ISSN 2466-135X
Vol.10, No.1

The 10th International Conference on

BIG DATA APPLICATIONS AND SERVICES (BIGDAS2022)

PROCEEDING

November 24–26, 2022
Jeju Island, Korea

Hosted by
Korea Big Data Service Society

THE KOREA BIG DATA SERVICE SOCIETY
한국빅데이터서비스학회

충북대학교 빅데이터연구소

제주대학교
JEJU NATIONAL UNIVERSITY
SW융합교육센터
Software Convergence Education Center

Table of Contents

Deep learning-based model for rapid prediction of in-hospital clinical deterioration

Trong-Nghia Nguyen[1], Ngoc-Tu Vu[1], Bo-Gun Kho[2], Guee-Sang Lee[1], Hyung-Jeong Yang[1], Soo-Hyung Kim[1,*], Aera Kim[1]

[1] Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju, Korea

[2] Pulmonology and Critical Care Medicine, Chonnam National University Hospital, Gwangju, Korea

{trongnghia7171, tuvungocnd, imdrkbg}@gmail.com,
{gslee, hjyang, shkim, arkim}@jnu.ac.kr

Abstract. In this study, we develop a deep learning application system with high interpretability and diversity in input features for the prediction of in-hospital clinical deterioration. The high ability to understand input features helps the system to stick to the actual context. The use of the Transformer structure has made our method superior to comparative models in many respects when testing on a large and challenging data set with a 0.652 F1-score, 0.77 sensitivity, 0.837 AUROC, and 0.839 AUPRC.

Keywords: Clinical Deterioration, Rapid Response System, Deep Learning, Machine Learning.

1 Introduction

Deterioration in hospitals is a serious problem for medical systems. Approximately 209,000 patients are treated for cardiac arrest in hospitals each year [1]. This index has increased since the emergence of the coronavirus pandemic 2019 (COVID-19), which has become a burden on public health [2]. In this context, rapid response systems (RRS) are constantly being researched and developed to prevent treatment delays that are caused by an overload due to large numbers of hospitalized patients. Many studies have focused on developing a "risk score" - an indicator to assess the loss of danger to a patient's clinical condition. In this investigation, we develop a Rapid Response System of applying deep learning and transformed architect techniques to improve the predictive quality of the rapid response system through a probability algorithm.

* Corresponding author.

1.1 Related works

When it comes to "risk score", it must be mentioned classic methods like Modify Early Warning Score [3], and National Early Warning Score [4]. These assessment methods are mainly based on the diagnostic experience of specialists and physicians. Through the input of patient vitals (Heart Rate (HR), Respiration Rate (RR), Body Temperature (BT), etc.) at a time, MEWS and NEWS can estimate the condition of that patient (Normal/Abnormal) at that moment. The limitation of the above methods is that the range of input features is too narrow in the context of the development of electronic health records (EHRs). This makes it impossible for us to take full advantage of the variety of input features. Thus, machine learning/deep learning application approaches were born.

In recent years, a large number of studies aimed at increasing the likelihood of predicting the early sign of clinical deterioration have been published. Typical can be mentioned MEWS ++, a variant of MEWS aimed at improving the ability to predict clinical developments in hospitalized patients through machine learning models [5]. This system trains three classical machine learning methods (random forest (RF), linear support vector machine, and logistic regression) on a large dataset. By comparison, it surpassed traditional MEWS with an increase of 37% in sensitivity, 11% in specificity, and 14% in the area under a receiver operating characteristic (ROC) curve (AUROC). Besides, DeepSigns [6]: a method for creating a computational model that can predict the deterioration of a patient's health in such a way that appropriate treatment can be started as soon as possible, has been developed based on the integration of Deep learning techniques, Recurrent Neural Networks [7], and the Long Short-Term Memory [8], were also introduced. Furthermore, we have also published a study [9] with the application of TabNet [10] - Interpretable Learning for tabular data on the same topic. Experimental results on a private dataset show that the proposed method outperforms machine learning models with 66% AUROC and 29.1% of the area under a precision-recall curve (AUPRC). However, this study still has many limitations such as limitations on input features type, and low sensitivity leading to the system skipping time points that should be alerted.

