Group12: Review on Biomedical/Clinical Text Processing

Saksham Gupta Saurab Jain Sathvik Bhagavan
Harshit Gupta Akshay Kumar Arya
170613, 170642, 170638, 170292, 17807074
{gupsak, jnsaurab, sathvikb, hargupta, akshayka}@iitk.ac.in
Indian Institute of Technology Kanpur (IIT Kanpur)

Abstract

In this report, we review recent advances in Biomedical/Clinical Text Processing. We look at formulations of different problems in this field and particularly study Named Entity Recognition and Clinical Coding in depth. We discuss various Deep Learning models and see how they extract contextual information to give state-of-the-art results. We also explore publicly available datasets in the biomedical domain, where data privacy is a massive bottleneck. Lastly, we provide different ideas to work on in the area of Clinical Coding.

1 Introduction

Biomedical/Clinical text processing (BioNLP) refers to the study of how natural language processing (NLP) methods may be applied to texts and literature of the biomedical and clinical domain. The exponential increase in the number of electronic health records (EHRs) has created a tremendous opportunity to derive previously unknown healthcare insights. EHRs contain a lot of information about patients, such as the patient disease history, treatment and results, interpretation of test images, etc. BioNLP has a wide range of real-life applications like gene-disease association, building EHR Question-Answer and cause of death classification, to name a few. In this paper, we first look at the different publicly available datasets for BioNLP. The privacy issues related to patients in EHRs make it challenging to access enough data to train the data-hungry deep learning models. Next, we look at some important problems in BionNLP and then discuss two tasks in detail: Named Entity Recognition and Clinical Coding. Lastly, we propose a model which may improve on the existing techniques of clinical coding in BioNLP.

2 Datasets used for BioNLP

There are many open datasets available which primarily consists of clinical reports of doctors. Also, many competitions have been organized in the past whose datasets are annotated and open to the public. Some of the datasets are:

- **MIMIC(Medical Information Mart for Intensive Care) III** [1]- This dataset was developed by Computational Physiology Lab in MIT. It includes discharge summaries, patient's progress notes written by a doctor. It contains data of around 60,000 patients in the intensive care unit. It is not annotated.
- **i2b2 datasets** [2] - These are the part of i2b2 challenges which began in 2006. These are publicly available and annotated, hence very beneficial. Also, different problem statements were released each year and hence different types of datasets are present. It also contains unannotated reports.
- **ShARe/CLEF eHealth** [3] - This was first developed in 2013 focused on clinical NLP and information retrieval tasks. It includes Radiology reports, discharge summaries, ECG reports etc. Again, it is annotated and hence beneficial to train and test supervised NLP algorithms.

- **Thyme corpus** [4] - This was developed as a project which aimed at extracting temporal relations which are present in the clinical reports. This corpus consists of radiology reports, related to brain and colon cancer. It is annotated for temporal relations between entities like events, time etc with other events etc. It also contains UMLS named entities.
- **UFAL Medical Corpus** [5] - This is a parallel corpus aimed at improving machine translation in the medical domain. We can use english sentences for downstream NLP tasks. They are extracted from various sources like EMEA (European Medical Agency, Medical Web Crawl etc)
- **MeDAL dataset** [6] - It stands for Medical Dataset for Abbreviation Disambiguation. Various abbreviations can have multiple meanings. Hence each sentence which has an abbreviations is expanded with the true choice for each sentence. This dataset can be used for Word Sense Disambiguation task.
- **PubMed PICO Element Detection Dataset** [7] - Each sentence is annotated with one of the following categories : Participant(P), Intervention(I), Comparison(C), Outcome(O).
- C**linical Abbreviation Sense Inventory** [8] - This dataset also includes acronym disambiguation and were manually annotated for 500 random instances in the corpus.
- **PubMedQA** [9] - This is a dataset which contains questions and their corresponding answers derived from PubMed abstracts. The answer to the questions are either yes, no or maybe. It contains around 1000 annotated by experts, 61200 unannotated and 211,300 artificially created questions.
- **BioScope Corpus** [10] - This is dataset annotated for negation detection. It contains clinical free texts, papers, etc. and its token level annotations.
- **CodiEsp 2020 corpus** [11] - This is a dataset which contains medical texts and its corresponding clinical codes. It is divided into two parts - one of them codes diseases and other codes procedures. This was released as a competition as a part of eHealth CLEF.
- **Inferring Which Medical Treatments Work from Reports of Clinical Trials** [12] - This is a dataset which comprises of reports of clinical trials (RCTs) and the task is to infer whether the treatment/drug/intervention actually works or not.

