國立臺灣大學電機資訊學院生醫電子與資訊學研究所

博士論文

Graduate Institute of Biomedical Electronics and Bioinformatics
College of Electrical Engineering and Computer Science
National Taiwan University

Doctoral Dissertation

以機器學習模型偵測一般病房住院病人惡化

Inpatient deterioration detection in general wards using machine learning model

蘇彰甫
Chang-Fu Su

指導教授：賴飛羆 博士

Advisor: Feipei Lai, Ph.D.

中華民國 112 年 7 月

July 2023

doi:10.6342/NTU202302444

國立臺灣大學博士學位論文
口試委員會審定書

PhD DISSERTATION ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

（論文中文題目）(Chinese title of PhD dissertation)

以機器學習模型偵測一般病房住院病人惡化

（論文英文題目）(English title of PhD dissertation)

Inpatient deterioration detection in general wards using machine learning model

本論文係 蘇彰甫 (姓名) D06945004 （學號）在國立臺灣大學 生醫電子與資訊學研究所 (系/所/學位學程)完成之博士學位論文，於民國 112 年 6 月 30 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Department / Institute of Graduate Institute of Biomedical Electronics and Bioinformatics on 30 (date) 06 (month) 2023 (year) have examined a PhD dissertation entitled above presented by Chang-Fu Su (name) D069459004 (student ID) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

賴飛羆 （指導教授 Advisor）	傅楸善	陳縕儂
陳信希	汪大暉	吳耀銘
黃樹林	孫維仁	[illegible]

系主任/所長 Director: 林致廷

誌謝

在我攻讀博士學位的過程中，最令我深刻感激的人，莫過於我的指導教授賴飛羆老師，他給予我悉心的指導與愛護。由於我是醫學背景跨足資訊領域，而且是在職生，求學之路一直充滿挑戰。賴老師除了在學業上給我全方面的指導外，更不遺餘力地提供資源來協助我的研究；在他的悉心指導下，我順利完成了研究著作和學位論文，我將永遠珍惜這份師恩。

身為一個在職學生，得到工作上長官的支持是極為重要的。我要感謝吳志雄執行長暨院長，他不斷鼓勵我在研究及學業上持續努力，並提必要的資源與支持。同時，我也要感謝張智星教授、邱淑怡教授在研究上的指導與協助，為我的研究打下了良好的基礎。另外，我要感謝傅楸善教授、趙坤茂教授、陳縕儂教授、孫維仁教授、陳晉興教授、吳耀銘教授、汪大暉教授和黃樹林教授，在論文口試時提供的寶貴建議。

此外，我要感謝醫品室的夥伴瓊儀、健銘、憶敏、志誠、培惠、瑄潞、郁婷、湘齡，以及資訊室騰越課長，他們在工作上及研究上給予我非常有價值的幫助。在攻讀博士學位的過程中，我要感謝原嘉、峰祺、佳東、富舜、裕仁、彥斌、沛甫、昌杰、哲瑋、昌恆及婉瑄等研究室夥伴們，我們互相鼓勵、一起學習，總是能碰撞出新的思維火花。

我要特別感謝我的太太貞儀，在整個求學過程中給予我無微不至的陪伴和支持。除了自己的工作外，她還兢兢業業地照顧孩子們的教育和成長，在學業與工作兩頭燒的時候，太太的支持是最可貴的。我也感謝我們的父母和家人給予我無盡的愛與自由發展的空間，正是因為他們的支持，我才能無後顧之憂地前進。

最後，我要感謝我生命中的每一位貴人，你們的幫助和支持讓我受益終生。我會永遠銘記在心，感謝你們。

蘇彰甫　謹誌　2023 年 7 月

doi:10.6342/NTU202302444

中文摘要

住院期間的心跳驟停（In-hospital cardiac arrest, IHCA）是嚴重的事件，常伴隨高死亡率，各大研究亦強調了早期識別和早期介入對於改善患者預後的重要性。部份的心跳驟停是突然地發生，沒有伴隨明顯徵兆，因此開發自動化的預測模型以識別高風險患者並及時進行介入是非常重要的。本研究引入了兩個創新的預測模型：『時間序列早期預警分數（Time-Series Early Warning Score, TEWS）』和『可解釋的時間序列早期預警分數（Explainable Time-Series Early Warning Score, TEWS-X）』。這兩個模型只使用常規量測的生命徵象資料來提供較為準確且可解釋的 IHCA 預測，使醫療提供者能夠採取主動措施，提高患者安全性。

TEWS 模型通過結合多個時間窗口的特徵，再加上類神經網路對於特徵趨勢和模式的處理能力，實現了更高的預測準確性。此外，TEWS-X 模型通過採用基於決策樹的機器學習方法和 SHAP 值，對醫療照顧者解釋其預測結果，使醫療照顧者可依此結果作出臨床決策。這些模型可以無縫地集成到現有的照護流程中，無需中斷工作流程，進而提升病人安全並優化資源分配。

關鍵字：住院病人心跳驟停、早期警訊系統、生命徵象、機器學習、可解釋人工智慧

doi:10.6342/NTU202302444

Abstract

In-hospital cardiac arrest (IHCA) is a critical event associated with high mortality rates. Early identification and intervention are crucial for improving patient outcomes. This study introduces two innovative predictive models: the Time-Series Early Warning Score (TEWS) and the Explainable Time-Series Early Warning Score (TEWS-X), designed to leverage vital signs data and provide accurate and explainable predictions of IHCA.

The TEWS model utilizes vital signs data from six time windows (48 hours) to predict IHCA occurrences and performs superior IHCA prediction performance compared to alternative classification algorithms. Incorporating features from multiple time windows significantly improves prediction accuracy, with an area under the receiver operating characteristic curve (AUROC) of 0.808, surpassing the performance of MEWS (AUROC of MEWS: 0.649).

The TEWS-X model incorporates a tree-based machine learning approach and SHAP values to enhance model explainability, enabling insights into feature importance and supporting transparent decision-making, facilitating an understanding of the critical factors influencing IHCA risk. These models can seamlessly integrate into existing care processes, improving patient safety without disrupting workflow.

doi:10.6342/NTU202302444

The TEWS and TEWS-X models represent significant advancements in IHCA prediction and explainability. By leveraging vital signs data and incorporating explainable modeling techniques, these models empower healthcare providers to identify patients at risk of IHCA and intervene promptly and proactively. Further research is needed to validate the models in diverse healthcare settings and explore additional data sources for enhanced predictive capabilities. Implementing the TEWS and TEWS-X models can improve patient outcomes and optimize resource allocation in the management of IHCA.

Keyword: IHCA, Early Warning Score, Vital sign, Machine Learning, Explainable AI

doi:10.6342/NTU202302444

CONTENTS

doi:10.6342/NTU202302444

doi:10.6342/NTU202302444

doi:10.6342/NTU202302444

doi:10.6342/NTU202302444

LIST OF FIGURES

doi:10.6342/NTU202302444

doi:10.6342/NTU202302444

LIST OF TABLES

doi:10.6342/NTU202302444

Chapter 1 Introduction

1.1 Motivation

In-hospital cardiac arrest (IHCA) poses a substantial risk to patient safety, despite its infrequency, and carries a high mortality rate. The Utstein resuscitation registry reporting template defines IHCA as providing chest compressions or defibrillation to patients in inpatient beds [1].

Annually in the United States, numerous cardiac arrests are reported among hospitalized patients, with an estimated incidence of approximately 0.92 per 1,000 bed days [2]. Unfortunately, outcomes following cardiac arrest are notably poor, reflected by post-discharge survival rates close to 25% in the United States and under 20% globally [3, 4]. A meta-analysis encompassing 40 studies reported a 1-year survival rate post-IHCA of 13.4% [5]. Moreover, within this meta-analysis, 17.6% of patients survived hospital discharge, implying that approximately 76% of patients who survived their hospital stay live for at least a year [5].

A preliminary analysis of 23 cohort studies has identified several factors associated with diminished survival odds post-in-hospital cardiac arrest. These include male gender, age 60 or above, presence of active malignancy, and history of chronic kidney disease [6]. In contrast, some factors are significantly linked to increased survival odds, such as

doi:10.6342/NTU202302444

witnessed arrest, monitored arrest, the daytime occurrence of the arrest, and an initial shockable rhythm. However, intubation during arrest and a resuscitation duration of at least 15 minutes are associated with reduced survival odds. Understanding the impact of intra-arrest factors on patient outcomes underscores the critical need to identify high-risk individuals. Improved survival rates largely depend on the preparedness and vigilance of the healthcare team.

Various early warning scoring systems have been established to identify hospitalized patients at high risk of clinical deterioration, meeting the urgent need for early recognition of such patients. These scoring systems typically incorporate relevant variables associated with predictive outcomes. The Modified Early Warning Score (MEWS) [7], which includes vital signs like temperature, heart rate, respiratory rate, and blood pressure, is a common approach to informing clinical decision-making. However, the area under the receiver operating characteristic curve (AUROC) for MEWS consistently falls below 0.7 in multiple studies, suggesting the need to explore the inclusion of additional clinical data such as laboratory results, demographics, and heart rate variability to augment predictive performance [8-13]. These efforts aim to enhance the precision and effectiveness of early warning scoring systems in identifying patients at elevated risk of deterioration and facilitating timely interventions to improve patient outcomes. Despite their promising

doi:10.6342/NTU202302444

results in accuracy, reduction in false alarms, and favorable detection rates, their applicability might be restricted in units where routine measurement of such clinical data is not the norm.

The advent of artificial intelligence (AI) and machine learning (ML) systems heralds a new chapter in biomedicine, transforming aspects ranging from molecular research to disease investigation. With their ability to autonomously analyze complex datasets, ML offers researchers the ability to extract valuable insights and uncover challenging patterns. Employing ML models could significantly enhance the accuracy of predictions either by utilizing existing data or optimizing features for optimal performance. By tapping into these advanced computational approaches, the biomedical community can open new knowledge pathways and advance disease understanding and management [14]. For example, Cho and Kwon developed a deep learning-based early warning score that accurately anticipates patient deterioration using vital signs recorded over 8 hours, specifically designed for patients in general wards. Similarly, in intensive care units (ICUs), some research has utilized ML techniques and continuous vital signs monitoring to anticipate deterioration [15, 16]. Nonetheless, it is crucial to acknowledge that continuous vital signs measurements might not always be available in general wards, posing a challenge to the widespread implementation of these approaches.

doi:10.6342/NTU202302444

doi:10.6342/NTU202302444

1.2 Purpose

Our principal goal is to advance the in-hospital cardiac arrest (IHCA) prediction field by harnessing the data generated through existing care processes. By capitalizing on these resources, our primary objective is to develop a refined and more accurate prediction model for IHCA.

In parallel, we recognize the critical importance of interpretability in healthcare applications. Although machine learning models have demonstrated remarkable potential in predicting IHCA, their complex nature, and limited transparency have hindered their widespread adoption in clinical settings. Consequently, our secondary objective involves creating a prediction model that achieves exceptional predictive performance and offers clear and comprehensible explanations for its predictions. By incorporating explainability into our model, we strive to enhance its utility and facilitate its integration into routine clinical practice.

It is important to note that our primary objective was not to identify every potential cardiac arrest patient but rather to identify individuals who may have been overlooked within the existing care processes and available resources. By focusing on patients who experienced in-hospital cardiac arrest (IHCA) but were not initially identified as high-

doi:10.6342/NTU202302444

risk individuals, we aimed to uncover specific characteristics that could help improve the identification and intervention strategies for these patients.

The ultimate goal was to provide healthcare professionals with an improved tool enabling proactive interventions, potentially averting adverse outcomes associated with IHCA.

doi:10.6342/NTU202302444

Chapter 2 Literature Review

2.1 Importance of IHCA Prediction

In-hospital cardiac arrest (IHCA) represents a significant health concern associated with considerable morbidity and mortality, although its incidence is relatively low. In the United States, an annual estimation of 290,000 cardiac arrest cases are attended to in hospitalized patients [2]. Survival rates during hospitalization and post-discharge for patients experiencing cardiac arrest remain deficient, with nearly 25% survival rates post-discharge in the United States and less than 20% globally [3, 4]. Several intra-arrest factors, such as a witnessed or monitored event and daytime occurrence, have been identified as predictive of enhanced survival rates [6]. Prompt recognition of high-risk patients is thus critical, given the considerable impact of healthcare team preparedness and responsiveness on post-cardiac arrest survival outcomes.

Nevertheless, achieving an accurate assessment of IHCA incidence and understanding its implications for patient outcomes is challenging. Previous estimates often rely on data from single institutions or small hospital clusters within similar geographical regions [17]. Such estimates have limited generalizability due to disparate IHCA definitions and substantial patient and hospital inclusion criteria variations. Furthermore, the absence of comprehensive and standardized data on all IHCA events within national registries and

doi:10.6342/NTU202302444

hospital databases obstructs the determination of overarching IHCA rates and potential shifts. Understanding the scale of this issue is essential, given the considerable variability in IHCA survival rates among hospitals, indicating significant opportunities for enhancing outcomes.

Data from the American Heart Association's Get with The Guidelines-Resuscitation (GWTG-R) registry indicates that the average age of IHCA patients is 66 years, with males constituting 58% of the cases [2]. Non-shockable rhythms such as asystole or pulseless electrical activity represent the most frequent presenting rhythms during IHCA, observed in approximately 81% of cases. Cardiac causes are implicated in most cardiac arrests, accounting for 50%-60% of cases, followed by respiratory insufficiency, contributing to 15%-40% of cases [3]. To forestall IHCA, it is crucial to implement robust systems to identify patients at risk of deterioration and swiftly deliver suitable interventions. Rapid response teams have proven effective in detecting deteriorating patients and initiating timely interventions. Essential treatment elements during cardiac arrest include high-quality chest compressions, adequate ventilation, early defibrillation when appropriate, and swift attention to reversible causes like hyperkalemia or hypoxia [18]. However, the evidence supporting more advanced treatment strategies, such as

doi:10.6342/NTU202302444

extracorporeal cardiopulmonary resuscitation (ECPR) or targeted temperature management, is still being determined and necessitates further research.

Post-cardiac arrest care represents a vital phase in managing IHCA patients. This phase prioritizes identifying and treating the underlying cause of cardiac arrest, providing hemodynamic and respiratory support, and potentially implementing neuroprotective strategies like targeted temperature management. Prognostication and decision-making in the post-cardiac arrest period require a comprehensive approach considering the potential for neurological recovery and ongoing multiorgan failure. It is important to abstain from prematurely withdrawing care without definitive prognostic signs during and after cardiac arrest. Advanced monitoring techniques and biomarkers may facilitate prognostication and guide decision-making in the post-resuscitation period [3].

