



TermInformer: unsupervised term mining and analysis in biomedical literature

Prayag Tiwari¹ · Sagar Uprety² · Shahram Dehdashti³ · M. Shamim Hossain⁴

Received: 17 June 2020 / Accepted: 2 September 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract

Terminology is the most basic information that researchers and literature analysis systems need to understand. Mining terms and revealing the semantic relationships between terms can help biomedical researchers find solutions to some major health problems and motivate researchers to explore innovative biomedical research issues. However, how to mine terms from biomedical literature remains a challenge. At present, the research on text segmentation in natural language processing (NLP) technology has not been well applied in the biomedical field. Named entity recognition models usually require a large amount of training corpus, and the types of entities that the model can recognize are limited. Besides, dictionary-based methods mainly use pre-established vocabularies to match the text. However, this method can only match terms in a specific field, and the process of collecting terms is time-consuming and labour-intensive. Many scenarios faced in the field of biomedical research are unsupervised, i.e. unlabelled corpora, and the system may not have much prior knowledge. This paper proposes the TermInformer project, which aims to mine the meaning of terms in an open fashion by calculating terms and find solutions to some of the significant problems in our society. We propose an unsupervised method that can automatically mine terms in the text without relying on external resources. Our method can generally be applied to any document data. Combined with the word vector training algorithm, we can obtain reusable term embeddings, which can be used in any NLP downstream application. This paper compares term embeddings with existing word embeddings. The results show that our method can better reflect the semantic relationship between terms. Finally, we use the proposed method to find potential factors and treatments for lung cancer, breast cancer, and coronavirus.

Keywords Term mining · Unsupervised learning · Term embeddings · Sequence labelling · GloVe · Biomedical literature

✉ Prayag Tiwari
prayag.tiwari@unipd.it

✉ M. Shamim Hossain
mshossain@ksu.edu.sa

Sagar Uprety
sagar.uprety@open.ac.uk

Shahram Dehdashti
shahram.dehdashti@qut.edu.au

³ School of Information Systems, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia

⁴ Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

¹ Department of Information Engineering, University of Padova, Padua, Italy

² The Open University, London, UK

1 Introduction

Term mining aims to mine terms from unstructured documents. Terminology is usually composed of multiple words and describes concepts in a particular domain that forms the basis of the domain. Unsupervised term mining aims to use automated algorithms to mine terms in the literature without relying on external resources and human intervention. Therefore, the method has a wide range of applications and can be used to process text data in any field. Existing research usually uses dictionary-based methods and supervised machine learning methods to mine terms in the literature. However, the dictionary-based method's limitation is that the method can only match terms in a specific field, and terms in the vocabulary are often different from expressions in the literature. For example, there are more than 5 ways in the literature to mention the disease "ischemic stroke". Terms in the literature contain many variants, so dictionary-based methods require constant vocabulary maintenance, which is costly and time-consuming.

The named entity recognition method has achieved good results in terms of recognition. However, the named entity recognition model requires a large number of labelled corpora for training and can only identify predefined types of terms, which is not suitable for this open term mining problem. Corpus annotation in the biomedical field is very challenging because of the high requirements for the knowledge background of the annotators, and the labelling process relies heavily on the knowledge of domain experts and annotation standards. Therefore, the method of manually labelling the term corpus for named entity recognition is very time-consuming and costly. In addition, using deep learning models to identify terms can also lead to increased computation costs. Based on these problems, this paper proposes an unsupervised term mining method. This method can automatically mine terms from the literature without the need to annotate the corpus manually, so it can work with word embedding training algorithms to produce term-level embeddings, not just word embeddings. We propose a multi-length term mining algorithm. Using this algorithm, we can fully mine terms from the text without requiring any external resources.

Existing term representation methods usually represent a term by average word embeddings [1]. However, this method cannot obtain the semantics of the term. This method only represents the term by each compound word. There are essential differences between terms and words, so this poses a limitation of term research. Faced with this problem, we apply the proposed method to train term embeddings. Based on the existing word embedding algorithm GloVe [2], we trained the mined terms and found that

the term representation can better represent the semantic relationship between terms. To evaluate the performance of the method, we created two datasets of different sizes. We compared the term representation composed of the original word embeddings with the term representation obtained based on our method. We observe that our method improves the representation of terms, can better characterize the relationships between terms, and better calculate the similarity between terms. This method can potentially be applied to any biomedical text mining system. The proposed method can also be used to build a term correlation graph. Finally, we explore the factors and treatments for lung and breast cancer using the proposed method. The results show that our method can find some key information for these diseases from the literature.

1.1 Contribution

The main contributions of this paper can be summarized below:

1. We propose an unsupervised term mining algorithm. The proposed method can be applied to any biomedical corpus, and the process can mine terms without manual annotation and can be used with word embedding training algorithms, which may become a scheme for solving term representation problems.
2. The proposed method improves the existing term representation based on word embeddings, and the obtained term embeddings can better characterize the relationship between terms.
3. This paper creates two-term mining datasets to evaluate the performance of the method.

