Fostering the use of Electronic Health Record data for better clinical management and patient safety: Methodology for the evaluation of early warning scores to predict in hospital sudden cardiac arrest

Joana Gonçalves[1,*]

Thesis to obtain the Master of Science Degree in Biomedical Engineering – November 2017

Supervisors: Prof. Dr. Mónica Oliveira[1] and Engineer João Leal[2]

[1]Instituto Superior Técnico, University of Lisbon, Portugal; [2]José de Mello Saúde, Department of Clinical Information and Business Intelligence *Email: joana.a.goncalves@ist.utl.pt

ABSTRACT: Early Warning Scores (EWS) are weighted scoring systems to activate emergency teams or direct patients to an Intensive Care Unit based on deterioration of vital signs or other physiological parameters. Some patients that have an in-hospital Sudden Cardiac Arrest (SCA) are known to deteriorate hours before the event. The National EWS (NEWS) was recently implemented in two hospitals managed by José de Melo Saúde however the applicability of NEWS on such context was not evaluated nor tested against other systems to predict SCA. Therefore, this thesis will start by identifying evidence based methods to predict in-hospital SCA, then assess completeness of EHR data based on EWS parameters and finally compare the performance of NEWS to other relevant EWSs using different methods to evaluate binary classifiers and interpreting its meaning for unbalanced data. The overall objective was to improve methods to evaluate and choose an EWS to predict in-hospital SCA. The proposed approach was applied to a database with 26367 patients and 39340 episodes observations from the Hospital CUF Infante Santo, 9 of which had a SCA. To compare NEWS against Cardiac Arrest Risk Triage on performance of prediction of SCA, Receiver Operating Characteristic (ROC), Precision Recall and other metrics were used. Results showed that almost all parameters were recorded in less than 60% of the episodes and that performance of EWSs should be determined for specific thresholds and tolerated tradeoffs, using metrics that take into account prevalence, instead of misleading commonly used summary metrics like area under the ROC.

KEYWORDS: Early Warning Scores; Sudden Cardiac Arrest; Electronic Health Records (EHR); Patient Safety; Binary Decision Problem; Unbalanced Data.

1. INTRODUCTION

In-hospital sudden cardiac arrest (SCA) survival rates are higher than out-of-hospital SCA survival rates with but still require an improvement [1], [2]. Is known that many in-hospital SCA are preceded by a progressive deterioration of the patient's condition, which can be translated as an abnormal change in vital sign values or other parameters [3]–[5]. Currently, there are several risk assessment tools published to predict SCA, unanticipated ICU admission and death that rely on a scoring system to activate emergency teams or direct patients to an ICU, based on deterioration signs, that are known as Early Warning Scores (EWS) [6]. Aggregate Weighted Scoring Systems (AWSS), such as EWS, are said to have a better discriminating ability that single parameter and multiparameter risk assessment tools [7]–[9].

Although studies have been conducted to compare the accuracy of these systems there is neither a defined "gold standard" system nor a unique method to use when validating an EWS system [10]. Additionally, the method that has been used by de majority of studies to compare EWSs, Area Under the Receiver Operating Characteristic (ROC), has been declared as poorly informative for skewed datasets, which happens when the prevalence of the outcome that is being predicted is very low. Studies suggest that Precision-Recall (PR) is a more informative plot for skewed datasets [11]. José de Mello Saúde (JMS) is a reference private operator of healthcare in Portugal established in 1945 and currently manages a network of healthcare units throughout the country. An EWS and respective plan of action was recently implemented in two hospitals managed by JMS: Hospital CUF Torres Vedras and Hospital CUF Porto [12]. NEWS is a well-known tool to predict deterioration and risk of an adverse outcome, such as SCA, and is used in the English National Healthcare System [10]. However, is not known how the system will perform in a Portuguese private healthcare provider scenario and neither if exists an EWS system that better adapts to this clinical setting.

The purposes of this study are to (1) define evidence based and good practice methods to predict in-hospital SCA, (2) assess completeness of data in the EHR necessary to calculate risk scores for SCA and (3) make a comparison between the recently introduced tool to predict deterioration and other potential better risk assessment tools (CART), using different methods to evaluate binary classifiers, with special focus on unbalanced data. The overall objective is to improve methods to evaluate and choose an EWS tools for in-hospital pre-SCA.

This study contributes to current literature in comparison between EWS systems because provides a comparison between two EWS, NEWS and CART, that have never been tested against each other, specifically for predicting SCA. It is also the first study, from the reviewed articles, that used Precision-Recall (PR) curves as one of the metrics to compare EWS and not only Receiver Operating Characteristic (ROC) curves, as well as other metrics. Furthermore, this is also the first study to apply retrospectively EWS scoring to a dataset obtained from a Portuguese private healthcare setting.

2. LITERATURE REVIEW

2.1. Risk Assessment Tools for SCA

Risk management in healthcare arose as a strategy to improve of quality of care and improve patient safety (Considine and Botti, 2004). Given the possible outcomes of a SCA and resource constraints it would be useful to stratify patients by risk of SCA.

Currently, there are more than 100 studies already published on scoring system to activate emergency teams or direct patients to an ICU based on deterioration signs [6]. The ideal trigger system should provide a maximum discrimination of patients' outcomes at a lowest trigger rate, therefore reducing the probability of missing any adverse event without an excessive workload and unnecessary call of emergency teams [13], [14].

When developing a AWSS the usual outcomes that one wants to predict are death, ICU transfer or SCA, being the last one is less predictable (can happen without any previous physiological alteration) [15].

With the previous knowledge that AWSS have a better discriminating ability that single parameter and multiparameter risk assessment tools [8], [7], [9] and given that NEWS, the system that is going to be implemented, is categorized as an AWSS the focus of these review is going to be in this kind of systems and how are their performance being compared. The search was made by key words in the scientific database PubMed.

The AWSSs with more references in literature were MEWS [17], ViEWS [18], NEWS [10] and CART [16]. Remaining articles had only one reference in the literature review. After analyzing outcomes that were being predicted (death, ICU transfer or SCA), the size of the sample, ratio between patients/episodes that suffered and did not suffered a SCA, methods that were used to assess performance/make a comparison and results it was decided that the EWS in study would be CART. This because CART is an AWSS developed to predict specifically SCA, could be calculated using the available data and that proved to be better than SEWS, ViEWS and MEWS at predicting SCA in some studies [6], [16]. NEWS and CART development and characteristics will be further explained, as well as an overview of all tools.

<u>NEWS</u>

NEWS was developed by the members of the RCPL National Early Warning Score Design and Implementation Group by modificating ViEWS based on clinical opinion, not having a computer modeling behind [10], [22]. It is actually used by the English National Health Care System and is recommended to be applied to all patients admitted to an acute hospital. To develop NEWS the first step was reviewing other existing EWS systems and related literature, making an initial draft. After discussing which parameters should be included the group decided on six of them, taking into account practicality of the measurement of each parameter. Similar to the previous AWSS, this system allocates a score to each physiological measurement made to patients inside the hospital based on how far the measurement is from the normal range, which is translated in a color-coded table (**Figure 1**).

The overall NEWS score for a patient results from the sum of all individual parameter score and its value will determine the recommended clinical response, that is divided in low risk (1-4), medium risk (5-6) and high risk (≥7), each creating a specific response such as increasing monitoring or calling emergency teams.

