

# Early Prediction of 30-Day ICU Re-admissions Using Natural Language Processing and Machine Learning

Zhiheng Li<sup>1</sup>, Xinyue Xing<sup>1</sup>, Bingzhang Lu<sup>2</sup>, Ying Zhao<sup>3</sup>, Zhixiang Li<sup>4,\*</sup>

<sup>1</sup>High School Division, Northeast Yucai Foreign Language School, Shenyang, China

<sup>2</sup>Junior High Division, Northeast Yucai School, Shenyang, China

<sup>3</sup>Department of Engineering Science and Applied Math, Northwestern University, Evanston, USA

<sup>4</sup>Department of Biomedical Engineering, Shenyang Pharmaceutical University, Shenyang, China

## Email address:

1479551965@qq.com (Zhiheng Li), 1045068673@qq.com (Xinyue Xing), 2268665795@qq.com (Bingzhang Lu),

yingzhao2018@u.northwestern.edu (Ying Zhao), yikuanli8@gmail.com (Zhixiang Li)

\*Corresponding author

## To cite this article:

Zhiheng Li, Xinyue Xing, Bingzhang Lu, Ying Zhao, Zhixiang Li. Early Prediction of 30-Day ICU Re-admissions Using Natural Language Processing and Machine Learning. *Biomedical Statistics and Informatics*. Vol. 4, No. 3, 2019, pp. 22-26. doi: 10.11648/j.bsi.20190403.11

**Received:** October 10, 2019; **Accepted:** November 18, 2019; **Published:** November 22, 2019

---

**Abstract:** ICU readmission is associated with longer hospitalization, mortality and adverse outcomes. An early recognition of ICU readmission can help prevent patients from worse situation and lead to lower treatment cost. As the abundance of Electronics Health Records (EHR), it is popular to design clinical decision tools with machine learning techniques manipulating on healthcare large scale data. To this end, we designed data-driven predictive models to estimate the risk of Intensive Care Unit (ICU) readmission. The discharge summary of each hospital admission was carefully represented by natural language processing algorithms. Unified Medical Language System (UMLS) was further used to standardize inconsistency of discharge summaries. 5 machine learning classifiers including naïve Bayes, support vector machine, logistic regression, gradient boosting decision tree and 2 feature representations including Bag-of-Words and Bag-of-CUIs were adopted to construct predictive configurations. The best configuration yielded a competitive AUC of 0.748. High contribution words and medical terms were further investigated to ensure that they were clinical meaningful. A comparative study between two feature representations were also discussed. Our work suggests that natural language processing of discharge summaries is capable to extract meaningful information from discharge summary automatically and to send clinicians the warning of unplanned 30-day readmission upon discharge.

**Keywords:** ICU Readmission, Machine Learning, Natural Language Processing, Unified Medical Language System (UMLS)

---

## 1. Introduction

ICU readmission describes the situation where patients get unplanned readmitted into the ICU within a relatively short interval (30 days as recognized in most cases) [1-3]. Such readmission is usually accompanied by deteriorated symptoms and thus longer ICU stay as well as higher mortality rate [4]. Predicting ICU readmission can help us recognize potential causes of such unwanted situation [5]. The most effective therapy will be concluded accordingly so as to prevent readmissions and the problems that may follow.

EHR (Electronic Health Record) stands for an

electronically stored data base containing health conditions of a broad range of population [6]. Demographics, medical history, medication, allergies, immunization status, laboratory results, radiology images and vital signs are usually digitalized and stored in EHR database or data warehouse [7]. Established originally for observational studies, EHR has now earned an irreplaceable position in the medical research world due to its abundance in easily accessible and comprehensible medical records. Due to its digital nature, EHR can be shared among institutions conveniently [8]. Its abundance in information also makes it a perfect source in medical decision making.

