Détection, prédiction et prévention des évènements indésirables liés aux soins via l'intelligence artificielle, en associant approche basée sur les règles de décision et l'apprentissage automatique

Antoine Saab

UNIVERSITE PARIS XIII –SORBONNE PARIS NORD

École doctorale Sciences, Technologies, Santé Galilée

Détection, prédiction et prévention des évènements indésirables liés aux soins via l'intelligence artificielle, en associant approche basée sur les règles de décision et l'apprentissage automatique

Detection, prediction and prevention of healthcare acquired adverse events with artificial intelligence, associating rules-based approach and machine learning

---

THÈSE DE DOCTORAT
présentée par

Antoine SAAB

Laboratoire d'informatique médicale et d'ingénierie des connaissances en e-Santé (LIMICS)

pour l'obtention du grade de
DOCTEUR EN INFORMATIQUE BIOMEDICALE

soutenue le 13 Décembre 2022 devant le jury d'examen constitué de :

CHAZARD Emmanuel, Université de Lille/ CHU de Lille, Rapporteur
MICHEL Philippe, CHU de Lyon, Rapporteur
AZZAG Hanene, Université Sorbonne Paris Nord, Examinatrice
LAMY Jean-Baptiste, Université Sorbonne Paris Nord, Directeur de thèse

To my children

Firas Elias,

Katerina Nour,

Maria Lydia

Acknowledgments

To the Professors Emmanuel CHAZARD and Philippe MICHEL who have amiably and generously accepted to evaluate this work.

To the Professor Hanene AZZAG who has kindly accepted to be part of the thesis jury.

To my thesis director, Professor Jean-Baptiste LAMY for his constant support, availability and genuine guidance and help all along this scientific journey. It has been an honor to work with you and to share difficulties and successes throughout these difficult years.

To my thesis co-director, Professor Mohammad KHALIL, for this opportunity and to future collaborations hopefully.

To the Lebanese Hospital Geitaoui -UMC directors, Sis. Hadia ABI CHEBLI and Pr. Pierre YARED for your support and trust.

To my friends and scientific co-pilots in this adventure, Melody SAIKALI, Cynthia ABI KHALIL and Mouin JAMMAL. May this open doors to future opportunities and achievements in our shared dream for better and safer patient care.

To my wife, Athina, for her support, patience and constant encouragement throughout these four years full of complicated challenges.

To my parents, for their love and sacrifices which made this moment possible.

Outline

List of Figures

List of Tables

1. Introduction

1.1. Background and significance

Patient safety is considered as one of the biggest public health priorities, with healthcare systems around the globe attempting to mobilize resources for the prevention of patient harms.

According to the World Health Organization's Global Patient Safety Action Plan 2021-2030 [5], research studies show that an average of one in ten patients is subject to an adverse event (AE) while receiving hospital care in high- income countries [6]. The estimate for low- and middle- income countries suggests that up to one in four patients is harmed, with 134 million AEs occurring annually due to unsafe care in hospitals, contributing to around 2.6 million deaths [7]. Overall, 60% of deaths in low-and middle-income countries from conditions amenable to health care are due to unsafe and poor-quality care [8].

Despite a positive momentum in the last twenty years [9]–[12], substantial opportunities for improvement remain, stemming from inefficient or inadequate implementation of safety actions [13], patient harm remaining one of the leading causes of morbidity and mortality in healthcare [14], causing further burden on healthcare systems such as increased costs, extended hospital stays and higher readmission rates [15]–[19]. It is estimated that 10-15% of healthcare expenditure is consumed by the direct sequelae of healthcare-related patient harms [20].

There are a number of reasons why the problem of the high incidence of patient harm is still prevailing despite more than two decades of awareness and research in the domain of patient safety [13]:

- Lack of reliable and efficient tools for the measure and surveillance of clinical AEs.
- Inconsistency in use of proven prevention techniques
- Lack of accessibility to proven prevention techniques
- Lack of usability of some technological solutions
- Areas with insufficient/no proven prevention techniques: the safety "blind spots"

Artificial Intelligence (AI) promises to hold an important potential in identifying, predicting and preventing patient harm. In fact, it can be applied to detect patient safety events and improve performance of clinical alarms, provide prediction of various patient safety events, and improve adherence to best practices.

While certain application domains of AI in medicine are getting gradually mature, namely medical imaging, many of the patient safety domain areas have still not witnessed widely

recognized AI-based applications. As for experimental applications that are published, the vast majority have still not been sufficiently validated for real-world use scenarios.

The systematic detection of clinical adverse events (AE), such as for example healthcare associated infections, post-surgical events, or medication related adverse events is fundamental to the measure of the level of patient harm arising from provision of care to patients. The availability of this measurement is in itself important to identify the evolution of incidences of various harm categories, that can be of interest at an individual (patient level), institutional (hospital level), regional (healthcare system) and global level. Without valid measurements of this type, it is almost impossible to assess the true impact of safety improvement actions on clinical outcomes.

Furthermore, if such events can be predicted in due time, a significant potential in terms of prevention can be unleashed for the best interest of the patients, the healthcare providers and the whole healthcare system.

The development of AI-based applications holds many challenges related to design, validation and user acceptance that need to be tackled before any such application can be transformed into a valuable tool for the improvement of patient safety.

1.2. Objectives

In this thesis, our objective is to explore the potential of AI when applied to the domain of prediction, measurement and prevention of clinical Adverse Events.

Our starting point will be a non-AI based trigger-tool system for the automated measurement of hospital acquired adverse events, designed before the thesis.

After identifying and analyzing the advantages and shortcomings of these systems, we will design two AI-based machine learning models, devoted to the prediction of hospital readmissions and patient clinical deterioration, respectively. We will also design an AI-based Clinical Decision Support application that associates the second machine learning model with clinical rules written by medical experts.

We will highlight specifically the added value that such systems could potentially produce both to improve current tools performances and to contribute to the patient safety efforts. In addition to that we will analyze the specificities of using machine learning versus rules-based algorithms in such applications, and how these two techniques can be used in synergy.

1.3.List of publications

- **Preliminary study**
  Saikali M, Békarian G, Khabouth, Mourad C, Saab A . Automated detection of patient harm: implementation and prospective evaluation of a real-time broad-spectrum surveillance application in a hospital with limited resources. (Accepted, September 2022, Journal of Patient Safety)

- **Contribution 1**
  Saab A, Saikali M, Lamy JB. Comparison of Machine Learning Algorithms for Classifying Adverse-Event Related 30-Day Hospital Readmissions: Potential Implications for Patient Safety. Stud Health Technol Inform. 2020 Jun 26;272:51-54. doi: 10.3233/SHTI200491. PMID: 32604598.

- **Contribution 2**
  Saab A, Abi Khalil C, Jammal M, Saikali M, Lamy JB. Early Prediction of All-Cause Clinical Deterioration in General Wards Patients: Development and Validation of a Biomarker-Based Machine Learning Model Derived From Rapid Response Team Activations. J Patient Saf. 2022 Sep 1;18(6):578-586. doi: 10.1097/PTS.0000000000001069. PMID: 35985042.

- **Contribution 3**
  Saab A, Abi Khalil C, Jammal M, Saikali M, Lamy JB. Design, implementation and validation of a Clinical Decision Support application for the identification and risk management of hospitalized patients at risk of clinical deterioration-the VIGIL project

  (in development, application developed, results of validation phase in progress)

1.4. Thesis outline

Hereafter, we have adopted the following outline for the thesis delineation.

First, we will present the state of the art regarding the measurement of the level of patient harm in healthcare institutions along with its practical and theoretical challenges (chapter 2, section 2.1), and the state of the art regarding the use of artificial intelligence to detect, predict and prevent patient harm (chapter 2, section 2.2). We will also present our preliminary, pre-PhD, study regarding the automated detection of in-hospital patient harm through non-AI based "triggers" (chapter 3).

Second, we will present the thesis contributions relative to the application of artificial intelligence in the detection and classification of patient harm in chapter 4 (applied to patient hospital readmissions) and in chapter 5 (applied to the prediction of patient clinical deterioration).

Third, we will present in chapter 6 an ongoing experience relative to the design, implementation and validation of a clinical decision support tool that was built based upon the machine learning model used in chapter 5 in addition to clinical rules and interventions, as a solution that can be used for the management of patients at risk of clinical deterioration in regular hospital floors.

In chapter 7, a final discussion and conclusion section will critically wrap up the important findings and commentaries stemming from the different contributions with regards to the findings of the state of the art.

2. State of the art

2.1. Current gaps in patient harm and clinical adverse events measurement

2.1.1. Hospital-acquired adverse events: incidence and evolution

Patient safety is defined as the absence of preventable harm to a patient and minimization of the risk of harm associated with the health care process [21]. In 1999, the publication of the Institute of Medicine's landmark report "To Err is Human: building a safer health system" [22] was transformational for patient safety awareness. It highlighted for the first time the extent of the problem of "patient harm", estimating that as many as 98 000 people died in American hospitals each year as a result of medical errors.

An increase in research about safety gaps in healthcare institutions followed in the few years after the report publication, and soon, patient safety was considered as one of the biggest public health priorities, with healthcare systems around the globe attempting to mobilize resources for the prevention of patient harms.

According to the World Health Organization's Global Patient Safety Action Plan 2021-2030[5], research studies show that an average of one in ten patients is subject to an adverse event (AE) while receiving hospital care in high- income countries [6]. The estimate for low- and middle- income countries suggests that up to one in four patients is harmed, with 134 million AEs occurring annually due to unsafe care in hospitals, contributing to around 2.6 million deaths [7]. Overall, 60% of deaths in low-and middle-income countries from conditions amenable to health care are due to unsafe and poor-quality care[8].

Two decades after this eye-opening report and the efforts deployed, there seems to be a significant improvement in the rates of AEs [9]–[12] in the United States but further research is needed to understand if this trend is similar across other countries, and across harm categories.

Despite this positive momentum, substantial opportunities for improvement remain, stemming from inefficient or inadequate implementation of safety actions [13], patient harm remaining one of the leading causes of morbidity and mortality in healthcare [14], causing further burden on healthcare systems such as increased costs, extended hospital stays and higher readmission rates [15]–[19]. It is estimated that 10-15% of healthcare expenditure is consumed by the direct sequelae of healthcare-related patient harms [20].

However, unsafe care is not only linked to hospital-based care; in fact, half of the global disease burden arising from patient harm seem to originate in primary and ambulatory care [23].

The majority of patient harms fall into the following categories: healthcare associated infections (HAIs), adverse drug events (ADEs), venous thromboembolism (VTE), surgical complications, pressure ulcers, falls, insufficient decompensation detection and diagnostic errors, including missed and delayed diagnosis[24], [25][24], [25].

A more detailed taxonomy of patient harm and adverse events categories has been elaborated by the Florida Hospital and illustrated in the Global Trigger Tool Implementation Guide for New Zealand [26]. Table 1 shows a slightly modified version of this taxonomy where we have added diagnostic errors and decompensation detection errors.

Table 1: Taxonomy of patient harm/clinical adverse events categories (based on the Florida Hospital Classification)

Events Related to Medications/Intravenous Fluids	
Kidney damage due to contrast dye	*Clostridium difficile* medication associated infection
Medication related renal insufficiency	IV volume overload/electrolyte imbalance
Medication related cardiac even/arrhythmia	Medication related hypotension
Medication related delirium, confusion, or over-sedation	Medication related glycemic events
Medication related allergic reaction	Medication related diarrhoea
Medication related bleeding	
Hospital Acquired Infections	**Events Related to Surgery or Other Procedure**
Catheter associated urinary tract infection	Abnormal bleeding following surgery or procedure
Central line associated blood stream infection	Blood clots and other occlusions related to surgery or procedure
Peripheral or central line non-blood stream infection	Complications related to peripheral venous or arterial puncture
Respiratory infection (non-ventilator associated)	Post-op acute renal failure
Surgical infection	Removal of retained foreign body
Ventilator associated pneumonia	Removal, injury or repair of organ
Other Hospital Acquired Infection	Cardiac complications related to surgery or procedure
Diagnosis related events	Prolonged post-op ileus
Misdiagnosis	Respiratory complications related to surgery or procedure
Delayed diagnosis	**Events Related to Patient Care**
Decompensation detection related events	Deep Vein Thrombosis/Venous Thromboembolism
Unidentified decompensation	Hospital acquired pressure ulcer /wound
Delayed identification of decompensation	Complications related to peripheral venous or arterial puncture
Inappropriate management of decompensation	Fall with injury
Other category of adverse events	
Other adverse events	

2.1.2. Measuring rates of clinical adverse events: current tools and challenges

One of the factors that may explain the insufficient level of desirable improvement of patient safety to this day is the lack of accurate measurement and regular monitoring of AE rates, both at national and institutional levels [13]. Hence, stemming from the "one cannot improve what one cannot measure" principle, it is crucial that healthcare systems measure AE levels timely, efficiently, reliably, and regularly in order to deploy appropriate actions to lower the level of harm.

Several tools have been developed in the literature to help detect AEs. These include morbidity and mortality conferences, autopsies, malpractice claims analysis, error reporting systems, administrative data analysis, chart review, and clinical surveillance through patient safety indicators[27]. However, all these methods are retrospective, limited in matter of detection scope, and resource intensive. In fact, voluntary reporting systems have been estimated to detect only 2-8% of actual AEs occurring during hospitalizations [28][29].

The latest recommended "gold standard" approach, currently used in many hospitals worldwide, is to perform a manual review of a sample of patient medical records, using a standardized approach such as the Institute for Healthcare Improvement's Global Trigger Tool (GTT) methodology[30].

Triggers are not harms in themselves – but 'clues' that harm may have occurred. When reviewing a medical record, triggers are used as 'flags' for potential harm. A trigger could be a laboratory result outside the normal range, a medication that has been prescribed or suddenly stopped, escalation to a higher level of care, or a return to interventional theatre. This enables a more efficient and streamlined approach for record review.
They have been identified as being associated with patient harms, but not all positive triggers necessarily identify an adverse event [31].

Although this method has shown to have a higher sensitivity versus other AE detection methods, its practical disadvantages include its resource-intensiveness due to time and personnel required, limitations in detection associated with the quality of clinical documentation, in addition to low inter-rater reliability, which are all limiting factors in its adoption[32], [33].

Recent research has suggested that automated AE detection methods using "data mining" techniques and "rules-based algorithms" are showing to be superior to manual tools [34]–[36] allowing for healthcare professionals to provide timely feedback and safety interventions [37]. Indeed, by design, automated AE detection methods can consistently screen large numbers of patients in real time to save valuable resources,

something which would be extremely tedious if done manually by reviewers with the same accuracy [36].

Efforts to automate the detection of AEs have been driven by the increase in the adoption of electronic medical records (EMR) worldwide over the last decade [38], with some incorporating a modified automated GTT methodology based on laboratory values, ICD-10 diagnosis codes and text mining of clinical documentation[32][39]. Since the safety benefits from EMRs have not been reaped as much as expected [40][41] (mainly because of a lack of "usability" of the solutions which increases burden to the healthcare workers, as well as a lack of intelligent clinical decision support functions that can help optimize care delivery and safety). Consequently, recent reports indicate that investments should now be focused on developing automated detection methods using EMR data, to routinely and continuously measure the frequency and types of patient harm [42].While most implemented automated AE systems detect only certain categories of AEs, such as hospital acquired infections, patient falls or adverse drug events [43]–[47], very few [48]–[50] offer the surveillance of a broad spectrum of AEs, and those who did were built on top of mature and complex EMR systems, not accessible to hospitals with limited resources.

Automated measurement tools for AEs are currently not available in the majority of commercially available solutions for hospitals and are still under development[51]. In fact, an ideal tool for the surveillance of patient safety events should provide validated, reliable, automated and real-time AE measurements that can be linked to improvement actions, and easily followed over time. Such a system should also use adequate and intuitive data representation tools in the form of a dashboard that is accessible to relevant managerial positions in the healthcare institution.

2.1.3. Main safety gaps remaining

There are a number of reasons why the problem of the high incidence of patient harm is still prevailing despite more than two decades of awareness and research in the domain of patient safety [13]:

- **Lack of reliable and efficient tools for the measure and surveillance of clinical AEs.** We cannot improve what we cannot reliably and consistently measure. Currently available tools for the measurement of clinical AEs do not permit healthcare institutions and systems to measure patient harm inside and outside the hospital, reliably, consistently and efficiently. Automated solutions are still under development and currently not available in the majority of commercially available solutions for hospitals. Without such tools, it is not possible to measure the evolution of progress in the prevention of AEs, in order to identify the impact of safety practices implementation.

- **Inconsistency in use of proven prevention techniques.** In fact, a number of prevention practices have demonstrated proof to have reduced the burden of certain categories of adverse events. For example, certain bundles of evidence-based practices were effective in drastically reducing certain types of hospital acquired infections (HAIs) [52], [53] such as Central line- associated bloodstream infections (CLABSI), Ventilator Associated Pneumonias (VAP) and Surgical Site Infections (SSI). Nevertheless, in recent studies hospital acquired infection rates remain high in most healthcare institutions in the USA[54], and this fact is thought to be related to the lack of adoption or inconsistency in the use of the aforementioned prevention techniques. The COVID-19 pandemic has had also a negative impact on the rate of HAIs in hospitals[55]. Hence, in order to improve the results of a domain of patient adverse events, interventions, tools or policies reinforcing the level of adoption and/or consistency in the implementation of the evidence-based practices are needed.

- **Lack of accessibility to proven prevention techniques.** For a number of AE domains, such as Medication errors, computerized clinical decision support solutions have been imagined, implemented and proven to be efficient to decrease the level of errors and consequently harms in this domain. Such solutions include computerized physician order entry (CPOE) that comprise modules to automatically check or flag allergies, drug-drug interactions, out-of-range doses, bar-code checking of drugs before dispensation and before administration. The level of adoption of such tools although increasing in developed countries, is still very limited in moderate- and low-income countries, mainly because of the high costs of such solutions. Lowering the barrier to accessibility to these solutions seem to be a main target in order to improve the safety results for this AE domain.

- **Lack of "usability" of some technological solutions.** Recent data[56] is suggesting that Clinical Decision Support function of existing Electronic Health Records (EHR) are not delivering the benefits that were promised in the early proof-of-concept studies. Main reasons for this poor performance are the disregard (in the design and implementation phases) of certain "non-technical" factors such as workflow adaptation, user-friendliness, training, organizational and culture issues. The term "usability" is currently being used to describe and measure the outcomes of the software functions in comparison to the needs of health care team members and the way they perform their work. EHRs with low usability scores seem even to be linked to bad patient outcomes and clinician burnout [57]. Hence, priority is to address these design and implementation issues of these solutions in order to reap the promised safety benefits. A number

of approaches involving disciplines such as human factors engineering, psychology, social sciences and user-oriented design are needed [58].

- **Areas with insufficient/no proven prevention techniques: the safety "blind spots".** A certain number of domains are emerging as important areas with significant incidence rates that lack evidence-based interventions proven to lower the associated level of harm, and thus require immediate attention from the patient safety scientific community. Such domains include: diagnostic errors, patient decompensation, inpatient falls, acquired pressure ulcers, outpatient safety and health information technology (HIT) induced errors.

  Diagnostic errors can stem from multiple factors, some are systemic (such as failures or breakdowns in communication, lack of coordination/handover or robust procedures) and others are individual (such as failures in data gathering or interpretation especially for abnormal patient results and patient history/status, overconfidence in diagnostic judgment and lack of knowledge/experience). Hence, no single physician's knowledge and decision making are sufficient to ensure an accurate diagnosis, especially when the diagnosis evolves across time and space and involves interactions between numerous team players.

  Clinical decompensation (or deterioration) of patients in clinical wards is a common event and can result in avoidable mortality. A patient who needs cardiac resuscitation is usually not expected to have a good prognosis. It is therefore important to have systems in place to detect the early signs of deterioration, so that mitigation can take place. Paper-based early warning systems, such as MEWS[59], have been elaborated in the early 2000 to assist nurses in the detection of clinical signs of deterioration. They have been widely implemented especially in the UK, USA and Australia, but much less in other developed countries, and even less in moderate- and low-income countries. The evidence base of the safety outcomes of these systems is very limited, and might be related to the relatively delayed detection of deterioration they offer. Currently, investments are made in electronic early warning systems that can offer a more reliable and earlier detection of patient deterioration.

  Patient falls are among the most common adverse events reported in hospitals and rates of fall are difficult to measure reliably since they depend mostly on reporting by nursing staff and there are no effective, systematic and independent approach to detect them. About 2% of hospitalized patients fall at least once during their stay. Recent data about the trends of adverse events rates in the US show that

Even though a number of successful quality improvement programs have been described (mostly relying on the risk evaluation of fall-risk in patients), most controlled studies of fall prevention did not yield positive results [60]. There is an urgent need for well-designed research studies in hospital fall prevention. Nurse staffing and even unit design considerations may play an important role into decreasing fall risk. The current nursing shortage witnessed on the international level can also negatively impact the efforts to reduce patient falls. There's a strong need to explore solutions to identify reliably inpatient falls and even offer a clinical decision support aid to flag patients at high risk, and maybe even detect early signs of high-risk physical activity that can lead to fall, hence preventing it.

Pressure ulcers are also among the adverse events with the highest incidence in hospitals. In the US, approximately 2.7% of hospitalized patients develop pressure ulcers that are largely preventable, thus incurring more than 28.2 billion (USD, 2019) financial burden which could be significantly reduced.

In a recent US national study[12], between 2010 and 2019, the decrease in the rates of pressure ulcers and falls was significantly lower than the other included AE categories, which may indicate a need for new initiatives related to the prevention of pressure ulcers and inpatient falls.

2.1.4. Preventable adverse events: the "holy grail" of patient safety

The measurement of patient harm is important in order to assess the level of patient safety in a healthcare institution. The objective of this measure is to help healthcare administrations and leaders to identify gaps and subsequently help derive specific improvement actions, with the objective to lower the level of harm in the identified safety areas. However, while this approach is important on a global and managerial level, a risk-predictive approach on the patient level could hold even greater potentials for harm prevention.

A certain number of AE categories are particularly eligible to this approach. These are the ones for which the risk of occurrence can be determined by independent factors related to the patient and other contextual variables that can be obtained in routine, in particular.

A number of risk assessment tools have been elaborated (mostly in paper form) in the last three decades for certain categories of AEs, in order to be used directly during patient care. Such examples include the Braden Scale [61] for the risk assessment of pressure ulcer, the Morse scale [62] for the risk assessment of patient fall, the Modified

Early Warning System score [59] for the stratification of clinical deterioration risk, and the Well's risk score to predict Deep Vein Thrombosis [63].

As clinical data availability is increasing with growing EHR adoption, risk scores can be elaborated for a number of additional AE domains. These CDSS tools can be constructed either through expert opinion or through data-driven techniques (e.g machine learning) or even in combining both methods. Possible candidate areas include risk of induced Acute Kidney Injury, acquired delirium, postoperative infections, risk of stroke, risk of pneumonia etc. The adoption of such tools by the healthcare professionals is conditioned by a better integration of these tools [64] in the EHR, taking into consideration the needs of the professionals and the findings of the CDS implementation science.

Ultimately, the timely incorporation of risk prediction CDS modules for the different types of AEs in the workflow of nurses, physicians and other healthcare could deliver personalized recommendations and alerts relative to the prevention of AEs that promise to impact the safety of care in more efficient ways than global improvement projects and protocols.

2.2. Use of artificial intelligence to detect and predict patient harm

2.2.1. The potential of AI in identifying, predicting and preventing patient harm

Recent reviews [25], [65], [66] of studies focusing on the use of AI in patient safety domains have highlighted the following conclusions regarding the potential of use of AI in the identification, prediction and prevention of patient harm:

- **Detect patient safety events and improve performance of clinical alarms**

Rules-based systems have been, since the end of the 1990s, first to be designed and evaluated in order to automatically detect adverse events. The later version of these systems used electronic triggers found in the EHR, some of which were inspired from the IHI Trigger Tools methodology [66]. Their detection performance was significantly better than the paper chart audit gold standard, but remained overall modest especially in terms of specificity and false alarm rates.

A number of recent studies have shown that machine-learning based systems could improve the detection performance of medication related adverse events, including prescription errors[67]–[69], [68]drug-drug interactions and medication reconciliation errors [70]–[72], reduce the false alarm rate of monitor vital sign data, and in particular alarms related to cardiac events (e.g arrythmias) [73]–[75].

Moreover, a number of Natural Language Processing (NLP) and Machine Learning (ML) based systems have been elaborated for automatic identification and classification of patient safety incidents and adverse events from incident reports and from the EHR clinical notes [76]–[78].

In the medical imaging field, numerous models and systems have been elaborated in order to automatically help identify relevant radiological findings including possible adverse events, thus contributing to reduce diagnostic errors [79], [80] .

- **Provide prediction of various patient safety events**

Reports have shown that several ML-based systems have been designed for the prediction of patient harm in multiple patient safety domains. Main prediction domains include for example: the onset of central-line associated bloodstream infections [81], Adverse Drug Events (ADE) and drug-drug interactions based on drug structural similarities and mechanism of action [68], [82], [83], predicting the personalized therapeutic dosage of digoxin [84] and warfarin through integrating certain relevant genomic sequencing data [85], [86], identifying inpatients at high-risk for Venous thromboembolism [87], [88], predicting postoperative blood loss [89], stratifying patient relative to the risk of developing pressure ulcers [90], predicting fall risk using wearables and computer vision [91], [92], and prediction of patient clinical deterioration [93], [94]

- **Improve adherence to best practices**

It was reported that patients only receive recommended care about 50% of the time [95], and unwarranted variation in clinician's decisions and action impair care delivery [96], [97] when they fail to use applicable guidelines[98]. AI-based applications can play a potential role in improving adherence to best practice through predicting user behavior and giving feedback, suggesting recommendations or enforcing certain actions at the right time in the clinician's workflow.

For example, ML algorithms combined with computer vision and data from other types of sensors were applied to monitor hand hygiene compliance both in outpatient and inpatient settings, with good reported performance in increasing the best practice compliance [99]–[101]. Similarly, in radiology, systems have been designed to enforce the appropriate use of diagnostic imaging modalities, for example in the appropriate prescription of imaging exams for the diagnosis of deep vein thrombosis and pulmonary embolism[102], [103].

2.2.2. Current concerns and challenges regarding the elaboration and use of AI-based models and applications

Machine learning and other artificial intelligence techniques are playing an increasing role in healthcare, across many clinical domains, and hold a big potential to improve patient safety gaps. This is being translated by the rapidly growing number of AI-based models and systems in this domain that are being published in the scientific literature, and focused upon in the media. While this is undoubtedly a sign of a paradigm shift in this domain that is leveraged by technology, the current "hype" shadows real concerns and gaps that are being raised in recent reviews and that should be handled before the commercial appetite for AI applications in healthcare introduces significant risks to the users and the patients. Three main challenges are affecting the adoption of AI-based solutions for patient safety, which are: 1) issues related to increasing the robustness of AI-based systems, 2) reporting standards to enable assessment of performance of AI models, and applications 3) acceptance challenges that need to be addressed for an institution to adopt these systems.

We have summarized and categorized hereafter the main points regarding these concerns and challenges.

- **Most studies are still preclinical**

  Most of the studies concerning models or systems powered by AI involve retrospective testing of algorithms on existing data. Only a small percentage of studies are prospective trials of AI-guided systems, and till date even a smaller percentage of AI-powered clinical solutions have got the approval of regulatory agencies (e.g FDA) [3], [104].

  This fact warrants the need for more attention to the characteristics of these systems and to the challenges and additional risks [105] they might be introducing, in order to prevent them and maximize their desired outcome in real-world contexts. In fact, while there is a big commercial pressure to implement AI-based solutions in healthcare, the implementation process should take its time, following two key steps: first, a well conducted and reported development and validation study, and then randomized clinical trials that evaluate usefulness in the real world[104].

- **Model data transparency and validation methods need to be scrutinized**

  A lack of transparency regarding the training data used for model development directly affects the reproducibility, generalizability, and interpretability of a proposed model [106]. Transparency is needed across 3 main categories: the population from which the data were acquired; model design and development, including training data; and model evaluation and validation.

  In fact, the performance of any AI model broadly depends on its reliability and its ability to generalize to the setting and population in which it is applied, rather than its performance represented by the training and test data alone. [107]

An empirical evaluation of 81 studies comparing AI models against clinicians showed major problems with lack of transparency, bias, and unjustified claims, likely because key details about the studies were often missing [104].

- **Existing predictive models reporting frameworks are insufficient for AI-based models**

The TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) framework [108], which is the current standard for reporting predictive models, is useful for standardizing the reporting of research-derived risk scores but presents important limitations when applied to ML models using observational data [3]. New reporting standards are currently being developed to better tackle AI model stability, bias adjustment, and interpretability. Examples include the AI-TREE framework, the MINIMAR framework, as well as an AI extension of TRIPOD [2], [109], [110].

- **Challenges concerning the evaluation of AI derived CDS tools**

Since AI models are often used as CDS in practice, reporting the actual implementation of these tools should adhere to standard approaches and frameworks. One important point to report is if the AI-derived CDS tool was tested side by side with existing tools, if available. Other important key issue for quality-improvement to be reported relates to how the tool was implemented in clinical care. Was it implemented at the point-of-care (for example such as a best-practice alert), or as a CDS in the form of data display (such as a risk dashboard), or was it implemented as a CDS with nudge-embedded support (such as presentation of risk or treatment choice)? [3]

Data assumptions, such as how missing data were handled and approaches to imputing data are also important to report.

- **The problem of result generalizability and adaptability to dataset-shift**

AI-based systems should be externally validated, stemming from the fact that any model that is developed within one dataset will reflect the idiosyncrasies and specificities of that dataset and will thus perform less well when the model is tried in new settings. However, this is not being routinely done in studies [111]. Also, the performance of algorithms can also degrade within the same institution as practices change or as demographics within that site change. Proactive learning algorithms aim to avoid using these types of unstable information. These algorithms proactively avoid learning site-specific biases and are therefore more robust when moved between institutions and when dataset shift occurs [112].

- **The problem of uncertainty of predictions**

A prediction can be uncertain (thus unreliable) for several reasons. For example, the learning model may not have had exposure to enough samples like the sample in the development data set (model uncertainty). Alternatively, the target outcomes that the model is predicting may be noisy (due to incomplete or uncertain input data) [3]. For example, there may be value in calculating and reporting the number of events used in the dataset used for algorithm learning phase per number of predictors, if only to give the reader a sense of model coverage and bias.

