IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116120

Diagnosis Prediction using Ensemble Learning

Parashiva Murthy B M [1] Akshar S Ramesh[2], Anurag Anand Vaidya[3], Varaprasad S[4], Yashas S[5]

Assistant Professor, Computer Science & Engineering, JSS Science & Technology University, Mysuru, India[1]

Student, Computer Science & Engineering, JSS Science & Technology University, Mysuru, India[2-5]

Abstract: The data in healthcare has increased in volume, intricacy and comprehensiveness. This growth leads to extensive application of artificial intelligence and machine learning in the healthcare sector. This study aims to examine the application of deep learning models and ensemble learning in diagnosis prediction. We apply Natural Language Processing techniques on medical notes to predict diagnosis. Real-life healthcare datasets, like MIMIC-2, contain tables with medical notes which can be pre-processed and used to train ML models. This paper presents an analysis of a diagnosis prediction algorithm. This facilitates the creation of autonomous medical systems which can be used to aid or act in place of healthcare professionals.

Keywords: MIMIC 2, EHR, Clinical notes, NLP, Bidirectional LSTM, BERT, Ensemble learning.

I INTRODUCTION

The ICU generates a huge quantity of data. This data has enormous promise since it can be utilized to create valuable decisions in a health care setup. It has the ability to improve healthcare by assisting clinicians such as doctors and other assisting staff in making more efficient clinical choices and optimizing hospital operations. Hospitals are subject to multiple pressures, including limited funds and healthcare resources. The intensive care unit (ICU) in particular has drawn considerable attention from the medical community due to its critically ill patients and costly resources. The ICU patient is highly monitored using electronic equipment to measure physiological data, which provides a rich opportunity for valuable clinical data analysis.

Electronic Health Records contain patient information such as symptoms, primary complaints, medical and pharmaceutical history, treatment, procedures and tests, initial diagnosis, final diagnosis, discharge medication, and care notes or referral notes. Medical Information Mart for Intensive Care (MIMIC) database is an open source, credentialed access Electronic Health Record. It provides a huge amount of information about patients staying in ICU. MIMIC contains data in 2 forms: structured and unstructured data. We are concerned with unstructured medical notes for diagnosis prediction. The version of MIMIC we are focussing on is MIMIC-2. We focus on the table note_events.csv which contains medical notes.

Deep Learning and Natural Language Processing in Healthcare

In this paper, we try to use the advances in Natural Language Processing and Deep Learning to extract knowledge from the unstructured medical notes. This knowledge can be used to predict diagnosis. In a clinical setup, this facilitates the creation of autonomous medical systems which can be used to aid or act in place of healthcare professionals.

Single concept extraction is used to extract critical information from clinical writing such as diagnoses, treatments, and procedures. Many researchers have used NLP approaches to match clinical notes with the top ten International Classification of Diseases codes, with varying degrees of success. However, given the intricacy of the clinical notes, there is tremendous space for improvement. In this study, we examine Single idea extraction (ie diagnosis) from clinical language using a cutting-edge NLP deep learning technique.

II METHODOLOGY

Architecture

In our study, we use the clinical notes table in the MIMIC 2 dataset. We apply necessary transformation to the dataset to obtain relevant features such as medical notes, ICD-9 codes etc.

This modified dataset is handled in different ways for Bi-LSTM and BERT. For Bi-LSTM, the data has to be pre-processed to remove noise from data. It is then passed to the Bi-LSTM model for prediction. For BERT, no pre-processing is required as the BERT model in itself has a pre-processing model pretended to the prediction model. The results from the two models undergoes ensembling to obtain the final result. Figure 1 shows the schematic of our system.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116120