PHYSIOLOGICAL PARAMETERS	3	2	1	0	1	2	3
Respiration Rate	≤8		9 - 11	12 - 20		21 - 24	≥25
Oxygen Saturations	≤91	92 - 93	94 - 95	≥96			
Any Supplemental Oxygen		Yes		No			
Temperature	≤35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥39.1	
Systolic BP	≤90	91 - 100	101 - 110	111 - 219			≥220
Heart Rate	≤40		41 - 50	51 - 90	91 - 110	111 - 130	≥131
Level of Consciousness				A			V, P, or U

Figure 1: National Early Warning Score (NEWS). The table was taken directly from source to maintain the color-codes recommended and not only the ranges and scores for each parameter [10].

<u>CART</u>

Cardiac Arrest Risk Triage score was developed by Churpek et al. in 2012 with the purpose of predicting SCA using vital signs and age. It was calculated using regression coefficients obtained from a database of patients, some of which suffered SCA, making the model contain cut-off points with the same sensitivity and specificity as the MEWS at cut-off value higher than 4. Similarly, to the remaining EWS the overall score for CART is calculated by adding together the scores associated with each physiological parameter, which can range from 0 to 22 (**Figure 2**). There was not any proposed strategy to apply this risk assessment tool in a real-live set and no threshold values with respective clinical response were recommended [6].

Figure 2: CART score [23].

	0	4	6	8	9	12	13	15	22
DBP (mmHg)	>49	40-49	35-39			<35			
HR (bpm)	<110	110-139					>139		
RR (bpm)	<21			21-23		24-25		26-29	>29
Age	<55	55-59			>69				

<u>Overall EWS</u>

The most used method for making a comparison between risk assessment was AUROC, even if almost all samples had imbalanced data (high ratio between number of patients/episodes/admissions and number of SCA). Additional comparison methods were efficiency curve, sensibility, specificity, PPV and NPV. In all the studies where the value of AUROC for SCA was calculated for a certain model, this value was always lower that the AUROCs calculated for the remaining outcomes, for the same model.

Comparison between traditional methods, such as logistic regression, and machine learning methods showed that the latter can be more accurate for studying and implementing real-time prediction tools and should be further explored [21].

2.2. Clinical Practice Guidelines

According to the definition proposed by Field and Lohr in 1990, clinical practice guidelines "are systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" [24]. Some of the recommendations included in these guidelines depend on risk calculations to make decisions (North, Fox and Chaudhry, 2016). There is a great number of clinical practice guidelines than have been addressed by different societies and organizations on the matter of SCA. In this review the focus is going to be the latest guidelines on the matter developed by the European Society of Cardiology (ESC), American Heart Association (AHA) and European Resuscitation Council (ERC), studying the topics related with prediction of SCA in each one (**Figure 3**).

2.3. Methods for evaluation of binary decision problems

In a binary decision problem the classifier labels an observation in a given set as positive or negative [27]. If the data used in the problem is imbalanced, the minority class is labeled as positive and the majority class is labeled as negative. The classifiers of a binary decision problem are evaluated by a square matrix 2x2 (**Figure 4**) where the columns represent the hypothesized class and the rows the true classes [28]

- $True\ Positive\ Rate\ (TPR) = \frac{TP}{P} = Recall = Sensitivity$
- $Positive\ Predictive\ Value\ (PPV\ ) = \frac{TP}{TP+FP} = Precision$
- $True\ Negative\ Rate\ (TNR) = \frac{TN}{FP+TN} = Specificity$
- $Trigger\ rate = \frac{TP+FP}{TP+TN+FP+FN}$

		True Class p	True Class n
Hypothesized Class	Y	True Positives (TP)	False Positives (FP)
	N	False Negatives (FN)	True Negatives (TN)
	Column Total	P	N

Figure 4: Confusion Matrix [29].

ROC Curve

ROC curve analysis is commonly used in medical decision making. Moreover, it gained great importance in evaluating the power of a certain diagnostic test to identify the true patient's state, find the optimal threshold for that test and compare alternative diagnostic tests between them, for the same conditions [30]. It can also be applied to risk assessment tools, by evaluating the ability of the tool to discriminate between individuals who will experience/not experience a certain event.

Each point in a ROC plot is a representation of (1-specificity) as a function of sensitivity, or FPR ,for a certain threshold value, which means that for each threshold there is a different confusion matrix built. [28] The diagonal line represents a model in which the class is randomly guessed, being better to have the curve closer to the superior left corner [29].

The Area Under the ROC (AUROC) is one of the indicators of ROC curve performance and summarizes the performance of a classifier for all possible thresholds, which makes the comparison between models' accuracy easier [28]. Since AUROC is always a percentage of the area of a unit square its value ranges between 0 and 1.

A data set is considered imbalanced if one of classes (minority class) contains a much smaller number of observations than the other class (majority classes). The problem is that the minority class is usually the main interest in models. Having a skewed class distribution can cause a misinterpretation of evaluation measures such as ROC [28]. ROC curves are not sensitive to the ratio of positives and negatives in a distribution (**Figure 5**).

Considering our problem, the predictive performance values will depend on the prevalence of the outcome that

		Importance of predicting SCA	Risk assessment of SCA by disease	Existence of EWS	Response to EWS by emergency teams	Specific EWS tools
2015 ERC Guidelines for Resuscitation	Section 3. Adult advanced life support.	•		•	•	•
	Section 4. Cardiac arrest in special circumstances		•			
	Section 10. Education and implementation of resuscitation	•		•	•	
2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death		•	•			
2015 AHA Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care	Part 4: Systems of Care and Continuous Quality Improvement	•		•	•	•
	Part 14: Education	•		•	•	

Figure 3: Topics mentioned on each chapter of each clinical practice guideline on SCA and risk assessment. Executive summaries, paediatric and neonatal related chapters were excluded. The remaining chapters that did not contain any of the relevant topics were omitted.

is being predicted in the study population, even if the pair sensitivity/1-specificity remain the same [31].

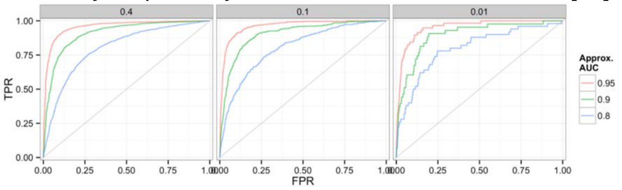

Figure 5: ROC curves for models tested with 3 different data sets with ratios positive/negative varying from 0.4, 0.1 to 0.01 [32].

<u>PR Curve</u>

The precision-recall (PC) plot is another visual representation method to evaluate the accuracy of a model and represents the tradeoff between precision and recall for each different threshold.

As in ROC, the area under the precision-recall curve (AUPRC) represents a summary of the information given by the PR curve, allowing the comparison between models [34]. The difference is that the baseline for ROC is fixed, while in PR the baseline is determined by the ratio of positives and negatives present in sample $P = \frac{P}{P+N}$.

To overcome this issue of unbalanced data, one can compare FP to the total number of TP given by a model for a given threshold and see the effect of having a large number of negative observations un the performance of the model by plotting a PC curve (**Figure 6**) [27], [32].

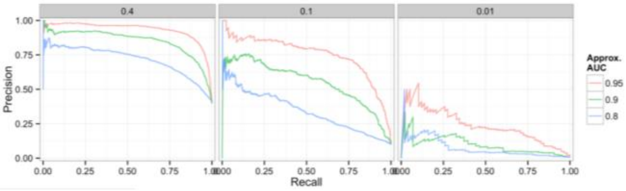

Figure 6: PR curves for models tested with 3 different data sets with ratios positive/negative varying from 0.4, 0.1 to 0.01 [32].