Information about uncertainty of predictions is very important for the user, as it might impact his decisions, as well as for the general reliability of the system. Bayesian inference is a common approach for obtaining uncertainty estimates, and new alternatives yield audit tools that can help determine both model and data uncertainty [113] .

- **Explainability and clinical usefulness of results**

CDS is generally considered beneficial when providers receive suggestions at specific times in their workflow and in ways that help them make better decisions[114]. Furthermore, it is critical to assess if the suggestion interrupts the clinician in his or her workflow. Generally, the goal is only to advise the clinician, including giving an assessment about uncertainty regarding a prediction, but leaving up to him/her the choice.

Decisions made by AI (and specifically by machine learning based systems) are often seen by clinicians as opaque due to the “black box” nature of the patterns derived by these techniques. Therefore, a principle of “explainability” is emerging as a requirement for adoption of AI solutions where it is based on the ethical and deontological requirement that humans affected by the decisions suggested by the AI solutions should be able to get an explanation why the decision is made in terms of language they can understand and they should be able to challenge the decision with reasoned arguments. The idea of explainable AI is that humans can understand how a CDS system has produced an outcome. This, however, without limiting the level of complexity of the algorithm, and with that negating the possible benefits of using AI. In some clinical applications, it might not be necessary to understand the exact details of the algorithm, but rather to have a sort of insight into factors that are important or decisive related to a specific prediction. What machine learning algorithms do is learn to assign weights to features in the data, in order to make optimal predictions based on that data. For clinicians, it is important to know which features are considered relevant by the algorithm and how much weight is assigned to this feature. Having that information, a clinician can judge whether the features that a CDS system picks out are indeed relevant or not. Using a system that

formalizes aspects of the reasoning process and explicates the factors that are combined, and with what weight, will support clinicians in developing their ability to articulate and justify their own reasoning process. [115]

- **Adoption barriers of AI-based CDS by clinicians and hospitals**

A number of CDS adoption barriers by clinicians and hospitals have been identified in recent studies, and should be taken into consideration in any design and implementation of such solutions: [116], [117]

Fragmented workflows: CDS systems can disrupt clinician workflow and yield decreased adoption, especially in the case of stand-alone systems. CDS systems disrupt workflow if designed without taking into consideration the "human factors" approach. Disrupted workflow can lead to increased cognitive effort, more time required to complete tasks, and less time face-to-face with patient, even if they are integrated within existing information systems. Studies have found that practitioners with more experiential knowledge are less likely to use, and more likely to override CDSS.

Alert fatigue and inappropriate alerts: Studies have found that the majority of CDS systems alerts can be inconsequential, and often physicians tend to disagree with or distrust alerts, or even simply override them. Also, if physicians are presented with excessive/unimportant alerts, they can suffer from alert fatigue.

CDS systems may be dependent on computer literacy. Lack of technological proficiency can be hindering when engaging with a CDSS. This can vary by the design details of the CDSS, but some have been found to be overly complex, relying too much on user skill.

System and content maintenance: Maintenance of CDS systems is an important but often neglected part of the CDS system life-cycle. This includes technical maintenance of systems, applications and databases that power the CDSS. Another challenge is the maintenance of knowledge-base and its rules, which must keep with the fast-changing nature of medical practice and clinical guidelines.

Trust in system accuracy and decision process: The accuracy of the content/predictions is identified as an important theme for clinicians to trust the CDS system. The uncertainty felt by clinicians about the quality and accuracy of evidence in addition to lack of explainability of the outcomes negatively impact their will to adopt the CDS solution.

Lack of transportability and interoperability Despite ongoing development for the better part of three decades, CDS systems (and even EHRs in general) suffer from interoperability issues. Many CDSS exist as cumbersome stand-alone systems, or exist in a system that cannot communicate effectively with other systems. Positively, interoperability standards are continuously being developed and improved, such as Health Level 7 (HL7) and Fast Healthcare Interoperability Resources (FHIR). These

are already being utilized in commercial EHR vendors. Several government agencies, medical organizations and informatics bodies are actively supporting and some even mandating the use of these interoperability standards in health systems.

2.2.3. Recent frameworks and practical recommendations for the design, evaluation and validation of AI-based models and CDS systems

Stemming from all the aspects and challenges described above, relative to AI-enabled CDS design and implementation, we select a bundle of recently published synthetic and practical conceptual frameworks and recommendations that can tackle the aforementioned challenges, and guide the efficient elaboration of such models and systems.

In a recent commentary by a number of the biggest names in the domain of healthcare CDS [1], a synthetic circular model depicting the major themes around the elaboration of CDS (Figure 1) has been suggested and explained. This model can serve as a reference for tackling the design challenges of AI-based models and CDS systems and for building a CDS design, evaluation and validation framework.

As for the reporting frameworks concerning the ML-based predictive models, Figure 2 shows an example list of a minimal set of reporting variables important to guarantee transparency and minimal bias, adapted from Hernandez-Boussard et al. [2]. Table 2, adapted from a study by Bates et al. [3] depicts a list of recommendations and minimal reporting themes relative to the design of studies and interventions involving ML and other AI techniques, and can serve a practical checklist for model design check and reporting methodologies fine tuning.

Other tools are aiming to help users and buyers of CDS solutions to evaluate the "usability" of these products. For example, the "Trust and Value Checklist" proposed by Silcox et al. regarding choice and adoption of AI-enabled CDS systems can be a practical tool to use to check the main issues related to adoption by clinicians and hospitals of AI-enabled CDS systems (Table 3).

Another recent example is the Medical Digital Score Solution[118], elaborated with the help of 130 French publishers of e-health solutions, health establishments, doctors, health authorities' recommendations and patient associations. This tool aims to help professionals and healthcare institutions in the evaluation of the clinical quality and relevance of eHealth Apps through 26 questions covering the following themes: the medical specialty and claims of the clinical application, the target users of the solution, the clinical evaluation of the solution, the editor (software publisher) and the likelihood of obtaining reimbursement.

Figure 1: Synthetic circular model depicting the major themes around the elaboration of CDS (Greenes et al., 2018 [1])

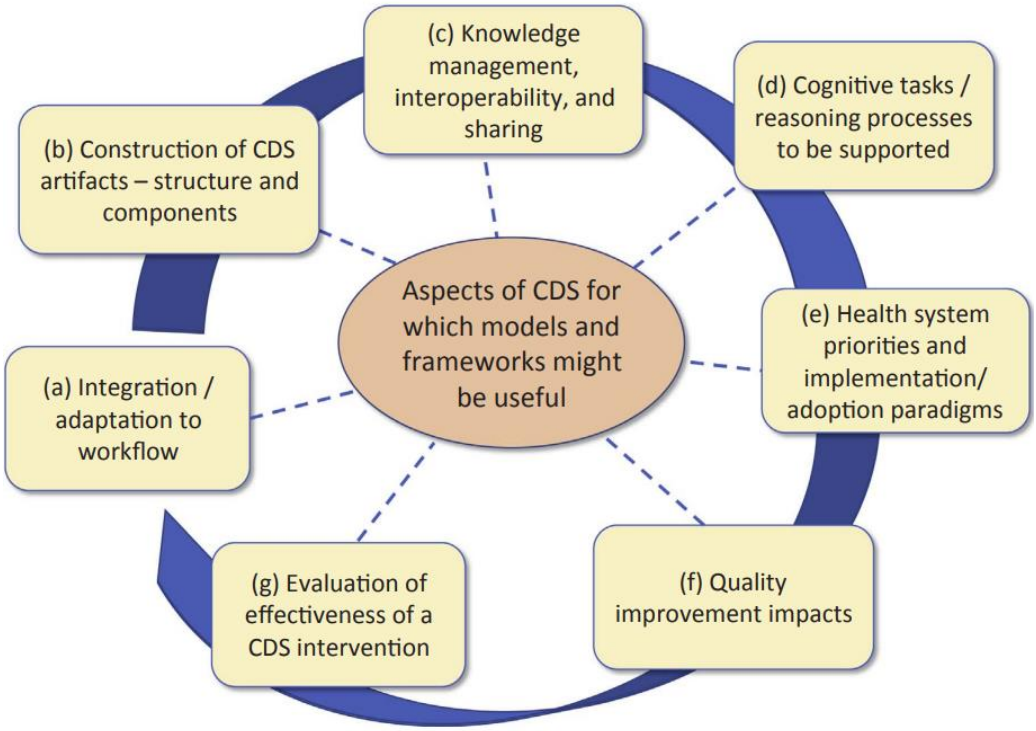


Figure 2: MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care (adapted from Hernandez-Boussard et al., 2020 [2])

Features	Description	Example[23]
1. Study population and setting		
Population	Population from which study sample was drawn	Patients undergoing elective surgery
Study setting	The setting in which the study was conducted (eg, academic medical left, community healthcare system, rural healthcare clinic)	U.S. academic, tertiary care hospital
Data source	The source from which data were collected	EHRs
Cohort selection	Exclusion/inclusion criteria	Adult patients; Patients were excluded if they died during hospitalization.
2. Patient demographic characteristics		
Age	Age of patients included in the study	Mean 58.34 y
Sex	Sex breakdown of study cohort	Female: 73.0%
		Male: 27.0%
Race	Race characteristics of patients included in the study	White: 69.0%
		Black: 3.1%
		Asian: 5.9%
Ethnicity	Ethnicity breakdown of patients included in the study	Hispanic: 13.2%
Socioeconomic status	A measure or proxy measure of the socioeconomic status of patients included in the study	Private: 31.9%
		Medicare: 47.8%
		Medicaid: 11.7%
3. Model architecture		
Model output	The computed result of the model	Postoperative pain scores
Target user	The indented user of the model output (eg, clinician, hospital management team, insurance company)	Risks scores produced by the model will be used by the hospital team for pain management
Data splitting	How data were split for training, testing, and validation	10-fold cross-validation
Gold standard	Labeled data used to train and test the model	100 manually annotated clinical notes and pain scores recorded in EHR
Model task	Classification or prediction	Prediction
Model architecture	Algorithm type (eg, machine learning, deep learning, etc.)	ElasticNet regularized regression
Features	List of variables used in the model and how they were used in the model in terms of categories or transformation	65 predictive features including age, race, ethnicity, sex, insurance type (as public and private) and preoperative pain (log transformation was applied)
Missingness	How missingness was addressed: reported, imputed, or corrected	Missing data were imputed using median of the variable distribution
4. Model evaluation		
Optimization	Model or parameter tuning applied	Generated vectors with a dimension of 300 and a window size of 5
Internal model validation	Study internal validation	Internal 10-fold cross-validation
External model validation	External validation using data from another setting	Not performed
Transparency	How code and data are shared with the community.	Code and sample data available via GitHub

Table 2: Recommendations for Study Design and Conduct and Reporting in Interventions Involving Machine Learning and Artificial Intelligence (adapted from Bates et al., 2020 [3])

Issue	**Recommendations for Study Design and Conduct**	**Recommendations: Reporting**
Validation	Validate performance in a separate data set, ideally from another site, and consider using a proactive algorithm accounting for differences between sites	Describe how the model was validated in a separate data set and how any other additional sites were chosen Describe whether a proactive algorithm was used to account for differences between the sites
	Describe any existing models that have been developed to predict the same outcome	Report how the model performed against any existing models
	Measure accuracy by using multiple segments of a data set	Describe how the model performed in the main segments of the data set
Uncertainty	Use approaches to let users know when specific predictions are more uncertain than others	Report whether the certainty of predictions was displayed to users and how this was done
Implementation	Use standard approaches for introducing CDS	Report which standard approaches for CDS introduction were used and the extent to which they were followed
Data	Determine which variables will be taken from standardized sources versus unstructured data or free text	Report which variables came from standardized sources vs. unstructured data or free text
	Make explicit data assumptions regarding missing data and censoring	Describe how missing data were addressed and what if any censoring was done
	Define how the study sample was obtained and the extent to which it was socio-demographically diverse	Report how the study sample was selected and the extent to which it was socio-demographically diverse

Table 3: The Clinician's AI-enabled CDS Software "Trust and Value Checklist" (adapted from Silcox et al., 2020 [4])

The Value of the CDS System
1- Does use of this AI- enabled CDS software or system yield clinically important improvement compared to the status quo?
2- Will use of this AI-enabled CDS software fit into my clinical workflow and/or my team's workflow? • Does this tool make sense within my current workflow? • Does it make recommendations in a timely manner? • Does it require me to manually enter information, move to another screen, or otherwise make additional "clicks"? • If so, does the clinical benefit warrant this additional activity and possible annoyance?
3- What information does the AI-enabled CDS software provide at the point of use about the logic behind the decisions or recommendations that it produces and the degree of certainty that these recommendations are correct?
The Data Behind the CDS System
4- What was the source of the data used to develop and train the AI-enabled CDS software?
5- How were the training data labeled?
6- Does this AI-enabled CDS software appropriately respect patient privacy?
Testing
7- Does this AI-enabled CDS software fall under the FDA's regulatory authority, and, if so, has it been cleared or approved by the agency?
8- If a software system is not under the FDA's authority, how was it tested, against what standard, and by whom (other than by its developers)?
9- Has this AI-enabled CDS software been tested on data from my own hospital or health system?
Maintaining and Improving the Software Over Time
10-What provisions has the developer made for monitoring and updating the AI-enabled CDS software over time to account for potential degradation in performance?

3. Preliminary study:

Automated detection of patient harm: implementation and prospective evaluation of a real-time broad-spectrum surveillance application in a hospital with limited resources

3.1. Introduction:

The detection and understanding of patient harm to improve patient safety and quality of care has become a top priority in healthcare recently, with healthcare systems around the globe attempting to mobilize resources for the prevention of such harms.

Although it has been over two decades since the eye-opening 1999 Institute of Medicine (IOM) report "To Err is Human"[22], which served to lay the foundations for the subsequent patient safety efforts to come, patient adverse events (AE) remain to be one of the leading causes of morbidity and mortality in healthcare [14]. Inpatient harm has also other negative consequences including increased costs, extended hospital stays and higher readmission rates [15]–[19]. It is estimated that 10-15% of healthcare expenditure is consumed by the direct sequelae of healthcare-related patient harms [20].

One of the factors that can explain this is the lack of accurate measurement and regular monitoring of AE rates, both at national and institutional levels [13]. Hence, stemming from the "one cannot improve what one cannot measure" principle, it is crucial that healthcare systems measure AE levels timely, efficiently, reliably, and regularly in order to deploy appropriate actions to lower the level of harm.

Several tools have been developed in the literature to help detect AEs. These include morbidity and mortality conferences, autopsies, malpractice claims analysis, error reporting systems, administrative data analysis, chart review, and clinical surveillance through patient safety indicators[27]. However, all these methods are retrospective, limited in matter of detection

scope, and resource intensive. In fact, voluntary reporting systems have been estimated to detect only 2-8% of actual AEs occurring during hospitalizations [28][29].

The latest recommended "gold standard" approach, currently used in many hospitals worldwide, is to perform a review of patient medical records, using a standardized approach such as the Institute for Healthcare Improvement's Global Trigger Tool(GTT) methodology[30]. Although this method has shown to have a higher sensitivity versus other AE detection methods, its practical disadvantages include its resource-intensiveness due to time and personnel required, limitations in detection associated with the quality of clinical documentation, in addition to low inter-rater reliability, which are all limiting factors in its adoption[32], [33]. Recent research has suggested that automated AE detection methods using "data mining" techniques are showing to be superior to manual tools [34], [35], [66] allowing for healthcare professionals to provide timely feedback and safety interventions [37]. Indeed, by design, automated AE detection methods can consistently screen large numbers of patients in real time to save valuable resources, something which would be extremely tedious if done manually by reviewers with the same accuracy [66].

Efforts to automate the detection of AEs have been driven by the increase in the adoption of electronic medical records (EMR) worldwide over the last decade [38], with some incorporating a modified automated GTT methodology based on laboratory values, ICD-10 diagnosis codes and text mining of clinical documentation[32][39]. Since the safety benefits from EMRs have not been reaped as much as expected [40][41], recent reports indicate that investments should now be focused on developing automated detection methods using EMR data, to routinely and continuously measure the frequency and types of patient harm [42].While most implemented automated AE systems detect only certain categories of AEs, such as hospital acquired infections, patient falls or adverse drug events [43]–[47], very few [48]–[50] offer the surveillance of a broad spectrum of AEs, and those who did were built on top of mature and complex EMR systems, not accessible to hospitals with limited resources.

The objective of this contribution is to prospectively validate and implement an application that fully automates the detection of broad categories of hospital AEs extracted from a basic hospital information system. This validation involves the measurement of the Positive Predictive Value (PPV) of the tool, as well as an estimated "detection sensitivity" by harm category, a metric

which, to the best of our knowledge, is not commonly observed in the validation studies of similar systems. This application is an enhanced version of a previously validated non-AI based method[124].

3.2.Materials and Methods

Study Design and Setting:

This is a prospective single-site observational study conducted in a 250 bed university hospital localized in Beirut (Lebanon), aiming to evaluate the implementation of an automated detection system built to identify patients with hospital acquired AEs.

Data extracted in the study ranged from October 2019 till June 2020. All hospital admissions during this period were included, with the exception of outpatient and 1-day non interventional hospitalizations.

This study was approved by the hospital's IRB and exempt from full-board review, due to the absence of human subjects research.

Data sources/ Measurements:

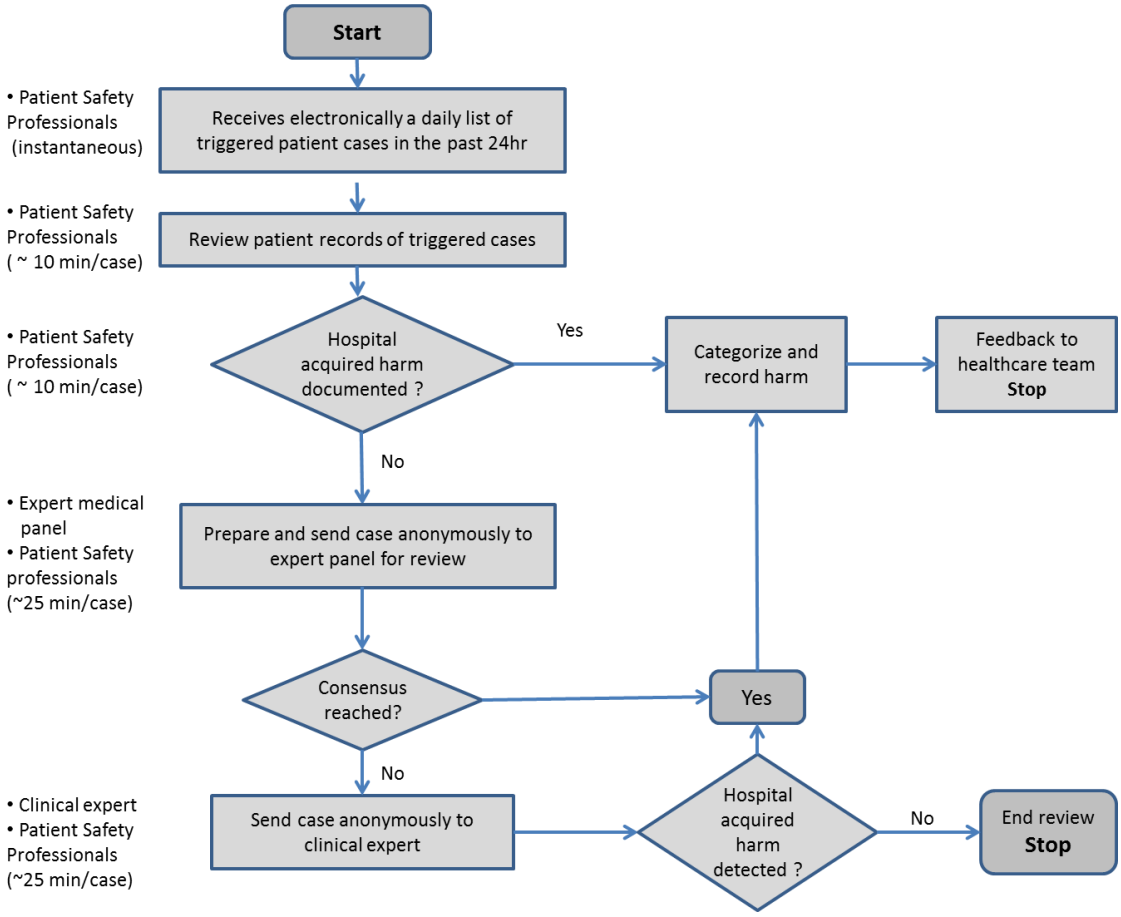
The application data was collected/refreshed in near real time (automatic run every 30 minutes) and comprised of the extraction of 14 indicative patient harm triggers, querying various databases of the Hospital Information System (HIS). The HIS is a basic information system which can be classified as stage 1 as per the Healthcare Information and Management Systems Society's (HIMSS) Electronic Medical Record Adoption Model (EMRAM). The team also receives a daily electronic report of all the triggered cases in the past 24 hours, classified per trigger type, as well as a push-notification of new results every 30 min.

The choice of triggers was derived from the Institute for Healthcare Improvement's Global Trigger Tools methodology, with specific inclusion/exclusion criteria and algorithms for each trigger developed in a previous study [124]. Such triggers include readmissions within 30-days, positive microbiological cultures, critical laboratory values, urgent diagnostic imaging and specific medications orders. They were selected to be indicative of a wide array of possible adverse event categories. Thus the system was intended to function as an automated broad-spectrum active adverse-event surveillance system.

A clinical validation of a representative sample of the results yielded by each trigger from the software, through clinical chart review methodology was performed. The sample for each trigger was drawn randomly during the data collection period, until a sample number relative to a confidence interval (CI) of 95% and an accuracy of 10% was achieved. For the microbiological culture triggers, the data collection period was extended up to December 2020 until the above-mentioned accuracy objective was reached.

Validating if the patient was subject to an AE consisted of a two-phase process as shown in Figure 3. First, a team of two patient safety professionals and a nurse reviewed systematically all cases identified by the triggers, and analyzed the patient file for any documented occurrence in the medical or nursing files of any associated patient harm, and when this evidence was available, this team validated the result of the trigger and classified the associated harm according to the Florida Hospital AE/Harm Classification[125][126]. For all cases where harm occurrence documentation was not available, validation moved to the second phase, where the team would submit the cases anonymously to a multidisciplinary consensus expert panel comprising, in addition to the team members, three experienced physicians (a surgeon, an internal medicine physician and a radiologist) for final validation. The panel was expanded when needed to include healthcare professionals from other specialized disciplines, depending on the expertise needed. All members would vote anonymously and the result for each trigger would be reached via panel consensus. Consensus was defined as the majority of panel members reaching the same vote. If no consensus was reached, the case would be submitted anonymously to a clinical expert opinion, deemed fit to review the case based on expertise and relevance of specialty. A re-vote would be conducted after briefing the expert's opinion. The core members of the panel remained the same throughout the study.

Figure 3: Data collection and triggered cases review process flowchart



The information was collected using an electronic data sheet with the following entries: trigger date, trigger type, AE occurrence status, patient name, age, medical file number, case number, admission and discharge dates, treating physician, case description, harm category, and level of harm according to the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) index [127]. Data collection was performed without any discrimination of sex, nationality, or socioeconomic status.

Outcome variables

The Positive Predictive Value (PPV) was calculated for each trigger as the ratio of detected AEs, relative to the total case alerts generated by the trigger.

Since the measurement of the system's sensitivity requires knowing the true incidence of patient harm in the institution through a thorough review of all inpatient files during the study period,

thus necessitating the investment of resources beyond the capabilities of the institution, an alternate method to approximate sensitivity was adopted. The sensitivity level was estimated by the ratio of the system's detected AEs per harm subcategory, to an estimated total number of AE cases for that subcategory. This estimate was extrapolated from incidence studies in the literature targeting the corresponding subcategory [128]–[145] after the application to the corresponding included hospital population (Table 7).

Statistical association between the occurrence of AEs and a number of demographic variables, namely age, sex, length of stay (LOS) was also conducted (Table 3). The tests were based on the comparison of these variables between the population with identified AEs and a representative sample of the overall included population.

Definitions:

Patient harm is defined as "death, temporary or permanent impairment of body function/structure requiring intervention"[14]

An *Adverse Event* is defined as any injury caused by medical management rather than by the underlying disease or condition of the patient and does imply harm [22].

A *trigger* is used as a 'flag' for potential harm, but is not a harm in itself. For example, an abnormal laboratory result, a medication antidote or escalation to a higher level of care [30].

Active surveillance is the review of the medical records while the patient is hospitalized, rather than retrospectively following discharge [146].

Statistical analysis:

Data analysis was performed using Addinsoft (2021) XLSTAT version 2021.4 statistical and data analysis solution, New York, USA (www.xlstat.com).

Demographic data are presented as means. Statistical significance was defined when the P value < 0.05. As the populations were verified not to have a normal distribution through a d'Agostino-Pearson goodness-of-fit test, a Wilcoxon- Mann Whitney rank sum test was preformed to examine if the differences in the mean age and length of stay (LOS) for those who were subject to an AE vs. those who were not, were statistically significant.

A chi-square test of independence was performed to examine the relationship between the variables sex and the occurrence of an AE.

3.3. Results

The data study period included a total of 8,760 admissions, corresponding to 9 months of daily active surveillance. After the application of inclusion and exclusion criteria for each trigger, the system yielded 946 cases that were subject to review (Figure 3). Of these 946 triggered cases, 394 were identified as AEs, occurring with 291 patients, yielding an overall PPV for the surveillance system of 42%, and an average of 1.4 AEs identified per surveillance day. 83% of patients were subject to one AE during their stay, 11% subject to two AEs, and 6% were subject to at least three AEs.

On average, 10 minutes were required to discuss each case by the team, and 25 minutes to discuss the case with the consensus panel. For all triggered cases, the primary data collection for case investigation was performed while the patients were still hospitalized.
The mean age of patients with an AE showed to be slightly higher than the mean age of the overall included patient population, M=62.7 years [SD=22.5] vs. M=59.4 years [SD=23.9] respectively, with fair evidence of results (p-value <0.1), where men were significantly more likely to be subject to an AE than women (X^2 (1, N = 291) = 10.5, with moderate evidence *p-value* <0.05)

Furthermore, the mean LOS of patients with an AE showed to be significantly higher than the mean LOS of the overall included patient population; with 42.1 days [41.8] vs. 8.7 days [11.8], respectively, with high evidence p-value <0.001. These results are summarized in Table 4.

Table 4: Demographic data of reviewed cases and corresponding statistical analyses

Variable	Overall population sample	Population with identified Adverse Events	Statistical Test and *p* value
Age	Mean =59.4 yrs SD =23.9 95% CI, 58.7- 60.0 (n=3007)	Mean =62.7yrs SD =22.5 95% CI, 60.2- 65.2 (n=394)	*Wilcoxon- Mann Whitney rank sum test* $p < 0.1$
Sex Men Women	 n=2695 n=2761	 228 165	 *Chi-Square* $p < 0.01$
LOS	8.7 days SD =11.8 95% CI, 8.4 -9.0 (n=3007)	42.1 days SD =41.8 95% CI, 37.5- 46.7 (n=394)	*Wilcoxon- Mann Whitney rank sum test* $p < 0.001$

LOS: Length of Stay; SD: Standard Deviation; n: sample number; CI: Confidence Interval

Results also show variability in the PPV among the five trigger categories/modules (Table 5). The highest average PPV was shown to be for the Healthcare Associated Infections Module with a PPV of 84.6%, whereas for the Medications, Radiology, Laboratory and Admission/Discharge Modules, the PPVs were 68.3 %, 42.4%, 18.5% and 11.2 %, respectively (Table 5).

Table 5: List of automated triggers and their Positive Predictive Values (PPV)

CATEGORY/ MODULE	TRIGGER	NUMBER OF CASES	NUMBER OF AEs	POSITIVE PREDICTIVE VALUE
Healthcare Associated Infections	Positive respiratory cultures (VAP detection algorithm)	69	57	82.6%
	Positive urinary cultures (CAUTI detection algorithm)	64	59	92.1%
	Positive blood cultures (CLABSI detection algorithm)	58	47	81.0%
	Clostridium Difficile Toxin A & B result	10	7	70.0%
	AVERAGE	**201**	**170**	**84.6%**
Admission/ Discharge Module	Readmissions within 30 days	419	47	**11.2%**
Laboratory Module	Hb drop of more than 25%	30	18	60.0%
	Creatinine raise (> 2 x baseline or raise of more than 0.3 in last 72 hrs)	55	29	52.7%
	INR> 6	37	6	16.2%
	AVERAGE	**541**	**100**	**18.5%**
Radiology Module	Doppler Ultrasound-Upper Limbs	23	16	70.0%
	Urgent CT or MRI exam ordered more than 48 h of admission	14	6	42.9%
	Urgent radiology exam ordered more than 4 days after admission	22	3	13.6%
	AVERAGE	**59**	**25**	**42.4%**
Medications Module	Pressure ulcer therapies order	116	88	75.9%
	High dose anticoagulants order	16	8	50.0%
	Anticoagulants reversal agents order	13	3	23.1%
	AVERAGE	**145**	**99**	**68.3%**
TOTAL		**946**	**394**	**41.7%**

AE: Adverse Event; CAUTI: Catheter Associated Urinary Tract Infection; CLABSI: Central line Associated Bloodstream Infection; CT: Computed Tomography; Hb: Hemoglobin; INR: International Normalized Ratio; MRI: Magnetic Resonance Imaging; VAP: Ventilator Associated Pneumonia

Table 6 shows the distribution of the severity of the detected AE, where the majority (78%) of the detected AE fell under Category F, followed by 18.5% under Category E, which is considered as an elevated level of harm.

