International Journal of Electrical and Computer Engineering (IJECE)
Vol. 11, No. 2, April 2021, pp. 1689~1696
ISSN: 2088-8708, DOI: 10.11591/ijece.v11i2.pp1689-1696

Machine learning model for clinical named entity recognition

Ravikumar J.[1], Ramakanth Kumar P.[2]
[1]Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India
[2]Department of Computer Science and Engineering, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India

Article Info

Article history:

Received Feb 22, 2020
Revised Aug 1, 2020
Accepted Sep 25, 2020

Keywords:

Clinical NER
Machine learning
Named entity recognition
Natural language processing

ABSTRACT

To extract important concepts (named entities) from clinical notes, most widely used NLP task is named entity recognition (NER). It is found from the literature that several researchers have extensively used machine learning models for clinical NER. The most fundamental tasks among the medical data mining tasks are medical named entity recognition and normalization. Medical named entity recognition is different from general NER in various ways. Huge number of alternate spellings and synonyms create explosion of word vocabulary sizes. This reduces the medicine dictionary efficiency. Entities often consist of long sequences of tokens, making harder to detect boundaries exactly. The notes written by clinicians written notes are less structured and are in minimal grammatical form with cryptic short hand. Because of this, it poses challenges in named entity recognition. Generally, NER systems are either rule based or pattern based. The rules and patterns are not generalizable because of the diverse writing style of clinicians. The systems that use machine learning based approach to resolve these issues focus on choosing effective features for classifier building. In this work, machine learning based approach has been used to extract the clinical data in a required manner.

Corresponding Author:

Ravikumar J.
Department of Computer Science and Engineering
Dr. Ambedkar Institute of Technology
Bengaluru, India
Email: ravij041@dr-ait.org

1. INTRODUCTION

The patient's data ranging from diagnoses, treatments, problems, medications to imaging and clinical notes like discharge summaries are available in electronic health records (EHR). For quality, billing and outcome structured data are important. On the other hand, narrative text is more engaging, more expressive and captures patient's data more accurately. Clinical notes also contain data indicating the level of concern and uncertainty to others who are reviewing the note. Hence, in order to obtain clear perspective on the condition of the patient, an analysis of narrative text needs to be done. But, the manual analysis of huge number of narrative text is time consuming and prone to errors.

To resolve this issue, machine learning based systems can be used. It can be observed from the literature that various machine learners have been used. support vector machines (SVMs) [1] and hidden markov model (HMM) [2] are examples of such learners. To understand the natural language [3], natural language processing that focuses on development of models is being used. The framework of NLP includes modules for syntactic processing like tokenization, parts of speech tagging and sentence detection. Modules for named entity recognition tagging, extraction of relation and concept identification are included in the NLP systems. An NLP system that has semantic processing models for extraction of pre-defined information is information extraction system. In the medical field, researchers are using NLP systems for identification of biomedical concepts and clinical syndromes from radiology reports [4] and discharge summaries [5].

Clinical researchers and other medical operations make use of important information extracted by analysis of clinical notes in detailed manner. These clinical notes provide rich and detailed medical information. In the present work, we have built a machine learning model for extraction of medical NERs namely disease, test and treatment. An analysis has been done from the text of doctor's notes and records generated during interaction with patient.

2. RELATED WORK

Decision tree based NER model was built by Sekine *et al.* [6] that used features such as part-of-speech tags extracted by morphological analyzer, specialized dictionary and character based information. This was developed for Japanese. Bikel *et al.* [7] used hidden markov model (HMM) for identification of named entity. Features like bi-gram and orthographic features like word case, word shape etc. were used. In his Ph.D thesis, Borthwick [8] used maximum entropy (MaxEnt) algorithm. McCallum *et al.* [9] extracted NER using algorithm based on conditional random fields. A semi Markov conditional random field algorithm was proposed by Sarawagi *et al.* [10] for extraction of named entity. The researches extended the semi Markov model with use of dictionary and notion of similarity function. An overall survey of NER research was provided by Naidu and Sekine [11].

Luu [12] proposed a framework that is based on different text mining and machine learning algorithms for addressing the challenges of clinical named entity recognition. The framework proposed has multiple levels and builds complex NER tasks. Different data sets-the CLEF 2016 challenge and BIONLP/NLPBPA 2004 were used for evaluation of the proposed method and the results validated the framework.

