

POLITECNICO DI TORINO

Facoltà di ingegneria

Master's degree course in Biomedical Engineering

Master's Degree Thesis

**Multimodal physiological  
time series analysis  
for outcome prediction  
in the intensive care unit**



POLITECNICO  
DI TORINO



Novo Nordisk Foundation  
Center for Protein Research

**Supervisors**

Prof. Filippo Molinari

**Candidates**

Davide Placido

**Internship Tutor**

Assoc. Prof. Sadasivan Puthusserypady  
M.D. Hans-Christian Thorsen-Meyer

ACCADEMIC YEAR 2018-2019

This work is subject to the Creative Commons Licence

*Alla mia famiglia*

# Summary

The intensive care unit (ICU) is a hospital department where critical patients are monitored and healed: the sensors used to check their vital conditions produce a huge amount of data that can be exploited to predict outcomes about their future conditions. In this work mortality prediction is performed using data recorded from more than 9000 patients in different hospitals of the Capital Region of Denmark from 2009 to 2016. Given the high number of physiological variables acquired in the ICU, a study of which variables subset provides more information is carried on, taking into account the number of patients who have that subset of physiological variables monitored. Both signal processing and statistic methods are used to extract information from the first 24 hours after the admission in the ICU; then a Long short-term memory (LSTM) model is used for classification and regression tasks. Finally, SHapley Additive exPlanations (SHAP) approach is used to increase the interpretability of the model.

Regardless of which task the model was designed for, age of admission is the most important feature; vital signs generally account the most with respect to arterial blood gas (ABG) measurements and no significant improvements are present when features from Empirical mode decomposition (EMD) are added to the features space. Moreover the model has comparable performances both when the training includes all the patients in the dataset or only the patients with measurements for every variable in the feature space.

# Sommario

La terapia intensiva è il reparto ospedaliero dove vengono fornite cure e assistenza a pazienti in condizioni critiche: i sensori utilizzati per controllare le condizioni vitali originano una quantità considerevole di dati. Tali dati possono essere sfruttati per poter predire le condizioni future dei pazienti. In questo progetto di tesi, dati raccolti da più di 9000 pazienti ricoverati nell'area di Copenaghen dal 2009 al 2016 sono stati utilizzati per la predizione della mortalità. Dato l'elevato numero di variabili fisiologiche monitorate nel reparto di terapia intensiva, è stato effettuato uno studio su quali variabili utilizzare per la predizione, tenendo in considerazione il numero di pazienti a cui effettivamente ciascuna variabile è stata monitorata. Metodi statistici e tecniche di analisi dei segnali sono state utilizzate per estrarre informazioni dalle prime 24 ore di acquisizione dopo l'ammissione nel reparto di terapia intensiva; una rete *long short-term memory* (LSTM) è stata usata per compiere classificazione e regressione. Infine, un metodo denominato "SHapley Additive exPlanations" è stato utilizzato per migliorare l'interpretazione del modello.

A prescindere da quale task il modello debba svolgere, l'età a cui il paziente viene ammesso nel reparto risulta essere la feature più determinante; le features estratte dai segnali vitali hanno un contributo superiore rispetto alle misurazioni ottenute dai prelievi di sangue arterioso, e l'aggiunta delle features estratte tramite *empirical mode decomposition* (EMD) non portano sostanziali miglioramenti alle prestazioni del modello. Infine le prestazioni sembrano essere analoghe indipendentemente che il modello utilizzi tutti i pazienti all'interno del dataset o solamente quelli da cui sono state acquisite solo le variabili contenute nello spazio delle feature.

# Acknowledgements

I would like to thank all the people who have given a contribution for the realization of this work. Sincere gratitude to Associate Professor Sadasivan Puthusserypady from Technical University of Denmark (DTU), for his constant help through all the period spent in Denmark. I would like to thank Professor Filippo Molinari from Politecnico di Torino for his supervision during my period abroad.

I am grateful to Research director and Professor Søren Brunak and the Brunak group from Novo Nordisk Foundation Center for Protein Research (CPR) for the opportunity they gave me to work with the ICU data. In particular sincere thanks to M.D. Hans-Christian Thorsen-Meyer and Post-doc Annelaura Bach Nielsen for their precious advise and the constructive feedback provided me.

Finally I am thankful to all the students and researcher from both DTU and CPR I have met during this period for sharing ideas and interests with me.

# List of Acronyms

<b>ICU</b> intensive care unit .....	4
<b>IMF</b> intrinsic mode function .....	12
<b>ECG</b> electrocardiogram .....	16
<b>BP</b> blood pressure .....	16
<b>HR</b> heart rate .....	26
<b>RR</b> respiratory rate .....	22
<b>SVM</b> support vector machine .....	17
<b>RNN</b> Recurrent neural network .....	13
<b>LSTM</b> Long short-term memory .....	4
<b>CNN</b> convolutional neural network .....	16
<b>FCN</b> fully convolutional network .....	19

<b>RBF</b> radial basis function .....	16
<b>EMD</b> Empirical mode decomposition .....	4
<b>PCA</b> Principal component analysis .....	18
<b>MLP</b> Multilayer perceptron .....	18
<b>GBT</b> gradient boosting trees .....	21
<b>MAP</b> Mean arterial pressure .....	13
<b>SP</b> Systolic pressure .....	24
<b>DP</b> Diastolic pressure .....	24
<b>SAPS</b> Simplified Acute Physiology Score .....	15
<b>GCS</b> Glasgow Coma Scale.....	27
<b>APACHE</b> Acute physiology and chronic health evaluation.....	15
<b>ABG</b> arterial blood gas.....	4
<b>MSE</b> Mean squared error.....	13
<b>SHAP</b> SHapley Additive exPlanations.....	4
<b>MCC</b> Matthews correlation coefficient.....	13

<b>AUC</b> Area under the curve.....	13
<b>ROC</b> Receiver operating characteristic.....	44
<b>SBE</b> Standard base excess.....	28
<b>KNN</b> k-nearest neighbors .....	21

# Contents

<b>List of Acronyms</b>	7
<b>List of Tables</b>	12
<b>List of Figures</b>	13
<b>1 Introduction</b>	15
1.1 Objectives . . . . .	15
1.2 Motivation . . . . .	15
<b>2 State of the art</b>	16
2.1 Feature engineering . . . . .	17
2.2 Feature learning . . . . .	18
<b>3 Physiology</b>	22
3.1 Vital signs . . . . .	22
3.1.1 Respiratory system . . . . .	22
3.1.2 Circulatory system . . . . .	23
3.1.3 Temperature . . . . .	26
3.2 Urinary output . . . . .	27
3.3 Glasgow coma scale . . . . .	27
3.4 Arterial blood gas measurements . . . . .	27
<b>4 Materials and Methods</b>	29
4.1 Model selected . . . . .	29
<b>5 Data preparation</b>	32
5.1 Preprocessing . . . . .	32
5.2 Feature extraction . . . . .	34
5.2.1 Statistic features . . . . .	35

5.2.2	Entropy measures . . . . .	36
5.2.3	Empirical mode decomposition . . . . .	37
5.3	Outcome prediction . . . . .	38
5.3.1	Classification . . . . .	38
5.3.2	Regression for 90 days of mortality . . . . .	38
5.3.3	Regression including censored patients . . . . .	39
5.4	Cross-validation . . . . .	40
<b>6</b>	<b>Results</b>	<b>42</b>
6.1	SHAP . . . . .	42
6.2	Classification . . . . .	44
6.3	Regression for 90 days of mortality . . . . .	49
6.4	Regression including censored patients . . . . .	54
6.5	Cox proportional-hazards regression model . . . . .	59
<b>7</b>	<b>Conclusion</b>	<b>61</b>

# List of Tables

3.1	Different ABG used as input.[30] [31] [29] [32] [33] . . . . .	28
5.1	Different input used for training. . . . .	34
6.1	Best performances for classification. Common Patients in Vital (Input 1), Common Patients in Vitals Static (Input 2), Common Patients in Vitals, ABG and Static (Input 3), Common Patients in Vitals, ABG and Static (intrinsic mode function (IMF)) (Input 4), All Patients in Vitals, ABG and Static (Input 5), All Patients in Vitals, ABG and Static (IMF) (Input 6). . . . .	44
6.2	Best performances for Regression on 90 days. Common Patients in Vital (Input 1), Common Patients in Vitals Static (Input 2), Common Patients in Vitals, ABG and Static (Input 3), Common Patients in Vitals, ABG and Static (IMF) (Input 4), All Patients in Vitals, ABG and Static (Input 5), All Patients in Vitals, ABG and Static (IMF) (Input 6). . . . .	49
6.3	Best performances for Regression including censored patients. Common Patients in Vital (Input 1), Common Patients in Vitals Static (Input 2), Common Patients in Vitals, ABG and Static (Input 3), Common Patients in Vitals, ABG and Static (IMF) (Input 4), All Patients in Vitals, ABG and Static (Input 5), All Patients in Vitals, ABG and Static (IMF) (Input 6). . . . .	54

# List of Figures

2.1	Structure of model. Reprinted from [8]. . . . .	19
2.2	Structure of model. Reprinted from [12] . . . . .	19
2.3	Structure of model. Reprinted from [13]. . . . .	20
2.4	Structure of model. Reprinted from [14]. . . . .	20
2.5	Structure of model. Reprinted from [15]. . . . .	21
3.1	Structure of model. Reprinted from [17]. . . . .	24
4.1	Recurrent neural network (RNN) unrolled over the time. Reprinted from [35] . . . . .	30
4.2	LSTM cell. Reprinted from [35]. . . . .	30
5.1	Shape the model receives as input. . . . .	33
5.2	Output of the model for regression of 90 days of mortality. . .	39
5.3	Kaplan-Meier estimation. . . . .	39
5.4	Output of the model for regression including censored patients.	40
6.1	Matthews correlation coefficient (MCC) and Area under the curve (AUC) for different input. . . . .	45
6.2	Features contribute for the first hour after the admission. . . .	46
6.3	Mean arterial pressure (MAP) contribute for the first hour after the admission. . . . .	47
6.4	Classification task: variables importance over 24 hours. . . . .	48
6.5	Classification task: features importance over 24 hours. . . . .	48
6.6	Standart deviation contribute for the first hour after the admission. . . . .	49
6.7	Mean squared error (MSE) and $R^2$ for different input. . . . .	50
6.8	Features contribute for the first hour after the admission. . . .	51
6.9	Regression task: variables importance over 24 hours. . . . .	52
6.10	Regression task: features importance over 24 hours. . . . .	53
6.11	Different metrics for the input tested with regression on 90 days.	53
6.12	MSE and $R^2$ for different input. . . . .	55
6.13	Features contribute for the first hour after the admission. . . .	56

6.14	Variables importance over 24 hours. . . . .	57
6.15	Features importance over 24 hours. . . . .	57
6.16	Different metrics for the input tested with regression including censored patients. . . . .	58
6.17	Cox model on two different regression tasks. . . . .	59

# Chapter 1

## Introduction

### 1.1 Objectives

The aim of the thesis is to exploit all the data usually recorded from the patients admitted in the ICU in order to develop a model able to predict any possibly outcome about the condition of these patients.

### 1.2 Motivation

The ICU is probably one of the most critical department in the hospital. Patients treated in this unit have severe and life threatening injuries that require a continuous monitoring of vital parameters. The outcome prediction of ICU patients can be a powerful tool used by doctors and technicians to better understand the effect of a specific therapy and to prevent the death of a patient. At the moment many severity scores, such as Simplified Acute Physiology Score (SAPS) II, SAPS III, Acute physiology and chronic health evaluation (APACHE) II, etc. are used to understand how severe the condition of a patient are after his admission in the ICU. These scores have some limitations both on the number of hours needed for their computation and the number of variables taken into account. The idea is building a model that is strong enough to not be affected by missing acquisitions of variables not recorded by medical staff, thus improving the performances of the prediction.

# Chapter 2

## State of the art

The aim of this chapter is to illustrate the main methodologies used to extract information from time series in order to perform classification: particular attention is placed on physiological time series since the domain is similar to the one this project is based on, therefore a better understanding of the choices taken in the following is possible. After a short introduction about general challenges in this research field two main approaches will be compared: methods based on feature extraction and end-to-end methods [1]. The concepts they are built on, together with their pros and cons, will be described afterwards.

Time series analysis is extensively used in order to predict possible outcomes in many research fields such as economy, meteorology, geology and others. Multi-modal analysis requires taking into account many issues related to the magnitude of data coming from different sensor devices. A robust model is necessary due to the presence of outliers, noise, artifacts and missing values, which are usually common when the dataset is composed by heterogeneous data. Another aspect to deal with is the extraction of information both from single modalities and from the relations between the different channels. Literature about multi-modal physiological time series contains analysis of many signals as electrocardiogram (ECG), heart rate, blood pressure (BP), respiratory rate as well as many supervised learning models as convolutional neural network (CNN), radial basis function (RBF) network and LSTM network. The choice of the model is probably the main aspect because it affects not only the results but also the set-up of the dataset.

## 2.1 Feature engineering

The approach presented in this section is probably the most traditional and is exploited through some examples.

The extraction and the selection of the features that the model is going to use for the classification are necessary in order to obtain information that are not expressed in raw data, especially when the model is not able to automatically extract significant features. Feature engineering approach requires a very good knowledge of the time series domain to select the features that better represent the events. A higher number of features does not imply a model with better performances. Sometimes a lot of features are redundant and the useful information are enclosed in a smaller subset. For this reason a lot of techniques can be exploited to reduce the dimensionality of the feature space. According to the literature, many fields of research have used feature extraction for time series analysis.

In [2], the feature extraction approach for time series classification is based on wavelet transform, fractal dimension and statistic methods. Wavelet transform is used to obtain the coefficients related to the different decompositions, and high frequency wavelet coefficients are selected as input features. Fractals are used to extract information about the complexity of the signal and parameters extracted from fractal spectrum are chosen as input features. Statistical features are mean value and standard deviation. Japanese vowels data is used to extract these features and the classification is achieved through a RBF network. This networks use radial basis functions as activation function. The same classifier is tested with different input features and the best accuracy is obtained from features coming from both wavelet, fractal and statistics.

In [3], feature extraction is applied for classification of physiological time series. Starting from the same principle of the EMD, the signal is decomposed in mode functions using an iterative convolutional filter. Statistical parameters are extracted from the different mode and used for classification. A support vector machine (SVM) model with recursive features elimination is used: all the features are used to train SVM and iteratively the features with lower weights are deleted.

Focusing on the challenge of this work, a lot of literature is present about mortality prediction in ICU; the Physionet challenge in 2012 employed MIMIC II Clinical Database [4] which contains data recorded in the ICU, therefore many authors contributed in this topic. In [5], this dataset is used to predict mortality using SVM; particular consideration is given to the imbalance of

the dataset. A cost sensitive Principal component analysis (PCA) is used to deal with the imbalance issue and to reduce the dimensionality of the features space.

In [6], another dataset containing ICU time series and clinical medical data is used for multi-label classification. Two models are compared, Multilayer perceptron (MLP) and LSTM network. The first model is trained with hand-engineered features while the second with raw data. The LSTM network, that used dropout and auxiliary outputs to avoid overfitting, definitely obtained better performances.

In [7], MIMIC II Database is used to predict re-admission in the intensive care unit. A set of features is extracted from the time series and then, an ensemble model is used to predict the outcome. The ensemble model consists on two different models that differently maximize the performance on two metrics: sensitivity and specificity. In this way there is no need to find one single model with a good trade-off between the two metrics since each input is tested on both. Their output are combined together depending on the aggregation criteria selected.

Physiological time series classification is applied not only for critical situations like patients in ICU but also for daily needs like emotion recognition. In [8], a LSTM network is used to recognize emotions from auditory, visual and physiological modalities. A high number of features are extracted: in order to reduce the number of features, PCA is applied and the results are used to train the model. The architecture is composed by a hidden layer, a temporal pooling layer, a multi-modal fusion layer and a LSTM layer (Figure 2.1). Temporal pooling layer works in a similar way to bi-dimensional pooling layer in CNN but in this case the input is added in a temporal sequence: in this way the model is trained with a higher diversity of features with respect to a raw features training.