3. METHODOLOGICAL APPROACH

3.1. Evidence based and good practice methods to predict in-hospital SCA

The first step to evaluate if a certain risk assessment tool is the ideal to be implemented is to identify which other tools exist, and how are they being tested. To do that it was necessary to analyze three distinct sources: clinical guidelines, clinical practice and literature review.

<u>Literature Review</u>

Analyzing guidelines on SCA and Resuscitation allow the identification of the recommendations made by several organizations on the adoption of EWS systems, which type of EWS should be used, what should be the monitoring for different patients in different areas, which factors (such as comorbidities and patient characteristics) are predictors of a SCA, among others.

Literature review on EWS systems allow the identification of EWSs that could potential have better performance that NEWS in the context in study and that could be calculated with the parameters that Gllint EHR system had recorded, as well as comparison methods between such systems.

The literature review on developed EWS, comparison of performance of several AWSS and of risk-assessment of SDA recommendations on clinical practice guidelines was already conducted in the previews section.

<u>Semi-structured interviews</u>

In order to gain a better understanding about data acquisition methods in practice, clinical practice guidelines and needs/ challenges related with in-hospital SCA in its three stages (pre-SCA, inter-SCA and post-SCA) semi-structured interviews to the previews healthcare professionals were designed and conducted: Nurse Rita Rego (Manager of the department of Risk Management and Patient Safety of JMS) and Dr. Lídia Sousa (Cardiologist working at JMS health care facilities).

<u>Surveys on in-hospital monitoring and risk assessment tools for SCA</u>

A survey was conducted in order to characterize level and characteristics of monitoring by hospital area and characterize knowledge on timely identification of signs of deterioration of the patient, with special focus on risk assessment tools.

The survey was developed using a web-based platform for collecting information, Typeform This platform has the option of downloading the resulting data as an Excel or CSV file or analyze directly data with a tool that generates automatic reports [35]. For the present analyses the reports will be processed using Excel. The survey can be accessed and completed by consulting the web page <u>https://joahag.typeform.com/to/T8wDNw</u>.

The target population for this survey were doctors and nurses from different hospital units that work in two hospitals managed by JML: Hospital CUF Infante Santo and Hospital CUF Descobertas. The objective was to get answers from at least one doctor and nurse working at three different units inside the hospital which had different monitoring levels and strategies.

3.2. Assessment of vital sign data completeness

Failure to identify patients deteriorating can be related to incomplete vital sign measurements (Stevenson et al., 2016). In order to calculate a score using risk assessment tools is necessary to have access to a patient updated history, information and vital signs , which should be present in the EHR [36]. Given that one of the objectives of this thesis is to compare the performance of two different EWSs with a provided data set is important to examine the documentation of vital signs in the EHRs of patients, with special attention to those who subsequently suffered a SCA. The completeness of vital signs documentation will be determined with respect to NEWS and CART.

<u>Data source and extraction</u>

This retrospective study was based in anonymized data acquired from the Glintt EHR nursing system, from another hospital managed by JMS, hospital CUF Infante Santo, and contains all registers from March 2016 to May 2017 with measured parameters, not discriminated by hospital unit. Hospital CUF Infante Santo has a total of 145 beds (single rooms and wards), an ICU, surgical block with 9 rooms and 70 consulting rooms for the different medical specialties [37].

The data on the patients that had suffered a SCA in that time period in Hospital CUF Infante Santo was made by doing a search by ICD-9 codes in diagnosis related groups. It was only possible to access data from controlled environments (e.g. ICU) and not from other hospital units, such as permanent service and emergency. In order to retrieve the codes related with SCA the ICD online search tool was used, browsing the string “sudden cardiac arrest” and excluding results without relevancy for the subject [39].

<u>Data processing and filtering</u>

The data was exported as two CSV files. One of the CSV files contained patient identification number, birthday date, name of the parameter that was measure, code of the parameter that was measured, number of the episode (each patient could have had several episodes in the given timeframe), the date on each the parameter was measured, the hour at which the parameter was measured and the value of that parameter. The second CSV contained the patient identification number, patient age, patient sex, name of the parameter that was measure, code of the parameter that was measured, number of the episode and other additionally information that was not used, only for patients that had suffered a SCA. All processing and analysis was performed with RStudio, version 1.0.153 software package [40].

Given that NEWS is only applicable to people older than or 16, and that there was no information about age restrictions for CART, all patients younger than 16 were excluded from the sample [10]. The parameters with interest for calculating NEWS and CART were HR, Temperature, GCS, SBP, DBP, RR SO2 and SO2 (Therapeutic).

3.3. Comparison between implemented and literature EWS tools

Although there are several comparison articles on risk assessment tools (some tools were even validated by comparison with previous tools) no comparison study between NEWS and CART score were found when reviewing literature, for neither outcome (SCA, ICU admission or death). Additionally, the tool that is most used in literature to evaluate EWS, ROC, provides poor information when dealing with unbalanced data, which is the present situation. It is recommended in literature that PR should be used instead.

The purpose of this section was to compare the accuracy of NEWS and CART score to predict SCA using different methodologies and the same database.

Calculate NEWS and CART scores

To do a retrospective calculus of NEWS and CART two functions were developed and implemented using RStudio version 1.0.153 software package [40] and the data that was previously processed. Additionally, all the episodes for which there was not at least one value to calculate NEWS or CART score were excluded from the sample.

The parameters required to measure CART are in **Figure 2** and for NEWS in **Figure 1**.

If there were no previous values for a measurement in an episode, then a median value was imputed, which means that that parameter would contribute with a score of 0 for the overall score. This strategy was used in other articles related with EWS [16].

Given that the measurements in the dataset provided were not acquired systematically with the same time period between them and that different episodes may have different parameters measured with different frequencies, differently from previous studies [13], the functions that calculated NEWS and CART score were designed to pair measurements by how close they were in time in order to add the scores associated with those measurements. For the purpose of this thesis only the maximum value among sum of those values was selected, which means that in the end each episode would be paired with only one NEWS score and one CART score, that would be the maximum value of all the overall scores calculated for that episode. It is guaranteed that the number of episodes for which is possible to calculate NEWS is the same as CART, which allows a proper comparison between them.

After calculating NEWS and CART score for all the episodes in the data set it was necessary to know in which of those episodes a SCA had happen. To do that the episodes from the dataset that contain information on patient who suffered SCA were matched with the episodes for which NEWS and CART score were calculated.

Statistical analysis

In order to calculate ROC and PR plots, and respective areas under the curve, for NEWS and CART score, as well as the respective areas under the curve, the ROCR package was used. ROCR is a package developed to visualize and evaluate the performance of scoring classifiers in R language. The tools allow the graphic visualization of the several cut-off points in both ROC and PR plots, as well as knowing easily the pairs of points used to construct each curve [42]. All analysis was again performed using RStudio, version 1.0.153 software package [40].

Given that NEWS has already well-defined cut-off points to take action a careful analysis is going to be made for the thresholds for medium (score higher than 5) and high risk (score higher than 7), specifically comparing how sensitivity, specificity and precision are affected in comparison with several cut-off points for CART score, which does not have proposed cut-off values. Additionally, the number of episodes that needs to be further evaluated (NNE) to detect an outcome of SCA and the trigger rate, which is going to be represented as the number of triggers per number of monitored episodes, are also going to be determined [33].