Mao *et al*. [13] opine that important clinical information related to diagnosis is available in Electronic medical record. By data mining of electronic medical record, recognition of medical named entity is done. In this research work, authors have taken ophthalmic electronic medical record as research object. In the beginning, under the guidance of specialist, training corpus is annotated. Later, trained HMM model is used in test set for recognition of entity. Finally, experiment is conducted for making comparison between the proposed algorithm and the algorithm based on word segmentation model. The results of the experimentation indicate that the algorithm achieves good results in the named entity recognition of electronic medical record. Li *et al*. [14] proposed a deep neural model BiLSTM-Att-CRF that is a combination of bidirectional long-short time memory network and attention mechanism. This improved the performance of NER in Chinese electronic medical records (EMRs). The proposed model achieved better results than other widely used models.

Qiu *et al*. [15] write that the goal of the clinical named entity recognition (CNER) is identification and classification of clinical terms like symptoms, exams, treatments, diseases. This is a crucial and fundamental task for clinical and translation research. In recent years, deep learning models have been successful in CNER tasks. These models depend on recurrent neural networks which maintain a vector of hidden activations that propagate through time. This causes too much time for model training. In the present work, the researchers have proposed a residual dilated convolutional neural network with conditional random field (RD-CNN-CRF) to solve it. In this method, dictionary features and Chinese characters are projected first into dense vector representations. Later, they are fed into the residual dilated convolutional neural network to capture contextual features.

Li *et al*. [16] proposed a model combining language model conditional random field algorithm (CRF) and bi-directional long short-term memory networks (BiLSTM) to realize automatic recognition and entity extraction in unstructured medical texts. The researchers crawled 804 specifications of drug for asthma treatment from the Internet. Later quantization is done for the normalized field of drug specification word by a vector as the input to the neural network. Experimentation indicated that recall, system accuracy and F1 value are improved by 5.2%, 6.18% and 4.87% compared to traditional machine learning model. The proposed model can be applied to extract named entity information from drug specification.

Summarising the concepts, the electronic medical record is a description of patients physical condition [17]. Named entity recognition is the method used for clinical data extraction. The NER was a combination of dictionary and rules [18]. In clinical decision, NLP has become recent trend [19]. Researchers have evaluated various machine learning algorithms with various features [20]. UMLS, Ctakes and Medline were introduced as characteristics and using semi-Markov model, an accuracy of 85.23% was achieved [21]. Wang *et al*. [22] constructed tagged symptom corpus including 11,613 chief complaints. Wang *et al*. [23] completed manual annotation for 12 data of liver cancer in 115 medical records. Yan *et al*. [24] put forward a united model of word segmentation and named entity recognition based on dual decomposition. Jianbo, *et al*., [25] selected 800 medical records and established named entity tagged corpus among which word segmentation and part-of-speech tagging utilize tools developed by Stanford University.

3. THE PROPOSED MODEL

The proposed model classifies clinical data and provides the data to concerned expert using machine learning framework and NLP technique. In the manual system, physicians and nurses have to go through the medical data and directs this data to concerned experts. It is time consuming, expensive and challenging task. The records of the patients include medical history, family history etc. The significant difference between classification of medical records and general text classification is word distribution. The proposed model uses machine learning framework for recognizing and extraction of concepts from clinical data. The framework includes an approach known as bidirectional long short tem memory-conditional random field (LSTM-CRF) initialized with general-purpose, off-the-shelf word embeddings. Figure 1 depicts the data flow used in the proposed model.

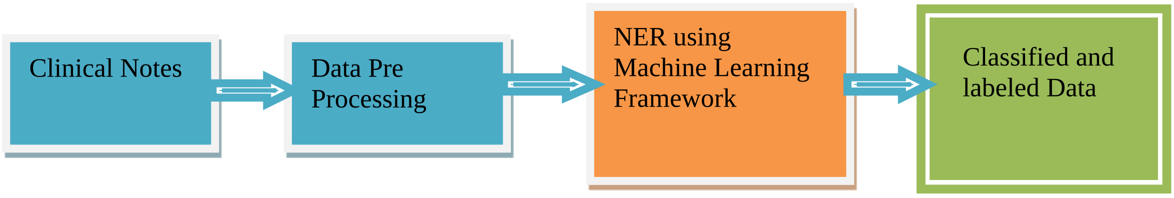


Figure 1. Machine learning framework for clinical NER

The input is $i = (i1, i2, i3, \dots . im)$ which indicate the words in a sentence
The output is $0 = \{o1, o2, o3, \dots om)$ which indicate named entity tags
Conditional probability is $P\ (o1, o2, o3, \dots om \mid i1, i2, i3, \dots . im\ )$
This can be done by defining feature map;

$$\Phi(\mathrm{i1}, \dots, \mathrm{im}, \mathrm{o1}, \dots, \mathrm{om}) \in \mathrm{Rd} \quad (1)$$