Table 6: Distribution of the severity of detected AEs using the NCC MERP index

Level of Harm (NCC MERP categories)	Number of AEs	% of AEs
Category A-Circumstances or events that have the capacity to cause error	5	1.3
Category D- An error occurred that reached the patient and required monitoring to confirm that it resulted in no harm to the patient and/or required intervention to preclude harm	3	0.8
Category E- An error occurred that may have contributed to or resulted in temporary harm to the patient and required intervention	73	18.5
Category F- An error occurred that may have contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization	309	78.4
Category I- An error occurred that may have contributed to or resulted in the patient's death	4	1.0
TOTAL	**394**	**100**

AE: Adverse Event; NCC MERP: National Coordinating Council for Medication Error Reporting and Prevention

The estimated sensitivity of the systems in detecting AEs varied among the four harm categories (Table 7). The estimated detection sensitivity relative to Hospital Acquired Infections was the highest (Ventilator Associated Pneumonia 80%, Central Line Associated Bloodstream Infection 75-96%, Catheter Associated Urinary Tract Infection 60-70%, Surgical Site Infection 40%), followed by events related to surgery or other procedures (13.3-36.3%), events related to patient care (4.9-44.2%) and events related to medications (5.7-21.1%).Some AE categories commonly found in hospitals, such as inpatient falls, and medication related constipation were not detected.

Table 7: Estimated sensitivity through comparison of detected cases per AE category to applied literature incidence ranges

AE/Harm Categories*	Target population	Incidence range in literature	No. of patients in target population	Estimated range of cases based on literature incidence	No. of detected cases	Estimated Detection Sensitivity	Triggers used (number of detected cases)
Events Related to Medications/ Intravenous (IV) Fluids							
Clostridium difficile medication associated infection	All inpatient admissions (excluding 1-day)	[0.73%-1.35%] of patient admissions	4238	30-57 (mean=43.5)	7	16.1%	*Clostridium Difficile* Toxin A & B (5) Readmissions within 30 days (1)
Kidney damage due to contrast dye	Diagnostic or interventional coronary angiography admissions	[5-7.5%] of diagnostic or interventional coronary angiography admissions	1127	56-84 (mean=70.0)	4	5.7%	Creatinine raise [> 2 x baseline or raise of more than 0.3 in last 72 hrs] (4)
Medication related renal insufficiency	All inpatient admissions	[0.5-1.25%] of patient admissions	8760	43-109 (mean=76)	16	21.1%	Creatinine raise [> 2 x baseline or raise of more than 0.3 in last 72 hrs] (16)
Medication related bleeding	All inpatients on high risk oral and IV anticoagulants	[6% to 10.2%] of patients on anticoagulants	2073	124-211 (mean=167.5)	10	5.9%	INR critical value (6) Use of anticoagulant reversal agents (3) Hb drop more than 25% (1)

Table 7: Estimated sensitivity through comparison of detected cases per AE category to applied literature incidence ranges (continued)

AE/Harm Categories*	Target population	Incidence range in literature	No. of patients in target population	Estimated range of cases based on literature incidence	No. of detected cases	Estimated Detection Sensitivity	Triggers used (number of detected cases)
Events Related to Patient Care							
DVT/VTE	All inpatient admissions (excluding 1-day)	[0.9% -1.3%] of patient admissions	4238	38-55 (mean=46.5)	7	15.0%	High dose anticoagulants (7)
Hospital acquired pressure ulcer	All inpatient admissions (excluding 1-day)	[1.4-4.5%] of medical and surgical admissions (non-ICU) [8.8-10.3%] of ICU admissions	4238 non-ICU 792 ICU	129-269 (mean=199)	88	44.2%	Use of pressure ulcer treatments (88)
Phlebitis Grade 3 and above (Visual Infusion Phlebitis Score)	All inpatient admissions	[1-6.4%] (Grade 3 and 4 phlebitis)	8760	87-556 (mean=321.5)	16	4.9 %	Doppler Ultrasound-Upper Limbs (16)

Table 7: Estimated sensitivity through comparison of detected cases per AE category to applied literature incidence ranges (continued)

AE/Harm Categories*	Target population	Incidence range in literature	No. of patients in target population	Estimated range of cases based on literature incidence	No. of detected cases	Estimated Detection Sensitivity	Triggers used (number of detected cases)
Hospital Acquired Infections							
Catheter Associated Urinary Tract Infection (CAUTI)	All inpatient admissions with inserted urinary catheters (excluding 1-day)	**Non ICU** [1.6-2.7/1000 device days] **ICU** [4.7-4.9/1000 device days]	Catheter days: 23,292 Catheter days: 8,352	37.2-62.8 (mean= 50) 39.2-40.9 (mean = 40)	59	65.5%	Positive urinary cultures (59)
Central Line Associated Blood Stream Infection (CLABSI)	All inpatient admissions with central lines (excluding 1-day)	**Non ICU** [4.1-4.3/1000 device days] **ICU** [4.1-4.3/1000 device days]	Catheter days: 7,851 Catheter days: 5319	32.2-33.7 (mean=32.9) 21.8-22.9 (mean=22.3)	47	85.1%	Positive blood cultures (47)
Surgical Site Infection (SSI)	All surgical admissions	[1.4-2.4%]of surgical admissions	3032	42-72 (mean=57)	23	40.3%	Readmissions within 30days (21) Radiological triggers (1) Hb drop more than 25% (1)
Ventilator Associated Pneumonia (VAP)	All inpatient on ventilators	**ICU** [20.7-27.1/1000 device days]	Ventilation days: 2970	61.5-80.5 (mean=71.0)	57	80%	Positive respiratory cultures (57)

Table 7: Estimated sensitivity through comparison of detected cases per AE category to applied literature incidence ranges (continued)

AE/Harm Categories*	Target population	Incidence range in literature	No. of patients in target population	Estimated range of cases based on literature incidence	No. of detected cases	Estimated Detection Sensitivity	Triggers used (number of detected cases)
Events Related to Surgery or Other Procedures							
Abnormal bleeding following surgery or procedure	All hospital surgical and interventional admissions	[0.82-2.5%]	4242	34.8-106.0 (mean=70.4)	15	21.3%	Readmissions within 30 days (3) Hb drop more than 25% (12)
Blood clots and other occlusions related to surgery or procedure	All hospital surgical and interventional admissions	0.26% of surgical and interventional admissions	4242	11.0	4	36.4%	Urgent radiological exam (1) Readmissions within 30 days (3)
Post-op acute renal failure	All hospital surgical admissions	[0.8-1%] of surgical admissions	3032	24-30.3 (mean=27.2)	9	33.1%	Creatinine raise [> 2 x baseline or raise of more than 0.3 in last 72 hrs] (9)
Respiratory complications related to surgery (pneumonia, respiratory failure, pneumothorax, ARDS, atelectasis, pleural effusion, etc.)	All hospital surgical admissions	[1.8-2.2]% of surgical admissions	3032	54.6-66.7 (mean=60.6)	8	13.2%	Readmissions within 30 days (7) Radiological trigger (1)

ARDS: Acute Respiratory Distress Syndrome; AE: Adverse Event; DVT/VTE: Deep Vein Thrombosis/Venous Thromboembolism; Hemoglobin: Hb; INR: International Normalised Ratio; IV: intra-venous; ICU: Intensive Care Unit

3.4.Discussion

Performance of the surveillance system and benchmarking with other comparable studies

The overall PPV of the system (42%) is comparable to other studies using trigger-tool based automated detection AE systems which have shown to have a median PPV of 40%[66] . Nevertheless, results showed inter and intra-trigger category PPV variability. The highest PPV belonged to the Healthcare Associated Infections triggers with an average PPV of 84.6%, followed by Pressure ulcer treatment (75.9%), Doppler Ultrasound for Upper Limbs (70%), Hemoglobin level drop (60%) and Creatinine raise (52.7%).

As found in other studies [147][148], LOS and age were also associated with a greater risk of AE occurrence.

On an individual trigger level, and compared with other automated trigger tools for the detection of AE[48], [66], a number of triggers showed an improved PPV. For example the mean PPV in the literature for the creatinine raise trigger was found to be approximately 45.2%, the pressure ulcer trigger 53%, the positive urinary culture 35%, and positive blood culture trigger 66.7 %, whereas our surveillance system showed a PPV of 52.7%, 75.9%, 92.1% and 81%, respectively. On the other hand, some triggers showed a lower PPV than that found in the literature[49], [66], [149], [150], such as the INR > 6 trigger (37.8% versus 16.2% in our study) and anticoagulant reversal agents trigger (29.5% versus 23.1% in our study).

For certain triggers, such as HAI triggers, the higher PPV is comparable to findings in the literature [43], [66], and can be partially attributed to the fact such triggers use guideline-based exclusion criteria and algorithms [151], that are applied to a rich set of objective variables in the HIS, thus mimicking the clinician's thought process for confirmation of such AEs without need for additional analysis of the patient case. On the other hand, some triggers (such as readmissions within 30 days and INR > 6) rely in their definition on the evidence of clinical symptoms (i.e fever, loss of consciousness, bleeding, pain etc.) and thus can be affected by the quality of documentation which may contribute towards having lower PPVs. This highlights

furthermore the need for the standardization of triggers and AE definition in the field [66], [152] in order to optimize benchmarking.

With regards to the estimated incidence detection, the tool showed variable levels of "estimated" sensitivity across and within categories, ranging relatively between high (> 50% sensitivity), moderate (between 20-50%) and poor (< 20 % sensitivity). For example, the tool seems to be detecting between 65% and 80% of the most common types of healthcare associated infections. Moderately detected harms include pressure ulcers (44%), blood clots and other occlusions related to surgery or procedure (36%) and post-op acute renal failure (33%). Whereas those poorly detected harms include DVT/VTE (15%), kidney damage due to contrast dye (5.7%), and phlebitis (4.9%). Several factors could explain this variability. In addition to documentation bias, the lack of prescription of confirmatory exams can affect the level of detection. For example, not prescribing systematically serum creatinine post cardiac angioplasty procedures can cause the system to miss certain acute kidney injuries due to contrast dye administration. Standardizing such clinical practices would help address this limitation.

Implementation and incorporation into routine patient safety surveillance and quality improvement

This application has been used daily by our quality and patient safety team for approximately two years, with the objective to lead to tangible outcomes and continuous quality improvement efforts and is currently an integral part of the hospital's AE detection efforts.

Automation of the data collection has reduced the daily time burden of the patient safety team required to collect and prepare data, from approximately two hours [124] to almost five minutes.

The application provided a near real-time triggering of cases through push notifications sent to the safety team, allowing for more timely investigations to be made close to the event date and time. This can maximize the chance of understanding factors that could have lead to the detected AE, that otherwise would be difficult to acquire using classical retrospective approaches. These investigations helped lead to proactive improvement efforts in various domains such as infection control, medication safety and management, nursing and medical practices.

Certain detected AEs were also subject to expert discussions, submitted for mortality and morbidity reviews or the modification of certain protocols. For example, the detection of pressure ulcers lead to the complete update of the pressure injury assessment tools and management protocols involving multi-disciplinary teams that included experts in nursing and plastic surgery.

Despite the increased number of automated AE detection systems being developed in recent years, most are targeting specific types of harms and there is a need to develop and validate systems with a broader scope of AE categories [42], [66]. The results of this study show that our system was able to detect diverse categories of AEs with promising detection performance (Table 4).

Moreover, this tool can have a potential use in monitoring clinical performance with minimal human effort and collection bias. The highly or moderately detected AE categories could even be used as a source of quality key performance indicators, such as 30-day readmissions for a number of standard surgical procedures, acute kidney injury rate, rate of HAIs, etc.

Finally, results have shown that the mean LOS of patients who have experienced an AE is almost five-fold more than the overall population and that the majority of detected AE are classified under severity MCCMERP *Category F*. We believe that collecting this severity index for detected AEs and relating it to associated costs can be useful to showcase to the hospital administration the financial impact of these events along with how targeted safety efforts can help alleviate this burden.

Furthermore, it has been found that at least 60% of hospitals between the US, Europe and Asia still have a basic EMR system[153], the majority being Stages 0-2 according to the HIMSS EMRAM model. The fact that our system stemmed from a basic HIS itself (Stage 1), it renders its results promising for hospitals with basic HIS and limited resources.

Limitations and future directions

In the calculation of the system's detection rate, a methodology based on extrapolated AE incidences from the literature was used instead of the "gold standard" chart review methodology. This choice was due to feasibility reasons and limited resources. However, it could have introduced some inaccuracies associated with the difference between the real

incidences of the AE categories found in the hospital and the benchmark. Moreover, the detection performance of certain triggers was affected by the prescription practice of clinicians as well as the quality of documentation. Increasing the sample number by prolonging the time period or conducting a multicentric study may also improve the results' reliability and generalizability. Furthermore, some triggers rely on the request of certain items that can be subject to change with time. This would require the algorithms used to be regularly reviewed.

Future efforts may be mobilized towards the integration of more advanced technologies and methods, such as text mining [43], [50] and machine learning into hospital AE detection applications[154], although this would require relatively more sophisticated technical resources. When used with standardized healthcare data exchange frameworks [155] for interoperability, such systems can have the potential to connect to broad types of EMR, regardless of their maturity level.

Finally, using the system's results to create live patient safety dashboards displaying the levels of certain hospital AEs can be interesting for hospital administrations to monitor clinical performance. Such advanced approaches may help address some of the previously mentioned limitations.

Conclusion

As the science of patient safety evolves, more tools are being developed for the detection of patient harm, in a more reliable and efficient manner. In this study, an internally developed automated AE detection surveillance application was able to identify AEs across a broad spectrum of harm categories, in a near-real time manner, with an overall 42% positive predictive value. The implementation of the system led to various clinical interventions and quality improvements actions. Such a system could serve as a promising patient safety tool, allowing for timelier, targeted and resource efficient interventions, even for hospitals with limited resources.

4. Contribution 1:

Comparison of Machine Learning Algorithms for Classifying Adverse-Event Related 30-Day Hospital Readmissions: Potential Implications for Patient Safety

4.1. Introduction

Nearly one in five patients is re-hospitalized within 30 days of discharge [156], incurring significant costs to the healthcare system [157]. Therefore, minimizing post-discharge adverse events has become a priority for many health care systems around the globe.

Many studies in the last decade have focused on identifying patients at risk of readmission using predictive models [158], [159]. However, few attempts have been made to identify potentially preventable readmissions [160], and, to the best of our knowledge, no studies have explored models to predict or identify readmissions related to hospital adverse-events. Moreover, study review highlights the need to use more standardized hospital information system (HIS) data related to the readmission (e.g. biological, radiological, billing and administrative data) instead of institution-specific or clinical judgement-based data, for an earlier and more benchmarkable prediction.

The main objective of the study was to construct a model, based on routinely available data from the HIS, that could determine, on a near real time basis, if the patient readmission within 30 days was associated with a hospital acquired adverse event that occurred in the previous admission (response variable).

4.2. Materials and Methods

The dataset used for training and testing of algorithms was built using the gold-standard approach, by a multidisciplinary consensus panel (internal medicine physician, radiologist, nurse, and patient safety professionals), expanded when needed to include physicians from specialized disciplines. The panel analyzed and classified 307 patient readmissions (within 30 days) extracted from the HIS from October 2019 till March 2020 (excluding readmissions of

oncology patients and elective readmissions) that occurred in a 250-bed university hospital in Beirut, Lebanon where the study took place. On average, 30 min were required for every case preparation, and 25 min for the panel discussion and classification of each medical case. 46 of these cases were labeled as related to adverse events, to which 47 cases non-related to adverse events were randomly chosen from the remaining dataset, to define a final balanced dataset of 93 cases, containing almost equal cases for each of the two classification classes: "Readmission related to adverse event", and "Readmissions not related to adverse events".

23 features (explanatory variables) were identified by the panel based on previous research results accomplished on this subject [161]and were extracted from the HIS. The features consist of both binary and continuous variables corresponding to the readmission and the previous admission cases.

A supervised learning approach was adopted using different machine learning algorithms: Random Forests (RF), Decision Tree (DT), Boosting (BT), Artificial Neural Networks (ANN) and Logistic Regression (LR).

Cross-validation of the model generated by each algorithm was performed using established methods (K-Folds=5 and Leave-One-Out) to avoid the results being influenced by the partitioning of the original dataset.

Results of the different algorithms were reported and compared using model classification accuracy, based on the information entropy. Variable importance was determined by calculating the relative influence of each feature on the classification (except for ANN where this method is complex and not standard).

This study was submitted to the hospital's institutional review board (IRB) and was exempt from further review since it does not directly involve human subjects.

4.3. Results

All algorithms showed good accuracy (>0.85) in the model training phase, which underlines their ability to fit data to a theoretical model using the proposed features.

Table 8 shows the accuracy obtained from different algorithms when the predictions of the generated models were compared against the test dataset. Among the different algorithms, ANN showed to be the most predictive (0.88), followed respectively by LR (0.62), RF (0.62), DT (0.60), and BT (0.55).

Table 9 shows the features' different weight importance, when this option was possible by the type of algorithm used, and highlights the features with significant weight (>5%).

Table 8: Accuracy result on evaluation for algorithms tested, with chosen algorithm parameters

Algorithm	**Mean accuracy**	**StdDev of accuracy**	**Parameters** (algorithm-specific function arguments used for result optimization, as per SKLEARN and KERAS libraries definitions)
ANN	0.88	*0.07*	*24/6/4/1 architecture, optimizer RMSprop, lr=0.0001, epochs=10000, batch size=30*
LR	0.62	*0.10*	*penalty='l2', dual=False, tol=0.0001, C=1, fit_intercept=False, intercept_scaling=1, class_weight='balanced', solver='lbfgs', max_iter=30000*
RF	0.62	*0.06*	*n_estimators = 30, criterion="entropy", max_depth=8, min_samples_leaf=2,min_samples_split=4,max_leaf_nodes=15,bootstrap=False*
DT	0.60	*0.06*	*criterion='entropy', max_depth=8, min_samples_leaf=2,min_samples_split=4*
BT	0.55	*0.07*	*XGBClassifier max_depth = 8, learning_rate = 0.001, gamma = 1,min_child_weight = 1.0, n_estimators=200*

Table 9: Feature importance by model (highlighted are those with mean importance ≥ 0.05)

Feature	RF	DT	BT	LR	Mean
Days since last discharge	0.15	0.15	0.25	0.01	0.14
White Blood Cells (WBC) value upon readmission	0.08	0.05	0.06	0	0.0475
Microbiological culture ordered in first 48h of readmission	0.03	0.1	0.08	0.07	0.07
C-Reactive Protein (CRP) value upon readmission	0.04	0	0.11	0	0.0375
Creatinine serum differential ratio	0.05	0.04	0.13	0.06	0.07
Potassium level differential	0.05	0	0.05	0	0.025
White Blood Cells (WBC) level differential	0.05	0.08	0.07	0	0.05
C-Reactive Protein (CRP) level differential	0.1	0.11	0	0	0.0525
Combined WBC-CRP levels	0	0	0	0.11	0.0275
Hemoglobin differential ratio	0.05	0.17	0.04	0.01	0.0675
Calcium level differential	0.03	0	0.06	0.01	0.025
Patient age	0.06	0.08	0	0.04	0.045
Patient sex	0.08	0	0.08	0	0.04
Paracetamol use on readmission	0	0	0	0.09	0.0225
Number of biological exams ordered in first 24h of readmission	0.07	0	0.04	0.01	0.03
Number of radiological exams ordered in first 24h of readmission	0.04	0	0.04	0.01	0.0225
Remarks in radiological requests ordered in first 24h of readmission	0.01	0	0	0.12	0.0325
Number of different types of medication during previous admission	0.06	0.06	0.05	0	0.0425
International Normalized Ratio (INR) raise	0.1	0.12	0	0.12	0.085
Number of surgical procedures performed in previous admission	0.04	0	0.11	0.04	0.0475
Cerebral CT-Scan performed on readmission	0	0	0	0.07	0.0175
Abdominal CT-Scan performed on readmission	0	0	0	0.06	0.015
Thoracic CT-Scan performed on readmission	0.03	0.06	0	0.11	0.05
Readmission type (Inpatient/ Outpatient)	0	0	0	0.08	0.02

4.4. Discussion and Conclusion

4.4.1. Predictive performance of the different algorithms

While accuracy levels across different algorithms seem to be moderate to low, they remain comparable to studies in the literature on 30-days readmissions [159], [162] . However, the significance of the features adopted in this study versus other models in the literature is that they permit a near real-time computation of the classification as soon as the patient is readmitted, thus allowing the possibility for immediate proactive administrative and/or clinical interventions to reduce the risk of any preventable adverse event. A few factors could potentially explain and lead to improving this result. First, the sample size can be improved with additional expert time and resources. Another factor is inherent to the nature of the classification outcome variable itself which can only be built on expert opinion and thus contains, despite all methods used to lower the risk of judgement bias, a residual level of uncertainty. Finally, the high number of harm categories in the medical field [31], and patient-specific influencing factors that should be taken into consideration, induce a need for a high number of features to encompass this domain's complexity.

4.4.2. Feature importance and possible interpretations

In contrast with other similar studies, the features were chosen in this study to be all directly extractable from a basic HIS, and not needing human intervention for data aggregation or interpretation. This choice is in line with the need to standardize such tools and benchmark results across different healthcare systems.

In the majority of tested models, the features impacting most of the results are: Days since last discharge, respective differential of CRP/WBC/INR/Creatinine serum, hemoglobin differential ratio, Thoracic CT-Scanner and Microbiological Cultures performed upon readmission. This result can be intuitively interpreted as a higher sensitivity towards detection of infections, hemorrhages and acute kidney injury cases. Interestingly, from the results obtained, the models insinuate also that readmissions occurring within 12 days of previous discharge are more prone to be associated with patient harm.

4.4.3. Potential practical implications for patient safety

Validating automated models for classifying 30-day readmissions can have some important implications for patient safety efforts in hospitals. Firstly, through correctly estimating the true level of nosocomial harm relative to readmissions, and using this information to analyze and improve clinical practices, hospitals will be able to measure the impact of deployed patient

safety efforts over time. Secondly, permitting a proactive management of such cases as soon as they enter the hospital, can help prevent any further harm and address any implications that may arise. Finally, the results can pave the way to more proactive models that can predict risks of preventable readmissions due to adverse-events before patients are physically discharged from the hospital, or identify the patients "at risk" and follow up with them by phone before they actually return to the hospital, thus preventing extra costs for the healthcare system and third-party payers.

4.4.4. Limitations

Given the time and effort that was needed to construct a training and testing dataset on this domain, the validated sample size used in this study was relatively small. Also, some specific data (such as radiology reports, medical notes) could not be extracted from the HIS at this point of the system's integration. These limitations will be taken into consideration in future studies to improve outcomes.

5. Contribution 2:

Early prediction of all-cause clinical deterioration in general wards patients: development and validation of a biomarker-based machine learning model derived from Rapid Response Team activations

5.1. Introduction

Delays in medical interventions in clinically deteriorating patients have been found to be associated with increased morbidity and mortality[163]–[165]. Therefore, early and continuous detection of gradually worsening patient conditions in hospital wards might allow for more rapid treatments, and thus, improved outcomes[166].

The most common forms of clinical deterioration are respiratory instability, hemodynamic instability, sepsis, bleeding, cardiac decompensation, and acute hepatic/renal failure[167]. Deteriorating patients often require transfer to a higher level of care (such as Intensive Care Units, ICU) and the urgent call for medical and nursing professionals for assessment and interventions.

Studies have documented that clinical signs and symptoms of patient deterioration (such as hypotension, bradycardia, tachypnea, tachycardia, altered level of consciousness, etc.) can be detected as early as six to eight hours before the deterioration event or cardiorespiratory arrest [168].

These findings, derived from the late 1990s, led to the development and wide implementation of specific hospitals Early Warning Systems (EWS) called “track-and-trigger” systems which can help predict clinical deterioration. These systems rely on the periodic observation of selected basic clinical signs (‘tracking’) with predetermined calling or response criteria (‘trigger’) for requesting the attendance of staff who have specific competencies in the management of acute illness and/or critical care [169]. In practice, most of these systems are based on the regular measurement of vital signs [59], that would serve to calculate a paper-based or electronic severity score with predetermined thresholds triggering a call for a rapid response team. This team then evaluates the patient and takes clinical actions to prevent or manage the deterioration[59]. “Track-and-Trigger” systems are currently still considered as the gold standard with regards to detecting and responding to clinical deterioration, and have been shown to increase the number of calls to the rapid response team, decrease the number of cardiac arrests and improve the response time of emergency medical teams [170].

However, these track-and-trigger systems have practical limitations. Firstly, the time from detection to actual deterioration is relatively short (0-8 hours), which provides a small window of opportunity for appropriate interventions that could prevent or mitigate the clinical risks. Secondly, the deterioration prediction score is sensitive to data quality and availability. Thus, any delays, omissions or errors in the measurement of vital signs, which are all human dependent factors, can potentially affect the performance of the deterioration prediction score. Moreover, automated versions of such track-and-trigger systems cannot be effectively implemented in hospitals with basic EMRs (i.e staged as 0, 1 or 2 according to the HIMSS EMRAM adoption model classification [171]), since they do not include an electronic nursing flowsheet documentation module. It is to be noted that the proportion of hospital with such basic EMRs is significant worldwide, especially in third-world countries[172].

A new and promising approach described in recent studies [173]–[177] involves the addition of physiological biomarkers measurements to the traditionally measured vital signs and demographic patient data routinely available in the EMR. Biomarkers are defined as biological characteristics (such as for example the C-Reactive Protein, Procalcitonin, Serum Creatinine, etc.) that are objectively measured and used as indicators of certain physiopathological processes [178]. This approach is based on the hypothesis that changes in certain biomarkers can precede the onset of clinical signs and symptoms, sometimes as early as 48-72 hours [179] [180], theoretically permitting an earlier prediction of deterioration than traditional track-and-trigger systems.

Moreover, most of the prediction models were trained according to cases with the following outcome variables: cardiorespiratory arrest/death/unexpected transfer to ICU. Only few studies [181] [182]have adopted the activation of the RRT as an outcome for the training and validation of the predictive model [94][93], even though such an outcome encompasses a broader and richer perspective of clinical deteriorations. In fact, a significant percentage (almost half) of RRT deterioration cases end with stabilization of patients on wards [183], [184], a clinical scenario otherwise not used by most systems.

Finally, recent studies have shown that machine learning–based early warning systems can achieve greater accuracy than aggregate-weighted early warning systems [93], thus their increased use in the derivation of new models.

The aim of this study is to elaborate and validate a biomarker-based model (without including vital signs data) based on absolute and differential biomarker values for the prediction of general (all-cause) clinical deterioration, using machine-learning (ML) algorithms as a derivation method, and expert-reviewed Rapid Response Team calls as the main outcome for model training and validation. Our hypothesis is that such a model could predict all-cause clinical deterioration earlier than track-and-trigger systems, without the need to use vital signs and other complex patient data (e.g. diagnosis, clinical notes…), thus allowing such an approach to be used in healthcare settings which have even the most basic EMR systems.

Ultimately, this may provide opportunities to intervene earlier, help allocate resources more effectively and potentially improve the patients' health outcomes.

5.2.Materials and Methods

The hospital Institutional Review Board deemed this study as "Exempt" from further review, as it does not directly involve human subjects.

Study Design and Setting

We conducted a retrospective single-institution cohort study of all consecutive adult (>18 years) hospitalized patients in non-critical wards for whom a Rapid Response Team (RRT) was called after 24h of their admission over more than a two years period (1 April 2018 through 30 June 2020).

The study took place in a 250-bed tertiary university hospital in Beirut, Lebanon. The hospital's EMR can be considered as basic (stage 1 as per the Healthcare Information and Management Systems Society's EMRAM model). The system contains admissions/discharge/transfer data, basic ancillaries with limited integration (laboratory, radiology and pharmacy), billing (procedures and consumables), but no electronic nursing or medical documentation, nor computerized physician order entry or clinical decision support applications.

Definitions

We have adopted the following complementary definitions for the clinically deteriorating patient: "one who moves from one clinical state to a worse clinical state which increases their individual risk of morbidity, including organ dysfunction, protracted hospital stay, disability, or death" [185] and "a dynamic state experienced by a patient compromising hemodynamic stability, marked by physiological decompensation accompanied by subjective or objective findings" [186].

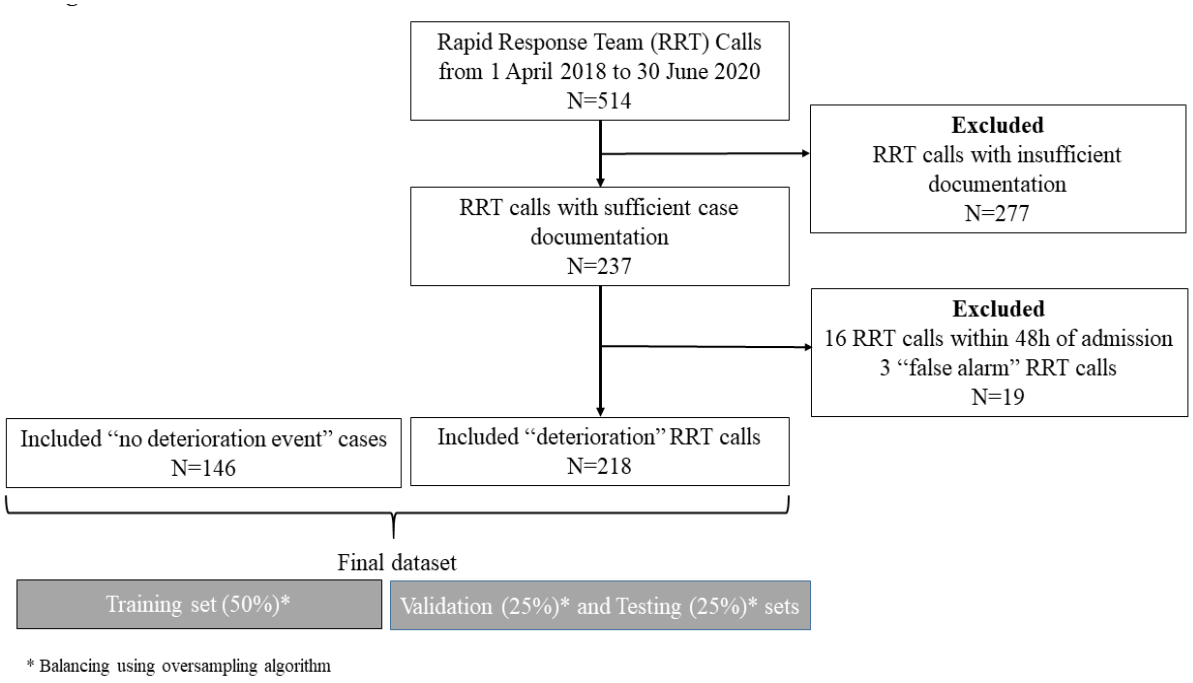
Data collection

A multidisciplinary expert consensus panel (an internal medicine physician, a nurse, two patient safety professionals, and a panel of physicians from specialized disciplines consulted on demand) analyzed all 514 RRT calls that were extracted from the hospital telephone log data. Of these 514, the panel selected the 237 cases where sufficient documentation about the event was found. Sixteen patients for whom an RRT call was initiated within the first 24 hours were excluded. The remaining dataset included 221 cases, for which the panel judged if a clinical deterioration occurred after a full review of the patient's medical file. The deterioration was also classified by the panel according to preset deterioration categories that are listed in

Supplementary Material. Then, after accounting for three false alarm calls, the final dataset included 218 deterioration cases.

