, A. Kandwal, Z. Nie, and L. Wang, "Deep learning intervention for health care challenges: Some biomedical domain considerations," *JMIR mHealth and uHealth*, vol. 7, no. 8, p. e11966, Aug. 2019. [Online]. Available: https://doi.org/10.2196/11966

[44] S. Ronicke, M. C. Hirsch, E. Türk, K. Larionov, D. Tientcheu, and A. D. Wagner, "Can a decision support system accelerate rare disease diagnosis? evaluating the potential impact of ada dx in a retrospective study," *Orphanet journal of rare diseases*, vol. 14, no. 1, p. 69, 2019.

[45] S. Aymé, B. Urbero, D. Oziel, E. Lecouturier, and A. Biscarat, "Information on rare diseases: The orphanet project," *La Revue de médecine interne / fondée ... par la Société nationale francaise de médecine interne*, vol. 19 Suppl 3, pp. 376S–377S, 02 1998.

[46] A. Mehta, Y. Enever, S. Evans, and L. Richfield, "Fabry outcome survey: the uk implementation experience," *Acta Paediatrica - ACTA PAEDIAT*, vol. 91, pp. 133–133, 11 2007.

[47] K. Jones, D. Ford, S. Thompson, and R. Lyons, "A profile of the sail databank on the uk secure research platform," *International Journal of Population Data Science*, vol. 4, 11 2019.

[48] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J.-L. Vincent, and D. C. Angus, "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801–810, 02 2016. [Online]. Available: https://doi.org/10.1001/jama.2016.0287

[49] M. Gupta, P. Bansal, and P. Sehrawat, "Sepsis and septic shock," *Anaesthesia, Pain Intensive Care*, 01 2020.

[50] K. Reinhart, R. Daniels, N. Kissoon, F. Machado, R. Schachter, and S. Finfer, "Recognizing sepsis as a global health priority — a who resolution," *New England Journal of Medicine*, vol. 377, 06 2017.

[51] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J.-L. Vincent, and D. C. Angus, "The third international consensus definitions for sepsis and septic shock (sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801–810, Feb 2016, pMC4968574[pmcid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26903338

[52] J. Calvert, D. Price, U. Chettipally, C. Barton, M. Feldman, J. Hoffman, M. Jay, and R. Das, "A computational approach to early sepsis detection," *Computers in Biology and Medicine*, vol. 74, 05 2016.

[53] J.-L. Vincent, S. M. Opal, J. C. Marshall, and K. J. Tracey, "Sepsis definitions: time for change," *The Lancet*, vol. 381, no. 9868, pp. 774–775, Mar. 2013. [Online]. Available: https://doi.org/10.1016/s0140-6736(12)61815-7

[54] J. Harrington, "Table 1. sirs criteria - acep now," 2020. [Online]. Available: https://www.acepnow.com/article/focus-on-systemic-inflammatory-response-syndrome-can-interfere-with-early-sepsis-detection/feature-story_pg19c-3/

[55] P. E. Marik and A. M. Taeb, "Sirs, qsofa and new sepsis definition," *Journal of thoracic disease*, vol. 9, no. 4, p. 943, 2017.

[56] J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. Reinhart, P. Suter, and L. G. Thijs, "The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure," 1996.

[57] M. Zhang, F. Yin, B. Chen, Y. Li, L. Yan, T. Wen, and B. Li, "Pretransplant prediction of posttransplant survival for liver recipients with benign end-stage liver diseases: A nonlinear model," *PloS one*, vol. 7, p. e31256, 03 2012.

[58] M. A. Reyna, C. Josef, S. Seyedi, R. Jeter, S. P. Shashikumar, M. B. Westover, A. Sharma, S. Nemati, and G. D. Clifford, "Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019," in *2019 Computing in Cardiology (CinC)*. IEEE, 2019, pp. Page–1.

[59] S. Fielding, P. M. Fayers, A. McDonald, G. McPherson, M. K. Campbell, R. S. Group *et al.*, "Simple imputation methods were inadequate for missing not at random (mnar) quality of life data," *Health and Quality of Life Outcomes*, vol. 6, no. 1, p. 57, 2008.

[60] K. Wang, S. Xie, K. Xiao, P. Yan, W. He, and L. Xie, "Biomarkers of sepsis-induced acute kidney injury," *BioMed Research International*, vol. 2018, p. 6937947, Apr 2018. [Online]. Available: https://doi.org/10.1155/2018/6937947

[61] N. M. Paige and G. T. Nagami, "The top 10 things nephrologists wish every primary care physician knew," *Mayo Clinic proceedings*, vol. 84, no. 2, pp. 180–186, Feb 2009, pMC2664589[pmcid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19181652

[62] "dealing with missing values in healthcare data," 2020. [Online]. Available: https://www.biosymetrics.com/missing-values-healthcare-data/

[63] M. Ramsey, "Noninvasive automatic determination of mean arterial pressure," *Medical and Biological Engineering and Computing*, vol. 17, no. 1, pp. 11–18, Jan 1979. [Online]. Available: https://doi.org/10.1007/BF02440948

[64] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for DNA microarrays ," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 06 2001. [Online]. Available: https://doi.org/10.1093/bioinformatics/17.6.520

[65] R. Ananda, A. R. Dewi, and N. Nurlaili, "A comparison of clustering by imputation and special clustering algorithms on the real incomplete data," *Jurnal Ilmu Komputer dan Informasi*, vol. 13, no. 2, pp. 65–75, 2020.