This is a mapping of entire input sequence paired with an entire state sequence to some dimensional feature vector. The probability as a log-linear model with the parameter vector has been modeled as

$$\omega \varepsilon Rd \quad (2)$$

$$\mathrm{P}(\mathrm{o}|\mathrm{i}; \mathrm{w}) = \frac{\exp\ (\omega . \Phi(\mathrm{i,i})}{\sum \mathrm{of} \exp\ (\omega . \Phi(\mathrm{i,of}))} \quad (3)$$

where o ranges over all possible output sequences. The expression $w.^{\Phi}(i, i) = score\ crf\ (i, o)$ indicates a scoring how well the state sequence fits the given input sequence. Hence score can be defined as,

$$score\ lstm - crf\ (\ i, o) = \sum_{j=0}^{n} W\ oj - 1, oj\ . LSTM(\ i)j + b\ oj - 1, o\ j \quad (4)$$

where $Oj - 1$,
oi are weight vector
b is the bias corresponding to the transition from $oi - 1$ to oj1espectively.

The algorithm used for the overall process is given in Figure 2. Medical records that consist of test conducted, patient’s health status, response to the treatments and diseases are given as input. In the next stage, concepts like medical tests, diagnosis and treatments mentioned in the clinical records are classified into categories. Later, the records are divided into training data and testing data. 70% of data is used as training data and it is fed to the model. Testing data (30% of data) that consists of patient’s information are fed to the model. Once the model is tuned for accuracy, the model will be ready to receive the real data. Then, the real data which is actually clinical records are fed to the pre developed model. The output includes list of words that indicate test conducted, problem diagnosed or treatment given. From the list of diseases and test conducted, the specializations are classified and displayed. The benefit of this is that the experts in specific area need not read all clinical record, they can directly read summary which saves lot of time. Using LSTM method which is based on machine learning, extraction of diagnosis and test names is extracted. NLP has been used for this. The screenshot is shown in Figure 3.

Step 1	**Start** Input: Medical records consisting of tests conducted, patient’s health status, diseases and response to the treatments.
Step 2	**Classification Model development** Concepts like medical tests, diagnosis and treatments mentioned in the clinical records are classified into categories.
Step 3	**Model building using training data** The records are divided into training data and testing data. 70% of data is used as training data and it is fed to the model.
Step 4	**Testing the model accuracy** Testing data (30% of data) that consists of patient’s information are fed to the model.
Step 5	**Input Medical records** The real data (clinical records) are fed to the pre developed model.
Step 6	**Obtain output** The output includes list of words that indicate test conducted, problem diagnosed or treatment given.
Step 7	**Classify** From the list of diseases and test conducted, the specializations are classified and displayed.
Step 7	**End**

Figure 2. Algorithm for classification

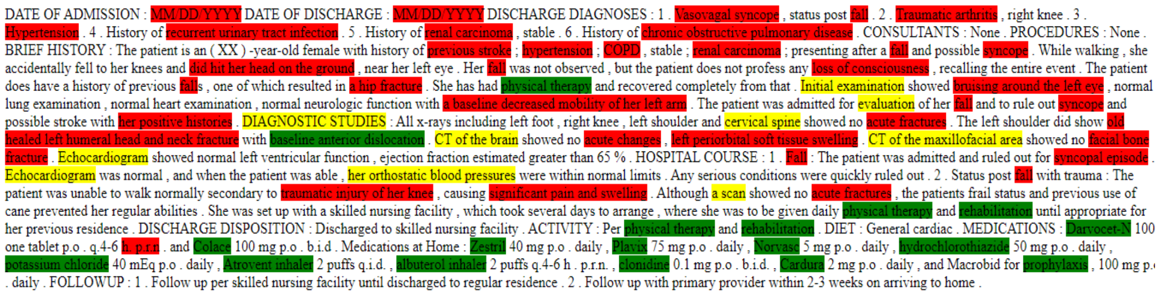
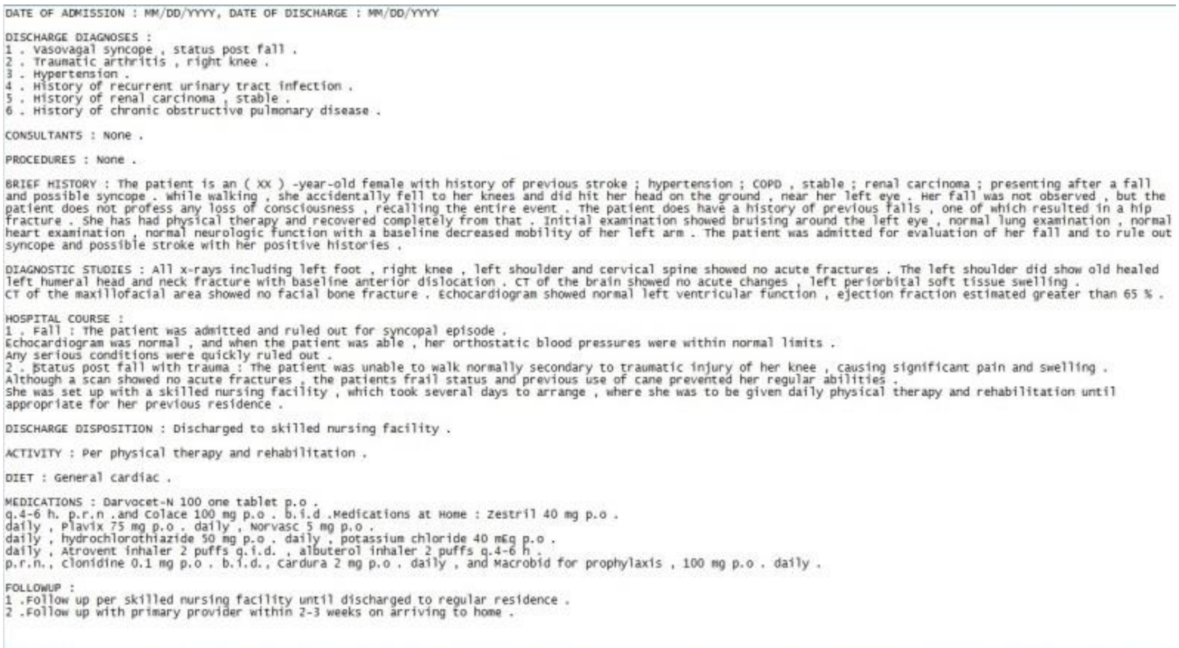