4. RESULTS AND ANALYSIS

4.1. Semi-structured interviews

The semi-structured interviews were logically divided in three sections, which corresponded with the three stages of in-hospital SCA: pre-SCA, inter-SCA and post-SCA.

It was agreed in both interviews that the places in the hospital where is more frequent to happen a SCA are the ICU and the ER, although it can happen everywhere. In an ICU there is the possibility of having a continuous monitorization, while in a ER the decision to monitor or not a patient depends on his necessities; overall the frequency of monitoring is always based on the patient's clinical condition. The monitoring of vital signs was related with EWS, which was said to be in use in only some healthcare facilities managed by JMS. It was also mention that monitors can provide triggers based on user defined individual parameters values, some of which are already pre-defined for type of patient (e.g. adults vs. children). Emergency teams do not usually assist units such as the ICU and the criteria used to call them is defined by DGS guidelines [43]. Both AHA and ERC guidelines for SCA and resuscitation were mentioned.

Although the focus of this thesis was pre-SCA, there are some important notes. It was said that is not a trivial task to identify retrospectively patients that had a SCA, unless there is clear codification (e.g. ICD-9, ICD-10) that specifies it.

4.2. Surveys

After extracting the resulting set of answers from the web-based platform as an Excel file it was analyzed using Excel. It was possible to obtain five responses: two from a doctor and a nurse working in permanent service (low level monitoring), two from a doctor and a nurse working in an emergency room (medium level monitoring) and one from a doctor working at an ICU (high level monitoring).

All five respondents acknowledged that the measurements of vital signs acquired were registered in the EHR; additionally, data acquired in the ICU was also registered in a monitoring central. While in permanent service and ER the data was said to be registered by nurses, in the ICU it can be directly imported from the monitors, being validated by the nurses. In the ICU monitoring can be continuous or periodic, depending on the patient needs; on ER and permanent service the decision to have or not periodic monitoring ant its frequency also depends on the patient's clinical condition. In all three units there is the possibility of measuring HR, RR, Temperature, SPO2, GCS, SBP, and performing ECGs; additional parameters were mentioned such as capnography, pulse wave reflex, glycaemia, and ketonemia.

80% of the respondents knew what an EWS was. The percentage of responders who knew each type of EWS is shown in **Figure 7**.

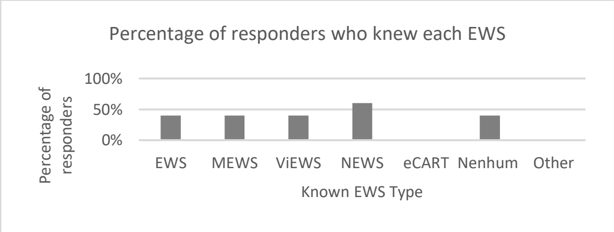

Figure 7: Percentage of responders who knew each EWS.

All respondents considered that the introduction of an EWS in the respective unit would be useful, however it was mentioned that it should be implemented as an automatic process that would retrieve data directly from an EHR system and would not require multiple registries of the same data by healthcare professionals. It was also said to be a good way of increasing patient safety, early recognition of potential deaths and comparison between objective data for later analysis.

Finally, the measures mentioned as potential improvements of outcomes and patient safety related with SCA focused on the prevention of such events, the uniformization of practice, the systematic retrieval of data to be able to calculate EWS and the existence of trained emergency teams. Having automatic triggers based on an EWS scale to alert emergency teams was also said to be a positive measure.

4.3. Assessment of vital sign data completeness

The initial dataset provided contained measurements of different parameters for 33863 patients and 51181 episodes. After filtering patients by age, the dataset contained 29111 patients and 43950 episodes. These are going to be considered when assessing completeness of data. After the final filtering by parameters of interest, the dataset contained 26367 patients and 39340 episodes. Of this data set 9 had a registered SCA.

Table 1 shows the completeness (documentation present) and omission (documentation not present) of each parameter in the given data set, filtered just for age. It is considered that a parameter is documented for a certain episode if there is at least one registry for that parameter.

The percentage of episodes within which the parameters were recorded at least once ranged from 0.9 % (SO_2 (therapeutic)) to 81.7% (Temperature). All parameters except temperature were recorded in less than 60% of the episodes, with RR having a particular low level of documentation (1.4%). HR, temperature, DBP and SBP were recorded for all the episodes with a registered SCA.

Table 1: Number of episodes, n (%), in which each parameter was or was not documented in the EHR, for overall sample and for episodes with registered SCA, and medium value and standard deviation of the values of each parameter.

Parameter	Documented, n (%)	Not Documented, n (%)	Documented for SCA, n (%)	Not Documented for SCA, n (%)	Medium Value	SD
GCS	23739 (54.0)	20211 (46.0)	7 (77.8)	2 (22.2)	14.53	1.50
HR (bpm)	20140 (45.8)	23810 (54.2)	9 (100.0)	0 (0.0)	76.25	14.82
RR (bpm)	613 (1.4)	43337 (98.6)	1 (11.1)	8 (88.9)	24.05	21.66
SO_2 (mmHg)	5499 (12.5)	38451 (87.5)	5 (55.6)	4 (44.4)	96.15	4.15
SO_2 (therapeutic) (mmHg)	374 (0.9)	43576 (99.1)	0 (0.0)	9 (100.0)	95.72	7.64
Temperature (°C)	35925 (81.7)	8025 (18.3)	9 (100.0)	0 (0.0)	36.49	0.72
DBP (mmHg)	13722 (31.2)	30228 (68.8)	9 (100.0)	0 (0.0)	68.95	14.34
SBP (mmHg)	13754 (31.3)	30196 (68.7)	9 (100.0)	0 (0.0)	126.5	22.29

A total of 4610 episodes (10%) had none of the necessary parameters to calculate CART or NEWS, with the exception of age and the lack of SO_2 (therapeutic). These will not be considered for the remaining analyses.

Fewer number of episodes had all the parameters necessary to calculate NEWS (n=299) than CART (n=595) (**Figure 7** and **Figure 8**) For the episodes in which a SCA was identified 8 had enough measurements to calculate 3 parameters and 1 to calculate 4 parameters, for CART. For NEWS the distribution was the following: 3 episodes had enough measurements to calculate 6 parameters, 1 episode to calculate 4 parameters, 4 episodes to calculate 5 and 1 episode to calculate 7 parameters.

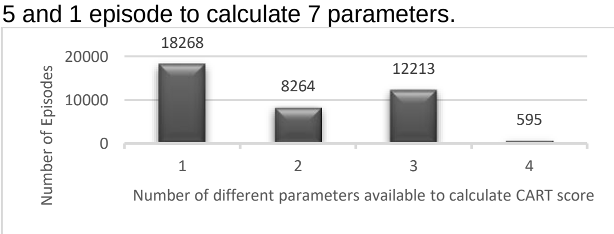

Figure 2: Distribution of episodes by the number of different parameters each has available to calculate CART score.

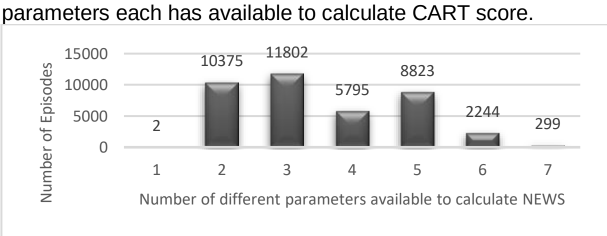

Figure 3: Distribution of episodes by the number of different parameters each has available to calculate NEWS score.