To foster improvement of IHCA outcomes on a broader scale, it is strongly recommended that hospitals actively engage in national quality-improvement initiatives. These initiatives strive to ameliorate the care delivered to IHCA patients by implementing evidence-based guidelines, refining resuscitation training for healthcare providers, and fostering a culture of continuous quality improvement. Exchanging best practices, participation in standardized reporting systems, and involvement in collaborative research

doi:10.6342/NTU202302444

endeavors can aid in identifying strategies to decrease IHCA rates, enhance survival rates, and improve long-term patient outcomes [19].

doi:10.6342/NTU202302444

2.2 Modified Early Warning Score (MEWS)

The Modified Early Warning Score (MEWS) is the foremost prediction model for detecting patient deterioration in healthcare practice [7]. This scoring system is an invaluable tool in assessing and monitoring the clinical status of patients, enabling healthcare professionals to promptly identify those at risk of critical conditions or adverse events, such as cardiac arrest or sepsis. MEWS plays a pivotal role in facilitating timely interventions to mitigate potential harm.

Early warning scores were conceived in the early 1990s to address the imperative need for early recognition of deteriorating patients [20]. MEWS was developed as an adapted version of the traditional Early Warning Score (EWS) to enhance its predictive value and clinical utility. Since its inception, MEWS has garnered widespread recognition and has been widely implemented in healthcare institutions across the globe.

Primarily employed in hospital settings, particularly in general wards, MEWS focuses on the vigilant monitoring of patients' vital signs to detect any indications of deterioration. It encompasses a comprehensive set of physiological parameters, with each parameter assigned a score based on the degree of abnormality observed. These individual scores are aggregated to generate an overall MEWS score, indicating the

doi:10.6342/NTU202302444

patient's risk of clinical deterioration. Higher MEWS scores are associated with a greater likelihood of deterioration.

While the specific variables and scoring system employed in MEWS may exhibit minor variations across different healthcare institutions, several typical variables are commonly included. These variables encompass respiratory rate, heart rate, systolic blood pressure, body temperature, and level of consciousness. Each variable is assigned a score according to predefined thresholds or ranges. For instance, if the respiratory rate falls within the normal range, it may be given a score of 0. However, deviations from the normal range may result in a higher score, signifying an elevated risk level. The Modified Early Warning Score (MEWS) classification rule is depicted in Table 1, providing a comprehensive visual representation of its categorization system.

Table 1. This is a table displaying the Modified Early Warning Score (MEWS).

	3	2	1	0	1	2	3
SBP (mmHg)	<70	71-80	81-100	101-199		≥200	
HR (bpm)		<40	41-50	51-100	101-110	111-129	≥130
RR (bpm)		<9		9-14	15-20	21-29	≥30
BT (°C)		<35		35-38.4		≥38.5	

doi:10.6342/NTU202302444

AVPU score	A	V	P	U

Note: 'SBP' stands for Systolic Blood Pressure, 'HR' represents Heart Rate, 'RR' denotes Respiratory Rate, 'BT' refers to Body Temperature. The consciousness level is represented as 'A' for Alert, 'V' for Reacting to Voice, 'P' for Reacting to Pain, and 'U' for Unresponsive.

Healthcare professionals regularly monitor patients' vital signs and calculate their MEWS scores at predetermined intervals. A higher MEWS score triggers an appropriate response protocol, such as notifying the medical team or increasing the frequency of vital signs monitoring. These interventions ensure timely and effective care, thereby preventing further deterioration and optimizing patient outcomes.

MEWS has demonstrated its effectiveness in identifying patients at risk of clinical deterioration and has significantly contributed to improved patient outcomes through early intervention [21]. Its simplicity and user-friendly nature have facilitated widespread adoption in healthcare institutions worldwide. Nevertheless, it is crucial to acknowledge that MEWS is just one component of a comprehensive approach to patient monitoring and should be utilized in conjunction with clinical judgment and the expertise of healthcare professionals.

doi:10.6342/NTU202302444

2.3 Model of IHCA Prediction

Several scoring systems have been employed to identify hospitalized patients at a heightened risk of clinical deterioration [2, 9-11, 16, 22, 23]. These systems are typically developed by selecting relevant variables associated with predictive outcomes. Most commonly used early warning scoring systems, such as the Modified Early Warning Score (MEWS) [8], rely on vital signs, including temperature, heart rate, respiratory rate, and blood pressure, for clinical assessments. However, the areas under the receiver operating characteristic curve (AUROC) for MEWS have often been reported to fall below 0.7 in numerous studies. Consequently, researchers have sought to enhance prediction performance by incorporating additional clinical data such as laboratory results, demographics, and heart rate variability [9-11, 13, 14, 24]. These approaches have improved accuracy, reducing false alarms and more reliable detection. Nonetheless, the feasibility of these methods may be limited in clinical units where regular measurement of such clinical data is not practical. Table 2 provides a comprehensive overview of research on predicting in-hospital cardiac arrest (IHCA), offering valuable insights into the diverse studies conducted in this field.

doi:10.6342/NTU202302444

Table 2. Comparison of Studies on IHCA Detection in Hospital Settings

	Cho et al[16]	Kim et al[15]	Kwon et al[25]	Bartkowiak et al[10]	Green et al[9]	Churpek et al[8]
Publication Year	2020	2019	2018	2018	2018	2016
research subject	ward	ICU	ward	ward (Surgical)	ward	ward
Interval of Vital Sign	8 hours	6 hours	8 hours		4 hours	8 hours
AUROC for MEWS	0.684	0.746	0.603	0.750	0.698	0.698
AUROC for research model	0.865 (DEWS)	0.896 (FAST-PACE)	0.850 (DEWS)	0.790 (eCART)	0.801 (eCART)	0.801 (Random Forest)
SBP	V	V	V	V	V	V
HR	V	V	V	V	V	V
RR	V	V	V	V	V	V
BT	V	V	V	V	V	V

doi:10.6342/NTU202302444

DBP		V		V	V	V
SpO2		V		V	V	V
AVPU score				V	V	V

The expanding adoption of artificial intelligence and machine learning (ML) systems has fundamentally transformed the field of biomedicine from the molecular level to disease management. ML enables automated analysis of highly complex data and generates meaningful insights. ML models can potentially enhance prediction accuracy using the same dataset or reduce features while maintaining performance [14]. Cho and Kwon developed a deep learning-based early warning score that accurately predicts deterioration in patients within general wards by leveraging vital signs recorded over 8 hours. Some studies have employed ML techniques with continuous vital signs to predict deterioration in intensive care units (ICUs) [15, 16]. However, continuous monitoring of vital signs may not be readily available in general wards.

Therefore, recognizing the limitations posed by the availability of continuous vital signs measurements in general wards, our study aimed to develop a more precise machine-learning model for predicting clinical deterioration, leveraging only five commonly measured vital signs: heart rate, systolic blood pressure, diastolic blood pressure, respiratory rate, and body temperature. By focusing on these vital signs, which are

doi:10.6342/NTU202302444

routinely monitored in general wards, our model offers a practical and accessible approach to risk assessment. This approach presents a potential alternative to the Modified Early Warning Score (MEWS) system, which relies on a broader range of variables.

Including heart rate, systolic blood pressure, diastolic blood pressure, respiratory rate, and body temperature in our model is supported by their clinical significance in reflecting a patient's physiological state and overall health. Heart rate is a vital indicator of cardiac function and can provide insights into autonomic regulation and overall cardiovascular well-being. On the other hand, systolic blood pressure offers valuable information about perfusion and tissue oxygenation, highlighting a patient's circulatory status. Lastly, respiratory data, including respiratory rate, is crucial in assessing respiratory function and detecting signs of respiratory distress or compromise.

The potential benefits of our proposed model are noteworthy. By utilizing the regularly measured vital signs, healthcare providers in general wards can readily implement this approach without additional resources or specialized monitoring equipment. Moreover, our model can improve prediction accuracy compared to scoring systems such as MEWS. By harnessing the power of machine learning, we aim to uncover intricate

doi:10.6342/NTU202302444

patterns and relationships within the vital signs data that might not be apparent through conventional approaches.

It is important to acknowledge that further validation and evaluation of the proposed model will be essential to ascertain its effectiveness and clinical utility. Prospective studies and comparative analyses against existing scoring systems are warranted to establish the superiority of our model in accurately predicting clinical deterioration within general ward settings. Ultimately, our objective is to provide healthcare professionals with a reliable tool that can enhance their ability to identify patients at risk of deterioration promptly, enabling timely interventions and improved patient outcomes.

doi:10.6342/NTU202302444

Chapter 3 Materials and Method for TEWS (Time-Series Early Warning Score)

Our primary objective is to develop an enhanced prediction model for in-hospital cardiac arrest (IHCA) by utilizing the existing data derived from current care processes. The conventional approach to IHCA prediction within the general ward setting involves using all available physiological data within a specific period. In our initial study, we postulated that incorporating information from multiple time points could yield a more comprehensive understanding of patients' physiological changes and trends. By integrating data from different time points, we sought to capture potential early warning signs of deteriorating health. This innovative approach was anticipated to fortify the model's predictive capabilities and facilitate the timely identification of patients at risk of cardiac arrest.

By utilizing existing data and refining predictive capabilities, we aspire to enhance patient safety and optimize outcomes in the general ward setting. Our overarching goal is to equip healthcare professionals with an improved tool that enables proactive interventions with the potential to avert adverse outcomes associated with IHCA.

doi:10.6342/NTU202302444

3.1 Ethics Declarations

This retrospective cohort study was granted approval by the En-Chu-Kong Hospital Institutional Review Board (IRB) with the assigned number ECKIRB1071001. We affirm that all experiments conducted adhered to applicable guidelines and regulations. The data utilized in this study were extracted from de-identified electronic health records (EHRs) by an IT specialist, ensuring that patient identities remained unlinked to the research team. Given this cohort study's retrospective nature and de-identified data utilization, the En-Chu-Kong Hospital IRB (ECKIRB1071001) waived the requirement for written informed consent.

3.2 Setting and Study Population

The research was executed within a community-based general hospital, involving a cohort drawn from the inpatient population. The data analyzed were sourced from the electronic health records (EHRs) of adult inpatients, all 20 years or older, who presented at the facility for care between August 2016 and September 2019. All identifiable patient information was de-identified and anonymized before the analysis phase to maintain confidentiality.

doi:10.6342/NTU202302444

Five core vital signs — systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), respiratory rate (RR), and body temperature (BT) — were utilized as the predictive features for this study. In line with standard medical practice, these vital signs were measured and recorded by healthcare staff, typically two to three times per day, which included measurements taken during the day, night, and early morning hours. For the study, we delineated the time window (TW) for these measurements into 8-hour segments, constituting three TWs within each 24-hour day. Each TW captured a complete set of the aforementioned vital signs.

The features captured during each TW were leveraged at three distinct TW intervals: 1, 3, and 6 TWs (corresponding to 8, 24, and 48 hours, respectively). For each TW, one complete set of features was contained. A visual representation of the study's methodological process can be found in Figure 1, which succinctly illustrates the step-by-step progression of the research.

doi:10.6342/NTU202302444

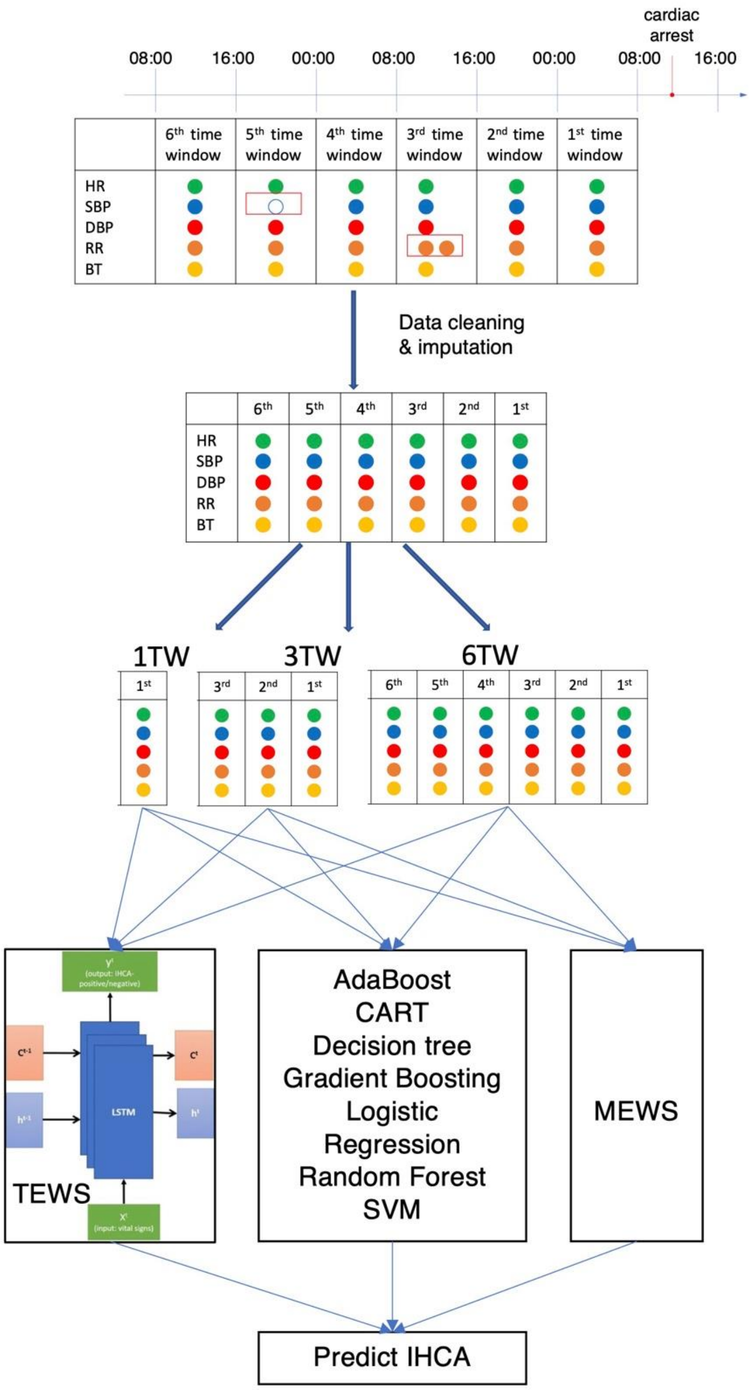


Figure 1. Research Procedure. TW refers to the time window.

doi:10.6342/NTU202302444

The hospital data was partitioned by date into a training and validation set (August 2016–November 2018) and a testing set (December 2018–September 2019). The training and validation set was utilized for developing the Modified Early Warning Score (MEWS), while the testing set was used to determine the MEWS performance. For binary classification, we employed the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC) as evaluation metrics. The pertinent characteristics of the study cohort are comprehensively presented in Table 3, providing essential details regarding the demographic and clinical attributes of the participants.

