DEEP LEARNING AND NLP APPROACH FOR ICD-9 RECOMMENDATION SYSTEM

Muthukumar Subramani[1], Nithishkumar K R[2]

[1] Assistant Professor,Dept of IT,MepcoSchlenk Engineering College,Sivakasi,Tamilnadu-626005,India
[2] UG Scholar,Dept of IT,MepcoSchlenkEngineering College,Sivakasi,Tamilnadu-626005,India

Received: 14 March 2020 Revised and Accepted: 8 July 2020

ABSTRACT: Healthcare becomes one of the largest sectors in the world. It involves both revenue and employment. It comprises of hospitals, medical devices, telemedicine, health insurance, diagnostic report etc.

ICD stands for the International Classification of Diseases listed by the World Health Organization (WHO), and its codes hold critical information about epidemiology, managing health, and treating conditions. ICD-9 is the Ninth version of ICD Code and ICD-10 is the 10thversion of ICD code. Both are widely used to describe patient diagnosis.We found that the convolution is the most effective component in our work. Experimental results show that convolution neural network combined with NLP approach improves the performance.

KEYWORDS: ICD-9, Convolution Neural Network, Natural Language Processing.

I. INTRODUCTION

The patient suffers from any health issue will be provided with the ICD code. The ICD code may vary by their version to version. In this paper, the ICD-9 Code[1][2]has beenused. This ICD code will be used in reporting the diseases and health conditions, assisting medical reimbursement decisions and collecting morbidity and mortality and used for applying for medical insurance. Medical coders are the person who will represent the ICD Codes[8] using the diagnostic report of the patient. Since it is a manual, coding it becomes more expensive, time-consuming and inefficient. Recently, deep learning becomes a great success.

Since due to lack of medical coders, the hospital staff used to do code that May results wrong.Therefore, In this study, we have used MIMIC III v1.4 dataset[3] with convolution neural network for the prediction of ICD codes [4] of patient's report. We use CNN model after preprocessing with NLP approach[5] on the medical data to have the best feature on the text . We build three models for the prediction of ICD Codes i.e. the top10, top11-20 and top50 codes on it. Initially, the preprocessing done with the discharge summary of the ICU patients found in that dataset. That relatively collects the report of 52, 962 documents with 6,984 ICD Codes. More than 32,000 reports the top 10 and top 20 ICD Codes were present. There are 55091 documents in Top 50 ICD codes. Applied the CNN Model on it and deployed as GUI Application using flask framework in python.

II. RELATED WORKS

2.1 NLP (Natural Language Processing)

The researcher predominantly addressed through NLP methods, which provides a good outcome. To achieve the automated ICD-9 coding, the paper[1] proposed a methodology by using electronic health record data, whereby raw clinical data is mapped into a feature set, and based on which supervised learning algorithms are trained.The system[6] implemented by using a messaging application with auxiliary Natural Language Processing (NLP) [5] library. The system was compared with the conventional ICD-10 application by using the Analytic Hierarchy Process (AHP).

2.2Deep Learning Approach

Deep Labeler that combines both CNN[15] and D2V for the prediction on the MIMIC III and MIMIC-II dataset. The direct CNN model with the D2V[13] part combined to predict the ICD-9 codes. The NLP approach can improve on those medical data with greater accuracy than the present. Convolution Neural Network and the preprocessed data with NLP makes better result

III. OVERVIEW OF THE PROPOSED WORK

convolution neural networkis widely employed in various Natural Language Processing (NLP) problems and provides promising results in sentence modeling and sentence classification[14] problems.In this work, we use for the prediction of ICD codes.

3.1 DATASET

The datasetMIMIC-III (http://mimic.physionet.org/)[3], which comprises over 58,000 hospital admissions. Each patient in the MIMIC dataset is having a list of ICD-9 codes which are use as labels in the experiments. Each code represents category of disease, diagnostic or symptom, injury or a treatment procedure. Table 1 illustrates the statistical information of the dataset.

Total discharge summary Reports	52,962
Total ICD Codes in the Reports	6,984
The average number of words per discharge summary	1,524
Average Code per patient	11
Maximum number of words per discharge summary	7,980
Maximum number of codes per patient	39
Minimum number of words per discharge summary	9
Minimum number of codes per patient	1

Table 1MIMIC III Dataset statistics

3.2 PROPOSED METHOD

For collecting the data, that is the report of a patient. MIMIC III v1.4 contains many other data too in it. We have considered two files,

1. NOTEEVENTS.CSV
2. DIAGNOSES_ICD.CSV

In the NOTEEVENTS.csv contains the patient details like identity through an id number, Category in the patient is under admission and the medical text.In the DIAGNOSES_ICD.csv contains the identification number and the corresponding equivalent ICD-9 Codes.We group by with their category as discharge summary and we got 59,652 records with equivalent with ICD-9 Code. A maximum ICD-9 code to a single patient is 39 and the minimum is one.Then process each text with top10_text.csv that contains the top10 ICD Code text with their equivalent ICD-9 Code and a similar way for the top20_text.csv file. Cleaned each text by removing numbers, special characters removed the stop words through NLTK library and done lemmatization.Vectorized the ICD-9 code associated with each text as one-hot encoding array to supply in the CNN model. Also vectorized the report text with each unique word points to a value. We used GloVe (Global Vectors for Word Representation) for creating the embedding matrix with the dimension of 100.

