



NATURAL LANGUAGE PROCESSING FOR EARLY PREDICTION OF SEVERE SEPSIS IN THE EMERGENCY DEPARTMENT

Po Hsiang Lin^{*1}, Wei-Chun Huang¹, Jer-Guang Hsieh², Wang-chuan Juang¹

1. Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan



Kaohsiung Veterans General Hospital,
TAIWAN

INTRODUCTION

Severe sepsis remain the leading cause of death no matter in the emergency department or intensive care unit admissions. Early recognition of severe sepsis is the first step for timely management and resuscitation in the emergency settings.

Nature language processing becomes more and more important as the growth of the electronic health record data and the improvement of the deep learning method.

OBJECTIVES

We aim to predict whether if severe sepsis for the patients admitted from the emergency department by the information of the triage..

METHOD

We conducted a cohort study of adults Hospitalized from the emergency department in The Kaohsiung Veterans General Hospital during January 2017 and December 2017. We randomly split the data to training and validation part under the ratio 4:1. The definition of severe sepsis is defined using ICD-9 which related with sepsis and organ failure. The included features obtained from the triage are respiratory rates, body temperature, pulse rates, blood pressure, coma scale, past history. Besides, we also used texts from the triage nurse by the nature language processing method. We choose Tensorflow to create deep learning model which can handle different types of data included binary, categorical, and texts.

In addition, to translate the triage nurse records written in Traditional Chinese into English, we employed EasyNMT which applied the model “mbart50_m2en” published by Facebook research.

RESULTS

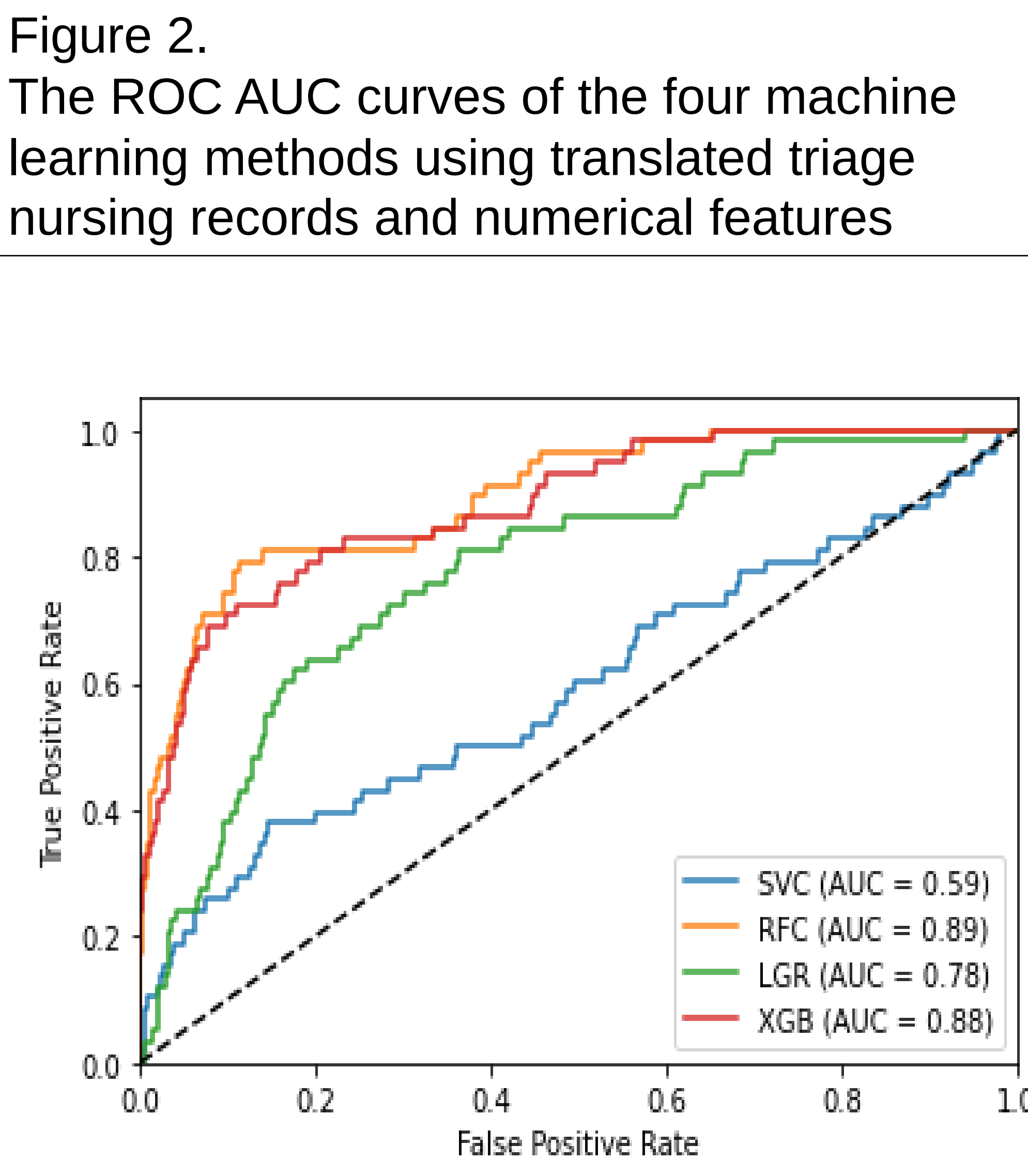
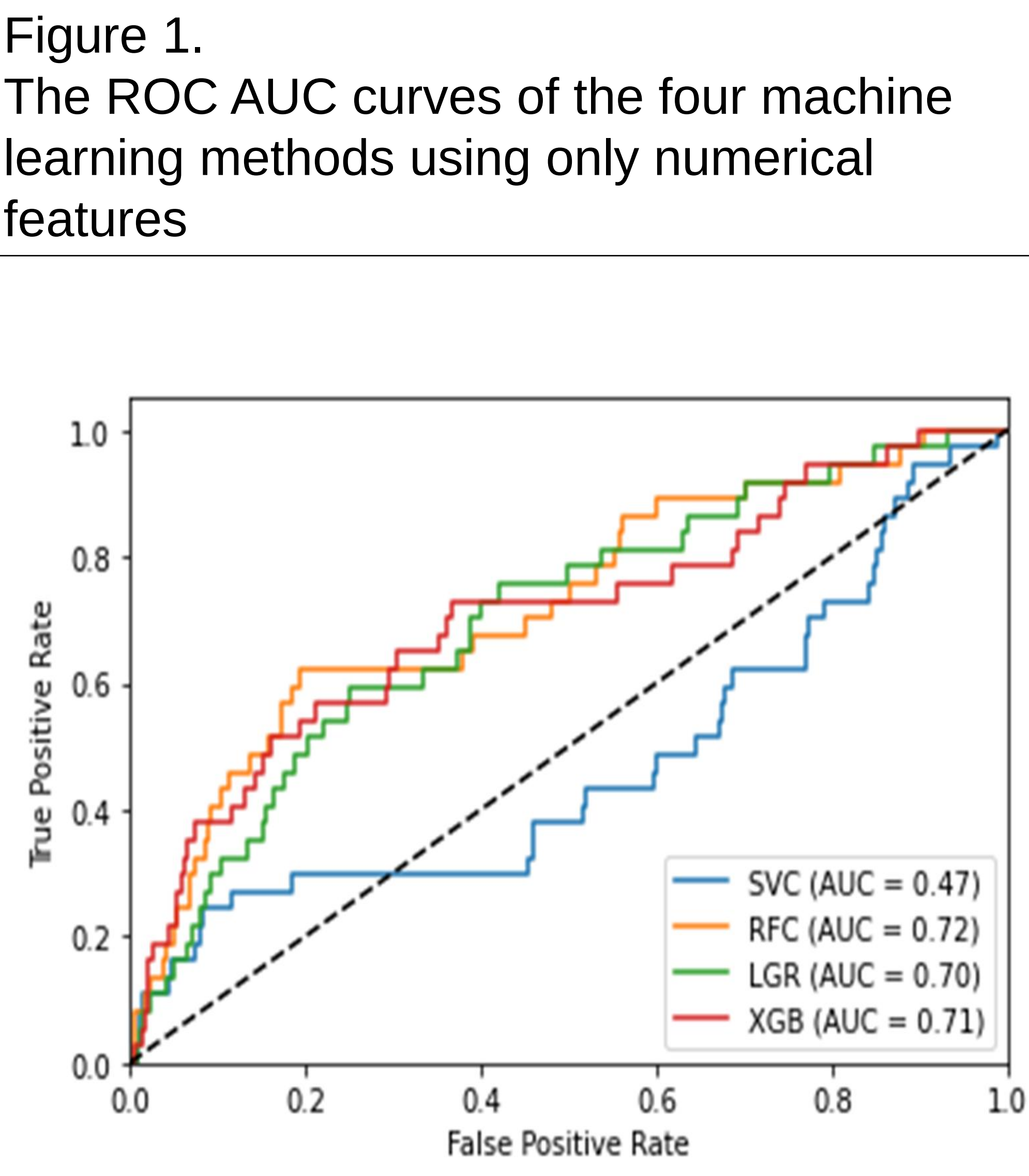
This study included 13180 participants, and only 377 patients suffered from severe sepsis under the definition of ICD-9.