## 2.2 Feature learning

A methodology that is getting common is the end-to-end approach: this is partly due to the complexity that the current model can handle. The models based on this approach do not need the design of a feature space since the raw input is directly fed into the model: it is the model itself able to find significant information from the input. In this way good results can be obtained in spite of neglecting a lot of pre-processing steps [9]. Indeed, a great advantage of deep learning is the ability of the model to find the best

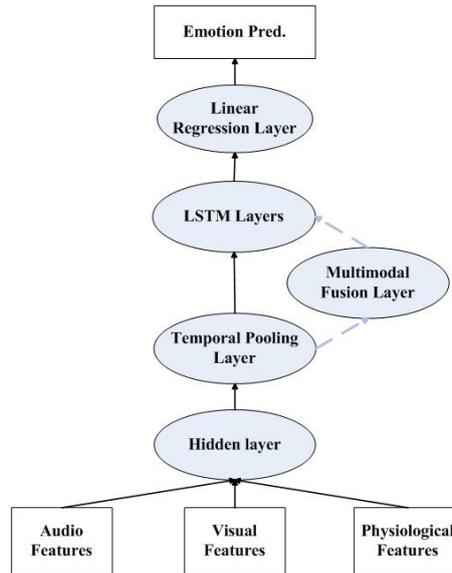


Figure 2.1: Structure of model. Reprinted from [8].

features for the specific classification task without the interference of human knowledge. Several types of architectures are used to perform the time series classification task [10].

In [11], an analysis between MLP, fully convolutional network (FCN) and ResNet is carried out trying to achieve a standard for time series classification. All the models are trained with raw data on multiple datasets and FCN resulted to have the best performances. Two choices enhanced the performances of FCN: global pooling, which is used instead of fully connected layer before softmax, and batch normalization.

In [12], multimodal time series analysis for sleep stage classification is performed. Also in this case an end-to-end approach is followed. All the modalities are processed in different pipelines, each of them consisting on a spatial filter, convolutional non-linear operations and max pooling. The outputs from the different modalities are concatenated and fed into a softmax classifier (Figure 2.2).

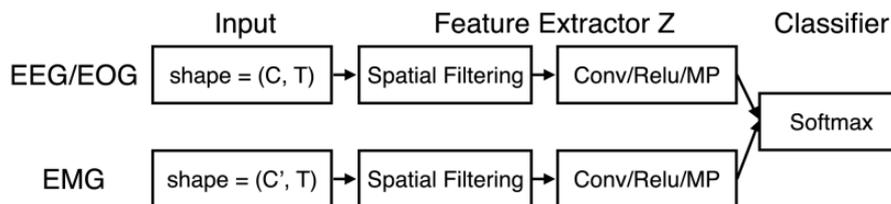


Figure 2.2: Structure of model. Reprinted from [12]

Some attempts are made in order to improve the performance of both LSTM and FCN trying to combine them together. In [13], FCN module is augmented with LSTM cell: the idea is to combine the ability of FCN to extract features and of LSTM to consider temporal dependencies. The input is fed in parallel to LSTM and FCN cell: the two output are concatenated and used by softmax layer for prediction (Figure 2.3).

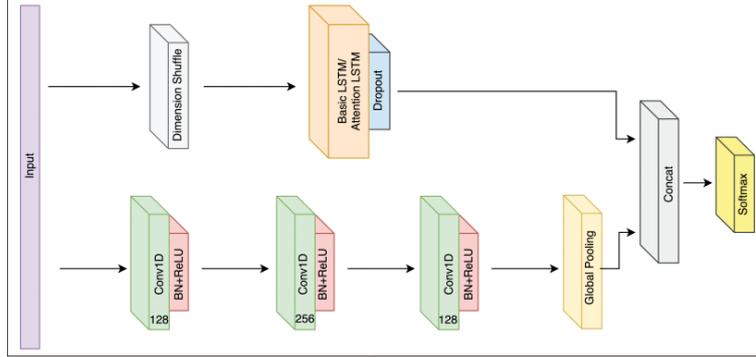


Figure 2.3: Structure of model. Reprinted from [13].

Different architectures implementations are available to extract information from multimodal data. A model like the one used in [14] can be a valid alternative when cross-modal information need to be extracted. In this case a cross-modal LSTM network is used for weight objective prediction. The different modalities are trained in different LSTM structures but, from the second layer on, the features from one LSTM can be concatenated to the others of another channel (Figure 2.4).

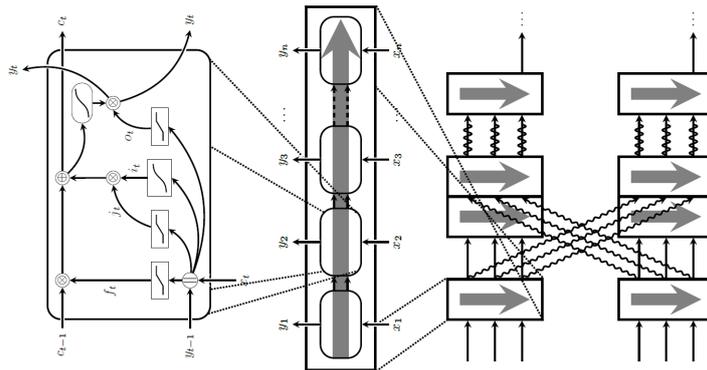


Figure 2.4: Structure of model. Reprinted from [14].

In [15], an interpretable model is built for outcome prediction in ICU. The idea is to adopt an approach called "knowledge distillation". A deep

learning model is used to train the data, as long as it is accurate enough to transfer its knowledge to a smaller and faster model. In this case the employed "student" model is gradient boosting trees (GBT): it has the advantage to have a good learning capacity and at the same time high interpretability. This is an important feature in the health applications like ICU where the comprehension of the features can help to define decision rules (Figure 2.5).

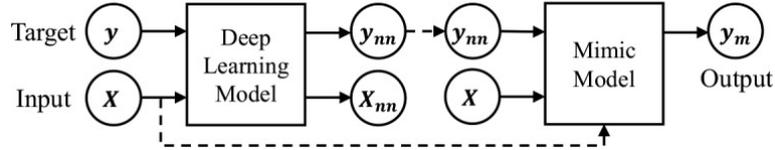


Figure 2.5: Structure of model. Reprinted from [15].

Finally, in [16] two interesting approaches are used for the early detection of sepsis in the ICU using the MIMIC III dataset. The first method consists of the combination of multitask gaussian processes and temporal convolutional network: gaussian processes are used as adapter to directly feed the temporal convolutional network even if the data are not equally spaced in time. This approach is compared to the combination of dynamic time warping and k-nearest neighbors (KNN): the KNN is trained over the distance matrix obtained from the dynamic time warping algorithm. Both the approaches obtained good results but the first method is preferred since dynamic time warping requires significant computational power and storage space for computing the distance among all the time series.

# Chapter 3

## Physiology

A lot of sensors are used in the ICU in order to immediately detect critical situations that need doctors' intervention. Some vital signs are monitored for all the patients, other parameters instead are monitored depending on the specific patient's injuries. The dataset used for this study contains a huge number of variables acquired from the patients, but only a subset of these variables, thus reducing the degree of freedom of the system. Moreover some variables are not included since they are acquired for a small subset of patients with particular conditions.

### 3.1 Vital signs

#### 3.1.1 Respiratory system

**Respiratory rate** The respiratory rate (RR) is the one of most important parameters concerning the respiratory system. It provides information about the balance of oxygen and carbon dioxide in the body. A high number of diseases can affect the RR, even if they are not related specifically with the respiratory system like head-injury and heart problems: this is the reason why this parameter is useful to detect critical conditions. A normal range for adults is 12-20 breaths per minute [17]. The simplest way to measure the RR is probably looking the patient's chest and count the number of times it rises in 15 seconds; the rate per minute is obtained multiplying the result by four. It is evident that it is not possible to apply this method in a hospital unite that requires a continuous monitoring. The most common automatic methods to measure the RR are methods based on airflow and movement detection. In the first case a sensor placed in the airway measures

the temperature variation with a thermistor or the pressure variation with a pressure transducer. In the second case strain gauges or impedance methods can be applied: they detect the volume chest variation to determine the RR [18].

**Oxygen saturation** Oxygen saturation is the percentage of hemoglobin bounded to oxygen: this measurement is commonly known as SpO<sub>2</sub>.

$$SpO_2 = \frac{O_2Hb}{O_2Hb + HHb} \quad (3.1)$$

Pulse oximeter is the device usually devoted to this measurement. It works accordingly to Lambert-Beer law which states that the absorbance of monochromatic light by a chromophore depends on the absorption coefficient at the specified wavelength, the optical distance and the chromophore concentration.

Considering that the O<sub>2</sub>Hb and HHb absorb red and near-infrared (IR) light in a different way, it is possible to define the concentration of these molecules measuring the rate between emitted light and received light at the two different wavelengths [19].

$$Absorbance_{\lambda 1} = \log \frac{I_0}{I} = \alpha_{\lambda 1 O_2 Hb} \times C_{O_2 Hb} \times d + \alpha_{\lambda 1 HHb} \times C_{HHb} \times d \quad (3.2)$$

A standard range for healthy subjects is between 94% and 98% [17]. Low levels of SpO<sub>2</sub> mean lack of oxygen in the peripheral circulation and therefore slowed cellular metabolism. The cause of this condition called hypoxia might be due to either respiratory or heart impairments.

### 3.1.2 Circulatory system

The heart is the main organ of the circulatory system. It is composed by four chambers: two atria and two ventricles for left and right side. The vena cava brings the de-oxygenated blood from the systemic circulation to the right atrium: the blood is pumped in the right ventricle and then is pumped again in the pulmonary circulation where it is re-oxygenated. At this point the blood full of oxygen flows into the left atrium and is pumped in the left ventricle. The contraction of the left ventricle, which is the most muscular part of the heart, lets the blood flow in the whole body [17].

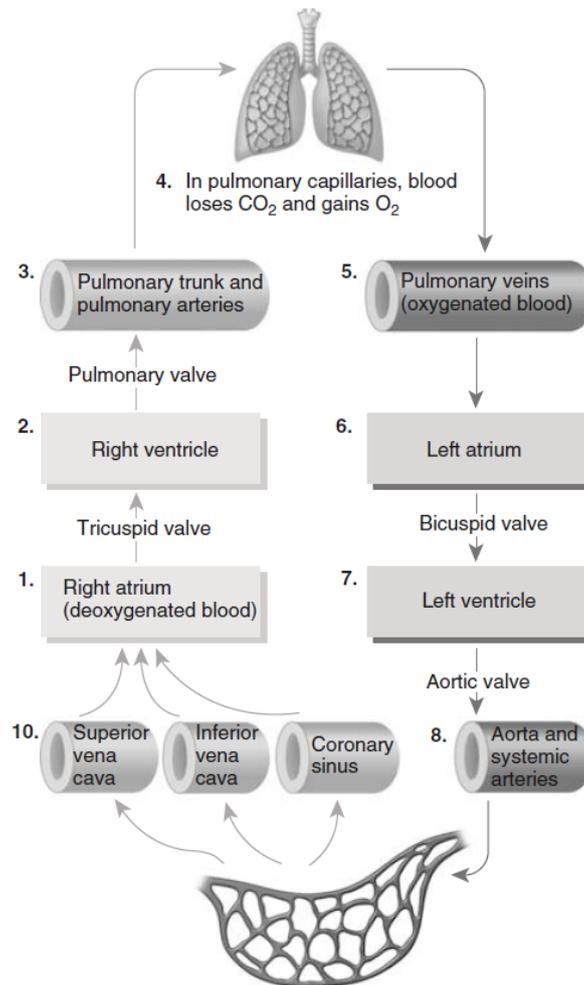


Figure 3.1: Structure of model. Reprinted from [17].

**Blood Pressure** The blood pressure is a measure of the pressure on the wall of large arteries. The measure unit is *mmHg*. Different type of pressure can be measured depending on the specific department:

- *Systolic pressure (SP)*: the pressure in the large arteries and is related to the systolic contraction. A common value is around 120 mmHg.
- *Diastolic pressure (DP)*: the pressure in the large arteries and is related to the diastolic contraction. A common value is around 80 mmHg.
- *Pulse pressure*: the difference between SP and DP.
- *MAP*: considered as the mean value of pressure in the arteries. It is

computed as follow:

$$\text{MAP} = \text{DP} + \frac{(\text{SP} - \text{DP})}{3} \quad (3.3)$$

Abnormal values of pressure can be due either to physiological reasons (i.e. physical activity, aging) or pathological reasons. In the last case the detection of the problem origin requires a wide range investigation since BP is determined by many factors:

- *Blood volume*: the higher the blood volume the higher the pressure on blood vessels.
- *Stroke volume*: the volume ejected by the left ventricle during the contraction.
- *Compliance*: the ability of blood vessels to increase their volume thanks to their elasticity when BP rises up. This characteristic in arteries is important because the elastic energy absorbed by vessels' walls is used during the diastolic phase to pump the blood: in this way a continuous blood flow guarantee a reasonable value of pulse pressure.
- *Viscosity and Peripheral resistance*: viscosity is an intrinsic characteristic of the blood that depends on the molecules inside the plasma. This parameter, together with the vessels' diameter, influences the fluid dynamic resistance and hence the pressure inside the vessel. The pressure is proportional both to fluid dynamic resistance and blood flow:

$$\Delta P = \text{Resistance} \times \text{Blood Flow} \quad (3.4)$$

A normal value of pressure ensures that the capillary hydrostatic pressure is high enough to allow the flow from the capillaries to the tissues: indeed it is the gradient pressure that drives nutrients from the capillaries to the cells. The opposite force that drives nutrients from tissues to capillaries is the osmotic pressure: nutrients flow from an area with high concentration to another with low concentration [20].

In the ICU the main techniques for BP monitoring are based on invasive and non-invasive measurements. In the first case invasive arterial blood pressure is recorded: a cannula needle inserted in an artery brings a column of fluid from the artery to a transducer that transforms the pressure in an electric signal [21]. In the second case non invasive blood pressure measurements based on oscillometric techniques are adopted. An arm cuff is wrapped

around the brachial artery and the pressure is brought to a value higher than the SP. The cuff is slowly deflated and the its pressure is measured constantly: a first event is detected when the pressure inside the cuff is equal to the SP and a second one when it is equal to the DP [22].

**Heart rate** The heart rate (HR) is a measure of how many times per minute the heart performs a contraction. A lot of literature has shown that HR is a good indicator for severe illnesses [23]: this is the reason why this parameter is also used in scoring systems to predict mortality in ICU. HR can be easily measured by technicians by checking the manual pulse or can be automatically extracted from device measurements as ECG or pulse oximeter.

### 3.1.3 Temperature

Everytime a chemical reaction occurs in the human body there is an exchange of energy. The sum of all these reactions represents the metabolism. The temperature inside the human body is not equal in all the sites: upper and lower limbs are more affected by external conditions like exposure and peripheral vasomotion; temperature inside the inner parts of the body such as central nervous system, chest and abdomen is regulated in a more strict way since sudden changes of temperature in these departments can have critical physiological effects. The sites where temperature can be measured are different and so the way to measure it: usually depending on the accuracy required a proper site near the core is selected [24].

The most common instrument for measuring the temperature is the thermometer: it is based on the characteristic of some fluids to change their volume depending on the temperature where they are placed. The most common fluid used in the last years was mercury: because of its toxicity and long response time of the instrument other electronic solution are now used to measure this vital sign [25].

The devices based on electronic components usually use one of the following tools:

- Resistance temperature: some resistances change their value depending on the temperature where they are placed. Knowing the temperature coefficient of the material it is possible to measure indirectly the temperature.
- Thermistor: the principle is similar to resistance temperature but in this case the fabrication material are semiconductors. The advantage of these

materials consists on the non-linear response they have: indeed even small changes of temperature can reflect in a big change of resistance.