Second, these cases were complemented by 146 "non-event" patient cases where no deterioration event had occurred during hospitalization, which were randomly chosen from a pool of patients admitted in the same study period, to the same wards and discharged home after a hospital stay between 3 and 7 days (5 days being the median Length of Stay (LOS) of patients admitted to the included general wards).
This constructed dataset was later split into three separate parts that were used respectively for the training, validation and testing of the model. We used an oversampling algorithm (SMOTE)[187] to balance the dataset distribution, after dataset splitting. Figure 4 illustrates the dataset selection and inclusion steps.

Figure 4: Flowchart of cases recruitment and dataset construction



Explanatory variables (model features)

Forty-four explanatory variables (model features) available in the EMR that could potentially be early predictors of the patient deterioration outcome were identified by the expert panel based on a literature review
[188][189][190][191][192][193][194][195][196][197][198][199][200][201] of the predictors of most common in-hospital clinical deterioration situations. These variables included demographic patient data (e.g. age and sex), laboratory values (absolute value and difference from the previous value, noted Δ) and use of specific medical devices or interventions on the

patient (such as BiPAP, mechanical ventilation), but did not include vital signs. The complete list of variables is listed in the Supplementary digital content 1.

Measurement and Prediction timing

Several time points for prediction were considered to account for the model's time dependency. Time of prediction, Tp, was defined as the time prior to T_0 at which the prediction was generated, where T_0 is the time of the deterioration event.

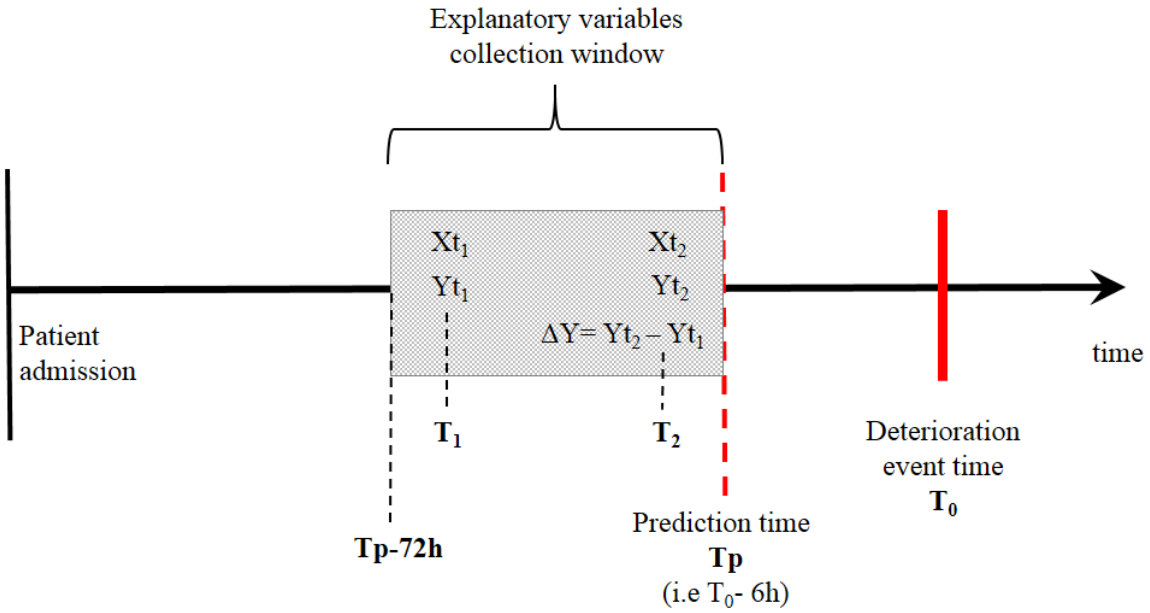
For each patient, we selected measurements (values of explanatory variables) at the following prediction time points Tp: T_0-3h, T_0-6h, T_0-12h, T_0-18h, T_0-24h, T_0-30h, T_0-36h, T_0-42h, T_0-48h. These prediction time points were chosen based on the frequency of patient clinical reevaluation (every 6h-8h) adopted for non-critical wards in clinical practice recommendations[202], and observed in most hospitals. For non-deteriorating patients, T0 was set as the time of discharge.

For each prediction time point (Tp), the most recent value relative to Tp of each explanatory variable was measured and documented, all the way up to three days (72h) before Tp, in line with similar studies [173]. This interval between Tp and Tp-72h will be called the explanatory variables collection (or sampling) window. In fact, this window was chosen to be wide enough in order to take into consideration the values of different laboratory exams that are not necessarily ordered by the medical team in the same day nor repeated with the same frequency as per clinical guidelines [203]. At the same time, that same window should be sufficiently limited in time (3 days) not to exceed the maximal predictive horizon of physiological biomarkers in the literature [204][205] relative to clinical deterioration (72h), hence close enough to the prediction time point so that the contained values of the requested exams can still be associated with the physiological and clinical status of the patient at the time of prediction.

Differential (delta) variables (for example ΔCRP) were defined as the difference between the available value of the variable closest to Tp in time and the available value of the variable furthest from Tp in time, all within the explanatory variables collection window.

Missing values among any explanatory variable in the window were imputed by using the mean value of the same variable over the entire cohort in the same time window. An illustration of the prediction timeline and its associated concepts can be found in Figure 5.

Figure 5: Timeline for prediction and concepts definition



Model training, validation, and testing/ Algorithms

The Python programming language was used for developing the scripts to create and analyze the models. A supervised learning approach was adopted using different machine learning algorithms: Random Forests (RF), Gradient Boosting (GB), Artificial Neural Networks (ANN) and Logistic Regression (LR). We used the implementation from the Sklearn Python module for RF and LR, XGBoost for GB, Keras for ANN.

50% of the dataset was used as a training set, and the rest of the dataset was equally split to be used for validation and testing using a 5-fold cross-validation.

Outcomes and Evaluation Metrics

The area under the receiver operating curve (AUROC) and the F1-score (defined as the harmonic mean of the precision and recall of the model outcome) were used for reporting the performance results of the different algorithms for each class of deterioration, calculated on the basis of a “one versus rest” approach.

In order to identify the important predictors of the model, variable importance was determined by calculating the relative influence of each explanatory variable on the algorithm classification results using the Python Sklearn library (Python Software Foundation. Python Language Reference, version 3.7).

The model parameters were fine tuned for the different algorithms using only the training and validation datasets (not the testing dataset) and using specific tools in the Python Sklearn model selection library, such as GridSearchCV.

5.3. Results

Descriptive statistics

Patient deterioration events in the study occurred in the following hospital departments: internal medicine (48%), infectious diseases (22%) and medico-surgical (30%).

The deterioration cases distribution by diagnosis and clinical outcome distribution (Stabilization on floor, Transfer to ICU, Code Blue, Not For Resuscitation) of the different deterioration cases by class are shown in Table 10.

Table 10: Distribution of the clinical outcomes of the deterioration cases included in the model

	Number of cases per outcome type					
Deterioration type (typical examples)	**Stabilized on floor**	**Transfer to ICU**	**Code Blue**	**Not For Resuscitation**	**Total Cases**	**Percentage**
Cardiological (Atrial fibrillation, Tachyarrythmia, Supraventricular tachycardia, Cardiac infarct)	38	8			46	21.1%
Pneumonia (Pneumonia, Aspiration pneumonia, Pneumonitis, Bronchiolitis)	32	21			52	23.9%
Pulmonary edema / Fluid overload (Heart Failure decompensation, Fluid overload)	16	4		1	21	9.6%
Sepsis (Sepsis / Severe sepsis / Septic shock)	25	31	3	2	62	28.4%
Hepatic / Pancreatic failure (Hepatic encephalopathy)	5	4			9	4.1%
Hypovolemia/ Hypovolemic shock (Hemorrhage)	3	6			9	4.1%
Other (Hospital induced/acquired conditions including hypoglycemia, medication errors/adverse effects, etc.)	9	8	1		16	7.3%
Total	**128**	**82**	**4**	**3**	**218**	**100.0%**

Model performance

Performance of the various algorithms was calculated and depicted in Table 11. The best performance was achieved with the Random Forests algorithm, with a maximal AUROC of 0.90 and F1-score of 0.85 obtained at prediction time T0-6h. This slightly decreases but is still acceptable at T0-42h, with an AUROC of 0.82 and an F1-score of 0.77

Table 11- Algorithms performance versus prediction time

Algorithm/ Number of test cases	Model parameters	Metrics	T_0-3h	T_0-6h	T_0-12h	T_0-18h	T_0-24h	T_0-30h	T_0-36h	T_0-42h	T_0-48h
Random Forest Classifier N=108	(n_estimators = 600, criterion="entropy", max_depth=12, min_samples_leaf=2,min_samples_split=4)	**precision (Deterioration/No deterioration)**	0.81/ 0.80	0.85/ 0.85	0.81/ 0.80	0.85/ 0.79	0.8/ 0.77	0.76/ 0.77	0.74/ 0.78	0.8/ 0.75	0.71/ 0.74
		recall (Deterioration/No deterioration)	0.79/ 0.80	0.85/ 0.85	0.79/ 0.80	0.77/ 0.87	0.75/ 0.81	0.77/ 0.75	0.79/ 0.72	0.74/ 0.81	0.75/ 0.70
		f1-score	0.81	0.85	0.81	0.82	0.78	0.76	0.75	0.77	0.73
		AUROC score	0.87	0.9	0.88	0.87	0.88	0.87	0.83	0.82	0.78
Boosting Classifier (XG-Boost) N=108	(max_depth = 12, learning_rate = 0.01, gamma = 0, min_child_weight = 1, n_estimators=600)	**precision (Deterioration/No deterioration)**	0.75/ 0.76	0.84/ 0.79	0.67/ 0.69	0.74/ 0.69	0.76/ 0.70	0.65/ 0.62	0.76/ 0.70	0.65/ 0.65	0.63/ 0.62
		recall (Deterioration/No deterioration)	0.77/ 0.74	0.77/ 0.85	0.72/ 0.64	0.66/ 0.77	0.66/ 0.79	0.58/ 0.68	0.66/ 0.79	0.64/ 0.66	0.60/ 0.64
		f1-score (Deterioration/No deterioration)	0.75	0.81	0.68	0.72	0.73	0.63	0.73	0.65	0.62
		AUROC score	0.85	0.86	0.81	0.83	0.85	0.76	0.79	0.73	0.72

Table 11- Algorithms performance versus prediction time (continued)

Algorithm/ Number of test cases	Model parameters	Metrics	T_0-3h	T_0-6h	T_0-12h	T_0-18h	T_0-24h	T_0-30h	T_0-36h	T_0-42h	T_0-48h
Artificial Neural Networks N=108	(architecture 20/8/1, loss='binary_crossentropy', optimizer='Adam', metrics=['accuracy'], BS=43, EPOCH=4000)	**precision (Deterioration/No deterioration)**	0.74/ 0.74	0.75/ 0.71	0.71/ 0.66	0.76/ 0.73	0.75/ 0.61	0.79/ 0.70	0.62/ 0.69	0.71/ 0.74	0.69/ 0.69
		recall (Deterioration/No deterioration)	0.74/ 0.74	0.68/ 0.77	0.60/ 0.75	0.72/ 0.77	0.45/ 0.85	0.64/ 0.83	0.75/ 0.55	0.75/ 0.70	0.68/ 0.70
		f1-score (Deterioration/No deterioration)	0.74	0.73	0.68	0.75	0.65	0.74	0.65	0.73	0.69
		AUROC score	0.78	0.78	0.82	0.79	0.76	0.8	0.72	0.78	0.75
Logistic Regression N=108	(penalty='l2', dual=False, tol=0.0001, C=1, fit_intercept=False, intercept_scaling=1, class_weight='balanced', random_state=None, solver='lbfgs', max_iter=30000,, warm_start=False, n_jobs=None, l1_ratio=None)	**precision (Deterioration/No deterioration)**	0.85/ 0.70	0.90/ 0.71	0.70/ 0.86	0.86/ 0.79	0.88/ 0.74	0.87/ 0.77	0.78/ 0.76	0.80/ 0.74	0.74/ 0.71
		recall (Deterioration/No deterioration)	0.62/ 0.89	0.62/ 0.93	0.70/ 0.86	0.77/ 0.88	0.68/ 0.91	0.73/ 0.89	0.75/ 0.79	0.71/ 0.82	0.70/ 0.75
		f1-score (Deterioration/No deterioration)	0.76	0.78	0.78	0.82	0.79	0.81	0.77	0.77	0.72
		AUROC score	0.81	0.85	0.82	0.87	0.86	0.88	0.83	0.81	0.8

Explanatory variables' importance

Explanatory variables' importance for the Random Forests model were calculated and represented in Figure 6, using a "heatmap" representation warm-to-cool color scheme, with the warm colors representing high-value impact of the variable and the cool colors representing a low-value impact.

The most contributing variables to the prediction result (in decreasing order) were the following: CRP, Lymphocytes count, Sodium minus Chloride, Sodium differential, Alkaline Reserve differential, Age, BUN differential, Potassium differential and Neutrophil-to-Lymphocyte ratio. Also, we illustrated in Supplemental material- 2 one example (among others) of a logical visualization of the decision-making process of the model using the Decision Tree algorithm at T0-12h, showing the above-mentioned variables and the model chosen thresholds.

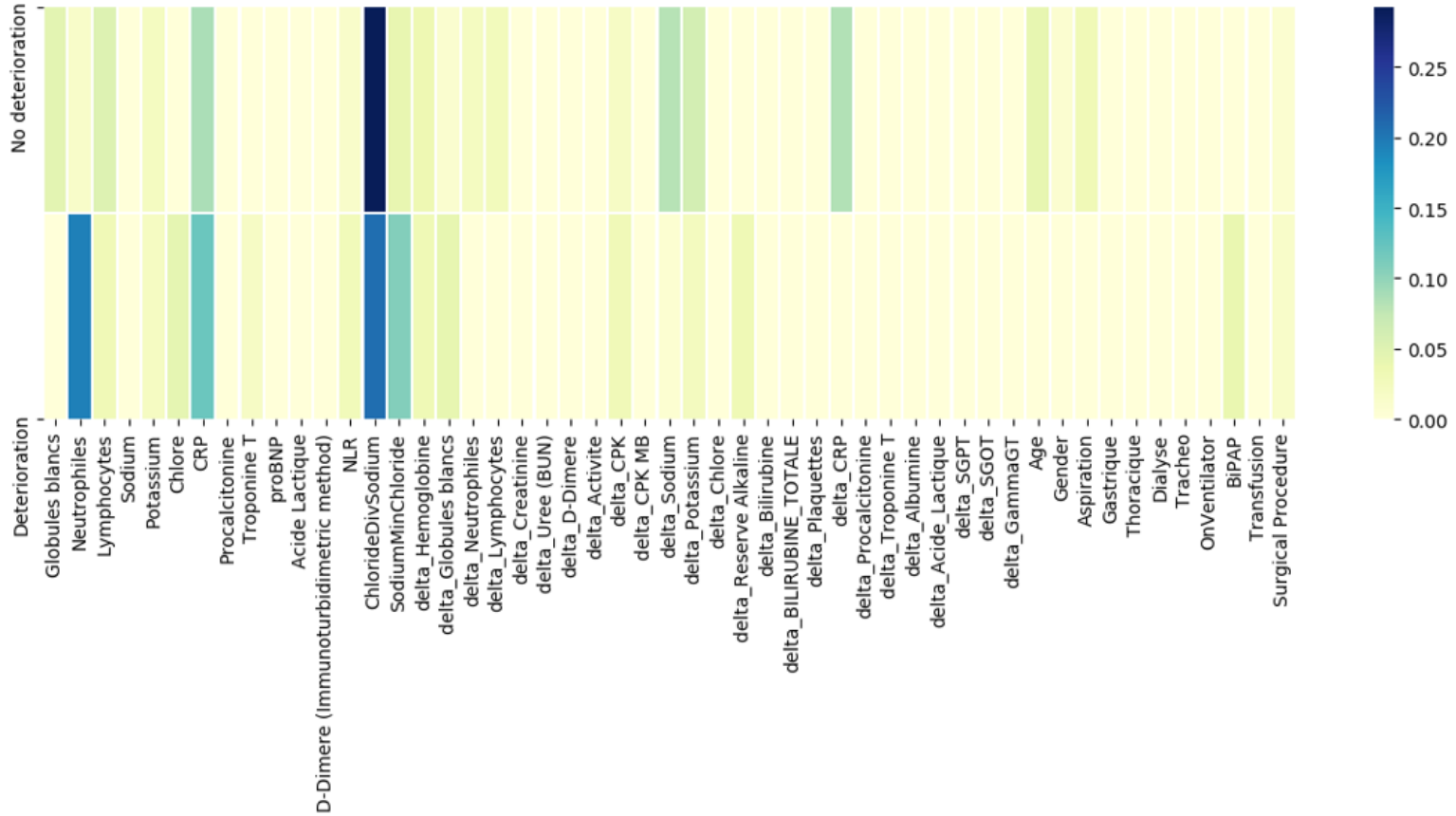
Benchmark against other all-cause deterioration models

Benchmarks to track-and-trigger (vital signs based) deterioration prediction models and to other more hybrid deterioration prediction models (vital signs, laboratory values, patient demographics, diagnosis, etc.) from the literature are given in Table 3, both in terms of performance metrics, outcome variables and best time-to-prediction.

The prediction model showed an earlier prediction horizon (up to 42h) with acceptable performance (AUROC more than 0.8), relative to most track-and-trigger systems (6h-24h), but also most hybrid all-cause deterioration prediction systems (12-48h).

The F1-Score (and specifically the positive predictive value) of the model is good (>0.8) and scored better than most track-and-trigger models, which could mean in practice a lower rate of false alarms generated.

Figure 6: Explanatory variables importances by deterioration class (Random Forest Classifier), prediction at T0-6h



(Globules blancs: White Blood Cells, Chlore: Chloride, NLR: Neutrophils to Lymphocytes Ratio, ChlorideDivSodium: Chloride to Sodium ratio, Uree: Urea, Activite: Partial Thromboplastin Time ratio, Reserve Alcaline: blood Bicarbonate, Plaquettes: Platelets)

Table 12: Benchmark relative to a number of recent studies and reviews with similar scope

Model category	Study/ Model name	Study phase	Study type	Statistical Methods Used for model derivation	Prediction performance	Types of variables used	Outcome measure	Prediction horizon (window)
Track-and-trigger models (vital signs based)	Campbell et al., 2020/ Q-ADDS [206]	Prediction model performance benchmark	Retrospective single-center cohort	Clinical consensus-based	0.71 (AUC)	Vital signs	Death/ unanticipated admission to intensive care	30h
	Kia et al., 2020 / MEWS++ [207]	Prediction model validation	Retrospective single-center cohort	Machine Learning algorithms	0.85 (AUROC)	Vital signs	Death/ unanticipated admission to intensive care	6h
	Kirkland et al., 2013 [181]	Prediction model validation	Retrospective single-center cohort	Multivariate regression analysis	0.71 (AUROC)	Vital signs, Braden score, Fall risk score	Rapid response team activation	2-12h
	Cho et al., 2020 [208]	Automated system performance benchmark	Retrospective single-center cohort	Machine Learning algorithms	0.86 (AUC)	Vital signs	Cardiac arrest/ unanticipated admission to intensive care	0.5h-24h
	Gerry et al., 2020 [120]		Systematic Review	AI and non-AI algorithms	0.55 to 0.96 (C-index)	Vital signs	Death/ unanticipated admission to intensive care	24h/Inpatient stay
	Fu et al., 2020 [209]		Systematic Review	AI and non-AI algorithms	0.71-0.96 (AUC)	Vital signs	Death/ unanticipated admission to intensive care	24h/Inpatient stay
	Peelen et al., 2021 [182]		Systematic Review	AI and non-AI algorithms	0.65-0.93 (AUC)	Vital signs	Rapid response team activation, cardiopulmonary resuscitation, unanticipated transfer to an ICU, or death	2h-24h
	Muralitharan et al, 2021 [93]		Systematic Review	Machine Learning algorithms	0.57 to 0.97 (AUC)	Vital signs	Cardiac arrest/ Death/ unanticipated admission to intensive care	4-24h

Table 12: Benchmark relative to a number of recent studies and reviews with similar scope (Continued)

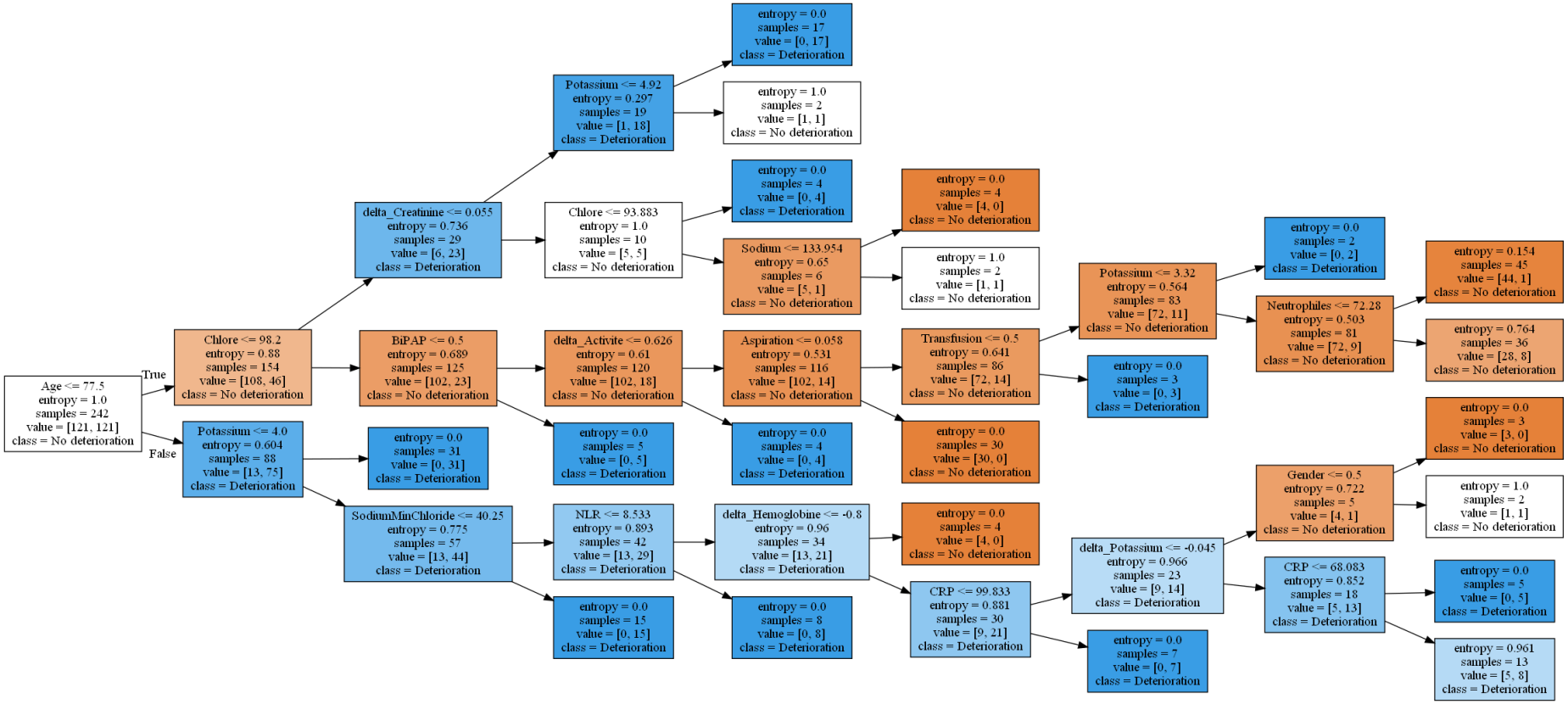
Model category	Study/ Model name	Study phase	Study type	Statistical Methods Used for model derivation	Prediction performance	Types of variables used	Outcome measure	Prediction horizon (window)
Hybrid deterioration prediction models (vital signs, biomarkers, and patient demographics data)	Jefferey et al., 2018 [210]	Prediction model validation	Retrospective single-center cohort	Machine Learning algorithms	0.85 (AUROC) 0.27 (F1-score)	Vital signs, Laboratory tests, ICD-10 diagnosis, Demographic data	Cardiopulmonary arrest	48h
	Churpek et al., 2016/ eCART [211]	Prediction model validation	Retrospective multicenter cohort	Machine Learning algorithms	0.77 (AUC)	Vital signs, Laboratory tests, Demographic data	Cardiac arrest/ Death/ unanticipated admission to intensive care	24h
	Kipnis et al., 2016 / AAM [173]	Evaluation of implemented system	Retrospective multicenter cohort	Discrete-time logistic regression	0.82 (AUC)	Vital signs, Laboratory tests, Severity of illness, Comorbidity index, Demographic data	Unanticipated admission to intensive care	12h-24h
	Pimentel et al., 2021/ HAVEN [212]	Evaluation of implemented system	Retrospective multicenter cohort	Machine Learning algorithms	0.90 (AUC)	Vital signs, Laboratory tests, Comorbidities index, Frailty	Cardiac arrest/ unanticipated admission to intensive care	24h-48h
	Blackwell et al., 2020 [188]	Prediction model validation	Retrospective single-center cohort	Multivariate regression analysis	0.71-0.84 (AUC) depending on outcome	Vital signs, Laboratory tests and continuous seven-lead electrocardiogram (ECG) signal	Unanticipated admission to intensive care	12h
Hybrid deterioration prediction models (Biomarkers, devices and demographics data only)	Our model	Prediction model validation	Retrospective single-center cohort	Machine Learning algorithms	0.82-0.9 (AUROC) 0.77-0.85 (F1-Score)	Laboratory tests, Attached devices, Demographic data	Rapid response team activation	24h-48h

Table 13 – List of explanatory variables by type

(Blue: biomarkers; Green: Patient demographic data; Yellow: clinical procedure status)

Absolute variables	**Differential variables**
White Blood Cells (WBC)	Δ_WBC
Neutrophils	Δ_Neutrophils
Lymphocytes	Δ_Lymphocytes
Sodium	Δ_Sodium
Potassium	Δ_Potasium
Chloride	Δ_Chloride
C-reactive protein (CRP)	Δ_CRP
Procalcitonin	Δ_Procalcitonin
Troponin T	Δ_Troponin_T
D-Dimer	Δ_D-Dimer
proBNP	Δ_Hemoglobin
NLR	Δ_Creatinine
ChlorideDivSodium	Δ_Urea
SodiumMinChloride	Δ_Activite
Age	Δ_CPK (creatine phosphokinase)
Gender	Δ_CPK MB (creatine phosphokinase myocardial band)
Aspiration in the last 24h	Δ_Bicarbonate
Tracheostomy status	Δ_Platelets
OnVentilator status	Δ_Albumin
BiPAP patient status	Δ_SGPT (serum glutamate-pyruvate transaminase)
Blood transfusion in the last 24h	Δ_SGOT (aspartate aminotransferase)
Surgical intervention in the last 24h	Δ_GammaGT (Gamma-glutamyltransferase)

Figure 7 – Example of model decision process visualization (using Decision Tree algorithm)



5.4.Discussion

In this retrospective, single-center study, we developed and evaluated a machine-learning model for the prediction of all-cause patient deterioration. The model's explanatory variables were mainly biomarkers values routinely available in basic EMRs, without inclusion of vital signs data.

1) Potential use of the model for predicting clinical deterioration and supporting clinical decision making

If transformed into an automated clinical decision support tool and applied systematically to all hospital inpatients, this model could potentially stratify patients based on their deterioration risk score, and proactively alert the healthcare team of patients possibly at high risk of deterioration within the next hours/days. The update or refreshing of the model data prediction result would basically rely on the arrival of new laboratory data, thus on the frequency of blood sample extraction, which in practice can range from 12h to 48h for most in-hospital patients.

This prediction is based on the capture of a rich "physiological picture" (mainly through biomarkers) which precedes chronologically the "clinical picture" (captured by track-and-trigger models, through observation of vital signs and clinical examination), hence an earlier prediction of deterioration.

This earlier prediction (up to 42h versus 6h to12h for track-and-trigger models) can give the healthcare team a window of opportunity to try to stabilize or manage at-risk patients on general wards, preventing as much as possible their transfer to the intensive care units or any further escalation in care. This information can also permit the medical and nursing team to selectively increase surveillance for patients at high risk of deterioration, hence trying to prevent or promptly mitigate expected deterioration events. In the context of a global shortage of health workers, this information can help in focusing resources on the patients that need those the most.

A complementary use of such a model can be for patient safety professionals in hospitals, who can make use of the prediction data on a daily basis to audit and verify the follow-up and safety actions taken by the healthcare team in order to manage the deterioration risks, including suitability of the level of care provided to the clinical status of the patient.

2) Model explainability and the road towards clinical validation and clinician adoption

Explainability, or the possibility to understand the model's classification logic is an important feature that can facilitate the "clinical interpretation" of the results by the clinicians.

In this study, the deterioration model permits a certain level of "explainability" for most algorithms applied and in particular Random Forests and Decision Tree, in the sense that it is possible to identify the main variables that influence most the model prediction results, along

with their respective weights. Further explainability can be obtained with Decision Tree algorithm where a visualization of the decision tree could be obtained, showing the logic behind the classification (Supplementary Digital Material 2).

Such data insight can help users understand the prediction results, and facilitate any future effort to clinically interpret and validate the model by an experienced panel of physicians. This "clinical validation" is an important step towards the practical adoption of the model by clinicians, where the latter are often reluctant to use "black box" models, even when they show good results.

3) Model specificities relative to other predictive models, and possible impact on results

While most deterioration models in the literature were derived from cases with specific outcomes of cardiac arrests, death and unplanned transfer to ICU (Table 3), the model elaborated in this study was trained and validated on deterioration cases linked to RRT activations that were confirmed by a panel of clinical experts. It is to be noted that RRT activation cases depict a broader image of clinical deterioration, since they include an additional outcome in clinical practice, which is the patient stabilization on the floor, amounting to almost half of deterioration cases (Table 1), in addition to the classical aforementioned outcomes.