4.4. Comparison between NEWS and CART scores

The maximum value of NEWS and CART scores was calculated for the same episodes. **Figure 9** and **Figure 10** show the distribution of maximum values of NEWS per episode and the distribution of maximum values of CART score per episode, respectively. The episodes for which a SCA was identified had the following maximum NEWS scores: 3 (2 episodes), 4 (3 episodes), 5, 6, 7 and 8. The episodes for which a SCA was identified had the following maximum CART scores: 0, 4 (4 episodes), 8, 17, 20 and 35.

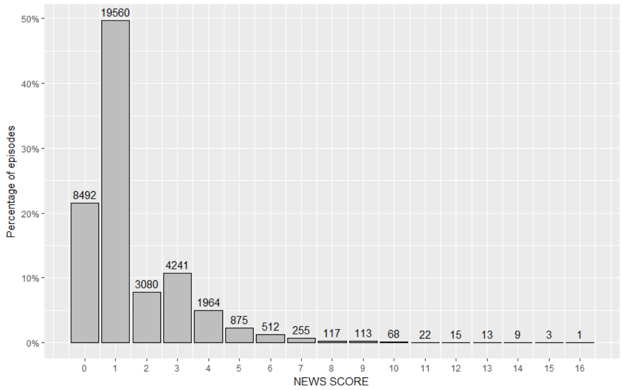

Figure 4: Distribution of maximum values of NEWS score per episode.

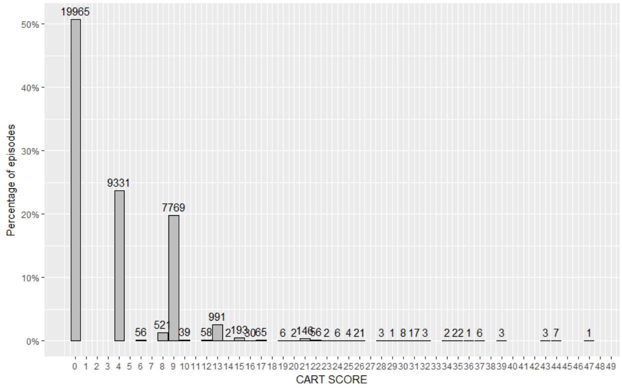

Figure 5: Distribution of maximum values of CART score per episode.

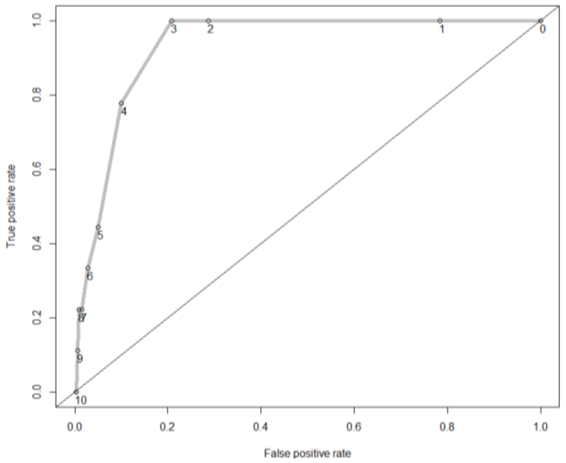

Figure 6: ROC plot (FPR vs TPR) for NEWS.

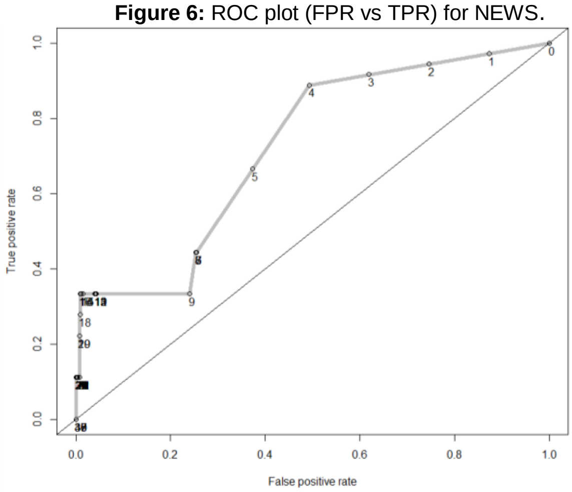

Figure 7: ROC plot (FPR vs TPR) for CART.

Figure 11 and **Figure 12** show the TPR and FPR points for NEWS (NEWS ROC curve) and CART (NEWS ROC curve), respectively.

The AUROCs for NEWS and CART were 0.93 and 0.72, respectively; this supposedly indicates that NEWS is a better prediction tool for SCA then CART for the all range of thresholds.

Figure 13 and **Figure 14** show the precision and recall points for NEWS (NEWS PR curve) and CART (CART PR curve), respectively. Both were plotted including all possible cutoff values. It shows that both NEWS and CART can distinguish themselves from a random classifier, being the baseline defined by 2.287×10^{-4}.

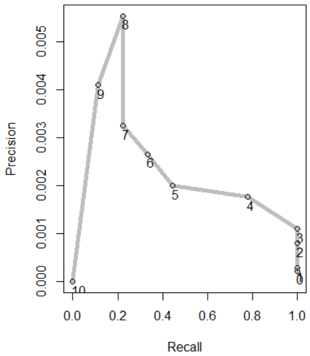

Figure 8: PR plot (Precision vs Recall) for NEWS.

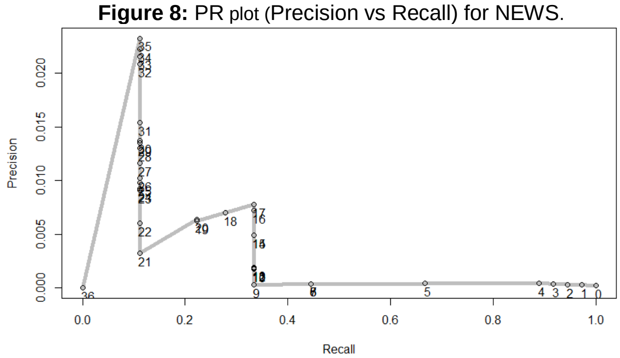

Figure 9: PR plot (Precision vs Recall) for CART.

The AUPRC for NEWS and CART were 0.00224 and 0.00281, respectively. This would mean that CART has a slightly better prediction performance than NEWS for the all range of thresholds; however, for all recall values the precision values in both plots are very low.

Given that NEWS had already well-defined cut-off points to take action and CART didn´t the next step was to match NEWS thresholds for medium (score higher than 5) and high risk (score higher than 7) with CART scores with the closest sensitivity to NEWS thresholds sensitivities that provided a higher precision (**Table 7**). The NEWS score 5 was matched with CART score 8. The NEWS score 7 was matched with CART score 17.

Table 1: Values of NEE, trigger rate, FPR, TPR and precision for selected cut-off values of CART and NEWS.

	Cut-off	NNE	Trigger Rate (%)	Specificity (%)	Sensitivity (%)	Precision (%)
CART	**4**	2422	49.3	50.8	88.9	0.041
	8	2497	25.4	74.6	44.4	0.040
	17	129	1.0	99.0	33.3	0.779
NEWS	**4**	567	10.1	89.9	77.8	0.177
	5	501	5.1	94.9	44.4	0.200
	7	308	2.8	98.4	22.2	0.325
	8	181	2.8	99.1	22.2	0.554

Comparing NEWS score 5 and CART score 8 it is possible to see that for equal sensitivity values the first one performs better in all the remaining metrics.