Table 3. Characteristics of the Research Sample expressed as mean ± standard deviation

Characteristic	Training & Validation Set	Testing Set
Duration of the Study	2016/8- 2018/11	2018/12 - 2019/9
Total patient number	11,762	5,103
Patient number with IHCA	81	37
Age	63.8 ± 19.9	63.7 ± 20.5
Gender: Male (%)	5,875 (49.9)	2,293 (44.9)
Body Weight (Kg)	63.2 ± 14.7	63.3 ± 17.6
Respiratory rate (1st TW)	18.9 ± 4.1	19.1 ± 5.0

doi:10.6342/NTU202302444

Diastolic blood pressure (1st TW)	73.6 ± 15.2	72.3 ± 20.1
Systolic blood pressure (1st TW)	133.2 ± 31.0	135.0 ± 40.8
Body Temperature (1st TW)	36.7 ± 4.4	37.5 ± 6.4
Heart rate (1st TW)	83.4 ± 21.5	84.9 ± 23.5

3.3 Main Outcome

The primary outcome under investigation in this study was cardiac arrest, defined as the absence of a detectable pulse accompanied by attempts at resuscitation. A meticulous examination of the electronic health records (EHRs) was conducted to ascertain the precise timing of each cardiac arrest event. This comprehensive analysis accurately identified and classified the selected inpatients into two distinct groups: positive and negative.

The positive group consisted of inpatients who experienced at least one cardiac arrest event while admitted to the general wards. In cases where patients encountered multiple cardiac arrest events during their hospitalization, only the initial event was considered for analysis. This approach ensured that each patient was represented by a single cardiac arrest event, maintaining the independence of observations and preventing duplication in the dataset.

doi:10.6342/NTU202302444

Conversely, the negative group comprised inpatients not admitted to the intensive care unit (ICU) and did not encounter any cardiac arrest events throughout the study period; including this negative group allowed for a comprehensive evaluation of the factors associated with the occurrence of cardiac arrest. By comparing the positive and negative groups, we aimed to provide a thorough understanding of the characteristics and circumstances surrounding cardiac arrest events in the general ward setting.

3.4 Model Development

3.4.1 Data Preprocessing

Given the inherent susceptibility to human or system errors in the compilation of electronic health records (EHRs), our dataset was potentially exposed to the problem of missing values. It is conceivable, for instance, that health personnel may not have recorded certain vital sign measurements within specific time windows (TWs), leading to incomplete TW data. To address this challenge, we implemented the method of multiple imputations by chained equations[24]. This technique, which effectively reintroduces the natural variability associated with missing data and accounts for the resultant uncertainty, is particularly effective for valid statistical inferences. In instances of duplicate data within the same TW, our protocol involved retaining the highest value.

doi:10.6342/NTU202302444

Additionally, we encountered a substantial distribution range in the feature values within our dataset, which could complicate the classifier training process. To surmount this hurdle, we employed standard scores (z-scores) to normalize the values of the features. This statistical transformation, by adjusting the distribution of feature values, aids in ensuring more reliable model training.

3.4.2 Handing Imbalanced Data

Imbalances in datasets are common in practical scenarios, especially within medical research, where class distributions frequently exhibit severe skewness. This issue of class imbalance similarly afflicted our dataset. It is essential to acknowledge that the effectiveness of most machine learning algorithms is maximized when the classes are balanced or nearly so. Unaddressed imbalances in datasets could undermine the classifier's performance, as biases in the training data could result in the neglect of underrepresented classes. There are several strategies, such as under-sampling (i.e., eliminating instances from the majority class) or oversampling (i.e., cases duplicating from the minority class), that are recommended to manage such imbalanced datasets. In our prior research [26], we mitigated the imbalance within the dataset by under-sampling negative samples to detect in-hospital cardiac arrest (IHCA).

doi:10.6342/NTU202302444

For this study, however, instead of adopting the oversampling or under-sampling techniques, we chose a modified weight-balancing method to address the class imbalance issue in our dataset. This method was particularly effective when one class significantly outnumbered the other. The modified weight balancing strategy adjusted the class weights according to the ratio of IHCA-positive to IHCA-negative samples, thereby ensuring equitable contributions from all classes during the loss computation process. In addition, we incorporated focal loss, a technique designed to assign greater weight to the minority class during training, to equalize the representation of classes further. By combining focal loss and the weight-balancing method, we successfully addressed the class imbalance issue within our imbalanced dataset while developing the Time-Series Early Warning Score (TEWS).

3.4.3 TEWS Model Development

Our TEWS model consisted of three recurrent neural network (RNN) layers utilizing Long Short-Term Memory (LSTM) units [27]. RNNs are neural networks designed to handle sequential data, making them well-suited for processing EHRs, which inherently possess a time-series structure [28]. The TEWS architecture, incorporating LSTM units, is depicted in Figure 2. LSTM units comprise a cell, an input gate, an output gate, and a forget gate. These components enable the cell to retain information over arbitrary time

doi:10.6342/NTU202302444

intervals while the gates regulate the flow of information into and out of the cell. Leveraging the strengths of LSTM in handling time-series data, TEWS effectively processes the temporal nature of the EHRs.

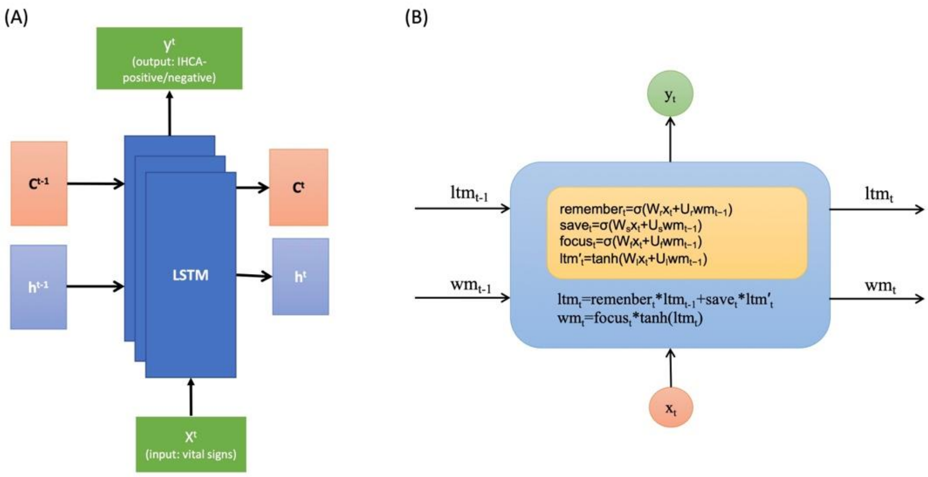


Figure 2**.** (A) Structure of TEWS. (B) Structure of LSTM cell

The dataset allocated for training was utilized to refine the Time-Series Early Warning Score (TEWS) model, with subsequent evaluation of the model's performance conducted using a validation dataset. Our TEWS model implemented a division of the training and validation datasets at an 8:2 ratio. The algorithm employed to define six distinct time windows for each vital sign measurement across the entirety of the inpatient population is depicted in Figure 3. This methodological approach provides a systematic and uniform process for creating time windows, facilitating an encompassing analysis of the vital sign data.

doi:10.6342/NTU202302444

Algorithm TW_data(*S*)

Input: a set of inpatients *S* with vital signs;

Output: 6 time windows of each inpatient in *TW*;

Method:

1. TW = ϕ ;
2. S^s ← sort(*S*); // sorted S by medical record numbers of inpatients
3. **FOR EACH** inpatient P ∈ S^s **DO**
4. P^s ← sort(P); // sorted P by time of measure of each vital sign
5. **FOR EACH** vital sign data B ∈ P^s **DO**
6. Create 6 time windows (1TW~6TW) by time for each inpatient P;
7. **FOR EACH** TW, data V ∈ B and V ∈ this TW
8. ∀ V, calculate the first and last value, maximum, minimum, mean, and standard deviation during the same time window;
9. **END FOR EACH**
10. **END FOR EACH**
11. **END FOR EACH**
12. **RETURN** *TW*;

Figure 3**.** The procedure for generating time interval data.

3.5 Performance Evaluation

The LSTM-based system for our study was implemented using the Python-based scikit-learn package [29]. At the same time, the neural networks were brought into effect using Keras, with TensorFlow as the backend engine. We employed several classification algorithms within the scikit-learn package for an extensive benchmarking procedure [30]. This suite of algorithms included AdaBoost [31, 32], random forest [33], logistic

doi:10.6342/NTU202302444

regression [34], gradient boosting [35], classification and regression tree [36], naïve Bayes [37], support vector machine (SVM) [38, 39], k-nearest neighbor [40, 41], and C4.5 decision tree [42].

Gradient boosting integrates weak prediction models to yield a single prediction model and can be interpreted as an optimization algorithm for an appropriate cost function[43]. Logistic regression is a statistical model that predicts the probability of a specific class utilizing a logistic function to model a binary dependent variable. Random forest, an ensemble learning method, has utility in classification, regression, and other tasks by constructing multiple decision trees during training and generating predictions based on the modal or mean values of the individual tree predictions.

Due to the imbalanced nature of our dataset, our proposed Time-Series Early Warning Score (TEWS) allocates class weights according to the ratio of IHCA-positive to IHCA-negative samples. Predicted probabilities for each instance in the validation dataset were calculated from each model to assess the predictive performance of the classification mentioned above algorithms and TEWS. The Modified Early Warning Score (MEWS) was also computed for comparative purposes. To contextualize the accuracy of these results against existing literature, we determined the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC)

doi:10.6342/NTU202302444

while considering whether an event occurred within an eight-hour window of each observation. These metrics are typically employed in the comparative evaluation of early warning scores.

In summary, we employed various contemporary algorithms and our proposed TEWS to benchmark and compare their prediction performance. The calculated predicted probabilities, AUROCs, and AUPRCs provide valuable insights into the accuracy and effectiveness of these models within the context of existing literature and contribute to the advancement of IHCA prediction in clinical practice.

3.6 Feature Selection

The feature selection process fundamentally identifies the most informative features from a pool of potentially useful features to efficiently distinguish between classes. The undertaking of feature selection can be effectuated through a process of elimination, which bifurcates into two primary methodologies: filter methods and wrapper methods. Wrapper methods leverage the performance of the predictor as the selection criterion. In this arrangement, the predictor is integrated into a search algorithm that identifies the subset of features that deliver the highest predictor performance. Sequential backward selection (SBS) algorithms are widely utilized within the framework of wrapper

doi:10.6342/NTU202302444

methods. SBS algorithms adopt a straightforward and greedy approach in their quest for feature selection. They systematically remove one feature from the complete set of features at a time while ensuring a slight decrement in predictor performance.

The SBS algorithm significantly benefits when the optimal feature subset consists of fewer features[44, 45]. This algorithm effectively pinpoints relevant features, contributing to reliable and efficient class discrimination.

doi:10.6342/NTU202302444

Chapter 4 Materials and Method for TEWS-X (Explainable Time-Series Early Warning Score)

Our primary objective is to develop an enhanced prediction model for in-hospital cardiac arrest (IHCA) by utilizing the existing data derived from current care processes. In addition to developing an enhanced prediction model, our second goal is to create an explainable prediction model based on our previous research. While deep learning models have shown great promise in predicting IHCA, the lack of interpretability often limits their adoption in clinical practice. Therefore, we aim to create a model that achieves high predictive performance and provides clear and understandable explanations for its predictions.

By incorporating techniques such as SHAP (SHaley Additive exPlanations) values, we can generate explanations for each prediction made by the model[46, 47]. These explanations will highlight the key contributing factors and their relative importance in determining the patient's risk of IHCA. This transparent and interpretable approach will enable healthcare providers to understand the underlying reasons behind the model's predictions, increasing their confidence in utilizing its outputs for clinical decision-making.

doi:10.6342/NTU202302444

Our research aims to develop an enhanced prediction model for IHCA using existing data and to ensure its interpretability and explainability. By achieving these goals, we hope to provide healthcare providers with a powerful tool that can assist in risk stratification and early intervention and ultimately improve patient outcomes in the context of IHCA.

4.1 Ethics Declarations

This retrospective cohort study was executed with the endorsement of the Institutional Review Board (IRB) at the En-Chu-Kong Hospital under the assigned IRB number ECKIRB1071001. The study was conducted meticulously, adhering to all relevant guidelines and regulations about human research. The dataset utilized in this study aligns with that employed in the initial investigation[48], and we attest to the complete compliance of all experiments with these established guidelines and regulations.

4.2 Setting and Study Population

The study cohort consisted of adult inpatients aged 20 years or older who received medical care at a community-based general hospital. The study dataset comprised these individuals' electronic health records (EHRs) from August 2016 to September 2019.

doi:10.6342/NTU202302444

Importantly, it should be noted that this timeframe predates the emergence of the COVID-19 pandemic, thereby minimizing any potential impact of the pandemic on healthcare processes and resource allocation that could confound the study results. For this investigation, we identified five vital signs as crucial predictors: systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, and body temperature. These vital signs were routinely assessed by healthcare professionals, with measurements conducted multiple times per day at varying periods, including daytime, nighttime, and early morning. To establish a standardized time window (TW) for data collection, we defined each TW as a duration of 8 hours. Thus, three TWs were identified within a single day, during which the aforementioned vital signs were documented.

4.3 Main Outcome

The principal aim of this study was to scrutinize the incidence of cardiac arrest as the primary outcome of interest. Cardiac arrest was delineated as the cessation of a discernible pulse coupled with resuscitation efforts. An in-depth appraisal of electronic health records (EHRs) was undertaken to ascertain the exact timing of each cardiac

doi:10.6342/NTU202302444

arrest event. This facilitated accurately identifying and categorizing the selected inpatients into positive and negative groups.

The positive group encapsulated inpatients who underwent at least one instance of cardiac arrest during their admission to the general wards. In scenarios where patients experienced multiple cardiac arrest events throughout their hospitalization, solely the first event was taken into account for analysis. This methodology ensured that each patient was represented by a singular cardiac arrest event, thereby maintaining the independence of observations and avoiding duplication within the dataset.

