UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL DOUTORADO

DENISE BANDEIRA

DEEPSIGNS:
A PREDICTIVE MODEL BASED ON DEEP LEARNING FOR THE EARLY DETECTION OF PATIENT HEALTH DETERIORATION

São Leopoldo
2020

Denise Bandeira

DEEPSIGNS:
A Predictive Model Based on Deep Learning for the Early Detection of Patient Health Deterioration

Dissertation presented as a partial requirement to obtain the Doctor's degree from the Applied Computing Graduate Program of the Universidade do Vale do Rio dos Sinos - UNISINOS

Advisor:
Prof. Dr. Cristiano André da Costa

São Leopoldo
2020

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Bandeira, Denise

DeepSigns: a predictive model based on deep learning for the early detection of patient health deterioration / Denise Bandeira — 2020.

74 f.: il.; 30 cm.

Tese (doutorado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, 2020.

"Advisor: Prof. Dr. Cristiano André da Costa, Unidade Acadêmica de Pesquisa e Pós-Graduação".

1. Predictive Scores. 2. Machine Learning. 3. Deep Learning. 4. LSTM. 5. Health Informatics. I. Título.

CDU 004.732

Bibliotecário responsável: Alessandro Dietrich — CRB 10/2338

(Esta folha serve somente para guardar o lugar da verdadeira folha de aprovação, que é obtida após a defesa do trabalho. Este item é obrigatório, exceto no caso de TCCs.)

To my father.

One study concluded that there were 1.7 errors per patient per day in America's ICUs. Of these errors, 29 percent could have caused clinically significant harm or death. Given that the average ICU length of stay is three days, this research suggests that nearly all patients hospitalized in the ICU sustain a potentially life-threatening mistake at some point during their stay,

— PETER PRONOVOST

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor, Dr. Cristiando André da Costa, for his support and encouragement during all stages of this work. His optimism and dynamism are inspiring.

I would also like to thank the entire group of professors, colleagues, and technical-administrative staff of the Graduate Program in Applied Computing at Unisinos. I especially thank the colleagues of SoftwareLAB, who have always been excited by each one of the groups' publications. I also thank Professor Dr. Rodrigo da Rosa Righi, who supported me in this journey's important moments.

I would like to thank Diogo Schmidt, colleague, friend, and partner of publications, for his tremendous help to develop this work.

With great affection, I thank my family, especially my sister, Carla, and my mother, Mariza, who supported me at all times on this journey, both from far and from near. Finishing my Ph.D. along with the COVID-19 pandemic, was not an easy task, and, with absolute certainty, without the support of these two wonderful women, I would not have succeeded.

The author would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) and Brazilian National Council for Scientific and Technological Development (CNPq) for supporting this work.

ABSTRACT

CONTEXT: The accurate and early diagnosis of critically ill patients depends on medical staff's attention and the observation of different variables, vital signs, and laboratory test results, among others. Seriously ill patients usually have changes in their vital signs before worsening. Monitoring these changes is essential to anticipate the diagnosis in order to initiate patients' care. Prognostic indexes play a fundamental role in this context since they allow us to estimate the patients' health status. Besides, Electronic Health Records' adoption improved data availability, which can be processed by machine learning techniques for information extraction to support clinical decisions. The volume and variety of data stored in the EHR make it possible to carry out more accurate analyzes that allow different types of health care assessments. Nevertheless, as the amount of data available is vast and complex, there is a need for new methods to analyze that data to explore significant patterns. The use of Machine Learning (ML) techniques to generate knowledge, search for information patterns, and support clinical decisions is one of the possibilities to address this problem. OBJECTIVE: this work aims to create a computational model able to predict the deterioration of patients' health status in such a way that it is possible to start the appropriate treatment as soon as possible. The model was developed based on the Deep Learning technique, a Recurrent Neural Networks, the Long Short-Term Memory, to predict patient's vital signs and subsequent evaluation of the patient's health status severity through Prognostic Indexes commonly used in the Health area. METHOD: The methodology of this work consists of the following steps carried out in sequence. The definition of the data source to be used in the creation of the model and the selection of the data, the pre-processing to create a database for the development of the model, the definition of the implementation of the model and its evaluation through comparison with other models. RESULTS: Experiments showed that it is possible to predict vital signs with good precision (accuracy $>$ 80%) and, consequently, predict the Prognostic Indexes in advance to treat the patients before deterioration. Predicting the patient's vital signs for the future and use them for the Prognostic Index' calculation allows clinical times to predict future severe diagnoses that would not be possible applying the current patient's vital signs (50% - 60% of cases would not be identified). CONCLUSION: This work's main scientific contribution is the creation of a method for predicting vital signs based on historical data with low Mean Squared Error and its following application in the calculation of prognostic indexes with effectiveness (50% - 60% of cases that would not be identified as severe). The differential presented by this proposal stems from the fact that few works predict vital signs. Most of the works focus on predicting specific health outcomes, such as specific diagnoses, considering the current vital signs. In this work, the proposal is to predict the evolution of vital signs in the future and use these predicted signs to calculate prognostic indexes.

Keywords: Predictive Scores. Machine Learning. Deep Learning. LSTM. Health Informatics.

RESUMO

CONTEXTO: O diagnóstico preciso e precoce do paciente crítico depende da atenção da equipe médica e da observação de diversas variáveis, sinais vitais, resultados de exames laboratoriais, entre outros. Pacientes gravemente enfermos geralmente apresentam alterações em seus sinais vitais logo antes de piorar. O monitoramento dessas mudanças é essencial para antecipar o diagnóstico e iniciar o atendimento ao paciente. Os índices prognósticos desempenham um papel fundamental neste contexto, uma vez que permitem estimar o estado de saúde dos pacientes. Além disso, a adoção do Registro Eletrônico de Saúde melhorou a disponibilidade de dados, que podem ser processados por técnicas de aprendizado de máquina para extração de informações para apoiar decisões clínicas. O volume e a variedade de dados armazenados no EHR possibilitam a realização de análises mais precisas que permitem diversos tipos de avaliações em saúde. No entanto, como a quantidade de dados disponíveis é vasta e complexa, há uma necessidade de novos métodos de análise desses dados para explorar padrões significativos. O uso de técnicas de Aprendizado de Máquina (AM) para gerar conhecimento, buscar padrões de informação e apoiar decisões clínicas é uma das possibilidades para enfrentar esse problema. OBJETIVO: este trabalho visa criar um modelo computacional capaz de predizer a deterioração do estado de saúde dos pacientes de forma que seja possível iniciar o tratamento adequado o mais rápido possível. O modelo foi desenvolvido com base na técnica de Deep Learning, uma Rede Neural Recorrente, a Long Short-Term Memory, para predizer os sinais vitais do paciente e posterior avaliação da gravidade do estado de saúde do paciente por meio de Índices Prognósticos comumente utilizados na área da saúde. MÉTODO: A metodologia deste trabalho consiste nas seguintes etapas realizadas em sequência. A definição da fonte de dados a ser utilizada na criação do modelo e a seleção dos dados, o pré-processamento para criar uma base de dados para o desenvolvimento do modelo, a definição da implementação do modelo e sua avaliação por comparação com outros modelos. RESULTADOS: Experimentos mostraram que é possível predizer sinais vitais com boa precisão (acurácia $>80\%$) e, conseqüentemente, predizer os Índices Prognósticos com antecedência para tratar os pacientes antes da deterioração. Prever os sinais vitais do paciente para o futuro e usá-los para o cálculo do índice prognóstico permite que os tempos clínicos prevejam diagnósticos graves futuros que não seriam possíveis aplicando os sinais vitais atuais do paciente (50% - 60% dos casos não seriam identificados) CONCLUSÃO: A principal contribuição científica deste trabalho é a criação de um método de predição de sinais vitais baseado em dados históricos com baixo Erro Quadrático Médio e sua posterior aplicação no cálculo de índices prognósticos com eficácia (50% - 60% dos casos que não seriam identificado como graves). O diferencial desta proposta decorre do fato de poucos trabalhos predizerem sinais vitais. A maioria dos trabalhos se concentra em predizer desfechos de saúde específicos, como diagnósticos específicos, considerando os sinais vitais atuais. Neste trabalho, a proposta é prever a evolução dos sinais vitais no futuro e utilizar esses sinais previstos para calcular índices prognósticos.

Palavras-chave: Índices Prognósticos. Aprendizado de Máquina. Aprendizado Profundo. LSTM. Informática na Saúde.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AaDO2	Alveolar-arterial oxygen tension difference
ANN	Artificial Neural Network
APACHE	Acute Physiology and Chronic Health Evaluation
APS	Acute Physiology Score
AUC	Area Under the Curve ROC
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CDSS	Clinical Decision Support System
CNN	Convolutional Neural Networks
ICD-9	International Classification of Diseases, Ninth Revision
DL	Deep Learning
EDI	Early Deterioration Indicator
EHR	Electronic Health Record
EWS	Early Warning Score
FiO2	Fraction of inspired oxygen
FHIR	Fast Healthcare Interoperability Resources
GCS	Glasgow Coma Scale
HR	Heart Rate
ICU	Intensive Care Unit
LHS	Learning Health System
LR	Logistic Regression
LSTM	Long Short-Term Memory
MAP	Mean Arterial Pressure
MEWS	Modified Early Warning Score
MIMIC	Medical Information Mart for Intensive Care
ML	Machine Learning
MSE	Mean Squared Error
NEWS	National Early Warning Score
OLS	Ordinary Least Squares
PaCO2	Partial Pressure of Carbon Dioxide
PaO2	Partial pressure of Oxygen
pH	Power of Hydrogen
qSOFA	quick Sepsis Related Organ Failure Assessment
quickSOFA	quick Sepsis Related Organ Failure Assessment

RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
RR	Respiratory Rate
SAPS	Simplified Acute Physiology Score
SIRS	Systemic Inflammatory Response Syndrome
SOFA	Sequential Organ Failure Assessment Score
SVM	Support Vector Machine
ViEWS	VitalPAC Early Warning Score

LIST OF SYMBOLS

^{o}F Graus Fahrenheit

CONTENTS

1 INTRODUCTION

Studies suggest that early intervention based on policies and care standards can improve patients' chances of survival (CHEN, 2018; JUNG et al., 2019; PAN et al., 2019; JANG et al., 2020) and Clinical Decision Support Systems (CDSS) can play a fundamental role in the identification of risk situations and take of action by hospital care teams to reduce cases of death (JONES et al., 2014; O'BRIEN et al., 2020). Sepsis, for example, is a critical condition and is the body's response to infection. In sepsis, the first three hours after its detection are crucial to start the treatment with antibiotics; it is the so-called sepsis golden hour (KUMAR et al., 2004).

1.1 Motivation

The volume and variety of data stored in the Electronic Health Record (EHR) make it possible to carry out accurate analyzes that allow different types of assessments in the area of health care (CURTIS et al., 2018; WANG et al., 2020). Some of the benefits of using Big Data in the Health area through EHR (SHAFQAT et al., 2018; ALLOGHANI et al., 2019) are prevention and identification of risk factors for specific diseases as also interventions to change population risk behaviors. The set of data collected from sensor-equipped devices, the idea of Internet of Health Things (COSTA et al., 2018a), the EHR and analysis results generated by CDSS, and the decisions physicians and their teams take every day increase the volume of data to be managed by systems and people. Thus, Machine Learning (ML) techniques can extract knowledge from the relationship between all of the data (PURUSHOTHAM et al., 2018; BRAND et al., 2018; CROWN, 2015; GIANNINI et al., 2019).

ML algorithms have been applied to support the accurate diagnosis of critically ill patients by analyzing their vital signs (CLIFTON et al., 2011; GULTEPE et al., 2014; RAJKOMAR et al., 2018; LIU et al., 2013, 2014; ALLOGHANI et al., 2019). However, the anticipation of diagnosis may depend on data not yet presented by the collection systems. Therefore, predicting the evolution of some of the vital signs, important for specific diagnoses, can constitute a reasonable alternative (FORKAN; KHALIL, 2017; JANG et al., 2020). The presented scenario indicates the possibility of applying ML algorithms to identify patients' risk to evolve to a more severe state of health and identify possible treatments. Although these extensive use of ML algorithms to analyze vital signs, to the best of our knowledge, only a few other works aim to predict the vital signs itself, e.g., (FORKAN; KHALIL, 2017). It is important to note that most works apply Machine Learning to predict a condition based on the patterns embedded in the signal, while this work proposes predicting the vital signs itself (ALLOGHANI et al., 2019). Besides, most ML systems are dedicated to predicting specific diagnoses through measured vital signs at one moment instead of with predicted vital signs.

Deep Learning (DL) (LECUN; BENGIO; HINTON, 2015) is a ML technique able to learn complex data abstractions through a hierarchical learning mechanism. Multiple abstraction

levels, composed of a cascade of modules processing layers from a raw data input layer, each successive layer transforms the input data into an output with a higher abstraction level (GOODFELLOW; BENGIO; COURVILLE, 2016). DL techniques have better performance since the creation of models can dispense the prepossessing stage, the discovery of knowledge happens through the relationship between data itself, without the need to identify the best predictors (SHICKEL et al., 2017; Pathak; Cai; Rajasekaran, 2018). Among DL techniques, Recurrent Neural Networks (RNN) are considered adequate for processing data sequences and have been applied in the Health area because they are appropriate for treating data such as text, speech, and DNA sequences (RAVI et al., 2017). In this type of network, each state of the network affects the next state's results — however, RNN presents memory loss problems for older data that may be important for prediction. The Long Short-Term Memory (LSTM) networks are RNN that preserve memory dependencies already discovered inside data, which has a more significant impact on prediction (GOODFELLOW; BENGIO; COURVILLE, 2016). The sequence of vital signs monitored for each hospitalized patient can be lengthy, and it is essential to use the entire available sequence so that the prediction model considers the history of patient outcome evolution. Such a strategy may add customization characteristics to the model towards the idea of precision medicine (BELLOT; SCHAAR, 2018).

Prognostic indexes (PHD et al., 2016) play a crucial role in establishing an estimate of the inpatients' progress. It can also help hospital units analyze their quality in care and performance to measure hospital mortality control compared to different hospitals' intensive care units. Besides, prognostic indexes can predict Intensive Care Unity (ICU) stay, length of hospital stay, and predict specific diseases such as heart attack or sepsis (ALAM et al., 2014; LE LAGADEC; DWYER, 2017; DOWNEY et al., 2017). These indexes consider the knowledge built in the Health area about the severity of patients' state of health and depict ranges of values for specific attributes and the relationships between them. Calculations are made to infer some element, usually a numerical index, about the patient's condition (GERRY et al., 2017; PRYTHERCH et al., 2010). Acute Physiology And Chronic Health Evaluation II (APACHE II) (WAGNER; DRAPER, 1984), Sequential Organ Failure Assessment Score (SOFA) (VINCENT et al., 1996a), some variations of Early Warning Score (EWS) (VINCENT et al., 1996b), Simplified Acute Physiology Score II (SAPS II) (LE GALL; LEMESHOW; SAULNIER, 1993) are the scores primarily used in medical care units. The one chosen for the firsts experiments in this work was the APACHE II because it has been applied all over the world (WAGNER; DRAPER, 1984).

1.2 Problem

It is crucial to detect the worsening situation as soon as possible to prevent the patient's condition from deteriorating. For severe cases, clinical protocols indicate that care teams' actions, if performed in the first hours of care, increase recovery chances (BOADA, 2016). Change in

vital signs can generate more precise prognostic information (BREKKE et al., 2019). However, the association between abnormal values of vital signs and the consequent outcome for hospitalized patients is not well known (ASIIMWE; VITTINGHOFF; WHOOLEY, 2020). Likewise, there is little information about the mortality rate among patients who present normal vital signs during hospitalization. The research carried out by (YAN; HSIA, 2019) identified that the following factors are common to patients who present normal vital signs at the time of the first visit, but later die: septicemia, fluid and electrolyte disorders, acute renal failure, psychosis, general weakness, and more advanced age. Thus, analyzing more information (e.g., laboratory tests), in addition to vital signs, it is a possibility to improve prognosis. Besides, predicting this information during patient care can improve the outcome prediction as well.

1.3 Research Question

The present work has as the main issue to study and apply ML techniques in the early identification of the worsening of the health status of patients admitted to hospital units. Therefore, some prognostic indexes were studied, as well as their limitations, were pointed out. Besides, some ML techniques were also analyzed, mainly considering the ones intended to EHR data, which have temporal information. Considering this context, of calculating prognostic indexes and the possibility of predicting the worsening of the patient's health employing ML techniques, we proposed the following research question to be answered by this work:

> *How would be an intelligent-based model to predict the information about the worsening of patients' health through different prognostic indexes calculated on clinical data predicted using Deep Learning (DL) techniques?*

Research carried out in the Health area can have an impact on clinical policies and practices in the area, possibly generating good results for patient health care (AKHNIF et al., 2017). Besides, they raise the idea of a Learning Health System (LHS) as an attractive alternative that can reduce service providers' costs (AKHNIF et al., 2017; LOWES et al., 2016; MURPHY et al., 2017; PAYNE et al., 2015; YADAV et al., 2018; PLATT; RAJ; WIENROTH, 2020). The research carried out in this work aims to generate positive impacts for the Health area, encouraging automatic and intelligent techniques based on data analysis to support patients' early diagnosis in serious condition.

Contributing to an LHS's idea, the affirmation of the EHR as an alternative for storing patients' health information during their hospitalizations and recording the history of the evolution of their health status (ROEHRS et al., 2017; WILLIAMS et al., 2017; WANG et al., 2020). Nevertheless, efforts to standardize and share EHR data allow us to glimpse a scenario where large volumes of data will be available to researchers to develop new research (JOHNSON et al., 2016; KAKKANATT et al., 2018; ROEHRS; COSTA; ROSA RIGHI, 2017).