- Thermocouple: in this case the temperature is computed measuring the voltage between the end of two wires fabricated with different materials.

A normal temperature range is between 36.5°C and 37.5°C.

## 3.2 Urinary output

Urine output is often used in ICU as a marker to detect possible failures in kidneys function but also as indicator of the intravascular volume status [26]. This value is usually acquired every hour by reading a graduate container connected to the patient's bladder through a catheter. The presence of the health-care staff is not necessary when devices capable of automatically measuring the patient's urinary output are used. Physiological values for adults are higher than 0.5 mL/kg/h: an urinary output lower than this threshold for more than 6 hours is associate to high risk of kidney failure [27].

## 3.3 Glasgow coma scale

The Glasgow Coma Scale (GCS) is a score used as indicator of the consciousness state of a patient. The score is between 3, which indicates complete unconsciousness, and 15, which indicates perfect consciousness. In order to compute this score both visual, verbal and motor response is tested [28]. This score is used as parameter to compute many other ICU scores such as SAPS and APACHE.

## 3.4 Arterial blood gas measurements

Arterial blood gases are analyzed to investigate possible failures in respiratory and metabolic system. The measures are done collecting blood from an arterial line (usually radial artery is used as access point) [29]. The blood gas used in this work are shown in the table 3.1:

Blood Glucose	Physiological range for FDA is lower than 100 mg/dL
Ionized Calcium	Physiological range si defined by each clinical laboratory
$FiO_2$	Fraction of inspired oxygen is important for patients with ventilation failure. The standard value is 0.21
$HCO_3$	Important to detect acid-base disorders. Physiological range 22 to 26 mEq/L
Hgb	Physiological range of hemoglobin for men is 13.8 to 18.0 g/dL
Lactate	Lactate is produced when anaerobic metabolism takes over after $O_2$ estinguish
methHgb	This hemoglobin can not bind oxygen. Physiological range of saturation 1-2 % Normal
$PaCO_2$	Partial pressure of carbon dioxide. Physiological range:38 to 42 mm Hg
$PaO_2$	Partial pressure of oxygen. Physiological range: 75 to 100 mm Hg
PF Ratio	Ratio between $PaO_2$ and $FiO_2$
pH	Physiological range: 7.38 to 7.42
Potassium	Physiological range: 3.5-5.2 mmol/L
SBE	Standard base excess (SBE). Physiological range is comprised between -2 and 2
Sodium	Physiological range: 135-145 mmol/L

Table 3.1: Different ABG used as input.[30] [31] [29] [32] [33]

# Chapter 4

## Materials and Methods

Python 3.5.2 has been used as programming language for all the steps of this work. Keras is a neural network library used in this work on top of TensorFlow to develop the deep neural networks. The data were already collected and organized in different files for every variable. The data were stored on Precision partition accessible through Computerome, a Danish national supercomputer used for life sciences research.

### 4.1 Model selected

LSTM model, introduced by [34], are particular kind of RNN that have shown particular abilities to remember both short and long term dependencies overcoming the problems present with RNN. Standard RNN differ from feed forward networks since they do not take into account only the current input but also the previous outputs. The basic concepts behind RNN are equal to a normal feed-forward network:

- The loss function, also known as cost function, is computed depending on the difference between the target and the current input.
- Partial derivatives are computed to define the gradient that minimizes the loss function.
- Each weight of the network is updated proportionally to its derivatives to minimize the loss function.

The problem that mostly affects RNN and partly the deep feed-forward neural networks is the vanishing (or exploding) gradient. To better explain

why RNN have this problem is easier to visualize the RNN as a feed forward network, where the recurrent connection are unrolled over the time, as shown in Figure 4.1.

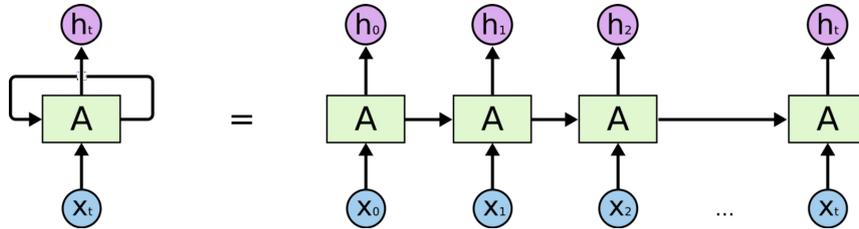


Figure 4.1: RNN unrolled over the time. Reprinted from [35]

When the gradient is computed with respect to a weight at the end of the network, the number of times the weights are multiplied each other is low enough not to cause troubles. Instead, when the gradient is computed with respect to a weight that resides in the earlier layers of the network, the number of terms multiplied together is too high: this results in a really unstable gradient that vanishes or explodes depending on the values multiplied together. When the values are lower than 1 the gradient vanishes rapidly while it explodes when they are higher than 1. The problem is particularly evident in RNN since the weight associated to the temporal loop is repeated through all the time steps [36] [37]. While RNN fails to remember long term dependencies because of the vanishing gradient, LSTM have proved to be particularly able to learn both long and short term dependencies.

The reason why LSTM fits so well this task resides in the structure of the cell state, shown in Figure 4.2.

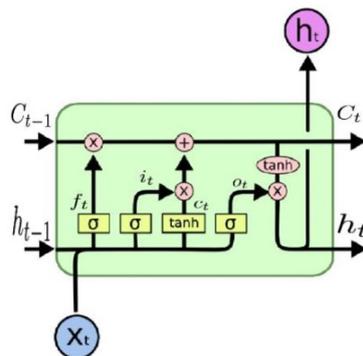


Figure 4.2: LSTM cell. Reprinted from [35].

The cell consists of four gates that manipulate the current information considering also output far back in time. The cell state runs through all the

time steps and can be modified by two gates: forget gate and input gate. The forget gate decides how much of the previous cell state keep or forget: 1 will leave the cell state unchanged while 0 will erase all the information content.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) = \sigma(W_{hf} \cdot h_{t-1} + W_{xf}x_t + b_f) \quad (4.1)$$

The next step is to decide how to update the new cell state: this is done through the input gate that decides how much of the new vector of candidates  $C'_t$  should be updated.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) = \sigma(W_{hi} \cdot h_{t-1} + W_{xi}x_t + b_i) \quad (4.2)$$

$$C'_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) = \tanh(W_{hC} \cdot h_{t-1} + W_{xC}x_t + b_C) \quad (4.3)$$

All the steps listed above are used to create the new cell state, as it is shown below:

$$C_t = f_t * C_{t-1} + i_t * C'_t \quad (4.4)$$

The new cell state is used together with the output gate to compute the new output of the cell state, that will be also used for the next time step.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) = \sigma(W_{ho} \cdot h_{t-1} + W_{xo}x_t + b_o) \quad (4.5)$$

$$h_t = o_t * \tanh(C_t) \quad (4.6)$$

The operation process of LSTM cell solves the problem of RNN thanks to its ability to shut down a gate, thereby preventing the cell state to be updated over many cycles. Even if the gate is left open, the cell state is not completely updated but it will store a weighted contribute of both previous and current state [35].

# Chapter 5

## Data preparation

### 5.1 Preprocessing

In this section the steps implemented to build the input for the training are described. The dataset contained different CSV files, each of them containing information for a specific variable (e.g. blood pressure, oxygen saturation etc.). Some variables consist of 'high frequency' data since their values are collected from the patient's bed at least every 30 seconds. The sampling frequency can be incremented automatically by the instrumentation if the conditions are considered critical. The 'low frequency' data are sampled without a fixed sampling rate: their rate usually goes from every hour up to once a day. The variables belonging to this category are all the ABG measurements, GCS and Urinary output, while all the vital signs are high frequency data.

The information related to a variable include acquisitions for all the patients involved in the study who have at least a measurement of that specific variable during their hospital stay. This means that a variable like heart frequency, which is the physiological data with the higher amount of details, will have a record for every acquisition, for every time-step and for all the patients independently on how long the acquisition has lasted. The whole dataset contains information from 13587 patients, but some of these patients are excluded since they do not satisfy the following requirements:

- The hospital stay is longer than 24 hours.
- The patient is older than 16 years.
- The patient has a record in the Danish National Patient Registry (necessary to have information about that patient).

The preprocessing steps are done in order to achieve a specific input shape. The number of rows represents the number of time steps, the number of columns corresponds to the features extracted from the raw data, and the depth of the matrix corresponds to the number of patients used as input. Figure 5.1 shows how the matrix is structured.

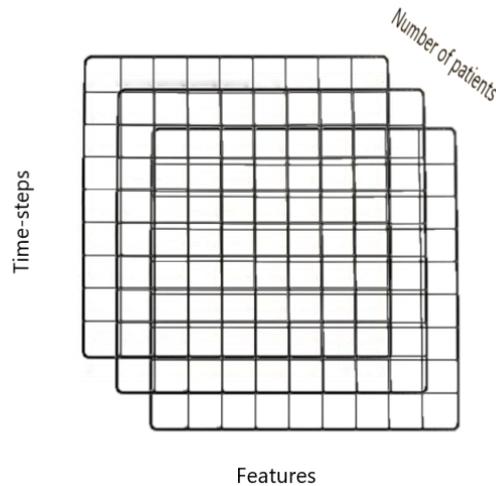


Figure 5.1: Shape the model receives as input.

Since the number of time steps is related to each hour after the admission in the ICU, the following arrangements are made:

- For every hour outliers are removed. A measurement is considered as outlier if it is outside the interquartile range (the same method used by boxplot). Outliers are usually due to misstatement of the instrumentation ( i.g an electrode is accidentally unplugged).
- After outliers are removed, the time axis is resampled every 15 seconds for high frequency data and every hour for low frequency data. What usually happens is that resampling produce many missing values since no values are found in that time range, so a linear interpolation is performed to avoid *NaN*. The resampling step, together with interpolation, fixes the problem related to filling the gap between two different acquisition (i.g. patients who went under surgery do not have parameters recorded in the meanwhile).

## 5.2 Feature extraction

Considering the shape of the input matrix fed into the model, the number of time steps is the only fixed dimension. Indeed features and number of patients were tested in different ways according to Table 5.1.

Number of patients	Number of features
Common patients	<ul style="list-style-type: none"> <li>- vital signs</li> <li>- vital signs and static information</li> <li>- vital signs, static information and arterial blood gas measurements</li> <li>- vital signs, static information and arterial blood gas measurements (with IMF)</li> </ul>
All patients	<ul style="list-style-type: none"> <li>- vital signs, static information and arterial blood gas measurements</li> <li>- vital signs, static information and arterial blood gas measurements (with IMF)</li> </ul>

Table 5.1: Different input used for training.

The input are not randomly selected but they are chosen starting from the easiest situation, then different hypothetical improvements are added to verify how the performances of the model are affected by these modifications.

Table 5.1 shows two different conditions regarding the number of patients: in the first case "Common patients" means that the input consists of all the patients who have at least one measurement for all the variables used as input. The second case, "All patients", means that the whole dataset is used as input, no matter which variables are used for the prediction. The first case is easier to manage since there is no need to deal with patients whose recordings are missing for the first 24 hours in a row. The problem of this approach is that, every time new variables for the input patients are added, the subset of patients with a measurement for all the variables shrinks, and so the benefit of having a big dataset is not fully exploited.

Indeed, if the whole dataset contains 9415 acceptable patients, when the subset is restricted to the patients with a measurement for all the vital signs the number drops to 5071. The addition of static information (age, gender and weight at the moment of admission) does not impact the size of patients set since these information are acquired for all the patients who are admitted

in the ICU, no matter what condition they have. The smallest number of patients included in the input occurs when also arterial blood gas measurements are used in addition to vital signs: in this case only 4031 patients are acceptable. It is important to take into account that the vital signs mentioned from the previous table on, include not only the vital signs listed in section 3.1 but also Urinary output and GCS.

The number of features depends on the variables used. As concern low frequency data the mean value of every hour is used as input at every time step. Static information do not change over time, so their value is just repeated for every time step. As regards high frequency data, a set of features explained in the following section is used.

### 5.2.1 Statistic features

Every hour of high frequency data consists of a set of values sampled every 15 seconds, therefore 240 samples are used to extract the features. The statistical features consist of:

- Arithmetic mean:

$$\bar{x} = \frac{1}{n} \left( \sum_{i=1}^n x_i \right) = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (5.1)$$

- Standard deviation:

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad (5.2)$$

- Skewness is used as measure of the symmetry of the probability distribution for all the recordings within the hour considered. A negative skewness appears with a distribution skewed to the right and vice-versa for positive skewness.

$$Skewness = \frac{m_3}{s^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (5.3)$$

- Kurtosis, as skewness, provides information on the shape of the distribution within the hour considered. In particular Kurtosis is the measure of the thickness or heaviness of the tails of a distribution with respect to the normal distribution.

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} - 3 \quad (5.4)$$

- Minimum value
- Maximum value

### 5.2.2 Entropy measures

- Shannon entropy:

$$H(X) = - \sum_{i=1}^n P(x_i) \ln(P(x_i)) \quad (5.5)$$

In equation 5.5,  $P(x_i)$  is the probability that the variable  $x_i$  occurs. If the signal has a probability distribution where most of the values have high probability to occur, the logarithm will be closer to zero and so the entropy. On the contrary, signals containing many rare values will have high entropy since the logarithm will approach minus infinity.

- Sample entropy:

$$\text{SampEn}(m, r, N) = - \ln \left[ \frac{A^m(r)}{B^m(r)} \right] \quad (5.6)$$

$$B^m(r) = \frac{1}{(N - m)} \sum_{i=1}^{N-m} B_i^m(r) \quad (5.7)$$

$$A^m(r) = \frac{1}{(N - m)} \sum_{i=1}^{N-m} A_i^m(r) \quad (5.8)$$

This measure was introduced in [38]. Given a signal,  $N$  is the number of samples,  $m$  is the embedding dimension and  $r$  is a tolerance value to admit two segments as similar. Two segments are considered similar if  $d[\mathbf{x}_m(i), \mathbf{x}_m(j)] \leq r$ , where the distance is defined as the maximum difference of their corresponding scalar components.  $B_i^m(r)$  represents the number of matches with dimension  $m$  over  $N - m - 1$ ;  $A_i^m(r)$  analogously, represents the number of matches for the embedding dimension  $m + 1$  over  $N - m - 1$ . A signal with the same number of matches for both  $m$  and  $m + 1$  embedding dimension will have a low entropy since the logarithm approaches 1, and the opposite will occur if the rate  $A^m(r)/B^m(r)$  is close to 0.

### 5.2.3 Empirical mode decomposition

The EMD is a method used to decompose a signal into other signals called IMF. The principle is similar to Fourier transform but, instead of using sine and cosine as basis functions, here an iterative algorithm is used to find the simple harmonic signals. The algorithm is explained in detail in [39] and is used by a wide number of authors to extract features for classifiers [40] [41] [42].

The IMF must respect the following rules to be a basis of the original signal:

- The number of extrema and the number of zero crossings must either be equal or differ at most by one.
- At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The procedure to extract the IMFs follows the steps listed below:

- Compute the upper envelope as the cubic-spline interpolation of local maxima
- Compute the lower envelope as the cubic-spline interpolation of local minima
- $h_i$  mode function is computed as the difference between the original signal and the average between the two envelopes.
- The procedure is repeated until a stopping criteria is satisfied

The stopping criteria used in this work is the one the python library uses by default:

$$SD = \sum_{t=0}^T \left[ \frac{|(h_{1(k-1)}(t) - h_{1k}(t))|^2}{h_{1(k-1)}^2(t)} \right] \quad (5.9)$$

When the  $SD$  reaches the value of 0.2 the iterations are stopped. In this work mean value, minimum value, maximum value and variance are extracted from IMF1 and IMF4 to increase the number of information from time series. These features are not used for all the input tested for the model but only for those reported in Table 5.1.