DATE OF ADMISSION : MM/DD/YYYY DATE OF DISCHARGE : MM/DD/YYYY DISCHARGE DIAGNOSES : 1 . Vasovagal syncope , status post fall . 2 . Traumatic arthritis , right knee . 3 . Hypertension . 4 . History of recurrent urinary tract infection . 5 . History of renal carcinoma , stable . 6 . History of chronic obstructive pulmonary disease . CONSULTANTS : None . PROCEDURES : None . BRIEF HISTORY : The patient is an (XX) -year-old female with history of previous stroke ; hypertension ; COPD , stable ; renal carcinoma ; presenting after a fall and possible syncope . While walking , she accidentally fell to her knees and did hit her head on the ground , near her left eye . Her fall was not observed , but the patient does not profess any loss of consciousness , recalling the entire event . The patient does have a history of previous falls , one of which resulted in a hip fracture . She has had physical therapy and recovered completely from that . Initial examination showed bruising around the left eye , normal lung examination , normal heart examination , normal neurologic function with a baseline decreased mobility of her left arm . The patient was admitted for evaluation of her fall and to rule out syncope and possible stroke with her positive histories . DIAGNOSTIC STUDIES : All x-rays including left foot , right knee , left shoulder and cervical spine showed no acute fractures . The left shoulder did show old healed left humeral head and neck fracture with baseline anterior dislocation . CT of the brain showed no acute changes , left periorbital soft tissue swelling . CT of the maxillofacial area showed no facial bone fracture . Echocardiogram showed normal left ventricular function , ejection fraction estimated greater than 65 % . HOSPITAL COURSE : 1 . Fall : The patient was admitted and ruled out for syncopal episode . Echocardiogram was normal , and when the patient was able , her orthostatic blood pressures were within normal limits . Any serious conditions were quickly ruled out . 2 . Status post fall with trauma : The patient was unable to walk normally secondary to traumatic injury of her knee , causing significant pain and swelling . Although a scan showed no acute fractures , the patients frail status and previous use of cane prevented her regular abilities . She was set up with a skilled nursing facility , which took several days to arrange , where she was to be given daily physical therapy and rehabilitation until appropriate for her previous residence . DISCHARGE DISPOSITION : Discharged to skilled nursing facility . ACTIVITY : Per physical therapy and rehabilitation . DIET : General cardiac . MEDICATIONS : Darvocet-N 100 one tablet p.o . q.4-6 h. p.r.n . and Colace 100 mg p.o . b.i.d . Medications at Home : Zestril 40 mg p.o . daily , Plavix 75 mg p.o . daily , Norvasc 5 mg p.o . daily , hydrochlorothiazide 50 mg p.o . daily , potassium chloride 40 mEq p.o . daily , Atrovent inhaler 2 puffs q.i.d. , albuterol inhaler 2 puffs q.4-6 h . p.r.n. , clonidine 0.1 mg p.o . b.i.d. , Cardura 2 mg p.o . daily , and Macrobid for prophylaxis , 100 mg p.o . daily . FOLLOWUP : 1 . Follow up per skilled nursing facility until discharged to regular residence . 2 . Follow up with primary provider within 2-3 weeks on arriving to home .