[66] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with lstm recurrent neural networks," *arXiv preprint arXiv:1511.03677*, 2015.

[67] S. Schellenberger, K. Shi, J. P. Wiedemann, F. Lurz, R. Weigel, and A. Koelpin, "An ensemble lstm architecture for clinical sepsis detection," in *2019 Computing in Cardiology (CinC)*, 2019, pp. Page 1–Page 4.

[68] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[69] K. R. Genga and J. A. Russell, "Update of sepsis in the intensive care unit," *Journal of Innate Immunity*, vol. 9, no. 5, pp. 441–455, 2017. [Online]. Available: https://www.karger.com/DOI/10.1159/000477419

[70] R. M. Dahl, L. Grønlykke, N. Haase, L. B. Holst, A. Perner, J. Wetterslev, B. S. Rasmussen, C. S. Meyhoff, and the 6S-Trial and TRISS Trial investigators , "Variability in targeted arterial oxygenation levels in patients with severe sepsis or septic shock," *Acta Anaesthesiologica Scandinavica*, vol. 59, no. 7, pp. 859–869, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/aas.12528

[71] N. Nesseler, Y. Launey, C. Aninat, F. Morel, Y. Mallédant, and P. Seguin, "Clinical review: The liver in sepsis," *Critical care (London, England)*, vol. 16, no. 5, pp. 235–235, Oct 2012, 23134597[pmid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23134597

[72] R. Hamaguchi, K. Sakurada, and R. Nakamura, "Rare event detection using disentangled representation learning," *CoRR*, vol. abs/1812.01285, 2018. [Online]. Available: http://arxiv.org/abs/1812.01285

[73] R. Kozor, S. Nordin, A. Abdel-Gadir, H. Bulluck, T. A. Treibel, C. Manisty, and J. C. Moon, "Ecg, lvh and t1 changes in fabry disease - implications for screening and understanding of the disease model," *Journal of Cardiovascular Magnetic Resonance*, vol. 18, no. Suppl 1, p. Q48, Jan 2016, pMC5032065[pmcid]. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5032065/

Appendix A

Implementation of a Relevant Algorithm

```
def utility(y_true, y_pred):

    # Set parameters
    dt_early   = -12
    dt_optimal = -6
    dt_late    = 3

    max_u_tp = 1
    min_u_fn = -2
    u_fp     = -0.05
    u_tn     = 0

    # Load labels and predictions.

    labels = y_true

    if y_pred.shape[1] == 2:
        predictions = tf.math.argmax(y_pred, axis=1).numpy()
    elif y_pred.shape[1] == 1:
        predictions = y_pred


    #unique_elements, counts_elements = np.unique(predictions, return_counts=True)
    #print()
    #print(np.asarray((unique_elements, counts_elements)))
    #print()

    # Check labels and predictions for errors.
```

```
29      if not (labels.shape[0] == predictions.shape[0]):
30          raise Exception('Numbers of labels and predictions for a file must be the
                same.')
31
32      num_rows = labels.shape[0]
33
34      def compute_prediction_utility(labels, predictions, dt_early=-12, dt_optimal=-6,
            dt_late=3.0, max_u_tp=1, min_u_fn=-2, u_fp=-0.05, u_tn=0, check_errors=True):
35
36          # Does the patient eventually have sepsis?
37          if np.any(labels):
38              is_septic = True
39              t_sepsis = np.argmax(labels) - dt_optimal
40          else:
41              is_septic = False
42              t_sepsis = float('inf')
43
44          n = labels.shape[0]
45
46          # Define slopes and intercept points for utility functions of the form
47          # u = m * t + b.
48          m_1 = float(max_u_tp) / float(dt_optimal - dt_early)
49          b_1 = -m_1 * dt_early
50          m_2 = float(-max_u_tp) / float(dt_late - dt_optimal)
51          b_2 = -m_2 * dt_late
52          m_3 = float(min_u_fn) / float(dt_late - dt_optimal)
53          b_3 = -m_3 * dt_optimal
54
55          # Compare predicted and true conditions.
56          u = np.zeros(n)
57          for t in range(n):
58              if t <= t_sepsis + dt_late:
59                  # TP
60                  if is_septic and predictions[t] == 1:
61                      if t <= t_sepsis + dt_optimal:
62                          u[t] = max(m_1 * (t - t_sepsis) + b_1, u_fp)
63                      elif t <= t_sepsis + dt_late:
64                          u[t] = m_2 * (t - t_sepsis) + b_2
65                  # FP
66                  elif not is_septic and predictions[t] == 1:
67                      u[t] = u_fp
68                  # FN
69                  elif is_septic and not predictions[t] == 0:
70                      if t <= t_sepsis + dt_optimal:
71                          u[t] = 0
72                      elif t <= t_sepsis + dt_late:
73                          u[t] = m_3 * (t - t_sepsis) + b_3
```

```
                # TN
                elif not is_septic and not predictions[t] == 0:
                    u[t] = u_tn


        # Find total utility for patient.
        return np.sum(u)

    # Compute utility
    observed_utilities = np.zeros(1)
    best_utilities     = np.zeros(1)
    worst_utilities    = np.zeros(1)
    inaction_utilities = np.zeros(1)

    best_predictions = np.zeros(num_rows)
    worst_predictions = np.zeros(num_rows)
    inaction_predictions = np.zeros(num_rows)

    if np.any(labels):
        t_sepsis = np.argmax(labels) - dt_optimal
        best_predictions[max(0, t_sepsis + dt_early) : min(t_sepsis + dt_late + 1,
            num_rows)] = 1
        worst_predictions = 1 - best_predictions

    observed_utilities = compute_prediction_utility(labels, predictions, dt_early,
        dt_optimal, dt_late, max_u_tp, min_u_fn, u_fp, u_tn)
    best_utilities = compute_prediction_utility(labels, best_predictions, dt_early,
        dt_optimal, dt_late, max_u_tp, min_u_fn, u_fp, u_tn)
    worst_utilities = compute_prediction_utility(labels, worst_predictions, dt_early,
         dt_optimal, dt_late, max_u_tp, min_u_fn, u_fp, u_tn)
    inaction_utilities = compute_prediction_utility(labels, inaction_predictions,
        dt_early, dt_optimal, dt_late, max_u_tp, min_u_fn, u_fp, u_tn)


    normalized_observed_utility = (observed_utilities - inaction_utilities) / (
        best_utilities - inaction_utilities)

    eval = best_utilities - observed_utilities

    return eval
```