## 5.3 Outcome prediction

In this work both classification and regression are tested. In the first case a binary classification model classifies as 1 or 0 depending on the mortality before or after 30 days. The regression model, instead, uses as output the number of days of survival after the admission in the ICU. The binary classification task is easier to model since no particular processes need to be done to define which is the output for every input. The regression task instead, requires taking into account the issue related to censored patients. Censored patients are the patients whose event of interest - in this case death - occurs after the end of the study time period. Indeed for these patients the output is not the real length of survival but just the number of days they are monitored within the study period: for these patients the event of interest could have happened whenever after that date. Two different strategies are implemented to deal with censored patients.

### 5.3.1 Classification

The binary classification is performed classifying with 1 the patients who died before 30 days after the admission in the ICU and with 0 all the others. An important consideration is that the number of patients belonging to the two classes is not the same, hence some methods need to be used to solve the imbalance between the two classes. In this work the *fit\_generator* Keras function is used to yield the same number of samples from both the classes every time the generator is called.

### 5.3.2 Regression for 90 days of mortality

In this case the model is designed to predict the survival time within 90 days after the admission. Indeed its output is a number between 0 and 9 if the death occurred within 90 days after the admission or 10 if the death occurred afterwards. Figure 5.2 shows how the survival time is mapped to the output.

In this case all the patients could be included in the training since no patient in the dataset is admitted earlier than 90 days before the end of the study time period. In this case the model cannot predict any outcome related to long-term survival hence it is supposed to focus more on short range dependencies.

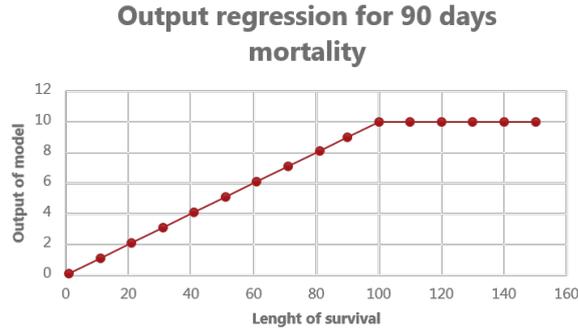


Figure 5.2: Output of the model for regression of 90 days of mortality.

### 5.3.3 Regression including censored patients

A high number of patients are censored patients since they survived after the end of the study time period. If the model aims at predicting the exact survival time without any limit, a strategy to include also censored patients needs to be found. The solution adopted is similar to the one used in [43], where censored patients who survived less than a threshold are weighted depending on how long they are monitored. This approach can be used for all those models that allow to weight each input. In this case, the threshold is chosen observing the survival function estimated with the Kaplan–Meier estimator [44] represented in Figure 5.3: 900 days is the survival time after which the survival probability is below 50%.

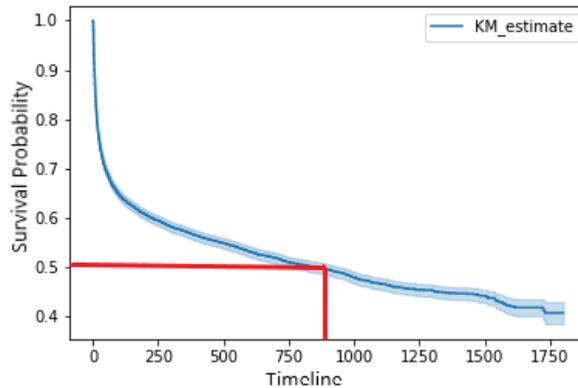


Figure 5.3: Kaplan-Meier estimation.

The not-censored patients and the censored patients with censored time higher the 900 days have weight equal to one since the model trusts their target output 100%. Censored patients who are monitored less than 900 days

are weighted depending on how high is the probability they have survived up to 900 days knowing that they have survived at least for the censored time. It turns out that patients monitored for a very short period have small weights because the date of their death is probably far from the observation period.

Since all the patients are used and the target is not restricted anymore, the survival time ranges from 1 to the maximum survival time (1800 days). Figure 5.4 shows how this range is mapped from 0 to 10 giving a higher importance to short length of survival.

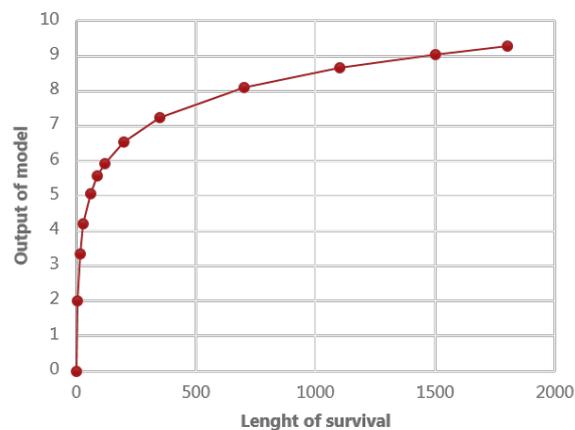


Figure 5.4: Output of the model for regression including censored patients.

The reason why the curve is not a straight line is that the same error in the prediction has a different impact depending on the value of the real target. Big errors are less relevant when the target is a long survival time.

## 5.4 Cross-validation

A parameter estimation using a grid search with stratified cross-validation is done to define the best model for classification and regression. In the following the parameters tuned for every model are listed:

- Number of LSTM layers [1, 3].
- Activation function of the lstm cell [sigmoid, tanh].
- Dimension of the output space [32,64,128, 180, 200, 256].

- Number of steps to yield from generator before declaring one epoch finished [1,3].
- Number of folds for kfold [5,10].

The classification task is evaluated depending on the MCC for each validation split. The best parameters are selected for every model and then performances are evaluated on their correlative test set. MCC was introduced in [45] and it is shown below:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5.10)$$

It is really meaningful when binary classification with classes considerably imbalanced is performed. Indeed, this score takes into account at the same time the influence of all the four sections of the confusion matrix providing an overall assessment of the performances [46].

The parameter used to evaluate the performances of regression is the MSE.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5.11)$$

MSE is used to evaluate the performances of an estimator: its value is always positive and performances are better when its value is closer to zero.

# Chapter 6

## Results

In this chapter the results obtained for all the models are illustrated. First, an introduction on a method used to evaluate the impact of the different features is introduced, then the results are listed in three sections depending on the prediction the model aims to.

### 6.1 SHAP

Usually deep models like the one used in this work, are referred as "black box" because of their lack of interpretability. Indeed, even though network architectures with high complexity allow to reach higher performances with respect to simple models, it gets difficult - if not impossible - to understand why the model returned that specific output. The SHAP technique, introduced in [47], allows to understand the contribute of each feature thanks to the SHAP value associate to the feature itself.

The contribution of a single feature,  $\phi_i$ , is calculated considering the difference between the output of the model when the subset  $S$ , extracted from the feature space  $F$ , includes or not the feature  $i$ . This is weighted for  $S!$  and  $(|F| - |S| - 1)!$  which respectively represent the number of permutations that can be obtained before and afterward the current feature is added to the subset  $S$ . Then this value is summed for all the possible combinations of  $S$  that can be obtained before the current feature is added. The result is normalized for the number of permutations that can be extracted from  $F$ .

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (6.1)$$

An easier way to understand the equation is considering two features  $x_1$  and  $x_2$ , with  $f(x_1) = 1$ ,  $f(x_2) = 2$  and  $f(x_1, x_2) = 4$ . If one wants to know the contribute of  $x_1$  when all the features are used the steps below should be followed:

- Calculate the output of the model when only  $x_1$  is used: in this case 1.
- Calculate the difference between  $f(x_1, x_2)$  and  $f(x_2)$ , which represents how the output changes when  $x_1$  is added to  $x_2$ . In this case the value of the difference is 2.
- $\phi_1$  is the sum of the previous two steps over the number of permutations that can be obtained with the current number of features: in this case, with two features, only two permutations can be obtained. So the contribute  $\phi_1$  for the variable  $x_1$  when all the features are used results to be 1.5. Similarly can be done for  $\phi_2$ .

Since the model used in this case is an LSTM network, the SHAP values matrix will have the dimension of  $\#samples \times \#timestep \times \#features$ . Each element of the matrix expresses how, for that patient at that specific timestep, that feature drives the prediction towards the predicted outcome. This means that, for every time step, the order of importance of the features will change: features that have big impact on the outcome prediction after 24 hours could be less significant after the admission in the ICU, and vice versa.

## 6.2 Classification

After all the parameters are tuned, the best model is selected depending on the validation metrics and then the final performances are evaluated on the test set. In Table 6.1, the best model selected for each classification input is explained:

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6
Validation split	10	10	10	10	5	10
Number of Layers	3	3	1	3	1	3
Number of steps	1	3	1	1	1	3
Activation function	Sigmoid	Sigmoid	Tanh	Sigmoid	Sigmoid	Tanh
Hidden units	256	64	64	180	64	64
MCC validation	0.46	0.55	0.56	0.60	0.55	0.54
MCC test	0.26	0.34	0.32	0.33	0.35	0.34

Table 6.1: Best performances for classification. Common Patients in Vital (Input 1), Common Patients in Vitals Static (Input 2), Common Patients in Vitals, ABG and Static (Input 3), Common Patients in Vitals, ABG and Static (IMF) (Input 4), All Patients in Vitals, ABG and Static (Input 5), All Patients in Vitals, ABG and Static (IMF) (Input 6).

As shown in Table 6.1, there is no particular parameter that is always the best for all the inputs. The MCC is used on validation set as metric to define the best set of parameters for that specific input. Another metric usually used in this application is the AUC: this value represents the integral of the Receiver operating characteristic (ROC) and can be interpreted as the probability that the classifier predicts with a higher value a random element from the class 1 with respect to a random element from class 0. Figure 6.1 shows a comparison of MCC and AUC for the different input.

In both Table 6.1 and Figure 6.1 it is possible to see that performances are definitively lower when only features extracted from vital signs are included (the label vital signs includes both the vital signs listed in the section 3.1 plus urinary output and GCS). The most important improvement is when static features are added as input of the model: performances do not change significantly when ABG measurements or IMF features are included. When the number of patients changes from "Common" to "All" it might be expected that performances would drop off because of the high number of patients with missing variables acquired in the hospital unit: actually this does not happen. Indeed performances are almost the same with the advantage that, in this case, the model works the same with every input it receives. The reason why performances are not affected can be explained by the fact that, in the ICU, the decision whether to record a variable or not from a patient

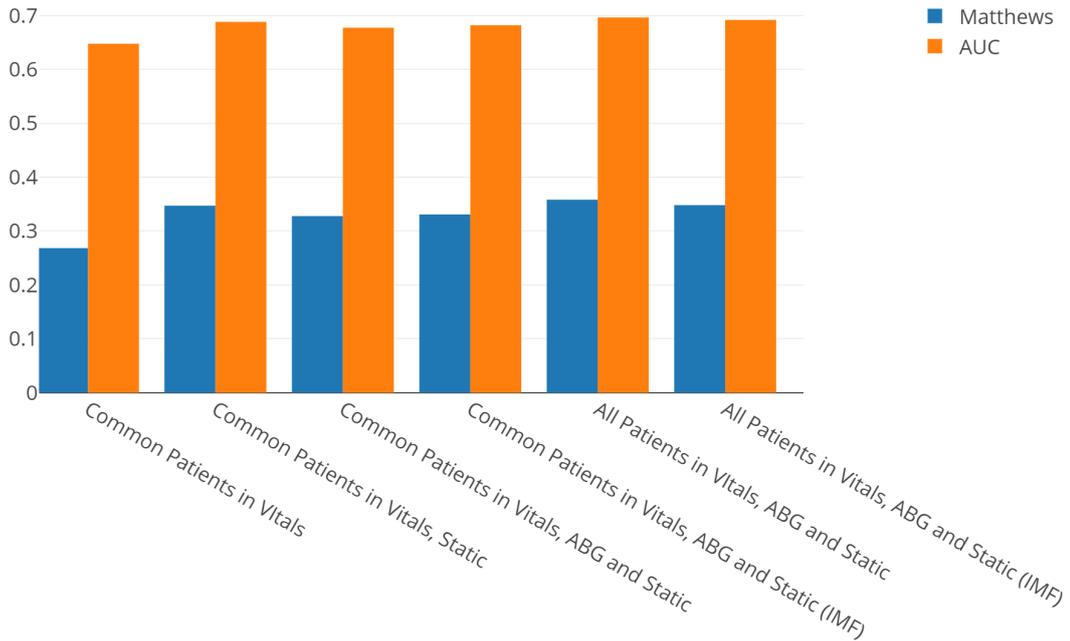


Figure 6.1: MCC and AUC for different input.

is up to the doctors: the fact that a variable is not present simply means that that physiological parameter is not relevant for the injury the patient is in the hospital for. Since the missing features are substituted with the mean value of all the patients at a specific time step, those features will just be considered as a normal condition with the advantage that more patients, maybe even easier to classify, can be included in the dataset.

The SHAP values are extracted for all the models selected but only the one for the input "All Patients in Vitals, ABG and Static with (IMF)" are shown below: in this way the importance of all the features, including the one extracted with the EMD, can be compared. Figure 6.2 shows, listed in order of importance, the features contribute and how they drive the prediction towards 1 or 0.

As shown in Figure 6.2, each feature has a different contribute that is not necessarily the same for the subsequent timestep. Each dot in the chart represents a patient: when the dot has a positive SHAP value it means that for that patient that variable contributes to have an output closer to 1 and so to death; vice versa when the dot has a negative value.

This can be easily demonstrated observing the first three variables: when the age of the admission is high, prediction is driven towards 1 and so to death. This is reasonable considering that young patients heal easier than older one. On the opposite, high value of GCS, which are associate to conscious patients,

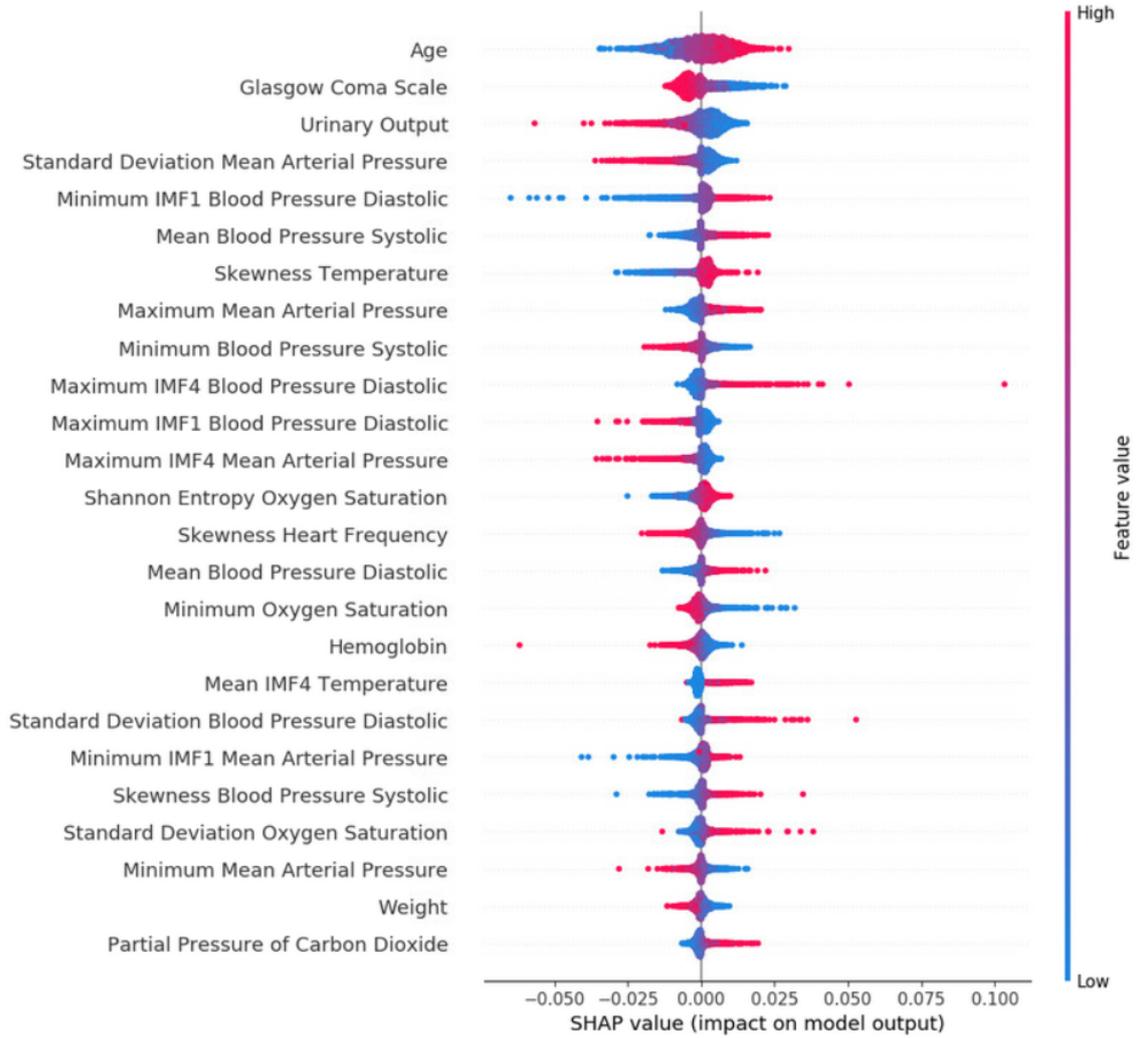


Figure 6.2: Features contribute for the first hour after the admission.

drive the prediction to 0, namely the patients who survive. The last example is the urinary output: low values of urinary output, usually associated to kidney failures, drive the prediction to 1 and the opposite for high values of urinary output.

