Author: Hermela Shimelis

NLP-Based Classification of ICU Discharge Summaries

ABSTRACT

Natural language processing (NLP) techniques can be used to extract information and analyze the unstructured text in electronic health records (EHRs). The objective of this project is to train deep learning (DL) models to classify free text from patient discharge summaries into one of 12 medical conditions relevant to intensive care units (ICU) re-hospitalization. A subset of the MIMIC-III data containing 1,610 discharge summaries that had been annotated with 12 medical conditions will be used for training DL models. The model developed here could serve as a benchmark for real-world practice using ICU discharge summaries to predict medical conditions related to recurrent readmissions to the ICU.

INTRODUCTION

Electronic health records (EHRs) are used to document data generated during patient care in hospitals. The data includes structured data, including diagnostic codes, tabulated lab results, as well as unstructured data, such as free text notes entered by care teams. The unstructured clinical note contains many valuable information, not effectively captured in the structured data. However, manual extraction of actionable clinical information from notes requires extensive amounts of time and is inefficient when the data volume is large. Automated extraction of clinical information from notes in the EHR can be achieved through processing of the large volume of unstructured text using machine learning techniques. Natural language processing (NLP) techniques can be used to convert free text into formats that can be used to train machine learning models for various tasks, including text classification, and content filtering. Deep learning (DL) models have been employed for disease classification of clinical notes, however significant improvement remains in at least two areas, 1) performance of current state of the art models show modest accuracy and 2) explainability, especially in the medical field, identification of informative text is important for determining use of accurate text snippet for classification of text. The objective of this project is to train explainable deep learning models to classify free text from 1,610 discharge summaries into one of 12 medical conditions.

BACKGROUNDS

A record amount of healthcare data documented in the EHR is continuously generated as part of patient care. This data has characteristic big data features as it includes variety of data, including structured and unstructured data, it is very large in volume, has a high veracity because it includes many manually entered erroneous or low-quality data; and it is a high velocity data as new information is generated every time patients visit the hospital. It is estimated that a single patient generates about 80 megabytes of data every year (*7*).

Up to 80% of data in the EHR, including discharge summaries, radiology reports , electrocardiograms, and medical videos are unstructured (*8*). This unstructured data can be used to extract many clinically actionable insights. For example, leveraging NLP and deep learning models, large amounts of free text clinical notes can be used for prediction of disease (*3*, *5*, *6*), length of hospitalization and mortality (*1*), enabling clinicians to identify patients in need of early intervention or change in management. Although recent deep learning architectures have shown promise with automated extraction of valuable information from free text notes, the above mentioned machine learning problems are challenging for a number of reasons, including long input text, sparsity, and computational cost (*3*, *6*).

Recent advancement in deep learning models for extraction of data from EHR were made partly due to the freely accessible repository of medical records, The Medical Information Mart for Intensive Care III (MIMIC-III) data, cited in over 3,500 articles (*9*–*11*). The database contains de-identified clinical data, including discharge summary from over 53,000 hospital admissions for adult patients to the intensive care units (ICU) at the Beth Israel Deaconess Medical Center from 2001 to 2012. The database also includes demographics, lab results, procedures, and medication history. The MIMIC-III database can be accessed through either Google Cloud Platform (GCP) or Amazon Web Services (AWS), which makes further downstream analysis of the data using big data tools such as BigQuery.

The current proposal focuses on classification of text from discharge summaries into one of 12 medical conditions that had been manually annotated by clinicians (*12*). Therefore, relevant deep learning models for classification of disease using MIMIC-III discharge summary and clinical notes are discussed here. A number of studies have used the MIMIC-III discharge summaries to extract International Classification of Diseases (ICD)-9 (ICD-9) codes from discharge summaries, essentially treating ICD-9 code prediction as a multiclass text classification, with about 8,900 codes for each medical condition included as labels (*2*, *4*–*6*). Meanwhile, a few studies have reported deep learning models for classification of selected (3-16) disease conditions , instead of all diseases (*1*, *3*). The latter approach requires availability of labeled data, which is labor insensitive for big data applications, consistent with the limited availability of annotated text data. Lu et.al., 2022 reported a comparative study on deep learning models for text classification into 16 medical conditions. The machine learning problem was analyzed as a 16 binary separate classification problem (*3*). The authors compared 6 deep learning models, including Convolutional Neural Network (CNN), a Transformer encoder, a pretrained BERT (Bidirectional Encoder Representations from Transformers), and four typical sequence neural networks models, namely, RNN (Recurrent Neural Network), GRU (Gated Recurrent Unit), LSTM (Long Short-Term Memory), and Bi-LSTM (Bi-directional Long Short-Term Memory) (*3*). In nearly all classification problems, the CNN model outperformed the other models, and had the least trainable parameters and training time (*3*), consistent with computational efficiency of 1-dimensional CNN models for text and time series classification. Similarly, the above-mentioned models, including RNN, LSTM, bi-directional LSTM and BERT have been used for ICD-9 multiclass classification of discharge summaries (*2*). In the ICD-9 multiclass (~8,900 labels) classification, BERT showed higher performance than the RNN based models; however, this model required high computation demand (*2*). Additionally, both the RNN based models and BERT are black box models and lack explainability.