Figure 3. Results of NER extraction, disease names (red), diagnosis (green) and tests (yellow)

Once NER with NLP is applied for extraction of entities and their relationships, further processing is done. The disease names, test, diagnosis test are fed as input to machine learning framework. An output of the model will be classified data labeled with specialization as shown in Figure 4. Figure 5 and Figure 6 shows the execution screenshot during classification.

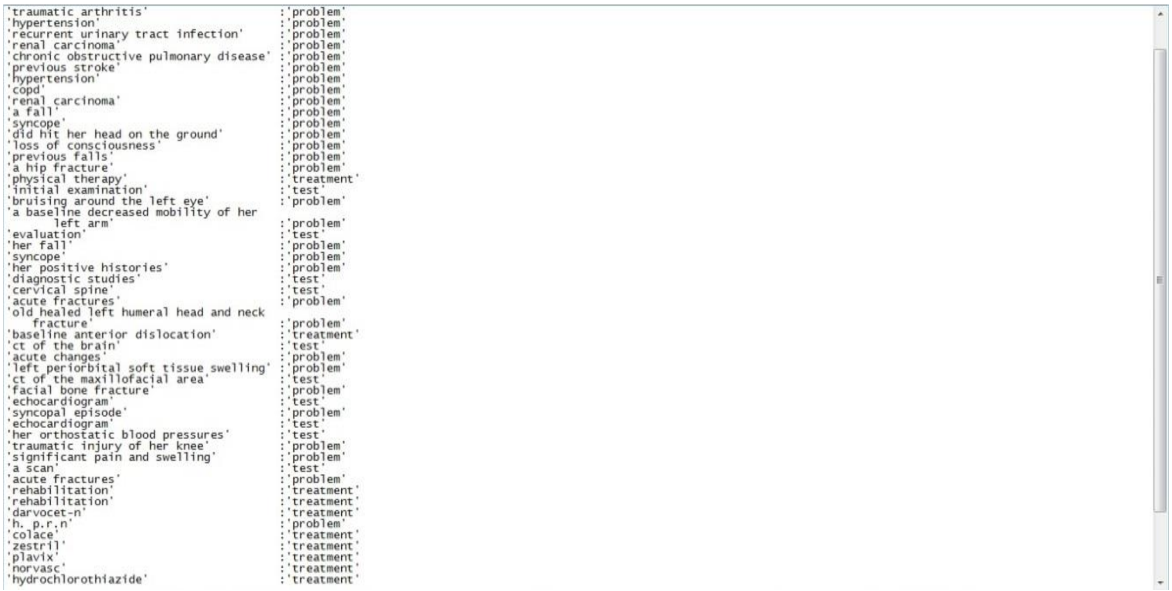
'vasovagal syncope'	Problem	Specialization 1
'fall'	Problem	Specialization 2
'traumatic arthritis'	Problem	Specialization 3
'hypertension'	Problem	Specialization 4
'physical therapy'	Treatment	
'evaluation'	Test	
'cervical spine'	Test	
'pain'	Problem	Specialization 5
‘traumatic injury of her knee’	Problem	Specialization 6
'hypertension'	Problem	Specialization 3
‘atrovent inhaler’	Treatment	
' a scan '	Test	