This is less interesting when the impact of a variable is expectable, but really useful when the impact of a variable is completely unknown. Considering work [48] as example, the changes in the dynamical structure of the MAP are analyzed to see the effect on mortality. As a result, the dynamics with low variability are associate to high risk of mortality and vice versa. It would be interesting to see if any features associated to variability for the MAP have any impact to the prediction.

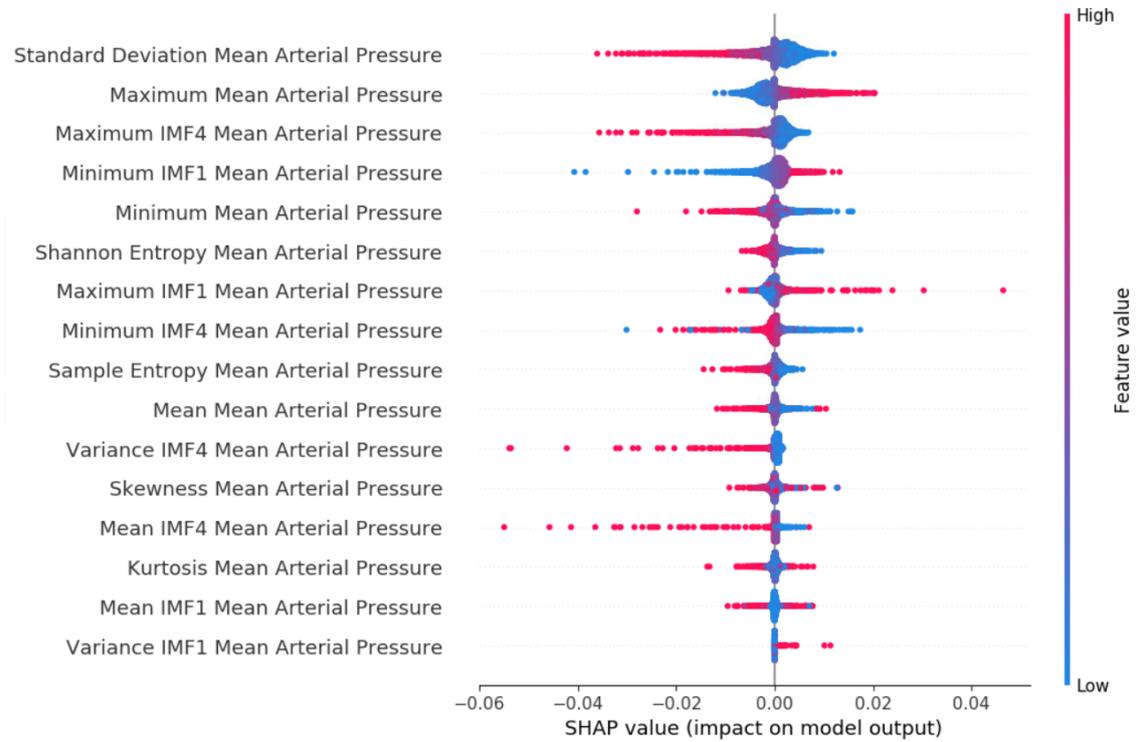


Figure 6.3: MAP contribute for the first hour after the admission.

In Figure 6.3 all the features extracted from MAP are ordered depending on the impact they have: it is clear that also in this study, high values of standard deviation are associated to low risk of mortality.

SHAP values provide also information on how the variables' importance changes over time: some variables that are significant in the early hours can be neglected after long time. In Figure 6.4, the variables are ordered depending on their importance and a box plot for each of them shows how they are spread over the 24 hours. The y-axis represents the position in the ranking: 1 correspond to the most important and 0 to the less significant. It is important to notice that the x-axis is not labelled with the single features but with the variables the features are extracted from: this is done gathering the SHAP values of all the features extracted from the single variable.

Some interesting observations can be deduced from Figure 6.4: age of admission results the most important variable for all the time step except for two of them, where it is though still important. The vital signs including GCS and Urinary output are placed in the right half of the x-axis, meaning that they are always more important than the ABG measurements, which are all in the left side of the x-axis.

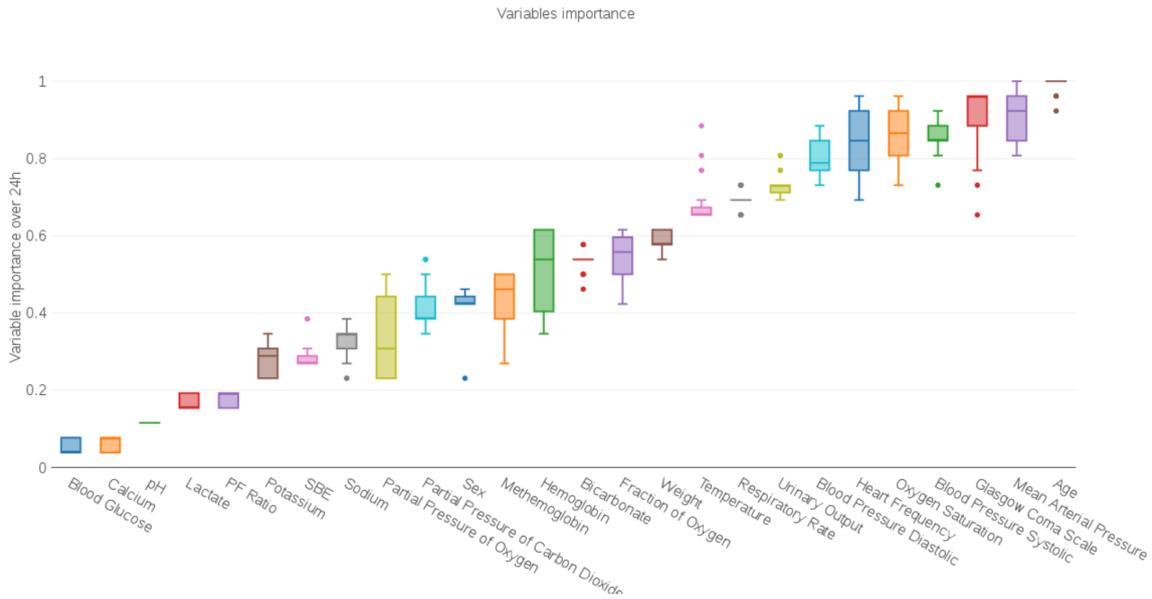


Figure 6.4: Classification task: variables importance over 24 hours.

Another interesting thing is to understand which features extracted from the different variables are more significant over the 24 hours.

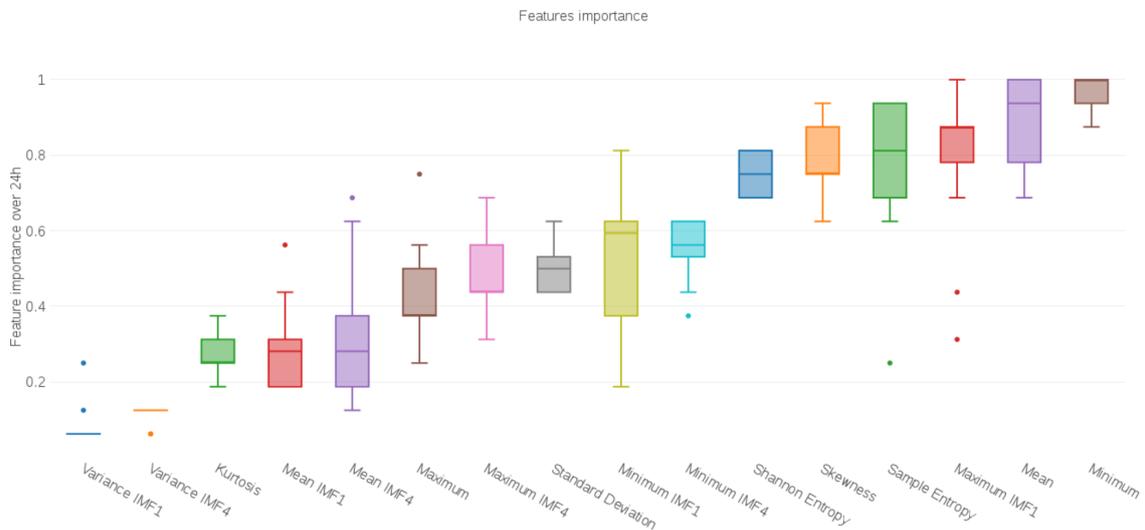


Figure 6.5: Classification task: features importance over 24 hours.

In this case the impact of all the features is shown ordered by importance. The feature that seems to provide more information is the minimum value. This does not necessarily mean that for all the variables the minimum value is the feature that accounts the most to predict mortality: it simply implies

that, considering the sum of the 'Minimum' features from all the variables, it has more influence rather than others. As can be observed in Figure 6.3, the standard deviation accounts the most for the MAP: anyway from Figure 6.5 standard deviation seems to be in the half of the ranking. Figure 6.6 shows that, even though standard deviation is the most important for MAP, it is not rather significant for other variables like heart frequency, respiratory rate and temperature.

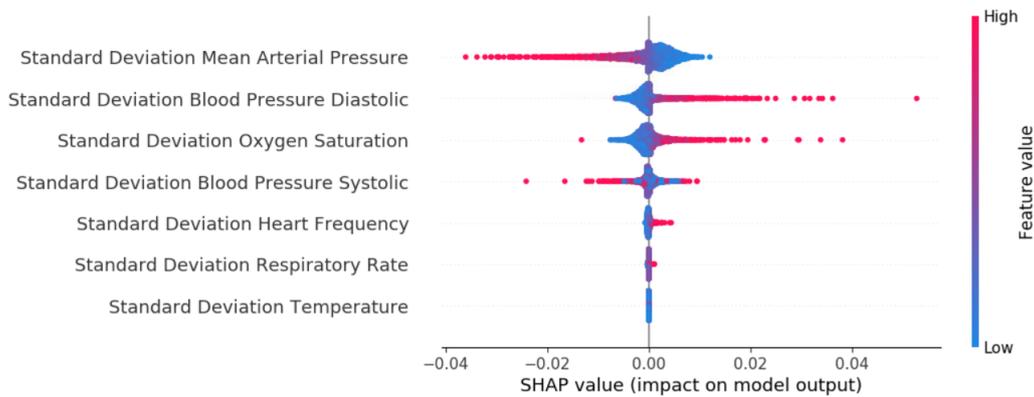


Figure 6.6: Standart deviation contribute for the first hour after the admission.

### 6.3 Regression for 90 days of mortality

In this section the results of the regression model are shown. As it is done for classification, the best set of parameters is selected depending on the best performances on validation and then the performances are checked on the test set (Table 6.2).

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6
Validation split	10	10	10	10	10	10
Number of Layers	3	1	1	1	1	1
Number of steps	1	1	1	1	1	1
Activation function	Sigmoid	Tanh	Tanh	Sigmoid	Tanh	Sigmoid
Hidden units	256	128	180	200	32	128
MSE validation	12.68	11.91	12.36	10.48	11.42	11.28
MSE test	15.89	14.45	13.28	14.00	13.67	12.99

Table 6.2: Best performances for Regression on 90 days. Common Patients in Vital (Input 1), Common Patients in Vitals Static (Input 2), Common Patients in Vitals, ABG and Static (Input 3), Common Patients in Vitals, ABG and Static (IMF) (Input 4), All Patients in Vitals, ABG and Static (Input 5), All Patients in Vitals, ABG and Static (IMF) (Input 6).

Table 6.2 leads to the conclusion that validation split, number of layers and number of steps are not affected by the different inputs: the parameters are respectively 10, 1 and 1. Activation function and hidden units still change a lot for the different inputs. In Figure 6.7 two scores are compared for this regression: mean squared error and  $R^2$ .

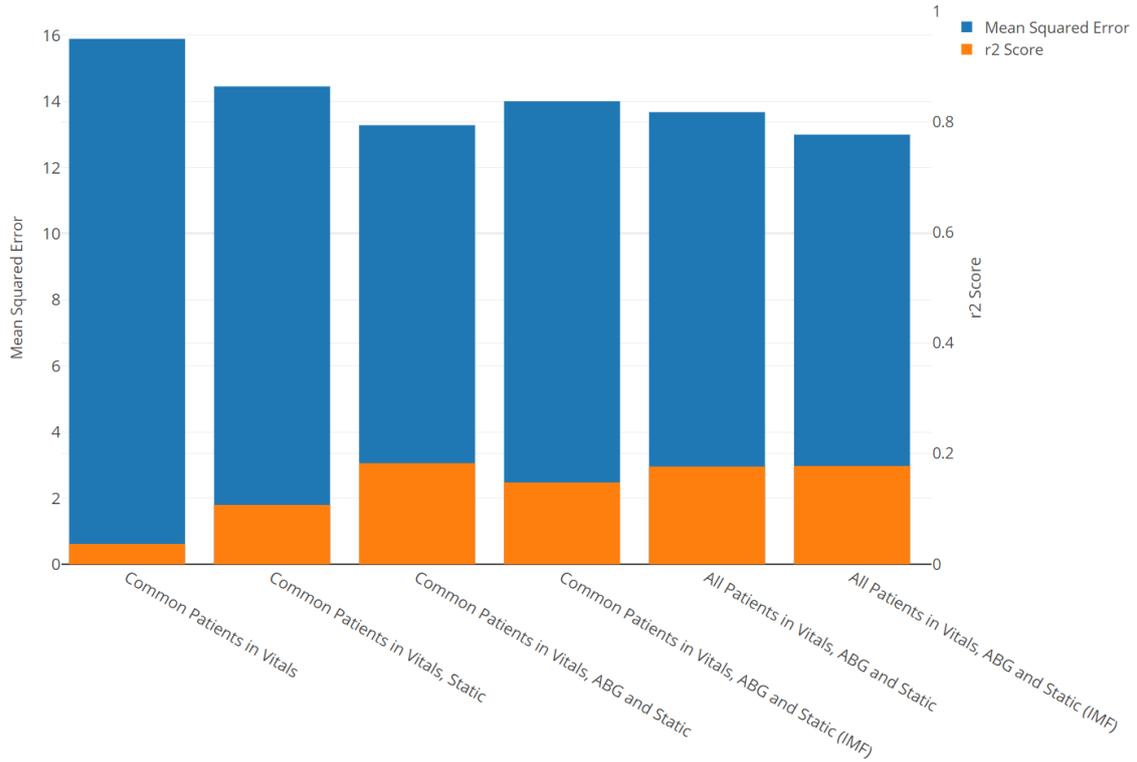


Figure 6.7: MSE and  $R^2$  for different input.