Finally, the finding that the evolution of computational resources and decrease in the cost involved with the storage and processing of data provided the investment in intelligent analysis techniques data and their application in the Health area also supports the research question proposed by this work (SHICKEL et al., 2017; YADAV et al., 2018). Among the classes of techniques employed, the ML techniques (KHARE et al., 2017), which have their comprehensive taxonomy, have in DL methods one of the techniques that currently have repercussions for its application with good results in the Health area (KAM; KIM, 2017; MANASHTY; THOMSON, 2017; RAJKOMAR et al., 2018; SHICKEL et al., 2017). DL techniques allow the discovery of knowledge only through the data itself, without the knowledge of specialists in the application area (RAVI et al., 2017), and for this reason, they are part of this work.

1.4 Objective and Scientific Contribution

The work presented here proposes the DeepSigns Model to work in conjunction with CDSS in hospital practice to predict the input information for the methods of severity assessment applied by hospitals. The DeepSigns Model aims to predict vital signs, through the use of LSTM networks, and to calculate the prognostic index for the early diagnosis of worsening patients' health status. Indeed, the DeepSigns model incorporates other variables besides vital signs, as results of laboratory tests, age, among others. DeepSigns considers the set of variables from APACHE II, presented in Figure 2. Therefore, in this work, referring to vital signs implies a broader set of variables.

Figure 1 depicts the traditional model for detecting deterioration of inpatients applied in hospitals versus the DeepSigns model proposed in this work. In the conventional model, inpatient vital signs are taken as input to the prognostic indexes, which will assess the patient health deterioration status. On the other hand, the DeepSigns Model aims at considering the predicted inpatient vital signs as inputs to the prognostic indexes calculation.

Many prognostic indexes, used in hospital units, could apply the DeepSigns predicted vital signs in their calculation; for instance, Manchester protocol (AZEREDO et al., 1995; COOKE; JINKS, 1999; STORM-VERSLOOT et al., 2011) in Emergency Departments or the scores typically used in ICU such as SAPS II, APACHE II, SOFA, or variations of EWS. Although there is a general assessment that prognostic indexes based on domain expertise (heuristics), such as those previously mentioned, are not as useful in identifying patients with a high risk of deterioration (CHURPEK et al., 2014; LINNEN, 2019), the hypothesis considered here is that the application of such methods with predicted data could be more effective.

The main scientific contributions of this work are as follows:

1. the development of a method for predicting patient data, feeding back to the model with new personal data, meeting the idea of the precision medicine idea;

2. the assessment of many prognostic indexes using the vital signs predicted instead of spe-

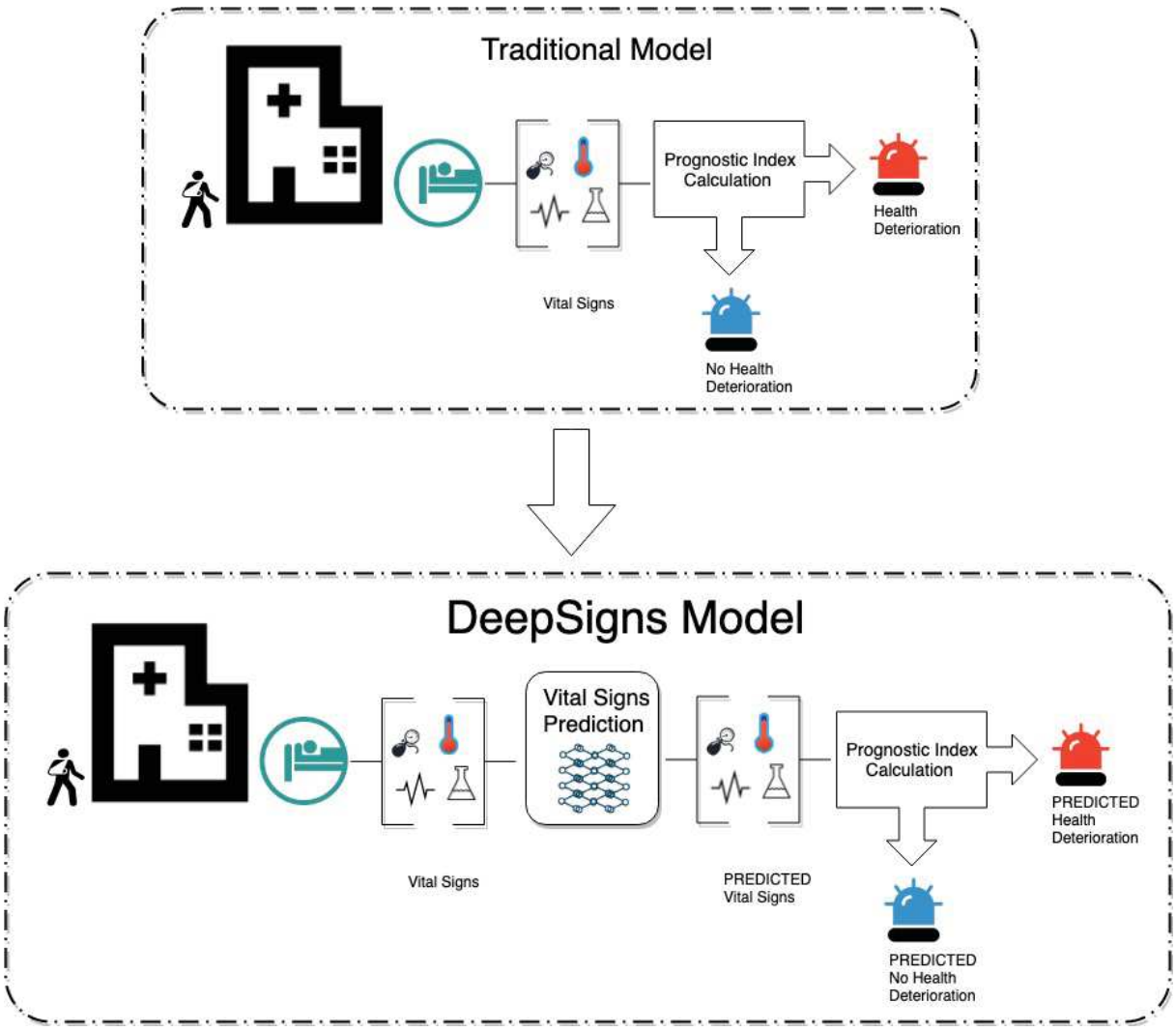

Figure 1 – Health Deterioration Prognostic: Traditional Model x DeepSigns Model

cific time measured values;

3. this work's main scientific contribution is the creation of a method for predicting vital signs based on historical data and its following application in the calculation of prognostic indexes;

4. the differential presented by this proposal stems from the fact that few works predict vital signs. Most of the works focus on predicting specific health outcomes, such as specific diagnoses, considering the current vital signs. In this work, the proposal is to predict the evolution of vital signs in the future and use these predicted signs to calculate.

1.5 Methodological Aspects

This section describes the present study's methodological aspects. The study has an applied nature and proposes a predictive model for early detection of patient health deterioration, using

Deep Learning, to assist the therapeutic decision-making in intensive medicine. The procedures and methods used in this work has a quantitative methodological approach. According to (AZEVEDO; MACHADO; SILVA, 2011), research with a quantitative focus aims to emphasize deductive reasoning by comparing measurable variables. Therefore, this work will have an experimental focus, analyzing the impacts of creating a prediction model based on vital signs and other patient's information, in conjunction with a heuristic-based prognostic index.

Defining the methodological procedures used in the research is essential, as stated by (GERHARDT; SILVEIRA, 2009). This research applied the following methods: bibliographic, documentary, and experimental research. The research process was composed of six steps, presented in Figure 2: problem characterization; research question formulation; study of state of the art related to the problem; DeepSigns model definition, development, and validation; results evaluation and discussion; and conclusions and future work propositions.

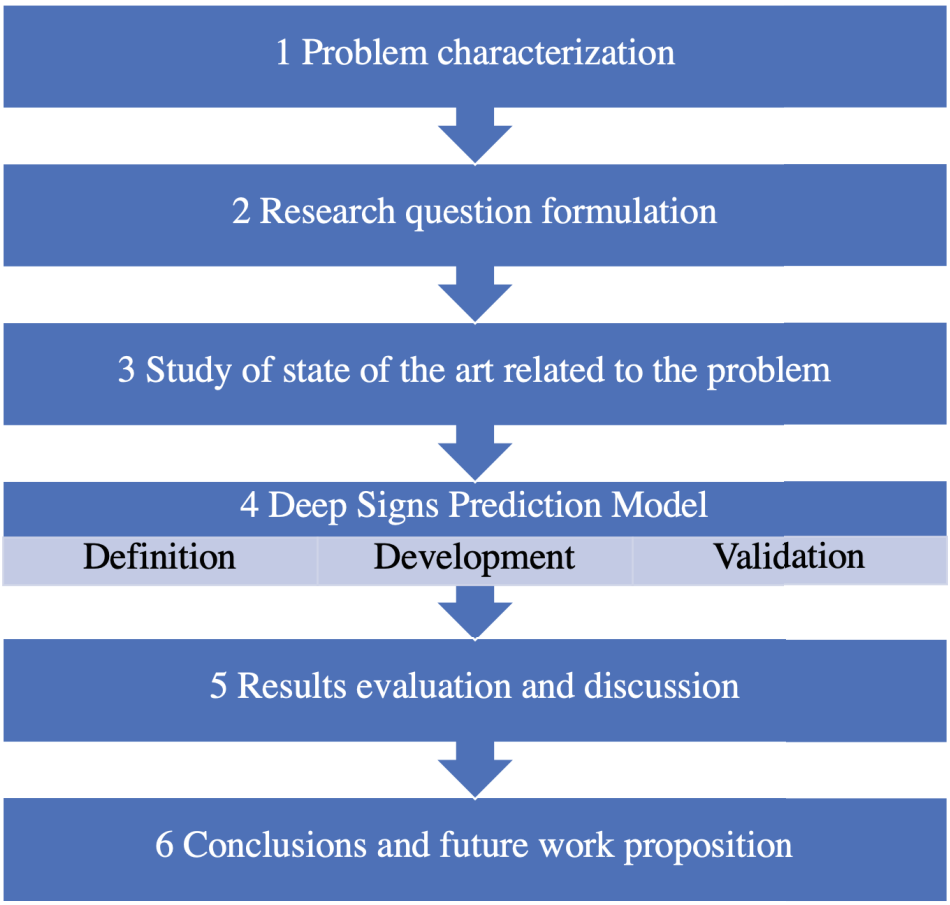

Figure 2 – Methodological Steps

Chapter 1 presented the two first steps of the methodology, according to Figure 2: the problem characterization and the research question formulation. The next chapters and sections will discuss steps three through six, as follows.

- Step 3: Chapters 2 and 3 discusses the main concepts underlying this research and present related work, discussing their results and conclusions;

- Step 4: Chapter 4 presents the DeepSigns model, the design decisions and architecture underlying the model and its development and validation process;
- Step 5: Chapter 5 discusses the results obtained through the work, its limitations, and opportunities.
- Step 6: Chapter 6 Discusses the lessons learned with this research and future directions.

1.6 Text Organization

To give the reader a better understanding of this text, the following is a brief description of the reminder chapters that are part of this work.

- Chapter 2 - Background: discusses the main concepts underlying the present work: Big Data in Health, Prognostic Indexes, and Machine Learning and Deep Learning.
- Chapter 3 - Related Works: presents works related to the present one, discussing their results and aspects which inspired this research.
- Chapter 4 - Proposed Model: presents the model proposed in this work.
- Chapter 5 - Results and Discussion: discuss the results obtained through the work, its limitations, and opportunities.
- Chapter 6 - Conclusion: Discusses the lessons learned with this research and future directions.

2 BACKGROUND

This chapter presents the theoretical framework that provides the basis for this study. The bibliographic review searched for other scientific works that deal with the themes addressed in this work. Returning to the research question proposed in this work, "What are the possible impacts on health care by predicting, in advance, the worsening of patients' health status, through different prognostic indexes calculated on predicted clinical data using Deep Learning techniques? (DL)?", we see the need to study different themes. Some areas of study, the target of this work, are more clearly portrayed in the text of the research question, namely: the prognostic indexes currently used to assess the evolution of the health status of patients admitted to hospital units; the demands of the Health area concerning the volume of data generated daily within hospitals, which is attached to the Big Data area; and the techniques to analyze these data in order to generate new prognoses, alternative to those currently employed, in order to improve the quality of treatments and promote the recovery of patients. The themes mentioned will be presented below and make up the body of knowledge covered by this work. Other themes are also related to the proposed research question and are addressed throughout this work to support some arguments and choices or promote discussion.

2.1 Big Data and Electronic Health Record in Patient Care

The possibility of obtaining new knowledge in the health field through the use of Big Data is causing the rapid digitization of a large amount of data (RAHMAN; SLEPIAN; MITRA, 2016; WANG; KUNG; BYRD, 2018). Increasingly, healthcare companies and organizations are using Big Data technology to improve health information systems' quality (RAHMAN; SLEPIAN; MITRA, 2016).

The EHR initiative arose to meet the demand for data for the Health area, maintain data on the regular operation of health institutions, and promote the creation of data collections that allow research activities (YADAV et al., 2018; SCHLOEFFEL, 2002). The volume and variety of data stored in the EHR make it possible to carry out more accurate analyzes that allow different types of health care assessments (CURTIS et al., 2018). The benefits of using Big Data in healthcare, through the EHR, can be seen from disease prevention and identification of risk factors to the design of interventions to change risk behaviors in the population (SHAFQAT et al., 2018).

This context motivated the construction of databases, with real and detailed data, specifically created to support clinical decision support research. In the field of intensive care, several databases, commercial (APACHE Outcomes, Project IMPACT, and Philips eICU) and non-commercial (MIMIC - Medical Information Mart for Intensive Care) (JOHNSON et al., 2016), were developed to assess the severity of health conditions, patient health, as well as the costs involved in treatments, among other purposes (CELI et al., 2013).

For the development of a Big Data project, the literature recommends six steps. First, it is necessary to identify the problem to be solved. The second step is to find the best way to collect the data. The third is to perform the pre-processing of data (clear the data), that is, to store meaningful data. The fourth is to analyze the data for the production of knowledge. The fifth step is to present the results found through analytical reports that are easy to understand. Finally, the sixth is to evaluate the project to verify if the developed project solves the problem (HUANG et al., 2015). The present work relates to Big Data applied to Health Area.

2.2 Machine Learning

According to (FANG et al., 2016), the amount of data available in the Health area is not only vast but complex, requiring new ways to analyze this data. It is necessary to explore significant patterns when analyzing large volumes of data. Applying ML techniques to generate knowledge, search for information patterns, and support clinical decisions is one way to address this problem (LEE et al., 2017).

As previously indicated, this work's focus is to deal with health data, which exists in considerable volume and analyze them to obtain assertive results concerning diagnoses and patient care. ML provides techniques that allow the analysis of large volumes of data so that the data itself can demonstrate specific situations, such as the patient's health status, for example. The ML allows the development of models called data-based models, which, unlike the prognostic indexes presented in section 2.3, are not heuristic models (GERRY et al., 2017).

According to (REZENDE, 2003), ML intends to build learning systems that automatically acquire knowledge. According to the authors, automated learning systems are computational algorithms that try to solve problems, making decisions based on accumulated experiences, based on successful decisions previously made. Also, according to (MITCHELL, 1997), transforming observational data into a model that predicts future events or explaining data not yet observed, improving its performance through experience is the objective of ML.

ML algorithms can be supervised and unsupervised (FANG et al., 2016). In supervised learning, the algorithm seeks to develop a general prediction rule using a set of samples (training base) where the class (label) of each sample is known, to predict the class of new data not yet labeled. Generally, the training base is formed by a set of samples (inputs) that are described by a vector of characteristics or attributes (variables) and the label of the associated class (answer or output). Supervised learning algorithms are categorized into: Classification problems (for class with discrete value labels) or Regression problems (for labels with continuous values). The type of ML addressed in this work is the supervised one.

In unsupervised learning, the classes' labels are unknown, and the objective of the algorithm is to find a way to group the input data by searching for similar characteristics. This type of problem is known as Clustering.

The learning process consists of finding a classifier with a greater generalization capacity,

finding a general function (model) that covers the classification, not only of the existing samples in the training base but also the largest number of new data. However, the model cannot be too specialist (overfitting) in the training data used, as it may present a low level of accuracy for new data (MITCHELL; MICHALSKI; CARBONELL, 2013).

According to (FANG et al., 2016), among the most widely used classification techniques in the medical field are Decision Trees, Rules-Based Algorithms, Support Vector Machine, Sparse Representation, Ensemble Algorithms (set of classifiers), Artificial Neural Networks (ANN), and DL.

2.2.1 Deep Learning

DL is an ML technique that is capable of learning complex data abstractions through a hierarchical learning mechanism. The technique provides multiple levels of abstraction, composed of a cascade of modules, or simple processing layers, in which, from a raw data input layer, each successive layer transforms the representation of the input data into an output with a higher level of abstraction. This process's final result (output layer) represents raw data in useful information (GOODFELLOW; BENGIO; COURVILLE, 2016). DL techniques have the best performance since the creation of the models can happen without a pre-processing step, since the discovery of knowledge occurs through the data, without prior identification of the best predictors (SHICKEL et al., 2017).

Although it is often challenging to train deep architectures efficiently, they have been the subject of several research works, as they present two essential advantages: they promote the reuse of resources and can lead to progressively more abstract comprehensions in the higher layers. The ability to build multiple levels of representation is related to the depth of the architecture. In deep circuits, the number of paths, or the way to reuse different parts of it, grows exponentially with depth, allowing, by changing the definition of what each node can calculate, the circuit's depth to be changed (BENGIO; COURVILLE; VINCENT, 2013).

Among DL techniques, Recurrent Neural Networks (RNN) are considered adequate for processing data strings and have been applied in the Health area because they are appropriate for treating data such as text, speech, and DNA sequences (RAVI et al., 2017). In this type of network, each state of the network affects the next state's result because of feedback connections in its structure.

