Machine Learning, Clinical Notes and Knowledge Graphs for Early Prediction of Acute Kidney Injury in the Intensive Care

Iacopo VAGLIANO[a,1], Wei-Hsiang HSU[a] and Martijn C. SCHUT[a]

[a] *Dept. of Medical Informatics, Amsterdam UMC, Location AMC, The Netherlands*

Abstract. Acute kidney injury (AKI) is an abrupt decrease of kidney function which is common in the intensive care. Many AKI prediction models have been proposed, but an analysis of what is the added value of clinical notes and medical terminologies has not yet been conducted. We developed and internally validated a model to predict AKI that includes not only clinical variables, but also clinical notes and medical terminologies. Our results were overall good (AUROC > 0.80). The best model used only clinical variables (AUROC 0.899).

Keywords. Acute kidney injury, clinical models, ICU, natural language processing

1. Introduction

Acute kidney injury (AKI) is an abrupt decrease of kidney function, with a prevalence of up to 50% in the intensive care unit (ICU). Early recognition of AKI is crucial as the efficacy of intervention greatly depend on it. Several AKI prediction models have been built using clinical variables, e.g., vitals and laboratory measurements [2]. Few studies used the rich information contained in clinical notes, possibly because it is not well known how to represent such information in AKI prediction models.

Knowledge graphs may enrich the representation of clinical notes in prediction models. A knowledge graph is a graph-based abstraction of knowledge to represent data from diverse sources. Medical terminologies, which are inherently complex, are often expressed as knowledge graphs, for example, the Unified Medical Language System (UMLS) [5] and SNOMED CT [6]. To the best of our knowledge, one study has used notes and knowledge graphs to predict AKI, but without using clinical variables [7]. So, a comparison of which type of information among clinical variables, clinical notes, and medical terminologies is most effective in AKI prediction modeling is not yet available.

We aim to investigate whether adding information extracted from clinical notes and knowledge graphs into machine-learning prediction models to predict AKI can improve predictive performance. We developed and internally validated a model to predict AKI within the first 48 hours of admission, which includes not only clinical variables, but also clinical notes to provide a comprehensive view of the patients' pathophysiologic condition. Furthermore, we aim to study what is the effect of enriching notes with external knowledge from UMLS or SNOMED CT on the models' performance.

---

[1] Corresponding Author, Iacopo Vagliano, Dept. of Medical Informatics, Amsterdam UMC, Location AMC, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands; E-mail: i.vagliano@amsterdamumc.nl.

2. Method

Data and population: We used data from the publicly available critical care database, Medical Information Mart for Intensive Care III (MIMIC-III). This database integrates comprehensive clinical data of patients admitted to an ICU at the Beth Israel Deaconess Medical Center in Boston, during 2001 to 2012 [8]. AKI was defined according to the KDIGO guidelines [9]. We included patients who aged 18 years or older at the time of ICU admission, have at least one measurement of serum creatinine or urine output, and whose length of stay in the ICU was at least 48 hours. To make sure patients have both clinical variables and notes available for our model, only ICU stays containing at least one note were retained. Data preprocessing is illustrated in the supplementary material.[2]

Model development: Our AKI prediction model is based on Long Short Term Memory (LSTM) networks [10]. First, two independent LSTM models were built separately, one uses clinical notes as input and the other uses clinical variables. Then, the output of these two models were concatenated and given as input to a final layer to deliver an overall prediction which combines information from both clinical variables and clinical notes (the latter, optionally enriched with external information from knowledge graphs).

Clinical notes were represented with Word2vec. To enrich notes with external knowledge from UMLS and SNOMED CT, we extracted three subgraphs, i.e., UMLS synonyms, SNOMED CT synonyms, and SNOMED CT parent-child relationships. We used retrofitting to generate a refined representation of clinical notes, using relational information from each subgraph to encourage linked words to have similar representations. More details are available in Section 1.1 of the supplementary material.[2]
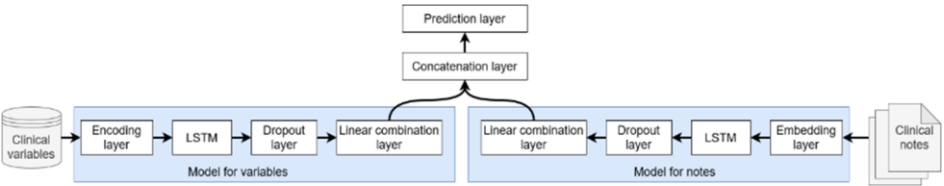

Figure 1. Architecture of the LSTM models used.

The architecture of the overall LSTM model is shown in Figure 1. We used the same architecture in the two LSTM models for variables and notes, except that the LSTM for clinical notes has an embedding layer, whilst the LSTM for clinical features has an encoding layer. The former layer acts as a lookup table to return the word embeddings learnt as parameters by the model during training, while the latter is a linear layer to compress the high-dimensional and sparse input variables into a lower-dimensional continuous representation, easier to manipulate by the model. The LSTM layers are followed by a dropout layer, to help prevent models from overfitting, and another linear layer, which creates a second learning filter for our model before concatenating the output of the two models, when relying on both clinical variables and clinical notes. The overall model ends with a projection layer that returns the probability of a patient having or developing AKI. The model parameters are shown in the supplementary material.[2]

Internal validation and performance measures: The dataset was randomly split into 80% training, 10% validation, and 10% test sets. We measured discrimination with the

---

[2] https://osf.io/thp53/, last access October 1, 2021.

area under the receiver operating curve (AUROC) and the area under the precision-recall curve (AUPRC); calibration with calibration curves.