Machine learning has long been a popular measure of constructing prognostic models on the basis of abundant data [9]. Given the collaborative ability of machine learning, and that the EHRs cannot be easily integrated manually, machine learning is widely used to process EHRs and further contribute a great deal in medical decision making [10] and medical predictive models [11]. To this end, we consider associating the plentiful resources of EHR and the mighty models of machine learning to construct comprehensive analytical models and therefore produce reliable predictions of ICU readmissions.

## 2. Related Works

There has been growing interest in the prediction of ICU readmission with machine learning approaches. A machine learning ICU readmission clinical decision support tool using patient EHR data may provide clinicians with additional information at discharge time that patients might be readmitted in the following 30 days [12]. Prior researches have abundant discussion on using machine learning to predict ICU readmission on both the methodology and algorithm side [13]. However, these researches have following two limitations. In one way, they focus on a subgroup of patients, such as patients with cardiovascular disease [14], post-surgery patients. In the other way, the prediction models were highly relied on lab results or biomarkers [15] to be input features, which easily causes missing data or extra treatment cause. To these limitations, we would like to extract clinical meaningful representations from discharge summaries, that every patient has it naturally, to predict patients 30-day unplanned readmission at discharge.

## 3. Methods

Detailed steps of how we built the 30-day unplanned ICU readmission predictive model can be found in Figure 1. We will introduce it step by step in the following subsections.

### 3.1. Dataset

Data for this study was a subset extracted from the public accessible critical care database, named MIMIC-III [2]. This dataset contains over 40,000 patients' de-identified data when the patients were in ICU unit. We first extracted all patients' admission data from the table named ADMISSIONS. Next, we excluded patients with the admission type of NEWBORN, as the newborn patients' information might be archived in another database resulting in these patients have a high missing rate of discharge summaries. Then, a hospital-expiration-flag, which describes whether patients pass away during hospitalization, was used to discard those patients do not have re-admission due to mortality. Then, if one ICU stay has a straight following ELECTIVE admission, we would seek the next "non-ELECTIVE" admission to be the following readmission, as we focused on unplanned readmission. We should notice that patients with multiple

ICU stays were treated as independent subjects. We used *subject-id* and *row-id* to link ADMISSION table with CHARTEVENTS (notes) table. The discharge summary of each selected subject was then found. Some odd cases, such have multiple discharge summaries of one stay, were then excluded. Finally, a cohort of 45305 subjects were generated, among which 5.26% has unplanned 30-day readmission. The distribution of readmission time interval can be found in Figure 2.

### 3.2. Word and Concept Representation

The discharge summaries were first pre-processed with Porter stemming [16] to reduce inflectional variations. In order to downstream to the machine learning classifiers, the raw text of summaries should be converted to vectors. We employed Bag-of-words (BoW) as the vectorization techniques for its simply implementation and great success in language modeling and document classification. With "BoW", each word in the summaries can be represented into a value while one summary can be represented into a vector. As "BoW" aims at extracting important words from one document to distinguish this document from others with the help of these important words. Some exclusion criteria should be applied to improve the performance and highlight important words. We excluded 313 stop words as the reference from NCBI [17]. The words occurred in more than 95% summaries were also discarded. To avoid that excessive features would bring difficulty to the machine learning classifiers, words occurred less than 5 summaries were also excluded. Term frequency-inverse document frequency (tf-idf) weighting adjustment was also applied. Finally, 22,405 words were included to form the vocabulary. In other words, each discharge summary was represented to a 22,405-long vector.

Considering about the variation and written habits when generating the discharge summaries, there was plenty of inconsistency to claim a same clinical meaning. This brought great difficulties to over model as "BoW" only care about the words. To address this problem, we employed the phenotyping system named MetaMap server from Unified Medical Language System (UMLS) [18] to identify medical concepts from the discharge summaries. Each concept has a unique identifier named Concept Unique Identifiers (CUIs). We also generated Bag-of-CUIs as the feature representation from discharge summaries.

Briefly, two feature sets: Bag-of-Words and Bag-of-CUIs were generated by interpreting discharge summaries using NLP techniques. These 2 feature sets would be transferred downstream to the machine learning classifiers.