Convolutional Neural Networks (CNN) are another type of DL mechanism that does not depend on any previous prediction information. For example, in the stock market prediction task (SELVIN et al., 2017), the prediction depends more on the moment's dynamics than any previous information, and CNN would suit this type of application well.

One difficulty with RNN networks is that they have memory loss problems for older data, which can be important for some prediction tasks. LSTM networks (HOCHREITER; SCHMIDHUBER, 1997) are RNN that preserve the memory of dependencies already discovered among

the data, which has a more significant impact on prediction (GOODFELLOW; BENGIO; COURVILLE, 2016).

In this work, we chose to apply LSTM networks because of the analysis of long sequences of data, which portray the variety of patients' vital signs.

2.3 Prognostic Indexes

Prognostic indexes (PHD et al., 2016) play a crucial role in establishing an estimate of the inpatients' progress. It can also help hospital units analyze their quality in care and perform as a measure of hospital mortality to compare different hospitals' intensive care units. Besides, prognostic indexes can predict Intensive Care Unity (ICU) stay, length of hospital stay, and predict specific diseases such as heart attack or sepsis (ALAM et al., 2014; LE LAGADEC; DWYER, 2017; DOWNEY et al., 2017). These indexes take into account the knowledge considered in the Health area about the severity of patients' state of health and depict ranges of values for specific attributes and the relationships between them. The calculations intend to infer some element, usually a numerical index, about the severity of the patient's condition (GERRY et al., 2017; PRYTHERCH et al., 2010). Acute Physiology And Chronic Health Evaluation II (APACHE II) (WAGNER; DRAPER, 1984), Sequential Organ Failure Assessment Score (SOFA) (VINCENT et al., 1996a), some variations of Early Warning Score (EWS) (VINCENT et al., 1996b), Simplified Acute Physiology Score II (SAPS II) (LE GALL; LEMESHOW; SAULNIER, 1993) are the scores primarily used in medical care units. The one chosen for the firsts experiments in this work was the APACHE II because it has been applied all over the world (WAGNER; DRAPER, 1984). Table 1 shows a comparison between the variables that are applied by some prognostic scores currently used by hospitals.

2.3.1 APACHE II

APACHE II is a mathematical clinical index system, published initially in (KNAUS et al., 1991). The first version of the index was developed as a good and reliable classification system based on multiple physiological variables, obtained from medical records, to estimate the disease's severity in critically ill patients.

The original version used the sum of the score of forty-three physiological variables to calculate the index, which represented the degree of severity of the disease - Acute Physiology Score (APS), and an assessment of the patient's health status before admission to the ICU, such as medical visits, work activity, ability to perform routine activities and the presence or absence of malignant diseases (KNAUS et al., 1991).

Due to the complexity of the first version of the index, it was considered difficult to apply, as it required many variables, (KNAUS et al., 1991) proposed a modification in the system to make it simpler and increase its predictive power. In the second version, called APACHE II, the

Table 1 – Common Scores Variables

Measure / Score	SAPS II	APACHE II	SOFA	NEWS	qSOFA
Age	X	X			
Temperature	X	X		X	
Blood Pressure (Diastolic)		X	X		
Blood Pressure (Systolic)	X	X	X		X
Heart Rate	X	X		X	
Respiratory Rate		X		X	X
Oxygenation FiO2	X	X	X	X	
Oxygenation PaO2	X		X		
Oxygenation SpO2				X	
Arterial pH		X			
Sodium	X	X			
Potassium	X	X			
Bicarbonate	X				
Urea	X				
Creatinine		X	X		
Hematocrit		X			
Leukocytes	X	X			
Platelets			X		
Bilirubin	X		X		
Urine Output	X		X		
Glasgow Coma Scale	X	X	X		X
APVU				X	
Presence of Comorbidity	X	X			
Type of Admission	X				
Number of variables	16	15	9	6	3

index started to use a smaller number of twelve variables, routinely collected in the ICU.

The APACHE II index is calculated based on the worst values recorded in the first twenty-four hours after admission to the ICU. Its application requires a minimum period of twenty-four hours of hospitalization of the patient. Physiological and laboratory variables are used, each with points from zero to four. Points are also assigned according to age (from zero to six points) and the presence or absence of comorbidities (zero to five points). The variables used by APACHE II are shown in Figure 3.

It is possible to estimate the risk of death with the APACHE II index, through the sum of physiological variables, age, and the presence of comorbidities (organ failure or immunocompromised evident before hospital admission). The result of the index can vary from zero to seventy-one points, where the higher the value of the index, the greater the probability of death.

There is a significant increase in death probability for every five points of increase in the index (KNAUS et al., 1991).

APACHE II is based on the variables presented in Table 2, some of which are vital signs, and other are laboratory exams results. Figure 3 shows its calculation algorithm. Table 1 shows a comparison between variables applied by prognostic scores currently used by hospitals. APACHE II variables are similar to the ones from other scores, e. g. the SAPS II (LE GALL; LEMESHOW; SAULNIER, 1993), but it has a broader set of variables than most of the other indexes. APACHE II was chosen for experimenting at first because the proposal presented in this work considers attaching different prognostic scores to the prediction model, as chapter 4 will discuss a broader set of vital signs to be part of the model seems adequate.

Table 2 – Vital signs present in the prediction model

Measure	Abbr.
Temperature	T
Mean Arterial Pressure	MAP
Heart Rate	HR
Respiratory Rate	RR
Oxygenation - Fraction of inspired oxygen	FiO2
Oxygenation - Partial pressure of oxygen	PaO2
Oxygenation - Alveolar-arterial oxygen tension difference	AaDO2
Arterial pH	ApH
Serum Sodium	SS
Serum Potassium	SP
Creatinine	C
Hematocrit	H
Leukocytes	L
Points on the GCS	G
Age	A

2.3.2 NEWS

The National Early Warning Score (NEWS), initially proposed in 2012 (PHYSICIANS, 2017), was developed to be used in the British health care system as a scoring system to predict mortality and severe worsening health status of patients (SILCOCK et al., 2015). The most recent version of this scoring system is NEWS2 (PHYSICIANS, 2017). Figure 4 shows the weight scale assigned to each patient's measurement when calculating the index. In the end, there is the summation of weights, and the index is defined; its value can vary between 0 (zero) and 6 (six). Considering m as the patient's measurement for each sign shown in Table 4; n as the number of signals (lines in Table 4) involved in the calculation of the index, and p represents the

	Acore APACHE II Points	High abnormal range +4	+3	+2	+1	+0	Low abnormal range +1	+2	+3	+4
Physiologic Variables	Temperature (C)	>41	39-40.9		38.5-38.9	36-38.4	34-35.9	32-33.9	30-31.0	<29.9
	Mean Aterial Pressure (mmHg)	>160	130-159	110-129		70-109		50-69		<49
	Heart Rate (/min)	>180	140-179	110-139		70-109		55-69	40-54	<39
	Respiratory Rate (/min)	>50	35-49		25-34	12-24	10-11	6-9	<5	
	Saturation FiO_2<50%: PaO_2					70 61-69.9		55-60	<55	
	Saturation FiO_2>50%: A-aDO_2	>500	350-499	200-349		199				
	Arterial pH	>=7.7	7.69-7.7		7.5-7.59	7.33-7.49		7.25-7.32	7.15-7.24	<7.15
	Na+ Serum Sodium (mM)	>=180	160-179	155-159	150-154	130-149		120-129	111-19	<=110
	K+ Serum Potassium (mM)	>7	6-6.9		5.5-5.9	3.5-5.4	3-3.4	2.5-2.9		<2.5
	Creatinene (mg/100ml)	>3.5	2-3.4	1.5-1.9		0.6-1.4		<0.8		
	Haemocrit (%)	>60		50-59.9	46-49.9	30-45.9		20-29.9		<20
	White Blood Count	>40		20-39.9	15-19.9	3-14.9		1-2.9		<1
Glasgow Coma Scale Score		+15 - GCS Score								
A. Total acute physiologic score										
B. Age points (years) <44=+0, 45-54=+2, 55-64=+3, 65-74=+5, >=75=+6										
C. Chronic health points (elective postoperative = +2, nonoperative or emergency postoperative patients = +5)										
Total APACHE Score = A + B + C. Minimum score = 0; maixmum score = 71.										
0-4 = 4% death rate; 10-14 = 15% deth rate; 20-24 = 40% death rate; 30-34 = 75% death rate										
5-9 = 8% death rate; 15-19 = 25% death rate; 25-29 = 55% death rate; Over 34 = 85% death rate										

Figure 3 – APACHE II Calculation

weight of said measure in Table 4; Equation 2.1 presents the calculation of the patient's NEWS index.

$$\sum_{k=1}^{n} p, m \tag{2.1}$$

NEWS Score	3	2	1	0	1	2	3
Respiratory Rate, min	≤8		9-11	12-20		21-24	>=25
SpO2	<=91	92-93	94-95	>=96			
Any Supplemental Oxygen		Yes		No			
Temperature, °C	<=35.0		35.1-36.0	36.1-38.0	38.1-39.0	≥39.1	
Blood pressure, mm Hg	<=90	91-100	101-110	111-129			>=220
Heart Rate, min	<=40		41-50	51-90	91-110	111-130	≥131
							V Reacting to voice
				A			P Reacting to pain
Consciousness (APVU)				Alert			U Unresponsive

Figure 4 – NEWS Calculation

2.3.3 MEWS

The Modified Early Warning Score (MEWS) is an evolution of NEWS, including the difference of patient's blood pressure to its expected value, plus the amount of urine (SUBBE, 2001). The index calculation remains the same as the NEWS index mentioned in the previous section but considering Figure 5.

MEWS Score	3	2	1	0	1	2	3
Heart Rate, min		<40	41-50	51-100	101-110	111-129	≥130
Respiratory Rate, min		≤8		9–14	15-20	21-29	>29
Blood pressure, mm Hg	<45%	<30%	<15%	Standar for patient	>15%	>30%	>45%
Temperature, °C		<35		35.0-38.4		≥38.5	
				A	V	P	
Consciousness (APVU)				Alert	Reacting to voice	Reacting to pain	U Unresponsive
Unine output		<0.5 mL/kg/h	<1 mL/kg/h		>3 mL/kg/h		

Figure 5 – MEWS Calculation

2.3.4 ViEWS

The VitalPAC Early Warning Score (ViEWS) implements some changes on top of MEWS concerning the ranges of variables measured in each index level. The authors justify the changes due to prior knowledge about such measures' behavior and their relationship with possible diagnoses and previous research (PRYTHERCH et al., 2010). The index calculation remains the same as that mentioned for the NEWS and MEWS scoring systems, now considering Table 6.

ViEWS Score	3	2	1	0	1	2	3
Heart Rate, min		<=40	41-50	51-90	91-110	111-130	≥131
Respiratory Rate, min	≤8		9-11	12-20		21-24	>=25
Temperature, °C	<=35.0		35.1-36.0	36.1-38.0	38.1-39.0	≥39.1	
Blood pressure, mm Hg	<=90	91-100	101-110	111-249	>=250		
SpO2	<=91	92-93	94-95	>=96			
Any Supplemental Oxygen				No			Yes
							V Reacting to voice
				A			P Reacting to pain
Consciousness (APVU)				Alerta			U Unresponsive

Figure 6 – ViEWS Calculation

2.3.5 SAPS II

The Simplified Acute Physiologic Score (SAPS II) is a typical risk stratification score. It was first described in 1984 as an alternative to the APACHE scoring system. SAPS II was developed and validated in European and North American cohorts. It is calculated from the worst value of 12 routine physiological measurements during the first 24 hours of patient admission, information about previous health status, and some information obtained at admission. This calculation results in an integer point score between 0 and 163 and predicts hospital mortality between 0% and 100%. (GRANHOLM et al., 2016; GODINJAK, 2016). Figure 7 shows the calculation with the values of the variables.

2.3.6 SOFA and qSOFA

Sepsis is an organ dysfunction caused by a harmful organism's response to infection, leading to death (SINGER et al., 2016). Sepsis is one of the biggest causes of mortality in serious

Parameter	Value (score)						
Heart Rate (bpm)			< 40 (11)	40-69 (2)	70-119 (0)	120-159 (4)	> 160 (7)
Systolic Blood Pressure (mmHg)			< 70 (13)	70-99 (5)	100-199 (0)	> 200 (2)	
Temperature (C)					< 39 (0)	> 39 (3)	
PaO_2 (mmHg))/FIO_2 (%)	< 100 (11)	100-199 (9)	>200 (6)				
Urine output (ml)		< 500 (11)	> 500 (4)		> 1000 (0)		
Serum Urea (g/L)					< 28 (0)	28-83 (6)	>84 (10)
Na+ Serum Sodium (mEq/L)				< 125 (5)	125-144 (0)	> 145 (1)	
K+ Serum Potassium (mEq/L)				< 3 (3)	1-20 (0)	> 20 (3)	> 20 (3)
Bilirubin (mg/dL)					< 4 (0)	4-5.9 ($)	> 6 (9)
Bicarbonate (mEq/L)			< 15 (6)	15-19 (3)	> 20 (0)		20-29.9
White Blood Count (cells/mm^3)				<1 (12)	1-20 (0)		
Glasgow Coma Scale	< 6 (26)	6-8 (13)	9-10 (7)	11-13 (5)	14-15 (0)	> 20 (0)	
A. Total acute physiologic score							
B. Age points (years) (<40=+0, 40-59=+7, 60-69=+12, 70-74=+15, >=75-79=+16,>=80=+18)							
C. Chronic desease (metastatic carcinoma=+9, hemantologic malignancy=+10, AIDS=+ 17)							
D.Type of admission (scheduled surgical=+0, medical=+6, emergency surgical=+8)							
Total SAPS II Score = A + B + C + D. Minimum score = 0; maximum score = 163.							
10 = 1% death rate; 20 = 3.7% deth rate; 30 = 10.6% death rate; 40 = 24.7% death rate; 50 = 46.1% death rate							
60 = 68.1% death rate; 70 = 83.8% death rate; 80 = 92.5% death rate; 90 = 96.7%; >142 = 100% death rate							

Figure 7 – SAPS III Calculation

diseases worldwide (VINCENT et al., 1996b; FLEISCHMANN et al., 2016). Due to the severity of this health condition and to identify it early, there are studies to clearly define the medical terms related to sepsis as criteria for diagnosis (SEYMOUR et al., 2016). The most recent effort in this direction was made, in 2014, by a task force, called by the European Society of Intensive Care Medicine and Society of Critical Care Medicine, and made up of 19 specialists in intensive care medicine, infectious diseases, pneumology surgery (SINGER et al., 2016). The work carried out by the commission was methodologically planned and endorsed by several international societies around the world (SINGER et al., 2016).

It has five possible values assigned to the patient according to some objective criteria: the range of measures related to 6 systems of the human organism: respiratory, blood, hepatic, cardiovascular, central nervous, and renal (VINCENT et al., 1996b). Data used by the SOFA index are usually in the EHR. The set of variables recovered from the EHR and which make the calculation possible are partial oxygen pressure (PaO2), inspired oxygen fraction (FiO2), platelets, bilirubin, mean arterial pressure (MAP), Glasgow Coma Scale (GCS), Creatinine, urine quantity (SEYMOUR et al., 2016). Figure 8 presents the ranges of values for each measure used to calculate the SOFA index.

SOFA score requires laboratory examinations, and this motivated the proposition of qSOFA (quick Sepsis Related Organ Failure Assessment), also known as quickSOFA. It is based on only three variables, respiration rate, systolic blood pressure, and mental status, each having a single threshold to determine a subscore of 0 or 1, shown in Figure 9.

SOFA Score	0	1	2	3	4
PaO2/FiO2, mm Hg (kPa)	≥400 (53.3)	<150	<300 (40)	<50	<20
Platelets, ×103/µL	≥150	<150	<100	<50	<20
Bilirubin, mg/dL (µmol/L)	<1.2 (20)	1.2-1.9 (20-32)	2.0-5.9 (33-101)	6.0-11.9 (102-204)	>12.0 (204)
Cardiovascular system		PAM <70 mm Hg	Dopamina <5 ou Dobutamina (qualquer dose)	Dopamine 5.1-15 or Epinephrine ≤0.1 or Norepinephrine ≤0.1	Dopamine >15 or Epinephrine >0.1 or Norepinephrine >0.1
Glasgow Coma Scale	15	13-14	10-12	6-9	<6
Creatinine, mg/dL (µmol/L)	<1.2 (110-170)	1.2-1.9 (110-170)	2.0-3.4 (171-299)	3.5-4.9 (300-440)	>5.0 (>440)
Urine output, mL/d				<500	<200

Figure 8 – SOFA Calculation

Criterion	Score
Respiration rate ≥ 22	1
Systolic blood pressure ≤ 100 mmHg	1
Altered mental status, GCS < 15	1

GCS = Glasgow Coma Scale

Figure 9 – qSOFA Calculation

3 RELATED WORKS

Monitoring the evolution of vital signs, and analyzing their interrelationships is one of the possibilities presented to diagnose patients' health status changes. Few studies were found focused on predicting vital signs, at least to the knowledge of the author. The following are some of the works whose research reflects a data-based approach to analyzing the patient's vital signs' behavior, summarized in Table 3.

Forkan and Khalil (FORKAN; KHALIL, 2017) present a method for clinical decision support by monitoring and predicting the vital signs of monitored patients in their homes. The analysis of continuously measured signals and the interrelationship between them provide the prediction of patients' vital signs. The work of (CLIFTON et al., 2011) employs ML in the early detection of deterioration of vital signs of monitored patients hospitalized. The technique used is a Support Vector Machine (SVM) of one class. The vital signs' values and their combinations that do not indicate deterioration of patients' health are considered usual, and those outside this standard are regarded as abnormal. The performance of the model was tested with both synthetic data and data collected in a clinical study. The readmission of patients to the ICU can be indicative of different problems. Patients who are previously discharged and need to return to the ICU may present significant deterioration of their health status. Besides, readmissions may also indicate erroneous assessments when these patients are discharged. Considering this motivation, (LIN et al., 2019) carried out a comparative study between several methods to predict patients' readmission in the ICU. The data used by the study come from MIMIC-III, the same used in the present study, However, data considered in the study were notes on the patients' physiological conditions, previous diseases, and demographic data. The study experimented with LR (Logistic Regression), SVM, CNN (Convolutional Neural Networks) and LSTM.