1.2 Method and Contribution

From the above background and motivation, we aim to develop deep learning based RRS (Deep-RRS) to satisfy the following factors: (1) Diversity in input data types - Instead of using only numeric features (vital signs or laboratory tests), our method applies Transformer architecture [11] to optimize the interpretability and understanding of the model for textual features. (2) Increase the sensitivity of the system - Improve the ability to detect "abnormal time points".

2 Proposed Method

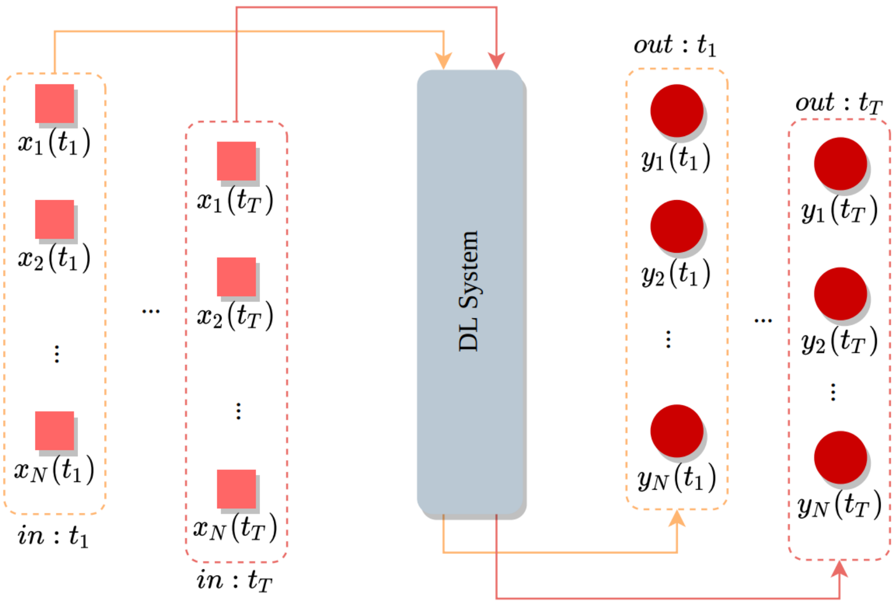

Figure 1: Overall pipeline of our proposed system. Input: patient's clinical variables (x) at timepoint t; Output: abnormal probability (y) corresponding to input measurement values x at time t.

The overall proposed system is illustrated in **Figure 1**. The main goal of the system is to assess whether the clinical condition of the patient at a given time is abnormal or normal through abnormal probability. Our system takes as input the patient's clinical indicators and general features at a certain time point. Here, the Machine Learning (ML)/Deep learning models (DL) trained on the large data set will process the input information and return the patient's abnormal probability at the corresponding time. By estimating a suitable threshold (0.5 in this problem), we can determine the final abnormal status from the above probability index. Thus, the "risk score" in our system is the abnormal probability.

2.1 Feature Selection

We mentioned the diversity of input data types as an advantage of this study over our previous method. However, more than simply using more features to increase the system's performance, this study performs physician-assisted feature selection so that these features really make sense for the medical side and are suitable for the emergency system.

From 37 features of the 2021 RRT CNUH dataset - the dataset used for this study, we conduct pre-processing, screening, and selecting the most 11 suitable and optimal features for the system. Used features include patient clinical variables and general information, which are divided into 2 types:

Numerical features (n = 6): Age (general) and vital signs (clinical): Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Hear Rare (HR), Respiration Rate (RR), Body Temperature (BT).

Categorical features (n = 5): All are the patient's general information: Gender, Diagnosis, Hospitalization Department, Hospitalization Route, and Inpatient Ward.