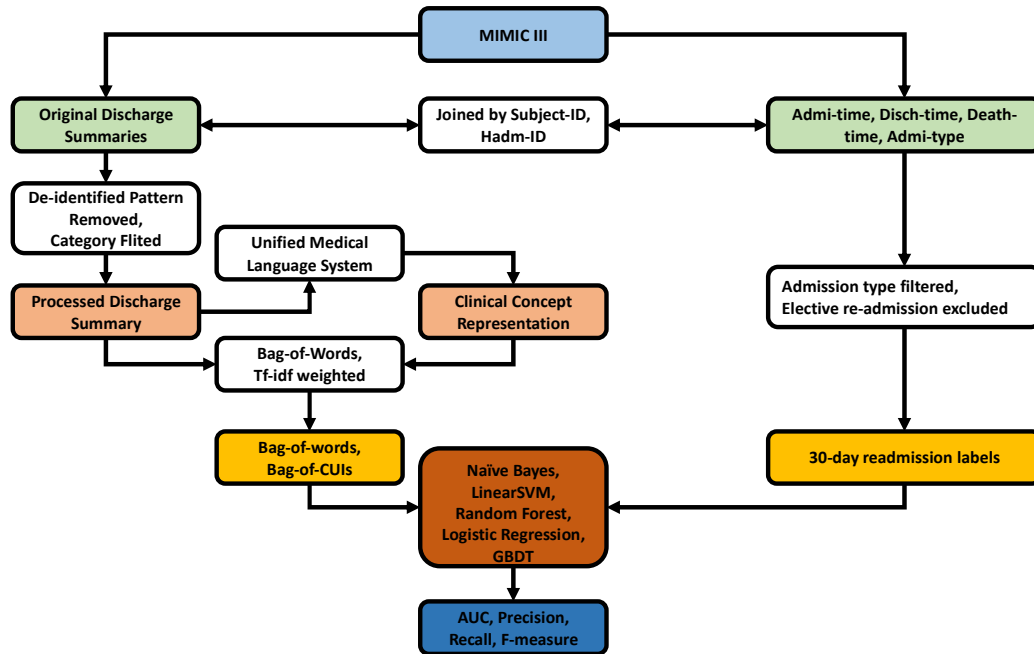
### 3.3. Machine Learning Classifiers and Pipeline

5 supervised learning classifiers were selected to build the predictive models, including 2 linear classifiers: logistic regression and support vector machine with linear kernel (LinearSVM), 2 ensemble models: random forest and gradient boosting decision trees (GBDT), and Naive Bayes as

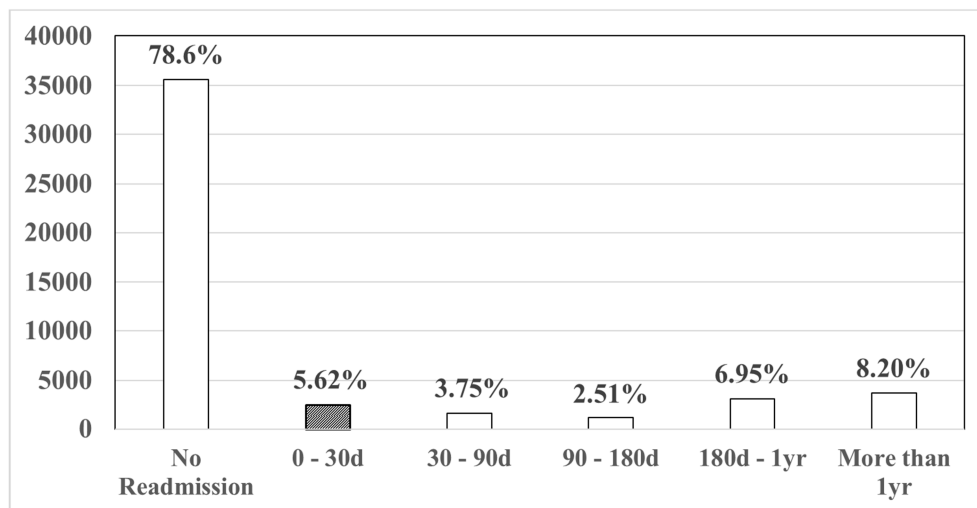
the benchmark. Entire dataset was first stratified to training set: held-out test set as 7:3. There were 5 algorithms on 2 feature sets, which were totally 10 configurations. We adopted area under the receiver operating characteristic (AUC) to evaluate the performance as the binary classification is highly imbalanced. Parameters of each algorithm were tuned with 5-fold cross-validation on the training set and then examined on the held-out test set.