In (RAJKOMAR et al., 2018), a prediction model, based on DL, is proposed on a clinical data structure, considering all the data stored on the patients and a specific set of data. The work employs a structure called Fast Healthcare Interoperability Resources (FHIR), where all patient data is stored and accessed by the prediction model. The argument presented is that the interrelationship between several data generates more accurate predictions, besides facilitating the pre-processing phase. The model was tested with its application in predicting mortality, hospital readmission within 30 days, and length of hospital stay.

Morales, Ribas, and Vellido (MORALES; RIBAS; VELLIDO, 2016) presents a study on the factors that influence the prognosis of sepsis. The work consists of two stages. The first one consists of studying the causal relationships between different measures registered in the systems to find relationships between these measures that indicate possible diagnoses (mortality and sepsis). This step is performed through a probabilistic approach, applying Bayesian networks. The second stage of the study is to predict the patient's risk of mortality and sepsis risk, leading to death or severe cognitive consequences. In this second stage, the work applied ANN (Artificial Neural Networks) and SVM (Support Vector Machine).

To early identify the imminent risk of death of a patient, (GHOSH et al., 2018) developed a predictor called the Early Deterioration Indicator (EDI), which seeks to detect signs of acute deterioration of the patient's health status through subtle changes in vital signs. According to the authors, these small variations often go unnoticed by prognostic indexes based on weight assignment. EDI uses a data-driven approach. The work of (CHURPEK; ADHIKARI; EDELSON, 2016) verified how much to include measures of trends in vital signs, rather than just the patient's current measurements, which can benefit the calculation of the prognostic indexes.

Table 3 – Related Works Summary

Related Work	Patients [1]	Measures [2]	Technique(s) [3]	Evaluation [4]	Results [5]
(LIN et al., 2019)	35,334	48,393	LSTM+CNN	AUC (Readmission)	79.1%
		48,393	LSTM	AUC (Readmission)	78.7%
		48,393	SVM	AUC (Readmission)	77.9%
		48,393	Logistic Regression	AUC (Readmission)	77.7%
(GHOSH et al., 2018)	*	14,282	Naive Bayes	AUC	76.0%
(RAJKOMAR et al., 2018)	114,004	326,221	LSTM	AUC (Mortality)	95.0%
			Attention-based	AUC (Readmission)	75.0%
			Time-aware Neural Network		
			Neural Network with Boosted	AUC (Lenght of stay)	86.0%
			Time-based Decision Stumps		
(FORKAN; KHALIL, 2017)	85	*	J48 Decision Tree	Hamming Score	90-95%
			Random Tree		
			SVM		
(CHURPEK; ADHIKARI; EDELSON, 2016)	*	269,999	Statistical Analysis	AUC	78.0%
(MORALES; RIBAS; VELLIDO, 2016)	412	*	Artificial Neural Networks	AUC (Mortality)	59.0%
			Artificial Neural Networks	AUC (Sepsis)	54.0%
			SVM	AUC (Mortality)	74.0%
			SVM	AUC (Sepsis)	77.0%
(CLIFTON et al., 2011)	19	1,500	SVM	Accuracy	95.0%
				Sensibility	98.0%
				Specificity	92.0%

[1] number of patients considered in the experiment;
[2] number of measures considered in the experiment;
[3] technique(s) considered by the model;
[4] evaluation measure(s) used to assess the results;
[5] result(s) of the evaluation technique(s) applied;
[*] not mentioned on source paper.

Table 3 summarizes this chapter of related works, recalling the volume of data considered in each experiment, the techniques applied, and evaluation results. The work of Forkan and Khalil (FORKAN; KHALIL, 2017) is the only multi-label classification model (PEREIRA et al., 2018), and the others are single-label classification methods, and they apply the AUC (Area Under The Curve ROC - Receiver Operating Characteristics) curve measure (OBUCHOWSKI; BULLEN, 2018) for the model evaluation. As stated before, the DeepSigns Model proposed here predicts the vital signs itself, instead of the specific diagnosis, which will serve as inputs to other methods, such as risk scores applied by hospitals. The values predicted are evaluated by the Mean Squared Error (MSE) because there is not one unique value to what the predicted one can be compared to, so the assessment evaluation score like AUC could not be applied.

The works presented in this chapter use vital signs in their models to monitor the patient's health status or some other related conditions. These works aim to detect a specific diagnosis, such as a specific disease or a specific situation, such as the patient's death or readmission to the ICU. When we do not know what we are looking for, this type of model may not be used effectively. On the other hand, there are a series of models based on heuristics, such as the prognostic indexes presented in section (O'BRIEN et al., 2020), used in hospital units' daily lives, but which lack data that benefit their application. Some studies have identified that prognostic indexes do not have efficient results because they consider data from a specific moment when patients arrive at the hospital. Ideally, it would be to have the patient monitored, and such an index calculated considering the value of their data, something that current hospital systems based on EHR can provide. Therefore, such scoring systems were not designed to consider the availability of EHR data, although they are used in hospitals and are well known by care teams. Taking advantage of this knowledge, added the possibility of providing data that make the indexes' calculation more efficient, is one of the gaps that the present work intends to fill.

4 THE DEEPSIGNS MODEL

This section presents the DeepSigns Model considering some design decisions that guided the model's architecture construction, and it also describes the Deep Learning algorithms chosen as part of the model's implementation. Although we present some methodological aspects in this chapter, the choice to do so was conscious and aimed at a better comprehension of the DeepSigns model.

4.1 Design Decisions

Different prognostic indexes are intended to predict the worsening of the patient's health status. However, most prognostic indexes consider vital signs measured at index calculation (CHURPEK; ADHIKARI; EDELSON, 2016). Therefore, incorporating vital signs' variability into the prognostic indexes' calculation presents an opportunity to improve such techniques. Nowadays, with the increasingly frequent use of EHR in hospitals, patients' historical data allows the application of computational techniques that consider such variability of vital signs measurements.

Some studies show that patient data measured over time has different relationships at distinct moments and influence each other over time, leading to alternative diagnoses, depending on the combination of specific values at a given time (FORKAN; KHALIL, 2017; CLIFTON et al., 2011). Therefore, the prediction of patients' data based on the combination of their variation in time may be an alternative to diagnose diseases. The focus is not predicting the diagnosis itself, but rather the evolution of patient data to identify a possible diagnosis for them in the future. There are few publications on this subject to the best of our knowledge, so this work considers such an alternative an excellent research opportunity.

Considering that patient data stored in EHR contains temporal information about each measurement stored, the EHR records can be interpreted as a time series and serve as inputs to techniques capable of processing such type of information. Besides, the time series' information allows considering historical data in the patient's health record to predict their future data.

4.2 Architecture

DeepSigns model is presented in Figure 10. The model aims to predict patients' vital signs based on historical EHR data from different sources and calculate different prognostic indexes based on predicted vital sign values. The purpose is that medical teams can compare the results of different prognostic indexes and choose the referral considered more appropriate for the patient.

Figure 10 depicts a scenario where there is a vital signs prediction model trained with patient data from EHR. So, as patients arrive in a hospital unit, ICU, for example, their data are stored in

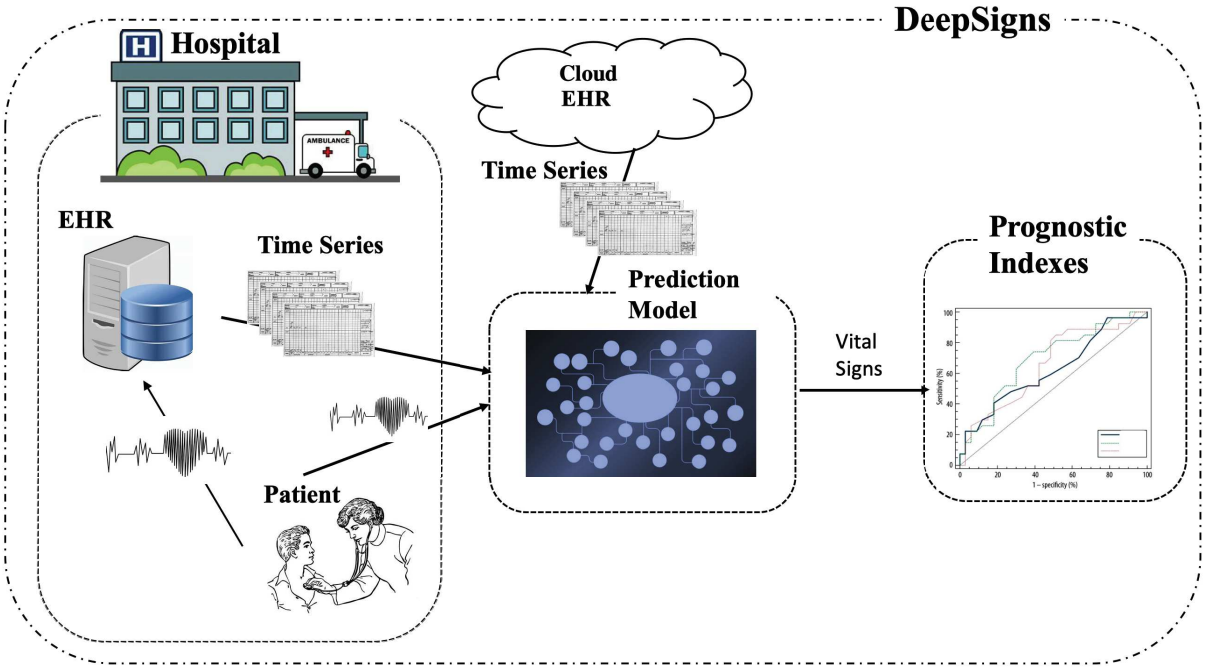

Figure 10 – DeepSigns - General View

the hospital's EHR and sent to the DeepSigns model to have their data predicted. The different prognostic indexes used in the hospital can apply the predicted vital signs to predict the patient's health status severity.

4.3 Deep Learning Algorithms

The DeepSigns model predicts the evolution of patients' data and was constructed based on LSTM networks, consuming EHR data organized in time series. Figure 11 presents an overview of this prediction model. At the first stage, we transform the EHR data in a time series format and train n LSTM networks, one for each predicted data.

DeepSigns model uses a different LSTM network to predict each vital sign, with all other vital signs as inputs. Each line in Table 4 presents a different LSTM network, its inputs, and the predicted attribute in the last column. Our model uses APACHE II attributes, which works with 15 attributes, including age. The DeepSigns model has 15 minus 1 LSTM networks because it does not predict age.

Deep Learning models have to set initial parameter values, which are not learned through the model execution, called the hyperparameters (GOODFELLOW; BENGIO; COURVILLE, 2016). These hyperparameters are reset many times as the model performance is evaluated until the best configuration is achieved. The LSTM Network used the following initial hyperparameters configuration (Table 5). Table 5 shows the first configuration use by DeepSigns Model.

The first stage concerns the extraction of EHR data, transforming them in time series format, and using these series to train and validate LSTM networks, one for each data to be predicted

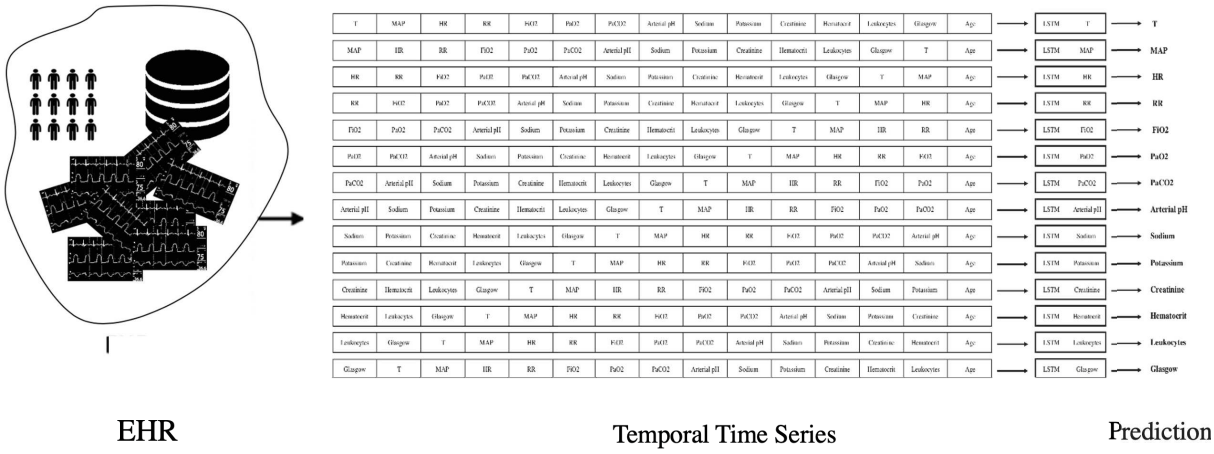

Figure 11 – DeepSigns Model - LSTM networks, consuming EHR data formated as time series

Table 4 – DeepSigns Model: LSTM networks, its inputs, and output - attribute predicted

	att1	**att2**	**att3**	**att4**	**att5**	**att6**	**att7**	**att8**	**att9**	**att10**	**att11**	**att12**	**att13**	**att14**	**Age Attribute**	**Predicted**
1	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Age	Temperature
2	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	Age	MAP
3	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	Age	HR
4	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	Age	RR
5	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	RR	Age	FiO2
6	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	RR	FiO2	Age	PaO2
7	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	RR	FiO2	PaO2	Age	PaCO2
8	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Age	Arterial pH
9	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Age	Sodium
10	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Age	Potassium
11	Creatinine	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Age	Creatinine
12	Hematocrit	Leukocytes	GCS	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Age	Hematocrit
13	Leukocytes	GCS	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Age	Leukocytes
14	GCS	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	Age	GCS

Table 5 – LSTM Configuration

Hyperparameter	Value
layers	3 (input, hidden, output)
LSTM Units	50
Loss Function	Mean Absolute Error
Optimizer	Adam
Activation	ReLU
Batch Size	72

by the model. In this study, the data to be predicted are the patients' vital signs presented in Table 2.

The model application, presented in Figure 12, starts on the arrival of a new patient in the hospital environment. Vital signs of the new patient will be measured at the arrival time and will enter the model to predict that patient's vital signs. After the prediction phase, the predicted data can be passed on to any prognostic index calculation system based on such vital signs. Thus, the calculation of the index will be performed for the patient's predicted data, which will make it possible to anticipate what will happen to the patient in the future, thus allowing the action of care teams. The predictions with a positive impact on decision making will feedback to the

model.

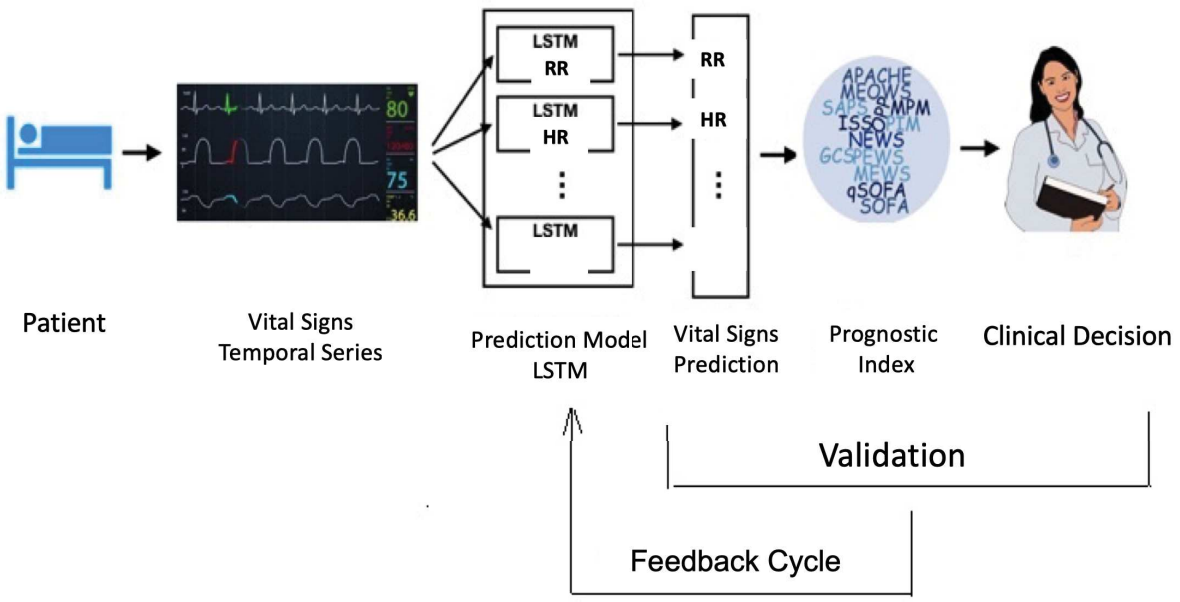

Figure 12 – Decision support model based on prediction of vital signs.

The differential of the proposed model in this study lies in the fact that the prognostic index calculation will be performed based on the predicted values of vital signs, based on the historical data series. With this, it is possible to predict what will be the patient's situation in the future. Commonly, prognostic indexes are used to identify the risks of patients' condition based on the current measurements of their vital signs (CHURPEK; ADHIKARI; EDELSON, 2016). However, this form of calculation, for some diseases, considering the severity of the patient's state, can only be diagnosed when there are few resources to promote recovery.

Figure 13 shows the applicability of the proposed model. The cases of patients whose prognostic indexes indicate low severity (normal health status) at time t but high severity at time $t + 1$ are those that could benefit from the proposed model since they would already be identified as severe at time t and due treatment could be initiated immediately.