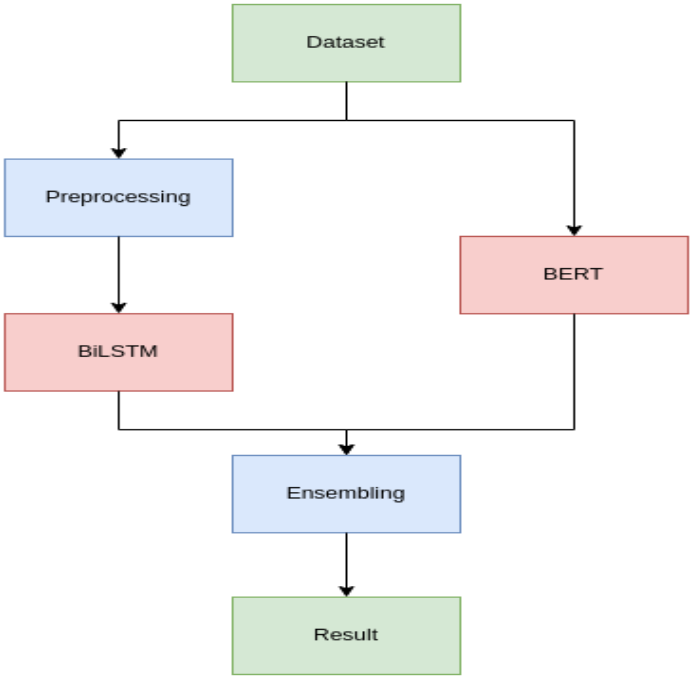

Fig. 1 System Architecture

a. Dataset

MIMIC-II (Multiparameter Intelligent Monitoring in Intensive Care II) is a database that has been created by collecting the data from the ICU patients. The MIMIC-II database consists of the following information: patient demographics, laboratory test results, vital sign recordings, fluid and medication records, charted parameters and free-text. As of the end of 2012, over 500 users have been approved for access to the MIMIC-II relational database, which reflects researchers' interest in the clinical data of MIMIC-II. Numerous innovative and significant studies on a broad range of topics are based on MIMIC-II and establish its importance. The software tools that make it feasible for a large worldwide community of investigators to draw on MIMIC-II are essential contributors to its value and utility for intensive care research.

The MIMIC-II relational database contains records from over 32,000 subjects, including over 7,000 neonatal patients. The raw data is stored in various base tables, generally organized by subject, hospital and ICU-stay IDs. Importantly, due to the sensitive nature of the data, all database records have been thoroughly and completely de-identified by the elimination of Protected Health Information (PHI) and altering the dates of the records.

Accessing and processing data from MIMIC-II is complex. It is highly recommended that studies based on the MIMIC-II database be conducted as collaborative efforts that include clinical, statistical, and relational database expertise.

b. Preprocessing

Data preprocessing is an important technique in data mining that helps transform raw data into an understandable format. Data preprocessing helps check the quality of data. The quality of data directly impacts the performance of machine learning models. Good data quality can lead to better prediction performance, thus forming a crucial step. Data preprocessing is achieved through various techniques. The data preprocessing techniques adopted in our methodology include:

- Data cleaning - Data cleaning is the process to remove incorrect data, incomplete data and inaccurate data from the datasets, and it also replaces the missing values. The major task was handling missing values. Since the occurrence of records with missing values was significantly low, the technique employed to handle missing values was dropping such records.
- Data integration - The note events table consisted of the clinical note texts, along with the diagnosis code while the icd9 table consisted of the mapping between diagnosis code and its description. Thus, an integration of the two tables was necessary, which was made possible by joining the tables.
- Textual preprocessing - Careful observation of clinical note texts revealed noise in the data. Natural Language Processing (NLP) techniques were employed in order to clean the texts.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116120

c. Bidirectional LSTM

Bidirectional Long Short Term Memory, abbreviated as BiLSTM, is a class of Recurrent Neural Network (RNN), primarily used in Natural Language Processing (NLP) tasks. It's a powerful tool that identifies dependencies between words and phrases in both directions of the sequence. Tensorflow enables us to define a custom bidirectional LSTM model. We begin by defining an Embedding layer, that obtains vector representation for words. This feeds into the Bidirectional LSTM layer, with 128 units. A dropout layer is added next to prevent overfitting. Finally, we have an output layer that outputs the probability estimates for the different classes.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, None, 200)         11245200