$R^2$  is a metric used for regression models that adds some more information than the MSE alone.

$$\hat{R}^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (6.2)$$

From Equation 6.2, it can be noticed that the fraction represents the rate between the MSE and the Variance: if the mean value of all the squared errors is comparable to the variance, then the  $R^2$  will be close to 0. If the MSE is considerably smaller than the variance, then the rate is closer to 0 and so the  $R^2$  close to 1. It is important to remember that, while MSE is

associate to good performances as it is close to 0,  $R^2$  has the best performance when it is close to 1.

Also for regression SHAP values can be used to interpret the importance of the features. It should not be assumed that the variables and the features have the same importance as in the classification task: indeed, even though the values of the input are the same, the output function of the model and the target are different. In Figure 6.8 the importance of the features for the first hour after the admission is shown.

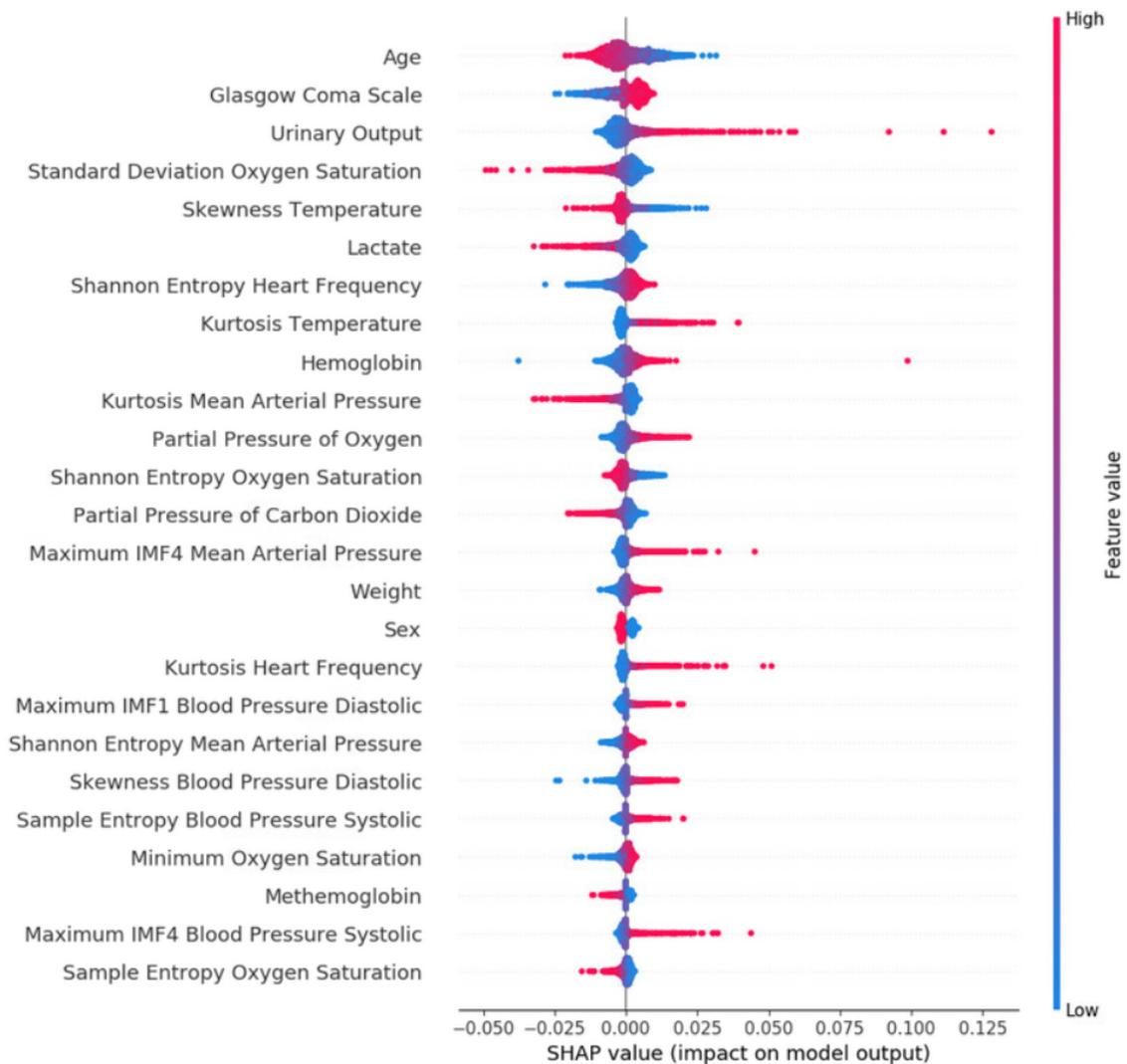


Figure 6.8: Features contribute for the first hour after the admission.

Although the order of the features could be similar to Figure 6.2, here the SHAP values are the opposite: positive values drive the prediction toward

high output number and so to longer length of survival while negative values for shorter length of survival. Considering age, which results to be in both cases the most important, the red dots associated to older people have negative SHAP values meaning that the output will be pushed toward short length of survival. Analogously to Figure 6.4 for classification, Figure 6.9 shows the importance for all the variable over the first 24 hours.

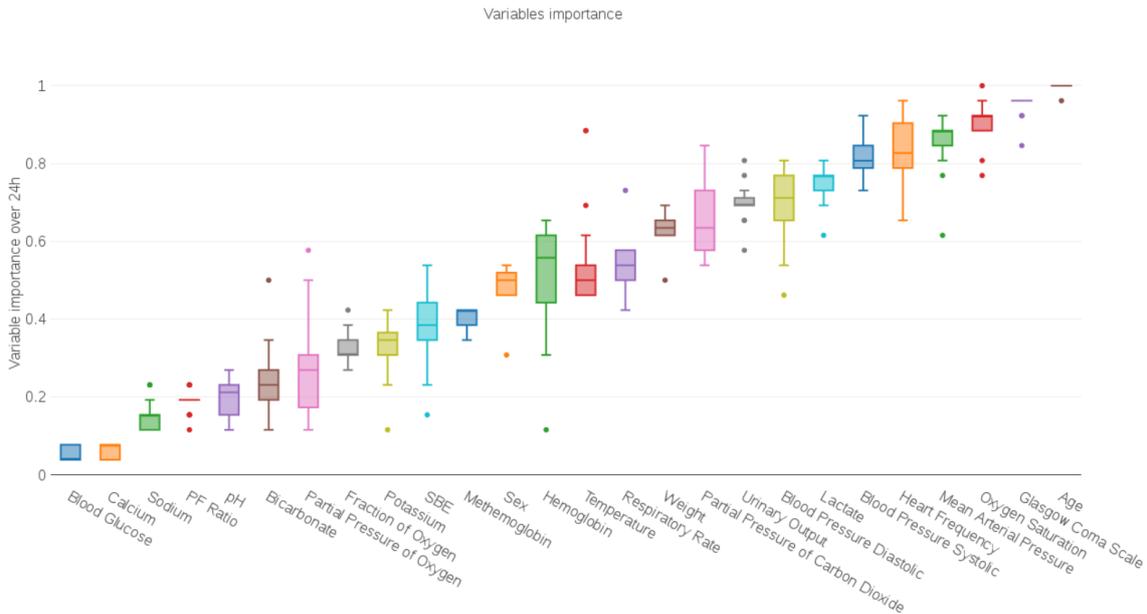


Figure 6.9: Regression task: variables importance over 24 hours.

There are no substantial differences between these two figures. Even if some variables are switched of one or two positions, the two macro categories "vital signs with Urinary output and GCS" and "ABG measurements" are well separated in the two halves of the x-axis. The only variable that seems to have a bigger impact in this kind of prediction, enough to be placed among the vital signs, is Lactate. Since here the prediction is based on the exact number of days, Lactate might be more influential for specific length of survival.

The importance of all the features extracted from the variables is shown in Figure 6.10.

Figure 6.10 shows that entropy measures have more importance for this prediction task. It would be interesting to understand whether the regression performs better or not with respect to classification. For a specific time range, prediction and actual length of survival can be mapped as binary class depending if they are below or over the specific time range. This allows, in addition to compare classification and regression, to see which is the time

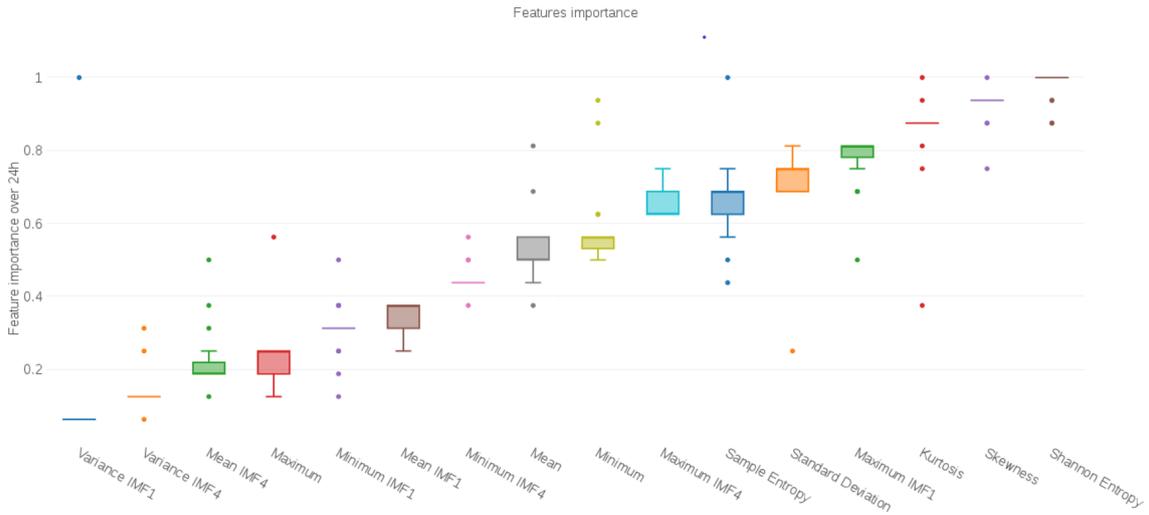


Figure 6.10: Regression task: features importance over 24 hours.

length the model performs better. Figure 6.11 shows, for all the inputs, three different metrics: precision, recall and MCC.

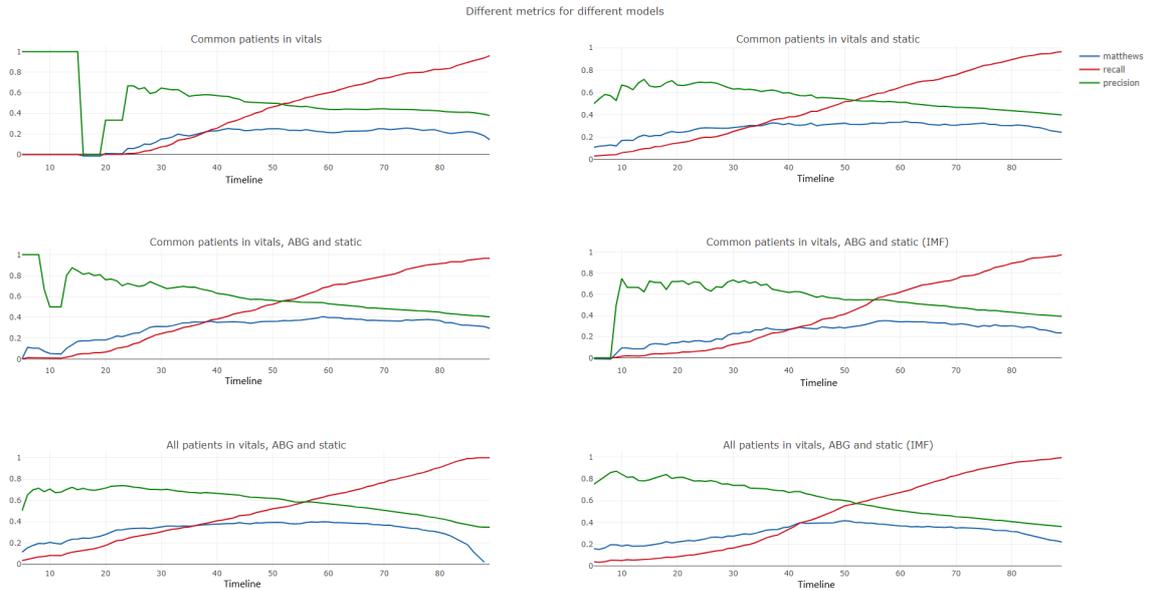


Figure 6.11: Different metrics for the input tested with regression on 90 days.

As explained in the subsection 5.3.2, the output of the model ranges from 0 days to 9 if the death occurred within 90 days, otherwise the output is 10. For all the models, as the survival time increases, precision decreases and recall increases. This is reasonable since almost all the patients are predicted with a value lower than 9 (90 days), so while the number of True Positive

increases in favour of recall, the number of False Positive increases at the expense of precision. The best compromise for all the models is placed at about 50 days of mortality which is also where the peak of MCC is placed. The best MCC for the classification task was 0.35 while now it is around 0.4 (both are evaluated on the test set).

## 6.4 Regression including censored patients

In the following section results will be shown in the same way as the previous sections: first Table 6.3 is used to show which are the best parameters and the results obtained for the different models. In Figure 6.12, MSE and  $R^2$  are compared over all the different models. Then Figure 6.13, 6.14 and 6.15 illustrate respectively the importance of the features for the first hour after the admission, the importance of the different variables over the first 24 hours and the importance of the features extracted from those variables over the first 24 hours.

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6
Validation split	10	10	10	10	10	10
Number of Layers	1	1	3	1	1	1
Number of steps	1	1	1	1	1	1
Activation function	Tanh	Sigmoid	Sigmoid	Tanh	Sigmoid	Tanh
Hidden units	200	128	200	64	256	256
MSE validation	5.19	4.30	3.96	4.46	4.52	4.52
MSE test	6.96	6.27	6.49	6.43	5.58	5.58

Table 6.3: Best performances for Regression including censored patients. Common Patients in Vital (Input 1), Common Patients in Vitals Static (Input 2), Common Patients in Vitals, ABG and Static (Input 3), Common Patients in Vitals, ABG and Static (IMF) (Input 4), All Patients in Vitals, ABG and Static (Input 5), All Patients in Vitals, ABG and Static (IMF) (Input 6).

Similarly to Table 6.2, Table 6.3 shows almost identical values for the parameters in all the models regarding number of layers, number of steps and validation split. As concern the number of hidden units and activation function, they change specifically for every input without any correlation. The fact that some parameters present the same value for all the models suggests that they are not affected by the size of the input or the prediction target, unlike activation function and hidden units.