4.4 Data Source

Data used in this work came from unidentified medical records in the MIMIC-III (Medical Information Mart for Intensive Care) database (JOHNSON et al., 2016). MIMIC-III has data from approximately 58,000 hospitalizations of more than 38,000 patients hospitalized between 2001 and 2012 in one of the adult ICU at Beth Israel Deaconess Medical Center: Coronary Unit, Cardiac Surgery Recovery Unit, Medical ICU, Surgical ICU and Surgical ICU - Trauma. The database includes demographic information, vital signs, laboratory tests, procedures, medication administration, discharge records, imaging examinations, and external government sources' mortality information.

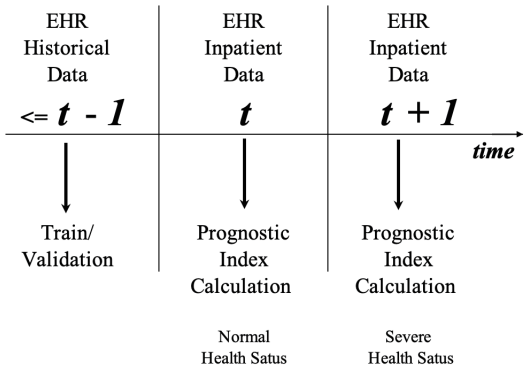

Figure 13 – Applicability of the proposed model

Figure 14 provides a general overview of the MIMIC-III data source. The data from the different sources are gathered in a repository and, before being included in the MIMIC database, they are deidentified and converted into the formats required for the MIMIC-III scheme.

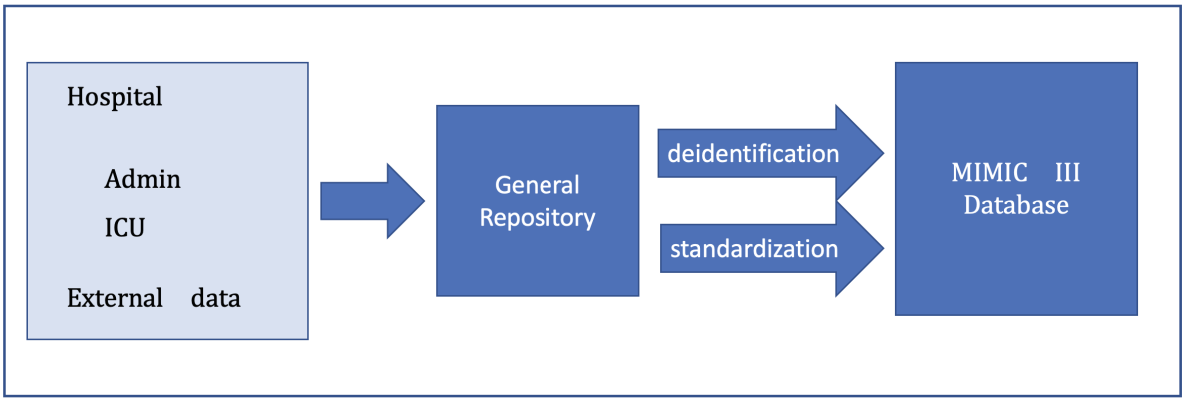

Figure 14 – Overview of MIMIC-III - data sources

MIMIC-III data are organized around 4 main elements: patients, hospitalizations, ICU admissions and patient events. Figures 15 and 16 show a schematic view of the MIMIC-III database scheme, showing the tables, their attributes and the relationships between them. Figure 16, specifically, presents a view of how events are attributed to patients; whe attributes subject_id and hadm_id play a central role in identifying the records and relationships between the tables.

In summary, Table 6 presents the main concepts stored in MIMIC-III and the relations where these concepts are stored.

MIMIC-III database was chosen at the beginning of this research, and it demanded some time and several procedures to use it effectively. MIMIC-III scheme is relatively detailed, so as the data set, it was not straightforward to apply it to the DeepSigns model. Currently, the MIMIC-IV (version v0.4) (JOHNSON et al., 2020) contains data from patients admitted between 2008 and 2019 to the Israel Deaconess Medical Center's emergency department or ICU.

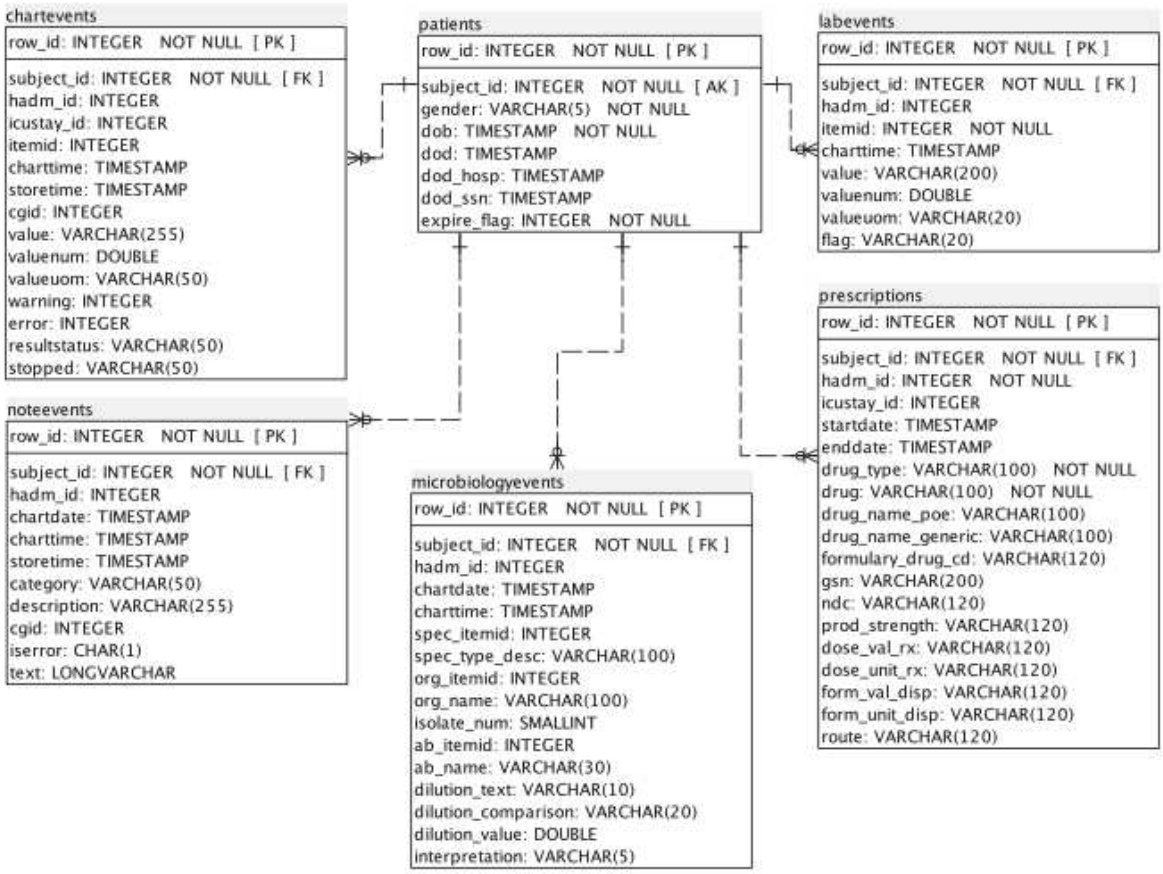

Figure 15 – Relational scheme of MIMIC-III - Patient data

Table 6 – MIMIC-III Concepts and Tables

Concept	**Database Table**
Measurement type	D_ITEMS
Patient hospitalizations	ADMISSIONS
Sequence of patient interactions in the hospital	TRANSFERS
Patient demographics	ADMISSIONS
ICD-9 code [1]	DIAGNOSES_ICD
Medication administration	OUTPUTEVENTS INPUTEVENTS_CV INPUTEVENTS_MV
Prescriptions	PRESCRIPTIONS

[1] International Classification of Diseases, Ninth Revision

4.5 Data Selection

The data retrieved from MIMIC-III for the development and validation of the model proposed in this work are those that constitute the calculation of APACHE II and presented in Table 2. Initially, it was necessary to recover the MIMIC-III data related to the patients' mea-

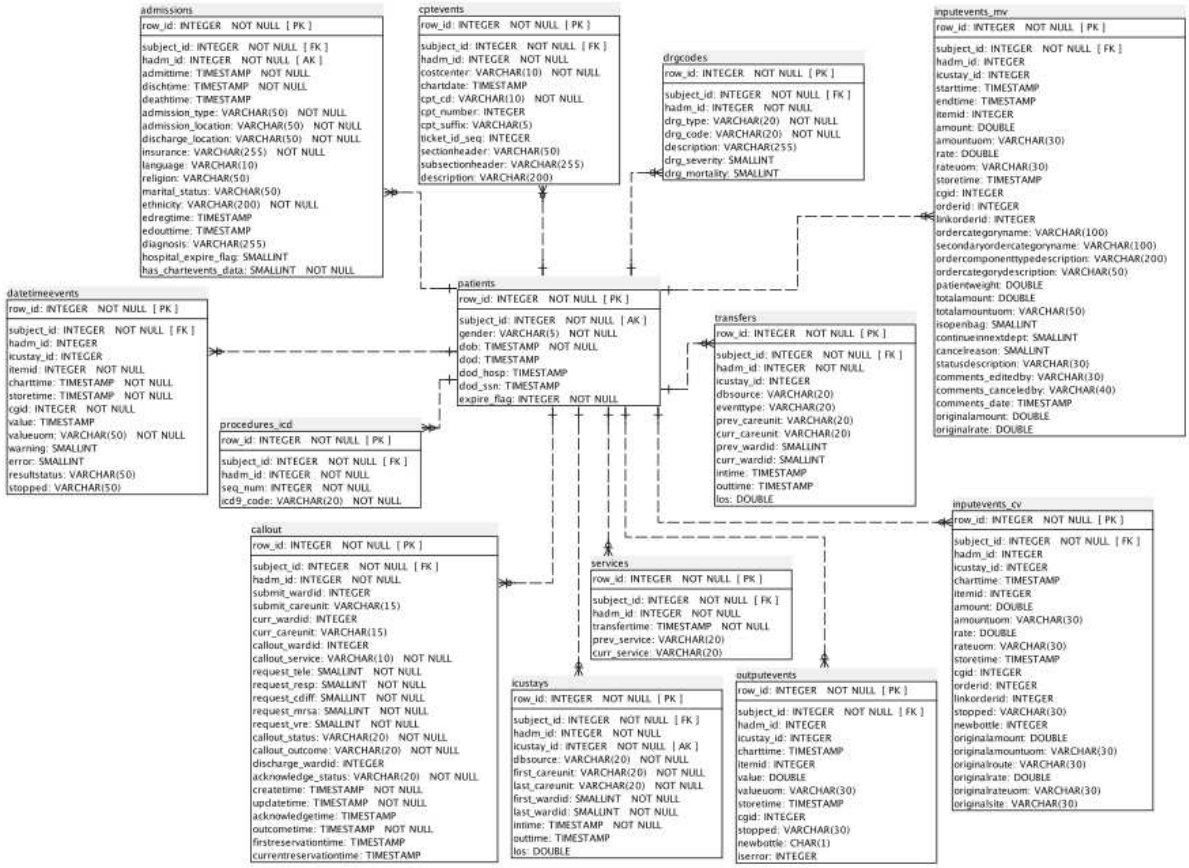

Figure 16 – Relational scheme of MIMIC-III - Patient events

surements at each hospitalization and structure them in time series. Besides, in order to perform the validation, the diagnostic data were also retrieved from the database. The model allows data from other data sources to be used, following the same methodology presented here.

The attributes related to the hospitalization identification, measurement moment, measure, measure numeric value, and unit of the measure were retrieved from MIMIC-III's tables. The measure retrieved were the ones of interest for the time series assembly, those representing the variables considered in the APACHE II calculation. Table 7 shows an example dataset resulting from this first step.

4.6 Preprocessing

Real-world health data often present noises, distortions, or missing values that impact the performance of algorithmic solutions (FANG et al., 2016). Therefore, predictive models' construction requires deciding in advance which criteria should be adopted for data selection. So, the first filter applied in data recovery consisted of records of adult patients (age greater than or equal to 18 years) and hospitalization time greater than 24 hours. Besides, only records containing non-zero values were retrieved for all variables of interest.

EHR data can be interpreted as a time series (CHE et al., 2017), but it is necessary to

Table 7 – Example Dataset for Inpatient 100,006

hadm_id [1]	**charttime** [2]	**itemid** [3]	**valuenum** [4]	**valueuom** [5]
100,006	06/04/08 16:30	618	22	BPM
100,006	06/04/08 16:30	456	100.667000	mmHg
100,006	06/04/08 16:30	678	97,000	Deg. F
100,006	06/04/08 16:30	198	15,000	points
100,006	06/04/08 17:00	618	20	BPM
100,006	06/04/08 17:00	456	85.333298	mmHg
100,006	06/04/08 18:00	456	88.000000	mmHg

[1] hadm_id is the MIMIC-III's attribute name for the hospitalization identification;
[2] charttime is the MIMIC-III's attribute name for the moment the measure was inserted on database;
[3] itemid is the MIMIC-III's attribute name for the measure taken from patient;
[4] valuenum is the MIMIC-III's attribute name for the measure value taken from patient;
[5] valueuom is the MIMIC-III's attribute name for the measure unit.

restructure such data, which were stored according to the relational model (Table 7), as it is the case of MIMIC-III, in the form of a time series. For this purpose, a pre-processing phase has performed the steps as follows:

- organization of data on a temporal basis;
- division into fixed time windows;
- calculation of the average of each measure within the time window;
- imputations of values for missing data;
- separating the data set into a training, test, and validation data.

In the work presented here, we experimented with many variations of time windows and imputation forms, considering that this may affect the results of the prediction (LIPTON; KALE; WETZEL, 2016) and the section 5 presents the results. Each time window has data from one admission from the same patient.

4.7 Computational tools

For this work, experiments were performed using TensorFlow framework (version 1.7.0), Keras API interface (version 2.1.5), Python programming language (version 2.7.12), and Jupyter Notebook web application (version 5.4.1), on a computer with Intel Core Processor (Haswell, no TSX, IBRS) – 64 bits - 2GHz - 8 cores - 12GB RAM – and Ubuntu 16.04.4 LTS.

4.8 Evaluation

For the vital signs prediction validation, considering that data was in the form of a time series, the measure employed for evaluation was the Mean Squared Error (MSE), which is the sum of the differences between the estimated and the actual value of the data, divided by the number of terms. The objective is to measure how close an erroneous prediction is to the real value. That is, to evaluate the distance between the predicted (estimated) value and the actual value of the measurement.

In order to ensure the generalization capability, i.e., to produce accurate predictions when analyzing new unlabeled data, ML models need to be trained, validated, and tested on independent data sets (HAYKIN et al., 2009). In the case of DL, the default validation method is Holdout. This method consists of dividing the data into two independents (mutually exclusive) subsets: training and validation. Typically, 60% of data are used for training (learning), 10% to evaluate performance during training (validation) through accurate measurements and error, and 30% for a final evaluation of the predictions (generalizations) performed by the model (HAYKIN et al., 2009). For DeepSigns, We divided our dataset into two, 70% of data were used for training the model, 60% for learning and 10% to evaluate, and 30% for testing. The datasets were independents, as Figure 17 shows.

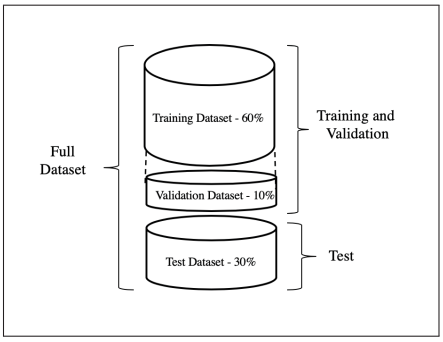

Figure 17 – Holdout method for training and testing

In detail, the evaluation of the prediction model proposed in this work involved the following steps, also detailed in Figure 18.

- Training of LSTM networks with patient data before time $t-1$;
- A new patient will be considered as the one with the data at time $t-1$;
- application of the model to predict the patient's vital signs for time t;
- calculation of prognostic index with vital signs predicted for time t;
- validation of the calculated prognostic index compared to what happened to the patient at time t, from the historical information.

In Figure 18, time t refers to the last vital signs' measures recorded for the patient; so, time $t-1$ refers to the measures before it.

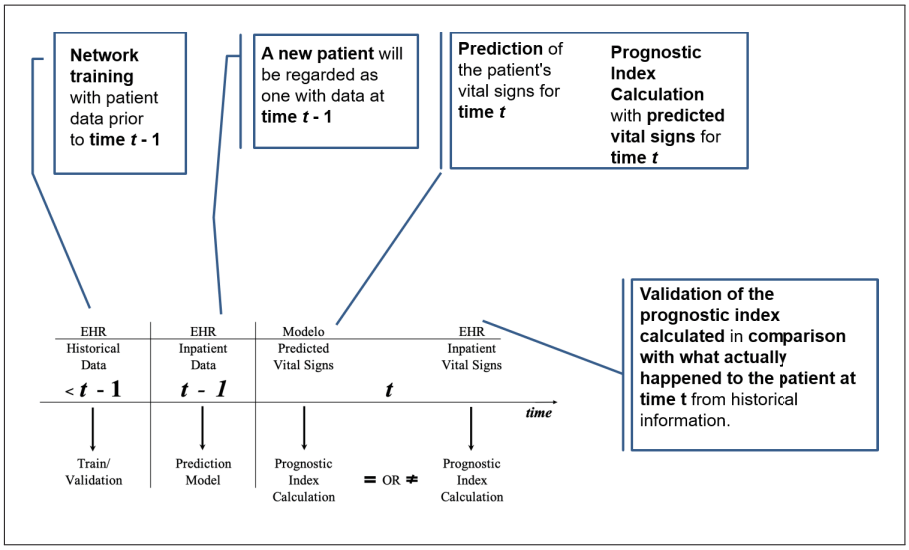

Figure 18 – Validation

4.9 Chapter Summary

This chapter introduced the DeepSigns model, detailing the following elements:

- Design Decisions: time series from EHR data.
- Architecture: Prognostic indexes calculations through predicted vital signs.
- Data Source: MIMIC-III database.
- Data Selection: variables applied in scoring systems' calculation.
- Preprocessing: organization and aggregation of data.
- Computational tools: TensorFlow framework, Python programming language and Jupyter Notebook.
- Evaluation: practical application evaluation.