These were some of the examples of datasets, there are many more and are used for various tasks of NLP. There is one major issue in making any clinical data public - privacy. Before releasing any data, people have to make sure that patients are anonymized and their details are not made public. Also, many companies/hospitals do not want to release any data, hence the exchange of data is difficult and also in order to access any public data, you have to register on their website and give the purpose for using the data.

In perspective of India, there are no datasets maintained by anyone related to clinical text processing that we could find. We can understand the reason behind this as in India, still many hospitals do not use Electronic Health Records especially in rural areas which makes it difficult to make a proper dataset. One way to go about this is to create our own dataset by scraping through open medical journals and reports.

3 Problem Statements in BioNLP

There are many problems which can be tackled using BioNLP. Some of the important problems are:

- **Negation Detection** - This is defined as the problem of identifying whether any statement in the clinical narrative has a negation or not. This is an important problem as many clinical statements are written as the absence of certain disease. For example - a doctor can write "No inflammation detected in his eye was detected.", so in this statement, "inflammation" was not detected. This might seem trivial for human beings but it is important for other downstream NLP tasks like Information Retrieval etc. Formally, Negation detection is a classification task. There are many algorithms that have been studied, implemented and tested. Some of them are - NegEx which is an algorithm based on regular expressions [13], NegFinder which is a rule-based system [14], an algorithm based on Naive Bayes and Decision Trees [15] etc. Recently many deep learning based approaches have also been studied like [16] which also detected the scope of the negation, i.e. the tokens which are affected by the negation using CNNs, [17] used LSTMs, [18] used BERT and transfer learning.

- **Word Sense Disambiguation (WSD)** - This is the task to determine the sense of a particular word used in a context. It could be seen as a classification problem where choices are its multiple senses. For e.g.: "An accident happened near the bank". Here bank could sense "riverbank" or "Organization". The broadly classified approaches to WSD are knowledge based approaches [19], [20] which use the knowledge present in lexical resources, supervised ML algorithms (such as Naive Bayes [21], SVM [22], Neural networks(DNN, Bi-LSTM) [23])[24] which use a classifier for each token/word, trained on corpus having sense annotated instances and semi-supervised ML algorithms [25] Supervised algorithms outperform knowledge based algorithms but they need lots of annotated data [26]. The important application areas include "Disambiguation of Biomedical Abbreviations" [27], "Summarization of biomedical documents" [28] etc.
- **Information Retrieval (IE)** - This is the task of extracting and encoding information from clinical narratives, journals, EHRs, discharge summaries, or medical reports. The IR system extracts concepts, events, entities from free text and determines the relation between them [29]. The extracted and encoded information can be used for performing specific downstream tasks. IE contains sub-tasks which extract semantic information like NER, Negation signal scope detection WSD, Relation extraction, etc. Most of the earlier research used rule based algorithms for IE. Later, ML techniques such as Support Vector Machines (SVM) [30], Conditional random field (CRF) [31], Decision Trees (DT) [32] were used because of their efficiency and accuracy. Recently, Neural Networks based architectures (LSTM, GRU, Attention models) [33] are being used for state of the art performance. The performance of IE systems differs a lot as different types of information are extracted in different tasks. It has many important applications like disease study, drug related studies etc. The tools available for IE are MetaMap [34], cTAKES [35], MedTagger [36], etc.
- **Named Entity Recognition (NER)** - This is explained in detail in the next section.
- **Clinical Coding** - This is also explained in detail in the next to next section.