After this process, we split the data using train_val_test_splitfunction, which uses the train_test_split function from sklearn. Then implemented the CNN model with all data to be trained with 5 epochs. So taking the top 10, top 20, total top 50 data we reach the good accuracy of 87.4 % and 91.3 %, 93% respectively after running with the CNN model.

3.3OVERALL WORKFLOW

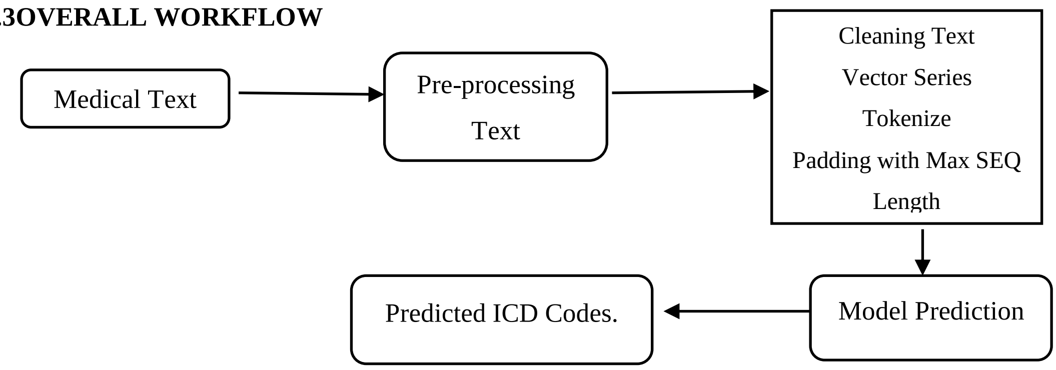

Fig 3.1 Workflow

Figure 3.1 Show the workflow in the deployment stage. Provided with sample data can be loaded from the sample data from the local system and also provided the GUI to take the input as the medical report to produce the respective ICD Code under the top codes.

3.4 DEEP CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

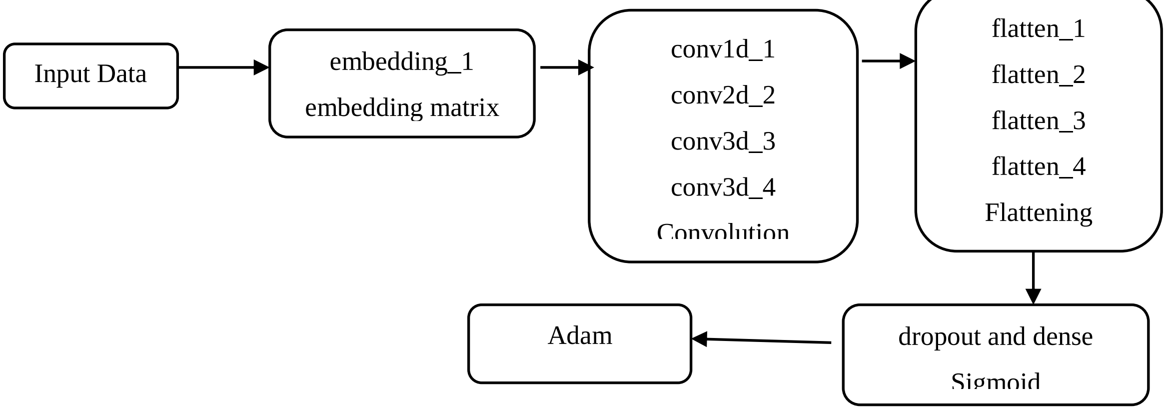

Fig 3.2 Deep CNN Architecture

Fig 3.2 shows Deep CNN Architecture with all layers by applying convolving techniques.This work implements a CNN for text classification.A convolution is a combined integration of two functions that shows you how one function modifies the other.Once the pooled featured map is obtained, the next step is to flatten it. Flattening involves transforming the entire pooled feature map matrix into a single column, whichis then fed to the neural network for processing. Max pooling works by placing a matrix of 2x2 on the feature map and picking the largest value in that box. The 2x2 matrix is moved from left to right through the entire feature map picking the largest value in each pass. We developed the deep learning[4][7] that can be used to classify based the ICD codes based on the number of classes[12] in it.