Figure 4. Classification as per specialization



```
DATE OF ADMISSION : MM/DD/YYYY, DATE OF DISCHARGE : MM/DD/YYYY

DISCHARGE DIAGNOSES :
1 . Vasovagal syncope , status post fall .
2 . Traumatic arthritis , right knee .
3 . Hypertension .
4 . History of recurrent urinary tract infection .
5 . History of renal carcinoma , stable .
6 . History of chronic obstructive pulmonary disease .

CONSULTANTS : None .

PROCEDURES : None .

BRIEF HISTORY : The patient is an ( XX ) -year-old female with history of previous stroke ; hypertension ; COPD , stable ; renal carcinoma ; presenting after a fall
and possible syncope . While walking , she accidentally fell to her knees and did hit her head on the ground , near her left eye . Her fall was not observed , but the
patient does not profess any loss of consciousness , recalling the entire event . The patient does have a history of previous falls , one of which resulted in a hip
fracture . She has had physical therapy and recovered completely from that . Initial examination showed bruising around the left eye , normal lung examination , normal
heart examination , normal neurologic function with a baseline decreased mobility of her left arm . The patient was admitted for evaluation of her fall and to rule out
syncope and possible stroke with her positive histories .

DIAGNOSTIC STUDIES : All x-rays including left foot , right knee , left shoulder and cervical spine showed no acute fractures . The left shoulder did show old healed
left humeral head and neck fracture with baseline anterior dislocation . CT of the brain showed no acute changes , left periorbital soft tissue swelling .
CT of the maxillofacial area showed no facial bone fracture . Echocardiogram showed normal left ventricular function , ejection fraction estimated greater than 65 % .

HOSPITAL COURSE :
1 . Fall : The patient was admitted and ruled out for syncopal episode .
Echocardiogram was normal , and when the patient was able , her orthostatic blood pressures were within normal limits .
Any serious conditions were quickly ruled out .
2 . Status post fall with trauma : The patient was unable to walk normally secondary to traumatic injury of her knee , causing significant pain and swelling .
Although a scan showed no acute fractures , the patients frail status and previous use of cane prevented her regular abilities .
She was set up with a skilled nursing facility , which took several days to arrange , where she was to be given daily physical therapy and rehabilitation until
appropriate for her previous residence .

DISCHARGE DISPOSITION : Discharged to skilled nursing facility .

ACTIVITY : Per physical therapy and rehabilitation .

DIET : General cardiac .

MEDICATIONS : Darvocet-N 100 one tablet p.o .
q.4-6 h. p.r.n .and Colace 100 mg p.o . b.i.d .Medications at Home : Zestril 40 mg p.o .
daily , Plavix 75 mg p.o . daily , Norvasc 5 mg p.o .
daily , hydrochlorothiazide 50 mg p.o . daily , potassium chloride 40 mEq p.o .
daily , Atrovent inhaler 2 puffs q.i.d. , albuterol inhaler 2 puffs q.4-6 h .
p.r.n., clonidine 0.1 mg p.o . b.i.d., Cardura 2 mg p.o . daily , and Macrobid for prophylaxis , 100 mg p.o . daily .

FOLLOWUP :
1 .Follow up per skilled nursing facility until discharged to regular residence .
2 .Follow up with primary provider within 2-3 weeks on arriving to home .
```

Figure 5. Clinical record



```
'traumatic arthritis'                      :'problem'
'hypertension'                             :'problem'
'recurrent urinary tract infection'        :'problem'
'renal carcinoma'                          :'problem'
'chronic obstructive pulmonary disease'    :'problem'
'previous stroke'                          :'problem'
'hypertension'                             :'problem'
'copd'                                     :'problem'
'renal carcinoma'                          :'problem'
'a fall'                                   :'problem'
'syncope'                                  :'problem'
'did hit her head on the ground'           :'problem'
'loss of consciousness'                    :'problem'
'previous falls'                           :'problem'
'a hip fracture'                           :'problem'
'physical therapy'                         :'treatment'
'initial examination'                      :'test'
'bruising around the left eye'             :'problem'
'a baseline decreased mobility of her
      left arm'                            :'problem'
'evaluation'                               :'test'
'her fall'                                 :'problem'
'syncope'                                  :'problem'
'her positive histories'                   :'problem'
'diagnostic studies'                       :'test'
'cervical spine'                           :'test'
'acute fractures'                          :'problem'
'old healed left humeral head and neck
   fracture'                               :'problem'
'baseline anterior dislocation'            :'treatment'
'ct of the brain'                          :'test'
'acute changes'                            :'problem'
'left periorbital soft tissue swelling'    :'problem'
'ct of the maxillofacial area'             :'test'
'facial bone fracture'                     :'problem'
'echocardiogram'                           :'test'
'syncopal episode'                         :'problem'
'echocardiogram'                           :'test'
'her orthostatic blood pressures'          :'test'
'traumatic injury of her knee'             :'problem'
'significant pain and swelling'            :'problem'
'a scan'                                   :'test'
'acute fractures'                          :'problem'
'rehabilitation'                           :'treatment'
'rehabilitation'                           :'treatment'
'darvocet-n'                               :'treatment'
'h. p.r.n'                                 :'problem'
'colace'                                   :'treatment'
'zestril'                                  :'treatment'
'plavix'                                   :'treatment'
'norvasc'                                  :'treatment'
'hydrochlorothiazide'                      :'treatment'
```

Figure 6. Classified data before labeling specialization

4. RESULTS AND DISCUSSIONS

In the proposed model machine learning algorithms used are support vector machine (SVM), naïve bayes, logistic regression, decision tree, random forest and light GBM. The screenshot related to accuracy of these algorithms is shown in Figure 7. The accuracy of the algorithms used is presented in graphical form in Figure 8. The model proposed can be used for extraction of medical data using NER and NLP technique. The machine learning model built into medical automation systems can be a good resource for medical experts as it saves lot of time spent for referring clinical records in detail. Also, administrative tasks can be easier as the model separates the diseases and treatment in to specializations.