Conversely, the negative group consisted of inpatients not admitted to the intensive care unit (ICU) and did not experience any cardiac arrest event throughout the study period. A comprehensive evaluation of the factors associated with cardiac arrest occurrence was possible by including positive and negative groups. This comparative analysis allowed for a thorough examination of the predictors and risk factors related to cardiac arrest in the general ward setting.

doi:10.6342/NTU202302444

4.4 Model Development

4.4.1 Data Preprocessing

In addition to the five vital signs measured across six times windows, we also incorporate the calculation of vital signs differences between different time windows. For instance, "HR1-6" represents the difference in heart rate between the first and the sixth times windows. Furthermore, age and gender variables are included in our dataset. Consequently, our augmented dataset encompasses 82 features, expanding from the original set of 30 features.

A notable distinction in the second study pertains to the nomenclature employed for the time windows, which differs from that used in the initial research. In the second study, we adopted a naming convention aligned with the chronological sequence of the intervals. Consequently, the first time window corresponds to the farthest period preceding the event, whereas the sixth time interval represents the immediate timeframe preceding the event. This revised ordering facilitates a clearer understanding of the temporal relationship between the intervals and the occurrence of the event under investigation. The nomenclature of features within the augmented dataset is presented in Table 4, providing a comprehensive overview of the assigned names for reference and analysis.

doi:10.6342/NTU202302444

Table 4. List of features of augmented dataset, besides age and gender

	Respiratory rate	*Body temperature*	*Heart rate*	*Systolic pressure*	*Diastolic pressure*
1st TW	RR1	BT1	HR1	SBP1	DBP1
2nd TW	RR2	BT2	HR2	SBP2	DBP2
3rd TW	RR3	BT3	HR3	SBP3	DBP3
4th TW	RR4	BT4	HR4	SBP4	DBP4
5th TW	RR5	BT5	HR5	SBP5	DBP5
6th TW	RR6	BT6	HR6	SBP6	DBP6
1st - 2nd	RR1-2	BT1-2	HR1-2	SBP1-2	DBP1-2
1st - 3rd	RR1-3	BT1-3	HR1-3	SBP1-3	DBP1-3
1st - 4th	RR1-4	BT1-4	HR1-4	SBP1-4	DBP1-4
1st - 5th	RR1-5	BT1-5	HR1-5	SBP1-5	DBP1-5
1st - 6th	RR1-6	BT1-6	HR1-6	SBP1-6	DBP1-6
2nd - 6th	RR2-6	BT2-6	HR2-6	SBP2-6	DBP2-6
3rd - 6th	RR3-6	BT3-6	HR3-6	SBP3-6	DBP3-6
4th - 6th	RR4-6	BT4-6	HR4-6	SBP4-6	DBP4-6
5th - 6th	RR5-6	BT5-6	HR5-6	SBP5-6	DBP5-6

doi:10.6342/NTU202302444

MAX-MIN	RRMAX-MIN	BTMAX-MIN	HRMAX-MIN	SBPMAX-MIN	DBPMAX-MIN

4.4.2 Handling Imbalanced Data

In various practical scenarios, particularly in the medical domain, datasets often exhibit imbalanced class distributions, where the prevalence of one class far outweighs the other. Similarly, the dataset utilized in our study also suffered from imbalanced class distribution. However, the performance of many machine learning algorithms is optimized when the number of samples in each class is relatively balanced. Neglecting the management of imbalanced datasets can significantly impact the effectiveness of classifiers. In machine learning classifiers, the biases present in training datasets may result in ignoring minority classes altogether. Therefore, our study aimed to address this imbalanced data issue by leveraging our three tree-based algorithms' inherent "weight" function.

Incorporating weight functions within these algorithms allowed us to account for the imbalanced nature of our dataset. By assigning appropriate weights to the samples from different classes, we aimed to ensure that the classifiers were not biased toward the majority class. This approach enabled us to mitigate the potentially detrimental effects of imbalanced data on the performance of our classifiers. By carefully managing

doi:10.6342/NTU202302444

imbalanced data, we sought to enhance the robustness and reliability of our machine learning models in addressing the specific challenges posed by imbalanced class distributions.

4.4.3 TEWS-X Model Development

Our second goal is to develop an explainable prediction model based on our initial research findings. We employed three tree-based systems implemented using the scikit-learn package in Python, which encompassed logistic regression, random forest, and XGBoost algorithms.

Our study used the three aforementioned tree-based algorithms to build our prediction model. To account for the imbalanced nature of our dataset, we applied weight adjustments during the model training process, ensuring that the minority class (i.e., the less prevalent outcome) receives higher emphasis.

By adjusting the weight for imbalanced data, we aimed to address the potential bias that could arise from our dataset's unequal distribution of outcomes. This approach allows the model to capture patterns better and make accurate predictions for both the majority and minority classes, ultimately improving the overall performance and reliability of the model.

doi:10.6342/NTU202302444

Furthermore, we comprehensively compared the datasets containing 82 features and 30 features. The dataset with 82 features incorporated the five vital signs in six times windows and included additional information such as age and gender. This augmented dataset aimed to provide a more comprehensive representation of the patient's characteristics and physiological changes over time.

4.5 Performance Evaluation

The evaluation of the developed Explainable Time-Series Early Warning Score (TEWS-X) was performed using the dedicated testing set, employing well-established performance metrics, including the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC). Given the inherent imbalance within our dataset, the AUPRC was employed as an additional evaluation criterion, offering a comprehensive assessment of our model's performance. To assess the impact of feature augmentation on the predictive accuracy of the models, we conducted a comparative analysis, examining the performance of the three tree-based algorithms on both datasets. This analysis aimed to determine whether adding additional features improved the models' ability to discriminate between different outcomes and enhance the accuracy of predictions.

doi:10.6342/NTU202302444

4.6 Expandability with SHAP

We employed the Shapley Additive Explanations (SHAP) method to analyze the feature importance of our three tree-based algorithms[46, 47]. This approach allowed us to gain insights into the global and local importance of features in each case, focusing on positive-prediction instances. By utilizing the SHAP method, we aimed to understand the overall contribution of features in our models. The global feature importance provided us with a comprehensive view of the relative significance of each input variable in predicting outcomes. This information was valuable in identifying the key factors driving the predictions made by our tree-based algorithms.

Additionally, we examined the local importance of features in positive-prediction cases. This analysis helped us understand the specific factors influencing individual instances where our models successfully predicted positive outcomes. We better understood these predictions' underlying mechanisms and reasoning by investigating the local importance.

doi:10.6342/NTU202302444

Chapter 5 Result

5.1 Result for TEWS (Time-Series Early Warning Score)

This study carefully selected a cohort of 16,865 adult admissions for analysis. Among these admissions, 118 individuals (0.7%) encountered cardiac arrest within a general ward setting, as depicted in Table 5. To provide a comprehensive understanding of the data, we describe the characteristics of both IHCA-positive and IHCA-negative cases in Figure 4.

Table 5. Demographic Information of the Research Sample

Characteristic	**Data for Training and Validation**	**Data for Testing**
Duration of the Study	August 2016-November 2018	December 2018-September 2019
Number of Patients	11,762	5,103
IHCA	81	37
Age	63.8 ± 19.9	63.7 ± 20.5
Gender: Male (%)	5,875 (49.9)	2,293 (44.9)
Body Weight	63.2 ± 14.7	63.3 ± 17.6

doi:10.6342/NTU202302444

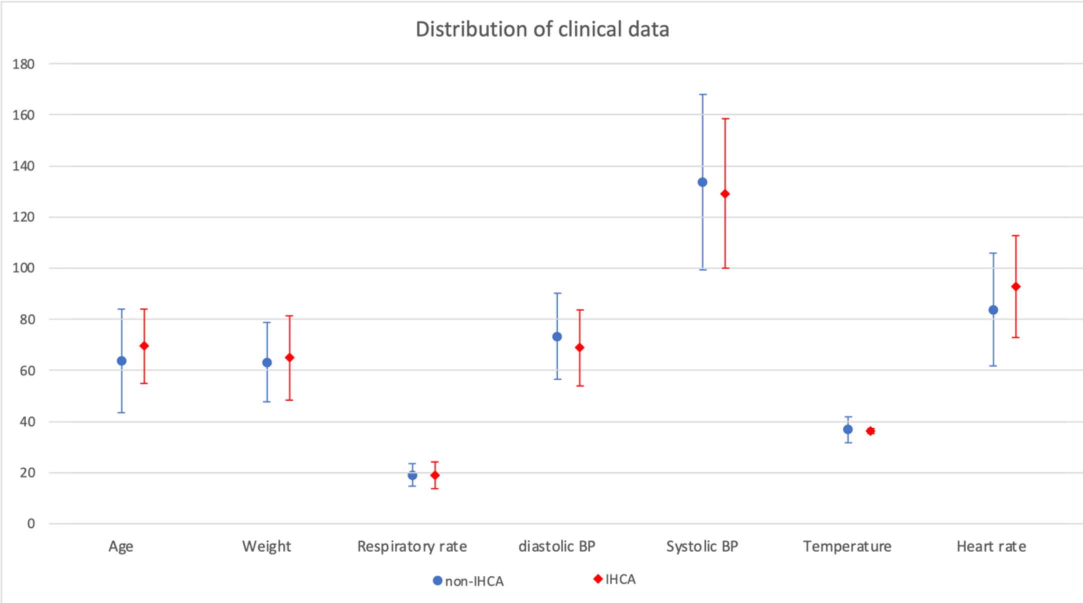


Figure 4. Distribution of vital sign data in general ward patients, expressed as mean ± standard deviation (SD).

Note: 'SD' denotes standard deviation, 'IHCA' designates the group experiencing an IHCA event, while 'non-IHCA' represents the group that did not have an IHCA event.

We employed two distinct tasks to assess the efficacy of our proposed Time-Series Early Warning Score (TEWS). Subsequently, we conducted a comparative analysis between the TEWS model and these classification approaches. The tasks encompassed the following components, each shedding light on the TEWS model's performance and predictive capabilities.

doi:10.6342/NTU202302444

5.1.1 Performance with one, three, and six time windows.

In the cohort characterized by one time window (TW) or an 8-hour interval, the Time-Series Early Warning Score (TEWS) model was employed. This model uses five vital signs from a specific TW to anticipate in-hospital cardiac arrest (IHCA) events. Figure 5 illustrates a comparative evaluation of the TEWS model, the Modified Early Warning Score (MEWS), and other classifiers. Remarkably, the Support Vector Machine (SVM) and logistic regression algorithms indicated the most elevated area under the receiver operating characteristic curve (AUROC) values (0.729 and 0.721, respectively). This was followed by gradient boosting (0.712) and the TEWS (0.688). Nonetheless, all classifiers performed within the range demarcated by the MEWS.

Within the cohort characterized by 3 TWs or a 24-hour interval, features derived from three TWs (24 h) were utilized to predict IHCA events using the TEWS model. Each TW contained a unique set of vital signs, culminating in 15 features. Notably, the TEWS model demonstrated a superior AUROC value (0.762), outperforming logistic regression (0.730), random forest (0.676), MEWS (0.649), and other algorithms.

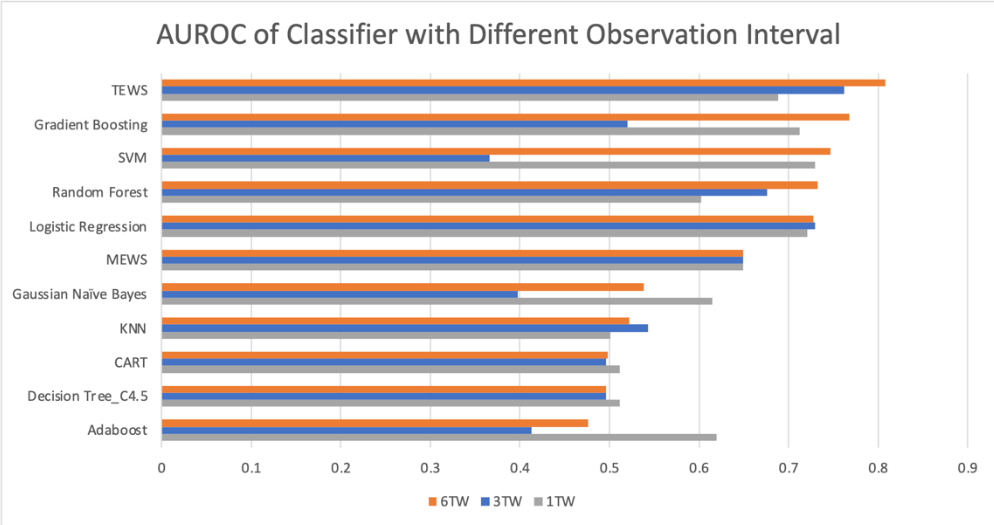
In the cohort characterized by 6 TWs or a 48-hour interval, the TEWS model employed features derived from six TWs (48 h) to anticipate IHCA events. The TEWS model exhibited a higher AUROC value (0.808) when compared to gradient boosting (0.768),

doi:10.6342/NTU202302444

SVM (0.747), random forest (0.733), and other algorithms, affirming its consistent high performance across the 1TW, 3TW, and 6TW groups.

While most classification algorithms demonstrated comparable performance levels when using features from a single TW, with AUROC values ranging between 0.62 and 0.73 (AUROC of MEWS: 0.65), the predictive abilities of specific classifiers improved when integrating data from multiple TWs. Our TEWS model exhibited superior performance in the 6TW group (AUROC = 0.808, AUPRC = 0.052) compared to the MEWS (AUROC = 0.649, AUPRC = 0.015).

doi:10.6342/NTU202302444



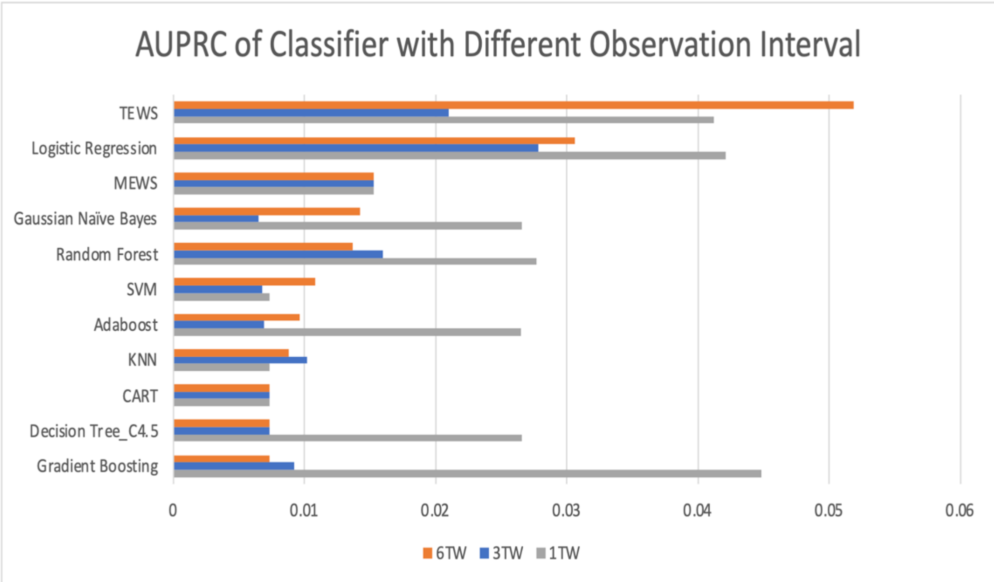


Figure 5**.** Values of AUROC and AUPRC for classifiers utilizing one, three, and six time windows. 'TW' stands for time window and 'TEWS' refers to the time-series early warning score.

doi:10.6342/NTU202302444

5.1.2 Performance with Features Chosen via SBS Algorithm

For the preliminary task of anticipating in-hospital cardiac arrest (IHCA), the Time-Series Early Warning Score (TEWS) exhibited optimal performance when integrating data from six time windows (TWs), which represented a duration of 48 hours. This methodology, however, required 30 features from the six TWs, prompting us to explore strategies to streamline the feature set without compromising on performance. For this purpose, a Sequential Backward Selection (SBS) algorithm was employed to discern the most critical elements within the six TWs. Figure 6 portrays the features selected, with the initial TW being the most proximate to the cardiopulmonary resuscitation time for IHCA-positive patients. Heart rate, respiratory rate, and systolic blood pressure emerged as essential features in predicting IHCA events. Notably, the outstanding features were the heart rate within the first, fourth, and fifth TWs, respiratory rate, and systolic blood pressure during the initial TW.