5 RESULTS AND DISCUSSION

This section presents the results of the experiments with the DeepSigns model. The first stage, as described in Section 4.5, is concerned with retrieving the desired data from the MIMIC-III database. All data from MIMIC-III, whose APACHE variables were collected and registered, were included in this study, resulting in 12,735,811 records, representing measures from distinct patients' admissions taken during hospitalization. These measures are likewise the ones presented in Table 8 for patient 100,006 as an example. Table 8 presents a view of what would be the whole dataset since it is impossible to present the dataset itself. Table 8 also shows that data from MIMIC-III used in this work varies significantly since most of them are continuous variables.

Table 8: Example Dataset for all Patients

Record	**hadm_id**	**charttime**	**itemid**	**valuenum**	**valueuom**
1	100,006	06/04/08 16:30	618	22	BPM
2	100,006	06/04/08 16:30	456	100.667000	mmHg
3	100,006	06/04/08 16:30	678	97,000	Deg. F
4	100,006	06/04/08 16:30	198	15,000	points
5	100,006	06/04/08 17:00	618	20	BPM
6	100,006	06/04/08 17:00	456	85.333298	mmHg
7	100,006	06/04/08 18:00	456	88.000000	mmHg
8	100594	04/08/17 21:00	198	15	points
9	100594	04/08/17 22:00	211	84	BPM
10	100594	04/08/17 22:00	456	80	mmHg
...	100594	...	...	...	...
110	100594	07/08/17 13:00	211	68	BPM
111	100594	07/08/17 13:00	618	17	BPM
112	100594	07/08/17 14:00	211	70	BPM
113	100594	07/08/17 14:00	618	19	BPM
...	...	...	...	...	...
10,250	102432	22/09/02 05:30	211	75	BPM
10,251	102432	22/09/02 05:30	456	63	mmHg
10,252	102432	22/09/02 07:00	211	72	BPM
10,253	102432	22/09/02 07:00	456	64.66670227	mmHg
10,254	102432	22/09/02 07:00	678	96.80000305	Deg. F
10,255	102432	22/09/02 07:00	198	15	points
...	102432	...	...	...	...
10,900	...	...	...	...	...

...	...	...	...	...	...
200,000	...	...	...	...	...
...	...	...	...	...	...
500,000	...	...	...	...	...
...	...	...	...	...	...
1,675,000	...	...	...	...	...
...	...	...	...	...	...
9,567,400	...	...	...	...	...
...	...	...	...	...	...
...	199727	...	...	...	...
12,735,807	199727	23/06/35 17:00	211	81	BPM
12,735,808	199727	23/06/35 17:00	456	65	mmHg
12,735,809	199727	23/06/35 19:00	456	68	mmHg
12,735,810	199727	23/06/35 19:00	618	28	BPM
12,735,811	199727	23/06/35 19:00	211	79	BPM

Complementary, as presented in section 4.6, the data were organized in 27,736 admissions. After that, we randomized these admissions, every time, to select the volume of data to enter the model. We created six different volume datasets with 100, 250, 500, 1000, 2,500 and 5,000 admissions each. Finally, we employed the holdout method for validation and repeated our results for three iterations. The next stage was to split the data set in fixed time windows to create the time series datasets. We experimented with fixed time windows of 1, 5, 10, 15, and 30 minutes.

We split it in fixed time windows for each volume dataset and created the time series to enter the LSTM networks. Table 9 shows a time series example for the patient 100,006 built from data presented in Table 7 and time window of 1 minute.

As aforementioned, we experimented with fixed time windows of 1, 5, 10, 15, and 30 minutes. DeepSigns has an approach to predict vital signs (and other variables) to feed the calculation of scoring systems. There are many scores for different diagnoses, and each disease evolves differently. Besides, more than one score can be recommended to help diagnose the same disease or condition.

Sepsis, for example, is caused by the body's response to an infection and leads the patient to a critical and life-threatening condition (RHODES et al., 2017). The reference committees responsible for defining sepsis and the possible conduct of medical teams recommend using more than one prognostic index, in this case, Systemic Inflammatory Response Syndrome (SIRS) (BONE; SIBBALD; SPRUNG, 1992), SOFA, and qSOFA, for its early diagnosis. qSOFA is a prognostic index that is easy and quick to calculate and employs only three variables (respiratory rate, systolic blood pressure, and index on the Glasgow coma scale). However, qSOFA does

Table 9 – Temporal Serie for patient 100,006

line	Temperature	MAP	HR	RR	FiO2	PaO2	PaCO2	Arterial pH	Sodium	Potassium	Creatinine	Hematocrit	Leukocytes	GCS	Age	APACHE II
1	99	90	89	18	21%	90	40	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
2	97	101	115	22	21%	90	40	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
3	97	85	117	20	21%	90	40	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
4	97	85	117	20	21%	90	40	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
5	97	88	122	18	21%	90	40	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
6	97	88	122	18	21%	90	40	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
7	97	88	117	23	21%	90	40	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
8	97	88	117	23	21%	189	48	7,3	139.5	4.45	1.0	38%	8.95	15	49	5
9	97	105	117	18	21%	189	48	7,3	139.5	4.45	1.0	38%	8.95	15	49	5
10	97	105	117	18	21%	189	48	7,3	139.5	4.45	1.0	38%	8.95	15	49	5
11	97	80	118	20	21%	189	48	7,3	139.5	4.45	1.0	38%	8.95	15	49	5
12	97	80	118	20	21%	189	48	7,3	139.5	4.45	1.0	38%	8.95	15	49	5
13	97	82	115	18	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
14	97	82	115	18	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
15	97	100	117	20	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
16	97	100	117	20	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
17	98	80	112	22	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	5
18	98	80	112	22	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	3
19	98	81	105	21	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	3
20	98	81	105	21	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	3
21	98	85	103	19	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	3
22	98	85	103	19	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	3
23	98	84	106	21	21%	69	37	7,4	139.5	4.45	1.0	38%	8.95	15	49	3
24	98	84	106	21	21%	69	37	7,4	139.5	4.45	0.6	31%	13.60	15	49	4
25	97	88	109	18	21%	69	37	7,4	131.0	3.70	0.6	31%	13.60	15	49	4
26	97	88	109	18	21%	69	37	7,4	131.0	3.70	0.6	31%	13.60	15	49	4
27	97	88	100	16	21%	69	37	7,4	131.0	3.70	0.6	31%	13.60	15	49	4
28	97	88	100	16	21%	69	37	7,4	131.0	3.70	0.6	31%	13.60	15	49	4
29	97	89	95	19	21%	69	37	7,4	131.0	3.70	0.6	31%	13.60	15	49	4
30	97	89	95	19	21%	69	37	7,4	131.0	3.70	0.6	31%	13.60	15	49	4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
228	98	115	105	19	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	6
229	98	103	122	17	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	6
230	98	103	122	17	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	7
231	98	95	119	26	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	7
232	98	95	119	26	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	6
233	98	96	136	21	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	6
234	98	96	126	21	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	6
235	98	96	124	20	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	6
236	98	96	124	20	40%	105	53	7,4	134.0	4.40	0.6	31%	9.8	15	49	6

not perform best in identifying sepsis (MD et al., 2018; GREEN et al., 2019). Other indexes, such as SOFA, are more complex and require variables in their calculation that are not always available at the bedside. The application of the DeepSigns model would be an alternative, as the model could predict the unavailable variables.

In sepsis, the first hour after its detection is crucial to start the treatment with antibiotics (DUGAR; CHOUDHARY; DUGGAL, 2020). So, smaller time windows can guarantee more data to the network training. If we take 30 minutes time window, three hours generates six samples. If the patient staying on the hospital is recent, let us say 24 hours, it would lead to only 48 samples to be used by the prediction model. Otherwise, smaller time windows could benefit the model with more training samples. Furthermore, we chose these time windows after discussions with physicians involved in the project. They think that they are useful for Emergency Departments and Intensive Care Unities.

The option for the 1-minute window was the one with the best results observed. The analysis presented in Figures 19 and 20 shows the results obtained by successive evaluations, and the best

results are always the ones observed for 1-minute windows for all the training datasets sizes. Forkan, Khalil, and Atiquzzaman (2017) (FORKAN; KHALIL; ATIQUZZAMAN, 2017) also carried out studies about an adequate window size for selecting the data to train the prediction model. They performed tests with windows of 8, 10, 12, 16, 20 minutes, and evaluated the window size of 10 minutes as the ideal size. Even the two models being different, the second smaller time window tested yielded the best results.

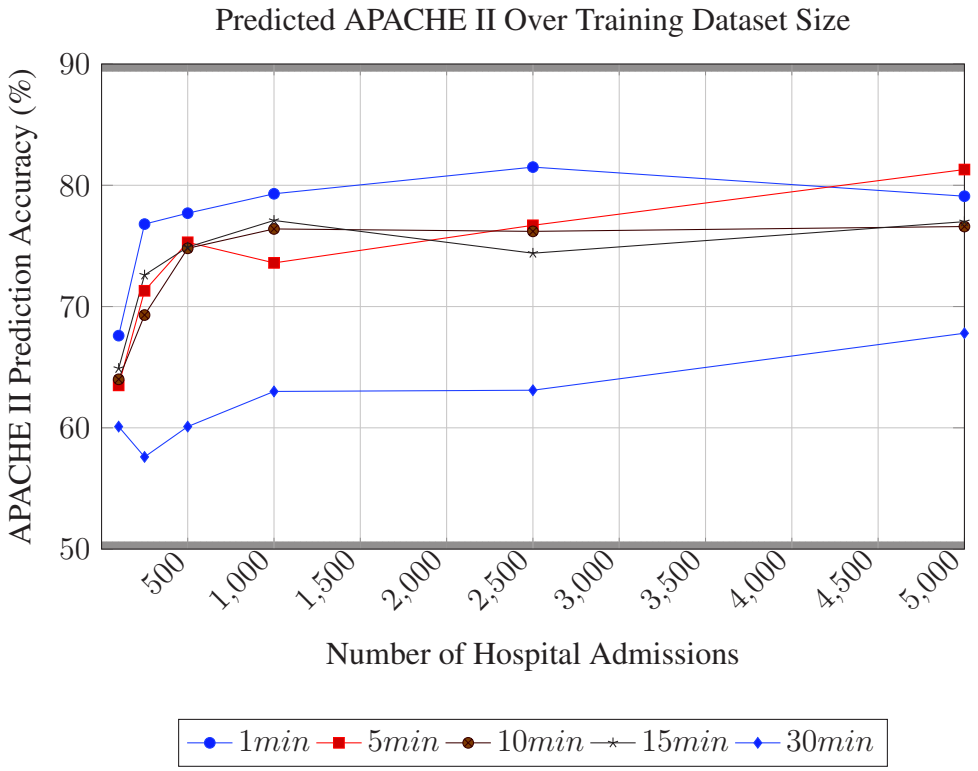

Figure 19 – Accuracy of APACHE II with predicted data over the training dataset size range from 0 - 5.000 hospital admissions

Figure 19 depicts the percentage of tested cases over the training dataset size, where the APACHE II index was calculated correctly. Besides, Figure 20 shows the best results like the ones trained with the biggest dataset sizes and the lowest window time (1 min - 2.500 admissions and 5 min - 5.000 admissions).

We also performed four different experiments about inputation values, replacing the missing values by the reference value for each vital signal, value zero, or constant values -1 or -100. The choice of such values is considered in the literature (LIPTON; KALE; WETZEL, 2016). As can be seen in Figure 21, which presents the MSE for each vital sign considered in this study, the best prediction (lowest value) was obtained when the reference value for each vital sign was inputted. In this work, the reference value is considered an expected value for vital signs and was defined as the median value from the APACHE II normal range for each vital sign. For example, in the APACHE II calculation, the standard temperature value is in the range from 36.0 to 38.4; in this case, we used the value 37.2 as the reference value.

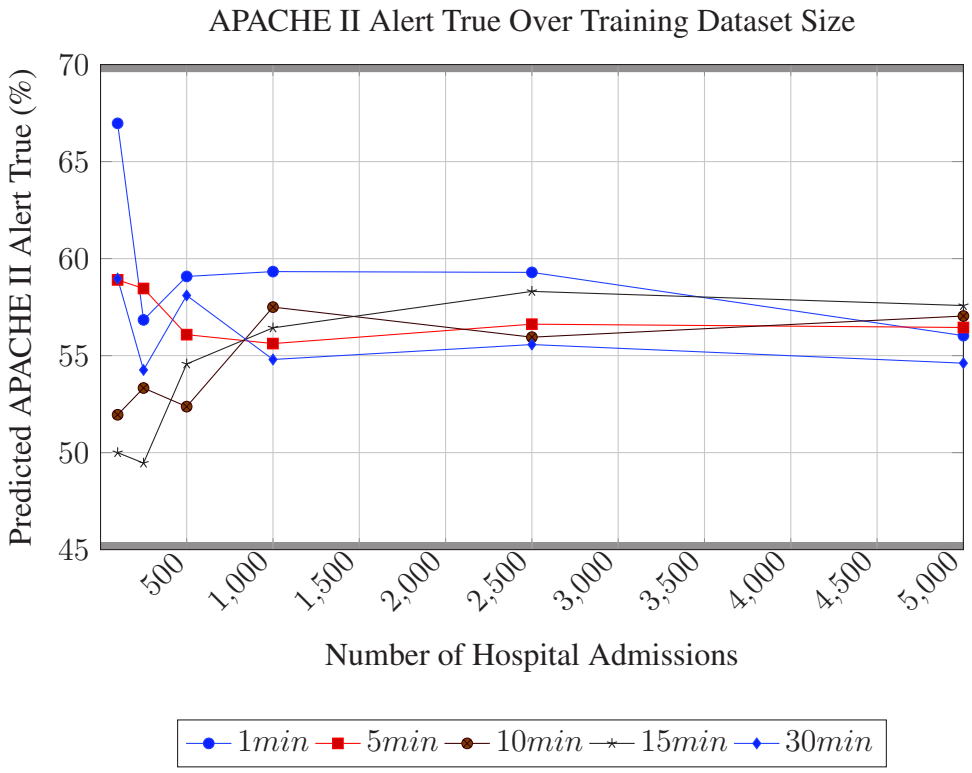

Figure 20 – percentage of APACHE II Alert True over the training dataset size of hospital admissions.

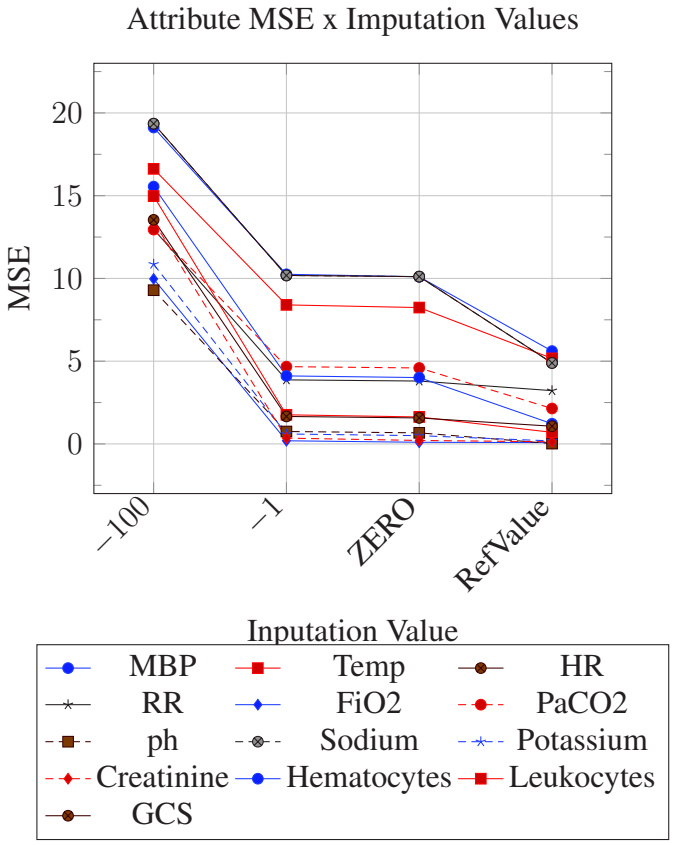

Figure 21 – MSE calculated for each predicted attribute and the corresponding imputation value

Dataset size and time window for aggregating data from the EHR are critical definitions to provide a viable procedure for training LSTM networks. Combining larger dataset sizes and smaller time windows lead to better results and a much longer time training the model. After experimenting with different datasets and time window sizes, we observed that results did not get much improvement after dataset size greater than 1,000. Recall Figures 19 and 20. Figure 22 plots the tradeoff between prediction accuracy and time (minutes) that takes to train the model. The best accuracy occurs in the graph's far-right point, and the more time-consuming training process occurs at the highest point.

