Applied Medical Informatics
Original Research Vol. 46, Suppl. 2 /2024, pp: S5-S8

Is it Beneficial to Use Different Thresholds Over Time for Early Prediction Model?

Sungsoo HONG[1,*], Hyunwoo CHOO[1], Kyung Hyun LEE[1], Sungjun HONG[2], Ki-Byung LEE[1,3], and Chang Youl LEE[3]

[1] AITRICS, Inc., 218 Teheran-ro, Gangnam-gu, 06221 Seoul, Republic of Korea
[2] Medical AI Research Center, Research Institute for Future Medicine, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, 06351 Seoul, Republic of Korea
[3] Division of Pulmonary, Allergy and Critical Care Medicine, Hallym University Chuncheon Sacred Heart Hospital, 77 Sakju-ro, 24253 Chuncheon, Republic of Korea
E-mails: sshong@aitrics.com; hwchoo@aitrics.com; lkh256@aitrics.com; hsj2864@skku.edu; hasej@aitrics.com; doclcy@hallym.or.kr

* Author to whom correspondence should be addressed

Received: 12 July 2024/Accepted: 11 September 2024/ Published online: 21 November 2024

Abstract
In production settings, deep learning models often rely on fixed thresholds. This study investigates whether using varying thresholds over time enhances predictive accuracy and clinical utility, especially for early sepsis prediction. We retrospectively analyzed EMR data from Hallym University Chuncheon Sacred Heart Hospital (2018-2022), excluding patients aged under 18 or without vital signs. Utilizing the AITRICS-VC SEPS deep learning model, which predicts sepsis using six vital signs, eleven lab results and patient information, we examined prediction thresholds at one-hour intervals before sepsis onset. Optimal thresholds for each interval were identified using the Youden index. Net benefit and decision curve analysis compared the performance of time-varying versus global thresholds. Results show interval-specific thresholds yield higher net benefits and increased true positive detections: 456 (0-1 hour), 122 (1-2 hours), 41 (2-3 hours), and 29 (3-4 hours) before sepsis onset. This suggests dynamically adjusting thresholds over time can improve early sepsis detection and patient outcomes.

Keywords: Deep learning; Early prediction; Sepsis; Threshold adjustment; Net benefit

Introduction

It is common to use the area under the receiver operating characteristic curve to present or evaluate the performance of deep learning models. However, in production settings, a specific threshold is typically set to evaluate performance. This raises the question: is this the best convention? For time-series prediction models, such as those used for early prediction of sepsis, it may be beneficial to use different thresholds over time. The model's prediction values may be lower when far from the onset time and relatively higher when close to the onset time. In this study, we explore whether using different thresholds over time is beneficial. We apply the concepts of net benefit and decision curve analysis to evaluate this approach.

Materials and Methods

Study Design

Our objective is to demonstrate the benefits of applying different thresholds over time in a deep learning model focused on the early prediction of sepsis onset. We hypothesize that the optimal threshold varies over time intervals before the onset, which can be particularly beneficial for early detection and intervention.

To determine if the optimal threshold varies over the specified time intervals, we split each episode into the intervals of 0-1 hour, 1-2 hours, 2-3 hours, and 3-4 hours before the onset. For negative episodes (those without an onset time), we used the timestamp of the last observation recorded instead. The Youden index, defined as the maximum of the sum of sensitivity and specificity scaled by subtracting 1, is a useful for finding the optimal threshold considering the balance. It is used to identify the optimal threshold for each interval [1].

Data Collection and Processing

We retrospectively collected EMR data in Hallym University Chuncheon Sacred Heart Hospital in period of from 2018-01-01 to 2022-12-31. Patients who are below 18 years of age or who have no vital signs reported were excluded. Collected data was refined into episode per each patient. Episodes were segmented into one-hour intervals relative to the sepsis onset time.

Deep Learning Model for Early Prediction of Sepsis

In this study, we utilized a deep learning model called AITRICS-VC SEPS, which has been approved by the Korea's Ministry of Food and Drug Safety as a medical AI solution that predicts future sepsis occurrence [2]. The model gets six vital sign, eleven lab results, and patient information, refer to the past in time-series manner, and then returns future sepsis risk probability.

Net Benefit and Decision Curve Analysis

To evaluate whether a model does more good than harm when used in clinical practice, we utilize the concept of net benefit [3-4]. Net benefit provides a measure that balances the true positive outcomes against the false positive outcomes, adjusted by the odds of a given threshold probability. The formula for net benefit is defined as:

$$Net\ benefit = \frac{True\ Positive}{N} - \frac{False\ Positive}{N} \times \frac{p_t}{1 - p_t} \quad (1)$$

where N is the total sample size and p_t is a threshold probability to define when a patient is positive. The number of true positives and false positives are determined based on the threshold probability p_t. Additionally, Standardized net benefit is more interpretable when compared with net benefits, and the outcome prevalence of each group is used to weight.

The decision curve is plotted with threshold probabilities on the x-axis and their corresponding net benefit values on the y-axis.

Results

Demographic Information

During the period, we collected 46,842 patients. Of which 22,317 (47.6%) were female and 24,525 were male (52.4%). Mean age of all patients was 58.2 (Standard deviation 18.3). The prevalence of sepsis was 6.7% (n=3,120).