1.2 Organization

The rest of the paper is organized as follows. Related works are discussed in Sect. 2. Section 3 describes the method proposed in this paper, followed by the proposed algorithm. Experiment results are explained in Sect. 4 consisting of experimental results, including dataset description, evaluation metric, analysis of term similarity, analysis of term relationship. Finally, we discuss the conclusion and possible future work in Sect. 5.

2 Related work

Named entity recognition models are widely used for recognizing biomedical terms. Settles et al. [3] use conditional random fields (CRF) [4] to recognize the gene and protein mentions in biomedical abstracts. Leaman et al. [5] propose the BANNER framework to recognize biomedical

entities, which aims to improve the generalization ability for this task. Habibi et al. [6] use the LSTM-CRF model to recognize genes, chemicals, and diseases mentions. Tang et al. [7] study three different types of word representation methods and analyze their performance for biomedical NER on JNLPBA and BioCreAtIvE II BNER tasks. Yao et al. [8] propose a deep learning model that consists of multiple CNN layers and achieves improvement on the GENIA dataset. Wang et al. [9, 10] propose a multitask learning approach to recognize biomedical entities by using training data collectively consisting of distinct types of entities. Yoon et al. [11] propose CollaboNet to resolve the issue due to lack of data and entity-type misclassification by using the integration of multiple NER frameworks. Cho et al. [12] use BiLSTM and CRF to propose contextual LSTM networks with CRF to capture all the contextual information.

Lafferty et al. [13] use conditional random fields (CRF) to build probabilistic models for sequence labelling problem. Nadeau et al. [14] investigate the feature engineering-based NER models and systems. Collobert et al. [15] use CNN to solve many NLP tasks. Lample et al. [16] adopt LSTM-CRF to solve the sequence labelling problem. Chiu et al. [17] propose to use bidirectional LSTM-CNNs to resolve the sequence labelling problem. Ma et al. [18] use lstm-cnns-crf model to recognize entities. Akbik et al. [19] propose to use character-level language modelling to improve performance.

Pre-trained language models, such as ELMo [20] and BERT [21], have also been applied in the clinical NLP field. Beltagy et al. [22] train the SciBERT to enhance downstream NLP tasks. Alsentzer et al. [23] train the BERT model based on both clinical notes and discharge summaries. Lee et al. [24] proposed BioBERT, a model that retrains BERT on PubMed and PMC corpora, which can improve the results of downstream BioBLP tasks. These studies focus on word-level representations without considering the term representation. Context-dependent representations make the same word have different representations in different sentences. However, this paper aims to obtain context-independent representations, so we do not adopt these methods.

3 Methods

3.1 Multi-length term mining algorithm

This section explains this algorithm. After statistical analysis, we found that terms composed of 2, 3, 4, and 5 words are the more common, so we mainly mine terms of the above length. Word vectors can directly represent terms with only one word. The input of the algorithm is the

original text, and the output is the mined terms. We do not need to use any external resources to apply it to any corpus in this process. We first perform word segmentation and part-of-speech tagging on the original text and then mine the terms.

As shown in Algorithm 1, the first line of the algorithm is mainly to initialize 4 dictionaries for storing terms of different lengths and input the corpora. The method then starts processing each sample, that is, each article. The fourth line performs word segmentation for these articles. Word segmentation is to divide these articles into a sequence of words and identify the punctuations, which can prevent the words and punctuations from being connected, resulting in an irregular vocabulary and inaccurate words.

Algorithm 1 Multi-length term mining algorithm

Require: The corpus D

Ensure: The mined terms

```

1: Initialize dictionary  $d2, d3, d4, d5$  and corpus  $D$ 
2: Initialize the word segmentation and pos tagging modules
3: for  $i \leftarrow 0, D.length - 1$  do
4:    $tokens = \text{Tokenize}(D[i])$ 
5:    $pos = \text{POSTagging}(D[i])$ 
6:    $\text{Extract}(tokens, pos, d2, 2)$ 
7:    $\text{Extract}(tokens, pos, d3, 3)$ 
8:    $\text{Extract}(tokens, pos, d4, 4)$ 
9:    $\text{Extract}(tokens, pos, d5, 5)$ 
10: end for
11: return  $d2, d3, d4, d5$ 

```

The fifth line is mainly used for part-of-speech (POS) tagging of words. We adopt the LSTM-CRF neural network for POS tagging. The detailed mechanism of this model will be introduced in Sect. 3.2. The POS features help in improving the term mining in the algorithm. Lines 6 to 9 are used to mine terms of different lengths. We mainly focus on terms composed of 2, 3, 4, and 5 words.

The term extraction algorithm is described in Algorithm 2. The second line splits the input word sequence into phrases of a certain length. The third line is to extract the POS tags of the phrase from the recognized POS tag sequence. In line 4, this algorithm matches the POS of the target phrase with our predefined rule. This rule considers the extraction of medical terms where the first two words are adjective, and noun, respectively, and the last term is also a noun. If the phrase matches successfully, the phrase will be treated as a potential term and added to the dictionary. Then, this algorithm counts how often the term appears. If the term has appeared in the dictionary, increase the term frequency by 1. If the term does not appear in the dictionary, the term has a frequency of 1. The purpose of counting term frequency is to extract meaningful terms. By setting thresholds, we have the flexibility to mine a certain number of terms. Terms are often repeated in the literature, and if a phrase appears only once, we do not consider it as a

term. If we set a higher threshold, it means that we can get more confident terms, which also demonstrates that these terms are more common.