Listing A.1: An implementation a modified version of the evaluation metric that calculates clinical utility.

Appendix B

Supplementary Data

Missing Values	Initially	After Interpolation			
		Sepsis	Non-Sepsis	Total	Exclude
HR	7,7%	0,0%	0,0%	0,0%	
O2Sat	12,0%	0,1%	0,0%	0,0%	
Temp	66,2%	1,3%	0,8%	0,9%	
SBP	15,2%	6,1%	1,2%	1,3%	
MAP	10,2%	0,0%	0,0%	0,0%	
DBP	48,1%	6,1%	1,2%	1,3%	
Resp	9,8%	0,1%	0,1%	0,1%	
EtCO2	100,0%	100,0%	100,0%	1,0%	Yes
BaseExcess	89,6%	24,7%	40,4%	40,1%	
HCO3	92,0%	14,0%	16,6%	16,6%	
FiO2	85,8%	23,5%	42,6%	42,2%	
pH	88,5%	23,3%	38,5%	38,1%	
PaCO2	91,2%	25,2%	41,0%	40,7%	
SaO2	95,0%	59,0%	65,2%	65,1%	
AST	98,5%	57,3%	75,6%	75,2%	
BUN	91,8%	13,9%	16,1%	16,1%	
Alkalinephos	98,5%	57,8%	76,2%	75,8%	
Calcium	95,0%	19,6%	34,3%	34,0%	
Chloride	91,7%	13,9%	16,7%	16,6%	
Creatinine	93,4%	14,8%	17,8%	17,8%	
Bilirubin_direct	99,9%	94,8%	97,6%	97,5%	Yes
Glucose	87,8%	13,0%	14,0%	13,9%	
Lactate	96,6%	39,8%	64,9%	64,3%	
Magnesium	92,2%	14,9%	22,7%	22,6%	
Phosphate	95,0%	19,3%	33,7%	33,4%	
Potassium	89,1%	12,3%	15,0%	14,9%	
Bilirubin_total	98,8%	58,1%	76,4%	76,0%	
TroponinI	99,9%	98,5%	97,9%	97,9%	Yes
Hct	88,2%	12,8%	13,2%	13,2%	
Hgb	91,2%	14,8%	16,3%	16,2%	
PTT	95,2%	24,2%	33,1%	32,9%	
WBC	92,5%	15,6%	18,3%	18,3%	
Fibrinogen	99,2%	81,5%	88,6%	88,5%	Yes
Platelets	93,5%	15,9%	18,2%	18,1%	
Age	0,0%	0,0%	0,0%	0,0%	
Gender	0,0%	0,0%	0,0%	0,0%	
Unit1	48,9%			48,9%	Yes
Unit2	48,9%			48,9%	Yes
HospAdmTime	0,0%	0,0%	0,0%	0,0%	
ICULOS	0,0%	0,0%	0,0%	0,0%	
SepsisLabel	0,0%	0,0%	0,0%	0,0%	

Figure B.1: Missing Values before and after interpolation