First, we construct four different machine learning models including support vector classifier(SVC), logistic regression(LGR), random forest classifier(RFC), and XGBoost(XGB) under the training dataset and validated on the remaining 1/5 dataset without texts from triage nurse records. The metrics about the area under the receiver operating characteristic curve(AUC), average precision score(APS), sensitivity, specificity, are calculated.

Then in order to deal with texts of triage nurse records written in Traditional Chinese, we employed EasyNMT which applied the model “mbart50_m2en” published by Facebook research for translation to English. Besides, we used the model “Bio_ClinicalBERT” for feature extraction. After we extracted the features from the last four layers, we added these features with the previous numerical features for computing. The following two tables present the results of numerical data calculations and the combined NLP method calculations(NLP+), respectively.

Among these four machine learning methods, RFC performed the best, and it can be observed that after incorporating the NLP method, the AUC increased to 0.89 due to the additional information from the triage nurse records.

Table 1.					Table 2.				
The performance of the four machine learning methods using only numerical features.					The performance of the four machine learning methods using translated triage nursing records and numerical features.				
Numerical	SVC	LGR	RFC	XGB	NLP+	SVC		RFC	XGB
Accuracy	0.97	0.73	0.89	0.97	Accuracy	0.97	0.73	0.91	0.97
AUC	0.47	0.70	0.72	0.71	AUC	0.59	0.78	0.89	0.88
Sensitivity	0	0.70	0.62	0	Sensitivity	0	0.70	0.69	0
Specificity	1	0.60	0.80	1	Specificity	1	0.74	0.91	1
APS	0.05	0.06	0.09	0.08	APS	0.07	0.09	0.42	0.43



CONCLUSIONS

By employing the NLP method, not only can nursing records in different languages be translated, but feature extraction can also be utilized to integrate machine learning techniques. This approach aims to improve the prediction of severe sepsis in emergency department.