Algorithm 2 Term extraction function

```

Require: word segmentation tokens, POS tags pos, Term dictionary dict, term length leng
Ensure: The updated dictionary
1: for  $i \leftarrow 0, tokens.length - 1$  do
2:    $term = tokens[i:i+leng]$ 
3:    $tag = pos[i:i+leng]$ 
4:   if  $matched(tag)$  then
5:     if  $term$  in dict then
6:        $dict[term]++$ 
7:     else
8:        $dict[term]=1$ 
9:     end if
10:  end if
11: end for
    
```

In the following, we briefly discuss the asymptotic complexity of our approach. The algorithm describes the processing of a large number of samples, where the number of samples is related to the corpus’s size. Therefore, we analyze the term mining process for only one sample. The algorithm first performs word segmentation and then performs part-of-speech tagging and then calls Algorithm 2. Since the complexity of Algorithm 2 is $\mathcal{O}(n)$, the complexity of Algorithm 1 mainly depends on the two steps of word segmentation and part-of-speech tagging. Therefore, our algorithm is approximately equal to the complexity of segmentation and part-of-speech tagging. The proposed method does not significantly increase computational costs.

3.2 LSTM-CRF

This subsection introduces the LSTM-CRF sequence labelling model for POS tagging, as shown in Fig. 1. This model does not depend on feature engineering; instead, it adopts the words and characters as input. This design can increase the generality for processing any dataset. Then, textual input is projected to the word embeddings, which is a way to encode the prior knowledge of semantics into a dense vector representation [25–27].

3.2.1 Long short-term memory

The long short-term memory (LSTM) network is a kind of recurrent neural network (RNN). It uses the LSTM unit [28] to solve the exploding and vanishing gradient problems encountered in the traditional RNNs. The formulations of LSTM unit are as follows.

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{1}$$

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \tag{2}$$

$$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \tag{3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{4}$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \tag{5}$$

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

where $\sigma(\cdot)$ is a sigmoid activation function. x_t represents the input vector at time step t , and h_t denotes the hidden state containing the context information in the former time

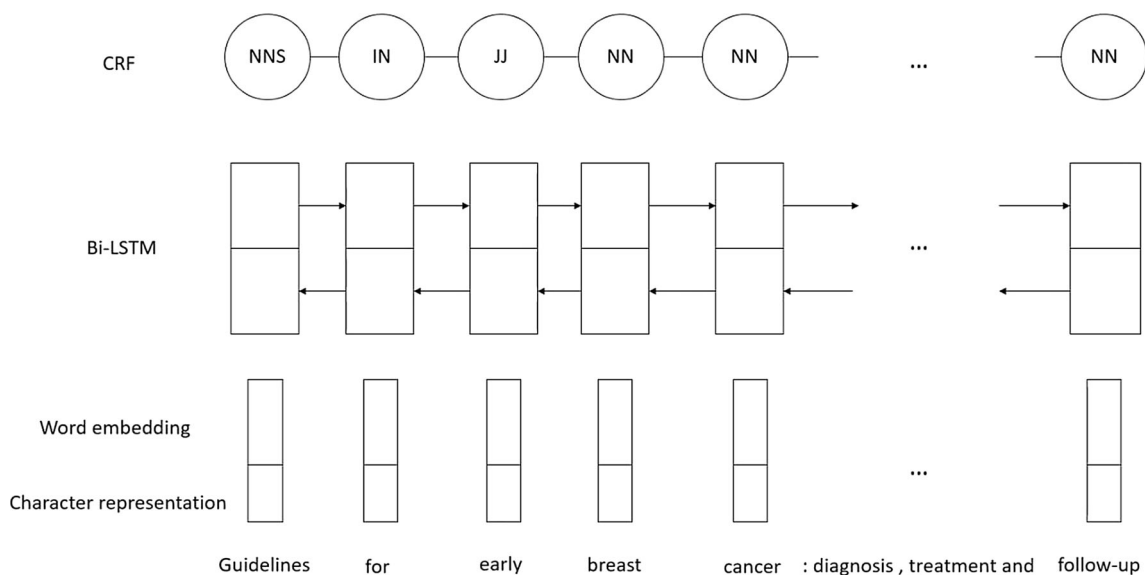


Fig. 1 LSTM-CRF network overview

steps. W and b are weight and bias parameters. i_t, f_t, c_t and o_t are the input gate vector, forget gate vector, cell state, and output gate vector, respectively. However, the hidden state in a forward LSTM can only capture the context information in the left side of current step [13, 16, 18]. Bidirectional LSTM (Bi-LSTM) has an operation to reverse the order of input sequence and concatenate the hidden state in each time step, which can capture the context information from the left and right side of the current step.