Furthermore, almost all of the deterioration models in the literature which include laboratory variables (such for example those of the LAPS-2 score [173]) use the absolute form of the exam values. To the best of our knowledge [182], our model is among a few (if not the only one) that use differential (or delta) biomarker variables in deterioration prediction models. It is known however that changes in biomarker values (delta) within a specific timeframe can indicate certain underlying pathophysiological changes, such as for example in case of bleeding (delta in hemoglobin values) or acute kidney injury (delta in creatinine values).

The analysis of variable importances (Figure 5) shows that a number of differential variables (for example: ΔSodium, ΔPotassium, ΔCRP) have a significant weight in the model prediction function.

We believe that the results of the prediction model were impacted to a certain extent by these specificities, but also the broad choice of biomarkers that intended to cover multiple deterioration mechanisms that are common to various deterioration etiologies. These mechanisms include but are not limited to respiratory and metabolic acidosis/alkalosis, systemic inflammation, electrolyte imbalance, volume imbalance and hypoperfusion/ischemia.

Finally, we believe that the exclusion of vital signs data from the model might have in a certain way contributed to an earlier prediction horizon. In fact, in pathophysiological processes leading to clinical deterioration, changes in biomarkers usually occur hours before clinical signs and symptoms. Furthermore, even in hybrid models (where variables comprise vital signs, laboratory data and other patient data), the importance of biomarkers could have been eclipsed

by the direct association (however late in matter of prediction) between the occurrence of clinical signs (vitals) with the deterioration event outcome. Further research might be needed to better elucidate the relation between the choice of variable type and the impact this has on the prediction horizon of clinical deterioration models.

Limitations

The study was conducted in a single center, which might have amplified the effect of certain factors on the results, such as the quality of the medical documentation and the specific practice of exam prescriptions for diagnosis and monitoring. An external and a prospective validation of the study model should be a undertaken to understand its performance in a real clinical context, before it can be implemented as a clinical decision support system.

Also, the number of deterioration events per explanatory variable is relatively small, which might have impacted to a certain extent the performance metrics and the statistics of the variables' importance. This is due to the limited sample size of the study. However, it corresponded to almost two years of systematic data collection of deterioration events in our hospital, and a thorough and time consuming validation by an expert panel of the cases outcome.

Conclusion

We have developed and validated an explainable prediction model for inpatient deterioration in general wards, trained on expert validated deterioration events with rapid response team activation. The model is mainly based on biomarkers, without use of vital sign data. The model performed better than most gold standard track-and-trigger systems, both in prediction performance and prediction horizon. Such a model can also be suitable for hospitals with limited resources and a basic EMR. Further increase of the data sample could contribute to improving its performance, and the model would gain to be externally and prospectively validated.

6. Contribution 3:

Implementation and validation of a CDS application for the early identification and risk management of hospitalized patients at high-risk for clinical deterioration on regular floors- The "VIGIL" project

Note: this study is still in progress regarding the data collection for the validation phase. Design and implementation of the clinical decision support application are completed and described below.

6.1. Introduction

Delays in identifying and managing clinically deteriorating patients have been found to be associated with increased morbidity and mortality[163]–[165] and this theme was identified as one of the emerging priorities in patient safety [13].

The current "gold standard" solutions adopted in hospitals worldwide to address this issue are the Rapid Response System (RRS), which are comprised of an afferent recognition limb, known as Early Warning System (EWS), and an efferent limb which handles the response to the alert, in matter of escalation and clinical interventions. RRS vary in their design and escalation mechanism, but most involve a nurse that receives or identifies the EWS alert and escalates to the medical team or to Medical Emergency Team (MET).

The increase in adoption of Electronic Medical Records (EMR) has rendered available the data that was once calculated manually by the nurses to identify the patient deterioration scores, and more and more automated EWS are available in commercial solutions and published in the literature. EWS data was revolving in the early years around the patient vital signs and some basic demographic data, but has since been extended to more categories, such as medical diagnosis, laboratory data, patient medications and medical imaging data. This data enrichment has permitted the elaboration of more complex and refined prediction models, and allowed the use of advanced statistical and computational techniques (such as machine learning) for their

derivation. It was reported that such complex models can achieve greater accuracy than classical aggregate-weighted EWS scores [93],

A recent review [94] has shown that although there is an increase in the number of automated and more complex EWS, the vast majority of these system have not been subject to implementation and validation in real clinical workflows, and therefore data on the real impact of these automated systems on patient outcomes is still scarce and more outcome studies are needed in this field.

Another even more recent review [213] did not detect improvements in patient outcomes following the implementation of automated real-time deterioration alerts, but recommended that more attention be given to the RRS efferent limb (response to the alert) rather to the alert itself, through reviewing the workflow of alert recipients and incorporating model features into the decision process to improve clinical utility.

In this study, we report a real-world implementation of a relatively new approach in RRS, which tries to bridge the gap between the deterioration alert and the clinical decision. The approach can be summarized as follows: 1) regarding the afferent limb, proposing an explanation of the results of a machine learning deterioration prediction model, using both the presentation of its features' values, and a clinical interpretation of these values through the identification of "clinical risks" via expert rules, and 2) regarding the efferent limb, formulating specific recommendations both to medical and nursing teams, relative to each "clinical risk" identified. These recommendations being are based on evidence-based medical and nursing guidelines.

We also present an evaluation protocol for this system, on many layers: validity of components, clinical usefulness and system usability. The evaluation study is still recruiting and the results are not available.

We adopted the framework described by Greenes et al. [1] for the characterization of the application design and implementation aspects.

6.2. Materials and Methods

6.2.1. CDS application description

a. The CDS application structure and components

A hybrid AI/Rules-based approach was adopted, in an attempt to combine in one application the capabilities of a machine learning model for clinical deterioration prediction with a rules-based module for clinical risk identification and management.

The application applies the machine learning model described in the previous chapter [214] for the stratification of the clinical deterioration risk of patients in non-critical wards (component 1), and combines it with an expert rules-based model (component 2) that was developed in collaboration with a panel of internal medicine physicians, which identifies a number of "clinical risks" for the patient, based on preset interpretation rules applied to the values of the explanatory variables of component 1 in addition to additional variables imported from the computerized patient medical record.
These "clinical risks" are not to be confused with the patient diagnosis, which in terms of causality is the primary and holistic reason explaining the patient condition. In fact, they are to be understood as describing the dynamic physiological paths that are secondary to this diagnosis, to the patient evolution or even to the hospital care, and that can directly trigger patient deterioration mechanisms if left untreated.

For each of the identified clinical risks, the CDS application then suggests a number of best-practice medical and nursing interventions (component 3) associated to each category of clinical risk, interventions that were validated by an expert panel based on the review of international practice guidelines. Figure 8 illustrates the different components of "VIGIL".

These "clinical risks" were identified further to a literature review in addition to a discussion with clinical experts. To illustrate through an example (Figure 9), an 85-year-old patient can be admitted for a community acquired pneumonia (diagnosis), but throughout his hospital stay his clinical status can deteriorate for example due to "clinical risks" such as sepsis and increasing oxygen requirements, both complications related to his initial diagnosis, or acute kidney injury further to administration of nephrotoxic drugs, which is a hospital-acquired condition. These risks can trigger a deterioration mechanism that will eventually translate into a clinically noticeable deterioration (such as hypotension, desaturation, tachycardia or decrease level of consciousness) requiring the intervention of a medical emergency team (MET) and in some conditions a transfer to a higher level of care.

All in all, the three different components would offer for each hospitalized patient a prediction score of the clinical deterioration risk, the identification of the clinical risks that can potentially explain and trigger his clinical deterioration, and finally suggest best-practice interventions in order to prevent or mitigate the effects of these risks, hopefully before the patient shows noticeable signs of clinical deterioration and necessitate transfer to a higher level of care.

Figure 8: "VIGIL" CDS application components

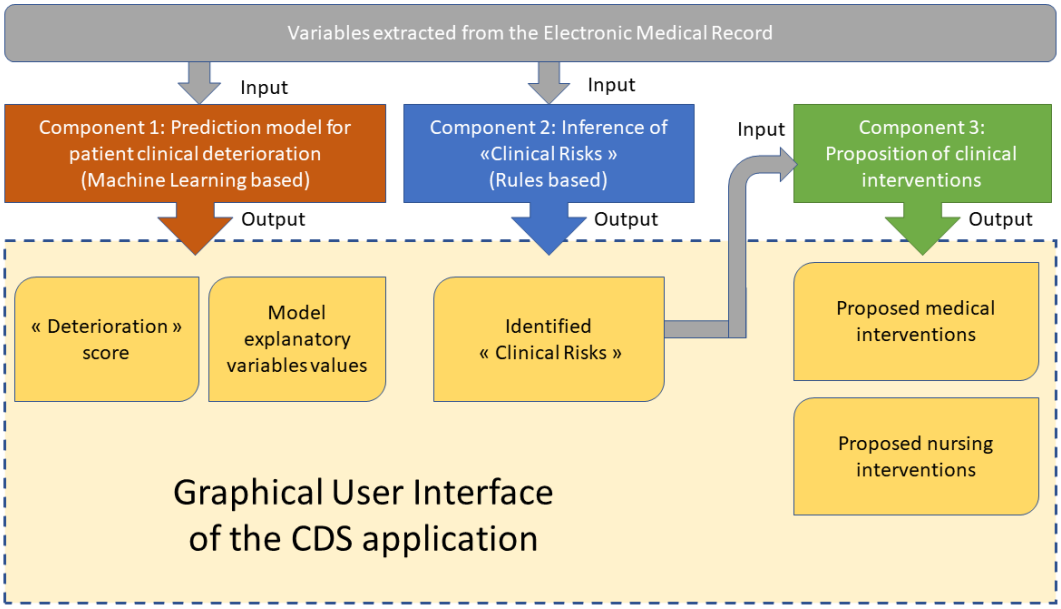
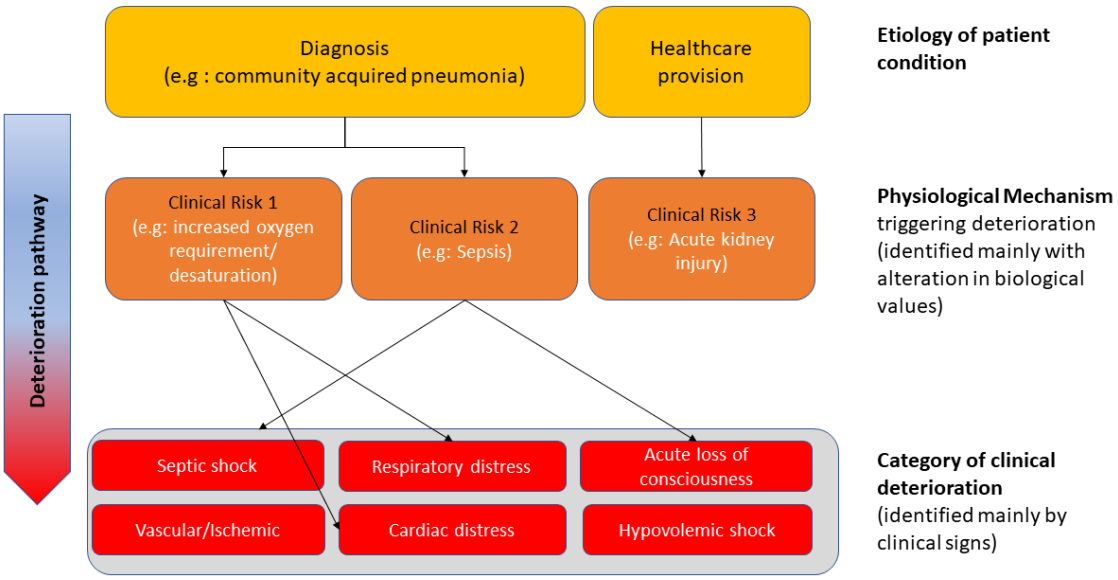


Figure 9: Illustration of the adopted conceptual framework regarding clinical deterioration



b. Cognitive tasks/ reasoning processes supported

The objective of the application is 1) to help the healthcare team identify the level of "clinical severity" of the patients based on a machine learning based prediction algorithm that "aggregates and synthesizes" a rich set of routinely available variables, mainly biological, 2) to support, through automated screening of data, the medical and the nursing team in identifying in a *"systematic and reliable way"* a number of common "clinical risks" that are associated with the patient condition, and 3) to *"remind"* the medical and the nursing team about the main clinical interventions to keep in mind for each of the "clinical risks", in order to ensure adherence to best practice guidelines.

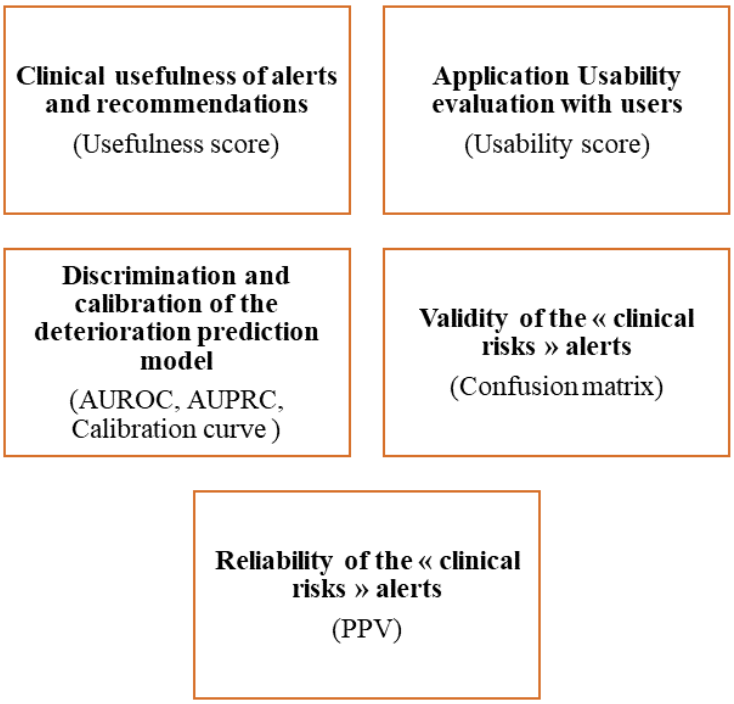
6.2.2. CDSS application evaluation

a. A multifaceted evaluation for the CDS application performance

In order to evaluate the performance of the CDS application, we will evaluate each component's outcome, in terms of reliability (were there any data-related issues that caused a false alarm?) and validity (was there clinical justification for this alarm?), as well as the clinical usefulness (was the alert clinically useful to the healthcare team?) of the application alerts and recommendations for a panel of users. Finally, a usability measurement method will be used to evaluate user satisfaction of the application and identify its potential strengths and areas for improvement. These different evaluation methods are listed below and summarized in Figure 10.

An impact measurement of the application relative to its ability to prevent or better manage clinical deterioration risks in hospitalized patient is planned but is outside the scope of this study.

Figure 10: Evaluation methods of the different application components and output



b. Validation of the deterioration prediction component

Data regarding deterioration cases identified by Rapid Response Team (RRT) interventions was systematically collected between February and August 2022. The daily prediction results of the application for the patients in this same period was then crossmatched with the cohort of deteriorated patients' cases. A deterioration case was considered as correctly predicted by the application if a prediction score above a certain threshold (e.g: > 0.7) was given by the application on any day up to 96 hours from the deterioration date and time. Otherwise it was considered that the application did not predict the deterioration event.
The area under the receiver operating curve (AUROC) and the area under the precision-recall curve (AUPRC) were used to reporting the performance results of the prediction model.

In order to identify the optimal classification threshold for the prediction score was calculated with the objective of maximizing sensitivity and specificity. Such threshold is important for the practical implementation of the application and the stratification of patients. Given the described threshold, the confusion map for this specific threshold was calculated.

Finally, the prediction model calibration was assessed using a calibration curve on the same dataset described above.

c. Validation of the clinical risks alerts

Based on a random sample of patients hospitalized in non-critical wards from September to December 2022, a multidisciplinary team composed of a senior patient safety officer and two physicians adopted the following protocol. For each patient, the team discussed with the treating attending physician the case of the patient (including reason for admission, diagnosis and evolution), and asked the physician to identify the current applicable types of clinical risks to the patient at this point of his hospitalization, from the predefined list of clinical risks illustrated in Table 13.
This list was compared to the list of risks identified by the CDSS, and whenever a discrepancy was found, the concerned identified risk(s) were discussed until a consensus was reached. The differences between the list identified by the attending physician and that of the CDSS were documented and analyzed.
In case the "clinical risk" alert was false due to a data reliability issue, the reliability score of the alert was assigned as Zero.
The reliability results were then reported for each type of "clinical risk" alarm.
The validity results were reported in the form of a confusion matrix for each type of clinical risk, where true positive, true negative, false positive and false negative rates of each type of "clinical risk" alarm was calculated.

Furthermore, for the "clinical risk" alerts deemed "valid" by the reviewers and the treating physician, the "clinical usefulness" of the alerts were assessed after comparative review of the patient medical prescription sheets just before and one day after the discussion with the treating physician. A "usefulness" score was then given as follows to the alert: Zero = alert already taken into consideration in the medical prescription sheets; 1 = alert did not induce relevant changes in the prescriptions; 3 = relevant changes in the medical prescription sheets noticed after discussion of alert.
The "clinical usefulness" results were reported for each type of clinical alert as a distribution of these scores.

d. Evaluation of the proposed interventions identified by the application

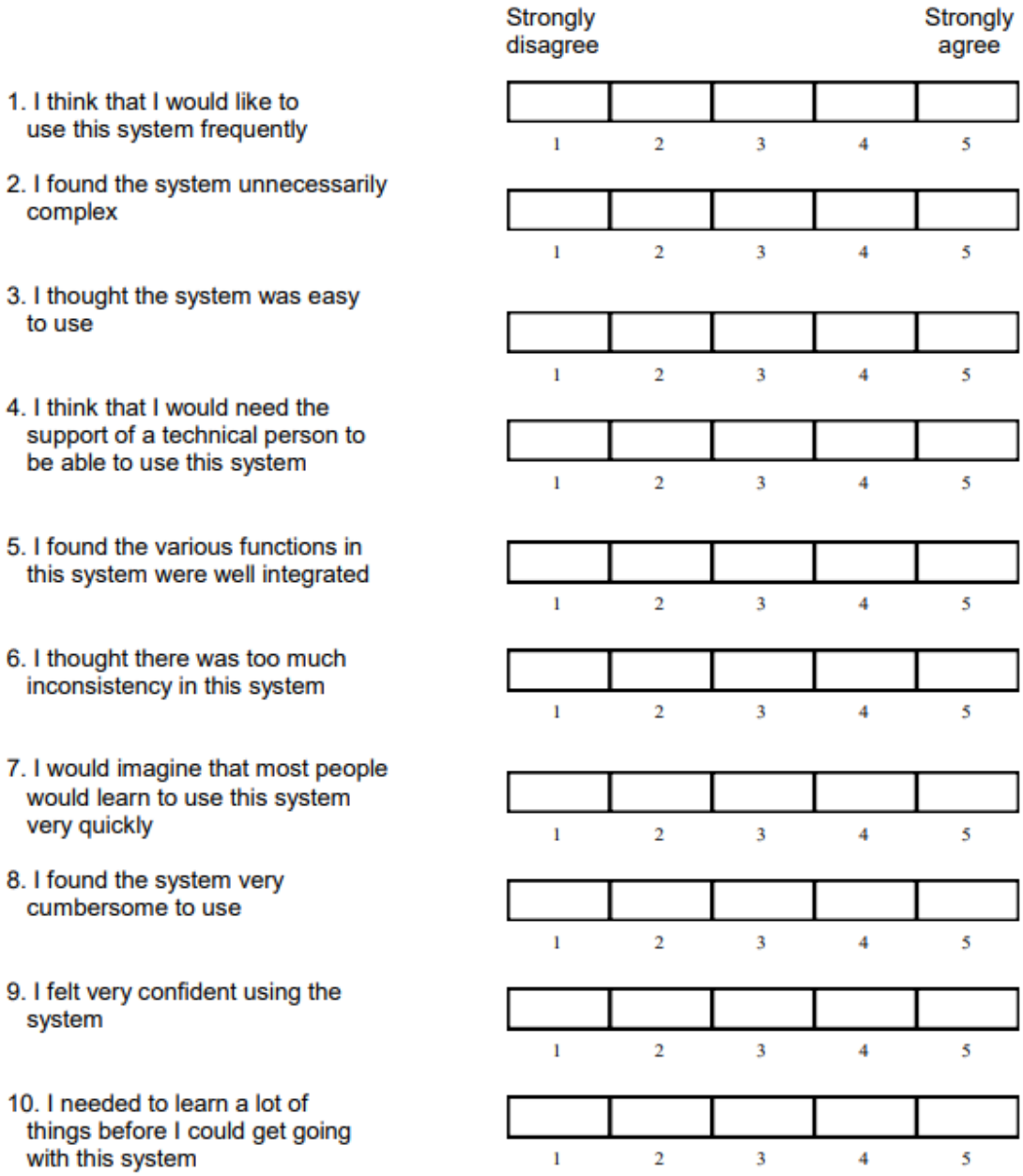
For the same sample of patients aforementioned, and for the "clinical risk" alerts deemed "valid" both by the reviewers and the treating physician, a discussion of the relative proposed medical and nursing interventions was respectively performed with the treating physician and the nurse in charge. The "clinical usefulness" of these automatically proposed medical and nursing interventions were assessed after a comparative review of the patient medical prescriptions and nursing plan just before and one day after the

discussion. A score was then given as follows to each proposed intervention: Zero= recommendation deemed not applicable or not indicated for the patient case, 1= recommendation applicable to the patient case and already ordered or executed by the healthcare team, 2= recommendation applicable to the patient case but not ordered by the healthcare team, 3= recommendation applicable to the patient case and ordered by the healthcare team post-discussion.
The "clinical usefulness" results were reported for each type of medical and nursing interventions as a distribution of these scores.

e. Application usability evaluation with a panel of users

Usability testing was undertaken in order to determine whether the application was judged as usable, effective and acceptable to users. We adopted the 'Think Aloud' qualitative method in order test the usability on a one-to-one basis for a group of physicians and a group of nurses. The sessions were moderated by one of the authors using a standardized, structured worksheet combined with a semi-structured discussion using open-ended questions to evaluate each tool component. Participants were encouraged to 'think aloud' and verbalize their thoughts about the component being tested. We also administered to each participant a validated, 10-item System Usability Scale [215] to assess their satisfaction regarding the usability of the tool.

Figure 11: System Usability Scale questionnaire



6.3.Results

The CDS application, which we named “VIGIL”, is a web application that can be accessed through an internet browser inside the hospital. The application, developed in Flask Python web framework [216], incorporated the predictive model described in the second contribution, with the addition of the two additional rules-based modules for the “clinical risk” evaluation and the medical and nursing interventions recommendations. The application is in the piloting phase before launching a wider evaluation before the end of 2022.

6.3.1. Identification of “clinical risks” and recommended “clinical interventions”

The medical panel identified 41 “clinical risks” and defined for each of these risks a formal practical definition (trigger) based on values of certain variables directly imported from the patient medical record, most of which are the explanatory variables of the deterioration prediction score (component 1). For these identified risks, the medical panel identified more than 119 best-practice medical interventions, paralleled by more than 105 nursing interventions, identified by the specialist nursing expert panel.

The list of “clinical risks”, their formal triggers and the associated medical and nursing interventions are listed in Table 13.

6.3.2. Integration of the solution/ adaptation to the clinical workflow

“VIGIL” is a standalone web application that is one-way integrated to the hospital information system (HIS) (reads only from the HIS databases). At preset intervals (by default every day at 10 am), the application automatically imports from the HIS data that is required to load the variables required by the components 1 and 2. The application renders its outputs within 90 minutes of process initiation, a timing which coincides with the morning round of the medical teams on the floors.

In each clinical department (with the exception of critical wards), the application is accessible on the PCs in the physicians’ room to the members of the medical team and on the PCs of the nursing desk for the members of the nursing team.

Hence, the application can be ideally consulted in the beginning of the medical rounds, when the attendings and residents usually check the updates of lab exams of their patients before the physical round.

Access to the application is also granted to the Senior Patient Safety Officer and to the Medical and Nursing Administrators, for auditing and overview purposes.

6.3.3. Features displayed in the User Interface

Once required HIS data is extracted, the CDSS computes and presents the results of all patients in a table where each row is a patient entry. The main results of the models are presented in eleven specific columns:

Patient information: In this column are displayed the patient's name, age, admission date
Case number: displays the unique identifier related to the medical case of the patient in the hospital
Treating doctor: displays name of the patient's treating physician
Patient bed: displays the current bed number of the patient
Diagnosis, recent procedures and medications: this section aggregates contextual data on the patient, such as his current and last ICD-10 diagnosis, the dates and names of the interventional procedures the patient had performed in the last month, as well as the list of medications ordered for the patient in the last 3 days.
Deterioration prediction score: this is the score yielded by the prediction model (component 1) that should be correlated with the risk of clinical deterioration.
Deterioration dynamic: relative to the last prediction score, this column shows the evolution and labels it as "improving", "status quo" or "worsening".
Verify the following risks: displays the inferred "clinical risks" based on the patient data.
Proposed medical interventions: displays the automatically proposed medical best practice interventions relative to the identified "clinical risks".
Proposed nursing interventions: displays the automatically proposed nursing best practice interventions relative to the identified "clinical risks".
All variables: displays all important variables taken into consideration by the different components, highlighting in red critical values of the variables.

An example of the information presentation is given in Figure 12.

Figure 12: VIGIL user interface with a case example

List of patients sorted by deterioration score, updated at 2022-09-29 11 hours

	PATIENT DETAILS	CASE NUMBER	TREATING DOCTOR	PATIENT BED	DIAGNOSIS,RECENT PROCEDURES AND MEDICATIONS	DETERIORATION PREDICTION SCORE	PREDICTION DYNAMIC	VERIFY THE FOLLOWING RISKS	CONSIDER THE FOLLOWING MEDICAL INTERVENTIONS	CONSIDER THE FOLLOWING NURSING INTERVENTIONS	ALL VARIABLES
2	Patient name: [redacted] Age: [redacted] Admission Date: [redacted]	[redacted]	[redacted]	[redacted]	Current diagnosis: Previous diagnosis: Acute renal failure Invasive procedures in the last 30days: {} Polypharmacy index: 23 Medications administered in the previous 3 days: Methylprednisolone (Lisamethyle) 40mg/2ml MontageSimple170cm+prised'air+ssbulb+enY (IV-SVFBPI) Becozyme Inj (Vitamine B Complex) (LASA) Combivent (Duolin) 2.5ml (LASA) GastriMUT Comp 20mg 14/bte (LASA) Crestor (Rosucor)20mg LASIx 250mg/25ml inj (Furosemide Renaudin) (LASA) Allevin 10cmx10cm GLUCOnate De Calcium Inj 10ml/fl (LASA) Risek (ulcazal 40mg) IV Hemoglucotest (Accucheck) Lancet Softclix (safe T pro Uno) Concor (Bisoprolol) 5mg Atrovent 500mcg/2ml (LASA) Resonium poudre (Resical)15g/sachet Fucitalmic Gel 10mg 5gr Flamazine (silvederma) 1% Creme 50gr/tube Seringue Insuline 1ml Hemoglucotest (Bandelettes) Gmate Piperacillin/ Tazobactam (Yanoven) 4.5g Humulin N 100IU/ml 10ml Heparine Sodique 25.000UI /5ml Bicarbonate de Sodium 8.4% (20ml) inj	0.934008	Status quo	Renal insufficiency risk [Uremia] Renal insufficiency risk [Renal Failure] Metabolic risk [Dehydration / Osm] Metabolic risk [Hypernatremia] Metabolic risk [Hypokalemia] Respiratory distress risk [High oxygen requirements] Respiratory distress risk [BiPAP/Risk of hypercapnia] Postoperative event risk [Recent procedure] Respiratory distress risk [Tracheostomy]	Determine etiology of hypernatremia (dehydration, osmotic diuretics, diarrhea, etc.) Identify and stop possible nephrotoxic drugs Verify/Adapt patient diet Repeat potassium level (obtain heparinized sample). Obtain stat ECG and examine for changes of hypokalemia.Correct hypokalemia with potassium supplementation if needed IV hydration and electrolytes monitoring	When giving potassium supplement through IV, ensure controlled delivery of medication to prevent bolus effect and reduce associated discomfort such as burning sensation at IV site Monitoring q4h of vital signs, consciousness level, patient color- alert if confusion or decreased level of consciousness Monitoring q4h of vital signs, assess for tachycardia, hypoventilation and altered neuromuscular function such as tetany, paresthesia, apathy, drowsiness, irritability Monitor urine output Ensure good patient hydration (PO/IV) Verify/Adapt patient diet card and tray	Hemoglobine : 10.5 (Δ : 0.1) Plaquettes : 262000.0 (Δ : -56000.0) Globules blancs : 12000.0 (Δ : 1300.0) Neutrophiles : 81.0 (Δ : -6.0) Lymphocytes : 12.0 (Δ : 3.0) NLR : 6.75 Sodium : 147.0 (Δ : 1.0) Potassium : 2.93 (Δ : -0.51) Chlore : 96.9 (Δ : 0.2) Reserve Alkaline : 29.1 (Δ : 1.3) ChlorideDivSodium : 0.66 SodiumMinChloride : 50.1 Creatinine : 5.45 (Δ : -0.61) Uree (BUN) : 171.0 (Δ : -1.0) proBNP : 5625.0 Δ Activite : -1.0 CRP : 48.0 (Δ : -12.0) Albumine : 3.2 Age : 79.0 Gender : 1.0 BiPAP : 1.0 Surgical Procedure : 1.0 Tracheo : 1.0 On_Oxygen_High_Requirements : 1.0

Table 13: Definition of clinical risks adopted in the study, and listing of corresponding medical and nursing risk management interventions

Category of clinical risks	**Classes of clinical risks**	**Practical rule**	**Medical risk management**	**Nursing risk management**
Infectious/ Septic risk	Onset of bacterial infection	(Increase of CRP of at least 30pts OR Increase of Procalcitonin of at least 0.5 pts) AND increase of WBC of at least 4000pts	"Blood, sputum, urine and urine analysis, wound drainage and stool cultures depending on the presenting signs", "CBC with differential, routine chemistries, LFTs, CRP, procalcitonin, and according to situation: troponin, lactic acid level, coagulation profile ",	"Monitor according to severity and need: vital signs, fever, consciousness level, patient color- alert if hypotension, fever, hypoxia/cyanosis, oliguria/anuria, confusion or decreased level of consciousness" “Verify patient labs to monitor evolution”
	Onset of viral infection	(Increase of CRP of at least 20pts OR WBC<4000) AND increase of %Lymphocytes of at least 20pts	“Consider patient isolation”	“Apply patient isolation after verification with medical team”

	Acute infection	(WBC>12000 OR WBC<4000) AND CRP>40 AND 5<Neutrophil-to-Lymphocyte ratio<10	"Blood, sputum, urine and urine analysis, wound drainage and stool cultures depending on the presenting signs", "CBC with differential, routine chemistries, LFTs, CRP, procalcitonin, and according to situation: troponin, lactic acid level, coagulation profile ", Order Urine analysis", Start empirical ATB coverage and readapt according to culture results", "Monitor evolution of inflammation markers and adapt ATB type/dose",	"Monitor according to severity and need: vital signs, fever, consciousness level, patient color- alert if hypotension, fever, hypoxia/cyanosis, oliguria/anuria, confusion or decreased level of consciousness" "Verify patient labs to monitor evolution"

	Sepsis	CRP>100 AND Neutrophil-to-Lymphocyte ratio >10	"CBC with differential, routine chemistries, LFTs, CRP, procalcitonin, and according to situation and severity: troponin, lactic acid level, coagulation profile ", "Blood, sputum, urine and urine analysis, wound drainage and stool cultures depending on the presenting signs", Order Urine analysis", Start empirical ATB coverage and readapt according to culture results", "Fluid hydration if hypotensive/ consider transfer to ICU if non-responsive, "Monitoring at least q4h and according to needs: vital signs, consciousness level, patient color", "Monitor evolution of inflammation markers and adapt antibiotherapy",	"Verify good function of IV access", "Urine output monitoring", "Administer first ATB doses timely without delay" "Monitor according to severity and need: vital signs, fever, consciousness level, patient color- alert if hypotension, fever, hypoxia/cyanosis, oliguria/anuria, confusion or decreased level of consciousness" "Verify patient labs to monitor evolution"
	Acute Sepsis	(CRP>70 AND	Same interventions as for "Acute infection/ Sepsis" +	Same interventions as for "Acute infection/ Sepsis" +

		NLR>10 AND Increase in CRP of more than 30 pts AND (Decrease in platelets count >50 000 units OR Decrease in Serum Bicarbonate of more than 5 pts OR Increase in Serum Creatinin of more than 0.5 pts)) OR (CRP>100 AND NLR>10 AND (Platelets count <150 000 units OR Serum Bicarbonate<20))	"Consider transfer to intensive care unit"	"Monitoring q2h of vital signs, consciousness level, patient color" "Consider continuous monitoring of vitals"
	Severe Neutropenia	Neutrophil count<750	"Check for onset of infection or adverse drug effect", "In case of infection, start empirical ATB coverage and readapt according to culture results",	"Consider protective isolation",

			"Consider protective isolation"	
	Severe Leukopenia	WBC <2000 AND Hemoglobin>10 AND Platelets>100000	"Check for onset of infection or adverse drug effect- consider protective isolation"	"Implement protective isolation following physician request"
Respiratory distress risk	High oxygen requirements/Risk of desaturation	Increase in oxygen requirement (device billing) OR Use of high flow oxygen device	"Perform ABGs and chest imaging" "Perform clinical examination of the patient's lungs. "CBC with differential, routine chemistries, LFTs, CRP, procalcitonin, and according to situation: troponin, lactic acid level, coagulation profile ", "Order sputum or aspirate culture" "Start empirical ATB coverage or readapt existing ATB according to culture results", "Perform clinical examination to R/O pulmonary embolism or DVT", "Monitoring q4h of vital signs, consciousness level, patient color", "Consider transfer to ICU if aggravation"	"Monitor the effects of sedation and analgesics on the patient's respiratory pattern; use judiciously". "Suction as necessary", "Help patient deep breathe and perform controlled coughing" "If the patient is permitted to eat, provide oxygen to the patient but differently (changing from mask to a nasal cannula)." "Maintain an oxygen administration device as ordered, attempting to maintain oxygen saturation at 90% or greater."