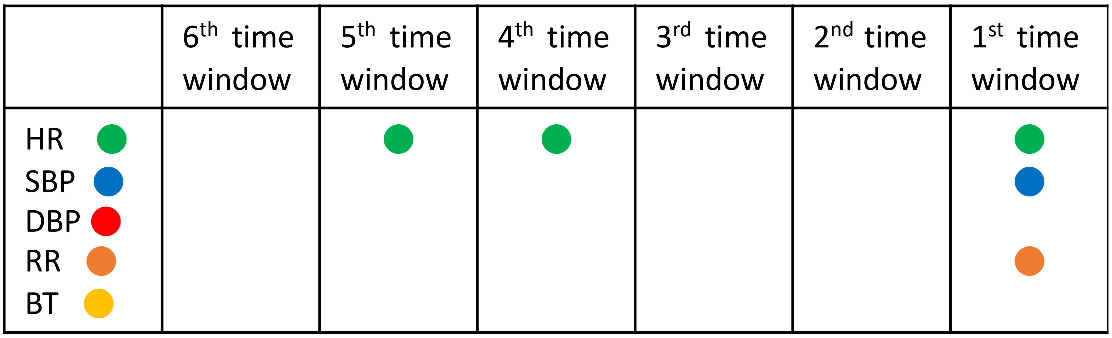


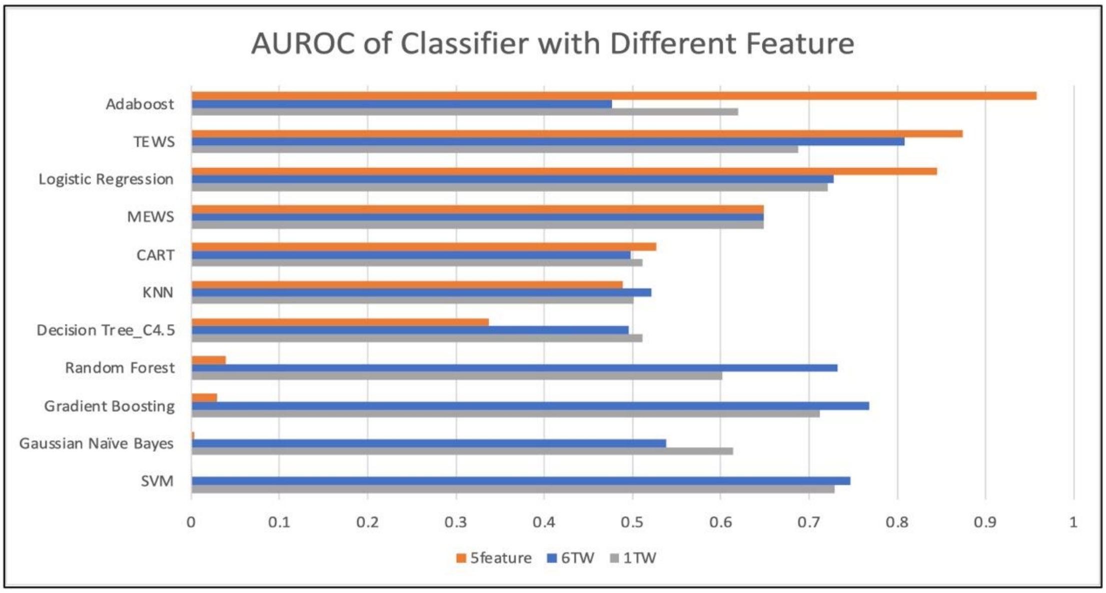
Figure 6. The top 5 features identified using the SBS (Sequential Backward Selection) algorithm.

doi:10.6342/NTU202302444

Upon incorporating the identified quintet of features, we appraised their efficacy when integrated into the Time-Series Early Warning Score (TEWS) model and alternative algorithms. To ascertain the predictive prowess of these algorithms employing the condensed feature set, we contrasted their performance against the Modified Early Warning Score (MEWS) and other classifiers, as depicted in Figure 7.

The TEWS model displayed its paramount performance, achieving a notable AUROC value of 0.875 alongside an AUPRC value of 0.087. Adaboost demonstrated potent performance, with an AUROC of 0.958 and an AUPRC of 0.110. The logistic regression also delivered praiseworthy performance, with an AUROC of 0.845 and an AUPRC of 0.050. These findings emphasize the efficacy of the selected features when paired with the TEWS model and alternative algorithms for the prediction of in-hospital cardiac arrest (IHCA).

doi:10.6342/NTU202302444



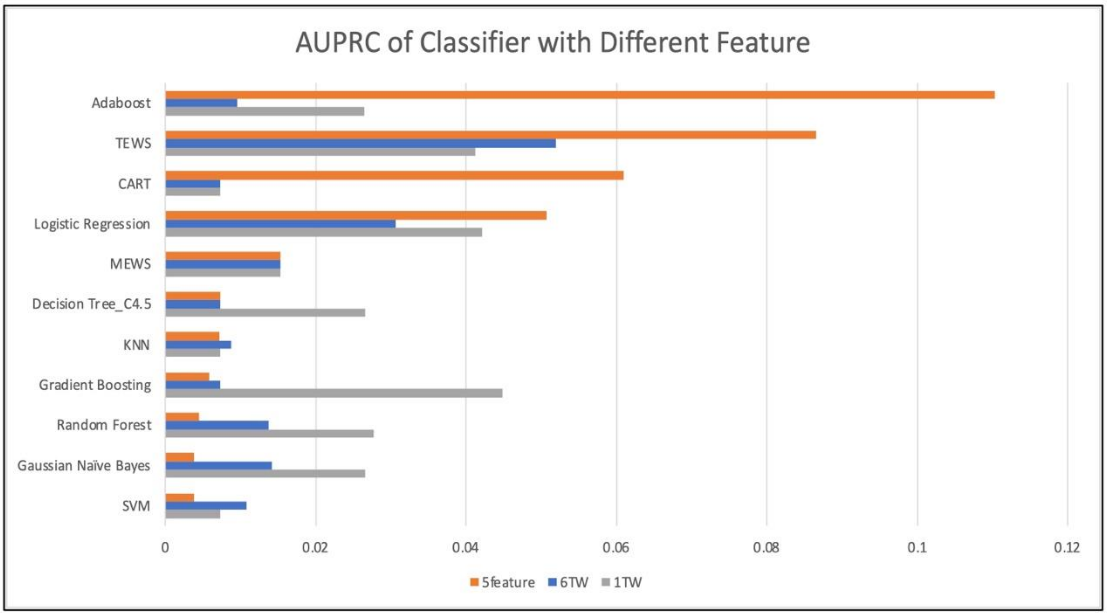


Figure 7**.** Values of AUROC and AUPRC for the classifier utilizing five chosen features at one and six time windows (TWs). 'TW' refers to time window, 'TEWS' is the acronym for time-series early warning score, and '5feature' denotes the five selected features.

doi:10.6342/NTU202302444

5.2 Result for TEWS-X (Explainable Time-Series Early Warning Score)

5.2.1 Performance of TEWS-X

The performance metrics, specifically the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC), achieved by each algorithm for the corresponding feature set and class weighting are presented in Table 6 and Table 7. For instance, in Logistic Regression, the accuracy scores range from 0.750 to 0.773 for the 30-feature sets and from 0.803 to 0.836 for the 82-feature sets. Random Forest achieves accuracy scores ranging from 0.721 to 0.779 for the 30-feature sets and 0.754 to 0.839 for the 82-feature sets. XGBoost attains accuracy scores ranging from 0.739 to 0.770 for the 30-feature sets and 0.776 to 0.834 for the 82-feature sets.

Table 6. AUROC for IHCA Prediction with Varying Class Weight

Weight		1:25	1:50	1:100	balanced
Original features (30)	LR	0.75052	0.75562	0.76169	0.7729
	RF	0.77915	0.74375	0.72176	0.72145
	XGB	0.77026	0.73864	0.75273	0.75389

doi:10.6342/NTU202302444

Feature augmentation (82)	LR	0.80302	0.81353	0.82362	0.83579
	RF	0.83949	0.81057	0.78873	0.75421
	XGB	0.83391	0.80607	0.78575	0.77578

Table 7. AUPRC for IHCA Prediction with Varying Class Weight

		1:25	1:50	1:100	balanced
Original features (30)	LR	0.05403	0.06346	0.06438	0.06237
	RF	0.02478	0.01681	0.01247	0.01117
	XGB	0.03094	0.02922	0.01573	0.01415
Feature augmentation (82)	LR	0.06215	0.06706	0.06905	0.07060
	RF	0.08605	0.04260	0.02203	0.01499
	XGB	0.03776	0.02593	0.01817	0.01558

The results demonstrate that the performance of the algorithms varies depending on the feature set and class weighting. Overall, Random Forest consistently exhibits higher accuracy scores than Logistic Regression and XGBoost across most feature sets and class weightings.

These findings emphasize the importance of considering feature selection and class weighting when applying machine learning algorithms. The table provides valuable

doi:10.6342/NTU202302444

insights into the relative performance of the algorithms under different settings, aiding in selecting the most suitable approach based on the specific requirements of the problem at hand. Considering the AUROC and AUPRC metrics, the random forest model with a class weighting of 1:25 and utilizing 82 features achieved the best prediction performance.

doi:10.6342/NTU202302444

5.2.2 Feature Importance Difference Visualization

We utilized the SHAP beeswarm plot, a concise and information-rich visualization tool, to explore the influence of the top features in our dataset on the model's output. This plot provides valuable insights into the significance of these features about predicting in-hospital cardiac arrest (IHCA).

Figure 8 presents the feature impact on our IHCA prediction model using the 30-feature and 82-feature datasets. Notably, our observations highlight the substantial impact of heart rate measurements from different time windows on IHCA prediction. Specifically, the 5th and 6th time windows demonstrate notable contributions, suggesting that heart rate values or trends during these periods contain valuable information for accurate IHCA prediction. This emphasizes the importance of considering heart rate measurements at different time points when assessing the risk of IHCA.

doi:10.6342/NTU202302444

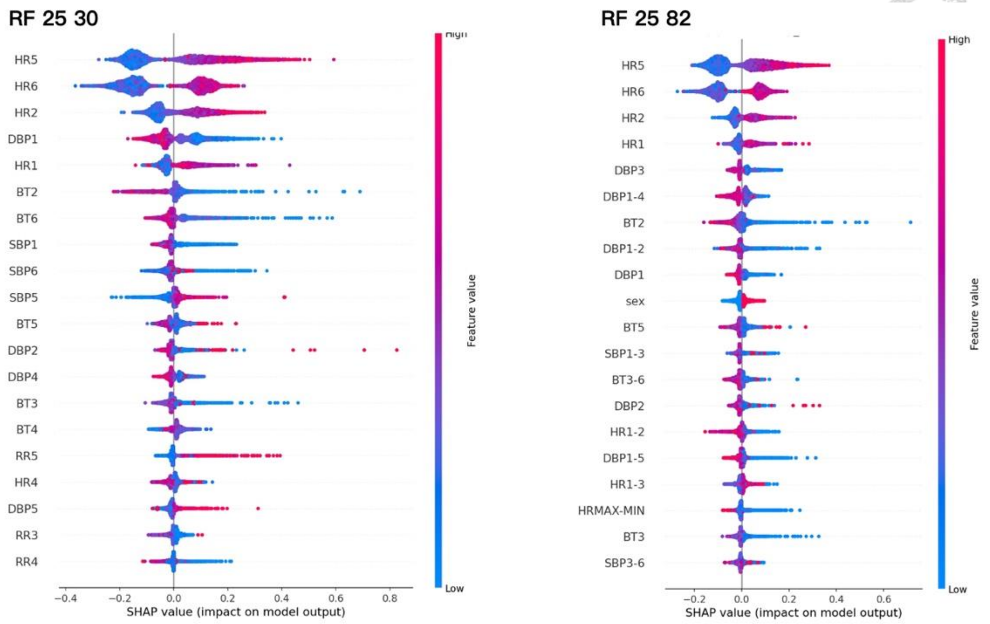


Figure 8. Feature importance by SHAP beeswarm plot, 30 features vs.82 features. (Random Forest with 1:25 class weight)

Furthermore, the order of feature importance remains consistent between the 30-feature set and the augmented 82-feature set. This consistency implies that the additional features do not significantly alter the relative importance of the top features of IHCA prediction.

In addition, the SHAP plot reveals that features related to diastolic blood pressure (DBP) show a consistent impact direction. Specifically, the decreases in DBP1-4, DBP1-2, and DBP1-5 are associated with an increased likelihood of IHCA. This observation suggests that a drop in diastolic pressure may contribute to IHCA.

Overall, the SHAP beeswarm plot provides valuable insights into the relationship between specific features and IHCA prediction, highlighting the critical role of heart

doi:10.6342/NTU202302444

rate measurements from different time windows and the consistent impact of decreased diastolic blood pressure on IHCA.