### 3.4. Tools

We built our configurations over Python 3.6.3. Machine learning classifiers, cross-validation, data pre-processing, parameters tuning was implemented by Scikit Learn package. The experiments were deployed on a Google Cloud Server, which has 8 CPUs and is capable of tuning parameters and conducting cross-validation in a timely manner.



**Figure 1.** The study design: We extracted patients' discharge summaries from MIMIC III database. Exclusion and inclusion criteria were applied to extract a cohort of patients with label of whether has a 30-day unplanned hospital readmission. We integrated 2 different data representation methods and 6 machine learning algorithms to generate predictive models. Area under receiver operating characteristic curve (AUC) was used to evaluate the model performance.



**Figure 2.** Readmission distribution. The distribution shows that there were 45305 subjects in the cohort, among which 2548, 5.62% subjects have 30-day unplanned ICU readmission.

## 4. Results and Analysis

### 4.1. Results

In this part, we will report the evaluation results of 5

supervised learning classifiers over 2 feature sets. From Table 1, we can see that except Naive Bayes as the benchmark, all other 4 classifiers yielded an AUC over 0.7. Logistic regression over Bag-of-CUIs had the best performance of AUC: 0.748, followed by LinearSVM over the same feature

yielded AUC: 0.744. The best performance of Bag-of-Words was 0.743, which is the 3rd place among all 10 configurations.

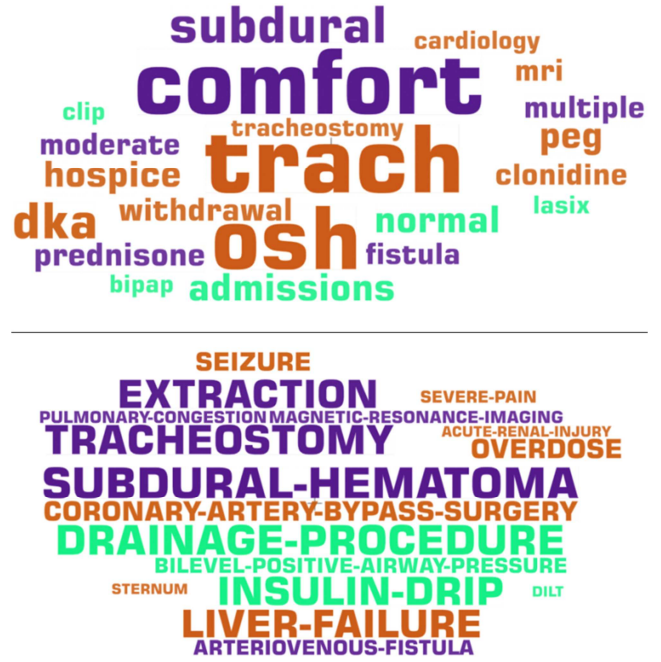
We further compared the performance of 2 features. Bag-of-CUIs beat BoW on 2 linear classifiers while lost on 2 ensemble classifiers. This might because that each word or phrase was mapped to multiple medical concepts in Metamap, which would also result in more features in Bag-of-CUIs than BoW. According to the experience of previous experiment, linear classifiers were better than ensemble classifiers at handling larger scale feature space. The performance of this study also validated that logistic regression and LinearSVM were the best classifiers when integrating Bag-of-words and machine learning to build clinical predictive models.