	Frequent need for suction/ Risk of obstruction	Use of aspiration tube (device billing)		Suction as necessary Monitor patient's behavior and mental status for the onset of restlessness, agitation, confusion, and (in the late stages) extreme lethargy
	BiPAP/Risk of hypercapnia	Use of BiPAP (device billing)	"Monitor with ABGs regularly"	"Verify BiPAP mask face sealing" "Monitor water level in BiPAP machine" "Apply prevention for BiPAP related pressure ulcer"
Cardiac risk	Elevated troponin	Troponin>0.04 AND Age>50 AND Serum Creatinine<2	"Order EKG" "repeat troponin cycle +EKG", "Evaluate for chest pain and cardiac risk factors" "Order cardiac consult" "Monitor vital signs and report chest pain and tachycardia"	"Vitals monitoring q2h and monitor for chest pain",
	Elevated and Increasing troponin	Increase in Troponin>0.04 AND Age>50 AND Serum Creatinine<2	"Order continuous monitoring" "Consider transfer to ICU if aggravation"	"Verify good function of IV access", "Implement continuous monitoring, assess for chest pain"
	Heart Failure	proBNP>900 AND Serum Creatinine<1.5	"Check for fluid overload" "Check for respiratory distress, dyspnea"	"Monitor fluid status", "Assess respiration rate and depth and report any distress"

			“R/O infection or ischemic changes” “Start diuretics and monitor urine output” “Monitor regularly vital signs and assess for desaturation, tachycardia, hypotension” “Re-evaluate HF treatment”	“Daily body weight”, "Verify/Adapt patient diet (low sodium diet)”, “Promote daily activity within patient tolerance”
Renal insufficiency risk	Acute Kidney Injury	Increase of Serum Creatinine >0.5mg/dl within 3 days	"Urine output monitoring", "IV hydration", "Identify and stop possible nephrotoxic drugs", "Correction of electrolyte and acid-base abnormalities", "Ultrasound of kidneys to evaluate kidney size and presence of obstruction", "Monitor kidney function through lab panel (Serum Creatinine, BUN, electrolytes and Serum Bicarbonate)"	“Monitor fluid intake and output” “Assess for presence of blood in urine”, "Verify/Adapt patient diet", “Check electrolyte balance to monitor evolution, and restore fluid balance according to prescription”, “Assess and report early signs of infection such as fever”, “Assess and report oliguria/ anuria”, “Verify with physician that medications are adapted for renal doses”,

	Uremia	Blood Urea Nitrogen (BUN) >50 mg/dl or increase in BUN of more than 20 points within 3 days	"Urine output monitoring", "IV hydration", "Identify and stop possible nephrotoxic drugs", "Correction of electrolyte and acid-base abnormalities", "Monitor kidney function through lab panel (Serum Creatinin, BUN, electrolytes and Serum Bicarbonate)" “Consider acute dialysis in severe cases”	“Monitor fluid intake and output” “Assess and report oliguria/ anuria”, “Assess for presence of blood in urine, anorexia, confusion, lethargy, bleeding, itching, fetor mouth odor” “Check electrolyte balance to monitor evolution, and restore fluid balance according to prescription”, “Assess and report early signs of infection such as fever”, “Verify with physician that medications are adapted for renal doses”,
	Renal Failure	Serum Creatinine>3	"Identify and stop possible nephrotoxic drugs", "Verify/Adapt patient diet" “Monitor kidney function through lab panel (crea, BUN, electrolytes and Bicarbonate)”, "Urine output monitoring",	“Verify with physician that medications are adapted for renal doses and not nephrotoxic” “Monitor fluid intake and output” “Limit fluid intake (IV and oral) as ordered”

			"intervention20":"Correction of electrolyte and acid-base abnormalities", “Assess for edema, swelling of feet and ankles, and shortness of breath, and chest pain”	“Assess and report edema, swelling of feet and ankles, and shortness of breath, and chest pain” "Adopt a renal diet that avoids high in sodium, high in potassium, high in protein foods", “Monitor and record vital signs to obtain baseline data: report any hypertension” Assess labs for hyperkalemia, electrolyte imbalances and creatinine/BUN raises” “Assess and report nausea and vomiting, ammonia odor on breath, diarrhea, clotted fistula, and signs of infection” “Assess the thrill and the condition of the AV fistula” “Daily weight of the patient” “Monitor glucose levels in diabetic patients” “Seek immediate medical reevaluation in severe hypertension, hyperkalemia,

				dyspnea, altered mental status and acute anuria”
Hepatic insufficiency risk	Risk of liver function perturbation	Increase in SGPT>10 OR Increase in SGOT>10 OR Increase in GGT>20 OR SGPT>100 OR SGOT>100 OR GGT>30	"Monitor LFTs daily, monitor glycemia, monitor chemistry panels, serum albumin and PT/INR as needed", ”Adapt medication treatment to liver function”, "Identify possible etiology for hepatic dysfunction. Order ultrasound imaging if necessary"	“Monitor for icteria, lower limb edema, ascites, petechiae, bleeding and glycemia" “Monitor patient nutritional status and adapt his diet accordingly as per physician’s recommendations” “Assess for any alteration in the patient’s consciousness level.”
	Risk of severe liver dysfunction	(Increase in SGPT>10 OR Increase in SGOT>10 OR Increase in GGT>20 OR) AND decrease of Prothrombin activity of more than 20 pts	“Monitor serum glucose” “Order a high-calorie and a medium to high protein diet” « Restrict fluids and sodium”	"Monitor vital signs, fever, consciousness level, patient color- alert if hypotension, fever, hypoxia/cyanosis, oliguria/anuria, confusion or decreased level of consciousness"

				« Monitor intake and output », “Monitor for icteria, lower limb edema, ascites, petechiae, bleeding and glycemia" “Monitor patient nutritional status and adapt his diet accordingly as per physician’s recommendations” “Assess for any alteration in the patient’s consciousness level.”
Postoperative event risk	Risk of postoperative events	Invasive procedure in the last 48h	"Order and monitor evolution of inflammation markers and hemoglobin level” “Inspect wound for any signs of infection, pain or dehiscence” “Perform physical exam to rule out any surgical complication”	"Monitor vital signs, fever, consciousness level, patient color- alert if hypotension, fever, hypoxia/cyanosis, oliguria/anuria, confusion or decreased level of consciousness" “Verify patient labs for indicators of systemic infection” “Regularly inspect the wound(s), observing its qualities and integrity. Inspect

				breaks or skin irritation around surgical site” “Change wound dressing as per hospital protocol using stringent aseptic techniques” “Assess postoperative pain (characteristics, location, intensity) every 2h and when needed” “Monito function and note output of indwelling catheters and drainage tubes”
Metabolic risk	Dehydration	1.86*(Sodium+Potassium) +1.15*4.75+0.357*BUN+14 >300	“Identify etiology of dehydration and correct accordingly” "IV hydration” "Monitor kidney function through lab panel (Serum Creatinin, BUN, electrolytes and Bicarbonate", "Correction of electrolyte and acid-base abnormalities"	"Urine output monitoring", “Encourage fluid intake”, “Monitor and record vital signs to obtain baseline data: report any hypotension, tachycardia or change in mental status” “Assess skin turgor”, “Monitor active fluid loss from wound drainage, tubes, diarrhea, bleeding, vomiting, etc.”

				“Assess labs for hyperkalemia, electrolyte imbalances and creatinine/BUN raises” “Weight patient daily”
	Fluid overload	1.86*(Sodium+Potassium)+1.15*4.75+0.357*BUN+14 <265	“According to patient case, consider loop diuretics and reduce IV hydration” “Identify and correct etiology (renal, hepatic, cardiac, etc.)” “Monitor urine output” “Correction of electrolyte and acid-base abnormalities", “Monitor vital signs, and report tachycardia and hypertension”,	« Monitor intake and output », “Verify and adapt IV serum flow rate”, “Daily weight of the patient”, “Monitor vital signs, especially for tachycardia and hypertension”, “Assess labs for electrolyte imbalances”, “Educate patient and enforce fluid restriction”, “Administer diuretics after prescription by medical team”, “Review dietary restriction depending on cause of fluid overload (heart failure, renal failure, etc.)”,

				“Assess patient and report: cough, shortness of breath, any altered mental status or anxiety, edema, swelling of feet and ankles”
	Metabolic acidosis	Serum Bicarbonate<20	"Identify etiology of metabolic acidosis (dilutional, renal insufficiency, sepsis, ischemia, severe diarrhea/fluid loss from lower GI tract, etc.) and correct according to cause " "Monitor kidney function through lab panel (Serum Creatinin, BUN, electrolytes and Bicarbonate"), “Monitor urine output” "Correction of electrolyte and acid-base abnormalities", “Monitor vital signs and assess for hypotension, desaturation, respiratory rate“ “Perform ABGs”	"Monitoring q4h of vital signs, consciousness level, heart rhythm, patient color- alert if tachycardia, change in respiratory rate or depth, headache, seizures, confusion or decreased level of consciousness" “Monitor fluid intake and output” “Monitor for electrolytes, especially potassium” “Ensure availability of Sodium Bicarbonate if needed” "Consider patient as high risk for fall"
	Metabolic alkalosis	Serum Bicarbonate>30	"Identify etiology of metabolic alkalosis (dehydration, excess use of diuretics, etc.) and correct according to cause" “Monitor urine output”	"Monitoring q4h of vital signs, consciousness level, heart rhythm, patient color, neuromuscular status- alert if tingling/numbness, severe vomiting, change in respiratory rate or depth,

			"Correction of electrolyte and acid-base abnormalities", “Monitor vital signs and assess for hypotension, desaturation, respiratory rate“ “Perform ABGs”	headache, seizures, confusion or decreased level of consciousness" "Ensure good patient IV hydration" “Restrict oral intake and encourage intake of foods high in potassium and calcium” “Identify potassium -losing drugs such as thiazide and furosemide and discuss their discontinuation with medical team”
	Hyponatremia	Sodium<135	“Order blood and urine osmolarity, and urine electrolytes “ "Determine etiology of hyponatremia (fluid overload, infection, SIADH, etc.) and correct according to cause" “Adapt type and quantity of IV hydration according to etiology” “Stop medications that can aggravate hyponatremia (certain families of pain killers, SSRIs, etc.)”	"Verify/Adapt patient diet card and tray", “Fluid restrictions to prevent dilution of sodium” "Monitor fluid intake and output; Calculate fluid balance to prevent overload", "Monitoring vital signs, consciousness level- alert if hypotension, confusion, headache, irritability or decreased level of

			“Correct hyponatremia gradually according to guidelines” “Assess for neurological status alteration (coma, somnolence, confusion…)”	consciousness and seizure, edema or muscle cramps” “Apply seizure and fall precautions”
	Hypernatremia	Sodium>145	“Order blood and urine osmolarity, and urine electrolytes “ "Determine etiology of hypernatremia (dehydration, osmotic diuretics, diarrhea, etc.) and correct according to cause " “Check for severe dehydration” “Adapt type and quantity of IV hydration according to etiology” “Correct hypernatremia gradually according to guidelines” “Stop medications that can aggravate hypernatremia (diuretics, IV saline perfusion, etc.)” “Assess for neurological status alteration (coma, somnolence, confusion…)”	"Verify/Adapt patient diet card and tray" (avoid food rich in sodium”, “Retrict sodium intake” “Apply regular mouth care” "interventionN10": "Ensure good patient hydration (PO/IV)", "interventionN11":"Monitoring of vital signs, consciousness level, patient color- alert if hypertension, hypotension, confusion or decreased level of consciousness", “Monitor fluid intake and output” “Apply seizure and fall precautions”

	Hypokalemia	Potassium <3	"Repeat potassium level” “Obtain stat ECG and examine for changes relative to hypokalemia.” “Perform neurological and muscular clinical assessment.” “Correct hypokalemia with potassium supplementation if needed"	"Monitoring q4h of vital signs, assess for tachycardia, hypoventilation and altered neuromuscular function such as tetany, paresthesia, apathy, drowsiness, irritability", "Verify/Adapt patient diet card and tray (potassium rich food)", “IV potassium supplementation after order from physician”, "When giving potassium supplement through IV, ensure controlled delivery of medication to prevent bolus effect and reduce associated discomfort such as burning sensation at IV site"
	Hyperkalemia	Potassium>5.5	"intervention37a":"Repeat potassium level (check for sample hemolysis)”, “Obtain stat ECG and examine for changes of hyperkalemia”, “Correct hyperkalemia with calcium gluconate or insulin with dextrose if needed",	“Monitoring q4h of vital signs, heart rate and rhythm, respiratory rate- assess for bradycardia, hypoventilation and decreased level of consciousness or neuromuscular function (muscular paresthesia or weakness)", “Monitor urine output”,

			“Check for heart rate irregularities”	"Verify/Adapt patient diet card and tray (reduce source of potassium and encourage intake of carbohydrates", “Limit or stop medications containing potassium after discussion with medical team”
	Hyperlactatemia/lactic acidosis	Serum lactate >4	"Identify hyperlactatemia etiology (sepsis, severe ischemia and/or shock, etc.)", "Perform ABGs to check for acidosis", "Consider transfer to intensive care unit"	"Monitoring q2h of vital signs, consciousness level, patient color" "Consider continuous monitoring of vitals" “Start 2 large bore IVs for fluid resuscitation” “Monitor fluid intake and output” “Administer oxygen if saturation is less than 94%” “Place patient on cardiac monitor” “Check labs to ensure patient lactate levels are dropping”

				“Check peripheral pulses”
Vascular/Circulatory/ Hematological risk	Thrombocytopenia	Platelets<50 000 AND Hemoglobin >10 AND WBC>4000	"Identify thrombocytopenia etiology: adverse drug effect, hepatic dysregulation, etc.", "Check for signs of hemorrhage (purpura, petechia…)", “Review anticoagulation medications”,	"Monitor for bruises", “Minimize patient activity that can cause bleeding or fall, and educate the at-risk patient and caregivers about precautionary measures to prevent tissue trauma or disruption of the normal clotting mechanisms. “Review laboratory results for coagulation status as appropriate.” “Avoid use of restraints; obtain a physician’s order if restraints are needed”.
	Pancytopenia	Platelets<50 000 AND Hemoglobin <9 AND WBC<2000	“Check for signs of hemorrhage (purpura, petechia…)”, “Check for adverse drug effect”, “Review anticoagulation medications”, “Monitor for signs of bleeding”, “Consider protective isolation”	"Monitor for bruises", “Minimize patient activity that can cause bleeding or fall, and educate the at-risk patient and caregivers about precautionary measures to prevent tissue trauma or disruption of the normal clotting mechanisms.

				“Review laboratory results for coagulation status as appropriate.” “Avoid use of restraints; obtain a physician’s order if restraints are needed”. “Apply protective isolation precautions”

	Disseminated Intravascular Coagulation	Platelets <100000 AND decrease of Prothrombin activity of more than 20 pts AND D -Dimer>1000	"Identify etiology of DIC and provide treatment for the underlying disorder (infectious, oncological, etc.)" "Consider transfer to higher level of care"	"Assess for changes in the level of consciousness" "Assess the respiratory depth, rate, and rhythm, in addition to breath sounds. Assess cough for signs of bloody sputum." "Assess for tachycardia, shortness of breath, and use of accessory muscles." "Monitor oxygen saturation and assess arterial blood gases (ABGs)." "Assist with coughing or suction as indicated." "Anticipate the need for intubation and mechanical ventilation." "Change patient position every 2hours and position the patient in a high-Fowler's position as indicated" "Maintain an oxygen administration device as ordered."

	Hemorrhage	(Decrease in Hemoglobin of more than 1.5 pts AND (1.86*(Sodium differential+Potassium differential) + 0.357* (BUN differential)) > (-20) AND Chloride differential <7) OR (Decrease in Hemoglobin of more than 1.5 pts AND (1.86*(Sodium+Potassium)+1.15*4.75+0.357*BUN+14) <300 AND (1.86*(Sodium+Potassium)+1.15*4.75+0.357*BUN+14)>265)	"Perform clinical evaluation to R/O bleeding- Obtain blood group, order coagulation panel and medical imaging if necessary", “Assess the effect of use of anticoagulants and NSAIDs that can induce or affect bleeding” “Collect urine and stool samples and test for occult blood” “Monitor vital signs and assess for hypotension or tachycardia”	"Verify good function of IV access", "Monitor q2h and according to severity and need: vital signs, consciousness level, patient color- alert if hypotension, hypoxia/paleness, oliguria/anuria, confusion or decreased level of consciousness" “Check stool and urine for occult blood”, “Assess skin for signs of petechia, signs of bruising”, "Consider patient as high risk for fall"
	Anemia	(Hemoglobin <8 AND (1.86*(Sodium differential+Potassium differential) + 0.357* (BUN differential)) > (-20) AND Chloride differential <7) OR (	"Identify anemia etiology/perform anemia workup", "Monitoring q4h of vital signs, consciousness level, patient color", "Consider blood transfusion"	“Assist the patient in prioritizing activities and establishing balance between activity and rest that would be acceptable to the patient” “Encourage a healthy diet that is packed with essential nutrients” "Monitor vital signs, consciousness level, patient

		Hemoglobin <8 AND (1.86*(Sodium+Potassium)+1.15*4.75+0.357*BUN+14) <300 AND (1.86*(Sodium+Potassium)+1.15*4.75+0.357*BUN+14)>265)		color- alert if hypotension, hypoxia/paleness, oliguria/anuria, confusion or decreased level of consciousness" “Assess for vertigo” “Consider patient as high risk for fall”
	Risk of blood clotting disorder/Emboli formation	D-Dimer > 500	"Adjust/ Add anticoagulants", "Monitoring q4h of vital signs, consciousness level, patient color"	"Monitor signs of heat, redness, edema or pain in limbs", "Monitor vital signs, consciousness level, patient color. Observe changes in cardiac rhythm and respiration rate and rhythm”
	Acute risk of blood clotting disorder/Emboli formation	Increase in D-Dimer > 500	"Perform clinical and if necessary radiological examination to R/O pulmonary embolism or DVT", "Adjust/ Add anticoagulants", "Monitoring q4h of vital signs, consciousness level, patient color"	"Monitor vital signs, consciousness level, patient color. Observe changes in cardiac rhythm and respiration rate and rhythm” “Arrange for someone to stay with the patient, as indicated”.
	Coagulation disorder	Decrease in Prothrombin Activity of at least 20%	“Check for anticoagulant treatment effect and correct accordingly “ “Assess liver function through liver enzymes and coagulation profile”	“Assess the patient for any signs of bleeding and teach the patient and family how to lower the risk of bleeding (trauma, risk of fall, accidental cuts, etc.) and what symptoms to look for that require

			“Check for bleeding signs and manage bleeding if necessary” “Consider administration of Vitamine K if needed”	medical attention (headache or altered mental status, vomiting blood, dark urine, change in heart rhythm or blood pressure, joint pain, etc.).”
Nutritional risk	Nasogastric tube feeding/Increased risk of aspiration pneumonia	Use of nasogastric tube (device billed)	“Consider ordering ACE inhibitors for the reduction of risk of aspiration pneumonia in at-risk patients who also require blood pressure control” “If aspiration is doubted, order to put several drops of blue or green food coloring in tube feeding to help test for aspiration”	“Check placement before feeding, using tube markings, x-ray study (most accurate), pH of gastric fluid, and color of aspirate as guides.” “If ordered by physician, put several drops of blue or green food coloring in tube feeding to help indicate aspiration. In addition, test the glucose in tracheobronchial secretions to detect aspiration of enteral feedings.” “Elevate the head of bed to 30 to 45 degrees while feeding the patient and for 30 to 45 minutes afterward if feeding is intermittent. Turn off the feeding before lowering the head of bed. Patients with continuous feedings should be in an upright position.” “Position patients with a decreased level of consciousness on their side.”

				"Apply good oral hygiene in elderly patients to prevent aspiration pneumonia"
	Severe hypoalbuminemia/Increased risk of edema/loss of muscle/pressure ulcer	Serum albumin < 3	"Determine and manage etiology of hypoalbuminemia through performing LFTs, urine albumin, protein measurement, BNP/NT-proBNP" and inflammation markers"	"Encourage patient to eat a balanced diet full of dairy, protein and whole-grain carbohydrates or taking supplements to increase the amount of protein and calories in his diet, and to removing foods high in sodium (salt) from his diet", "Assess for peripheral edema in the lower extremities" "Monitor urine output", "Monitor level of serum albumin in order to assess outcome of interventions"
Interpretation risk	Possible serum contaminated blood sample	Increase of Serum Chloride of more than 10 pts	Order another sample	Withdraw another sample

6.3.4. Validation results of the different application components

Results still under process

6.3.5. Usability evaluation results

Results still under process

6.4. Discussion (part of the discussion will be waiting the completion of the data collection phase)

6.4.1. Possible applications of the tool in the clinical workflow

"VIGIL" consolidates in one view a digest of information regarding the patient case, its overall deterioration risk score, and the different clinical risks that are stemming from his/her clinical status. It can be used by the medical team for example in medical "sitting rounds" in order to discuss on a daily basis the evolution of the patient's clinical risks, and decide upon adequate actions to counter or manage those risks. The tool can provide a personalized risk assessment and management report for each patient, against which the clinicians can critically compare their evaluations and clinical decisions.

On a more administrative level, the tool can be exploited to identify "high risk" patients on regular floors or post transfer from critical care units, in order to verify medical and nursing staffing issues, provide appropriate monitoring and risk management, or even study appropriateness of transfer decisions.

Moreover, it can be possible to use the tool to filter "high risk" patients and round on them at night by the hospital night supervisors, so as to verify their condition and the care they are receiving.

In addition to that, "VIGIL" can be used as a practical educational tool for medical or nursing interns, since it has the advantage of presenting real and up-to-date cases of patients in the hospital.

7. Discussion, Perspectives, and Conclusions

7.1. Potential impact of the elaborated AI-enabled models on patient safety practice

- **Improving the detection performance of AEs relative to gold standard tools**

  The rules-based tool developed in the first study was tested prospectively and used in the real workflow on daily basis, as a tool for the Patient Safety Department. Till date, few such tools have been used routinely in practice, beyond validation or pilot study.
  This tool mimics to a certain extent the reasoning process of a clinical auditor when trying to identify an AE in the patient medical file.
  When comparing the results given by this tool in matter of AE detection, to the standard technique which is used by most hospitals to detect AEs, namely the incident reporting system, it was found that in 9 months, 394 AEs were identified, occurring with 291 patients, with an average of 1.4 AEs identified per surveillance day. Whereas only x% of these AEs were reported through the incident reporting system, which shows a significant gain in the detection performance of AEs. This result is also well documented in the literature, when comparing automated AE detection tools to AEs detected through incident reporting systems (reference).
  When compared to another audit tool used for determining AE incidence in a hospital or healthcare system, which is the paper based chart review, this rules-based automated method permits a systematic, reliable and routine use for AE detection surveillance, which cannot be performed with chart audit due to its resource intensiveness.

  The ML-based model relative to the AE-related readmissions developed in the second study portrays a potential of machine learning algorithms in improving the detection performance of trigger-based or even rule-based systems. In fact, data driven ML algorithms have performance advantages over rules-based approaches, as they allow simultaneous consideration of multiple data sources to identify predictors and outcomes [25]. Regarding the identification of AI-related 30-days patient readmissions, the potential is to improve the positive predictive value of the current detection trigger (which is estimated to be around 12%) to at least 50% in timely manner and using immediately available variables in the HIS. For this purpose, the two methods gain to be used consecutively to obtain the best results.

- **Provide prediction of various patient safety risks and events**

  The third study relative to the prediction of clinical deterioration highlights the potential of using ML-algorithms with a complex array of biomarkers, and an expert labeled high quality dataset to derive a model for the prediction of the risk of clinical deterioration in patients. In order to further improve the performance of the model, a study of its performance relative to subgroups of patient deterioration diagnosis (e.g sepsis, pneumonia, heart failure decompensation, etc.). This will be performed in the prospective evaluation of the model (fourth study – "VIGIL").
  The model can be also used to evaluate the risk of readmission at the point where a discharge decision is made by the treating team, to further examine the conditions of discharge and prevent hospital-related complications or AEs which can lead to readmission within 30 days and the associated clinical and financial burden.

- **Improve adherence to best practice in terms of managing clinical deterioration risks**

  In the fourth study, the application "VIGIL" combined the deterioration prediction capacity with a rules-based identification of clinical risks that could lead to clinical deterioration. Based on these identified risks, the application used a second rules-based module based on clinical practice guidelines to recommend medical and nursing minimal interventions. These recommendations can be useful both for the medical teams (especially residents or interns) and nursing teams as a reference that provides timely information relative to their patients. This tool can be used in the clinical workflow but also as an educational tool in the training curriculum of both medical and nursing professionals.
  These recommendations do not replace the physician's judgement nor are a mandatory step in the clinical workflow. However, as an independent reference, they can be used to double check the clinical management, or guide any investigation in case of a suspected preventable patient deterioration.

- **The potential link between diagnostic aid and prediction of deterioration: towards clinical practice evaluation**

  Theoretically, the variation of the clinical deterioration prediction score can be correlated to the clinical evolution of the patient. In cases where clinical deterioration can be potentially reversed, measuring the dynamics of the patient's deterioration score plot during his hospital stay can be evaluated as an outcome indicator of clinical practice. For example, it can be used to compare between clinical outcomes of same type of patient diagnosis between physicians or between healthcare institutions.

The medical and nursing processes generally follow the following logical steps that form a cycle: clinical evaluation-diagnosis- care planification- care implementation- and finally clinical reevaluation.
Therefore, the evaluation and diagnosis steps play a very important role in the whole clinical process and affect directly the process outcomes.
Coupling a module for evaluation/diagnostic aid to clinicians to a clinical outcome score in the same application (such as in the fourth study, "VIGIL") promises to open a new approach to the evaluation of clinical practices and to patient safety research, which is lately focusing on the diagnostic errors phenomenon [13], its mechanisms and impact.