 bidirectional (Bidirectiona  (None, 256)              336896
 l)

 dropout (Dropout)           (None, 256)               0

 dense (Dense)               (None, 5)                 1285

=================================================================
Total params: 11,583,381
Trainable params: 11,583,381
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2 BiLSTM Model Architecture

d. BERT

BERT stands for Bidirectional Encoder Representations from Transformers, a state of the art open-source model by Google. Transformers consist of an encoder that reads the input text data and decoder that predicts the word for the task. Superiority of the BERT over other models comes from its training phase of the model; it uses two training strategies.
Bert has multiple variants like small, base, large and even a few expert variants for specific domains. However due to limited computing resources, considering the tradeoff between training time and accuracy, we decided to use 'small_bert/bert_en_uncased_L-2_H-128_A-2' and trained the model for 15 epochs.
L = Number of Transformers (2)
H = Embedding Length (128)
A = Number of Attention Heads per Transformer Layer (2)

```
Model: "model"
__________________________________________________________________________________________________
 Layer (type)                Output Shape              Param #     Connected to
==================================================================================================
 text (InputLayer)           [(None,)]                 0           []

 preprocessing (KerasLayer)  {'input_mask': (None,     0           ['text[0][0]']
                             128),
                              'input_type_ids':
                             (None, 128),
                              'input_word_ids':
                             (None, 128)}

 BERT_encoder (KerasLayer)   {'encoder_outputs':       4385921     ['preprocessing[0][0]',
                              [(None, 128, 128),                    'preprocessing[0][1]',
                              (None, 128, 128)],                    'preprocessing[0][2]']
                              'default': (None,
                             128),
                              'pooled_output': (
                             None, 128),
                              'sequence_output':
                              (None, 128, 128)}

 dropout (Dropout)           (None, 128)               0           ['BERT_encoder[0][3]']

 classifier (Dense)          (None, 5)                 645         ['dropout[0][0]']

==================================================================================================
Total params: 4,386,566
Trainable params: 4,386,565
Non-trainable params: 1
__________________________________________________________________________________________________
```

Fig. 3 BERT Model Architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116120

e. Ensembling

Ensembling is a general approach in machine learning that combines the predictions from many models, with the aim of achieving better predictive performance. The fundamental reasoning behind such an approach is that a diverse group of models would capture different patterns within the data and combining them would strengthen generalization, whereas a single model would identify only a subset of these patterns and miss out on the others, thus diminishing its generalisability. There are numerous techniques to perform ensembling. A simple, yet powerful ensemble technique, called Weighted Averaging was employed.

III RESULTS

In this section, we analyze the performance of the different models on the task of diagnosis prediction.

a. BiLSTM Results

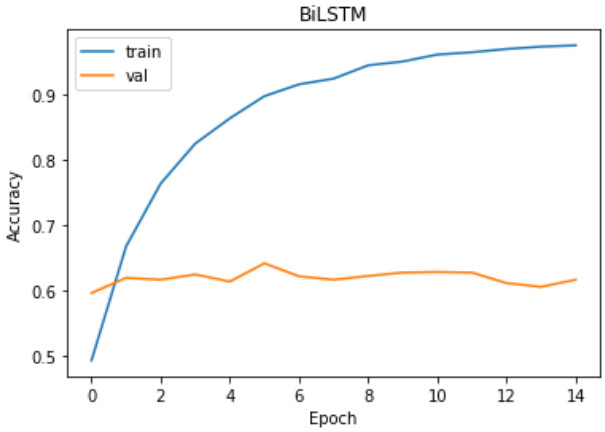

Fig. 4 Accuracy for BiLSTM

Figure 4 depicts a plot of accuracy of the Bidirectional LSTM model over epochs. We observe an accuracy between 60% to 63% for the model.