The first model, that uses only the vital signs with urinary output and GCS, is the worst one while the others, that use also other variables and features, have better performances, even more when all the patients in the

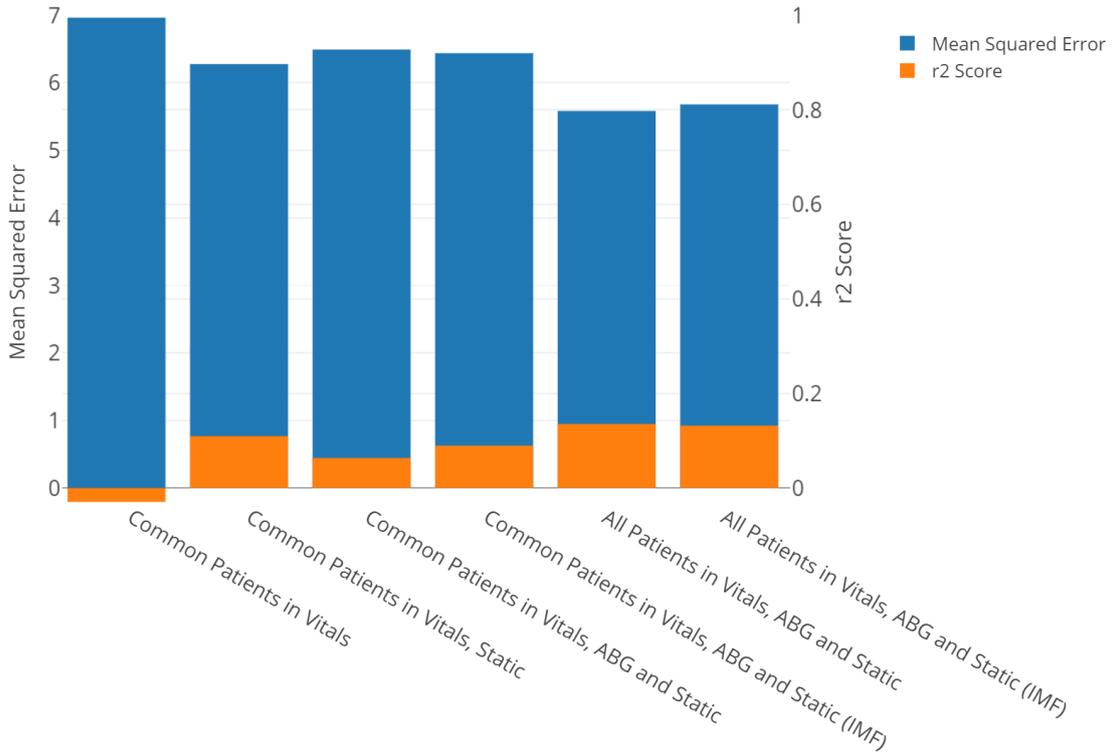


Figure 6.12: MSE and  $R^2$  for different input.

dataset are used for the training.

In this case, the order of the input features is not exactly the same as the one present in Figures 6.2 and 6.8: anyway no general conclusion about the impact of the features can be done observing only the first hour after admission. Also in this case, as well as for 90 days regression, higher values are related to an output pushed towards longer length of survival and vice versa.

While the contribute for the physiological variables is almost the same as the two tasks showed previously, the importance of the features extracted from the variables is slightly different. Indeed, even though six out of eight features extracted from the IMF are ranked as the worst, two of them results to have a big importance.

Finally, binary metrics can be extracted from the regression results for the different models (Figure 6.16).

In this case the metrics are evaluated in 1000 days and the trend of the precision recall and MCC are quite similar to the regression model on 90 days.

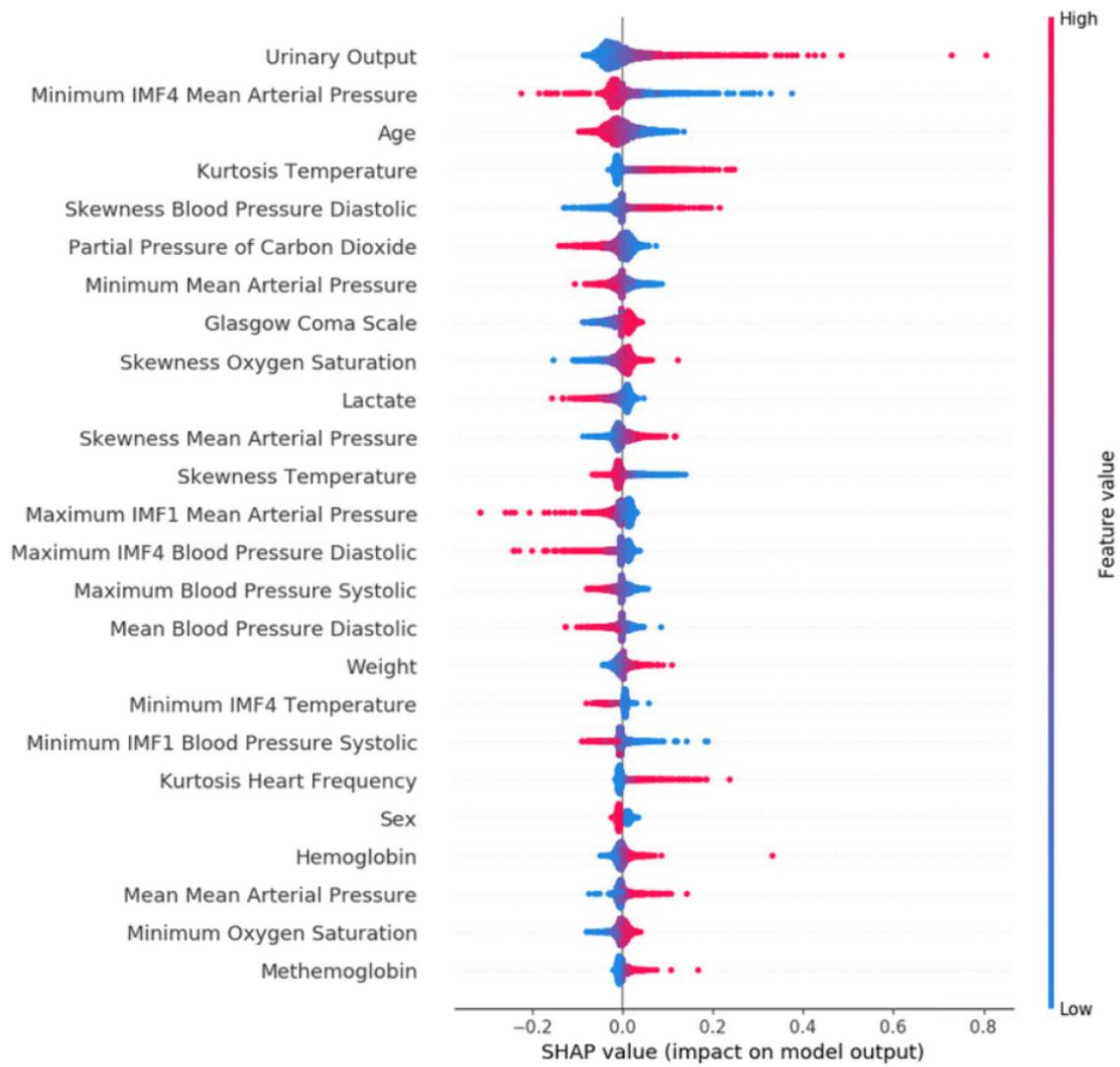


Figure 6.13: Features contribute for the first hour after the admission.

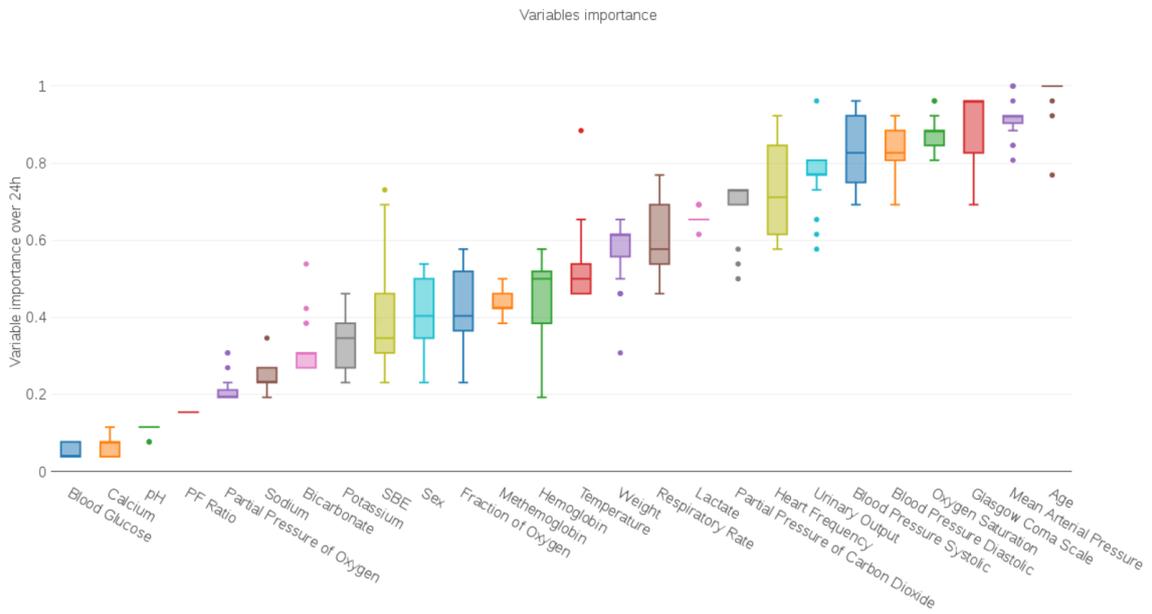


Figure 6.14: Variables importance over 24 hours.

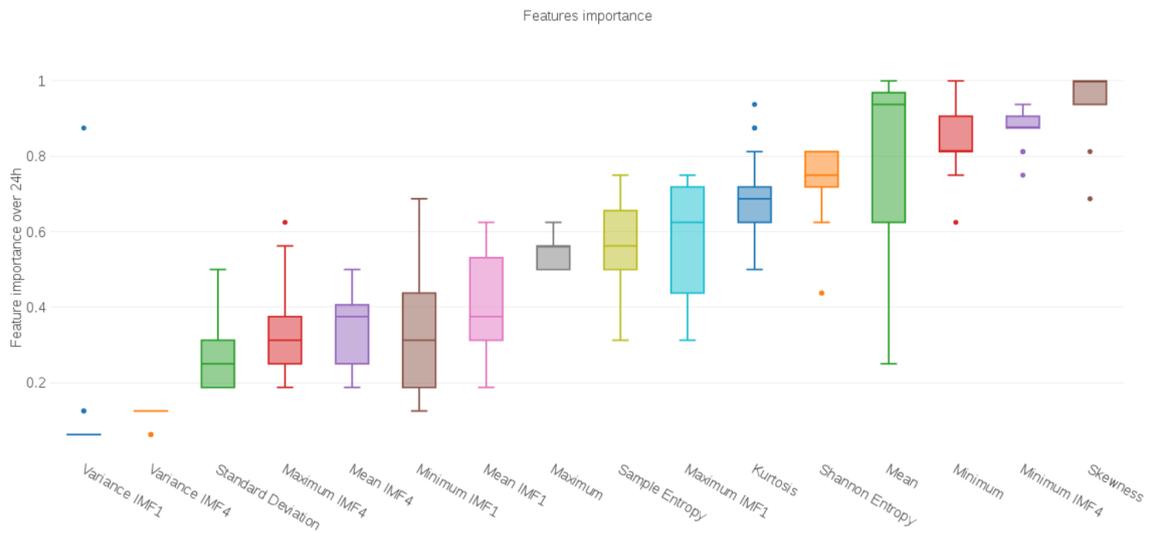


Figure 6.15: Features importance over 24 hours.

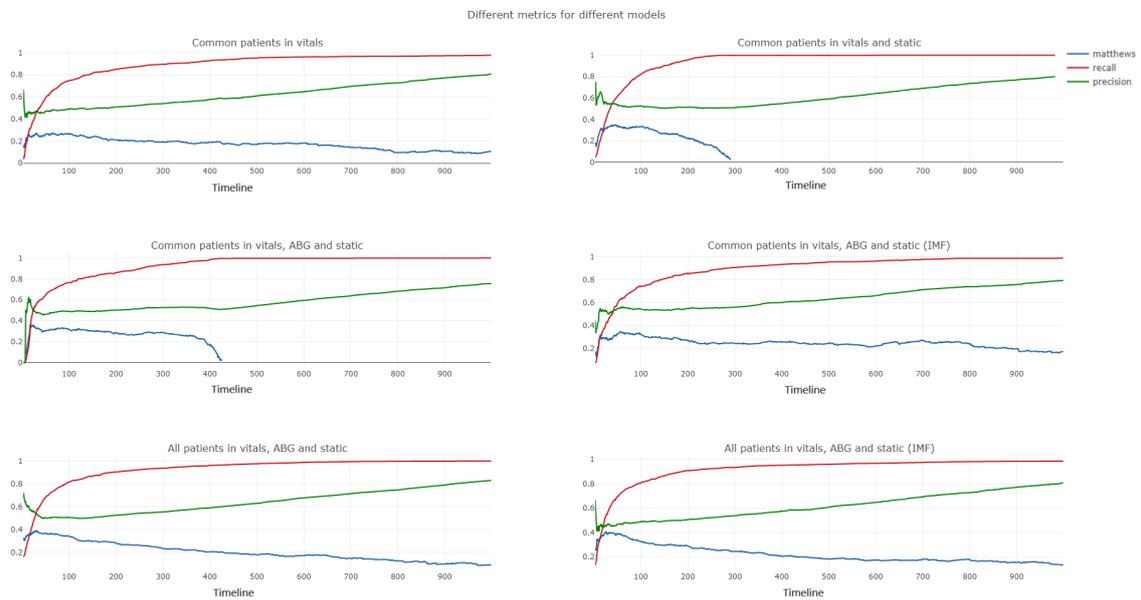


Figure 6.16: Different metrics for the input tested with regression including censored patients.

## 6.5 Cox proportional-hazards regression model

The cox model is generally a statistical model that can be used to understand how some covariates affect the probability that a particular event occurs at a specific time point. In this case the event of interest is death and the idea is to understand how the output of the model really affects the survival probability of the patient: for this reason the last layer of the LSTM network is used as input variable of the survival model; many studies have used deep learning models combined with a survival model [49] [50]. If the model has optimal performances, the output of the survival model is obvious: the survival probability would be close the 100% until the day the patient is predicted to die, then the survival probability would drop off around 0. If the output of this LSTM model is used as input of the survival model, the result will be a survival probability that decreases as fast as the number returned by the LSTM model is low.

The function used to estimate the survival probability is the hazard function  $h(t)$ , which can be interpreted as the probability that the event of interest occurs at the time  $t$ .

$$h_k(t) = h_0(t)e^{\sum_{i=1}^n \beta x} \quad (6.3)$$

In Equation 6.3,  $\beta$  is the weight associated to each covariate  $x$ , and  $h_0(t)$  is the baseline hazards where all the covariates are set to zero. The survival model can be trained with other static features besides the output of the LSTM model, which is not trained with information about other possible diseases the patient is affected.

The Cox-model is trained and tested on both the regression on 90 days and the regression with censored patients.

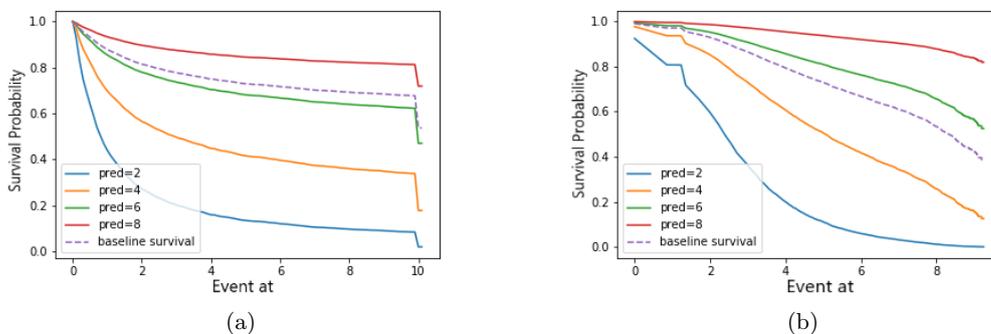


Figure 6.17: Cox model on two different regression tasks.

Figure 6.17a shows the Cox model trained with the output of the regression on 90 days: the x-axis range represents the length of survival as shown in Figure 5.2. When the LSTM output is 2 (blue line), the survival prediction is about 30% at 20 days and about 15% at 90 days. When LSTM output is higher, such as 8, the survival probability decreases much more slowly, indeed after 90 days it is about 90%. The same principle is shown in Figure 6.17b, where the output used is the LSTM model with the censored patients.