Deep learning models with less computational demand, better performance and explainability have been developed for medical text classification. Several studies have used the MIMIC-III text data for ICD-9 disease classification. Mullenbach et al., developed a model combining CNN and attention mechanism to select part of the text that are most relevant for classification (*12*). The Convolutional Attention for Multi-Label classification (CAML) model achieved a macro AUC of 0.896 for 8,922 ICD-9 prediction. The text snippet's importance for the classification can be determined, hence, the model can explain why it predicted each code (*12*). Hu et al., reported a similar CNN-based model, Shallow and Wide Attention convolutional Mechanism (SWAM) for multiclass ICD-9 classification of discharge summaries. The best model, SWAM-CAML model, showed macro AUC 0.90 (*4*). Lin et al., reported another CNN-based model, Multi-Head Label decoding (MHLD), for ICD-9 disease classification of admission and discharge summaries. This model showed slightly improved performance, with macro AUC of 0.923 (*5*). Overall, the CNN-based models with attention mechanism are explainable and have better model performance, compared to the RNN-based models (*11*).

These CNN-based models with attention mechanism discussed above are explainable, computationally less intensive, compared to the RNN-based models, and showed better model performance. Therefore, the best model combining CNN and attention mechainsm, MHLD, will be used for the current multilabel text classification problem.

METHODS

Dataset

Discharge summaries for 12 medical condition classification will be extracted from the MIMIC-III database (*7–9*) which is freely accessible through either Google Cloud

Platform (GCP) or Amazon Web Services (AWS). The database contains de-identified clinical data, including discharge summary from over 53,000 hospital admissions. This big EHR data has been stored as a relational database with 26 tables in Amazon S3 cloud storage. Each table is linked using subject ID as a key identifier. Therefore, the acquisition of the big data has been handled by the MIMIC-III database. The 1,610 subject IDs of discharge summaries that have been annotated with 12 disease labels are separately provided as a CSV file on Physionet platform that host MIMIC-III data, and they can be accessed via the link: https://www.physionet.org/content/phenotype-annotations-mimic/1.20.03/. The corresponding discharge summaries are located on the MIMIC-III database, and they are accessed via GCP. For extraction of discharge summaries and preprocessing is carried out using BigQuery, implemented in GCP. Text date preprocessing and exploratory data analysis will also be performed using BigQuery.

Deep learning model development

CNN-based model, Multi-Head Label decoding (MHLD) model, will be trained in bigquery. The selected model is appropriate for addressing the research objective of training multiclass classification models that is explainable. If the model performance is not satisfactory, or they are computationally costly, 1-D CNN will be used as a base model; since CNN, without attention mechanism, showed reasonable performance in previous studies. One limitation of the proposed model is that the above studies were employed for multiclass ICD-9 classification, with over 8,000 labels, whereas the current classification problem involves only 12 classes. Therefore, the model performance may not be similar to the published studies.

REFERENCES

1. hhit, Medical record data storage & retrieval, *Healthcare Data Management Software & Services | Harmony Healthcare IT* (2020). https://www.harmonyhit.com/health-data-volumes-skyrocket-legacy-data-archives-rise-hie/.

2. H.-J. Kong, Managing unstructured big data in healthcare system. *Healthc. Inform. Res.* **25**, 1–2 (2019).

3. H. Lu, L. Ehwerhemuepha, C. Rakovski, A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. *BMC Med. Res. Methodol.* **22**, 181 (2022).

4. S. Hu, F. Teng, L. Huang, J. Yan, H. Zhang, An explainable CNN approach for medical codes prediction from clinical text. *BMC Med. Inform. Decis. Mak.* **21**, 256 (2021).

5. Y.-T. Lin, S.-L. Wei, H.-H. Huang, H.-C. Wang, H.-H. Chen, “Disease classification on admission and on discharge with residual CNN-transformer” in *IEEE/WIC/ACM International Conference on Web Intelligence* (ACM, New York, NY, USA, 2021; https://dl.acm.org/doi/10.1145/3486622.3493946).

6. J. Zhang, L. D. Bolanos Trujillo, A. Tanwar, J. Ive, V. Gupta, Y. Guo, Clinical utility of automatic phenotype annotation in unstructured clinical notes: intensive care unit use. *BMJ Health Care Inform.* **29**, e100519 (2022).

7. A. Johnson, T. Pollard, R. Mark, MIMIC-III clinical database, PhysioNet (2023); https://doi.org/10.13026/C2XW26.

8. A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, R. G. Mark, MIMIC-III, a freely accessible critical care database. *Sci. Data* **3**, 160035 (2016).

9. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, H. E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**, E215-20 (2000).

10. E. Moseley, L. A. Celi, J. Wu, F. Dernoncourt, Phenotype annotations for patient notes in the MIMIC-III database, PhysioNet (2020); https://doi.org/10.13026/TXMT-8M40.

11. I. Aden, C. H. T. Child, C. C. Reyes-Aldasoro, International Classification of Diseases prediction from MIMIIC-III clinical text using pre-trained ClinicalBERT and NLP deep learning models achieving state of the art. *Big Data Cogn. Comput.* **8**, 47 (2024).

12. J. Mullenbach, S. Wiegreffe, J. Duke, J. Sun, J. Eisenstein, Explainable prediction of medical codes from clinical text, *arXiv [cs.CL]* (2018). http://arxiv.org/abs/1802.05695.