2.2 Deep Learning Model Architecture

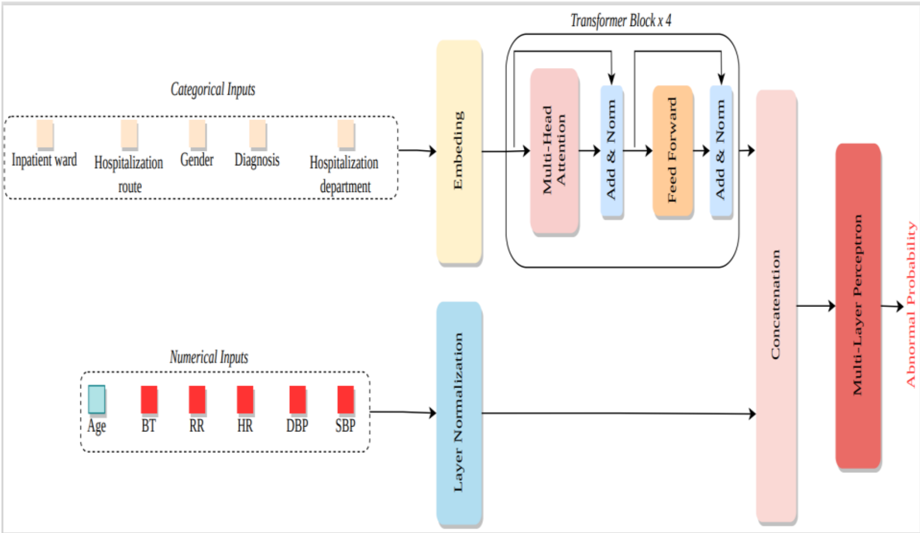

Figure 2: Proposed Transformer based model for in-hospital abnormal status prediction in Deep-RRS.

Our proposed DL model (**Figure 2**), is based on TabTransformer's structure [12], consisting of 3 main elements: an embedding layer, 4 Transformer layers, and a multi-layer perceptron. Which, each Transformer layer consists of a multi-head attention layer, followed by a positional transition layer. Among the input features, we have Diagnosis - a field containing the doctor's assessment and comment about the patient's condition through a sentence. Therefore, we use Transformer constructs because of their powerful natural language processing capabilities.

Assume that we have a set of input features $x = \{x_{cat}, x_{num}\}$, where $x_{num} \in \mathbb{R}^n$ denote all n numerical features and $x_{cat} = \{x_1, x_2, \ldots, x_m\}$ illustrates all m categorical features. model's pipeline can be summarized into the following steps:

1. x_{num} is normalized by Layer Normalization while x_{cat} is embedded to get the embedded features $E_\varphi(x_{cat})$.
2. Embedded features $E_\varphi(x_{cat})$ are fed into 4 Transformer blocks for obtaining contextual embedding features. The sequence Transformer layers could be

formulated as S_ϕ which processes $E_\varphi(x_{cat})$ to get the contextual embedding $\mathbb{C} = \{c_1, ..., c_m\}$ corresponding for each x_i, $i \in \{1, ..., m\}$.

3. Concatenating contextual embedding features $\mathbb{C}$ and normalized numerical features together.
4. Concatenation is used as input of MLP (denoted as $\mathcal{M}_\vartheta$) to get the prediction probability y.

Let $\mathcal{K}$ be the cross-entropy for the whole classification task, our model minimizes the loss function $\mathcal{L}(x, y)$ for enhancing all model's parameters by the first-order gradient approaches. Which, the model defines φ as the column embedding, ϕ for the Transformer layer, and ϑ for the top of MLP.

$$\mathcal{L}(x, y) = \mathcal{K}(\mathcal{M}_\vartheta(S_\phi(E_\varphi(x_{cat})), x_{num}), y). \quad \textbf{(1)}$$

3 Experiment Results

3.1 Dataset & Experiment configuration

For the experimental process, we performed the study on the 2021 Rapid Response Team Chonnam National Hospital dataset (RRT-CNUH). This dataset was collected and screened by the Rapid Response Team of Hakdong Chonnam National University Hospital during the period from February 1, 2021, to November 30, 2021. **Table 1** aggregates demographics on training/test sets and some used characteristics of the dataset.