Comparing NEWS score 7 and CART score 17 it is possible to see that for equal sensitivity values the last one performs better in all the remaining metrics. However, sensitivity of 33% is a very low value and in the present situation would mean that only 3 of the 9 episodes in which SCA happened would be identified as positives.

5. DISCUSSION

5.1. Evidence based and good practice methods to predict in-hospital SCA

There are few to none recommendations on the use of EWS by the European Society of Cardiology (ESC), American Heart Association (AHA) and European Resuscitation Council (ERC). Although all of them state the importance of predicting SCA and the majority of detecting early signs of deterioration, not only but also to predict a SCA event, only 2015 AHA and 2015 ERC specify the existence of EWS tools to identify patients deteriorating and respective response of emergency teams to deterioration, specifying MEWS [25], [26]. However, neither suggest that a certain EWS system should be used instead of another one. 2015 AHA are the only guidelines where a direct recommendation on the use of EWS systems was made, stating that the use of these systems may be considered for adults and children, although that recommendation is only classified as Class IIb and LOE C-LD, which means that is a week recommendation (benefit higher or equal to risk) and the level of evidence is based on limited data on the matter. [25] Despite this, all respondents to the survey said that the introduction of an EWS in the respective unit they work in would be useful and a good method to increase patient safety, early recognition of potential deaths and comparison between objective data for later analysis. Although the sampled in study was low, healthcare professionals seem open and motivated to the introduction of EWS, nonetheless it was mentioned that its introduction should not have a great impact on workload.

The increasing number of risk assessment tools already developed to detect early signs of deterioration further difficult the choice of a well validated EWS system. First of all, EWSs perform differently depending on whether the outcome that one pretends to predict is SCA, ICU admission or death, or even a combination of all of them. According to the review studies, when testing the ability of a EHR to predict each of these outcomes all systems identified SCA worse than the remaining outcomes, which can be also explained by the extremely low incidence of such events despite its importance. Secondly, there are different methods for developing EWS such as logistic regression, and machine learning methods. Traditional EWS were developed either by empirically defining a model and test it after against a medical database or derive the model from those medical databases, which make them dependent on the dataset they were generated from. Recently the focus has been on machine learning methods, that have been shown to be more accurate for studying and implementing real-time prediction tools and should be further explored (Churpek et al., 2017), however may be more difficult to currently implement in a real-life setting. Thirdly, EWS need to have application in the context it is being inserted into, given that the population characteristics may condition a poor performance of the system or that the healthcare organization may not have enough resources to implement it. One of the limitations is the parameters that scoring system can include: these parameters need to be easily assessed and available when calculating the score, reason why most scoring systems are only based on routinely acquired data, such as vital signs. Is necessary that the health care organizations that decide to implement a EWS develop or adapt a protocol to use it in daily practice and instruct healthcare professionals to follow it. Pre-defined cut-off scores of that same EWS should trigger some sort of action, being it an increase on monitoring frequency or the consultancy with expert professionals. A clear advantage of NEWS over the remaining EWS reviewed is the existence of a well-defined and systematic plan of action for the implementation of such system that is already being used in several organizations.

Most of the reviewed articles revealed EWS were compared against each other by plotting ROC and determine AUROC for each of the EWS and each outcomes in study. However recent studies argue the applicability of AUROC as a method of evaluating this specific binary decision problem because it does not take into account the prevalence of the outcome that is intended to measure, which in this case is SCA, and only provides a summary value of all different cut-off values, most of which may not even be used, or have a clinical applicability (Romero-Brufau et al., 2015). Therefore additional evaluation methods that are sensible to data skewness should be used, which may change the results and conclusions of the previous reviewed articles no EWS (Saito and Rehmsmeier, 2015).

As it was also mentioned in the interviews and surveys different units inside the hospital have different strategies and levels of monitoring: it is easier to implement a EWS system in a well monotonised unit (e.g. ICU) however is also more useful to implement a EWS system in places where is less probable to identify a patient who will suffer an in-hospital SCA. Strategies for implementing such systems would need to be discussed independently but the main limiting factors would be

5.2. Assessment of vital sign data completeness

By analyzing the completeness/omission of each parameter it was possible to identify that all parameters with exception of temperature were recorded in less than 60% of the episodes, which indicates that improvements on data collection should be made. Focusing only in the patients that have had a SCA the results show better completeness of data, with all parameters except RR being recorded more that 50% of the time. This makes sense knowing that the measurements for the patients who suffered SCA came from hospital units with higher monitoring. Comparing these results with the ones obtained from a study that solely focused in the completeness of data for patients who suffered a SCA it was possible to see that patients who suffered a SCA in the present study had BP (100% vs. 29.8%), HR (100% vs.78.5%), temperature (100% vs. 64.9%) and Glasgow scale (77.8% vs. 15.4%) registered in more episodes that the literature study, with the majority of the parameters even being recorded for all patients. However, is important to notice that the literature study had a larger sample of patients that had suffered SCA. For the overall episodes in the present study temperature (81.7% vs. 64.9%) and Glasgow scale (54.0% vs.15.4%) were even registered for more episodes than the literature study [44].

RR had a particular low level of documentation for both overall population (1.4%) and episodes for which a SCA was registered (11.1%), being one of the parameters with the lowest documentation level. It was already reported in literature that some healthcare professionals underestimate the importance of RR, leading to a deficient recording of this parameter, although it is used for the calculation of both NEWS and CART score. The introduction of an EWS system into the organization can be a good strategy to decrease the percentage of missing data on this and other parameters as it requires the systematized assessment of parameters in other to calculate the overall scores [45], [46].

Possible reasons for the general incompleteness of data are the complexity of systems to enter data on vital signs even if they were measured and the posterior difficult assess to the documented data by the healthcare professionals to assist on decision making, which decreases the motivation to register such values instead of keeping track of the individual values [44]. Even data on which patients had a SCA was said not to be a trivial task do get, unless there is clear codification (e.g. ICD-9, ICD-10) that specifies it. Lack of time to introduce such parameters can also be an additional reason.

The implication of missing data for the comparison of CART and NEWS performance on predicting SCA is that it may be based on calculated CART and NEWS score that do not correspond to reality given that the maximum scores for each patient of each EWS system were calculated based on the dataset being analyzed. It can consequently decrease the conclusions that can be taken and provide a wrong fitting of the models to a data set [49].

The percentage of episodes in this study with at least one measurement for each of the parameters necessary to data to calculate NEWS and CART score was very low, being 1.5% for CART and 0.8% for NEWS, another reason for why the remaining episodes were not excluded. NEWS requires a higher number of parameters (n=7) than CART (n=4) to be assessed, which explains why fewer number of episodes had all the parameters necessary to calculate NEWS (n=299) than to calculate CART (n=595). The episodes for which a SCA was identified all had registered more than half the measurements necessary to calculate both NEWS and CART score.

5.3. Comparison between NEWS and CART scores

The ratio between episodes in which there was and there was not a registered SCA ($9{:}39340 \cong 2.3 \times 10^{-4}$) is lower than the expected from the estimates retrieved from the surveys to the healthcare professionals and samples used in the studies that were reviewed. One of the possible explanations is the difficult process of identifying which patients had a SCA unless there is clear codification indicating so, which was mentioned in the conducted interviews.