5.2.3 Feature Impact Amplitude and Distribution Visualization

We employed SHAP bar and beeswarm plots to visually depict the feature order and impact values for predicting in-hospital cardiac arrest (IHCA), providing valuable insights into the significance of different features in IHCA prediction.

Figure 9 presents the feature impact amplitude and distribution in our IHCA prediction model using the 82-feature datasets. Our analysis shows that heart rate, body temperature, and changes in diastolic blood pressure rank among the top 10 features that significantly contribute to IHCA prediction. These findings indicate the importance of these features in identifying patients at risk of IHCA. Heart rate is the most prominent among these influential features, exerting the most substantial impact on IHCA prediction. This underscores the crucial role of heart rate as a key indicator in assessing the likelihood of IHCA.

doi:10.6342/NTU202302444

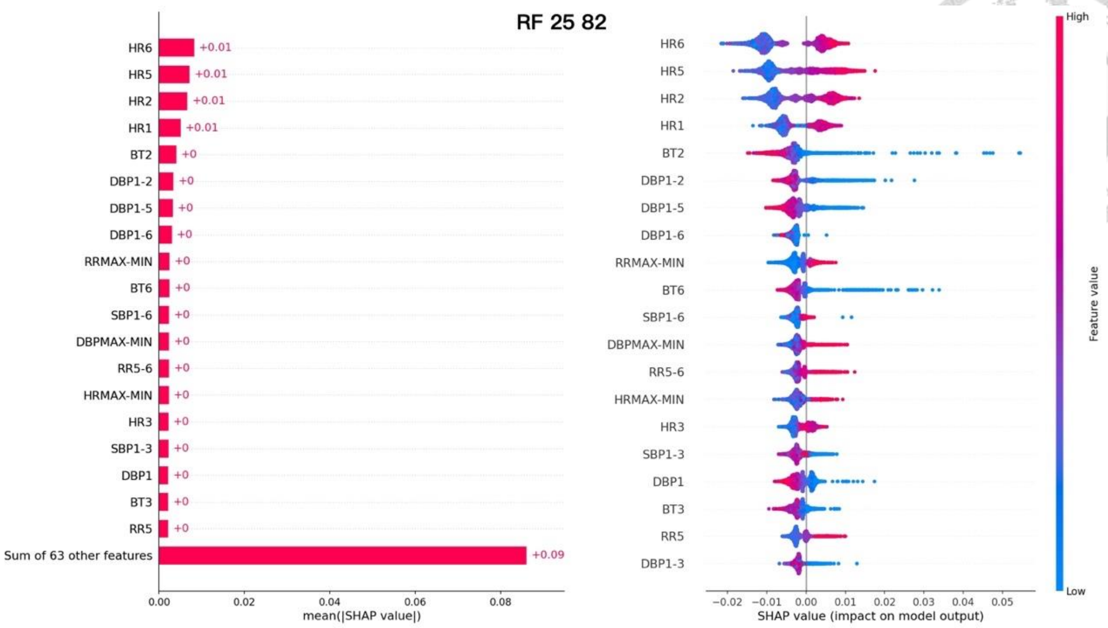


Figure 9. Feature importance by SHAP bar plot and beeswarm plot, 82 features. (Random Forest with 1:25 class weight)

doi:10.6342/NTU202302444

5.2.4 Feature Impact Distribution Visualization

We use the SHAP aggregated force plot to provide insights into the impact of feature values on the prediction of in-hospital cardiac arrest (IHCA). By examining the plot in our study, we can determine the cutoff points for each feature, indicating the threshold at which they significantly contribute to IHCA prediction.

Figure 10 in our study showcases the cutoff points observed for HR5 and HR6 at 80 beats per minute. This finding suggests that tachycardia, characterized by an elevated heart rate, may significantly influence the prediction of in-hospital cardiac arrest (IHCA). Figure 11, on the other hand, demonstrates that DBP1-2 and DBP1-5 exhibit a cutoff point at -10 mmHg. Crossing this threshold, a decrease in diastolic blood pressure indicates an increased probability of IHCA prediction. Furthermore, BT2 reveals a cutoff point at 35.7 degrees Celsius, meaning that a low body temperature preceding the assessment day can potentially impact IHCA prediction. These observations provide important insights into the relationship between these vital sign indicators and the likelihood of IHCA.

doi:10.6342/NTU202302444

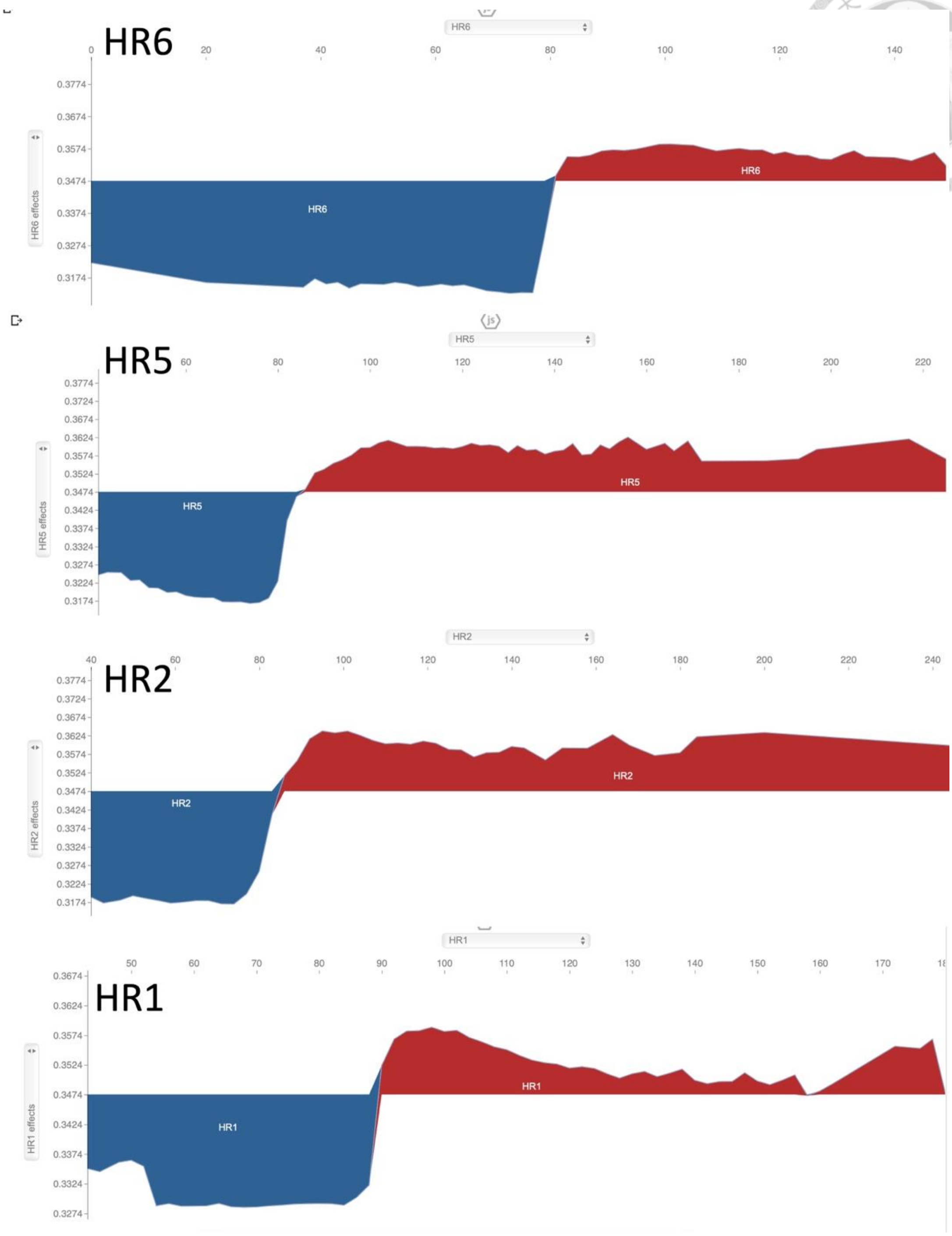


Figure 10. SHAP aggregated force plot of Heart rate of the different time windows. (Random Forest with 1:25 class weight and 82 features)

doi:10.6342/NTU202302444

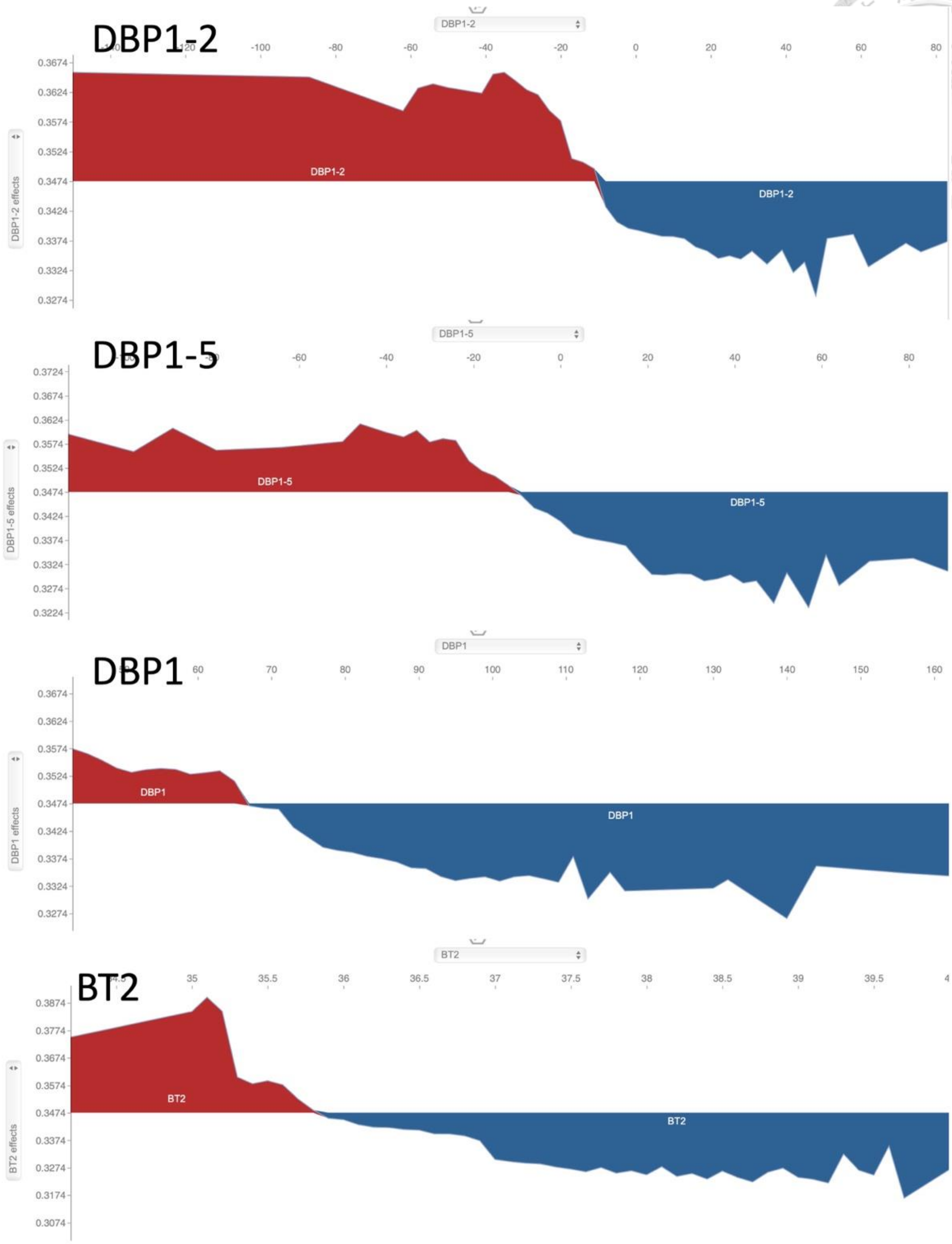


Figure 11. SHAP Aggregated Force plot of DBP difference and DBP1 and BT2 of the different time windows. (Random Forest with 1:25 class weight and 82 features)

It is important to note that while individual features have cutoff points indicating their impact, the overall contribution of a single feature remains minimal in the context of

doi:10.6342/NTU202302444

IHCA prediction. The SHAP Aggregated Force plot provides a comprehensive understanding of the relationship between feature values and their impact on IHCA prediction. It highlights the significance of multiple features working in combination rather than relying on a single feature alone.

5.2.5 Local Feature Importance Visualization

The SHAP force plot, a powerful analytical tool, allows for a detailed examination of the local impact of individual features on a single case. We gain valuable insights into the factors by considering a specific case from the positive group correctly predicted as positive. In Figure 12, we observe the varying importance of features between the 30-feature and 82-feature models. Notably, the 82-feature model highlights the discrepancy in diastolic blood pressure as the primary determinant of IHCA prediction. Despite DBP1-5 being ranked sixth in global importance, it substantially influences the local prediction outcome. We acquire additional information that enhances our understanding of the prediction results by employing the SHAP force plot in conjunction with the 82-feature model.

doi:10.6342/NTU202302444

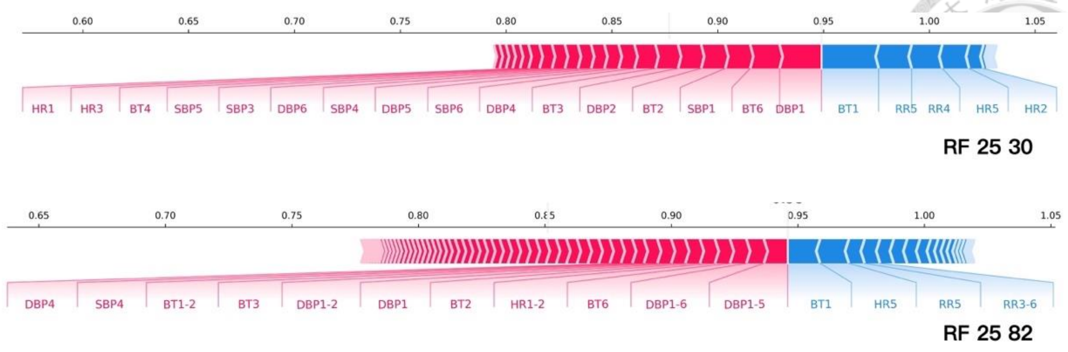


Figure 12. Feature importance by SHAP force plot, 30 vs. 82 features for a single case. (Random Forest with 1:25 class weight)

By employing the SHAP force plot and the augmented feature set, we gain deeper insights into the specific factors driving the IHCA prediction in this positive case. This comprehensive analysis enhances our understanding of the local impact of individual features and their contribution to the accurate prediction of IHCA.

doi:10.6342/NTU202302444

Chapter 6 Discussion

6.1 TEWS (Time-Series Warning Score)

Our initial investigation aimed to exploit vital signs data over a forty-eight-hour window to predict the incidence of cardiac arrest[48]. Following rigorous examination, we substantiated the superior efficacy of the Time-Series Early Warning Score (TEWS) model, which utilizes features extracted from six distinct time windows (TWs), in contrast to alternative classification algorithms. Table 8 and Table 9 provide a comprehensive synopsis of the performance of various classifiers across divergent datasets.