The existing NLP systems for NER using clinical data consist of syntactic processing modules like sentence detection, tokenization, part-of-speech tagging etc. The semantic modules include concept identification, entity recognition, relation extraction and anaphoric resolution etc. So far, in the literature, it is observed that systems exists for extraction of named entities like disease, treatment etc which was useful for doctors to read summary information without reading complete clinical records. But, the proposed model goes one step further by classifying the named entities as per specialization. This can be embedded in health automation system for efficient delivery of services saving lot of time. Hence the proposed system can be a good candidate for the research in the area of NER in medical field.

```
Accuracy for training set for svm = 0.9256198347107438
Accuracy for test set for svm = 0.8032786885245902

Accuracy for training set for Naive Bayes = 0.8677685950413223
Accuracy for test set for Naive Bayes = 0.7868852459016393

Accuracy for training set for Logistic Regression = 0.8636363636363636
Accuracy for test set for Logistic Regression = 0.8032786885245902

Accuracy for training set for Decision Tree = 1.0
Accuracy for test set for Decision Tree = 0.7704918032786885

Accuracy for training set for Random Forest = 0.987603305785124
Accuracy for test set for Random Forest = 0.7540983606557377

Accuracy for training set for LightGBM = 0.9958677685950413
Accuracy for test set for LightGBM = 0.7704918032786885
```

Figure 7. Accuracy of algorithms

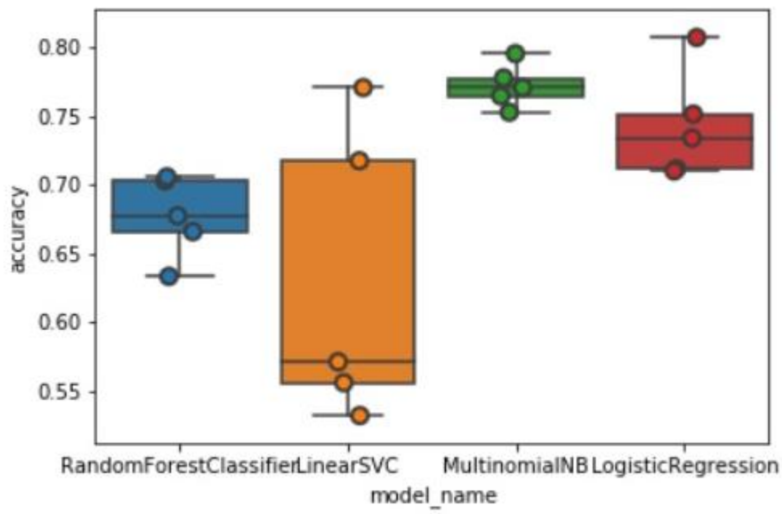


Figure 8. Accuracy comparison of algorithms

5. CONCLUSION

Because of diverse writing style of clinicians, the rules and patterns are not generalizable. These issues can be addressed by making use of technologies like machine learning. Named entity recognition is grouped into three approaches. Machine learning based approaches, rule-based approaches and dictionary based approaches. The systems that use machine learning based approach focus on choosing effective features for classifier building. Several researchers have extensively used machine learning models for clinical NER. Databases such as PubMed which include medical publications have generated lot of interest among researchers for applying information extraction techniques to medical literature. In an attempt to contribute to the research in this area, this work proposed a machine learning model for clinical NER. The model proposed perfomed better compared to some of the existing methods.