# Chapter 7

## Conclusion

The aim of the project was to exploit the huge amount of data acquired by the patients admitted in the ICU, and to extract possible outcomes about their future condition. In this work the attention was mainly focused on vital signs with high frequency data content, since the quantity of information that can be extracted is higher. Later, other variables were added to the vital signs to enrich the knowledge of the model. The same procedure was carried on for both the classification task and the regression tasks. Even though the performances do not outstand the current state of the art, the work might be still considered interesting for understanding the impact of some features and variables in the final prediction. Indeed the lack of interpretability of a model is one of the reasons why a tool might not be used by clinicians inside the hospitals: the comprehension of which are the elements that drive the prediction toward a specific outcome can be more interesting than the outcome itself. In this way the doctors have a tool able to provide information on which variables need to be monitored and where to take action.

# Bibliography

- [1] J. Wu, L. Yao, and B. Liu, «An overview on feature-based classification algorithms for multivariate time series», in *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Apr. 2018, pp. 32–38. DOI: [10.1109/ICCCBDA.2018.8386483](https://doi.org/10.1109/ICCCBDA.2018.8386483).
- [2] L. Ge and L.-J. Ge, «Feature extraction of time series classification based on multi-method integration», *Optik - International Journal for Light and Electron Optics*, vol. 127, no. 23, pp. 11 070–11 074, 2016, ISSN: 0030-4026. DOI: <https://doi.org/10.1016/j.ijleo.2016.08.089>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030402616309652>.
- [3] D. Mao, Y. Wang, and Q. Wu, «A new approach for physiological time series», *CoRR*, vol. abs/1504.06274, 2015. arXiv: [1504.06274](https://arxiv.org/abs/1504.06274). [Online]. Available: <http://arxiv.org/abs/1504.06274>.
- [4] I. Silva, G. Moody, D. J. Scott, L. A. Celi, and R. G. Mark, «Predicting in-hospital mortality of ICU patients: The physionet/computing in cardiology challenge 2012», in *2012 Computing in Cardiology*, Sep. 2012, pp. 245–248.
- [5] J. Liu, X. Chen, Z. Fang, K. Tong, L. Fang, and J. Li, «ICU mortality prediction using modified cost-sensitive pca and chaos pso», in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Oct. 2017, pp. 1–6. DOI: [10.1109/HealthCom.2017.8210761](https://doi.org/10.1109/HealthCom.2017.8210761).
- [6] Z. C. Lipton, D. C. Kale, C. Elkan, and R. C. Wetzell, «Learning to diagnose with LSTM recurrent neural networks», *CoRR*, vol. abs/1511.03677, 2015. arXiv: [1511.03677](https://arxiv.org/abs/1511.03677). [Online]. Available: <http://arxiv.org/abs/1511.03677>.

- [7] R. Viegas, C. M. Salgado, S. Curto, J. P. Carvalho, S. M. Vieira, and S. N. Finkelstein, «Daily prediction of icu readmissions using feature engineering and ensemble fuzzy modeling», eng, *Expert Systems With Applications*, vol. 79, pp. 244–253, 2017, ISSN: 18736793, 09574174. DOI: [10.1016/j.eswa.2017.02.036](https://doi.org/10.1016/j.eswa.2017.02.036).
- [8] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, «Long short term memory recurrent neural network based multimodal dimensional emotion recognition», in *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, ser. AVEC '15, Brisbane, Australia: ACM, 2015, pp. 65–72, ISBN: 978-1-4503-3743-4. DOI: [10.1145/2808196.2811634](https://doi.org/10.1145/2808196.2811634). [Online]. Available: <http://doi.acm.org.proxy.findit.dtu.dk/10.1145/2808196.2811634>.
- [9] N. Tajbakhsh and K. Suzuki, «Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs», *Pattern Recognition*, vol. 63, pp. 476–486, 2017, ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2016.09.029>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320316302795>.
- [10] J. C. B. Gamboa, «Deep learning for time-series analysis», *CoRR*, vol. abs/1701.01887, 2017. arXiv: [1701.01887](https://arxiv.org/abs/1701.01887). [Online]. Available: <http://arxiv.org/abs/1701.01887>.
- [11] Z. Wang, W. Yan, and T. Oates, «Time series classification from scratch with deep neural networks: A strong baseline», *CoRR*, vol. abs/1611.06455, 2016. arXiv: [1611.06455](https://arxiv.org/abs/1611.06455). [Online]. Available: <http://arxiv.org/abs/1611.06455>.
- [12] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, «A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series», *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 758–769, Apr. 2018, ISSN: 1534-4320. DOI: [10.1109/TNSRE.2018.2813138](https://doi.org/10.1109/TNSRE.2018.2813138).
- [13] F. Karim, S. Majumdar, H. Darabi, and S. Chen, «LSTM fully convolutional networks for time series classification», *IEEE Access*, vol. 6, pp. 1662–1669, 2018, ISSN: 2169-3536. DOI: [10.1109/ACCESS.2017.2779939](https://doi.org/10.1109/ACCESS.2017.2779939).

- [14] P. Veličković, L. Karazija, N. D. Lane, S. Bhattacharya, E. Liberis, P. Liò, A. Chieh, O. Bellahsen, and M. Vegreville, «Cross-modal recurrent models for weight objective prediction from multimodal time-series data», in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth '18, New York, NY, USA: ACM, 2018, pp. 178–186, ISBN: 978-1-4503-6450-8. DOI: [10.1145/3240925.3240937](https://doi.org/10.1145/3240925.3240937). [Online]. Available: <http://doi.acm.org/10.1145/3240925.3240937>.
- [15] Z. Che, S. Purushotham, R. G. Khemani, and Y. Liu, «Interpretable deep models for ICU outcome prediction», *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2016, pp. 371–380, 2016.
- [16] M. Moor, M. Horn, B. Rieck, D. Roqueiro, and K. Borgwardt, «Temporal convolutional networks and dynamic time warping can drastically improve the early prediction of sepsis», *ArXiv preprint arXiv:1902.01659*, 2019.
- [17] J. Smith and R. Roberts, *Vital signs for nurses: An introduction to clinical observations*. John Wiley & Sons, 2011. DOI: [10.1002/9781119139119.ch7](https://doi.org/10.1002/9781119139119.ch7).
- [18] F. AL-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, «Respiration rate monitoring methods: A review», *Pediatric Pulmonology*, vol. 46, no. 6, pp. 523–529, DOI: [10.1002/ppul.21416](https://doi.org/10.1002/ppul.21416). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ppul.21416>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ppul.21416>.
- [19] E. D. Chan, M. M. Chan, and M. M. Chan, «Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations», *Respiratory Medicine*, vol. 107, no. 6, pp. 789–799, 2013, ISSN: 0954-6111. DOI: <https://doi.org/10.1016/j.rmed.2013.02.004>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S095461111300053X>.
- [20] J. Betts, P. Desaix, J. Johnson, O. Korol, D. Kruse, B. Poe, J. Wise, M. Womble, and K. Young, *Anatomy & physiology*, ser. Open Textbooks. OpenStax College, Rice University, 2013, ISBN: 9781938168130. [Online]. Available: <https://books.google.dk/books?id=jD00ngEACAAJ>.
- [21] H. L. Li-wei, M. Saeed, D. Talmor, R. Mark, and A. Malhotra, «Methods of blood pressure measurement in the ICU», *Critical care medicine*, vol. 41, no. 1, p. 34, 2013.

- [22] R. Jaafar, H. M. Desa, Z. Mahmoodin, M. R. Abdullah, and Z. Zaharudin, «Noninvasive blood pressure (NIBP) measurement by oscillometric principle», in *2011 2nd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering*, Nov. 2011, pp. 265–269. DOI: [10.1109/ICICI-BME.2011.6108622](https://doi.org/10.1109/ICICI-BME.2011.6108622).
- [23] R. S. Hoke, U. Müller-Werdan, C. Lautenschläger, K. Werdan, and H. Ebel, «Heart rate as an independent risk factor in patients with multiple organ dysfunction: A prospective, observational study», *Clinical Research in Cardiology*, vol. 101, no. 2, pp. 139–147, Feb. 2012, ISSN: 1861-0692. DOI: [10.1007/s00392-011-0375-3](https://doi.org/10.1007/s00392-011-0375-3). [Online]. Available: <https://doi.org/10.1007/s00392-011-0375-3>.
- [24] D. I. Sessler, «Temperature monitoring and perioperative thermoregulation», *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 109, no. 2, pp. 318–338, 2008.
- [25] M. R. Neuman, «Measurement of vital signs: Temperature [tutorial]», *IEEE Pulse*, vol. 1, no. 2, pp. 40–49, Sep. 2010, ISSN: 2154-2287. DOI: [10.1109/MPUL.2010.937907](https://doi.org/10.1109/MPUL.2010.937907).
- [26] M. Legrand and D. Payen, «Understanding urine output in critically ill patients», *Annals of Intensive Care*, vol. 1, no. 1, p. 13, May 2011, ISSN: 2110-5820. DOI: [10.1186/2110-5820-1-13](https://doi.org/10.1186/2110-5820-1-13). [Online]. Available: <https://doi.org/10.1186/2110-5820-1-13>.
- [27] R. Bellomo, C. Ronco, J. A. Kellum, R. L. Mehta, P. Palevsky, and the ADQI workgroup, «Acute renal failure – definition, outcome measures, animal models, fluid therapy and information technology needs: The second international consensus conference of the acute dialysis quality initiative (adqi) group», *Critical Care*, vol. 8, no. 4, R204, May 2004, ISSN: 1364-8535. DOI: [10.1186/cc2872](https://doi.org/10.1186/cc2872). [Online]. Available: <https://doi.org/10.1186/cc2872>.
- [28] G. Teasdale and B. Jennett, «Assessment of coma and impaired consciousness: A practical scale», *The Lancet*, vol. 304, no. 7872, pp. 81–84, 1974.
- [29] H. M. Mohammed and D. A. Abdelatif, «Easy blood gas analysis: Implications for nursing», *Egyptian Journal of Chest Diseases and Tuberculosis*, vol. 65, no. 1, pp. 369–376, 2016, ISSN: 0422-7638. DOI:

- <https://doi.org/10.1016/j.ejcdt.2015.11.009>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0422763815301175>.
- [30] H. T. Le, N. S. Harris, A. J. Estilong, A. Olson, and M. J. Rice, «Blood glucose measurement in the intensive care unit: What is the best method?», *Journal of diabetes science and technology*, vol. 7, no. 2, pp. 489–499, 2013.
- [31] S. K. Aberegg, «Ionized calcium in the icu: Should it be measured and corrected?», *Chest*, vol. 149, no. 3, pp. 846–855, 2016, ISSN: 0012-3692. DOI: <https://doi.org/10.1016/j.chest.2015.12.001>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0012369215002937>.
- [32] S. Gupta, A. K. Gupta, K. Singh, and M. Verma, «Are sodium and potassium results on arterial blood gas analyzer equivalent to those on electrolyte analyzer?», *Indian journal of critical care medicine: Peer-reviewed, official publication of Indian Society of Critical Care Medicine*, vol. 20, no. 4, p. 233, 2016.
- [33] J. Szrama and P. Smuszkiewicz, «An acid-base disorders analysis with the use of the stewart approach in patients with sepsis treated in an intensive care unit», *Anaesthesiology Intensive Therapy*, vol. 48, no. 3, pp. 180–184, 2016, ISSN: 1642-5758. DOI: [10.5603/AIT.a2016.0020](https://doi.org/10.5603/AIT.a2016.0020). [Online]. Available: [https://journals.viamedica.pl/anaesthesiology\\_intensivetherapy/article/view/AIT.a2016.0020](https://journals.viamedica.pl/anaesthesiology_intensivetherapy/article/view/AIT.a2016.0020).
- [34] S. Hochreiter and J. Schmidhuber, «Long short-term memory», *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735). eprint: <https://doi.org/10.1162/neco.1997.9.8.1735>. [Online]. Available: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [35] C. Olah, *Understanding lstm networks*, 2015. [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [36] S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber, *et al.*, *Gradient flow in recurrent nets: The difficulty of learning long-term dependencies*, 2001.
- [37] Y. Bengio, P. Simard, and P. Frasconi, «Learning long-term dependencies with gradient descent is difficult», *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, Mar. 1994, ISSN: 1045-9227. DOI: [10.1109/72.279181](https://doi.org/10.1109/72.279181).

- [38] J. S. Richman and J. R. Moorman, «Physiological time-series analysis using approximate entropy and sample entropy», *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, H2039–H2049, 2000, PMID: 10843903. DOI: [10.1152/ajpheart.2000.278.6.H2039](https://doi.org/10.1152/ajpheart.2000.278.6.H2039). eprint: <https://doi.org/10.1152/ajpheart.2000.278.6.H2039>. [Online]. Available: <https://doi.org/10.1152/ajpheart.2000.278.6.H2039>.
- [39] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, «The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis», *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998. DOI: [10.1098/rspa.1998.0193](https://doi.org/10.1098/rspa.1998.0193). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1998.0193>. [Online]. Available: <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1998.0193>.
- [40] R. Djemili, H. Bourouba, and M. A. Korba, «Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic eeg signals», *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 285–291, 2016, ISSN: 0208-5216. DOI: <https://doi.org/10.1016/j.bbe.2015.10.006>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0208521615000777>.
- [41] A. Rojas, J. M. Górriz, J. Ramírez, A. Gallix, and I. A. Illán, «Empirical mode decomposition as a feature extraction method for alzheimer’s disease diagnosis», in *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, Oct. 2012, pp. 3909–3913. DOI: [10.1109/NSSMIC.2012.6551897](https://doi.org/10.1109/NSSMIC.2012.6551897).
- [42] P. F. Diez, V. Mut, E. Laciari, A. Torres, and E. Avila, «Application of the empirical mode decomposition to the extraction of features from eeg signals for mental task classification», in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep. 2009, pp. 2579–2582. DOI: [10.1109/IEMBS.2009.5335278](https://doi.org/10.1109/IEMBS.2009.5335278).
- [43] I. Štajduhar and B. Dalbelo-Bašić, «Learning bayesian networks from survival data using weighting censored instances», *Journal of Biomedical Informatics*, vol. 43, no. 4, pp. 613–622, 2010, ISSN: 1532-0464. DOI: <https://doi.org/10.1016/j.jbi.2010.03.005>. [Online].

- Available: <http://www.sciencedirect.com/science/article/pii/S1532046410000389>.
- [44] E. L. Kaplan and P. Meier, «Nonparametric estimation from incomplete observations», *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.
- [45] B. Matthews, «Comparison of the predicted and observed secondary structure of t4 phage lysozyme», *Biochimica et Biophysica Acta (BBA) - Protein Structure*, vol. 405, no. 2, pp. 442–451, 1975, ISSN: 0005-2795. DOI: [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0005279575901099>.
- [46] S. Boughorbel, F. Jarray, and M. El-Anbari, «Optimal classifier for imbalanced data using matthews correlation coefficient metric», *PloS one*, vol. 12, no. 6, e0177678, 2017.
- [47] S. M. Lundberg and S.-I. Lee, «A unified approach to interpreting model predictions», in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- [48] L. H. Lehman, S. Nemati, R. P. Adams, G. Moody, A. Malhotra, and R. G. Mark, «Tracking progression of patient state of health in critical care using inferred shared dynamics in physiological time series», in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2013, pp. 7072–7075. DOI: [10.1109/EMBC.2013.6611187](https://doi.org/10.1109/EMBC.2013.6611187).
- [49] T. Ching, X. Zhu, and L. X. Garmire, «Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data», *PLoS computational biology*, vol. 14, no. 4, e1006076, 2018.
- [50] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, «Deepsurv: Personalized treatment recommender system using a cox proportional hazards deep neural network», *BMC medical research methodology*, vol. 18, no. 1, p. 24, 2018.