The maximum NEWS score value for the episodes in which a SCA was identified were lower than the threshold for medium risk (score 5) for 5/9 episodes and lower than the threshold for high risk (score 7) in 7/9 episodes, which could indicate that the values used to calculate the score did not correspond to the actual reality, meaning that they either were not measured when the parameter further distance from the normal range or were not recorded. It could also indicate a poor performance of NEWS to identify correctly patients deteriorating to the point of a SCA. As it was mentioned before no thresholds were defined for CART but given the results NEWS score 5 was matched with CART score 8 and NEWS score 7 was matched with CART score 17, which was also the value for which CART was considered to have a better performance in a previous study [6]. With this assumption, the maximum CART score value for the episodes in which a SCA was identified were lower than the threshold for both medium risk (score 8) and high risk (score 17) for 5/9 episodes, with a score having even a 0 score, the minimum possible value.

An ideal EWS should provide a maximum discrimination of patients' outcomes at a lowest trigger rate, therefore reducing the probability of missing any adverse event without an excessive workload and unnecessary call of emergency teams. (Rothschild et al., 2010; Smith et al., 2016). This shows than when choosing an EWS system to implement and when reporting its performance is always necessary to make a trade-off between the model's capability of detecting a high percentage of individuals with an outcome of interest and the percentage of false-positive outcomes. The first can be translated as the model's sensitivity because it shows the percentage of the total number of outcomes of interest that it was capable of identifying. For the percentage of false-positive outcomes there were some metrics that could be used; however, the ones that were chosen should reflect the prevalence of the outcome that is being predicted because in this case is a care event.

The values of AUROCs for NEWS and CART obtained in this study were 0.93 and 0.72, respectively. This may have leaded to choose NEWS instead of CART, given that the first has a higher AUROC value.

As suggested in literature, PR curve was also determined given that is more informative then ROC for skewed datasets and plots recall against precision [11]. When considering diagnostics tools having a low precision is very undesirable because the patient might be incorrectly classified as having a disease; for EWS it would not be so problematic because it would only mean that a patient would require more frequent monitoring and attention, although increases workload [33].

The values of AUPRC for NEWS and CART were 0.00224 and 0.00281, respectively. This means that CART has a slightly better prediction performance than NEWS for the all range of thresholds; however, for all recall values the precision values in both plots are very low. This conclusion is the opposite of the one made using AUROC values. Which means that this metrics alone can be misleading: the solution was too analyze more carefully cut-off points that are relevant and study how each one will affect the workload and the percentage of patients that are going to be correctly identified.

NNE was determined for the specific cut-off points because although it is mathematically given by the inverse of precision it can be more explanatory when discussing possible tradeoffs with healthcare professionals and hospital managements. Trigger rate, which is the number of triggers per number of monitored episodes, was also determined for the same reasons.

It was possible to conclude that for the cut-off value for medium risk NEWS would be a better option than CART because for equal sensitivity values, meaning episodes in which a SCA was correctly identified, the first one performed better in all the remaining metrics. For the cut-off value for high risk the opposite happened, meaning that a decision between both scoring systems focusing only in this cut-off values would rely on the sensitivity values that that healthcare organization/ unit would be willing to tolerate, given the associated burden.

6. CONCLUSION

This thesis intended to provide a better methodology for the comparison between EWS systems with the intent of being implemented in a certain healthcare organization. In this situation specifically the EWS that was recently implemented in two of the hospitals managed by JMS (NEWS) was compared with the EWS chosen after the compilation of evidence based and good practice methods to predict in-hospital SCA (CART), and using a data set containing measurements of parameters acquired from the EHR of another hospital managed by JMS.

In conclusion, the decision to opt by an EWS instead of another cannot focus in metrics that provide the overall performance of a tool or do not take into account the prevalence of the outcome that is being predicted. The decision should be made taking into account metrics that explicit show the tradeoff between the percentage of outcomes correctly predicted (precision) and the burden (e.g. NNE, precision, trigger rate), for possible and relevant cut-off values. The healthcare organization/ unit needs to decide which sensitivity values would be willing to tolerate, given the associated burden. Additionally, is important to have a well systematized plan of action associated with each cut off value defined., as well as an efficient way to apply the scoring system in daily practice, like being able to integrate it with the EHR system or monitors instead of using paper charts.

REFERENCES

[1] M. E. Kleinman *et al.*, “Part 5: Adult basic life support and cardiopulmonary resuscitation quality: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care,” *Circulation*, vol. 132, no. 18, pp. S414–S435, 2015.

[2] D. Mozaffarian *et al.*, *Heart disease and stroke statistics-2015 update : A report from the American Heart Association*, vol. 131, no. 4. 2015.

[3] R. M. H. Schein, N. Hazday, M. Pena, B. H. Ruben, and C. L. Sprung, “Clinical antecedents to in-hospital cardiopulmonary arrest,” *Chest*, vol. 98, no. 6, pp. 1388–1392, 1990.

[4] C. Franklin and J. Mathew, “Developing strategies to prevent inhospital cardiac arrest: analyzing responses of physicians and nurses in the hours before the event.,” *Resuscitation*, vol. 28, pp. 167–168, 1994.

[5] J. Nurmi, V. P. Harjola, J. Nolan, and M. Castrén, “Observations and warning signs prior to cardiac arrest. Should a medical emergency team intervene earlier?,” *Acta Anaesthesiol. Scand.*, vol. 49, no. 5, pp. 702–706, 2005.

[6] M. M. Churpek, T. C. Yuen, S. Y. Park, D. O. Meltzer, J. B. Hall, and D. P. Edelson, “Derivation of a cardiac arrest prediction model using ward vital signs,” *Crit. Care Med.*, vol. 40, no. 7, pp. 2102–2108, 2012.

[7] C. P. Subbe, A. Slater, D. Menon, and L. Gemmell, “Validation of physiological scoring systems in the accident and emergency department,” *Emerg. Med. J.*, vol. 23, no. 11, pp. 841–845, 2006.

[8] B. L. Albert and L. Huesman, “Development of a Modified Early Warning Score Using the Electronic Medical Record,” *Dimens. Crit. Care Nurs.*, vol. 30, no. 5, pp. 283–292, Sep. 2011.

[9] C. Mathukia, W. Fan, K. Vadyak, C. Biege, and M. Krishnamurthy, “Modified Early Warning System improves patient safety and clinical outcomes in an academic community hospital,” *J. Community Hosp. Intern. Med. Perspect.*, vol. 5, no. 2, p. 10.3402/jchimp.v5.26716, 2015.

[10] Royal College of Physicians, *National Early Warning Score (NEWS): Standardising the assessment of acute- illness severity in the NHS.* 2012.

[11] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLoS One*, vol. 10, no. 3, pp. 1–21, 2015.

[12] JMS, “José de Mello Saúde,” 2017. [Online]. Available: http://www.josedemellosaude.pt/.

[13] G. B. Smith *et al.*, “A Comparison of the Ability of the Physiologic Components of Medical Emergency Team Criteria and the U.K. National Early Warning Score to Discriminate Patients at Risk of a Range of Adverse Clinical Outcomes.,” *Crit. Care Med.*, pp. 2171–2181, 2016.

[14] J. M. Rothschild, E. Gandara, S. Woolf, D. H. Williams, and D. W. Bates, “Single-Parameter Early Warning Criteria to Predict Life-Threatening Adverse Events,” *J. Patient Saf.*, vol. 6, no. 2, pp. 97–101, 2010.