IV. IMPLEMENTATION DETAILS

4.1 NLP

The medical text will have a large volume of data regarding the patient's report. Thus this need to be pre-processed to reduce the noise in the medical text like removing the date, etc. Then vectorize the text present in the document and are tokenized, subsequently padded with the maximum sequence length to maintain the same shape[9]. Finally, the text is predicted with the model.

4.2 CONVOLUTIONAL NEURAL NETWORK AND NLP

A CNN consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolution layers, pooling layers and normalization layers.Convolution layers apply a convolution operation to the input, passing the result to the next layer. Max pooling layers uses the maximum value from each of cluster neurons at the prior layer, average pooling layer uses the average value from each of a cluster of the neuron at the prior layer. The score calculated as follows,

Micro Precision $MiP = \frac{\sum_{m=1}^{M}\sum_{i=1}^{N} y_i^m \hat{y}_i^m}{\sum_{m=1}^{M}\sum_{i=1}^{N} \hat{y}_i^m}$

Micro Recall $MiR = \frac{\sum_{m=1}^{M}\sum_{i=1}^{N} y_i^m \hat{y}_i^m}{\sum_{m=1}^{M}\sum_{i=1}^{N} y_i^m}$

$$Micro\ F - measure = \frac{2 \times MiP \times MiR}{MiP + MiR}$$

Precision as fraction of true positive among all the positive’s recalled. Recall as fraction of true positives among all the correct events. F-Score (Balanced) as the harmonic mean of the Precision and Recall.

4.3 DEPLOYMENT OF MODEL IN FLASK

Once the models trained, downloaded for the deployment. Here, Flask is a web application framework used in python and a webapp is developed. For the implementation of the backend, it uses Python code and uses the model for doing the prediction on the given report. The equivalent ICD-9 code can be changed to ICD-10 code by using web scraped data from the internet using the requests and bs4. In this project, we used https://www.icd10data.com/Convert/(icd-code) to get the equivalent ICD-10 [6] code for the present ICD-9 code from the present result. Figure 4.2 shows the result on the web app with selection of top-10 model.

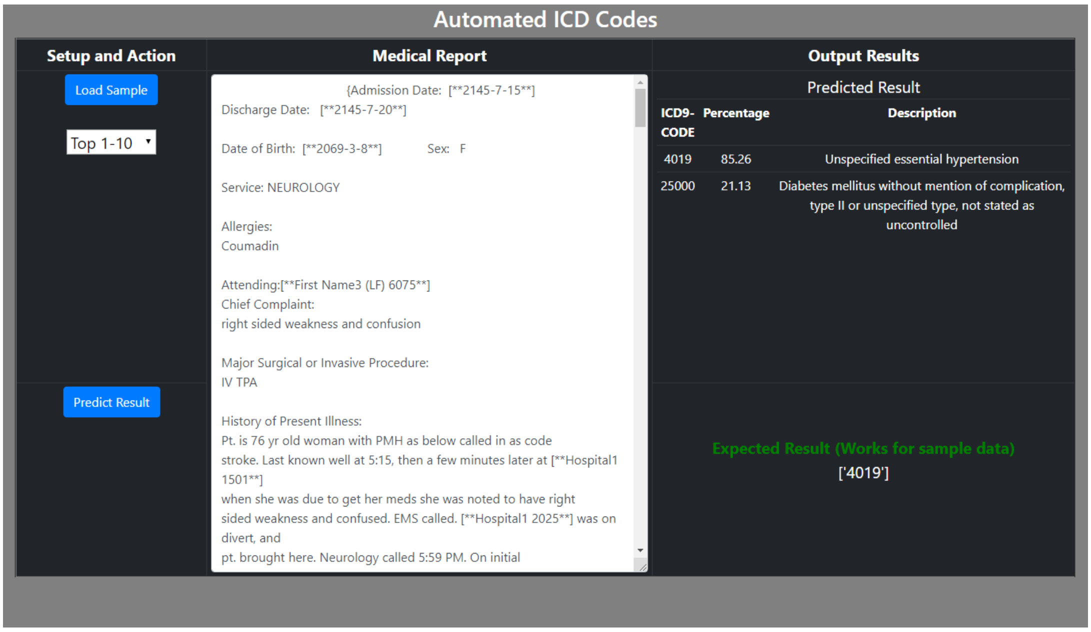

Fig 4.2 Web app result

V. EXPERIMENTAL RESULTS

In this work, we have done experiments with three different models on three different set of classes[12]. |It takes each class on the ICD-9 code on top priority. On training the model, it produces the accuracy on 87.4 % on the first model to shows up with top 1-10 ICD-9 codes and the other model takes 91.3 % accuracy on it and for top 50 codes, it takes 93% accuracy. It improves some scores in another model after having the limited codes to a patient. It works well for the top 1-10 model; it gives out the promising result as expected as we can see in the fig 7.1 GUI output. When we train our model with lot amount data it, outcomes will be great at work. When we train with full text on the data it will consume a large amount of time during the process than training with sample result. We take out the ICD-10 data from the ICD-9 code that we get in the result through the internet from the conversion of ICD-9 to ICD-10.The f1-score we got with training on the validation set and testing data with top10 ICD-9 Codes model i.e. for model-1