#### 4.2. Analysis

We further examined our models by presenting the top 20 features with greatest feature importance in logistic regression model over both BoW and Bag-of-CUIs as Figure 3. The font size of each feature is associated with its importance towards the models. Most of the words in both feature space was clinical meaningful. In Bag-of-Words, we could find "Tracheotomy" [19] (which is a surgical procedure which consists of making an incision on the anterior aspect of the neck and opening a direct airway through an incision in the trachea) and "fistula" [20] (an abnormal connection between organs). In Bag-of-CUIs, we could find "CORONARY-ARTERY-BYPASS-SURGERY" (which is a surgical procedure to restore normal blood flow to an obstructed coronary artery) and "SUBDURAL-HEMATOMA" [21] (A pool of blood between the brain and its outermost covering). All these words and concepts are capable to result in unplanned readmission due to severe condition or post-surgical morbidity. We also found some features emerged in both feature space. Such as "bipap" in BoW is the abbreviation of "BILEVEL-POSITIVE-AIRWAY-PRESSURE" in Bag-of-CUIs; and "fistula" in BoW is the second word of "ARTERIOVENOUS-FISTULA" in Bag-of-CUIs. These words and concepts presented high correlation with unplanned ICU readmission, can also remind clinicians that these subgroups of patients have a high possibility of readmission. Internal consistency of two feature sets also suggested that clinical meaningful keywords extracted from discharge summaries can build unplanned readmission predictive models and bag-of-words model was able to extract those keywords from discharge summaries.

**Table 1.** The performance of 5 machine learning classifiers on ICU readmission prediction using 2 different features as input.

|                                  | Bag-of-Words | Bag-of-CUIs |
|----------------------------------|--------------|-------------|
| Naive Bayes                      | 0.612        | 0.591       |
| LinearSVM                        | 0.743        | 0.744       |
| Logistic Regression              | 0.739        | 0.748       |
| Random Forest                    | 0.713        | 0.688       |
| Gradient Boosting Decision Trees | 0.721        | 0.704       |



**Figure 3.** Top 20 Words and CUIs have the greatest feature importance in the predictive model. Font size of each feature is associated with the importance. Upper: Bag-of-Words, Lower: Bag-of-CUIs.

#### 4.3. Limitation and Future Work

Obviously, there are limitations of our study that could be improved in the future work. Firstly, we only applied traditional machine learning classifiers other than deep neural networks. Recently, deep learning has shown its promising abilities in medical domain. We could apply 1-D convolutional neural network or recurrent neural network in the future experiment. Secondly, our dataset was limited. We only involved subjects from MIMICIII dataset. Future work on more generalized dataset could be done to make our predictive model more solid and robust. Thirdly, we only took discharge summaries into consideration. However, other clinical notes, such as radiology reports, physicians' diagnosis are also very essential to describe patients' condition within ICU stay. We are planning to involve more clinical notes under other categories. Finally, we only adopted Metamap as the phenotyping system to identify medical concepts from free text. Further experiments could be done with other system such as cTAKE, CLAMP [22] to examine whether Metamap is the most ideal tool in readmission prediction with discharge summaries.

## 5. Conclusion

In our study, we proposed and demonstrated that carefully representing of discharge summaries is useful to estimate 30-day unplanned ICU readmission. Bag-of-Words were applied to extract clinical meaningful information from discharge summaries. Metamap was further used to obtain unified medical term from free text and form another feature representation. 5 supervised learning classifiers were experimented on those 2 representations, respectively. Our

best result yielded an AUC of 0.748 when using logistic regression as classifier and Bag-of-CUIs as feature representation. Further analysis was conducted to validate that high contribution words and medical terms were clinical meaningful. Our model showed its strength of without relying on extra lab or radiology results to predict readmission status. Future work could be done by enlarge dataset to cover more subjects, enable more vectorization methods such as cTAKE and enrich classifiers to deep neural networks.