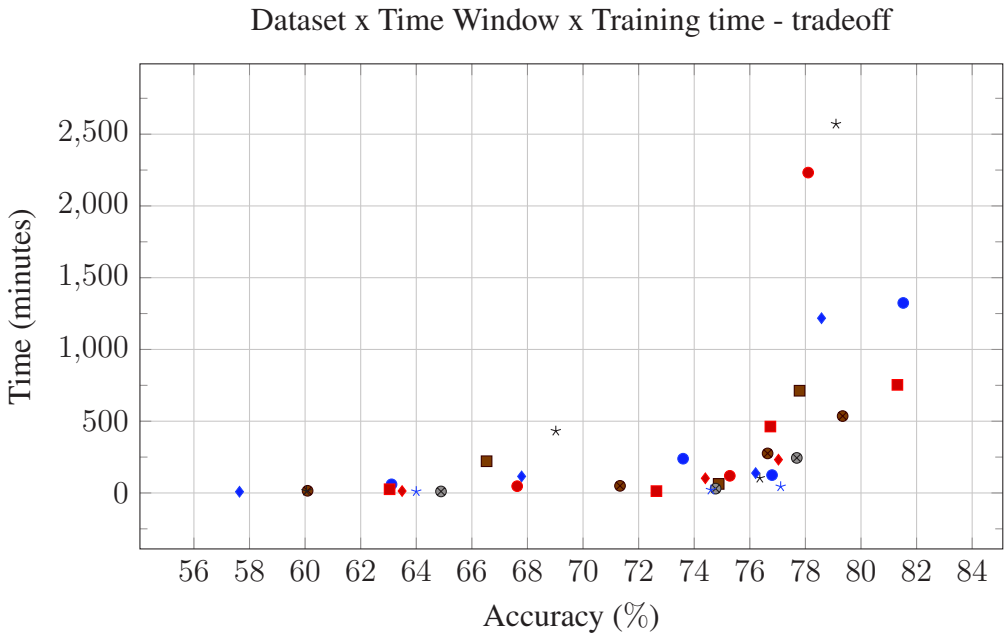

Figure 22 – Dataset x Time Window x Training time - tradeoff

Table 10 presents the top 10 best results plotted in Figure 22. The best result is the first line representing a dataset of 2,500 admissions, and it takes 22 hours to train. In contrast, a less time-consuming training process is the line with 1,000 admissions, and it takes 1 hour to train. Considering the DeepSigs model expects to aggregate more data over time, as new patients are admitted to the hospital, the time to retrain the model is an essential and critical parameter to DeepSigns to work correctly. That is why the next experiments will use the dataset with 1,000 patients' admissions and a 15-minute time window.

In the next step, trying to achieve a better network configuration for the problem of CDSS that we are dealing with in this work, we experimented with different number of neurons and epochs configurations. We experimented with 10, 15, 20, 25, and 50 epochs and 50, 75, and 100 neurons. Figure 23 summarizes the results, showing the best results for the APACHE II prediction accuracy with 50 epochs and 50 neurons LSTM configuration.

After training the networks with a dataset with 1,000 patients' admissions, a 15-minute time window, 50 epochs, and 50 neurons, Table 11 presents an evaluation of each network, showing the MSE calculated for each of them. The magnitude of MSE value varies according to the

Table 10 – Tradeoff Top 10

Admissions (Num)	Time Window (min)	Accuracy (%)	Time (h)
2500	**1**	**81.50**	**22**
5000	5	81.30	13
1000	1	79.30	9
5000	1	79.10	43
10000	10	78.60	20
10000	5	78.10	37
10000	15	77.80	12
500	1	77.70	4
1000	**15**	**77.10**	**1**
5000	15	77.00	4

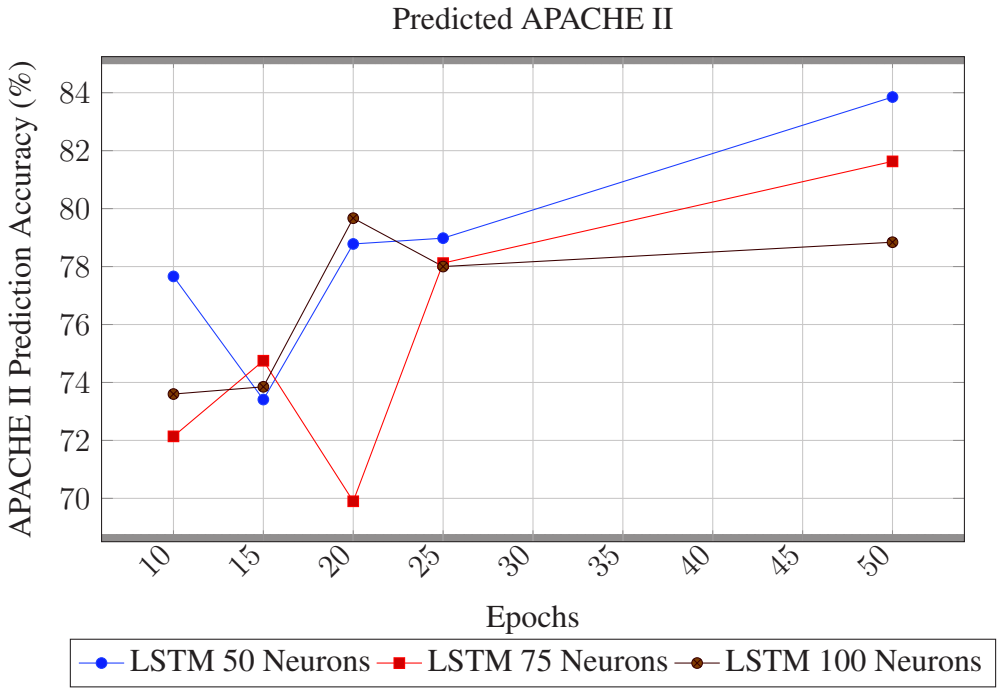

Figure 23 – Accuracy of Predicted APACHE II over Epochs.

attribute predicted since MSE is a measure that considers the distance between the predicted value and the real value. The table shows the reference value for the attribute predicted by the network, the range of values considered as typical values by the APACHE II score, and also the MSE of the network. Thus, it is possible to observe how much each network allowed the approximation between the real and the predicted value.

We compared our model's results to the Ordinary Least Squares (OLS) (HUTCHESON, 1999), and the results show that DeepSigns outperforms OLS (Figure 24). OLS is a statistical

Table 11 – LSTM MSE

	LSTM	Reference Value	Range	MSE
1	**T**[1]	37.2	36.0 - 38.4	0.09
2	**MAP**[2]	89.5	70 - 109	1.43
3	**HR**[3]	89.5	70 - 109	1.91
4	**RR**[4]	18	12 - 24	0.76
5	**FiO2**	0.21	0.21*	0.01
6	**PaO2**	90	80 - 100	4.13
7	**PaCO2**	40	35 - 45	0.29
8	**Arterial pH**	7.395	7.33 - 7.49	0.01
9	**Sodium**	139.5	130 - 149	0.17
10	**Potassium**	4.45	3.5 - 5.4	1.60
11	**Creatinine**	1	0.6 - 1.4	0.03
12	**Hematocrit**	37.95	30 - 45.9	0.27
13	**Leukocytes**	8.95	3 - 14.9	0.31
14	**GCS**	15	15**	0.16

[1] Temperature [2] Mean Arterial Pressure [3] Heart Rate [4] Respiratory Rate
* In ambient air, at sea level, we have a FiO2 of 21%. ** The minimum score is a three, indicating a deep coma or a brain-dead state. The maximum is 15, which indicates a fully awake patient.

method to estimate unknown parameters in a linear regression model. It does the prediction minimizing the sum of the squares of the differences between the variable's values being observed in a given dataset and those predicted by the linear function. Selvin et al. (SELVIN et al., 2017), also compared his model developed based on RNN with a linear prediction model, the ARIMA model (BOX et al., 2015), and obtained better results.

As proposed in Figure 13 of Section 4.8, the model applicability validation considered the number of tested cases where the prognostic index calculation based on predicted vital signs indicates a significant variation (KNAUS et al., 1991). These are the cases where an anticipated intervention would possibly revert the patient's health deterioration. Figure 20 depicts the percentage of tested cases over the training dataset size, where an alert would pop in anticipation according to the above definition. As shown in Figure 20, the number of new alerts generated based on predicted vital signs would be between 50% and 60%.

The predictive model of vital signs presented here has a low cost of development because, as DL models make full use of data and its pattern of variation and relationships, there is no need to identify the most relevant predictors for the problem. DL techniques allow automatic generation of predictors, which reduces the need for human intervention at this stage of the process (RAVI et al., 2017).

Furthermore, the possibility of processing historical data of the patient's vital signs aligns

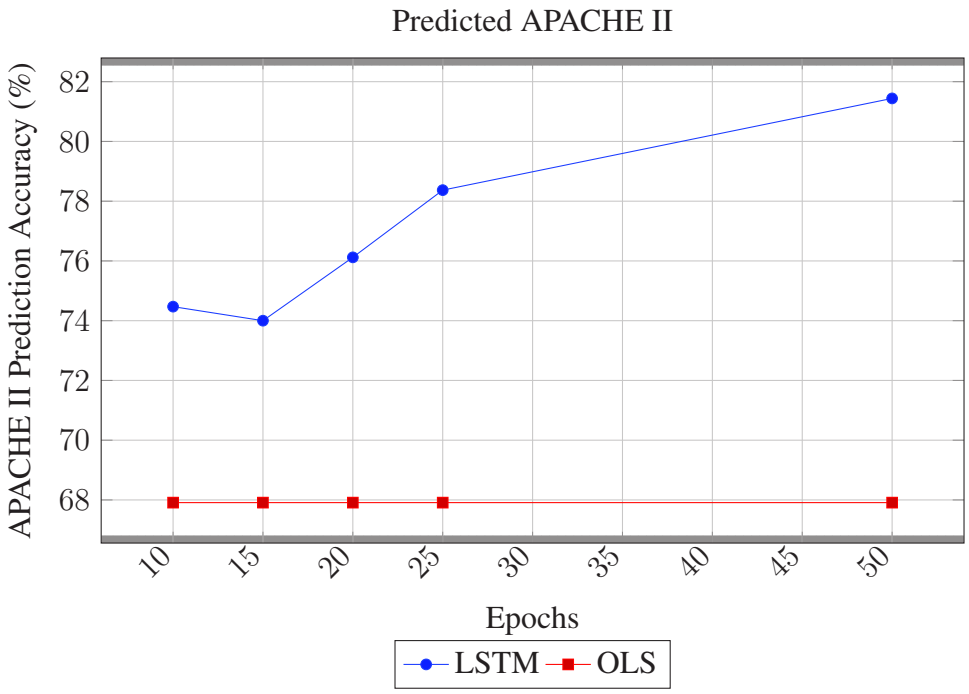

Figure 24 – LSTM X OLS

with results found by (CHURPEK; ADHIKARI; EDELSON, 2016), which indicates the importance of considering the variability of vital signs for the prognostic indexes calculation. Differently from their work, here, the vital sign's prediction, considering the variability of data, will be used in the index calculation.

During their stay in hospital wards, patients' data are dynamic, which means it can vary faster than expected. The constant monitoring of these data is essential and can provide better decision making. Moreover, predicting it can provide essential information to calculate prognostic indexes and anticipate the outcome as a deterioration status instead of a specific diagnostic, providing an indicator to continue patient treatment.

The early warning scoring systems take into account the real-time patient data in their calculation. Some works demonstrate that these early warning scores have poor performance than some of the new ML scoring systems in development (BRAND et al., 2018; TONEKABONI et al., 2018; JIN et al., 2018; PAN et al., 2019; LYU et al., 2018). In contrast, the early warning scoring system area is a mature subject and has substantial research about it for a long time. The work presented here developed an idea that embeds the two techniques, a prediction model based on Machine Learning (Deep Learning) and an early warning scoring system. It is essential to observe that the DeepSigns model can be attached to any scoring system considering the model is retrained for the appropriate variables.

6 CONCLUSION

This work presented the process of developing a computational model able to diagnose, in advance, the worsening of the health status of patients in such a way that it is possible to start the appropriate treatment as soon as possible and avoid deterioration of their health status. The model was developed using ML techniques, employing a RNN of type LSTM to predict vital signs of patients and posterior evaluation of the severity of the patient's health status through prognostic indexes commonly used in the Health area.

This work's main scientific contribution is the creation of a method for predicting vital signs based on historical data with low MSE and its following application in the calculation of prognostic indexes with effectiveness (50% - 60% of cases that would not be identified as severe). The differential presented by this proposal stems from the fact that few works predict vital signs. Most of the works focus on predicting specific health outcomes, such as specific diagnoses, considering the current vital signs. In this work, the proposal is to predict the evolution of vital signs in the future and use these predicted signs to calculate prognostic indexes.

The development of a method for predicting patient data, feeding back to the model with new personal data, meets precision medicine. Accordingly, the DeepSigns model highlights the opportunity that ML models have to assist in hospitals' decision-making, customizing care for patients' needs.

In the United States of America, the preventable inpatient deaths are estimated as between 44,000 and 98,000 per year (KOHN et al., 2000). A recent systematic review, which analyzed eight studies of hospitalized patients, concluded that 3.1% of 12,503 deaths included in the study were preventable according to physicians (RODWIN et al., 2020). The test results from DeepSigns identified between 50%, and 60% of tested cases would generate an alert that would not be produced otherwise. Considering also, 17% of these 50% and 60% tested cases died; it means that around 9% of deaths could be prevented with DeepSigns model application to the early intervention in ICU unities.

LSTM models are suitable for temporal series processing problems. Comparing ourselves with the OLS (Ordinary Least Squares) technique, DeepSigns results were superior. So, we decided to work with data loads and hyperparameters variations instead of experimenting with other network configurations, for instance, adding more layers, what would, probably, add more time to train the network. DeepSigns model was designed to be part of a CDSS (Clinical Decision Support Systems) and this demands the model to be retrained from time to time with data from new patients' arrivals. The less time to train, the better are the chances to improve the model performance in order to save people's lives. DeepSigns is a generic model and does not focus on specific diseases. So, the retraining mechanism plays an important role. Let us say we are experiencing COVID-19, just like today's pandemic situation. Furthermore, we get many patients arriving at hospital units with similar vital signs (or other information) patterns. DeepSigns model could be retrained once a day, for example, and some pattern could emerge

from data, guiding hospital teams to recognize a new disease.

The prediction of each variable included in this study was quite good. Chapter 5 presents an evaluation of each LSTM network, showing the MSE calculated for each of them. The results presented in Table 11 show the reference value for the attribute predicted by the network, the range of values considered as typical values by the APACHE II score, and also the MSE of the network. Thus, it is possible to observe how much each network allowed the approximation between the real and the predicted value. Considering these results, testing with other network compositions will be subject of future works.

As a result, the present work has generated some publications, and with that, we hope to contribute with ideas for improving health care, especially for hospitalized patients in critical situations. Some of them directly linked to the work presented (SILVA et al., 2020) and other related to previous works and collaborations (KAIESKI et al., 2020; SCHMIDT et al., 2018; COSTA et al., 2018b) carried out during the Ph.D. period.

As future work, we intend to experiment with the DeepSigns model with dynamic data obtained in real-time from patients in an ICU. With this data, we plan to discuss the results with the clinical staff and evaluate the possible impact on the patients' treatment. We also plan to compare the results produced by the DeepSigns model with a model that predicts the prognostic index itself using DL techniques and considering historical data.

REFERENCES

AKHNIF, E. et al. Scoping literature review on the learning organisation concept as applied to the health system. , [S.l.], p. 1–12, Feb. 2017.

ALAM, N. et al. The impact of the use of the early warning score (ews) on patient outcomes: a systematic review. **Resuscitation**, [S.l.], v. 85, n. 5, p. 587–594, May 2014.

ALLOGHANI, M. et al. Prospects of Machine and Deep Learning in Analysis of Vital Signs for the Improvement of Healthcare Services. In: **Nature-inspired computation in data mining and machine learning**. Cham: Springer International Publishing, 2019. p. 113–136.

ASIIMWE, S. B.; VITTINGHOFF, E.; WHOOLEY, M. Vital signs data and probability of hospitalization, transfer to another facility, or emergency department death among adults presenting for medical illnesses to the emergency department at a large urban hospital in the united states. **The Journal of Emergency Medicine**, [S.l.], Jan. 2020.

AZEREDO, T. et al. Efficacy of the manchester triage system: a systematic review. **Elsevier**, [S.l.], 1995.

AZEVEDO, D.; MACHADO, L.; SILVA, L. d. Metodos e procedimentos de pesquisa: do projeto ao relatório final. **UNISINOS**, [S.l.], 2011.

BELLOT, A.; SCHAAR, M. van der. A hierarchical bayesian model for personalized survival predictions. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. 23, n. 1, p. 72–80, Dec. 2018.

BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: a review and new perspectives. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v. 35, n. 8, p. 1798–1828, June 2013.

BOADA, C. M. Investigating prognostic factors in sepsis using computational intelligence methods. **Tese**, [S.l.], p. 1–46, Oct. 2016.

BONE, R. C.; SIBBALD, W. J.; SPRUNG, C. L. The ACCP-SCCM Consensus Conference on Sepsis and Organ Failure. **Chest**, [S.l.], v. 101, n. 6, p. 1481–1483, June 1992.

BOX, G. E. et al. **Time series analysis**: forecasting and control. [S.l.]: John Wiley & Sons, 2015.

BRAND, L. et al. Real time mortality risk prediction: a convolutional neural network approach. In: INTERNATIONAL CONFERENCE ON HEALTH INFORMATICS, 11., 2018. **Anais...** SCITEPRESS - Science and Technology Publications, 2018. p. 463–470.

BREKKE, I. J. et al. The value of vital sign trends in predicting and monitoring clinical deterioration: a systematic review. **PLOS ONE**, [S.l.], v. 14, p. e0210875–, Jan. 2019.

CELI, L. A. et al. “big data” in the intensive care unit. closing the data loop. **American journal of respiratory and critical care medicine**, [S.l.], v. 187, n. 11, p. 1157, 2013.

CHE, C. et al. **An rnn architecture with dynamic temporal matching for personalized predictions of parkinson’s disease**. Philadelphia, PA, 2017.

CHEN, J. In search of the ‘best’ rapid response early warning system – the journey has just begun. **Resuscitation**, [S.l.], v. 123, p. A1–A2, Feb. 2018.

