Deep Learning in the Medical Domain: Predicting Cardiac Arrest Using Deep Learning

Youngnam Lee[1]*, Joon-myoung Kwon[2]*, Yeha Lee[1], Hyunho Park[1], Hugh Cho[1], Jinsik Park[2]

[1]VUNO, Seoul; [2]Department of Emergency Medicine, Mediplex Sejong Hospital, Incheon, Korea

With the wider adoption of electronic health records, the rapid response team initially believed that mortalities could be significantly reduced but due to low accuracy and false alarms, the healthcare system is currently fraught with many challenges. Rule-based methods (e.g., Modified Early Warning Score) and machine learning (e.g., random forest) were proposed as a solution but not effective. In this article, we introduce the DeepEWS (Deep learning based Early Warning Score), which is based on a novel deep learning algorithm. Relative to the standard of care and current solutions in the marketplace, there is high accuracy, and in the clinical setting even when we consider the number of alarms, the accuracy levels are superior.

Key Words: artificial intelligence; cardiac arrest; deep learning; rapid response team

INTRODUCTION

In 2009, the Health Information Technology for Economic and Clinical Health Act was implemented to roll out electronic health records (EHRs) and it was rapidly adopted. A total of 30 million dollars was invested into this policy that incorporated incentives and penalties for widely adopting EHRs. EHRs include demographic information, present and past diagnoses, laboratory results, vital sign, and clinical notes to efficiently manage patients in the clinical setting and can leverage clinical informatics to predict diseases, patient length of stay, and re-admission rates.

Due to the variety of clinical information, providers can quickly access through EHRs; a rapid response team (RRT) can effectively reduce the number of in-hospital cardiac arrest and mortalities. However, with the current standard of care (Modified Early Warning Score [MEWS]), the RRT is suffering from low accuracy and false alarms. In terms of the accuracy, the RRT could only detect 30% of unplanned intensive care unit admissions, and the MEWS that the RRT team employs does not even reach an area under the receiver operating characteristic (AUROC) of 0.8 [1-3]. Previous algorithms have focused on accuracy; however, in the clinical environment false alarms are the most important factor when evaluating an algorithm [4,5]. Whenever alarms sound, in real time RRT reconfirm patients and determine intervention. Therefore, if there are too many false alarms, an unnecessary amount of time and costs are wasted.

To overcome the limits of MEWS, many traditional machine learning algorithms have been incorporated but still suffer from low accuracy and false alarms. In this article, we introduce

Lee Y, et al.   Deep Learning Based Early Warning Score

ACC

the Deep learning based Early Warning Score (DeepEWS), that is based on a novel deep learning algorithm [6]. Deep learning has demonstrated state of the art results for computer vision, machine translation and has demonstrated many successful results [7-9]. In the medical domain, deep learning has been limited to being applied to medical images. However, recently there have been new applications to EHR and biosignal research.

DeepEWS was compared to rule-based and the most promising traditional machine learning algorithms in the clinical setting. There were many different scenarios on which the results were compared and particularly, the sensitivity versus the number of alarms effectively demonstrated the superiority in accuracy and ease of clinical implementation for DeepEWS.

MACHINE AND DEEP LEARNING

Both machine learning and deep learning analyze data through self-learning to solve the task or problem at hand (i.e., predict,

KEY MESSAGES

- We introduce the DeepEWS (Deep learning based Early Warning Score), which is based on a novel deep learning algorithm.
- Relative to the standard of care and current solutions in the marketplace, there is high accuracy and low false alarms.

classify). However, machine learning requires feature engineering, whereas deep learning does not, and this is the essence of why deep learning provides higher accuracy [10]. Feature engineering strives to solve the problem in the model through pattern recognition by sensitizing various scenarios. In this process, domain knowledge and expertise can translate to significantly different results. However, deep learning uses deep neural networks to find these features automatically. As such, even without domain knowledge and specialist expertise, the model can self-learn. The deeper the neural

Figure 1. DeepEWS (Deep learning based Early Warning Score) operating on data from Electronic Medical Records. MEWS: Modified Early Warning Score; NIBP(S): non-invasive blood pressure (systolic); NIBP(D): non-invasive blood pressure (diastolic); HR: heart rate; RR: respiratory rate; SPO₂: blood oxygen saturation; AVPU: alert, verbal, pain, unresponsive.

Lee Y, et al.   Deep Learning Based Early Warning Score

ACC

network, the more diverse features are found, and this translates to higher performance and accuracy in results.

DEEP LEARNING BASED EARLY WARNING SCORE

In-hospital cardiac arrests alongside with sepsis is a major factor that determines patient safety. In the United States, every year there are 20,900 cardiac arrest patients, and globally the survival rate for cardiac arrest patients is below 20%. It is well documented in the literature, that 80% of cardiac arrest patients experience deterioration in health, 8 hours before the cardiac arrest event, and there have been many attempts to predict this onset. However, the current standard of care suffers from low accuracy and false alarms that renders the clinical implementation and practical use of these technologies challenging.

DeepEWS solved two major shortcomings that existing algorithms suffer from (low accuracy and false alarms) and provided clinically viable solutions. DeepEWS uses all four vital signs measured for 8 hours that can be easily measured in a hospital setting (systolic blood pressure, body temperature, heart rate, and respiratory rate). The details of the method can be confirmed in [6]. As shown in Figure 1, DeepEWS can measure real-time cardiac arrest risks (within 24 hours of the cardiac arrest event) from vital signs in Electronic Medical Records.

The study period of this paper was from June 2010 to July 2017, with 56,076 patients were admitted to the Mediplex Se-

jong Hospital, and 415 first cardiac arrest occurred. Four forty-eight patients were excluded because they were admitted or discharged outside the study period.