7.2.Rules-based approach and machine learning: the power of synergy

- **Using rules-based inferences to improve explainability of ML algorithms**

Explainability of ML algorithms is a major challenge that can affect the adoption of such a technology by clinicians and healthcare institutions. This challenge is also more or less complex depending on the type of algorithm involved. For example, multilayered perceptron neural networks and similar structures used in Deep Learning are considered as "black boxes".
Traditionally, two approaches are commonly used by developers and clinicians for AI-driven applications.
The first is to use exclusively "explainable algorithms", which are algorithms that are relatively easily understandable and interpretable by the users (such as Logistic Regression or Decision Tree). This, however, might limit the level of complexity of the algorithm, and with that negate the possible benefits of using AI.
The second is using more complex algorithms but with computing a statistical "weight" for each feature (explanatory variable) used in the algorithm. In fact, for clinicians, it is important to know which features are considered relevant by the algorithm and how much weight is assigned to this feature. Having that information, a clinician can judge whether the features that the system picks out are indeed relevant or not [115] and evaluate algorithm outcomes accordingly, even if he doesn't the exact explanation on how the outcome was decided by the system.

In our fourth study (the "VIGIL" application), where a Random Forest algorithm for the prediction score, we have adopted two complementary approaches. First, the classical approach, where we displayed the values of all main features (with significant weight) along with the overall prediction score in order for the clinician to try to interpret the result according to the values of the main features. Second, we adopted another approach, which can be described as adding an interpretative layer based on expert system inferences. Practically, a number of rules were programmed to yield a number

of inferences in the form of "clinical risks" based on the values and combination of values of the different variables used by the predictive model. In addition to that, the critical values of the model variables were highlighted in red to facilitate identification and interpretation by the user. This constitutes an example of how rules-based inferences can be used to improve the explainability of ML algorithms.

- **Using rules-based inferences to augment the dataset of ML algorithms and improve their prediction performance**

  One of the biggest challenges faced in the elaboration of ML-derived models is to ensure a quality dataset with a sufficient volume (rule of thumb of at least 10 event cases per feature). In the domain of clinical deterioration (third study), high quality labeling needs a verification by human experts, especially if subtypes of deterioration (respiratory deterioration, cardiac deterioration, infectious deterioration, etc.) need to be identified for further fine tuning of the model. This renders the process resource intensive and limits the speed of dataset constitution.
  Expert rules-based inferences about "clinical risks" based on the model features (such as the ones used in the "VIGIL" study) can help identify, in prospective or even retrospective implementation, certain high risk categories of patients that can be for example cross-matched with the database about transfers from regular floors to critical care in the hospital information system, in order to prepare new cohorts by type of deterioration risk, that can be added to the dataset and require less time intensive expert confirmation.
  Furthermore, it could be practical to add this possibility in the application, so that the clinician user can directly "label" a certain patient case and add it to the updated dataset of patient cases, while this case is fresh in his/her memory. This could render the application suitable to be "unclocked", with the possibility of automatically updating its algorithms.

- **Using ML algorithms to improve classification performance of rules-based models**

  Rules-based models are very useful in automating a clinical recommendation or a thought process. In the medical field however, every patient is unique when it comes to the combination of his clinical history, his comorbidities, and genetic signature that can impact the body response to diseases as well as to therapeutics.
  When formulating guidelines and recommendations, often in the form of decision trees, a margin is left for the clinician's judgement to adapt the recommendations to the patient case specificities or to identify non-indicated recommendations.
  Rules-based models have a real challenge when dealing with this complexity, and this translates in a lower performance than expected in the classification or predicting

functions. For example, the rules-based "AE- related 30 day-readmission" in the first study has a PPV of less than 12%. The reason behind this is partially the high number of scenarios for readmission that cannot be easily modeled through strict rules. For example, a hospitalized patient may be readmitted for a complementary already planned elective procedure (e. g planned removal or readjustment of a urinary stent) or further to new findings (e. g anapath results), which is not considered as an AE.
Where rules-based models can easily draw the "main" lines in the thought process and be as close as possible to clinical recommendations guidelines, a ML add-on can potentially help in handling this variability of sub-scenarios, since it can allow simultaneous consideration of multiple predictors and optimize fitting to the desired outcomes.

7.3. CDSS vs. AI Models: The map is not the territory

- **Insight into the maturation process from predictive model to a CDSS software**
  Very few predictive models grow to become CDS applications. And this is not mainly due to a lack of programming skills needed to transform the model into a software application. The reason behind this challenge lies more into a number of important but delicate steps that are mandatory to cross in this maturation process.
  The path from the model described in the third study to the application described in the fourth study included the following milestones.
    - *Setting a cutoff value for the classifier* relative to discrimination between classes of patients, relative to the risk of deterioration. From what prediction score value do we consider a patient to have a significant risk of clinical deterioration? This threshold was set through performing an evaluation of the model results on a prospective labeled dataset, and measuring the AUROC of the classifier. The threshold for the classifier was then identified by setting a minimal value for the true positive rate (TPR) and a maximal value for the false positive rate (FPR), and optimizing the threshold to fulfill this condition. Hence, the threshold was selected to be 0.85, for a minimal TPR of 0.7 and a maximal TPR of 0.3.
    - *Solving the missing data and uncertainty of prediction problems.* Predictive model are born good but incomplete data makes them corrupt!
    - In fact, the derivation process of models is performed on a complete dataset where each feature has a defined value. The algorithm then computes a model that fits at best the values of the features to the desired outcome variables. When used prospectively (for example in a CDS application), all features do not have always have values (e. g not all patients have a CRP exam in their labs ordered), and sometimes this can be the case with features that have a significant importance or weight for the model outcomes. We have opted in our application to identify and label as "uncertain" predictions where values for any of the top model features in terms of predictive importance are missing.

- *Addressing the explainability challenge.* As mentioned in section 7.2, explainability is key to improve clinician users' adoption of the application, and to support efficiently and reliably their decisions. The approach we adopted was based on highlighting the values of "important" features, in addition to formulating rules-based inferences ("clinical risks") based on the values of the same features as the predictive model.
- *Identifying the added value of the application for the users.* For the application to be adopted by clinicians, it needs to provide an added value to their work while introducing no extra burden to their workflow. We have opted to install the application in the medical team's office, where medical attendings and residents check the patient exam results, discuss the patient cases and document in the medical files. Also, it was installed on the nursing desk where nurses uses the HIS and document in the patient file. The application intent in this workflow would ideally be to play a role of a clinical reference for the medical and nursing teams in order to double check their evaluations, their identified risks for the patients and their management plan, against the data and recommendations provided by the application. This implementation will be the subject of a usability study with a sample of users, in order to better understand the value of the solution for the users and the barriers that need to be removed.

- **The need for a specific validation framework for AI-enabled CDS applications**

  The evaluation of an AI-enabled CDS involves a bigger complexity than the evaluation of a AI-derived model. In fact, multiple dimensions of the application need to be scrutinized and reported for the evaluation to be comprehensive.
  A validation framework for AI-enabled CDS applications is yet to be elaborated and adopted. However, the approach proposed by Greenes et al. [1] seems to identify the main components for such a framework.
  For our fourth study, we opted to tackle the evaluation challenge using this approach, combining a classical prospective evaluation of the prediction model using AUROC for discrimination performance and calibration plot for the goodness of fit of the model to real probabilities of events.
  For the application functionalities, each rules-based function ("clinical risks" and "medical recommendations" and "nursing recommendations") was tested against expert labeled data, using a confusion matrix.
  As for the application usability, it will be object to a usability study in order to determine its user friendliness, its added value in their workflow and their suggestions in order to optimize its usability.
  Finally a "human factors" analysis will be also planned with a focus group to evaluate the risks and impacts of this application on patients and on the clinician users.

Bibliography

[1] R. A. Greenes, D. W. Bates, K. Kawamoto, B. Middleton, J. Osheroff, and Y. Shahar, "Clinical decision support models and frameworks: Seeking to address research issues underlying implementation successes and failures," *Journal of Biomedical Informatics*, vol. 78. 2018. doi: 10.1016/j.jbi.2017.12.005.

[2] T. Hernandez-Boussard, S. Bozkurt, J. P. A. Ioannidis, and N. H. Shah, "MINIMAR (MINimum information for medical AI reporting): Developing reporting standards for artificial intelligence in health care," *Journal of the American Medical Informatics Association*, vol. 27, no. 12. 2020. doi: 10.1093/jamia/ocaa088.

[3] D. W. Bates, A. Auerbach, P. Schulam, A. Wright, and S. Saria, "Reporting and implementing interventions involving machine learning and artificial intelligence," *Ann Intern Med*, vol. 172, no. 11, 2020, doi: 10.7326/M19-0872.

[4] C. Silcox, S. Dentzer, and D. W. Bates, "AI-Enabled Clinical Decision Support Software: A 'Trust and Value Checklist' for Clinicians," *NEJM Catal*, vol. 1, no. 6, 2020, doi: 10.1056/cat.20.0212.

[5] World Health Organization, *Global patient safety action plan 2021–2030:Towards eliminating avoidable harm e health care*. 2021.

[6] L. Slawomirski, A. Auraaen, and N. S. Klazinga, "The economics of patient safety - Strengthening a value-based approach to reducing patient harm at national level: Organisation for Economic Co-operation and Development - OECD; 2017 [15th Jan. 2018].," *OECD Health Working Papers No.96*, no. 96, 2017.

[7] NASEM, *National Academies of Sciences Engineering and Medicine: Crossing the global quality chasm: Improving health care worldwide*. 2018.

[8] M. E. Kruk *et al.*, "High-quality health systems in the Sustainable Development Goals era: time for a revolution," *The Lancet Global Health*, vol. 6, no. 11. 2018. doi: 10.1016/S2214-109X(18)30386-3.

[9] Y. Wang *et al.*, "National Trends in Patient Safety for Four Common Conditions, 2005–2011," *New England Journal of Medicine*, vol. 370, no. 4, 2014, doi: 10.1056/nejmsa1300991.

[10] R. J. Baines *et al.*, "Changes in adverse event rates in hospitals over time: A longitudinal retrospective patient record review study," *BMJ Qual Saf*, vol. 22, no. 4, 2013, doi: 10.1136/bmjqs-2012-001126.

[11] C. P. Landrigan, G. J. Parry, C. B. Bones, A. D. Hackbarth, D. A. Goldmann, and P. J. Sharek, "Temporal Trends in Rates of Patient Harm Resulting from Medical Care," *New England Journal of Medicine*, vol. 363, no. 22, 2010, doi: 10.1056/nejmsa1004404.

[12] N. Eldridge *et al.*, "Trends in Adverse Event Rates in Hospitalized Patients, 2010-2019," *JAMA*, vol. 328, no. 2, pp. 173–183, Jul. 2022, doi: 10.1001/JAMA.2022.9600.

[13] D. W. Bates and H. Singh, "Two decades since to err is human: An assessment of progress and emerging priorities in patient safety," *Health Aff*, vol. 37, no. 11, 2018, doi: 10.1377/hlthaff.2018.0738.

[14] M. A. Makary and M. Daniel, “Medical error-the third leading cause of death in the US,” *BMJ (Online)*, vol. 353, 2016, doi: 10.1136/bmj.i2139.

[15] K. N. Slawomirski L, Auraaen A, “The economics of patient safety - Strengthening a value-based approach to reducing patient harm at national level: Organisation for Economic Co-operation and Development - OECD; 2017 [15th Jan. 2018].,” *OECD Health Working Papers*, no. 96, 2018.

[16] C. Zhan and M. R. Miller, “Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries during Hospitalization,” *J Am Med Assoc*, vol. 290, no. 14, 2003, doi: 10.1001/jama.290.14.1868.

[17] L. Adler *et al.*, “Impact of Inpatient Harms on Hospital Finances and Patient Clinical Outcomes,” *J Patient Saf*, vol. 14, no. 2, 2018, doi: 10.1097/PTS.0000000000000171.

[18] L. Tessier, S. J. T. Guilcher, Y. Q. Bai, R. Ng, and W. P. Wodchis, “The impact of hospital harm on length of stay, costs of care and length of person-centred episodes of care: A retrospective cohort study,” *CMAJ*, vol. 191, no. 32, 2019, doi: 10.1503/cmaj.181621.

[19] J. Kjellberg *et al.*, “Costs associated with adverse events among acute patients,” *BMC Health Serv Res*, 2017, doi: 10.1186/s12913-017-2605-5.

[20] J. C. Goodman, P. Villarreal, and B. Jones, “The social cost of adverse medical events, and what we can do about it,” *Health Aff*, vol. 30, no. 4, 2011, doi: 10.1377/hlthaff.2010.1256.

[21] World Health Organization, “Minimal Information Model for Patient Safety Incident Reporting and Learning System (user guide),” *SpringerReference*, 2016.

[22] L. Kohn, J. Corrigan, and M. S. Donaldson, “Committee on Quality of Health Care in America. To Err Is Human: Building a Safer Health System,” *National Academy Press*, 1999.

[23] D. M. Woods, E. J. Thomas, J. L. Holl, K. B. Weiss, and T. A. Brennan, “Ambulatory care adverse events and preventable adverse events leading to a hospital admission,” *Qual Saf Health Care*, vol. 16, no. 2, 2007, doi: 10.1136/qshc.2006.021147.

[24] A. K. Jha, I. Larizgoitia, C. Audera-Lopez, N. Prasopa-Plaizier, H. Waters, and D. W. Bates, “The global burden of unsafe medical care: Analytic modelling of observational studies,” *BMJ Qual Saf*, vol. 22, no. 10, 2013, doi: 10.1136/bmjqs-2012-001748.

[25] D. W. Bates *et al.*, “The potential of artificial intelligence to improve patient safety: a scoping review,” *npj Digital Medicine*, vol. 4, no. 1. 2021. doi: 10.1038/s41746-021-00423-6.

[26] “Global trigger tool implementation guide :: Health Quality & Safety Commission.” https://www.hqsc.govt.nz/our-work/system-safety/adverse-events/publications-tools-and-resources/tools/trigger-tools/global-trigger-tool-implementation-guide/ (accessed Aug. 09, 2022).

[27] E. J. Thomas and L. A. Petersen, “Measuring errors and adverse events in health care,” *J Gen Intern Med*, vol. 18, no. 1, pp. 61–67, 2003, doi: 10.1046/j.1525-1497.2003.20147.x.

[28] D. C. Classen *et al.*, “‘Global trigger tool’ shows that adverse events in hospitals may be ten times greater than previously measured,” *Health Aff*, vol. 30, no. 4, 2011, doi: 10.1377/hlthaff.2011.0190.

[29] D. J. Cullen, D. W. Bates, S. D. Small, J. B. Cooper, A. R. Nemeskal, and L. L. Leape, “The incident reporting system does not detect adverse drug events: a problem for quality improvement.,” *Jt Comm J Qual Improv*, vol. 21, no. 10, 1995, doi: 10.1016/S1070-3241(16)30180-8.

[30] F. Griffin and R. Resar, “IHI Global Trigger Tool for measuring adverse events,” *IHI Innovation Series white paper*, no. September, 2007.

[31] F. Griffin and R. Resar, “IHI Global Trigger Tool for measuring adverse events,” *IHI Innovation Series white paper*, no. September, 2007.

[32] P. Doupi, H. Svaar, B. Bjørn, E. Deilkås, U. Nylén, and H. Rutberg, “Use of the Global Trigger Tool in patient safety improvement efforts: Nordic experiences,” *Cognition, Technology and Work*, vol. 17, no. 1, 2015, doi: 10.1007/s10111-014-0302-2.

[33] K. Mevik, F. A. Griffin, T. E. Hansen, E. Deilkås, and B. Vonen, “Is inter-rater reliability of Global Trigger Tool results altered when members of the review team are replaced?,” *International Journal for Quality in Health Care*, vol. 28, no. 4, 2016, doi: 10.1093/intqhc/mzw054.

[34] H. J. Murff *et al.*, “Automated identification of postoperative complications within an electronic medical record using natural language processing,” *JAMA - Journal of the American Medical Association*, vol. 306, no. 8, 2011, doi: 10.1001/jama.2011.1204.

[35] G. B. Melton and G. Hripcsak, “Automated detection of adverse events using natural language processing of discharge summaries,” *Journal of the American Medical Informatics Association*, vol. 12, no. 4, 2005, doi: 10.1197/jamia.M1794.

[36] S. N. Musy *et al.*, “Trigger Tool–Based Automated Adverse Event Detection in Electronic Health Records: Systematic Review,” *J Med Internet Res*, vol. 20, no. 5, p. e198, May 2018, doi: 10.2196/jmir.9901.

[37] D. R. Murphy, A. N. D. Meyer, D. F. Sittig, D. W. Meeks, E. J. Thomas, and H. Singh, “Application of electronic trigger tools to identify targets for improving diagnostic safety,” *BMJ Quality and Safety*. 2019. doi: 10.1136/bmjqs-2018-008086.

[38] J. Liang *et al.*, “Adoption of electronic health records (EHRs) in China during the past 10 years: Consecutive survey data analysis and comparison of Sino-American challenges and experiences,” *J Med Internet Res*, vol. 23, no. 2, 2021, doi: 10.2196/24813.

[39] D. Classen, M. Li, S. Miller, and D. Ladner, “An electronic health record–based real-time analytics program for patient safety surveillance and improvement,” *Health Aff*, vol. 37, no. 11, 2018, doi: 10.1377/hlthaff.2018.0728.

[40] D. C. Classen *et al.*, “National Trends in the Safety Performance of Electronic Health Record Systems from 2009 to 2018,” *JAMA Network Open*, vol. 3, no. 5. 2020. doi: 10.1001/jamanetworkopen.2020.5547.

[41] K. M. Gomes and R. M. Ratwani, “Evaluating Improvements and Shortcomings in Clinician Satisfaction with Electronic Health Record Usability,” *JAMA Netw Open*, vol. 2, no. 12, 2019, doi: 10.1001/jamanetworkopen.2019.16651.

[42] D. W. Bates and H. Singh, “Two decades since to err is human: An assessment of progress and emerging priorities in patient safety,” *Health Aff*, vol. 37, no. 11, 2018, doi: 10.1377/hlthaff.2018.0738.

[43] C. M. Rochefort, D. L. Buckeridge, and A. J. Forster, “Accuracy of using automated methods for detecting adverse events from electronic health record data: A research protocol,” *Implementation Science*, vol. 10, no. 1, 2015, doi: 10.1186/s13012-014-0197-6.

[44] A. B. Chapman, K. S. Peterson, P. R. Alba, S. L. DuVall, and O. V. Patterson, “Detecting Adverse Drug Events with Rapidly Trained Classification Models,” *Drug Saf*, vol. 42, no. 1, 2019, doi: 10.1007/s40264-018-0763-y.

[45] Y. Huang *et al.*, “Automated safety event monitoring using electronic medical records in a clinical trial setting: Validation study using the VA NEPHRON-D trial,” *Clinical Trials*, vol. 16, no. 1, 2019, doi: 10.1177/1740774518813121.

[46] M. Govindan, A. D. Van Citters, E. C. Nelson, J. Kelly-Cummings, and G. Suresh, “Automated detection of harm in healthcare with information technology: A systematic review,” *Quality and Safety in Health Care*. 2010. doi: 10.1136/qshc.2009.033027.

[47] R. Freeman, L. S. P. Moore, L. García Álvarez, A. Charlett, and A. Holmes, “Advances in electronic surveillance for healthcare-associated infections in the 21st Century: A systematic review,” *Journal of Hospital Infection*, vol. 84, no. 2. 2013. doi: 10.1016/j.jhin.2012.11.031.

[48] K. J. O’Leary *et al.*, “Comparison of traditional trigger tool to data warehouse based screening for identifying hospital adverse events,” *BMJ Qual Saf*, vol. 22, no. 2, 2013, doi: 10.1136/bmjqs-2012-001102.

[49] D. C. Stockwell, E. Kirkendall, S. E. Muething, E. Kloppenborg, H. Vinodrao, and B. R. Jacobs, “Automated adverse event detection collaborative: Electronic adverse event identification, classification, and corrective actions across academic pediatric institutions,” *J Patient Saf*, vol. 9, no. 4, 2013, doi: 10.1097/PTS.0000000000000055.

[50] C. Sammer *et al.*, “Developing and Evaluating an Automated All-Cause Harm Trigger System,” *Jt Comm J Qual Patient Saf*, vol. 43, no. 4, 2017, doi: 10.1016/j.jcjq.2017.01.004.

[51] D. Classen, M. Li, S. Miller, and D. Ladner, “An electronic health record–based real-time analytics program for patient safety surveillance and improvement,” *Health Aff*, vol. 37, no. 11, 2018, doi: 10.1377/hlthaff.2018.0728.

[52] M. M. Shabot, M. R. Chassin, A. C. France, J. Inurria, J. Kendrick, and S. P. Schmaltz, “Using the targeted solutions Tool® to improve hand hygiene compliance is associated with decreased health care-associated infections,” *Jt Comm J Qual Patient Saf*, vol. 42, no. 1, 2016, doi: 10.1016/s1553-7250(16)42001-5.

[53] P. Pronovost *et al.*, “An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU,” *New England Journal of Medicine*, vol. 355, no. 26, 2006, doi: 10.1056/nejmoa061115.

[54] The Leapfrog Group, “2017 Healthcare- Associated Infections Leapfrog Hospital Survey Report,” 2018. https://www.leapfroggroup.org/sites/default/files/Files/Leapfrog-Castlight%202018%20HAI%20Report.pdf (accessed Aug. 10, 2022).

[55] L. M. Lastinger *et al.*, “Continued increases in the incidence of healthcare-associated infection (HAI) during the second year of the coronavirus disease 2019 (COVID-19) pandemic,” *Infect Control Hosp Epidemiol*, pp. 1–5, May 2022, doi: 10.1017/ice.2022.116.

[56] D. C. Classen *et al.*, “National Trends in the Safety Performance of Electronic Health Record Systems from 2009 to 2018,” *JAMA Network Open*, vol. 3, no. 5. 2020. doi: 10.1001/jamanetworkopen.2020.5547.

[57] J. L. Howe, K. T. Adams, A. Z. Hettinger, and R. M. Ratwani, “Electronic health record usability issues and potential contribution to patient harm,” *JAMA - Journal of the American Medical Association*, vol. 319, no. 12. 2018. doi: 10.1001/jama.2018.1171.

[58] R. M. Ratwani, “Electronic Health Records and Improved Patient Care: Opportunities for Applied Psychology,” *Curr Dir Psychol Sci*, vol. 26, no. 4, 2017, doi: 10.1177/0963721417700691.

[59] 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, doi: 10.1093/qjmed/94.10.521.

[60] J. H. LeLaurin and R. I. Shorr, “Preventing Falls in Hospitalized Patients: State of the Science,” *Clinics in Geriatric Medicine*, vol. 35, no. 2. 2019. doi: 10.1016/j.cger.2019.01.007.

[61] B. J. Braden, J. Maklebust, and J. Maklebust, “Preventing pressure ulcers with the Braden scale: an update on this easy-to-use tool that assesses a patient’s risk.,” *Am J Nurs*, vol. 105, no. 6, 2005.

[62] J. M. Morse, R. M. Morse, and S. J. Tylko, “Development of a Scale to Identify the Fall-Prone Patient,” *Can J Aging*, vol. 8, no. 4, 1989, doi: 10.1017/S0714980800008576.

[63] D. Scarvelis and P. S. Wells, “Diagnosis and treatment of deep-vein thrombosis,” *CMAJ. Canadian Medical Association Journal*, vol. 175, no. 9. 2006. doi: 10.1503/cmaj.060366.

[64] V. Sharma, I. Ali, S. van der Veer, G. Martin, J. Ainsworth, and T. Augustine, “Adoption of clinical risk prediction tools is limited by a lack of integration with electronic health records,” *BMJ Health and Care Informatics*, vol. 28, no. 1. 2021. doi: 10.1136/bmjhci-2020-100253.

[65] A. Choudhury and O. Asan, “Role of artificial intelligence in patient safety outcomes: Systematic literature review,” *JMIR Medical Informatics*, vol. 8, no. 7. 2020. doi: 10.2196/18599.

[66] S. N. Musy *et al.*, “Trigger tool–based automated adverse event detection in electronic health records: systematic review,” *Journal of Medical Internet Research*, vol. 20, no. 5. 2018. doi: 10.2196/JMIR.9901.

[67] G. D. Schiff *et al.*, “Screening for medication errors using an outlier detection system,” *Journal of the American Medical Informatics Association*, vol. 24, no. 2, 2017, doi: 10.1093/jamia/ocw171.

[68] G. Segal, A. Segev, A. Brom, Y. Lifshitz, Y. Wasserstrum, and E. Zimlichman, “Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting,” *Journal of the American Medical Informatics Association*, vol. 26, no. 12, 2019, doi: 10.1093/jamia/ocz135.

[69] J. Corny *et al.*, “A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error,” *Journal of the American Medical Informatics Association*, vol. 27, no. 11, 2020, doi: 10.1093/jamia/ocaa154.

[70] Q. Li *et al.*, “An end-to-end hybrid algorithm for automated medication discrepancy detection,” *BMC Med Inform Decis Mak*, vol. 15, no. 1, 2015, doi: 10.1186/s12911-015-0160-8.

[71] J. Long, M. J. Yuan, and R. Poonawala, “An Observational Study to Evaluate the Usability and Intent to Adopt an Artificial Intelligence–Powered Medication Reconciliation Tool,” *Interact J Med Res*, vol. 5, no. 2, 2016, doi: 10.2196/ijmr.5462.

[72] S. Hasan, G. T. Duncan, D. B. Neill, and R. Padman, “Automatic detection of omissions in medication lists,” *Journal of the American Medical Informatics Association*, vol. 18, no. 4, 2011, doi: 10.1136/amiajnl-2011-000106.

[73] W.-T. M. Au-Yeung, A. K. Sahani, E. M. Isselbacher, and A. A. Armoundas, “Reduction of false alarms in the intensive care unit using an optimized machine learning based approach,” *NPJ Digit Med*, vol. 2, no. 1, 2019, doi: 10.1038/s41746-019-0160-7.

[74] S. Ansari, A. Belle, H. Ghanbari, M. Salamango, and K. Najarian, “Suppression of false arrhythmia alarms in the ICU: A machine learning approach,” *Physiol Meas*, vol. 37, no. 8, 2016, doi: 10.1088/0967-3334/37/8/1186.

[75] L. Chen *et al.*, “Using supervised machine learning to classify real alerts and artifact in online multisignal vital sign monitoring data,” *Crit Care Med*, vol. 44, no. 7, 2016, doi: 10.1097/CCM.0000000000001660.

[76] M. Li, J. Health, D. Ladner, S. Miller, and D. Classen, “Identifying Hospital Patient Safety Problems in Real Time with Electronic Medical Record Data Using an Ensemble Machine Learning Model,” 2018.

[77] H. J. Murff *et al.*, “Automated identification of postoperative complications within an electronic medical record using natural language processing,” *JAMA - Journal of the American Medical Association*, vol. 306, no. 8, 2011, doi: 10.1001/jama.2011.1204.

[78] Y. Wang, E. Coiera, and F. Magrabi, “Using convolutional neural networks to identify patient safety incident reports by type and severity,” *Journal of the American Medical Informatics Association*, vol. 26, no. 12, 2019, doi: 10.1093/jamia/ocz146.

[79] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nature Medicine*, vol. 25, no. 1. 2019. doi: 10.1038/s41591-018-0300-7.

[80] K. G. van Leeuwen, S. Schalekamp, M. J. C. M. Rutten, B. van Ginneken, and M. de Rooij, “Artificial intelligence in radiology: 100 commercially available products and their scientific evidence,” *Eur Radiol*, vol. 31, no. 6, 2021, doi: 10.1007/s00330-021-07892-z.

[81] C. Beeler *et al.*, “Assessing patient risk of central line-associated bacteremia via machine learning,” *Am J Infect Control*, vol. 46, no. 9, 2018, doi: 10.1016/j.ajic.2018.02.021.

[82] W. Ogallo and A. S. Kanter, “Towards a Clinical Decision Support System for Drug Allergy Management: Are Existing Drug Reference Terminologies Sufficient for Identifying Substitutes and Cross-Reactants?,” in *Studies in Health Technology and Informatics*, 2015, vol. 216. doi: 10.3233/978-1-61499-564-7-1088.

[83] M. Zitnik, M. Agrawal, and J. Leskovec, “Modeling polypharmacy side effects with graph convolutional networks,” in *Bioinformatics*, 2018, vol. 34, no. 13. doi: 10.1093/bioinformatics/bty294.

[84] Y. H. Hu, C. T. Tai, C. F. Tsai, and M. W. Huang, “Improvement of Adequate Digoxin Dosage: An Application of Machine Learning Approach,” *J Healthc Eng*, vol. 2018, 2018, doi: 10.1155/2018/3948245.

[85] A. Pavani, S. M. Naushad, R. M. Kumar, M. Srinath, A. R. Malempati, and V. K. Kutala, “Artificial neural network-based pharmacogenomic algorithm for warfarin dose optimization,” *Pharmacogenomics*, vol. 17, no. 2, 2016, doi: 10.2217/pgs.15.161.

[86] Y. H. Hu, F. Wu, C. L. Lo, and C. T. Tai, “Predicting warfarin dosage from clinical data: A supervised learning approach,” *Artif Intell Med*, vol. 56, no. 1, 2012, doi: 10.1016/j.artmed.2012.04.001.

[87] P. Ferroni *et al.*, “Risk Assessment for Venous Thromboembolism in Chemotherapy-Treated Ambulatory Cancer Patients,” *Medical Decision Making*, vol. 37, no. 2, 2017, doi: 10.1177/0272989X16662654.

[88] T. Nafee *et al.*, “Machine learning to predict venous thrombosis in acutely ill medical patients,” *Res Pract Thromb Haemost*, vol. 4, no. 2, 2020, doi: 10.1002/rth2.12292.

[89] R. S. P. Huang *et al.*, “Post-operative bleeding risk stratification in cardiac pulmonary bypass patients using artificial neural network,” *Ann Clin Lab Sci*, vol. 45, no. 2, 2015.

[90] J. Alderden *et al.*, “Predicting pressure injury in critical care patients: A machinelearning model,” *American Journal of Critical Care*, vol. 27, no. 6, 2018, doi: 10.4037/ajcc2018525.

[91] J. Howcroft, J. Kofman, and E. D. Lemaire, “Prospective Fall-Risk Prediction Models for Older Adults Based on Wearable Sensors,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, 2017, doi: 10.1109/TNSRE.2017.2687100.