4 Named Entity Recognition in BioNLP

NER in BioNLP is the task of identifying entities like diseases, genes, chemicals, symptoms and drugs in the given text. NER is generally viewed as a sequence classification problem or a **word-level tagging**. The words in the input sequence are labelled using the **"BIO" format**: B (beginning), I (inside), O (outside). The task is a complex one as biological entities [37]:

1. continuously increase with new research
2. often have large numbers of synonyms
3. are often referred to using abbreviations
4. are often described using long phrases
5. are composed of punctuation and characters

Early entity recognition techniques were rule-based methods. They relied heavily on domain expertise and carefully hand-crafted features. Over the years, many statistical machine learning algorithms were used in biomedical NER. The popular ones include Hidden Markov Model (HMM) [38], SVM, CRF and Structured Support Vector Machine (SSVM). Among the above-mentioned algorithms, CRF is still used in many state-of-the-art systems. However, the performance of such supervised algorithms depends on the availability of a well-annotated training corpus and on the selection/creation of a relevant features. The paper [39] compared some of the statistical and semantic methods, commonly used for NER. It also proposed a combination of statistical and semantic methods which showed a significant improvement over either approach.

With the recent surge in the popularity of deep neural networks (DNN), these have also been increasingly applied to biomedical NER. DNNs have improved on the state-of-the-art of traditional machine learning methods. Several recent studies [37, 40, 41, 42, 43, 44], on biomedical NER using DNNs, have adopted a fairly common pipeline. The input word sequence is transformed into a matrix by concatenating the corresponding "word vectors" from the word embedding matrix. The paper [45] proves that **fine-tuning the embeddings** on small subset of training corpus improves accuracy. The

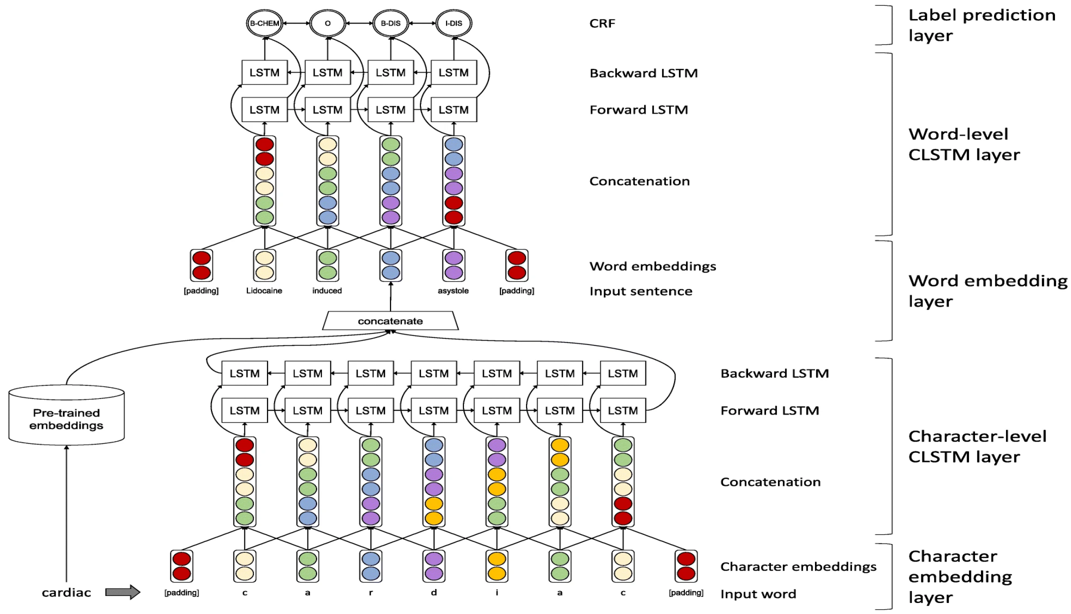


Figure 1: Model Architecture for NER using contextual information by [37]

most common DNN architecture used in biomedical NER is the Long Short Term Memory (LSTM). The final CRF layer makes tagging decisions according to the output of the LSTM layer.

The paper [37] incorporated a number of different techniques to acquire contextual information. The figure-1 gives a detailed graphical representation of the model architecture. The model used **character-level** and word level embeddings to feed into a BiLSTM layer. The accuracy of the model is superior to the BERT transformer on many datasets. Also, **few-shot learning** is being applied to counterbalance the lack of the datasets. Few shot methods generally involve using data similar to one given in the problem. On these lines, the paper [46] used a BiLSTM based model with fine-tuned and pre-trained word and character level embeddigs to improve on the state-of-the-art method.