REFERENCES

[1] T. Joachims, C. Nedellec, and C. Rouveirol. "Text categorization with support vector machines: learning with many relevant. In Machine Learning," *ECML-European Conference on Machine Learning*, 1998, pp. 137-142.
[2] L. Rabiner et al., "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, 1989, pp. 257-286.
[3] S. M. Meystre, G. K. Savova, K. C. Kipper-Schuler, J. F. Hurdle, "Extracting informa-tion from textual documents in the electronic health record," *Yearbook of Medical Informatics*, vol. 35, pp. 128-144, 2008.
[4] R. W. V. Flynn, T. M. Macdonald, N. Schembri, G. D. Murray, A. S. F. Doney, "Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes," *Pharmacoepidemiology and Drug Safety*, vol. 19, no. 2010, pp. 843-847, 2010.
[5] H. Yang, I. Spasic, J. A. Keane, G. Nenadic, "A text mining approach to the pre-diction of disease status from clinical discharge summaries," *Journal of the American Medical Informatics Association (JAMIA)*, vol. 16, no. 4, pp. 596-600, 2009.
[6] Sekine, S., "Nyu: Description of the Japanese NE System Used For Met-2," *Proc. of the Seventh Message Understanding Conference (MUC-7)*, 1998.
[7] Bikel, D. M., Schwartz, R., and Weischedel, R. M., "An algorithm that learns what's in a name," *Machine learning*, vol. 34, no. 1-3, pp. 211-231, 1999.
[8] Borthwick, A., "A maximum entropy approach to named entity recognition," PhD diss., New York University, 1999.
[9] McCallum, A and Wei L., "Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons," *In Proceedings of the seventh conference on Natural language learning at HLT-NAAC*L, vol. 4, 2003, pp. 188-191.
[10] Sarawagi, S. and Cohen, W. W., "Semi-markov conditional random fields for information extraction," In *Advances in Neural Information Processing Systems*, 2004, pp. 1185-1192.
[11] Cohen, W. W., and Sarawagi, S., "Exploiting dictionaries in named entity extraction: combining semi-markov extraction processes and data integration methods," In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 89-98.
[12] T. M. Luu, R. Phan, R. Davey and G. Chetty, "A Multilevel NER Framework for Automatic Clinical Name Entity Recognition," *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, New Orleans, LA, 2017, pp. 1134-1143.
[13] X. Mao, F. Li, H. Wang and H. Wang, "Named Entity Recognition of Electronic Medical Record Based on Improved HMM Algorithm," *2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC)*, Dalian, China, 2017, pp. 435-438.
[14] L. Li and L. Hou, "Combined Attention Mechanism for Named Entity Recognition in Chinese Electronic Medical Records," *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, Xi'an, China, 2019, pp. 1-2.
[15] J. Qiu, Q. Wang, Y. Zhou, T. Ruan and J. Gao, "Fast and Accurate Recognition of Chinese Clinical Named Entities with Residual Dilated Convolutions," *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain, 2018, pp. 935-942.
[16] W. Li, et al., "Drug Specification Named Entity Recognition Base on BiLSTM-CRF Model," *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, Milwaukee, WI, USA, 2019, pp. 429-433.
[17] R. C. Wasserman, "Electronic medical records (EMRs) epidemiology and epistemology: reflections on EMRs and future pediatric clinical research," *Academic pediatrics*, vol. 11, no. 4, pp. 280-287, 2011.
[18] D. Demner-Fushman, W. W. Chapman and C. J. McDonald, "What can natural language processing do for clinical decision support?," *Journal of biomedical informatics*, vol. 42, no. 5, pp. 760-772, 2009.
[19] A. R. Aronson and F. M. Lang, "An overview of MetaMap: historical perspective and recent advances," *Journal of the American Medical Informatics Association*, vol. 17, no. 3, pp. 229-236, 2010.
[20] M. Jiang, Y. Chen, M. Liu, et al., "A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 601-606, 2011.
[21] B. De Bruijn, et al., "Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 557-562, 2011.
[22] Y. Wang, Z. Yu, L. Chen, et al., "Supervised methods for symptom name recognition in free-text clinical records of traditional Chinese medicine: An empirical study," *Journal of biomedical informatics*, vol. 47, pp. 91-104, 2014.
[23] H. Wang, W. Zhang, Q. Zeng, et al., "Extracting important information from Chinese Operation Notes with natural language processing methods," *Journal of biomedical informatics*, vol. 48, pp. 130-136, 2014.
[24] Y. Xu, Y. Wang, T. Liu, et al., "Joint segmentation and named entity recognition using dual decomposition in Chinese discharge summaries," *Journal of the American Medical Informatics Association*, vol. 21, no. e1, pp. e84-e92, 2014.
[25] J. Lei, B. Tang, X. Lu, et al., "A comprehensive study of named entity recognition in Chinese clinical text," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 808-814, 2014.

BIOGRAPHIES OF AUTHORS

Ravikumar J. is presently working as an Assistant professor at Dr. Ambedkar Institute of Technology, Bengaluru. He has 8 years of Teaching and 1 year of industry Experience. His research interests include Digital Image processing, computer networks and IOT.

Ramakanth Kumar P. is presently working as Professor & HoD of Department of Computer Science and Engineering, R V College of Engineering. He has 25 years of Teaching and 14 years of R&D Experience. His research interests include Digital Image Processing, Pattern Recognition and Natural Language processing.