b. BERT Results

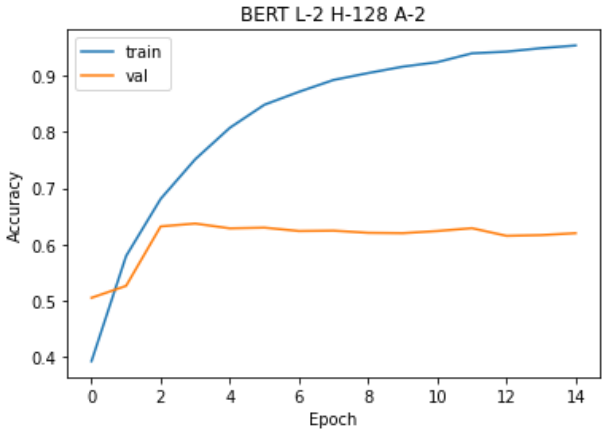

Fig. 5 Accuracy for BERT

Figure 5 depicts a plot of accuracy of the BERT model over epochs. We observe an accuracy around 63% for the model.

c. Ensembling

Mathematically, weighted averaging is given as,

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116120

$$P = \sum_{1}^{n} (w_i \times p_i)$$

To achieve the best possible results, weights of 20% and 80% were assigned to the Bidirectional LSTM and BERT models respectively.

IV. CONCLUSION

In summary, this study aims to obtain the effectiveness of the use of novel RNN architectures on EHR such as MIMIC 2. The novel RNN architectures in consideration are Bi-LSTM, BERT. The study uses Ensemble Learning to improve the accuracy of results. From the results, we observe that with these models one can achieve good results for classification and concept extraction (ex: Diagnosis) from clinical notes. The model uses NLP to extract information from text which is used to assign ICD-9 code to the text in consideration.

Our experiment will serve as the basis of future experimentations of RNN, Bi-LSTM, BERT and advanced ensembling techniques on medical notes from MIMIC-2 and MIMIC-3 dataset and for diagnosis prediction systems.

Automatic Diagnosis Prediction systems can save a lot of critical time in ICU and trauma situations thereby saving lives. This area of study can be useful in the near future. Hence, further research can be made to improve the accuracy and sophistication of the models.

REFERENCES

[1]. Lee J, Scott DJ, Villarroel M, Clifford GD, Saeed M, Mark RG. Open-access MIMIC-II database for intensive care research. Annu Int Conf IEEE Eng Med Biol Soc. 2011;2011:8315-8318. doi:10.1109/IEMBS.2011.6092050

[2]. Reza Sadeghi, Tanvi Banerjee, William Romine, Early hospital mortality prediction using vital signals, Smart Health, Volumes 9–10, 2018, Pages 265-274, ISSN 2352-6483, https://doi.org/10.1016/j.smhl.2018.07.001.S.

[3]. Awad A, Bader-El-Den M, McNicholas J, Briggs J. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. Int J Med Inform. 2017 Dec;108:185-195. doi: 10.1016/j.ijmedinf.2017.10.002. Epub 2017 Oct 5. PMID: 29132626.

[4]. Nuthakki, Siddhartha, Sunil Neela, Judy Wawira Gichoya and Saptarshi Purkayastha. "Natural language processing of MIMIC-III clinical notes for identifying diagnosis and procedures with neural networks." ArXiv abs/1912.12397 (2019): n. Pag.

[5]. Mascio, Aurelie, Zeljko Kraljevic, Daniel M Bean, Richard J. B. Dobson, Robert J Stewart, Rebecca Bendayan and Angus Roberts. "Comparative Analysis of Text Classification Approaches in Electronic Health Records." BIONLP (2020).

[6]. Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

[7]. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM networks. In Proceedings of the IEEE International Joint Conference on Neural Networks(IJCNN), Montreal, QC, Canada, 31 July–4 August 2005; Volume 4, pp. 2047–2052.

[8]. Devlin, Jacob & Chang, Ming-Wei & Lee, Kenton & Toutanova, Kristina. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

[9]. F. Huang, G. Xie and R. Xiao, "Research on Ensemble Learning," 2009 International Conference on Artificial Intelligence and Computational Intelligence, 2009, pp. 249-252, doi: 10.1109/AICI.2009.235.