3.2.2 Conditional random field

The conditional random field (CRF) layer performs the label sequence prediction [16, 18]. This model can be more effective than classifying each word independently because the word label is determined not only by itself but also by the context.

$$p(y|z; W, b) = \frac{\prod_{i=1}^n \exp(W_{y_{i-1}y_i}^T z_i + b_{y_{i-1}y_i})}{\sum_{y' \in Y(z)} \prod_{i=1}^n \exp(W_{y'_{i-1}y'_i}^T z_i + b_{y'_{i-1}y'_i})} \quad (7)$$

$$L(W, b) = \sum_i \log p(y|z; W, b) \quad (8)$$

$$y^* = \arg \max_{y \in Y(z)} p(y|z; W, b) \quad (9)$$

where $\{[z_i, y_i]\}, i = 1, 2, \dots, n$ denotes a set of words z with a label sequence y . W and b are weight and bias parameters. $p(y|z; W, b)$ is the probability of label sequence y over all possible sequences $Y(z)$ on the input z . In training stage, the objective is to maximize the log-likelihood $L(W, b)$. In prediction, the decoder will find the optimal label sequence as Eq. 9 by Viterbi algorithm [29].

3.3 Term embedding

Word vector approaches aim to project a large vector space of words into a much lower-dimensional space and generate dense representations. It has made significant contributions to enhance various NLP techniques and has been widely used in various downstream NLP tasks like sentiment analysis, document classification, etc., to achieve improved results. The word representations are computable and have actively promoted the development of deep learning NLP. Existing word vector training methods mainly embed words into a fixed-length vector, but cannot be directly used to learn term vector in sentences. For terms, the word embeddings are not enough to represent the overall meaning of the term. Existing methods have limitations on the problem of term representation.

Existing methods usually use average or max pooling of all word vectors to represent a multi-word term. The

problem is that the term representation obtained by this method will make the term most similar to each composed word. This method cannot reflect the relationship between different terms, but only the relationship between words contained in the term. This method belongs to a word-level learning method and cannot be used in the term level. Based on the above problems, we use the multi-length term mining algorithm proposed in Sect. 3.1 together with the word vector learning algorithm to fundamentally alleviate the problem of term representation and learn term embeddings. We use the GloVe algorithm to train term embeddings. However, the major difference is that our vocabulary consists of terms mined using the algorithms described above. This enables the GloVe algorithm to learn vectors for terms rather than individual words. For example, it will learn one single vector for the compound term “lung cancer”. GloVe will learn vectors for the two terms “lung” and “cancer”, and then one has to average them to obtain the vector for “lung cancer”. The training objective for GloVe embeddings is shown in equation (10).

$$J = \sum_{i,j=1}^v f(X_{i,j})(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij})^2 \quad (10)$$

where X_{ij} represents the co-occurrence frequency of the word_{*i*} and word_{*j*}. w_i and w_j denote vector representations of the word_{*i*} and word_{*j*}, respectively. b_i and b_j are the bias value for the word_{*i*} and word_{*j*}, respectively. $f(X_{i,j})$ is the weighting factor defined in equation (11).

$$f(x) = \begin{cases} (x/x_{\max})^\alpha, & \text{if } x < x_{\max} \\ 1, & \text{otherwise} \end{cases} \quad (11)$$

where x_{\max} and α are hyperparameters. These two parameters are set to $x_{\max} = 100$ and $\alpha = 0.75$.

4 Results

In this section, we conduct experiments on two datasets. We first introduce the datasets, evaluation methods. Then, we compare the mined term-based embeddings with the word embedding-based method. Finally, we analyze the experimental results.

Table 1 Number of terms mined on different datasets

Term length	2 words	3 words	4 words	5 words	Total
PubMed-10K	34,362	7304	1090	129	42,885
PubMed-100K	268,890	72,612	11,267	1,545	354,314

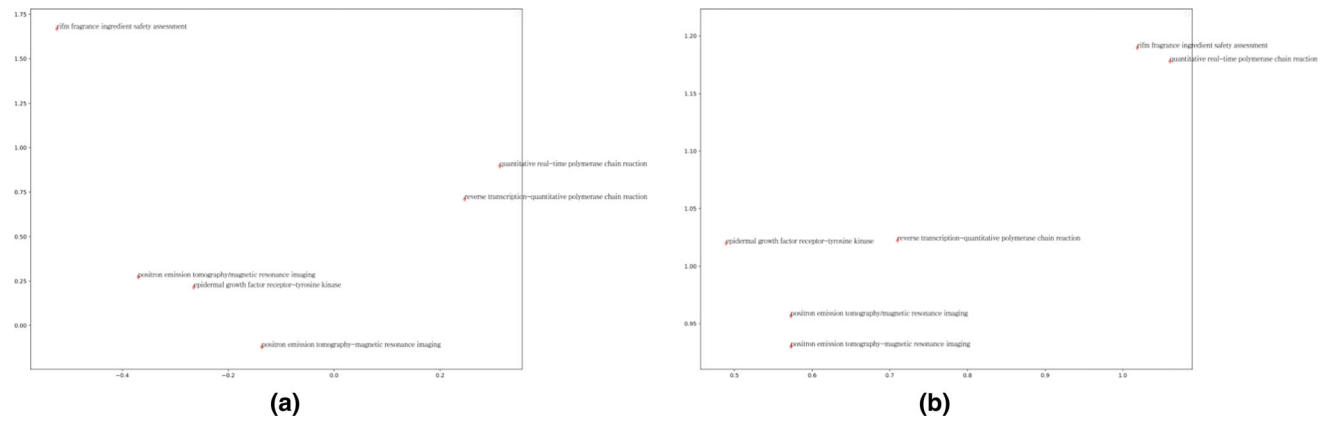


Fig. 3 Visualization of term embeddings where each term consists of 5 words, and a and b are based on our method and word embeddings