## References

- [1] Pronovost, P. J., et al., *Developing and pilot testing quality indicators in the intensive care unit*. Journal of critical care, 2003. 18 (3): p. 145-155.
- [2] Johnson, A. E., et al., *MIMIC-III, a freely accessible critical care database*. Scientific data, 2016. 3: p. 160035.
- [3] Higgins, T. L., et al., *Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III)*. Critical care medicine, 2007. 35 (3): p. 827-835.
- [4] Groenewegen, K. H., A. M. Schols, and E. F. Wouters, *Mortality and mortality-related factors after hospitalization for acute exacerbation of COPD*. Chest, 2003. 124 (2): p. 459-467.
- [5] Makris, N., et al., *Unplanned early readmission to the intensive care unit: a case-control study of patient, intensive care and ward-related factors*. Anaesthesia and intensive care, 2010. 38 (3): p. 723-731.
- [6] Tang, P. C., et al., *Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption*. Journal of the American Medical Informatics Association, 2006. 13 (2): p. 121-126.
- [7] Burton, L. C., G. F. Anderson, and I. W. Kues, *Using electronic health records to help coordinate care*. The Milbank Quarterly, 2004. 82 (3): p. 457-481.
- [8] Sox, H. C., et al., *Medical decision making*. 2007: ACP Press.
- [9] Bellazzi, R. and B. Zupan, *Predictive data mining in clinical medicine: current issues and guidelines*. International journal of medical informatics, 2008. 77 (2): p. 81-97.
- [10] Alić, B., L. Gurbeta, and A. Badnjević. *Machine learning techniques for classification of diabetes and cardiovascular diseases*. in *2017 6th Mediterranean Conference on Embedded Computing (MECO)*. 2017. IEEE.
- [11] Li, Y., et al. *Early prediction of acute kidney injury in critical care setting using clinical notes*. in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2018. IEEE.
- [12] Berenson, R.A., Pronovost, P.J. and Krumholz, H.M., 2013. Achieving the potential of health care performance measures. *Timely Anal Immed Health Pol*, (2013), p. 2.
- [13] Sundararaman, A., S. V. Ramanathan, and R. Thati, *Novel approach to predict hospital readmissions using feature selection from unstructured data with class imbalance*. Big data research, 2018. 13: p. 65-75.
- [14] Bardell, T., et al., *ICU readmission after cardiac surgery*. European journal of cardio-thoracic surgery, 2003. 23 (3): p. 354-359.
- [15] Vincent, J.-L., K. Donadello, and X. Schmit, *Biomarkers in the critically ill patient: C-reactive protein*. Critical care clinics, 2011. 27 (2): p. 241-251.
- [16] Willett, P., *The Porter stemming algorithm: then and now*. Program, 2006. 40 (3): p. 219-223.
- [17] Huang, Y.-F. and C.-H. Hsu, *PubMed smarter: Query expansion with implicit words based on gene ontology*. Knowledge-Based Systems, 2008. 21 (8): p. 927-933.
- [18] Aronson, A. R. *Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program*. in *Proceedings of the AMIA Symposium*. 2001. American Medical Informatics Association.
- [19] Cotton, R. T. and C. M. Myer III, *Contemporary surgical management of laryngeal stenosis in children*. American journal of otolaryngology, 1984. 5 (5): p. 360-368.
- [20] Maitland, C. G., *Perilymphatic fistula*. Current neurology and neuroscience reports, 2001. 1 (5): p. 486-491.
- [21] Frederickson, R. G., *The subdural space interpreted as a cellular layer of meninges*. The Anatomical Record, 1991. 230 (1): p. 38-51.
- [22] Zhang, Y., et al., *Psychiatric symptom recognition without labeled data using distributional representations of phrases and on-line knowledge*. Journal of biomedical informatics, 2017. 75: p. S129-S137.