Our results unveil that applying the TEWS model incorporating features derived from six TWs yielded an impressive predictive prowess, evidenced by an AUROC value of 0.808 and an AUPRC value of 0.052. These outcomes surpass those achieved when employing features from a single TW (AUROC = 0.688, AUPRC = 0.041) and the Modified Early Warning Score (MEWS) (AUROC = 0.649, AUPRC = 0.015). Such compelling evidence reinforces that including vital signs data from multiple TWs offers invaluable insights for accurately predicting cardiac arrest.

doi:10.6342/NTU202302444

Table 8. AUROC values for classifiers with one, three, and six time windows, *TW* time window, *TEWS* time-series early warning score.

ALGORITHM	1TW	3TW	6TW
ADABOOST	0.6195	0.4131	0.4765
DECISION TREE_C4.5	0.5110	0.4963	0.4963
CART	0.5112	0.4960	0.4980
KNN	0.5014	0.5426	0.5215
GAUSSIAN NAÏVE BAYES	0.6147	0.3976	0.5381
MEWS	0.6492	0.6492	0.6492
LOGISTIC REGRESSION	0.7213	0.7297	0.7281
RANDOM FOREST	0.6024	0.6761	0.7327
SVM	0.7292	0.3656	0.7469
GRADIENT BOOSTING	0.7122	0.5199	0.7678
TEWS	0.6883	0.7621	0.8080

Table 9. AUPRC values for classifiers with the first, the third, and the sixth times windows, *TW* time window, *TEWS* time-series early warning score.

ALGORITHM	1TW	3TW	6TW

doi:10.6342/NTU202302444

GRADIENT BOOSTING	0.0448	0.0092	0.0073
DECISION TREE_C4.5	0.0266	0.0073	0.0073
CART	0.0073	0.0073	0.0073
KNN	0.0073	0.0102	0.0088
ADABOOST	0.0265	0.0069	0.0096
SVM	0.0073	0.0068	0.0108
RANDOM FOREST	0.0277	0.0160	0.0137
GAUSSIAN NAÏVE BAYES	0.0266	0.0065	0.0142
MEWS	0.0153	0.0153	0.0153
LOGISTIC REGRESSION	0.0421	0.0278	0.0306
TEWS	0.0412	0.0210	0.0519

Previous studies have similarly designated respiratory rate, heart rate, age, and systolic blood pressure as pivotal predictors of clinical deterioration[8]. Our investigation introduces a TEWS model that utilizes merely five features from six TWs: respiratory rate, systolic blood pressure within the most recent TW, and three heart rate readings from distinct TWs. This strategy yields a notable AUROC of 0.875 and AUPRC of 0.087, surpassing other classification algorithms. These outcomes emphasize

doi:10.6342/NTU202302444

introducing analyzing trends in heart rate variation, rather than solely absolute heart rate values, to enhance prediction accuracy.

Our investigation showcases several advantages compared to prior research. Firstly, while specific deep learning-based early warning systems are proficient at predicting patient deterioration, particularly within intensive care settings, our TEWS model has broader applicability, including general wards and long-term care facilities. Secondly, we embraced a longer observation window of 48 hours for vital signs, paired with a deep learning-based approach, to amplify the precision of cardiac arrest prediction without additional variables. Lastly, our model exclusively employs vital sign data, making it universally compatible with any system configured for Modified Early Warning Score (MEWS) deployment. The implementation of TEWS merely necessitates a personal computer equipped for either manual vital signs input or automatic extraction from electronic health records (EHRs).

Despite these advantages, our investigation bears several limitations. Firstly, it was conducted at a solitary community general hospital, potentially curtailing the broad applicability of our results to varied healthcare settings. Secondly, despite the superior performance of our TEWS model when utilizing vital signs data across 48 hours, its predictive capacity on the first day of admission did not eclipse that of other early

doi:10.6342/NTU202302444

warning systems. However, wearable devices collecting prehospital heart rate data might be an alternative data source to enhance the model's performance. Lastly, our model demonstrated shortcomings in precisely predicting some instances of cardiac arrest within our dataset, particularly those characterized by sudden collapses such as pulmonary embolism following cesarean section or postoperative airway obstruction with hematoma. Moreover, the TEWS model is unable to detect deterioration between two time windows, which points to an inherent limitation of noncontinuous vital signs-based prediction models.

In summation, our investigation effectively exhibits the efficacy of the TEWS model in predicting cardiac arrest by leveraging vital signs data. Incorporating multiple time windows and emphasizing trend analysis of heart rate substantially enhance the model's performance relative to other classification algorithms. While our investigation offers invaluable insights, additional research is needed to validate these findings across varied healthcare settings and explore integrating alternative data sources to further enhance predictive capabilities.

doi:10.6342/NTU202302444

6.2 TEWS-X (Explainable Time-Series Early Warning Score)

Our primary objective in this study was to utilize vital signs data spanning two days to predict the occurrence of cardiac arrest accurately. Additionally, we aimed to develop a prediction model for in-hospital cardiac arrest (IHCA) that is both straightforward and explainable based on our initial research findings.

We opted for a tree-based machine learning method instead of deep learning to achieve explainability. Although this approach does not fully capture the time-series nature of the data like recurrent neural networks (RNNs) do, we preserved some time-series factors in our tree-based model by considering the sign of feature differences along the timeline. This enables us to retain certain aspects of the vital sign trends and their impact on IHCA prediction.

We employed an augmented dataset and adjusted the class weights to perform similarly to RNNs to compensate for the loss of time-series modeling capabilities. By incorporating techniques such as SHAP (SHapley Additive exPlanations) values, we could generate explanations for each prediction made by our model. These explanations provide insights into the key contributing factors and their relative importance in assessing the patient's risk of IHCA. This transparent and interpretable approach empowers healthcare providers to comprehend the rationale behind the model's

doi:10.6342/NTU202302444

predictions, thereby increasing their confidence in leveraging the model's outputs for clinical decision-making.

We utilized SHAP bar plots and beeswarm plots to ascertain the global feature importance, offering valuable insights into the model's inner workings. This enables us to employ these models for IHCA prediction and allocate additional attention to cases where the predictions align reasonably well. Furthermore, the SHAP force plot provides us with local feature importance, aiding in identifying the direction of potential treatment strategies.

Given that vital signs are the foundation of our prediction model, the local SHAP impact values for individual cases can be effectively described on the TPR (Temperature, Pulse, Respiration) sheet, which records daily vital signs. As part of the routine care process, when the care team records the most recent vital signs, this data can be seamlessly transmitted to the TEWS-X service and promptly displayed on the TPR sheet page. By incorporating an alarm system into the daily care routine, modifying or disrupting existing care processes is unnecessary.

Integrating the TEWS-X service into the TPR sheet allows for real-time monitoring and evaluation of patients' vital signs within their natural care setting. The SHAP impact values offer valuable insights into the influence of specific vital sign measurements on

doi:10.6342/NTU202302444

predicting in-hospital cardiac arrest (IHCA). By providing this information on the TPR sheet, the care team can easily interpret the significance of each vital sign measurement and identify any concerning trends or patterns.

The automatic display of the TEWS-X results on the TPR sheet serves as an additional layer of support for the care team, facilitating early detection of patients at risk of IHCA. This seamless integration ensures that the alarm system becomes integral to the daily care routine, allowing healthcare providers to respond promptly and appropriately when necessary.

Importantly, incorporating the TEWS-X service and alarm system does not necessitate any changes to the existing care processes. It seamlessly integrates into the routine documentation of vital signs, ensuring that healthcare providers can continue their daily tasks without disruption. By enhancing the TPR sheet with the predictive capabilities of the TEWS-X service, healthcare teams can optimize patient care and improve outcomes without compromising the established care workflow.

In summary, when integrated with the TPR sheet, the TEWS-X service can provide a user-friendly platform for displaying vital sign data and SHAP impact values. This integration allows for continuous monitoring and early detection of patients at risk of IHCA within the existing care routine. By incorporating these features, we enhance

doi:10.6342/NTU202302444

patient safety and facilitate informed decision-making by the care team, all without the need for any changes to the daily care process.

doi:10.6342/NTU202302444

6.3 Implementation of TEWS/TEWS-X

In-hospital cardiac arrest (IHCA) can be categorized into predictable and unpredictable. Predictable IHCA refers to cases with identifiable patterns or similarities to previous instances of IHCA. In these cases, valuable insights can be gleaned from previous experiences and medical knowledge, aiding in anticipating and managing such events. Common symptoms of predictable IHCA include chest tightness with dyspnea, hypotension with tachycardia, or alterations in consciousness. Recognizing these indicators allows for timely intervention and the allocation of appropriate resources, such as immediate transfer to the intensive care unit (ICU), to prevent further deterioration.

To guide the prioritization of ICU admission for predictable IHCA cases, specific ICU admission criteria have been developed[49]. These criteria serve as guidelines to ensure that patients with a higher likelihood of deterioration receive the necessary intensive care promptly. However, it is essential to note that the decision for ICU admission can be influenced by factors such as the availability of ICU beds and the awareness and judgment of the care team[50]. Limited ICU bed capacity or variations in individual clinical assessment may impact the decision-making process, potentially affecting the timeliness of ICU admission for those at risk of IHCA.

doi:10.6342/NTU202302444

While proactive measures can be taken to identify patients at risk of in-hospital cardiac arrest (IHCA) and transfer them to the intensive care unit (ICU) for close monitoring and intervention, it is essential to acknowledge that certain cases of IHCA remain unpredictable. These instances encompass rare and unforeseen events such as anaphylactic shock, acute post-operative hemorrhage, and amniotic fluid embolism. Unlike predictable IHCA cases, these events occur in advance without significant warning signs or evident indicators.

Figure 13 showcases the outcomes observed in general ward settings during routine daily care. Our primary objective is to identify individuals requiring heightened attention within the existing care processes and available resources. Through our research, we have developed an automated alert system utilizing vital sign data obtained during routine care. This method holds promise for implementation in various healthcare settings, including long-term care units, low-staffed care units, home healthcare, telemedicine services, and medical facilities in remote areas. The application of our method provides a compelling rationale for transferring patients to evacuation hospitals when necessary.

doi:10.6342/NTU202302444

By leveraging the power of technology and utilizing vital sign data, our approach enables the timely identification of patients at risk of adverse events in general ward settings. This proactive alert system serves as an additional layer of safety, ensuring that patients receive appropriate monitoring, interventions, and care based on their individual needs. Implementing our method in different healthcare settings, particularly those with limited resources or geographical constraints, can enhance patient outcomes and improve the overall quality of care.

The automatic alert system generated by our method serves as a valuable tool for healthcare providers, allowing them to respond to critical situations and allocate resources effectively and promptly. By identifying individuals who require closer monitoring or specialized interventions, our method aids in optimizing patient care and facilitating timely transfers to the evacuation unit when warranted.

Integrating our approach into routine care processes can significantly enhance patient safety and improve healthcare outcomes. By harnessing the capabilities of technology, we can proactively identify individuals in need of attention, ensure timely interventions, and ultimately improve patient outcomes in various healthcare settings.

doi:10.6342/NTU202302444

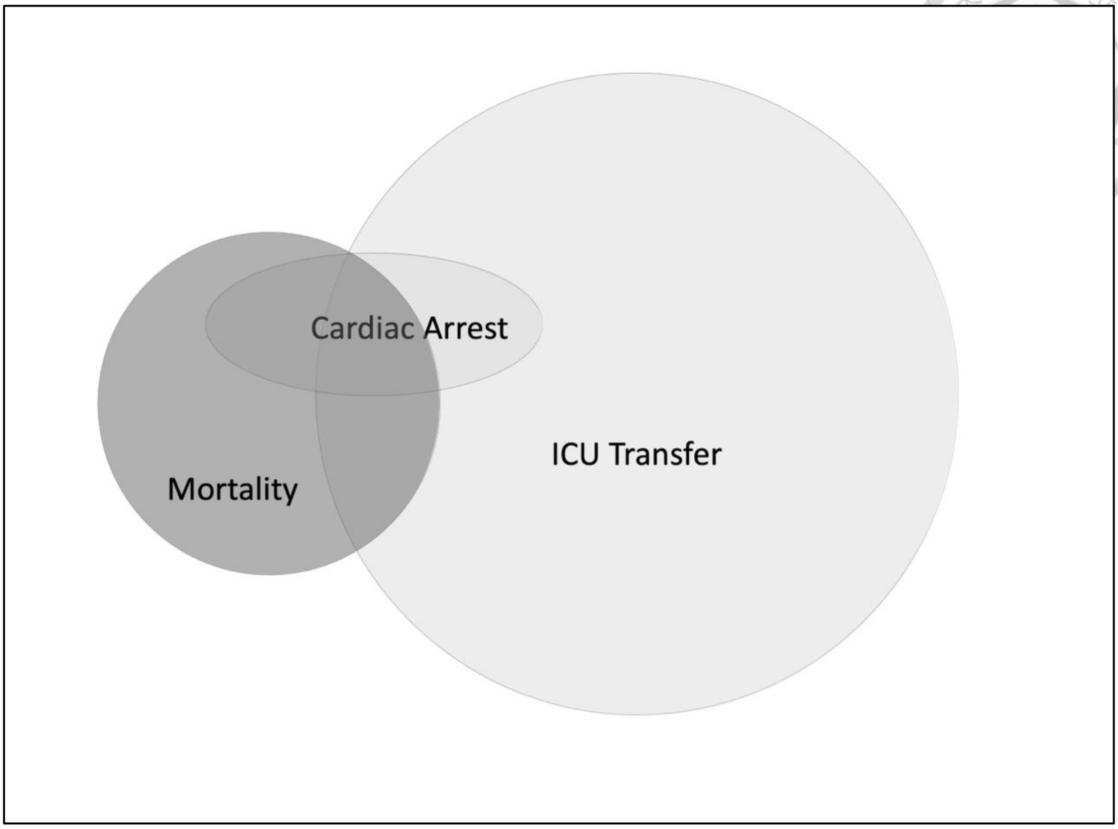


Figure 13.Venn Diagram Illustrating Outcome Events in General Wards.

doi:10.6342/NTU202302444

Chapter 7 Conclusion

Currently, healthcare providers are emphasizing disease treatment and symptom improvement to optimize patient outcomes. However, allocating healthcare resources requires a balanced approach, ensuring reasonable care for all patients rather than universal intensive care. Consequently, there is a possibility of unexpected cardiac arrest cases that may go unnoticed in general wards.