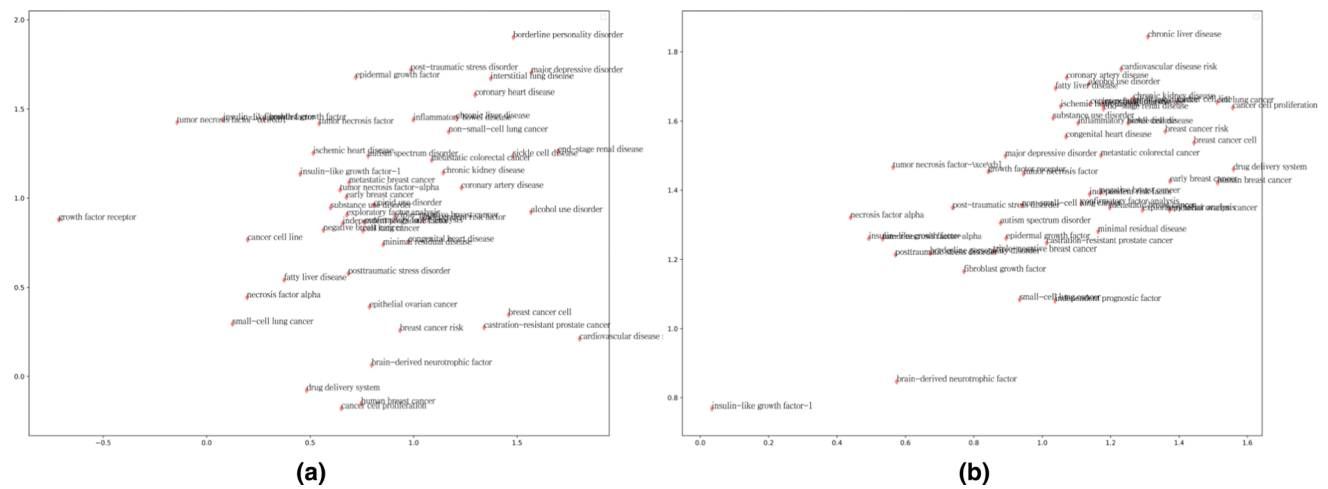


Fig. 4 Visualization of term embeddings where each term consists of 3 words, and a and b are based on our method and word embeddings

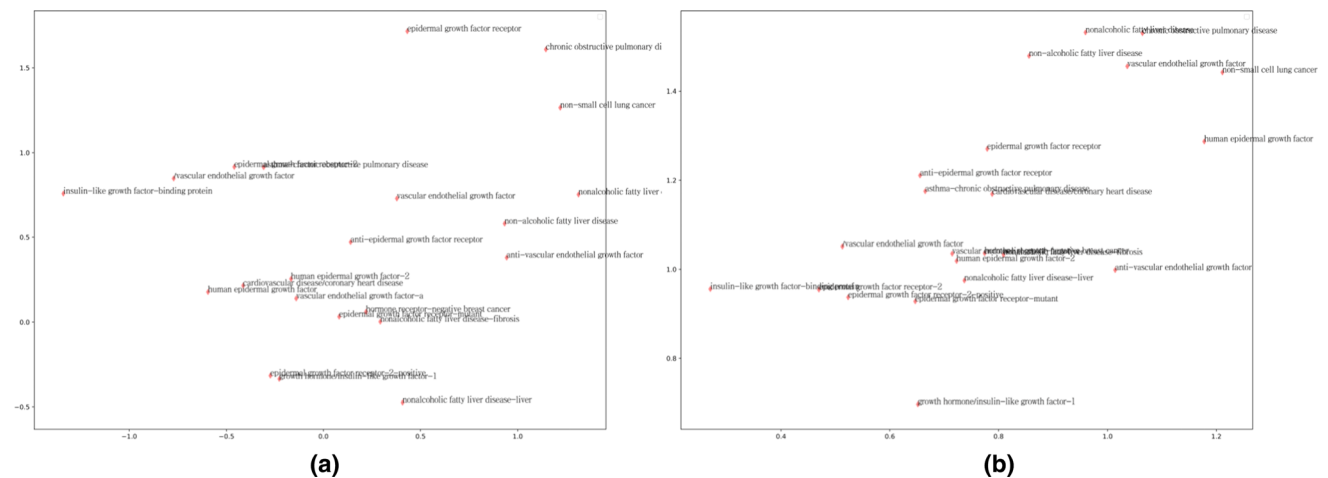


Fig. 5 Visualization of term embeddings where each term consists of 4 words, and a and b are based on our method and word embeddings