There are two significant results that the study demonstrated: the accuracy level and applicable in a clinical environment. Figure 2 represents the AUROC for DeepEWS and competing algorithms. DeepEWS shows an AUROC of 0.850 relative to MEWS's 0.603, a 40.9% improvement in accuracy; (currently, the second version of DeepEWS recorded an AUROC of 0.911, a vast improvement from its previous version). Relative to algorithms used in previous research such as random forest (0.780) and logistic regression (0.613), the accuracy levels are markedly higher as well.

Figure 3 shows the sensitivity relative to the number of alarms and demonstrates how effective DeepEWS is in the clinical environment by maintaining high levels of accuracy even at the low number of alarms. In the real-world setting, an RRT is required to cover many patients so keeping a high level of accuracy with a small number of alarms is quintessential. We can see that at points of the lower number of alarms the accuracy level is significantly higher for DeepEWS relative to other competing algorithms. DeepEWS is more accurate with a lower number of alarms. DeepEWS is 42.7% accurate when the number of alarms is equal to 40 (point B); MEWS is 22.7% accurate when the number of alarms is equal to 110 (point C).

CONCLUSION

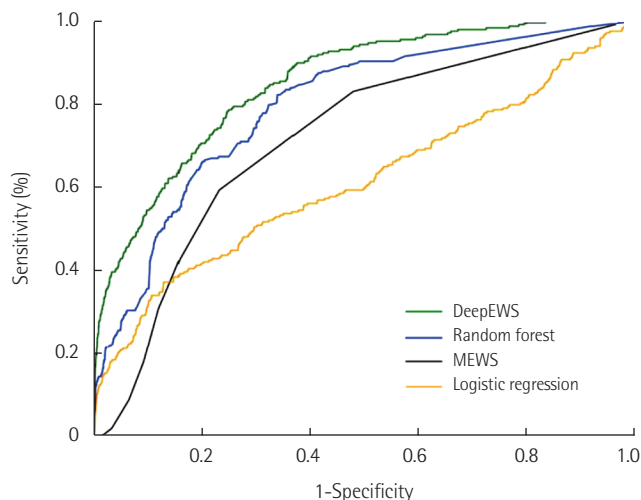This study has two significant implications. First, the first study

Figure 2. The area under the receiver operating characteristic of algorithms for detecting cardiac arrest. DeepEWS: Deep learning based Early Warning Score; MEWS: Modified Early Warning Score.
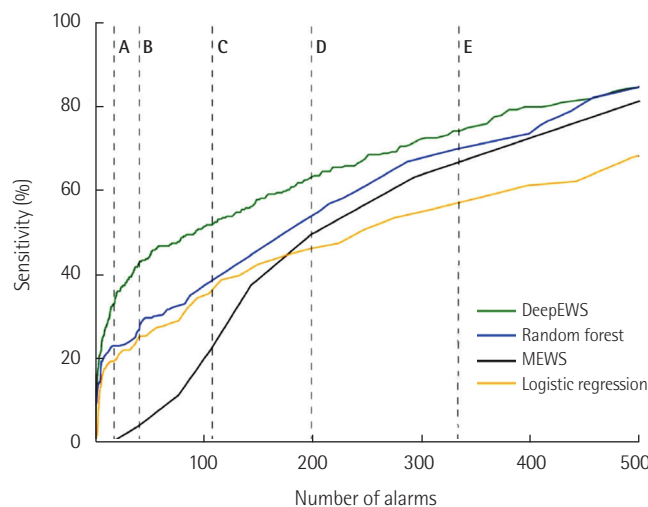

Figure 3. The sensitivity according to the number of alarms. The Y-axis indicates sensitivity, and the X-axis means the number of alarms per 1,000 patients in 1 hour (A: 20, B: 40, C: 110, D: 200, E: 330).

Lee Y, et al.   Deep Learning Based Early Warning Score

ACC

using deep learning has significantly improved accuracy. Second, it has been shown that DEWS is applicable in a clinical environment by comparing the accuracy according to the number of alarms. A large amount of false alarms is the factor for why MEWS and traditional machine learning algorithms could not be adopted in the clinical environment. DeepEWS predicts cardiac arrest with high accuracy and low false alarms. There are only four vital signs that are used as inputs that can easily be obtained in any clinical environment and augmenting these results with laboratory results, and clinical notes could enhance the accuracy even further.

CONFLICT OF INTEREST

Youngnam Lee, Joon-myoung Kwon, Yeha Lee, and Jinsik Park were involved in developing DEWS (Deep learning based Early Warning Score). VUNO provided support in the form of salaries for authors (Youngnam Lee, Yeha Lee, Hyunho Park, and Hugh Cho), but did not provide funding and did not have any additional role. No other conflict of interest relevant to this article was reported.

ORCID

Joon-myoung Kwon   https://orcid.org/0000-0001-6754-1010

REFERENCES

1. Hillman K, Chen J, Cretikos M, Bellomo R, Brown D, Doig G, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. Lancet 2005;365: 2091-7.

2. Trinkle RM, Flabouris A. Documenting rapid response system afferent limb failure and associated patient outcomes. Resuscitation 2011;82:810-4.

3. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016;44:368-74.

4. Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007;33:667-79.

5. Bell MB, Konrad D, Granath F, Ekbom A, Martling CR. Prevalence and sensitivity of MET-criteria in a Scandinavian University Hospital. Resuscitation 2006;70:66-73.

6. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc 2018;7:e008678.

7. Deo RC. Machine learning in medicine. Circulation 2015;132: 1920-30.

8. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402-10.

9. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. JAMA 2016;316: 2353-4.

10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521: 436-44.