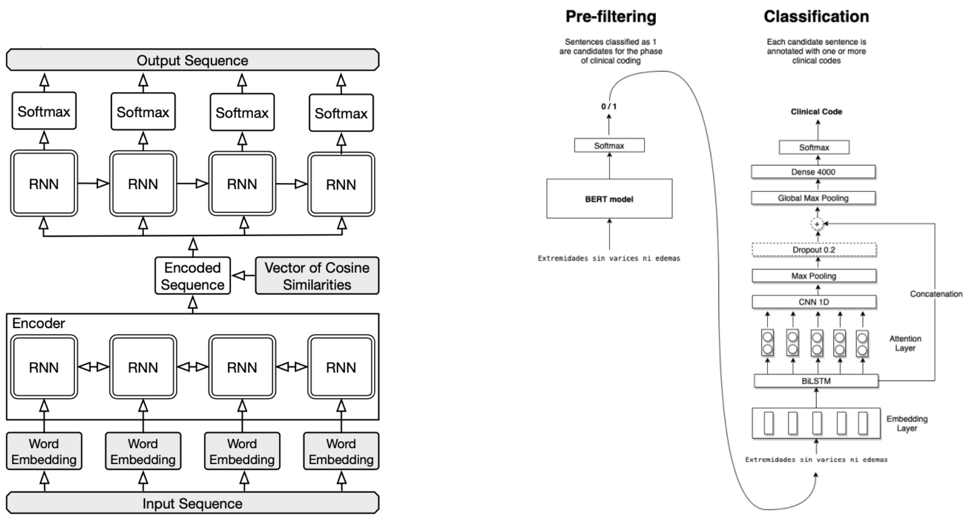
5 Clinical Coding

The task of clinical coding involves translating clinical statements into code as defined in International standards. This is helpful as clinical codes will be independent of any language which would facilitate doctors for reviewing any patient's medical history with ease and also will make the statements more concise and faster to interpret and process. Use of NLP in this domain can help in automatically extracting codes which will save a lot of time and if properly trained can outperform human coders as well. There are many types of clinical codes, important amongst them are **ICD10** (International Classification of Diseases) and **CPT** (Current Procedural Terminology).

Early Methods included K Nearest Neighbors, Bayesian Classifier etc [47]. Some people also proposed hybrid systems consisting of handcrafted rules and Decision Trees [48]. Recently, a number of deep learning based approaches have also been used to solve this. This problem can also be viewed as a machine translation problem, where text is translated from the language in which medical records were written, into a generalised clinical coding language. In the succeeding subsections, some of these deep learning techniques are presented, along with papers which utilized these techniques.

5.1 Using CNNs for ICD coding

CNNs along with attention mechanism[49] have been used in the past for predicting ICD codes. They utilised CNN layers to build document representations, and then frame the problem as a multi label



(a) Seq2seq model architecture used by [53]

(b) Fusion Model architecture used by [54]

Figure 2: Model architectures for clinical coding

classification problem. This paper[50] proposes a **multi filter residual CNN** to perform ICD coding. They built onto the previous state of the art[49]. The text fragments are first converted into a matrix using pre-trained word2vec embeddings. A convolutional layer is then used having multiple convolutional filters. This extra layer helps in overcoming the limitations of [49], by being able to handle text fragments of varying lengths. This is followed by a convolutional layer having multiple residual blocks. Following this, the attention layer and the output layer similar to [49] is used. The model was trained using the **MIMIC dataset**[51], and was compared to five state of the art models. It was concluded that the model obtained better performance than the baseline methods. The authors also highlighted that performance could have been improved further using recurrent transformers[52].

5.2 Sequence to sequence architecture

An encoder decoder model [55] is used for translating text into the clinical code. The input is a sequence of text in a given language which is required to be converted to the corresponding clinical code, the output. The encoder processes this input to generate a vector which ideally should capture the information present in this sequence. The decoder on the other hand uses this vector and tries to predict a sequence, which should be the clinical code. The encoders and decoders both are made up of RNNs. LSTMs are often used instead to capture the long term dependencies in the sequence.