[15] G. B. Smith, D. R. Prytherch, P. Meredith, P. E. Schmidt, and P. I. Featherstone, “The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death,” *Resuscitation*, vol. 84, no. 4, pp. 465–470, 2013.

[16] M. M. Churpek, T. C. Yuen, and D. P. Edelson, “Risk stratification of hospitalized patients on the wards,” *Chest*, vol. 143, no. 6, pp. 1758–1765, 2013.

[17] C. P. Subbe, M. Kruger, P. Rutherford, and L. Gemmel, “Validation of a modified Early Warning Score in medical admissions.,” *QJM*, vol. 94, no. 10, pp. 521–526, 2001.

[18] D. R. Prytherch, G. B. Smith, P. E. Schmidt, and P. I. Featherstone, “ViEWS-Towards a national early warning score for detecting adult inpatient deterioration,” *Resuscitation*, vol. 81, no. 8, pp. 932–937, 2010.

[19] C. Anne Kelly, A. Upex, and D. N. Bateman, “Comparison of consciousness level assessment in the poisoned patient using the alert/verbal/painful/unresponsive scale and the glasgow coma scale,” *Ann. Emerg. Med.*, vol. 44, no. 2, pp. 108–113, 2004.

[20] J. D. Groarke *et al.*, “Use of an admission early warning score to predict patient morbidity and mortality and treatment success,” *Emerg Med J*, vol. 25, no. 12, pp. 803–806, 2008.

[21] M. Churpek, T. Yuen, C. Winslow, D. O. Meltzer, M. Kattan, and D. Edelson, “Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards,” vol. 44, no. 2, pp. 298–305, 2017.

[22] T. Badriyah *et al.*, “Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS),” *Resuscitation*, vol. 85, no. 3, pp. 418–423, 2014.

[23] M. Churpek, T. Yuen, and D. Edelson, “Predicting clinical deterioration in the hospital: the impact of outcome selection,” vol. 84, no. 5, pp. 564–568, 2013.

[24] M. Field and K. Lohr, *Clinical Practice Guidelines: Directions for a New Program*, no. 8. 1990.

[25] AHA, *2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care*, vol. 132, no. 18. 2015.

[26] R. Greif *et al.*, “European Resuscitation Council Guidelines for Resuscitation 2015. Section 10. Education and implementation of resuscitation,” *Resuscitation*, vol. 95, pp. 288–301, 2015.

[27] J. Davis and M. Goadrich, “The relationship between Precision-Recall and ROC curves,” *Proc. 23rd Int. Conf. Mach. Learn.*, pp. 233–240, 2006.

[28] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, “Evaluation measures for models assessment over imbalanced data sets,” *J. Inf. Eng. Appl.*, vol. 3, no. 10, pp. 27–38, 2013.

[29] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett. 27*, vol. 27, pp. 861–874, 2006.

[30] K. Hajian-Tilaki, “Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation,” *Casp. J. Intern. Med.*, vol. 4, no. 2, pp. 627–635, 2013.

[31] C. M. Florkowski, “Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests.,” *Clin. Biochem. Rev.*, vol. 29, pp. S83–S87, 2008.

[32] A. Beger, “Precision-Recall Curves,” *SSRN Electron. J.*, pp. 1–7, 2016.

[33] S. Romero-Brufau, J. M. Huddleston, G. J. Escobar, and M. Liebow, “Why the C-statistic is not informative to evaluate early warning scores and what metrics to use,” *Crit. Care*, vol. 19, no. 1, p. 285, 2015.

[34] K. Boyd, K. H. Eng, and C. D. Page, “Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals,” in *Lecture Notes in Computer Science*, vol. 8190 LNAI, no. PART 3, 2013, pp. 451–466.

[35] Typeform, “Typeform,” 2017. [Online]. Available: https://www.typeform.com/.

[36] H. Martins and F. Mealhada, “Exames sem papel no coração: prevenção e monitorização de risco cardiovascular,” *Factores de Risco*, no. 42, pp. 38–40, 2016.

[37] CUF, “Infraestruturas,” 2017. [Online]. Available: https://www.saudecuf.pt/unidades/infante-santo/o-hospital/infraestruturas.

[38] WHO, “International statistical classification of diseases and related health problems - 10th revision.,” *World Heal. Organ.*, vol. 2, no. 5, pp. 1–252, 2016.

[39] “ICD9,” 2009. [Online]. Available: http://icd9cm.chrisendres.com/index.php.

[40] RStudio Team, “RStudio: Integrated Development for R. RStudio, Inc.,” 2015.

[41] D. Nedea, “AVPU Scale,” 2017. [Online]. Available: https://www.mdapp.co/avpu-scale-calculator-163/.

[42] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer, “ROCR: Visualizing classifier performance in R,” *Bioinformatics*, vol. 21, no. 20, pp. 3940–3941, 2005.

[43] DGS, “Normas e Circulares Normativas,” 2016. [Online]. Available: https://www.dgs.pt/directrizes-da-dgs/normas-e-circulares-normativas.aspx.

[44] J. E. Stevenson, J. Israelsson, G. C. Nilsson, G. I. Petersson, and P. a Bath, “Recording signs of deterioration in acute patients: The documentation of vital signs within electronic health records in patients who suffered in-hospital cardiac arrest.,” *Health Informatics J.*, vol. 22, no. 1, pp. 21–33, 2016.

[45] J. McBride, D. Knight, J. Piper, and G. B. Smith, “Long-term effect of introducing an early warning score on respiratory rate charting on general wards,” *Resuscitation*, vol. 65, no. 1, pp. 41–44, 2005.

[46] M. Odell *et al.*, “The effect of a critical care outreach service and an early warning scoring system on respiratory rate recording on the general wards,” *Resuscitation*, vol. 74, no. 3, pp. 470–475, 2007.

[47] World Health Organization Regional Office for Europe, “From Innovation to implementation eHealth in the WHO European Region,” 2016.

[48] P. B. Jensen, L. J. Jensen, and S. Brunak, “Translational genetics: Mining electronic health records: towards better research applications and clinical care,” *Nat. Rev. - Genet.*, vol. 13, pp. 395–405, 2012.

[49] K. B. Bayley, T. Belnap, L. Savitz, A. L. Masica, N. Shah, and N. S. Fleming, “Challenges in Using Electronic Health Record Data for CER,” *Med. Care*, vol. 51, pp. S80–S86, 2013.

[50] A. R. Patel, F. J. Zadravecz, R. S. Young, M. V. Williams, M. M. Churpek, and D. P. Edelson, “The Value of Clinical Judgment in the Detection of Clinical Deterioration,” *JAMA Intern. Med.*, vol. 175, no. 3, pp. 456–458, 2015.

[51] T. J. Hodgetts, G. Kenward, I. G. Vlachonikolis, S. Payne, and N. Castle, “The identification of risk factors for cardiac arrest and formulation of activation criteria to alert a medical emergency team,” *Resuscitation*, vol. 54, no. 2, pp. 125–131, 2002.

[52] G. D. Finlay, M. J. Rothman, and R. A. Smith, “Measuring the modified early warning score and the Rothman Index: Advantages of utilizing the electronic medical record in an early warning system,” *J. Hosp. Med.*, vol. 9, no. 2, pp. 116–119, 2014.

[53] JMS, “Política de Segurança do Doente,” 2017. [Online]. Available: https://www.josedemellosaude.pt/qualidade/politica-de-seguranca-do-doente.