[92] R. Alazrai, Y. Mowafi, and E. Hamad, “A fall prediction methodology for elderly based on a depth camera,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015, vol. 2015-November. doi: 10.1109/EMBC.2015.7319512.

[93] S. Muralitharan *et al.*, “Machine learning–Based early warning systems for clinical deterioration: Systematic scoping review,” *J Med Internet Res*, vol. 23, no. 2, 2021, doi: 10.2196/25187.

[94] K. D. Mann *et al.*, “Predicting patient deterioration: A review of tools in the digital hospital setting,” *Journal of Medical Internet Research*, vol. 23, no. 9. 2021. doi: 10.2196/28209.

[95] P. L. Yong and L. Olsen, *The healthcare imperative: Lowering costs and improving outcomes*. 2010.

[96] B. C. James, “Making It Easy to Do It Right,” *New England Journal of Medicine*, vol. 345, no. 13, 2001, doi: 10.1056/nejm200109273451311.

[97] C. R. Weir, P. Taber, T. Taft, T. J. Reese, B. Jones, and G. del Fiol, “Feeling and thinking: can theories of human motivation explain how EHR design impacts clinician burnout?,” *J Am Med Inform Assoc*, vol. 28, no. 5, 2021, doi: 10.1093/jamia/ocaa270.

[98] J. F. Sucher, F. A. Moore, S. R. Todd, R. M. Sailors, and B. A. McKinley, “Computerized clinical decision support: A technology to implement and validate evidence based guidelines,” *Journal of Trauma - Injury, Infection and Critical Care*, vol. 64, no. 2. 2008. doi: 10.1097/TA.0b013e3181601812.

[99] R. Geilleit *et al.*, “Feasibility of a real-time hand hygiene notification machine learning system in outpatient clinics,” *Journal of Hospital Infection*, vol. 100, no. 2, 2018, doi: 10.1016/j.jhin.2018.04.004.

[100] M. Meng, M. Sorber, A. Herzog, C. Igel, and C. Kugler, “Technological innovations in infection control: A rapid review of the acceptance of behavior monitoring systems and their contribution to the improvement of hand hygiene,” *American Journal of Infection Control*, vol. 47, no. 4. 2019. doi: 10.1016/j.ajic.2018.10.012.

[101] A. Haque, M. Guo, A. Milstein, and L. Fei-Fei, “Towards Vision-Based Smart Hospitals: A System for Tracking and Monitoring Hand Hygiene Compliance ,” in *Proceedings of Machine Learning for Healthcare*, 2017. Accessed: Aug. 13, 2022. [Online]. Available: https://www.semanticscholar.org/paper/Towards-Vision-Based-Smart-Hospitals%3A-A-System-for-Haque-Guo/6bd36ced896c8910a5b636cc569e25856f7305ec

[102] I. Banerjee *et al.*, “Development and Performance of the Pulmonary Embolism Result Forecast Model (PERFORM) for Computed Tomography Clinical Decision Support,” *JAMA Netw Open*, vol. 2, no. 8, 2019, doi: 10.1001/jamanetworkopen.2019.8719.

[103] J. Willan, H. Katz, and D. Keeling, “The use of artificial neural network analysis can improve the risk-stratification of patients presenting with suspected deep vein thrombosis,” *Br J Haematol*, vol. 185, no. 2, 2019, doi: 10.1111/bjh.15780.

[104] M. Nagendran *et al.*, “Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies in medical imaging,” *The BMJ*, vol. 368, 2020, doi: 10.1136/bmj.m689.

[105] M. O. Kim, E. Coiera, and F. Magrabi, “Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review,” *Journal of the American Medical Informatics Association : JAMIA*, vol. 24, no. 2. 2017. doi: 10.1093/jamia/ocw154.

[106] C. L. Andaur Navarro *et al.*, “Risk of bias in studies on prediction models developed using supervised machine learning techniques: Systematic review,” *The BMJ*, vol. 375. 2021. doi: 10.1136/bmj.n2281.

[107] R. D. Riley *et al.*, “External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges,” *The BMJ*, vol. 353, 2016, doi: 10.1136/bmj.i3140.

[108] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, “Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement,” *Eur Urol*, vol. 67, no. 6, 2015, doi: 10.1016/j.eururo.2014.11.025.

[109] S. Vollmer *et al.*, “Machine learning and AI research for patient benefit: 20 critical questions on transparency, replicability, ethics and effectiveness,” *ArXiv*, 2018.

[110] G. S. Collins and K. G. M. Moons, “Reporting of artificial intelligence prediction models,” *The Lancet*, vol. 393, no. 10181. 2019. doi: 10.1016/S0140-6736(19)30037-6.

[111] G. S. Collins *et al.*, “External validation of multivariable prediction models: A systematic review of methodological conduct and reporting,” *BMC Medical Research Methodology*, vol. 14, no. 1. 2014. doi: 10.1186/1471-2288-14-40.

[112] A. Subbaswamy, P. Schulam, and S. Saria, “Preventing failures due to dataset shift: Learning predictive models that transport,” 2020.

[113] P. Schulam and S. Saria, “Can you trust this prediction? Auditing pointwise reliability after learning,” 2020.

[114] E. J. Emanuel and R. M. Wachter, “Artificial Intelligence in Health Care: Will the Value Match the Hype?,” *JAMA - Journal of the American Medical Association*, vol. 321, no. 23. 2019. doi: 10.1001/jama.2019.4914.

[115] S. van Baalen, M. Boon, and P. Verhoef, “From clinical decision support to clinical reasoning support systems,” in *Journal of Evaluation in Clinical Practice*, 2021, vol. 27, no. 3. doi: 10.1111/jep.13541.

[116] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, “An overview of clinical decision support systems: benefits, risks, and strategies for success,” *npj Digital Medicine*. 2020. doi: 10.1038/s41746-020-0221-y.

[117] S. Khairat, D. Marc, W. Crosby, and A. al Sanousi, “Reasons for physicians not adopting clinical decision support systems: Critical analysis,” *JMIR Medical Informatics*, vol. 20, no. 4. 2018. doi: 10.2196/medinform.8912.

[118] N. Wagneur, P. Callier, J.-D. Zeitoun, D. Silber, R. Sabatier, and F. Denis, “Assessing a New Prescreening Score for the Simplified Evaluation of the Clinical Quality and Relevance of eHealth Apps: Instrument Validation Study,” *J Med Internet Res*, vol. 24, no. 7, Jul. 2022, doi: 10.2196/39590.

[119] A. H. sen Fang, W. T. Lim, and T. Balakrishnan, “Early warning score validation methodologies and performance metrics: a systematic review,” *BMC Med Inform Decis Mak*, vol. 20, no. 1, 2020, doi: 10.1186/s12911-020-01144-8.

[120] S. Gerry *et al.*, “Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology,” *BMJ*, 2020, doi: 10.1136/bmj.m1501.

[121] V. J. Major, N. Jethani, and Y. Aphinyanaphongs, “Estimating real-world performance of a predictive model: A case-study in predicting mortality,” *JAMIA Open*, vol. 3, no. 2, 2020, doi: 10.1093/jamiaopen/ooaa008.

[122] B. van Calster *et al.*, “Calibration: The Achilles heel of predictive analytics,” *BMC Med*, vol. 17, no. 1, 2019, doi: 10.1186/s12916-019-1466-7.

[123] L. Vonrueden *et al.*, “Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems,” *IEEE Trans Knowl Data Eng*, 2021, doi: 10.1109/TKDE.2021.3079836.

[124] M. Saikali, A. Tanios, and A. Saab, “Evaluation of a Broad-Spectrum Partially Automated Adverse Event Surveillance System,” *J Patient Saf*, p. 1, Nov. 2017, doi: 10.1097/PTS.0000000000000442.

[125] “Florida Hospital Global Trigger Tool Implementation Toolkit,” 2017. http://www.safetyleaders.org/pages/idPage.jsp?ID=4879

[126] “The Global Trigger Tool A Practical Implementation Guide for New Zealand District Health Boards,” 2012.

[127] NCC MERP, “National Coordinating Council for Medication Error Reporting and Prevention,” *Taxonomy of Medication Errors*, p. 19, 2001.

[128] (Who) World Health Organization, “Report on the Burden of Endemic Health Care-Associated Infection Worldwide.,” *WHO Library Cataloguing-in-Publication Data*, 2011.

[129] G. de Lissovoy, K. Fraeman, V. Hutchins, D. Murphy, D. Song, and B. B. Vaughn, “Surgical site infection: Incidence and impact on hospital utilization and treatment costs,” *Am J Infect Control*, vol. 37, no. 5, 2009, doi: 10.1016/j.ajic.2008.12.010.

[130] M. Zegers *et al.*, “The incidence, root-causes, and outcomes of adverse events in surgical units: Implication for potential prevention strategies,” *Patient Saf Surg*, vol. 5, no. 1, 2011, doi: 10.1186/1754-9493-5-13.

[131] S. S. Magill *et al.*, “Multistate Point-Prevalence Survey of Health Care–Associated Infections,” *New England Journal of Medicine*, vol. 370, no. 13, 2014, doi: 10.1056/nejmoa1306801.

[132] K. W. Poh, C. H. Ngan, J. Y. Wong, T. K. Ng, and N. Mohd Noor, “Reduction of central-line-associated bloodstream infection (CLABSI) in resource limited, nonintensive care unit (ICU) settings,” *Int J Health Care Qual Assur*, vol. 33, no. 2, 2020, doi: 10.1108/IJHCQA-11-2019-0195.

[133] V. D. Rosenthal *et al.*, “International Nosocomial Infection Control Consortium report, data summary of 50 countries for 2010-2015: Device-associated module,” *Am J Infect Control*, vol. 44, no. 12, 2016, doi: 10.1016/j.ajic.2016.08.007.

[134] M. Talaat *et al.*, “National surveillance of health care–associated infections in Egypt: Developing a sustainable program in a resource-limited country,” *Am J Infect Control*, vol. 44, no. 11, 2016, doi: 10.1016/j.ajic.2016.04.212.

[135] A. Daud and F. Mohamad, “Patient characteristics related to phlebitis in the east coast of peninsular Malaysia hospital,” *Jurnal Keperawatan Indonesia*, vol. 24, no. 1, 2021, doi: 10.7454/jki.v24i1.1097.

[136] V. Tagalakis, S. R. Kahn, M. Libman, and M. Blostein, “The epidemiology of peripheral vein infusion thrombophlebitis: A critical review,” *American Journal of Medicine*, vol. 113, no. 2. 2002. doi: 10.1016/S0002-9343(02)01163-4.

[137] A. Salgueiro-Oliveira, P. Parreira, and P. Veiga, “Incidence of phlebitis in patients with peripheral intravenous catheters: The influence of some risk factors,” *Australian Journal of Advanced Nursing*, vol. 30, no. 2, 2012.

[138] D. Berlowitz, “Incidence and Prevalence of Pressure Ulcers,” in *Pressure Ulcers in the Aging Population*, 2014. doi: 10.1007/978-1-62703-700-6_2.

[139] J. C. Gardiner, P. L. Reed, J. D. Bonner, D. K. Haggerty, and D. G. Hale, “Incidence of hospital-acquired pressure ulcers - a population-based cohort study,” *Int Wound J*, vol. 13, no. 5, 2016, doi: 10.1111/iwj.12386.

[140] J. M. Januel *et al.*, “Symptomatic in-hospital deep vein thrombosis and pulmonary embolism following hip and knee arthroplasty among patients receiving recommended prophylaxis: A systematic review,” *JAMA - Journal of the American Medical Association*, vol. 307, no. 3. 2012. doi: 10.1001/jama.2011.2029.

[141] K. Gunasekaran *et al.*, “A review of the incidence diagnosis and treatment of spontaneous hemorrhage in patients treated with direct oral anticoagulants,” *Journal of Clinical Medicine*, vol. 9, no. 9. 2020. doi: 10.3390/jcm9092984.

[142] A. Wonnacott, S. Meran, B. Amphlett, B. Talabani, and A. Phillips, “Epidemiology and outcomes in community-acquired versus hospital-acquired aki,” *Clinical Journal of the American Society of Nephrology*, vol. 9, no. 6, 2014, doi: 10.2215/CJN.07920713.

[143] A. Maskey, R. C. Kafle, and S. Lamsal, “Risk factors and incidence of contrast-induced acute kidney injury associated with diagnostic or interventional coronary angiography,” *Journal of Advances in Internal Medicine*, vol. 9, no. 1, 2020, doi: 10.3126/jaim.v9i1.29162.

[144] S. Morabito *et al.*, “Incidence of contrast-induced acute kidney injury associated with diagnostic or interventional coronary angiography,” *J Nephrol*, vol. 25, no. 6, 2012, doi: 10.5301/jn.5000101.

[145] S. M. Vindigni and C. M. Surawicz, “C. Difficile infection: Changing epidemiology and management paradigms,” *Clinical and Translational Gastroenterology*, vol. 6, no. 7. 2015. doi: 10.1038/ctg.2015.24.

[146] M. K. Szekendi *et al.*, “Active surveillance using electronic triggers to detect adverse events in hospitalized patients.,” *Qual Saf Health Care*, vol. 15, no. 3, pp. 184–90, 2006, doi: 10.1136/qshc.2005.014589.

[147] L. S.J., B. K.F., A. D., and V. C., “What is known about adverse events in older medical hospital inpatients? A systematic review of the literature,” *International Journal for Quality in Health Care*, vol. 25, no. 5, pp. 542–554, 2013.

[148] A. B. A. Sari, A. Cracknell, and T. A. Sheldon, “Incidence, preventability and consequences of adverse events in older people: Results of a retrospective case-note review,” *Age and Ageing*, vol. 37, no. 3. pp. 265–269, 2008. doi: 10.1093/ageing/afn043.

[149] U. Nwulu, K. Nirantharakumar, R. Odesanya, S. E. McDowell, and J. J. Coleman, “Improvement in the detection of adverse drug events by the use of electronic health and prescription records: An evaluation of two trigger tools,” *Eur J Clin Pharmacol*, vol. 69, no. 2, 2013, doi: 10.1007/s00228-012-1327-1.

[150] J. T. Patregnani, M. C. Spaeder, V. Lemon, Y. Diab, D. Klugman, and D. C. Stockwell, "Monitoring the harm associated with use of anticoagulants in pediatric populations through trigger-based automated adverse-event detection," *Jt Comm J Qual Patient Saf*, vol. 41, no. 3, 2015, doi: 10.1016/S1553-7250(15)41015-3.

[151] CDC - Center for Disease Control and Prevention, "Identifying Healthcare-associated Infections (HAI) for NHSN Surveillance," *http://www.cdc.gov/nhsn/PDFs/pscManual/2PSC_IdentifyingHAIs_NHSNcurrent.pdf*, 2015.

[152] C. Moore, J. Li, C. C. Hung, J. Downs, and J. R. Nebeker, "Predictive value of alert triggers for identification of developing adverse drug events," *J Patient Saf*, vol. 5, no. 4, 2009, doi: 10.1097/PTS.0b013e3181bc05e5.

[153] I. Kose *et al.*, "Adoption rates of electronic health records in Turkish Hospitals and the relation with hospital sizes," *BMC Health Serv Res*, vol. 20, no. 1, 2020, doi: 10.1186/s12913-020-05767-5.

[154] A. Saab, M. Saikali, and J. B. Lamy, "Comparison of machine learning algorithms for classifying adverse-event related 30- A y hospital readmissions: Potential implications for patient safety," in *Studies in Health Technology and Informatics*, 2020, vol. 272. doi: 10.3233/SHTI200491.

[155] HL7, "HL7 FHIR, Release 4," *HL7 FHIR*, 2019.

[156] S. F. Jencks, M. v. Williams, and E. A. Coleman, "Rehospitalizations among Patients in the Medicare Fee-for-Service Program," *New England Journal of Medicine*, vol. 360, no. 14, 2009, doi: 10.1056/nejmsa0803563.

[157] J. Kjellberg *et al.*, "Costs associated with adverse events among acute patients," *BMC Health Serv Res*, vol. 17, no. 1, 2017, doi: 10.1186/s12913-017-2605-5.

[158] A. Artetxe, A. Beristain, and M. Graña, "Predictive models for hospital readmission risk: A systematic review of methods," *Computer Methods and Programs in Biomedicine*, vol. 164. 2018. doi: 10.1016/j.cmpb.2018.06.006.

[159] D. Kansagara *et al.*, "Risk prediction models for hospital readmission: A systematic review," *JAMA - Journal of the American Medical Association*, vol. 306, no. 15. 2011. doi: 10.1001/jama.2011.1515.

[160] A. Garcia-Arce, F. Rico, and J. L. Zayas-Castro, "Comparison of Machine Learning Algorithms for the Prediction of Preventable Hospital Readmissions," *Journal for Healthcare Quality*, vol. 40, no. 3, 2018, doi: 10.1097/JHQ.0000000000000080.

[161] M. Saikali, A. Tanios, and A. Saab, "Evaluation of a Broad-Spectrum Partially Automated Adverse Event Surveillance System," *J Patient Saf*, 2017.

[162] L. Turgeman and J. H. May, "A mixed-ensemble model for hospital readmission," *Artif Intell Med*, vol. 72, 2016, doi: 10.1016/j.artmed.2016.08.005.

[163] L. T. Q. Cardoso *et al.*, "Impact of delayed admission to intensive care units on mortality of critically ill patients: A cohort study," *Crit Care*, vol. 15, no. 1, p. R28, Jan. 2011, doi: 10.1186/cc9975.

[164] C. B. Sankey, G. McAvay, J. M. Siner, C. L. Barsky, and S. I. Chaudhry, "'Deterioration to Door Time': An Exploratory Analysis of Delays in Escalation of Care for Hospitalized Patients," *J Gen Intern Med*, vol. 31, no. 8, pp. 895–900, Aug. 2016, doi: 10.1007/s11606-016-3654-x.

[165] M. M. Churpek, B. Wendlandt, F. J. Zadravecz, R. Adhikari, C. Winslow, and D. P. Edelson, "Association between intensive care unit transfer delay and hospital mortality: A multicenter investigation," *J Hosp Med*, vol. 11, no. 11, pp. 757–762, Nov. 2016, doi: 10.1002/jhm.2630.

[166] H. Brown, J. Terrence, P. Vasquez, D. W. Bates, and E. Zimlichman, "Continuous monitoring in an inpatient medical-surgical unit: A controlled clinical trial," *American Journal of Medicine*, vol. 127, no. 3, pp. 226–232, Mar. 2014, doi: 10.1016/j.amjmed.2013.12.004.

[167] J. N. Blackwell *et al.*, "Early Detection of In-Patient Deterioration," *Crit Care Explor*, vol. 2, no. 5, p. e0116, May 2020, doi: 10.1097/cce.0000000000000116.

[168] M. A. Rose, L. A. Hanna, S. A. Nur, and C. M. Johnson, "Utilization of electronic modified early warning score to engage rapid response team early in clinical deterioration," *J Nurses Prof Dev*, vol. 31, no. 3, pp. E1–E7, May 2015, doi: 10.1097/NND.0000000000000157.

[169] T. C. for C. P. at NICE, *Acutely ill patients in hospital: Recognition of and response to acute illness in adults in hospital*. 2007.

[170] K. K. Hall, A. Lim, and B. Gale, "The Use of Rapid Response Teams to Reduce Failure to Rescue Events: A Systematic Review," *J Patient Saf*, vol. 16, no. 3S Suppl 1, 2020, doi: 10.1097/PTS.0000000000000748.

[171] "Electronic Medical Record Adoption Model | HIMSS Analytics - North America." https://www.himssanalytics.org/emram (accessed Jun. 30, 2021).

[172] I. Kose *et al.*, "Adoption rates of electronic health records in Turkish Hospitals and the relation with hospital sizes," *BMC Health Serv Res*, vol. 20, no. 1, 2020, doi: 10.1186/s12913-020-05767-5.

[173] P. Kipnis *et al.*, "Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU," *J Biomed Inform*, vol. 64, pp. 10–19, Dec. 2016, doi: 10.1016/j.jbi.2016.09.013.

[174] G. J. Escobar *et al.*, "Piloting electronic medical record–based early detection of inpatient deterioration in community hospitals," *J Hosp Med*, vol. 11, no. 1, pp. S18–S24, Nov. 2016, doi: 10.1002/jhm.2652.

[175] T. Kamio, T. Van, and K. Masamune, "Use of Machine-Learning Approaches to Predict Clinical Deterioration in Critically Ill Patients: A Systematic Review," *International Journal of Medical Research & Health Sciences*, 2017.

[176] A. D. Jeffery *et al.*, "Advancing in-hospital clinical deterioration prediction models," *American Journal of Critical Care*, 2018, doi: 10.4037/ajcc2018957.

[177] M. M. Churpek, T. C. Yuen, C. Winslow, D. O. Meltzer, M. W. Kattan, and D. P. Edelson, "Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards," 2016. doi: 10.1097/CCM.0000000000001571.

[178] A. J. Atkinson *et al.*, “Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework,” *Clinical Pharmacology and Therapeutics*. 2001. doi: 10.1067/mcp.2001.113989.

[179] A. Orfanu, V. Aramă, C. Popescu, C. Tilişcan, A. Streinu-Cercel, and Ş. S. Aramă, “The Dynamical Assessment of Inflammatory Biomarkers in Predicting the Outcome of Septic Patients and the Response to Antimicrobial Therapy,” *The Journal of Critical Care Medicine*, 2020, doi: 10.2478/jccm-2020-0004.

[180] P. A. Kavsak, S. A. Hill, W. B. Supapol, P. J. Devereaux, and A. Worster, “Biomarkers for predicting serious cardiac outcomes at 72 hours in patients presenting early after chest pain onset with symptoms of acute coronary syndromes,” *Clin Chem*, 2012, doi: 10.1373/clinchem.2011.172064.

[181] L. L. Kirkland *et al.*, “A Clinical Deterioration Prediction Tool for Internal Medicine Patients,” *American Journal of Medical Quality*, vol. 28, no. 2, 2013, doi: 10.1177/1062860612450459.

[182] R. v Peelen, Y. Eddahchouri, M. Koeneman, and H. van Goor, “Algorithms for Prediction of Clinical Deterioration on the General Wards :,” no. June, pp. 1–8, 2021, doi: 10.12788/jhm.3630.

[183] J. Tirkkonen, T. Tamminen, and M. B. Skrifvars, “Outcome of adult patients attended by rapid response teams: A systematic review of the literature,” *Resuscitation*, vol. 112. 2017. doi: 10.1016/j.resuscitation.2016.12.023.

[184] E. M. Sørensen and J. A. Petersen, “Performance of the efferent limb of a rapid response system: An observational study of medical emergency team calls,” *Scand J Trauma Resusc Emerg Med*, vol. 23, no. 1, 2015, doi: 10.1186/s13049-015-0153-8.

[185] D. Jones, I. Mitchell, K. Hillman, and D. Story, “Defining clinical deterioration,” *Resuscitation*, vol. 84, no. 8, pp. 1029–1034, Aug. 2013, doi: 10.1016/j.resuscitation.2013.01.013.

[186] R. M. Padilla and A. M. Mayo, “Clinical deterioration: A concept analysis,” *J Clin Nurs*, vol. 27, no. 7–8, pp. 1360–1368, Apr. 2018, doi: 10.1111/jocn.14238.

[187] A. Fernández, S. García, F. Herrera, and N. V. Chawla, “SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary,” *Journal of Artificial Intelligence Research*. 2018. doi: 10.1613/jair.1.11192.

[188] J. N. Blackwell *et al.*, “Critical Care Explorations Critical Care Explorations Early Detection of In-Patient Deterioration: One Prediction Model Does Not Fit All,” 2020, doi: 10.1097/CCE.0000000000000116.

[189] L. M. Fleuren *et al.*, “Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy,” *Intensive Care Medicine*. 2020. doi: 10.1007/s00134-019-05872-y.

[190] D. Zeiberg, T. Prahlad, B. K. Nallamothu, T. J. Iwashyna, J. Wiens, and M. W. Sjoding, “Machine learning for patient risk stratification for acute respiratory distress syndrome,” *PLoS One*, vol. 14, no. 3, Mar. 2019, doi: 10.1371/journal.pone.0214465.

[191] P. Yang *et al.*, "A new method for identifying the acute respiratory distress syndrome disease based on noninvasive physiological parameters," *PLoS One*, vol. 15, no. 2, p. e0226962, Feb. 2020, doi: 10.1371/journal.pone.0226962.

[192] S. Le *et al.*, "Supervised Machine Learning for the Early Prediction of Acute Respiratory Distress Syndrome (ARDS)," *medRxiv*, p. 2020.03.19.20038364, Mar. 2020, doi: 10.1101/2020.03.19.20038364.

[193] X. F. Ding *et al.*, "Predictive model for acute respiratory distress syndrome events in ICU patients in China using machine learning algorithms: A secondary analysis of a cohort study," *J Transl Med*, vol. 17, no. 1, p. 326, Oct. 2019, doi: 10.1186/s12967-019-2075-0.

[194] W. Xiong, M. Xu, Y. Zhao, X. Wu, B. Pudasaini, and J. M. Liu, "Can we predict the prognosis of COPD with a routine blood test?," *International Journal of COPD*, 2017, doi: 10.2147/COPD.S124041.

[195] S. L. Hyland *et al.*, "Early prediction of circulatory failure in the intensive care unit using machine learning," *Nat Med*, 2020, doi: 10.1038/s41591-020-0789-4.

[196] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman, "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU," *Crit Care Med*, 2018, doi: 10.1097/CCM.0000000000002936.

[197] T. Desautels *et al.*, "Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach," *JMIR Med Inform*, 2016, doi: 10.2196/medinform.5909.

[198] Y. P. Tabak, X. Sun, C. M. Nunez, and R. S. Johannes, "Using electronic health record data to develop inpatient mortality predictive model: Acute Laboratory Risk of Mortality Score (ALaRMS)," *Journal of the American Medical Informatics Association*, vol. 21, no. 3, pp. 455–463, 2014, doi: 10.1136/amiajnl-2013-001790.

[199] G. J. Escobar, J. D. Greene, P. Scheirer, M. N. Gardner, D. Draper, and P. Kipnis, "Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases," *Med Care*, 2008, doi: 10.1097/MLR.0b013e3181589bb6.

[200] M. Karakioulaki and D. Stolz, "Biomarkers in pneumonia-beyond procalcitonin," *International Journal of Molecular Sciences*. 2019. doi: 10.3390/ijms20082004.

[201] C. S. Gori, L. Magrini, F. Travaglino, and S. di Somma, "Role of biomarkers in patients with dyspnea," *European Review for Medical and Pharmacological Sciences*. 2011.

[202] ACSQH, "National consensus statement: essential elements for recognising and responding to acute physiological deterioration (second edition)," 2017.

[203] A. Ambasta, S. Pancic, B. M. Wong, T. Lee, D. McCaughey, and I. W. Y. Ma, "Expert Recommendations on Frequency of Utilization of Common Laboratory Tests in Medical Inpatients: a Canadian Consensus Study," *J Gen Intern Med*, 2019, doi: 10.1007/s11606-019-05196-z.

[204] A. Orfanu, V. Aramă, C. Popescu, C. Tilişcan, A. Streinu-Cercel, and Ş. S. Aramă, "The Dynamical Assessment of Inflammatory Biomarkers in Predicting the Outcome of Septic

Patients and the Response to Antimicrobial Therapy," *The Journal of Critical Care Medicine*, 2020, doi: 10.2478/jccm-2020-0004.

[205] P. A. Kavsak, S. A. Hill, W. B. Supapol, P. J. Devereaux, and A. Worster, "Biomarkers for predicting serious cardiac outcomes at 72 hours in patients presenting early after chest pain onset with symptoms of acute coronary syndromes," *Clin Chem*, 2012, doi: 10.1373/clinchem.2011.172064.

[206] V. Campbell *et al.*, "Predicting clinical deterioration with Q-ADDS compared to NEWS, Between the Flags, and eCART track and trigger tools," *Resuscitation*, 2020, doi: 10.1016/j.resuscitation.2020.05.027.

[207] A. Kia *et al.*, "MEWS++: Enhancing the Prediction of Clinical Deterioration in Admitted Patients through a Machine Learning Model," *J Clin Med*, vol. 9, no. 2, p. 343, Jan. 2020, doi: 10.3390/jcm9020343.

[208] K.-J. Cho *et al.*, "Detecting Patient Deterioration Using Artificial Intelligence in a Rapid Response System," *Crit Care Med*, vol. 48, no. 4, pp. e285–e289, Apr. 2020, doi: 10.1097/CCM.0000000000004236.

[209] L. H. Fu *et al.*, "Development and validation of early warning score system: A systematic literature review," *Journal of Biomedical Informatics*, vol. 105. 2020. doi: 10.1016/j.jbi.2020.103410.

[210] A. D. Jeffery *et al.*, "Advancing in-hospital clinical deterioration prediction models," *American Journal of Critical Care*, 2018, doi: 10.4037/ajcc2018957.

[211] M. M. Churpek, T. C. Yuen, C. Winslow, D. O. Meltzer, M. W. Kattan, and D. P. Edelson, "Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards," 2016. doi: 10.1097/CCM.0000000000001571.

[212] M. A. Pimentel *et al.*, "Detecting Deteriorating Patients in Hospital: Development and Validation of a Novel Scoring System," *Am J Respir Crit Care Med*, 2021, doi: 10.1164/rccm.202007-2700oc.

[213] R. Blythe, R. Parsons, N. M. White, D. Cook, and S. McPhail, "A scoping review of real-time automated clinical deterioration alerts and evidence of impacts on hospitalised patient outcomes," *BMJ Qual Saf*, vol. 31, no. 10, pp. 725–734, Oct. 2022, doi: 10.1136/BMJQS-2021-014527.

[214] A. Saab, C. Abi Khalil, M. Jammal, M. Saikali, and J.-B. Lamy, "Early Prediction of All-Cause Clinical Deterioration in General Wards Patients: Development and Validation of a Biomarker-Based Machine Learning Model Derived From Rapid Response Team Activations," *J Patient Saf*, vol. 18, no. 6, pp. 578–586, Sep. 2022, doi: 10.1097/PTS.0000000000001069.

[215] "SUS: A 'Quick and Dirty' Usability Scale," in *Usability Evaluation In Industry*, 2020. doi: 10.1201/9781498710411-35.

[216] M. Grinberg, "Flask Web Development: Developing Web Applications with Python - Miguel Grinberg - Google Books," *Google Books*. 2018.