Assign ICD Coding to Discharge Summaries Utilizing NLP

Min Zhong

Dec. 2022

Abstract

Current ICD code assignment to clinical discharge summaries is a manual process which is time-consuming, high in labor cost and easy to make mistakes. The objective of this project is to develop an automated ICD coding system for easy and fast code assignments. To achieve the objective, three models were trained and compared for ICD coding using natural language processing. Among the three models, Attention-Based Convolutional Neural Network (CNN-ATT) achieved the highest F1-score for both validation and testing datasets. The trained models were implemented in a web application which allows health providers for easy and fast code assignments with user-provided discharge summaries. Further work is needed to deploy and configure the web application to the internet.

1. Background

Unstructured clinical text such as discharge summaries contains valuable information for clinical decision making. To carry out meaningful analysis, these medical records need to be converted into clinical codes. The International Classification of Diseases (ICD) is the most widely used medical classification list worldwide. The 9th revision ICD-9 was introduced in 1979 and its Clinical Modification (ICD-9-CM) is still the official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States (CDC, 2022). From 1999 to present, the 10th revision ICD-10 has been used in many countries to classify and code diagnoses, symptoms and procedures in all healthcare settings. The next generation of classification ICD-11 was released in 2019 but it is not yet implemented in hospital settings.

Currently, ICD code assignment to discharge summaries is a manual process which is time-consuming, high in labor cost and easy to make mistakes. With the transition from ICD-9 to ICD-10, the number of codes increased 18 times from approximately 3300 to 70,000 codes (Subotin & Davis, 2014), which makes manual coding a non-trivial task and could be erroneous. Moreover, the continuous evaluation of coding rules makes the manual assignment an even complex process. For example, the coming ICD-11 increases the complexity by introducing a new code structure and multiple new chapters.

2. Problem Statement

To overcome manual coding challenges and facilitate clinical coding assignments, there is a need for an automated ICD coding system. Several studies have explored and developed the automated ICD coding system using traditional machine learning or deep learning (Perotte, et al., 2014; Berndorfer & Henriksson, 2017; Li & Yu, 2020; Hsu, Chang, & Chang, 2020). In general, the automated coding system starts with preprocessing to remove unwanted information from the clinic text. Next, feature extraction is applied to obtain useful features from clinical reports. Then the discharge summaries are classified either using machine learning or deep learning models. Finally, model is assessed and selected based on evaluation metrics. Previous studies are primarily based on the Medical Information Mart for Intensive Care (MIMIC)-II or MIMIC-III dataset. Most previous studies focused on model training and development, no end user platform was developed and reported.

3. Solution

This project aims to develop a web application for ICD coding system that utilizes natural language processing to assign ICD-9 code for discharge summaries. This tool is expected to be used by health providers for easy and fast code assignments. The overall architecture diagram of this project is shown in Fig. 1.

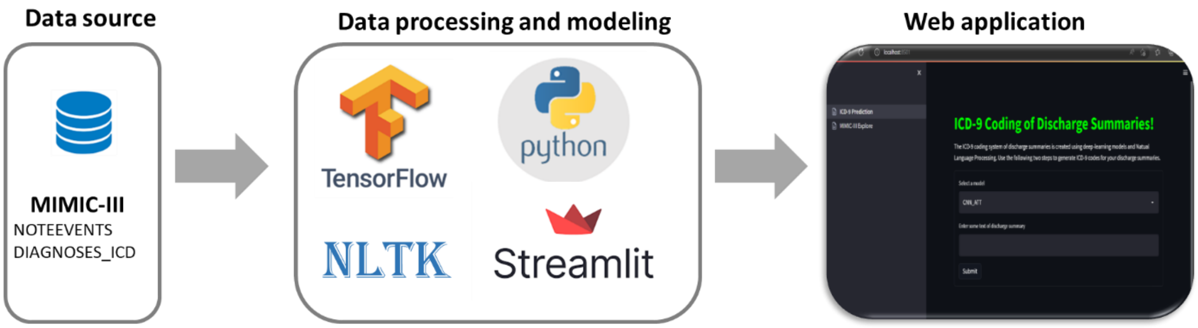

Figure 1. Architecture diagram

3.1 Data Source and Preprocess

Two data tables that contain discharge summaries and diagnoses ICD were obtained from the PhysioNet website. https://physionet.org/content/mimiciii/1.4/.

The following preprocessing was applied to the raw text of MIMIC-III discharge summaries: tokenize notes, remove punctuation and numeric-only token, lowercase all tokens, create training (70%), validation (15%) and test (15%) splits. Most of samples are concentrated in top 50 ICD-9 codes. Therefore, this work will focus on top 50 ICD-9 codes.

3.2 ICD-9 Coding Approaches

Three models were trained for ICD-9 coding: logistic regression, Convolutional Neural Network (CNN), and CNN-ATT. Since the dependent variable contains 50 categorical classes (top 50 ICD-9 codes), the F1 score is used as a measure to compare the performance of these three models. Python package, Tensorflow, was used for model training and evaluation.