ID	Terms	Ours	Baseline
1	chronic obstructive pulmonary disease	non-small-cell lung cancer copd obstructive sleep apnea tetanus congenital polyq prematurity coronary heart disease ulcerative colitis stereotactic	kidney heart acute syndrome thromboembolic ductal-dependent
2	nuclear factor-kappab	n-glycome 1a1 blattella pre-rrna leptochloa left- mca2+ atmosphere-surface acetoin shipbuilding	full-scope hash1 factor-κb reconstituting kappa-b resonance factor-erythroid magnetic κb
3	liquid chromatography-tandem mass spectrometry	lc-ms penalized coptotermes multi-slice spme -0.82 31p-nmr qpcr uhplc-hrms spectral-domain	chromatography energy-resolved chromatography-mass coupled ultraviolet-visible tandem thin-layer
4	alcohol use disorder	acute respiratory hyaluronic acid fundus motor femoropopliteal echocardiographic ptpe endoscopic ct artifacts	symptoms substance consumption chronic exposure opioid smoking

Fig. 6 The most closely related terms base on the PubMed-10k dataset

4.4 Analysis of term relationship

In this subsection, we analyze the difference between terms of various lengths for learning term embeddings. The baseline is the most commonly used term representation method based on word vectors. We apply principal component analysis (PCA) [30] to reduce the dimension of the learned term, which is convenient for a visualization based on low dimensions to observe and measure the semantic similarity between terms. As shown in Figs. 8 and 9, we found that term embeddings learned by our method can better reflect the relationship between terms. For example, disease-related terms are relatively close, but the word

vector-based term representation does not reveal the semantic relationship between terms well. The baseline method mainly retains the words' similarity, so this method cannot get the term similarity very well. We further find that the longer the term, the less accurate the term relationship based on the word vector method, and the more obvious the need for term embedding. So the proposed method can learn a reasonable representation for the longer term.

We analyze the performance of the proposed method on different types of terms. As shown in Fig. 8, for cancer terms, our method can achieve a more accurate semantic distribution, so that semantically related terms are closer. For example, we found “non-small cell lung cancer” is

ID	Terms	Ours	Baseline
1	toll-like receptor	tlr4 tlr2 tlr hsp27-mediated signaling pathway wnt nf- κ b nrf2 cannabinoid tlr9	tlr4 agonists signaling agonist antagonists antagonist tlr activation
2	human immunodeficiency virus	hiv hiv-1 persons living hiv infection hepatitis c virus tuberculosis hcv type infection	infection hiv-1 viral influenza hiv bovine hepatitis
3	chronic obstructive pulmonary disease	copd exacerbation exacerbations asthma multiple sclerosis allergic rhinitis chronic kidney disease ild inflammatory bowel disease heart failure	copd severe acute hypertension lung kidney
4	non-alcoholic fatty liver disease	nafld nonalcoholic fatty liver disease aplastic kpdm aspirin-exacerbated nash hbs b-acute lymphoblastic leukemia insulin resistance bpw	nafld nonalcoholic kidney fibrosis hepatic diseases
5	inflammatory bowel disease	ibd ulcerative colitis cardiomyopathy multiple sclerosis rheumatoid arthritis acute myeloid leukemia chronic obstructive pulmonary disease 13 biosimilar-treated colorectal cancer hbeag-negative	inflammation chronic ibd syndrome progression systemic pathogenesis

Fig. 7 The most closely related terms base on the PubMed-100k dataset

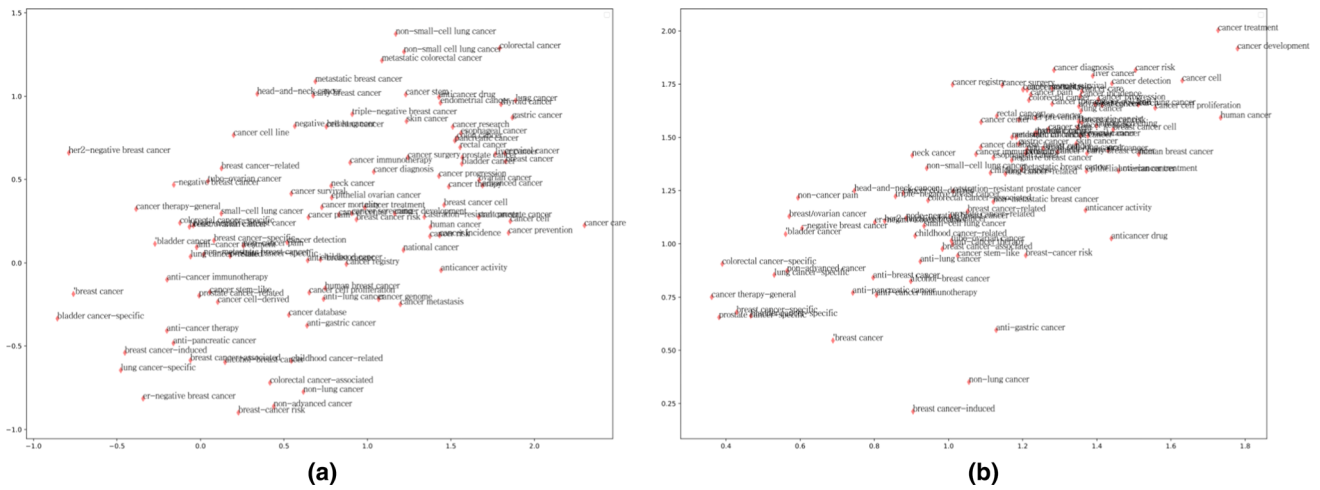


Fig. 8 Visualization of cancer and **a** and **b** are based on our method and word embeddings