The paper [53] uses a seq2seq architecture to perform ICD-10 coding of death certificates written in english language. For the encoder, bidirectional LSTMs were used to convert the input sequence into an encoded vector. The decoder was a left to right LSTM which predicted the output sequence from the encoded vector, as well as a cosine similarity vector. The cosine similarity vector was used to incorporate prior knowledge. The architecture is shown in figure 2(a). The dataset that was used was **the CLEF e-Health 2017 provided dataset**, which included a part of the **CepiDC Cause of Death Corpus** containing english text. The model performed better than the mean and median score of the other submitted models, even though no task specific feature engineering was done. This showed how deep learning models provide an edge over the traditional methods.

5.3 Transformers for clinical coding

Transformers changed the path of NLP when they were introduced in 2017. The applications of it are versatile and it can also be adapted towards solving the problem of clinical coding [51]. **BertXML**

[56] was introduced in 2020 which was trained on huge dataset, anonymous medical notes (around 7.5 million) of about 1 million patients. They trained BERT for clinical coding from scratch instead of using off-the-shelf models as the vocabulary, because they wanted one complete EHR to be taken in as input, which would not have fit in the 512 characters that BIO-BERT takes. So they trained their own BERT model with an input size of 1024 and also made a vocabulary from scratch. They used a multi-label output layer from **AttentionXML** [57]. After running various experiments on other models, they found out that BertXML outperformed BioBert and ClinicalBert in ICD code classification.

Some people also tried fusing Bert with LSTM and CNN. This was for a competition called as CodiEsp 2020 [54]. The text was split into many sentences. They were taken as input into the BERT model. The output was a probability that the sentence contained references to a clinical code. Those sentences which were classified as possible clinical code references were passed onto the next part of the model. The architecture is given in figure 2(b). They tried many models involving combinations of CNNs, LSTMs, attention mechanism instead of BERT for prefiltering. They also tried different word embeddings. In the end, they used Fasttext for Spanish Unannotated Corpora as their choice and BERT as their prefiltering choice.

6 Future Directions

- We plan to compare the impact and effect of using different word embeddings with the same model architecture in Clinical Coding. We will train the model using word2vec, BERT, BioBERT, and use a CNN architecture similar to [49], [50] for clinical coding.
- For future work, we plan to use character-level embeddings (proven to work in NER [45]) along with word embeddings in clinical coding tasks. We will train character-level embeddings and pass them into LSTM model before concatenating them with word embeddings.
- As annotated Indian datasets are very few and not publically available, we could use unsupervised learning algorithms (for e.g. [58]) for clinical coding on publically available Indian clinical narratives.

7 Conclusion

In this paper, we explored five major problems in the area of biomedical/clinical text processing. We studied Named Entity Recognition and Clinical Coding in detail. Although significant progress has been made in these area, a lot remains to be explored in this emerging field. The models presented in the papers we studied are well suited for academic purpose, most of them are not viable for large scale/industrial use. Yet, looking at tremendous pace of research in BioNLP, it will not be an overstatement to say that in the near future, the NLP based systems will perform better than humans.

8 Individual Contributions

Tasks	Team Member contributed
1. Abstract and Introduction	Saksham Gupta
2. Datasets for BioNLP	Sathvik Bhagavan, Harshit Gupta and Akshay Kumar Arya
3. Problem Statements	Akshay Kumar Arya, Sathvik Bhagavan
4. Named Entity Recognition	Saksham Gupta
5. Clinical Coding	Sathvik Bhagavan, Harshit Gupta and Saurab Jain
6. Future Directions	All members
7. Conclusion	Sathvik Bhagavan, Saksham Gupta

References

[1] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[2] Ö. Uzuner, B. R. South, S. Shen, and S. L. DuVall, "2010 i2b2/va challenge on concepts, assertions, and relations in clinical text," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 552–556, 2011.

[3] H. Suominen, S. Salanterä, S. Velupillai, W. W. Chapman, G. Savova, N. Elhadad, S. Pradhan, B. R. South, D. L. Mowery, G. J. Jones, *et al.*, "Overview of the share/clef ehealth evaluation lab 2013," in *International Conference of the Cross-Language Evaluation Forum for European Languages*, pp. 212–231, Springer, 2013.

[4] S. Bethard, L. Derczynski, G. Savova, J. Pustejovsky, and M. Verhagen, "Semeval-2015 task 6: Clinical tempeval," in *proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 806–814, 2015.