CHURPEK, M. M.; ADHIKARI, R.; EDELSON, D. P. The value of vital sign trends for detecting clinical deterioration on the wards. **Resuscitation**, [S.l.], v. 102, p. 1–5, May 2016.

CHURPEK, M. M. et al. Multicenter development and validation of a risk stratification tool for ward patients. **American Journal of Respiratory and Critical Care Medicine**, [S.l.], v. 190, n. 6, p. 649–655, Sept. 2014.

CLIFTON, L. A. et al. Identification of patient deterioration in vital-sign data using one-class support vector machines. **FedCSIS**, [S.l.], 2011.

COOKE, M. W.; JINKS, S. Does the manchester triage system detect the critically ill? **Emergency Medicine Journal**, [S.l.], v. 16, n. 3, p. 179–181, May 1999.

COSTA, C. A. da et al. Internet of health things: toward intelligent vital signs monitoring in hospital wards. **Artificial Intelligence in Medicine**, [S.l.], v. 89, p. 61–69, 2018.

COSTA, C. A. da et al. Internet of health things: toward intelligent vital signs monitoring in hospital wards. **Artificial intelligence in medicine**, [S.l.], v. 89, p. 61–69, 2018.

CROWN, W. H. Potential application of machine learning in health outcomes research and some statistical cautions. **Value in Health**, [S.l.], v. 18, n. 2, p. 137–140, Mar. 2015.

CURTIS, J. R. et al. Using electronic health records for quality measurement and accountability in care of the seriously ill: opportunities and challenges. **Journal of Palliative Medicine**, [S.l.], v. 21, n. S2, p. S–52–S–60, 2018.

DOWNEY, C. L. et al. Strengths and limitations of early warning scores: a systematic review and narrative synthesis. **International Journal of Nursing Studies**, [S.l.], v. 76, p. 106–119, 2017.

DUGAR, S.; CHOUDHARY, C.; DUGGAL, A. Sepsis and septic shock: guideline-based management. **Cleveland Clinic Journal of Medicine**, [S.l.], v. 87, n. 1, p. 53–64, Jan. 2020.

FANG, R. et al. Computational health informatics in the big data age. **ACM Computing Surveys**, [S.l.], v. 49, n. 1, p. 1–36, July 2016.

FLEISCHMANN, C. et al. Assessment of global incidence and mortality of hospital-treated sepsis. current estimates and limitations. **American journal of respiratory and critical care medicine**, [S.l.], v. 193, n. 3, p. 259–272, 2016.

FORKAN, A. R. M.; KHALIL, I. A clinical decision-making mechanism for context-aware and patient-specific remote monitoring systems using the correlations of multiple vital signs. **Computer Methods and Programs in Biomedicine**, [S.l.], v. 139, p. 1–16, Feb. 2017.

FORKAN, A. R. M.; KHALIL, I.; ATIQUZZAMAN, M. Visibid: a learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data. **Computer Networks**, [S.l.], v. 113, n. C, p. 244–257, Feb. 2017.

GERHARDT, T. E.; SILVEIRA, D. T. **Méodos de pesquisa**. [S.l.]: Plageder, 2009.

GERRY, S. et al. Early warning scores for detecting deterioration in adult hospital patients: a systematic review protocol. **BMJ Open**, [S.l.], v. 7, n. 12, p. e019268–6, Dec. 2017.

GHOSH, E. et al. Early deterioration indicator: data-driven approach to detecting deterioration in general ward. **Resuscitation**, [S.l.], v. 122, p. 99–105, Jan. 2018.

GIANNINI, H. M. et al. A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock. **Critical Care Medicine**, [S.l.], v. 47, n. 11, p. 1485–1492, Nov. 2019.

GODINJAK, A. G. Predictive value of SAPS II and APACHE II scoring systems for patient outcome in medical intensive care unit. **Acta Medica Academica**, [S.l.], v. 45, n. 2, p. 89–95, Dec. 2016.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT press, 2016.

GRANHOLM, A. et al. Predictive performance of the simplified acute physiology score (saps) ii and the initial sequential organ failure assessment (sofa) score in acutely ill intensive care patients: post-hoc analyses of the sup-icu inception cohort study. **PLOS ONE**, [S.l.], v. 11, p. e0168948–, Dec. 2016.

GREEN, S. L. et al. The Combined SIRS qSOFA (qSIRS) Score is More Accurate Than qSOFA Alone in Predicting Mortality in Patients with Surgical Sepsis in an LMIC Emergency Department. **World Journal of Surgery**, [S.l.], v. 44, n. 1, p. 21–29, Oct. 2019.

GULTEPE, E. et al. From vital signs to clinical outcomes for patients with sepsis - a machine learning basis for a clinical decision support system. **JAMIA**, [S.l.], 2014.

HAYKIN, S. S. et al. **Neural networks and learning machines**. [S.l.]: Pearson Upper Saddle River, 2009. v. 3.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, [S.l.], v. 9, n. 8, p. 1735–1780, 1997.

HUANG, T. et al. Promises and challenges of big data computing in health sciences. **Big Data Research**, [S.l.], v. 2, n. 1, p. 2–11, 2015.

HUTCHESON, G. Ordinary least-squares regression. In: **The multivariate social scientist**. [S.l.]: SAGE Publications, Ltd., 1999. p. 56–113.

JANG, D.-H. et al. Developing neural network models for early detection of cardiac arrest in emergency department. **The American Journal of Emergency Medicine**, [S.l.], v. 38, n. 1, p. 43–49, Jan. 2020.

JIN, M. et al. Improving hospital mortality prediction with medical named entities and multimodal learning. **CoRR**, [S.l.], v. cs.CL, 2018.

JOHNSON, A. E. W. et al. Mimic-iii, a freely accessible critical care database. **Scientific Data**, [S.l.], v. 3, p. 160035–9, May 2016.

JOHNSON, A. et al. **Mimic-iv (version 0.4)**. [S.l.]: PhysioNet, 2020.

JONES, S. S. et al. Health information technology: an updated systematic review with a focus on meaningful use. **Annals of Internal Medicine**, [S.l.], v. 160, n. 1, p. 48–54, Jan. 2014.

JUNG, K. et al. Predicting need for advanced illness or palliative care in a primary care population using electronic health record data. **Journal of Biomedical Informatics**, [S.l.], v. 92, p. 103115, Feb. 2019.

KAIESKI, N. et al. Um modelo para predição individualizada da deterioração da saúde de pacientes no âmbito da internet das coisas. In: PRINCIPAIS DO XX SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE, 2020. **Anais...** [S.l.: s.n.], 2020. p. 167–178.

KAKKANATT, C. et al. Curating and integrating user-generated health data from multiple sources to support healthcare analytics. **IBM Journal of Research and Development**, [S.l.], 2018.

KAM, H. J.; KIM, H. Y. Learning representations for the early detection of sepsis with deep neural networks. **Computers in Biology and Medicine**, [S.l.], v. 89, p. 248–255, Oct. 2017.

KHARE, A. et al. Machine learning theory and applications for healthcare. **Journal of Healthcare Engineering**, [S.l.], v. 2017, p. 1–2, 2017.

KNAUS, W. A. et al. The apache iii prognostic system. **Chest**, [S.l.], v. 100, n. 6, p. 1619–1636, 1991.

KOHN, L. T. et al. To err is human. **building a safer health system**, [S.l.], v. 600, p. 2000, 2000.

KUMAR, A. et al. Duration of shock prior to antimicrobial administration is the critical determinant of survival in human septic shock: 41. **Critical Care Medicine**, [S.l.], v. 32, n. 12, p. A11, 2004.

LE GALL, J.-R.; LEMESHOW, S.; SAULNIER, F. A new simplified acute physiology score (saps ii) based on a european/north american multicenter study. **Jama**, [S.l.], v. 270, n. 24, p. 2957–2963, 1993.

LE LAGADEC, M. D.; DWYER, T. Scoping review: the use of early warning systems for the identification of in-hospital patients at risk of deterioration. **Australian Critical Care**, [S.l.], v. 30, n. 4, p. 211–218, 2017.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, [S.l.], v. 521, n. 7553, p. 436, 2015.

LEE, C. et al. **Big healthcare data analytics**: challenges and applications. Cham, 2017.

LIN, Y.-W. et al. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. **PLOS ONE**, [S.l.], v. 14, p. e0218942–, July 2019.

LINNEN, D. Statistical modeling and aggregate-weighted scoring systems in prediction of mortality and icu transfer: a systematic review. **Journal of Hospital Medicine**, [S.l.], v. 14, n. 3, p. 161–9, 2019.

LIPTON, Z. C.; KALE, D. C.; WETZEL, R. Modeling missing data in clinical time series with rnns. **arXiv.org**, [S.l.], June 2016.

LIU, N. T. et al. Development and validation of a machine learning algorithm and hybrid system to predict the need for life-saving interventions in trauma patients. **Medical & Biological Engineering & Computing**, [S.l.], v. 52, n. 2, p. 193–203, Nov. 2013.

LIU, N. T. et al. Utility of Vital Signs, Heart Rate Variability and Complexity, and Machine Learning for Identifying the Need for Lifesaving Interventions in Trauma Patients. **Shock**, [S.l.], v. 42, n. 2, p. 108–114, Aug. 2014.

LOWES, L. P. et al. 'learn from every patient': implementation and early results of a learning health system. **Developmental Medicine & Child Neurology**, [S.l.], v. 59, n. 2, p. 183–191, Aug. 2016.

LYU, X. et al. Improving clinical predictions through unsupervised time series representation learning. **arXiv.org**, [S.l.], Dec. 2018.

MANASHTY, A.; THOMSON, J. L. A new temporal abstraction for health diagnosis prediction using deep recurrent networks. In: INTERNATIONAL DATABASE ENGINEERING & APPLICATIONS SYMPOSIUM, 21., 2017, New York, New York, USA. **Anais...** ACM Press, 2017. p. 14–19.

MD, S. Y. H. et al. Low accuracy of positive qsofa criteria for predicting 28-day mortality in critically ill septic patients during the early period after emergency department presentation. **Annals of Emergency Medicine**, [S.l.], v. 71, n. 1, p. 1–9.e2, Jan. 2018.

MITCHELL, R.; MICHALSKI, J.; CARBONELL, T. **An artificial intelligence approach**. [S.l.]: Springer, 2013.

MITCHELL, T. Machine learning, mcgraw-hill higher education. **New York**, [S.l.], 1997.

MORALES, C.; RIBAS, V. J.; VELLIDO, A. Applying conditional independence maps to improve sepsis prognosis. **ICDM Workshops**, [S.l.], 2016.

MURPHY, A. et al. Economics of early warning scores for identifying clinical deterioration—a systematic review. **Irish Journal of Medical Science (1971 -)**, [S.l.], v. 187, n. 1, p. 193–205, 2017.

O'BRIEN, C. et al. Development, Implementation, and Evaluation of an In-Hospital Optimized Early Warning Score for Patient Deterioration. **MDM Policy & Practice**, [S.l.], v. 5, n. 1, p. 238146831989966, Jan. 2020.

OBUCHOWSKI, N. A.; BULLEN, J. A. Receiver operating characteristic (roc) curves: review of methods with applications in diagnostic medicine. **Physics in Medicine & Biology**, [S.l.], v. 63, n. 7, p. 07TR01, 2018.

PAN, Z. et al. A self-correcting deep learning approach to predict acute conditions in critical care. **CoRR**, [S.l.], 2019.

Pathak, S.; Cai, X.; Rajasekaran, S. Ensemble deep timenet: an ensemble learning approach with deep neural networks for time series. In: IEEE 8TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL ADVANCES IN BIO AND MEDICAL SCIENCES (ICCABS), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 1–1.

PAYNE, T. H. et al. Report of the amia ehr-2020 task force on the status and future direction of ehrs. **Journal of the American Medical Informatics Association**, [S.l.], v. 22, n. 5, p. 1102–1110, Sept. 2015.

PEREIRA, R. B. et al. Correlation analysis of performance measures for multi-label classification. **Information Processing and Management**, [S.l.], v. 54, n. 3, p. 359–369, May 2018.

PHD, K. M. H. M. et al. Using patient admission characteristics alone to predict mortality of critically ill patients: a comparison of 3 prognostic scores. **Journal of Critical Care**, [S.l.], v. 31, n. 1, p. 21–25, Feb. 2016.

PHYSICIANS, R. C. of. National early warning score (news) 2. **Standardising the assessment of acute-illness severity in the NHS**, [S.l.], 2017.

PLATT, J. E.; RAJ, M.; WIENROTH, M. An analysis of the learning health system in its first decade in practice: scoping review. **J Med Internet Res**, [S.l.], v. 22, n. 3, p. e17026, Mar 2020.

PRYTHERCH, D. R. et al. Views—towards a national early warning score for detecting adult inpatient deterioration. **Resuscitation**, [S.l.], v. 81, n. 8, p. 932–937, 2010.

PURUSHOTHAM, S. et al. Benchmarking deep learning models on large healthcare datasets. **Journal of Biomedical Informatics**, [S.l.], v. 83, p. 112–134, July 2018.

RAHMAN, F.; SLEPIAN, M. J.; MITRA, A. A novel big-data processing framwork for healthcare applications - big-data-healthcare-in-a-box. **BigData**, [S.l.], p. 3548–3555, 2016.

RAJKOMAR, A. et al. Scalable and accurate deep learning for electronic health records. **arXiv.org**, [S.l.], Jan. 2018.

RAVI, D. et al. Deep learning for health informatics. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. 21, n. 1, p. 4–21, Jan. 2017.

REZENDE, S. O. **Sistemas inteligentes**: fundamentos e aplicações. [S.l.]: Editora Manole Ltda, 2003.

RHODES, A. et al. Surviving Sepsis Campaign. **Critical Care Medicine**, [S.l.], v. 45, n. 3, p. 486–552, Mar. 2017.

RODWIN, B. A. et al. Rate of preventable mortality in hospitalized patients: a systematic review and meta-analysis. **Journal of General Internal Medicine**, [S.l.], v. 324, n. 6, p. 370–8, Jan. 2020.

ROEHRS, A.; COSTA, C. A. da; ROSA RIGHI, R. da. Omniphr: a distributed architecture model to integrate personal health records. **Journal of Biomedical Informatics**, [S.l.], v. 71, p. 70–81, July 2017.

ROEHRS, A. et al. Personal health records: a systematic literature review. **Journal of Medical Internet Research**, [S.l.], v. 19, n. 1, p. e13–21, 2017.

SCHLOEFFEL, P. Electronic health record definition, scope and context. iso. **TC**, [S.l.], v. 215, 2002.

SCHMIDT, D. et al. Um modelo de predição de mortalidade em unidades de terapia intensiva baseado em deep learning. In: PRINCIPAIS DO XVIII SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE, 2018. **Anais...** [S.l.: s.n.], 2018.

SELVIN, S. et al. Stock price prediction using LSTM, RNN and CNN-sliding window model. **ICACCI**, [S.l.], p. 1643–1647, 2017.

SEYMOUR, C. W. et al. Assessment of clinical criteria for sepsis. **JAMA**, [S.l.], v. 315, n. 8, p. 762, 2016.

SHAFQAT, S. et al. Big data analytics enhanced healthcare systems: a review. **The Journal of Supercomputing**, [S.l.], v. 2, n. 1, p. 1–46, 2018.

SHICKEL, B. et al. Deep ehr: a survey of recent advances on deep learning techniques for electronic health record (ehr) analysis. **arXiv.org**, [S.l.], June 2017.

SILCOCK, D. J. et al. Validation of the national early warning score in the prehospital setting. **Resuscitation**, [S.l.], v. 89, p. 31–35, Apr. 2015.

SILVA, D. B. da et al. Deepsigns: a predictive model based on deep learning for the early detection of patient health deterioration. **Expert Systems with Applications**, [S.l.], p. 113905, 2020.

SINGER, M. et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). **JAMA**, [S.l.], v. 315, n. 8, p. 801–10, Feb. 2016.

STORM-VERSLOOT, M. N. et al. Comparison of an informally structured triage system, the emergency severity index, and the manchester triage system to distinguish patient priority in the emergency department. **Academic Emergency Medicine**, [S.l.], v. 18, n. 8, p. 822–829, Aug. 2011.

SUBBE, C. P. Validation of a modified early warning score in medical admissions. **QJM**, [S.l.], v. 94, n. 10, p. 521–526, 2001.

TONEKABONI, S. et al. Prediction of cardiac arrest from physiological signals in the pediatric icu. **MLHC**, [S.l.], 2018.

VINCENT, J. L. et al. The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. **Intensive Care Medicine**, [S.l.], v. 22, n. 7, p. 707–710, 1996.

VINCENT, J.-L. et al. **The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure**. [S.l.]: Springer, 1996.

WAGNER, D. P.; DRAPER, E. A. Acute physiology and chronic health evaluation (apache ii) and medicare reimbursement. **Health care financing review**, [S.l.], Nov. 1984.

WANG, L. et al. The application of unsupervised deep learning in predictive models using electronic health records. **BMC Medical Research Methodology**, [S.l.], v. 20, n. 1, p. 38–9, Feb. 2020.

WANG, Y.; KUNG, L.; BYRD, T. A. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. **Technological Forecasting & Social Change**, [S.l.], v. 126, n. C, p. 3–13, Jan. 2018.

WILLIAMS, R. et al. Clinical code set engineering for reusing ehr data for research: a review. **Journal of Biomedical Informatics**, [S.l.], v. 70, p. 1–13, June 2017.

YADAV, P. et al. Mining electronic health records (ehrs). **ACM Computing Surveys**, [S.l.], v. 50, n. 6, p. 1–40, Jan. 2018.

YAN, B. W.; HSIA, R. Y. Unexpected deaths among patients with normal initial vital signs. **Academic Emergency Medicine**, [S.l.], v. 26, n. 11, p. 1285–1288, Aug. 2019.