Decision Curve Analysis

Figure 1 illustrates the decision curves for interval-group-specific cutoffs alongside the global cutoff. For all interval groups, the optimal thresholds calculated using the Youden index correspond to higher net benefit values, compared to those derived from the global cutoff.

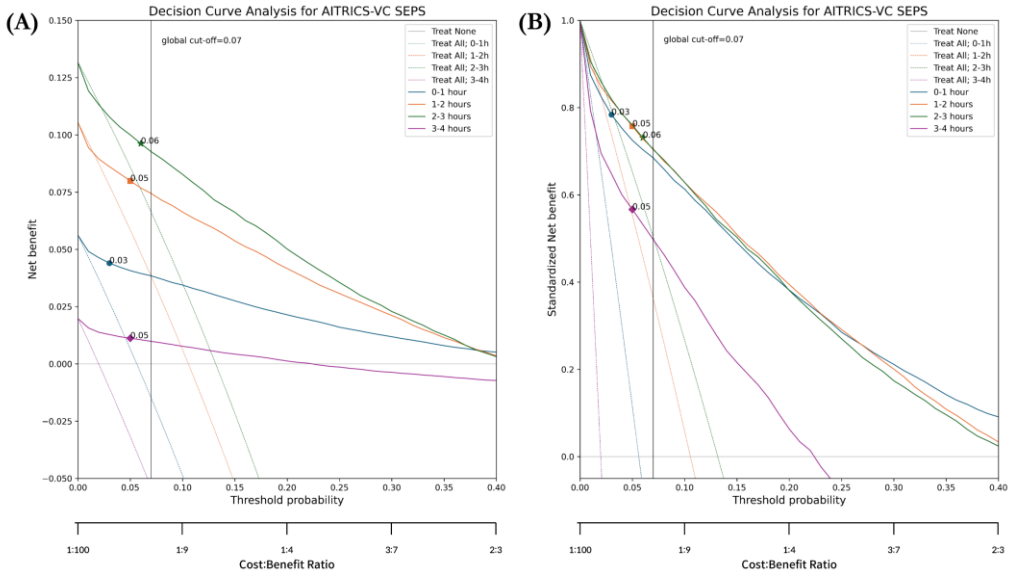

Figure 1. (A) Net benefit curves for each interval group. **(B)** Standardized net benefit curves. The global threshold (0.07) is represented by a vertical line. Optimal thresholds for each benefit line are indicated by numbers and distinct markers. The dashed line represents a policy of treating all patients regardless of their score.

Group-wise optimal thresholds enhanced the detection of true positives across time intervals. Increases were observed in the number of true positives: 456 (9.4%) within the 0-1 hour interval, 122 (4.1%) within 1-2 hours, 41 (1.8%) within 2-3 hours, and 29 (3.1%) within 3-4 hours.

Discussion

In this study, we explored that using varying threshold over time would be beneficial than using a single global threshold when deploying a deep learning model, especially for early sepsis prediction. To achieve this, we utilized the net benefit concept and decision curve analysis.

Our retrospective study highlights the benefits of using varying thresholds. However, a key limitation is the reliance on known onset times, which allows for grouping data by hour-based intervals. In real-world settings, the onset time cannot be predicted in advance. To address this, we suggest including informative messages when alarms are triggered using a global cutoff, indicating the observed findings.

Conclusions

This study demonstrates that applying time-varying thresholds in early sepsis prediction models significantly enhances detection accuracy and clinical decision-making benefits. The improved net benefits and increased true positive rates highlight the potential for this approach to be adopted in clinical settings, offering a more responsive and precise method for early intervention.

List of Abbreviations: Not applicable.

Author Contributions: SSH, HC defined the research's aim and the experiments' design. SJH participated in the design of the study. SSH carried out the experiments and performed the statistical analysis. SSH makes figure. HC and SSH write the manuscript. SSH and KHL collected the data and clean them. KHL, SJH critically reviewed the draft and helped manuscript to be completed.

Funding: This research received no funding.

Ethics Statement: This study is approved by the institute review board of Hallym University Chuncheon Sacred Heart Hospital (IRB No: CHUNCHEON 2023-03-007-002). The informed consent is waived as this is a retrospective study that has minimal risk to patients.

Data Availability Statement: The data utilized in this study were obtained under the supervision and with the grant support of Hallym Chuncheon Sacred Heart Hospital. For access to the data, individual contact with the hospital is required.

Conflict of Interest: HC, SSH, KHL and KBL are employee of AITRICS.

References

1. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J. 2005;47(4):458-472. doi: 10.1002/bimj.200410135.
2. Sung M, Hahn S, Han CH, Lee JM, Lee J, Yoo J, et al. Event Prediction Model Considering Time and Input Error Using Electronic Medical Records in the Intensive Care Unit: Retrospective Study. JMIR Med Inform. 2021;9(11):e26426. doi: 10.2196/26426.
3. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-574. doi: 10.1177/0272989X06295361.
4. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. J Clin Oncol. 2016;34(21):2534-2540. doi: 10.1200/JCO.2015.65.5654.