3.3 Web Application

A web application was developed using the Python Streamlit package. The web application contains two pages: **MIMIC-III data explore, and ICD-9 code prediction.** The first page provides a few summary details of the MIMIC-III dataset and the performance of the models. Users have the choice to view plots. The second page is designed to predict ICD-9 coding with user's input of discharge summary texts. Users have the option to choose one of three models for ICD-9 coding.

4. Outcome and Further Work

Table 1 shows the F1 score of three models for three datasets: training, validation, and testing. CNN-ATT has the highest F1 score for validation and testing datasets.

Table 1. F1-score of ICD-9 coding on three datasets

Model	Threshold	F1 score		
		Training	Validation	Testing
Logistic regression	0.14	0.99	0.48	0.48
CNN	0.49	0.67	0.52	0.51
CNN-ATT	0.32	0.77	0.63	0.62

Although CNN-ATT is the best model, all three models were included in the web application. Figure 2 shows a screenshot of this ICD-9 coding application.

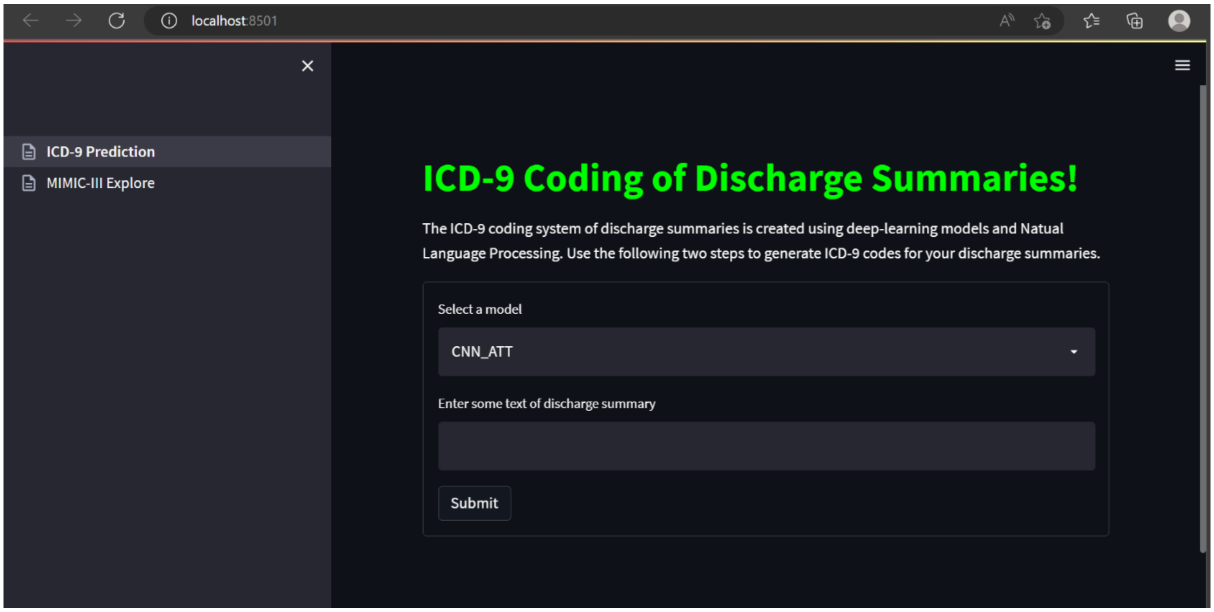

Figure 2. Screenshot of the ICD-9 coding application

This project successfully developed three ICD-9 coding models and created a web application for health providers for easy and fast code assignments. Future work is recommended to improve model performance using different data preprocessing approach and modeling algorithms. Currently, the web application can only be used in local environment. More effort is needed to deploy and configure the web application to the internet. In addition, the appearance and user interface of this web application can be further improved.

References

Berndorfer, S., & Henriksson, A. (2017). Automated Diagnosis Coding with Combined Text Representations. *Stud Health Technol Inform.*, 201-205.

CDC. (2022). *National Center for Health Statistics*. Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/nchs/icd/icd9cm.htm

Hsu, C.-C., Chang, P.-C., & Chang, A. (2020). Multi-Label Classification of ICD Coding Using Deep Learning. *2020 International Symposium on Community-centric Systems (CcS)* (pp. 1-6). doi: 10.1109/CcS49175.2020.9231498.

Li, F., & Yu, H. (2020). ICD Coding from Clinical Text Using Multi-Filter Residual Convolutional Neural Network. *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 8180-8187). https://doi.org/10.1609/aaai.v34i05.6331.

Perotte, A., Pivovarov, R., Natarajan, K., Weiskopf, N., Wood, F., & Elhadad, N. (2014). Diagnosis code assignment: models and evaluation metrics. *Journal of the American Medical Informatics Association : JAMIA*, 231–237.

Subotin, M., & Davis, A. (2014). A System for Predicting ICD-10-PCS Codes from Electronic Health Records. *Proceedings of BioNLP 2014* (pp. 59--67). Baltimore, Maryland: Association for Computational Linguistics.