[5] `https://ufal.mff.cuni.cz/ufal_medical_corpus`.

[6] Z. Wen, X. H. Lu, and S. Reddy, "MeDAL: Medical abbreviation disambiguation dataset for natural language understanding pretraining," in *Proceedings of the 3rd Clinical Natural Language Processing Workshop*, (Online), pp. 130–135, Association for Computational Linguistics, Nov. 2020.

[7] D. Jin and P. Szolovits, "PICO element detection in medical text via long short-term memory neural networks," in *Proceedings of the BioNLP 2018 workshop*, (Melbourne, Australia), pp. 67–75, Association for Computational Linguistics, July 2018.

[8] S. Moon, S. Pakhomov, and G. Melton, "Clinical abbreviation sense inventory," 2012.

[9] Q. Jin, B. Dhingra, Z. Liu, W. W. Cohen, and X. Lu, "Pubmedqa: A dataset for biomedical research question answering," *arXiv preprint arXiv:1909.06146*, 2019.

[10] V. Vincze, G. Szarvas, R. Farkas, G. Móra, and J. Csirik, "The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes," *BMC bioinformatics*, vol. 9, no. 11, pp. 1–9, 2008.

[11] A. Miranda, A. Gonzalez-Agirre, and M. Krallinger, "CodiEsp corpus: Spanish clinical cases coded in ICD10 (CIE10) - eHealth CLEF2020," Apr. 2020. Funded by the Plan de Impulso de las Tecnologías del Lenguaje (Plan TL).

[12] E. Lehman, J. DeYoung, R. Barzilay, and B. C. Wallace, "Inferring which medical treatments work from reports of clinical trials," *arXiv preprint arXiv:1904.01606*, 2019.

[13] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, "A simple algorithm for identifying negated findings and diseases in discharge summaries," *Journal of biomedical informatics*, vol. 34, no. 5, pp. 301–310, 2001.

[14] P. G. Mutalik, A. Deshpande, and P. M. Nadkarni, "Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the umls," *Journal of the American Medical Informatics Association*, vol. 8, no. 6, pp. 598–609, 2001.

[15] I. Goldin and W. W. Chapman, "Learning to detect negation with 'not'in medical texts," in *Proc Workshop on Text Analysis and Search for Bioinformatics, ACM SIGIR*, 2003.

[16] Z. Qian, P. Li, Q. Zhu, G. Zhou, Z. Luo, and W. Luo, "Speculation and negation scope detection via convolutional neural networks," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, (Austin, Texas), pp. 815–825, Association for Computational Linguistics, Nov. 2016.

[17] F. Fancellu, A. Lopez, and B. Webber, "Neural networks for negation scope detection," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, (Berlin, Germany), pp. 495–504, Association for Computational Linguistics, Aug. 2016.

[18] A. Khandelwal and S. Sawant, "Negbert: A transfer learning approach for negation detection and scope resolution," *arXiv preprint arXiv:1911.04211*, 2019.

[19] B. T. McInnes, T. Pedersen, Y. Liu, G. B. Melton, and S. V. Pakhomov, "Knowledge-based method for determining the meaning of ambiguous biomedical terms using information content measures of similarity," in *AMIA Annual Symposium Proceedings*, vol. 2011, p. 895, American Medical Informatics Association, 2011.

[20] E. Agirre, A. Soroa, and M. Stevenson, "Graph-based word sense disambiguation of biomedical documents," *Bioinformatics*, vol. 26, no. 22, pp. 2889–2896, 2010.

[21] R. Bruce and J. Wiebe, "Word-sense disambiguation using decomposable models," *arXiv preprint cmp-lg/9406005*, 1994.

[22] Y. K. Lee, H. T. Ng, and T. K. Chia, "Supervised word sense disambiguation with support vector machines and multiple knowledge sources," in *Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pp. 137–140, 2004.

[23] H.-L. Lai, H.-L. Hsu, J.-S. Liu, C.-H. Lin, and Y. Chen, "Supervised word sense disambiguation on polysemy with neural network models: A case study of bun in taiwan hakka," *International Journal of Asian Language Processing*, p. 2050011, 2021.