Table 1 Cohort demographics of the RRT-CNUH dataset.

		Total N (%)	***Training (%)***	***Test (%)***
Number patients		25,329	18,470 (80)	6,859 (20)
Normal (%)			15,775 (81.4)	5,873 (85.6)
Abnormal (%)			3,602 (18.6)	1,335 (14.4)
General Information	Age (mean±std)		64.2 ± 17.6	64.0 ± 18.1
	Gender (male/female)		9,752/8,718	3,605/3,254
Vital signs (mean±std)	SBP		110.5 ± 38.7	110.7 ± 38.9
	DBP		65.4 ± 27.3	65.5 ± 27.4
	HR		69.6 ± 27.3	69.4 ± 27.2
	RR		17.6 ± 6.5	17.6 ± 6.4
	SBT		34.6 ± 8.4	34.6 ± 8.0

We split this dataset by 80% for training and 20% for testing. The mean age of the patients in the data set fluctuates at 64 and the male/female ratio is balanced. However, the rate of clinical performance deterioration (abnormal) for the whole samples has a large difference (18.6% for train and 14.4% for test). Therefore, the models need to be sensitive enough to extract the minority samples from the entire data.

For comparison models, we divide them into 3 main groups:

Traditional Method: MEWS - This method will lose its advantage over other models because it only calculates based on 5 vital signs features.

Machine Learning Models: Extreme Gradient Boosting (XGBoost), Random Forest (RF), Multi-Layer Perceptron, Light Gradient Boosting - Effective machine learning models for tabular classification based on tree and boosting principles. These models make full use of input features and are hyper-parameterized by Sklearn's GridsearchCV library.

Deep Learning Models: TabNet [10], DANets [13] - Two novel models have been proven to outperform machine learning methods in recent years. TabNet is the method we applied in our previous research.

With the characteristic of the binary classification task and the unbalanced nature of the data, besides F1-score, we use AUROC and AUPRC as comparison metrics. Besides, to evaluate the sensitivity of the model, the Sensitivity score is also used.

3.2 Results

Table 2 Models performance metrics.

Method	*F1*	*Sensitivity*	*AUROC*	*AUPRC*
MEWS	0.471	0.01	0.526	0.465
XGBoost	0.599	0.65	0.705	0.693
RF	0.612	0.69	0.736	0.724
MLP	0.553	0.44	0.635	0.621
LGB	0.602	0.68	0.707	0.695
TabNet [10]	0.612	0.73	0.715	0.705
DANets [13]	0.532	0.37	0.597	0.573
Deep-RRS (proposed)	0.652	0.77	0.837	0.839

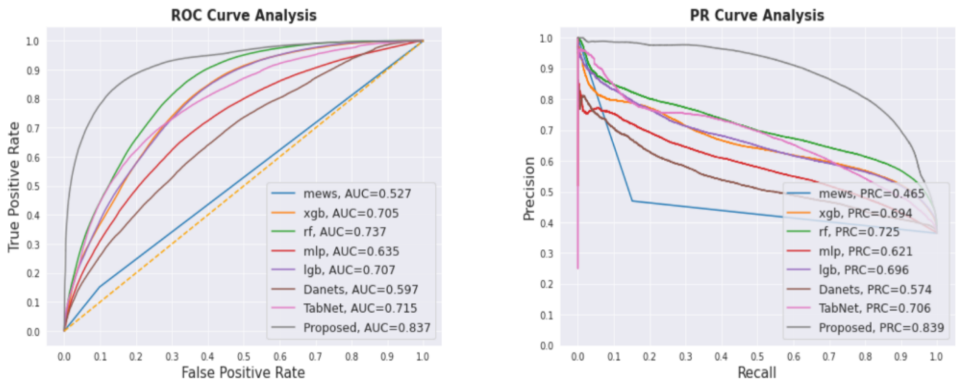

Figure 3 ROC & PRC curve analysis.