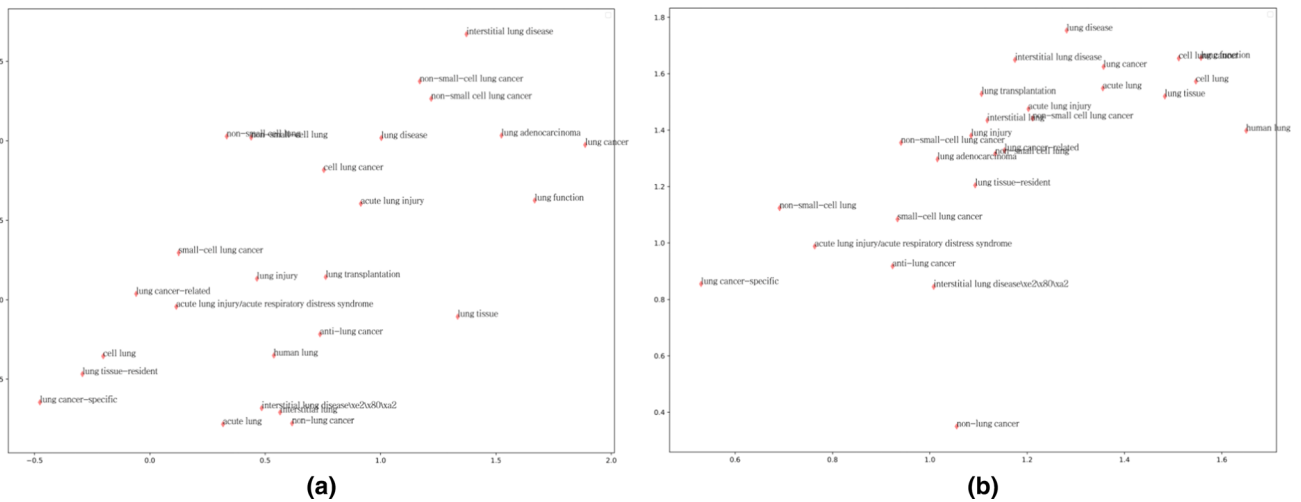


Fig. 9 Visualization of lung disease and **a** and **b** are based on our method and word embeddings

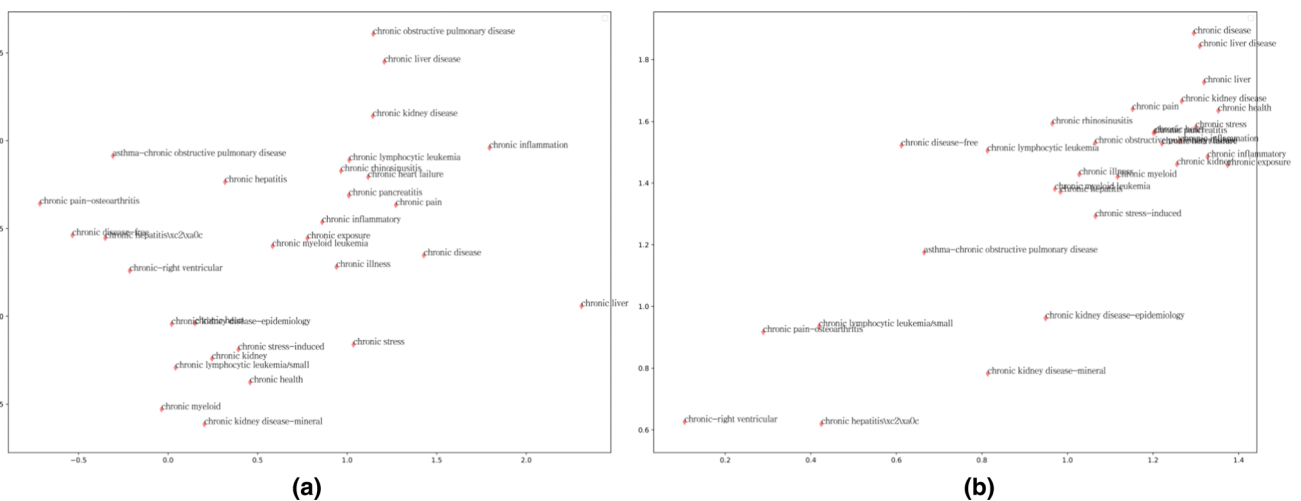


Fig. 10 Visualization of chronic diseases and **a** and **b** are based on our method and word embeddings