[24] H. Liu, V. Teller, and C. Friedman, "A multi-aspect comparison study of supervised word sense disambiguation," *Journal of the American Medical Informatics Association*, vol. 11, no. 4, pp. 320–331, 2004.

[25] D. Yuan, J. Richardson, R. Doherty, C. Evans, and E. Altendorf, "Semi-supervised word sense disambiguation with neural models," *arXiv preprint arXiv:1603.07012*, 2016.

[26] V. N. Garla and C. Brandt, "Knowledge-based biomedical word sense disambiguation: an evaluation and application to clinical document classification," *Journal of the American Medical Informatics Association*, vol. 20, pp. 882–886, 10 2012.

[27] M. Stevenson, Y. Guo, A. Alamri, and R. Gaizauskas, "Disambiguation of biomedical abbreviations," in *Proceedings of the BioNLP 2009 Workshop*, pp. 71–79, 2009.

[28] L. Plaza, M. Stevenson, and A. Díaz, "Improving summarization of biomedical documents using word sense disambiguation," in *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, pp. 55–63, 2010.

[29] J. Cowie and W. Lehnert, "Information extraction," *Communications of the ACM*, vol. 39, no. 1, pp. 80–91, 1996.

[30] A. Sarker and G. Gonzalez, "Portable automatic text classification for adverse drug reaction detection via multi-corpus training," *Journal of biomedical informatics*, vol. 53, pp. 196–207, 2015.

[31] Q. Li, S. A. Spooner, M. Kaiser, N. Lingren, J. Robbins, T. Lingren, H. Tang, I. Solti, and Y. Ni, "An end-to-end hybrid algorithm for automated medication discrepancy detection," *BMC medical informatics and decision making*, vol. 15, no. 1, pp. 1–12, 2015.

[32] K. Yadav, E. Sarioglu, M. Smith, and H.-A. Choi, "Automated outcome classification of emergency department computed tomography imaging reports," *Academic Emergency Medicine*, vol. 20, no. 8, pp. 848–854, 2013.

[33] X. Ji, Y. Tu, W. He, J. Wang, H.-W. Shen, and P.-Y. Yen, "Usevis: Visual analytics of attention-based neural embedding in information retrieval," *Visual Informatics*, 2021.

[34] 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.

[35] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications," *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 507–513, 2010.

[36] H. Liu, S. J. Bielinski, S. Sohn, S. Murphy, K. B. Wagholikar, S. R. Jonnalagadda, K. Ravikumar, S. T. Wu, I. J. Kullo, and C. G. Chute, "An information extraction framework for cohort identification using electronic health records," *AMIA Summits on Translational Science Proceedings*, vol. 2013, p. 149, 2013.

[37] H. Cho and H. Lee, "Biomedical named entity recognition using deep neural networks with contextual information," in *BMC Bioinformatics*, vol. 6, 2019.

[38] S. Zhao, "Named entity recognition in biomedical texts using an hmm model," in *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP)*, pp. 87–90, 2004.

[39] A. B. Abacha and P. Zweigenbaum, "Medical entity recognition: A comparaison of semantic and statistical methods," in *Proceedings of BioNLP 2011 Workshop*, pp. 56–64, 2011.

[40] W. Yonghui, "Clinical named entity recognition using deep learning models," in *Annual Symposium proceedings*, 2017.

[41] M. Gridach, "Character-level neural network for biomedical named entity recognition," in *Journal of Biomedical Informatics*, 2017.

[42] Z. Liu, M. Yang, and X. Wang, "Entity recognition from clinical texts via recurrent neural network," in *BMC Medical Informatics and Decision Making, vol. 17*, 2017.

[43] D. Zeng, C. Sun, L. Lin, and B. Liu, "LSTM-CRF for Drug-Named Entity Recognition," in *Entropy, vol. 19*, 2017.

[44] C. Wang, X. abd Yang and R. Guan, "A comparative study for biomedical named entity recognition," in *Int. J. Mach. Learn. Cyber., vol. 9*, 2018.

[45] H. El Boukkouri, O. Ferret, T. Lavergne, and P. Zweigenbaum, "Embedding strategies for specialized domains: Application to clinical entity recognition," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, (Florence, Italy), Association for Computational Linguistics, July 2019.