Threshold	Training	Validation	Testing
0.4	81.2%	77.4%	77.0%
0.5	80.4%	76.0%	75.9%

Table 5.1 F1-Score Result for the Model-1

The f1-score we got with training on the validation set and testing data with top11 - 20 ICD-9 Codes model i.e. for model-2.

Threshold	Training	Validation	Testing
0.4	72.5%	67.2%	65.1%
0.5	69.2%	64.2%	62.4%

Table 5.2 F1-Score Result for the Model-2

The f1-score we got with training on the validation set and testing data with top50 ICD-9 model Codes i.e. for model-3.

Threshold	Training	Validation	Testing
0.4	72.5%	67.2%	65.1%
0.5	69.2%	64.2%	62.4%

Table 5.3 F1-Score Result for the Model-3

VI. CONCLUSION

In this project, we have presented a Deep Learning with NLP based approach for automating the ICD code. We converted the ICD-9 Code to ICD-10 code, from the freely available MIMIC III v1.4 dataset. Health domain is an important sector around the globe so performing the work should provide accurate results on it. Therefore, we planned to choose the top data i.e. Top ICD codes to make models and deployed it in a web application.

VII. REFERENCES

[1] M. Li et al., "Automated ICD-9 Coding via A Deep Learning Approach," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 4, pp. 1193-1202, 1 July-Aug. 2019

[2] J. C. Ferrão, F. Janela, M. D. Oliveira and H. M. G. Martins, "Using Structured EHR Data and SVM to Support ICD-9-CM Coding," 2013 IEEE International Conference on Healthcare Informatics, Philadelphia, PA, 2013, pp. 511-516

[3] Johnson, Alistair & Pollard, Tom & Shen, Lu & Lehman, Li-wei& Feng, Mengling&Ghassemi, Mohammad & Moody, Benjamin &Szolovits, Peter &Celi, Leo & Mark, Roger. (2016). MIMIC- III, a freely accessible critical care database. Scientific Data. 3. 160035. 10.1038/sdata.2016.35.

[4] Medori, Julia &Fairon, Cédrick. (2010). Machine learning and features selection for semi-automatic ICD-9-CM encoding. 84-89.

[5] Pons, Ewoud & Braun, Loes&Hunink, Myriam& Kors, Jan. (2016). Natural Language Processing in Radiology: A Systematic Review. Radiology. 279. 329-343. 10.1148/radiol.16142770.

[6] N. Siangchin and T. Samanchuen, "Chatbot Implementation for ICD-10 Recommendation System," 2019 International Conference on Engineering, Science, and Industrial Applications (ICESI), Tokyo, Japan, 2019, pp. 1-6

[7] Medori, Julia &Fairon, Cédrick. (2010). Machine learning and features selection for semi-automatic ICD-9-CM encoding. 84-89.

[8] Zenobia Liendo, Guillaume De Roo, AmitabhaKarmakar "Classifying medical notes into standard disease codes"

[9] M. R. Hasan, M. Maliha and M. Arifuzzaman, "Sentiment Analysis with NLP on Twitter Data," 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2019

[10] YunZhi Chen, HuiJuan Lu, LanJuan Li "Automatic ICD-10 coding algorithm using an improved longest common subsequence based on semantic similarity"

[11] I. Goldstein, A. Arzumtsyan, and € O. Uzuner, "Three approaches to automatic assignment of ICD-9-CM codes to radiology reports," in Proc. Amer. Med. Informat. Assoc. Annu. Symp., 2007,Art. no. 279.

[12] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in Data Mining and Knowledge Discovery Handbook. Berlin, Germany: Springer, 2009, pp. 667–685

[13] J. H. Lau and T. Baldwin, "An empirical evaluation ofdoc2vec with practical insights into document embedding generation," in Proc. 1st Workshop Representation Learning NLP,2016, pp. 78–86.

[14] Z. Yang, D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, "Hierarchical attention networks for document classification," in Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Technol., 2016, pp. 1480–1489.

[15] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T.-S. Chua,"Disease inference from health-related questions via sparse deep learning," IEEE Trans. Knowl. Data Eng., vol. 27, no. 8, pp. 2107–2119, Aug. 2015.