Our primary objective is not to identify every potential cardiac arrest patient but to identify individuals who may be overlooked within the existing care processes and available resources. We strive to identify specific characteristics of patients who have experienced in-hospital cardiac arrest (IHCA) but were not initially identified as high-risk individuals. Our ultimate goal is to intervene and provide additional care to these potential patients, reducing unforeseen cardiac arrest events.

Recognizing resource allocation limitations, healthcare providers must prioritize patients based on their clinical condition and allocate intensive care resources accordingly. This approach ensures that patients with more severe conditions receive the required attention and care, including bedside monitoring or transfer to the intensive care unit (ICU) to prevent further deterioration.

doi:10.6342/NTU202302444

We aim to supplement the existing care processes by identifying patients who have experienced IHCA despite not being initially flagged as high-risk individuals. This approach allows us to implement targeted interventions and provide additional care, reducing the likelihood of unexpected cardiac arrest events. We can achieve these goals by optimizing resource allocation and improving patient outcomes without significantly disrupting the current care infrastructure.

In conclusion, our TEWS and TEWS-X models represent noteworthy advancements in IHCA prediction and comprehensibility. By harnessing vital signs data and incorporating explicable modeling techniques, these models empower healthcare providers to proactively identify individuals at risk of IHCA and intervene promptly and purposefully. Early identification and timely intervention are pivotal in diminishing IHCA mortality rates and ameliorating patient outcomes. Further research is warranted to validate the models across diverse healthcare settings and explore supplementary data sources for enhanced predictive capabilities. Implementing the TEWS and TEWS-X models can revolutionize IHCA management by providing healthcare providers with valuable tools to enhance patient care and optimize resource allocation. The integration of these models into clinical practice can heighten patient safety, curtail cardiac arrest incidents, and ultimately save lives.

doi:10.6342/NTU202302444

Reference

[1] G. D. Perkins *et al.*, "Cardiac arrest and cardiopulmonary resuscitation outcome reports: update of the Utstein Resuscitation Registry Templates for Out-of-Hospital Cardiac Arrest: a statement for healthcare professionals from a task force of the International Liaison Committee on Resuscitation (American Heart Association, European Resuscitation Council, Australian and New Zealand Council on Resuscitation, Heart and Stroke Foundation of Canada, InterAmerican Heart Foundation, Resuscitation Council of Southern Africa, Resuscitation Council of Asia); and the American Heart Association Emergency Cardiovascular Care Committee and the Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation," (in eng), *Circulation,* vol. 132, no. 13, pp. 1286-300, Sep 29 2015, doi: https://doi.org/10.1161/cir.0000000000000144.

[2] R. M. Merchant *et al.*, "Incidence of treated cardiac arrest in hospitalized patients in the United States," (in eng), *Critical care medicine,* vol. 39, no. 11, pp. 2401-6, Nov 2011, doi: https://doi.org/10.1097/CCM.0b013e3182257459.

[3] L. W. Andersen, M. J. Holmberg, K. M. Berg, M. W. Donnino, and A. Granfeldt, "In-Hospital Cardiac Arrest: A Review," *Jama,* vol. 321, no. 12, pp. 1200-1210, Mar 26 2019, doi: https://doi.org/10.1001/jama.2019.1696.

[4] L. M. Chen, B. K. Nallamothu, J. A. Spertus, Y. Li, and P. S. Chan, "Association between a hospital's rate of cardiac arrest incidence and cardiac arrest survival," (in eng), *JAMA internal medicine,* vol. 173, no. 13, pp. 1186-95, Jul 8 2013, doi: https://doi.org/10.1001/jamainternmed.2013.1026.

[5] M. Schluep, B. Y. Gravesteijn, R. J. Stolker, H. Endeman, and S. E. Hoeks, "One-year survival after in-hospital cardiac arrest: A systematic review and meta-analysis," *Resuscitation,* vol. 132, pp. 90-100, Nov 2018, doi: https://doi.org/10.1016/j.resuscitation.2018.09.001.

[6] S. M. Fernando *et al.*, "Pre-arrest and intra-arrest prognostic factors associated with survival after in-hospital cardiac arrest: systematic review and meta-analysis," *BMJ (Clinical research ed.),* vol. 367, p. l6373, Dec 4 2019, doi: https://doi.org/10.1136/bmj.l6373.

[7] 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. 521-6, Oct 2001, doi: https://doi.org/10.1093/qjmed/94.10.521.

[8] 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

doi:10.6342/NTU202302444

Conventional Regression for Predicting Clinical Deterioration on the Wards," *Critical care medicine,* vol. 44, no. 2, pp. 368-74, Feb 2016, doi: https://doi.org/10.1097/CCM.0000000000001571.

[9] M. Green, H. Lander, A. Snyder, P. Hudson, M. Churpek, and D. Edelson, "Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients," (in eng), *Resuscitation,* vol. 123, pp. 86-91, Feb 2018, doi: https://doi.org/10.1016/j.resuscitation.2017.10.028.

[10] B. Bartkowiak *et al.*, "Validating the Electronic Cardiac Arrest Risk Triage (eCART) Score for Risk Stratification of Surgical Inpatients in the Postoperative Setting: Retrospective Cohort Study," (in eng), *Annals of surgery,* Jan 12 2018, doi: https://doi.org/10.1097/sla.0000000000002665.

[11] M. M. Churpek *et al.*, "Multicenter development and validation of a risk stratification tool for ward patients," *American journal of respiratory and critical care medicine,* vol. 190, no. 6, pp. 649-55, Sep 15 2014, doi: https://doi.org/10.1164/rccm.201406-1022OC.

[12] F. Sessa *et al.*, "Heart rate variability as predictive factor for sudden cardiac death," (in eng), *Aging (Albany NY),* vol. 10, no. 2, pp. 166-177, 2018, doi: https://doi.org/10.18632/aging.101386.

[13] N. Liu *et al.*, "An Intelligent Scoring System and Its Application to Cardiac Arrest Prediction," *IEEE Transactions on Information Technology in Biomedicine,* vol. 16, no. 6, pp. 1324-1331, 2012, doi: https://doi.org/10.1109/TITB.2012.2212448.

[14] M. A. Akel, K. A. Carey, C. J. Winslow, M. M. Churpek, and D. P. Edelson, "Less is more: Detecting clinical deterioration in the hospital with machine learning using only age, heart rate, and respiratory rate," (in eng), *Resuscitation,* vol. 168, pp. 6-10, Nov 2021, doi: https://doi.org/10.1016/j.resuscitation.2021.08.024.

[15] J. Kim, M. Chae, H. J. Chang, Y. A. Kim, and E. Park, "Predicting Cardiac Arrest and Respiratory Failure Using Feasible Artificial Intelligence with Simple Trajectories of Patient Data," (in eng), *J Clin Med,* vol. 8, no. 9, Aug 29 2019, doi: https://doi.org/10.3390/jcm8091336.

[16] K.-J. Cho *et al.*, "Detecting Patient Deterioration Using Artificial Intelligence in a Rapid Response System," (in eng), *Critical care medicine,* vol. 48, no. 4, pp. e285-e289, Apr 2020, doi: https://doi.org/10.1097/ccm.0000000000004236.

[17] T. J. Hodgetts *et al.*, "Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital," (in eng), *Resuscitation,* vol. 54, no. 2, pp. 115-23, Aug 2002, doi: https://doi.org/10.1016/s0300-9572(02)00098-9.

doi:10.6342/NTU202302444

[18] P. G. Lyons, D. P. Edelson, and M. M. Churpek, "Rapid response systems," (in eng), *Resuscitation*, vol. 128, pp. 191-197, Jul 2018, doi: https://doi.org/10.1016/j.resuscitation.2018.05.013.

[19] B. K. Nallamothu *et al.*, "How Do Resuscitation Teams at Top-Performing Hospitals for In-Hospital Cardiac Arrest Succeed? A Qualitative Study," (in eng), *Circulation*, vol. 138, no. 2, pp. 154-163, Jul 10 2018, doi: https://doi.org/10.1161/circulationaha.118.033674.

[20] M. E. Smith *et al.*, "Early warning system scores for clinical deterioration in hospitalized patients: a systematic review," *Annals of the American Thoracic Society*, vol. 11, no. 9, pp. 1454-65, Nov 2014, doi: https://doi.org/10.1513/AnnalsATS.201403-102OC.

[21] J. Ludikhuize, M. Borgert, J. Binnekade, C. Subbe, D. Dongelmans, and A. Goossens, "Standardized measurement of the Modified Early Warning Score results in enhanced implementation of a Rapid Response System: a quasi-experimental study," (in eng), *Resuscitation*, vol. 85, no. 5, pp. 676-82, May 2014, doi: https://doi.org/10.1016/j.resuscitation.2014.02.009.

[22] S. H. Kim *et al.*, "Predicting severe outcomes using national early warning score (NEWS) in patients identified by a rapid response system: a retrospective cohort study," *Scientific Reports*, vol. 11, no. 1, p. 18021, 2021/09/09 2021, doi: https://doi.org/10.1038/s41598-021-97121-w.

[23] M. M. Churpek, T. C. Yuen, S. Y. Park, R. Gibbons, and D. P. Edelson, "Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*," (in eng), *Critical care medicine*, vol. 42, no. 4, pp. 841-8, Apr 2014, doi: https://doi.org/10.1097/ccm.0000000000000038.

[24] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate Imputation by Chained Equations in R," *J Stat Softw*, vol. 45, no. 3, pp. 1-67, 12/12/ 2011, doi: https://doi.org/10.18637/jss.v045.i03.

[25] J. m. Kwon, Y. Lee, Y. Lee, S. Lee, and J. Park, "An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest," *Journal of the American Heart Association*, vol. 7, no. 13, Jun 26 2018, doi: https://doi.org/10.1161/JAHA.118.008678.

[26] H.-K. Chang *et al.*, "Early Detecting In-Hospital Cardiac Arrest Based on Machine Learning on Imbalanced Data," *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 1-10, 10-13 June 2019 2019, doi: https://doi.org/10.1109/ICHI.2019.8904504.

doi:10.6342/NTU202302444

[27] S. Hochreiter and J. r. Schmidhuber, "Long short-term memory," (in eng), *Neural Comput*, vol. 9, no. 8, pp. 1735-80, Nov 15 1997, doi: https://doi.org/10.1162/neco.1997.9.8.1735.

[28] C. Esteban, O. Staeck, S. Baier, Y. Yang, and V. Tresp, "Predicting Clinical Events by Combining Static and Dynamic Information Using Recurrent Neural Networks," *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 93-101, 4-7 Oct. 2016 2016, doi: https://doi.org/10.1109/ICHI.2016.16.

[29] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. null, pp. 2825–2830, 2011, doi: https://doi.org/10.48550/arXiv.1201.0490.

[30] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1-37, 2008/01/01 2008, doi: https://doi.org/10.1007/s10115-007-0114-2.

[31] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning (ICML '96)*, pp. 148–156, 1996, doi: https://dl.acm.org/doi/10.5555/3091696.3091715.

[32] R. E. Schapire, "A brief introduction to boosting," *Proceedings of the 16th international joint conference on Artificial intelligence - Volume 2*, pp. 1401–1406, 1999, doi: https://dl.acm.org/doi/10.5555/1624312.1624417.

[33] H. Tin Kam, "Random decision forests," *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, pp. 278-282 vol.1, 14-16 Aug. 1995 1995, doi: https://doi.org/10.1109/ICDAR.1995.598994.

[34] D. Böhning, "Multinomial logistic regression algorithm," *Ann Inst Stat Math*, vol. 44, no. 1, pp. 197-200, 1992/03/01 1992, doi: https://doi.org/10.1007/BF00048682.

[35] J. H. Friedman, "Stochastic gradient boosting," *Comput Stat Data Anal*, vol. 38, no. 4, pp. 367-378, 2002/02/28/ 2002, doi: https://doi.org/10.1016/S0167-9473(01)00065-2.

[36] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification And Regression Trees*. 2017, pp. 1-358.

[37] D. J. Hand and K. Yu, "Idiot's Bayes: Not So Stupid after All?," *Int Stat Rev*, vol. 69, no. 3, pp. 385-398, 2001, doi: https://doi.org/10.2307/1403452.

[38] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995/09/01 1995, doi: https://doi.org/10.1007/BF00994018.

doi:10.6342/NTU202302444

[39] S. Abe, *Support vector machines for pattern classification* (Advances in pattern recognition). London: Springer, 2005, pp. xiv, 343 p.

[40] E. Fix and J. L. Hodges, "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties," *Int Stat Rev*, vol. 57, no. 3, pp. 238-247, 1989, doi: https://doi.org/10.2307/1403797.

[41] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining (2nd Edition)*. Pearson, 2018.

[42] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Elsevier Science, 2014.

[43] L. Breiman, "Arcing the Edge Technical Report 486; Statistics Department," *University of California: Berkeley, CA, USA*, 1997.

[44] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput Electr Eng*, vol. 40, no. 1, pp. 16-28, 2014/01/01/ 2014, doi: https://doi.org/10.1016/j.compeleceng.2013.11.024.

[45] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," *2014 Science and Information Conference*, pp. 372-378, 27-29 Aug. 2014 2014, doi: https://doi.org/10.1109/SAI.2014.6918213.

[46] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017, doi: https://doi.org/10.48550/arXiv.1705.07874.

[47] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56-67, 2020/01/01 2020, doi: https://doi.org/10.1038/s42256-019-0138-9.

[48] C.-F. Su, S.-I. Chiu, J.-S. R. Jang, and F. Lai, "Improved inpatient deterioration detection in general wards by using time-series vital signs," *Scientific Reports*, vol. 12, no. 1, p. 11901, 2022/07/13 2022, doi: http://doi.org/10.1038/s41598-022-16195-2.

[49] J. D. Chalmers *et al.*, "Severity assessment tools to guide ICU admission in community-acquired pneumonia: systematic review and meta-analysis," *Intensive Care Medicine*, vol. 37, no. 9, pp. 1409-1420, 2011/09/01 2011, doi: https://doi.org/10.1007/s00134-011-2261-x.

[50] R. Robert *et al.*, "Influence of ICU-bed availability on ICU admission decisions," *Annals of Intensive Care*, vol. 5, no. 1, p. 55, 2015/12/30 2015, doi: https://doi.org/10.1186/s13613-015-0099-z.

doi:10.6342/NTU202302444