[46] M. Hofer, A. Kormilitzin, P. Goldberg, and A. J. Nevado-Holgado, "Few-shot learning for named entity recognition in medical text," *CoRR*, vol. abs/1811.05468, 2018.

[47] L. S. Larkey and W. B. Croft, "Combining classifiers in text categorization," in *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '96, (New York, NY, USA), p. 289–297, Association for Computing Machinery, 1996.

[48] R. Farkas and G. Szarvas, "Automatic construction of rule-based icd-9-cm coding systems," in *BMC bioinformatics*, vol. 9, pp. 1–9, Springer, 2008.

[49] J. Mullenbach, S. Wiegreffe, J. Duke, J. Sun, and J. Eisenstein, "Explainable prediction of medical codes from clinical text," *arXiv preprint arXiv:1802.05695*, 2018.

[50] F. Li and H. Yu, "Icd coding from clinical text using multi-filter residual convolutional neural network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 8180–8187, 2020.

[51] H. Schäfer, "Multilingual icd-10 code assignment with transformer architectures using mimic-iii discharge summaries," 2020.

[52] Z. Dai, Z. Yang, Y. Yang, W. Cohen, J. Carbonell, Q. Le, and R. T.-X. Salakhutdinov, "Attentive language models beyond a fixed-length context. arxiv 2019," *arXiv preprint arXiv:1901.02860*.

[53] Z. Miftahutdinov and E. Tutubalina, "Kfu at clef ehealth 2017 task 1: Icd-10 coding of english death certificates with recurrent neural networks.," in *CLEF (Working Notes)*, 2017.

[54] M. Polignano, V. Suriano, P. Lops, M. de Gemmis, and G. Semeraro, "A study of machine learning models for clinical coding of medical reports at codiesp 2020," in *Working Notes of Conference and Labs of the Evaluation (CLEF) Forum. CEUR Workshop Proceedings*, 2020.

[55] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[56] Z. Zhang, J. Liu, and N. Razavian, "Bert-xml: Large scale automated icd coding using bert pretraining," 2020.

[57] R. You, Z. Zhang, Z. Wang, S. Dai, H. Mamitsuka, and S. Zhu, "Attentionxml: Label tree-based attention-aware deep model for high-performance extreme multi-label text classification," 2019.

[58] A. Sonabend, W. Cai, Y. Ahuja, A. Ananthakrishnan, Z. Xia, S. Yu, and C. Hong, "Automated icd coding via unsupervised knowledge integration (unite)," *International journal of medical informatics*, vol. 139, p. 104135, 2020.

[59] S. Sheikhalishahi, R. Miotto, J. Dudley, A. Lavelli, F. Rinaldi, and V. Osmani, "Natural language processing of clinical notes on chronic diseases: Systematic review," in *JMIR Med Inform.*, 2019.

[60] I. Spasic and G. Nenadic, "Clinical text data in machine learning: Systematic review," in *JMIR Med Inform.*, 2020.

[61] P. Colombo, W. Witon, A. Modi, J. Kennedy, and M. Kapadia, "Affect-driven dialog generation," in *NAACL-HLT*, (Minneapolis, USA), 2019.

[62] E. Moons, A. Khanna, A. Akkasi, and M.-F. Moens, "A comparison of deep learning methods for icd coding of clinical records," *Applied Sciences*, vol. 10, p. 5262, 07 2020.

[63] Y. Wang, K. Zheng, H. Xu, and Q. Mei, "Interactive medical word sense disambiguation through informed learning," *Journal of the American Medical Informatics Association*, vol. 25, no. 7, pp. 800–808, 2018.

[64] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Lingvisticae Investigationes*, vol. 30, no. 1, pp. 3–26, 2007.

[65] R. Morante, A. Liekens, and W. Daelemans, "Learning the scope of negation in biomedical texts," in *Proceedings of the 2008 conference on empirical methods in natural language processing*, pp. 715–724, 2008.

[66] R. Navigli, "Word sense disambiguation: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 2, pp. 1–69, 2009.

[67] N. Bach and S. Badaskar, "A review of relation extraction," *Literature review for Language and Statistics II*, vol. 2, pp. 1–15, 2007.