3. Results

The final dataset consisted of 46,985 ICU stays of 33,795 unique patients. Descriptive statistics of the population are in Section 2.1 of the supplementary material.[2]

Table 1 outlines the discrimination (AUROC and AUPRC) of the models. Using only clinical variables achieved the best results. Exploiting clinical variables and notes performed better than only notes, and models with external knowledge achieved similar results to models without external knowledge. The calibration curves of the models are available in the supplementary material. Using clinical variables and notes retrofitted with parent-child relations yielded the best calibration, with a slight improvement on the models with only notes and with notes retrofitted with SNOMED CT synonyms.

Table 1. Models' discrimination with various inputs. We used six different input sets. Syn. stands for synonyms.

Input sets	Clinical variables	Clinical notes	Variables + notes	Variables + notes + UMLS syn.	Variables + notes + SNOMED syn.	Variables + notes + SNOMED parent-child
AUROC	**0.899**	0.801	0.821	0.821	0.819	0.816
AUPRC	**0.957**	0.898	0.910	0.910	0.909	0.908

4. Discussion

Our results shows overall good performances (AUROC > 0.80). The best-performing model (AUROC 0.90) used only clinical variables. Clinical notes retrofitted with parent-child relations used together with clinical variable yielded the best calibration.

When exploiting clinical notes, we used only 3,000 words to represent each patient, while the patients' notes included over 13,000 words on average. This might be the main reason why clinical notes did not improve the discrimination. Clinical notes may still be a valuable source of information on AKI prediction. We examined important words by ranking the most frequent words in the clinical notes (see the supplementary material[2]). In most cases, these top ranking words appear to be clinically meaningful. For example, *heparin* is a medication used to prevent blood clotting during kidney dialysis. *Edema*, which is a sign of AKI, happens when failing kidneys do not remove extra fluid which builds up in the patient's body causing swelling in the legs, ankles, feet, and/or hands. *Lasix* is one diuretic that can treat fluid retention and swelling, which might be caused by kidney dysfunction. With more advanced approaches to preprocess raw text from the notes, it is possible that structured clinical features and unstructured clinical notes exist as complementary sources of information for machine learning models to predict AKI.

Similar issues could explain why knowledge graphs did not affect the models' discrimination. For example in UMLS, only 7,093 out of 359,080 medical concepts (roughly 0.02%) are single-word. Clinical notes often refer to multi-words concepts, such as *end-stage chronic renal failure* or *end-stage kidney disease*. Since our word embeddings represent single words, multi-word concepts cannot be captured. The representation of multi-word concepts may improve models' performance.

Our study has some limitations. First, the MIMIC III database includes US patients; thus our results may not generalize to other populations. Second, we performed a simple

train/validation/test split, which does not take into account the variability of train/validation/test sets. Third, we did not perform full hyperparameter tuning but relied on a set of parameters for our models pre-selected in preliminary experiments.

As a strength, our study relies on a public dataset to encourage reproducibility and our code is available at bitbucket.org/aumc-kik/ml-cn-kg-4-aki-prediction.

Future work includes exploring more-advanced preprocessing of notes, ranging from using more words per patient to represents multi-word concepts, as well as study different models, such as convolutional neural networks. A prospective validation is also needed to assess clinical utility and effect on patient outcomes.

5. Conclusions

Early recognition of AKI is essential for effective treatment of this disease in the ICU. In contrast to previous work, we used various types of information, i.e., clinical variables, clinical notes and knowledge graphs, and we studied what is the added value of clinical notes and knowledge graphs to predictive performance. We provided effective models to predict AKI in the ICU. All the models achieved good results. The best discrimination was achieved by using only clinical variables, the best calibration with retrofitted clinical notes and clinical variables. Our work contributed to combining clinical variables, notes and knowledge graphs, which may also be useful in other settings and populations.

Acknowledgments

References

[1]   Hoste EA, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive care medicine. 2015 Aug;41(8):1411-23.
[2]   Park S, Lee H. Acute kidney injury prediction models: current concepts and future strategies. Current Opinion in Nephrology and Hypertension. 2019;28(6):552-559.
[3]   Sun M, et.al. Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes and Structured Multivariate Physiological Measurements. MedInfo. 2019 Aug 21;264:368-72.
[4]   Le S, et.al. Convolutional Neural Network Model for Intensive Care Unit Acute Kidney Injury Prediction. Kidney international reports. 2021 May 1;6(5):1289-98.
[5]   Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research, 2004;32(suppl_1):D267-D270.
[6]   Rogers J, Bodenreider O. SNOMED CT: Browsing the Browsers. In KR-MED, 2008:30-36.
[7]   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 Dec 3 (pp. 683-686). IEEE.
[8]   Johnson AE, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May 24;3(1):1-9.
[9]   Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron Clin. Pract. 120. 2012;c179–c184.
[10] Hochreiter S, et al.. Long short-term memory. Neural Computation, 9(8):1735–1780.
[11] Mikolov T, et al. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems 2013 (pp. 3111-3119).
[12] Faruqui M, et al. Retrofitting word vectors to semantic lexicons. In North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2015. (pp. 1606-1615). ACL