Table 2 shows the overall performance of comparison approaches and **Figure 3** illustrates the ROC curve and Precision-recall curve analysis, respectively. Our proposed model outperformance comparison approaches in all evaluation metrics with 0.652 of F1-score, 0.77 of sensitivity, 0.837 of AUROC, and 0.839 of AUPRC. Deep-RRS showed significant improvement in the case of AUROC and AUPRC with an increase of up to 0.101 of AUROC and 0.115 of AUPRC (compared to RF (0.736 of AUROC and 0.724 of AUPRC) - best of the rest). The high sensitivity has also shown the model's ability to overcome the "missing alarm" problem.

From the above results, it could be seen that the ability to optimize the characteristics of the input data through extracting robust contextual embedding features has increased the efficiency of the prediction process. Exploiting the context of categorical features has enhanced the interpretability of the system. In particular, the use of the Transformed technique - a powerful structure in natural language processing has helped the model make the most of textual features such as Diagnosis instead of having to use limited features for MEWS or must encode them for machine learning models.

4 Conclusion

Deep-RRS has been shown to be superior in predicting clinical deterioration by time step. The advantage of being able to take advantage of a variety of input features helps the system to overcome comparison methods due to its high interpretability. Furthermore, the high sensitivity of the system has been verified through testing and evaluation. However, the limitation of this study is its output as the system can only predict each time step. Therefore, we will improve the versatility of the time series-based system in the future.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1A4A1019191). This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF)& funded by the Korean government (MSIT) (NRF-2019M3E5D1A02067961). The corresponding author is Soo-Hyung Kim.

References

1. Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R. Mackey, R. H. (2017). Heart Disease and Stroke Statistics- 2017 Update: A Report from the American Heart Association. Circulation, 135(10), e146–e603.
2. R. M. Padilla and A. M. Mayo: Clinical deterioration: A concept analysis, Journal of Clinical Nursing, vol. 27, no. 7-8, pp. 1360–1368, 2018.
3. C. P. Subbe, M. Kruger, P. Rutherford, and L. Gemmel, “Validation of a modified early warning score in medical admissions,” QJM – Monthly Journal of the Association of Physicians, vol. 94, no. 10, pp. 521–526, 2001.
4. Royal College of Physicians, “National Early Warning Score (NEWS) 2 Standardizing the assessment of acute-illness severity in the NHS: Additional Implementation Guidance,” Rcplondon. Ac. Uk, no. March 2020.

5. A. Kia, P. Timsina, H. N. Joshi, E. Klang, R. R. Gupta, R. M. Freeman, D. L. Reich, M. S. Tomlinson, J. T. Dudley, R. Kohli-Seth, M. Mazumdar, and M. A. Levin, "MEWS++: Enhancing the prediction of clinical deterioration in admitted patients through a machine learning model," Journal of Clinical Medicine, vol. 9, no. 2, 2020.
6. D. B. da Silva, D. Schmidt, C. A. da Costa, R. da Rosa Righi, and B. Eskofier, "DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration," Expert Systems with Applications, vol. 165, no. March 2020, p. 113905, 2021.
7. Rumelhart, David E; Hinton, Geoffrey E, and Williams, Ronald J (Sept. 1985). Learning internal representations by error propagation. Tech. rep. ICS 8504. San Diego, California: Institute for Cognitive Science, University of California.
8. Hochreiter, S., & Schmidhuber, J"urgen. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
9. T.-n. Nguyen, T.-h. Vo, B.-g. Kho, and G.-s. Lee, "Deep Interpretable Learning for a Rapid Response System," pp. 8–10.
10. S. O. Arik and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning," 2019.
11. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.
12. X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "TabTransformer: Tabular Data Modeling Using Contextual Embeddings," no. Rong 2014, 2020.
13. J. Chen, K. Liao, Y. Wan, D. Z. Chen, and J. Wu, "DANets: Deep Abstract Networks for Tabular Data Classification and Regression," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 4, pp. 3930–3938, 2022.