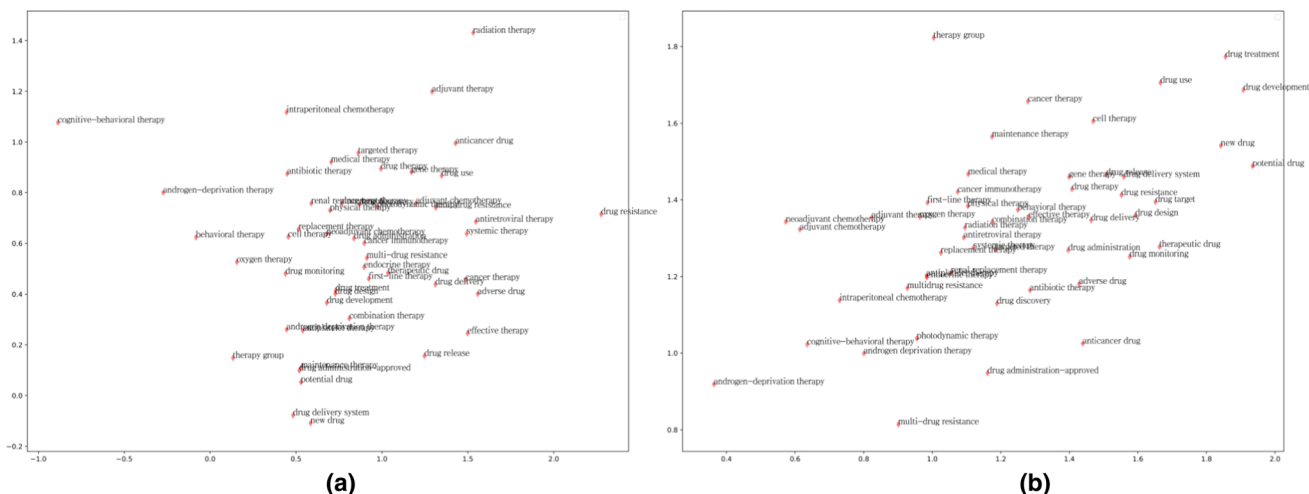


Fig. 11 Visualization of drug and therapy, and a and b are based on our method and word embeddings

closely related to “colorectal cancer”. There is potential relation between “epithelial ovarian cancer” and “neck cancer”, “early breast cancer” and “head-and-neck cancer”. This characteristic helps find related treatment schemes from other terms. However, the word vector-based method makes the terms mixed.

Figure 10 shows the terms related to chronic diseases. Our method brings together similar diseases. We find some potential links between chronic diseases, such as “chronic pancreatitis” and “chronic heart failure”, “chronic rhinosinusitis” and “chronic lymphocytic leukemia”. The baseline method does not generate this effect, so our

method’s advantage is that it can compare similar chronic diseases to find treatment options.

Terms related to drugs and treatments are shown in Fig. 11. Our method can put related drugs and treatment methods together, which helps medical researchers choose the corresponding treatment plan and recommend more treatment plans. We observe “neoadjuvant chemotherapy” and “cancer immunotherapy” are closely related. The baseline method has no obvious semantic characteristics.

Figure 9 shows the terms related to lung diseases. Our method reveals the relationship between lung diseases. The baseline method does not reveal this relationship.

Fig. 12 Visualization of factors for breast cancer (blue dot) (color figure online)

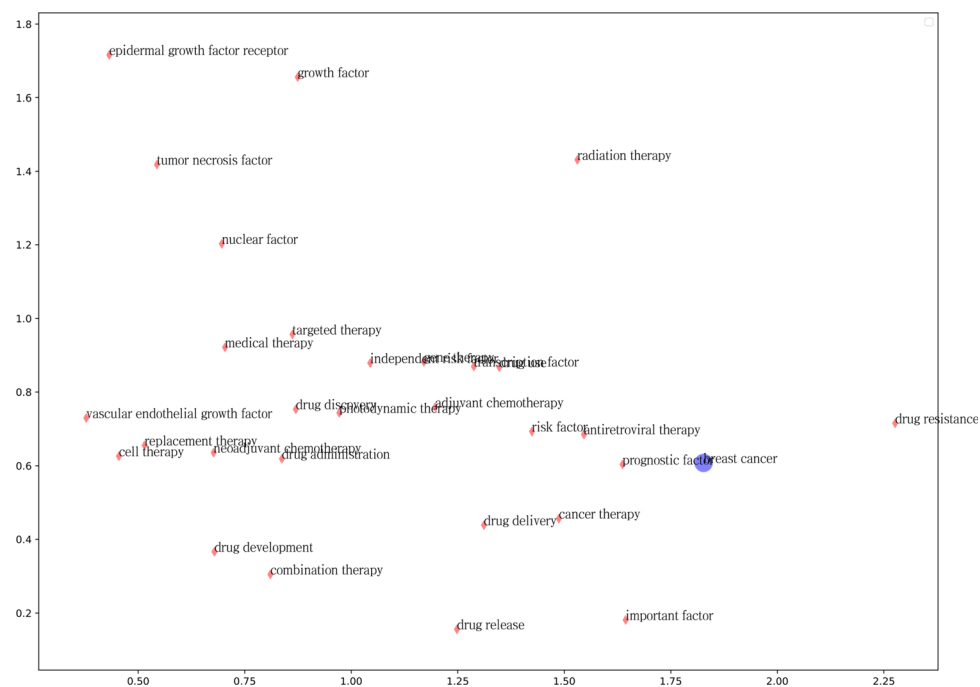


Fig. 13 Visualization of factors for lung cancer (blue dot) (color figure online)

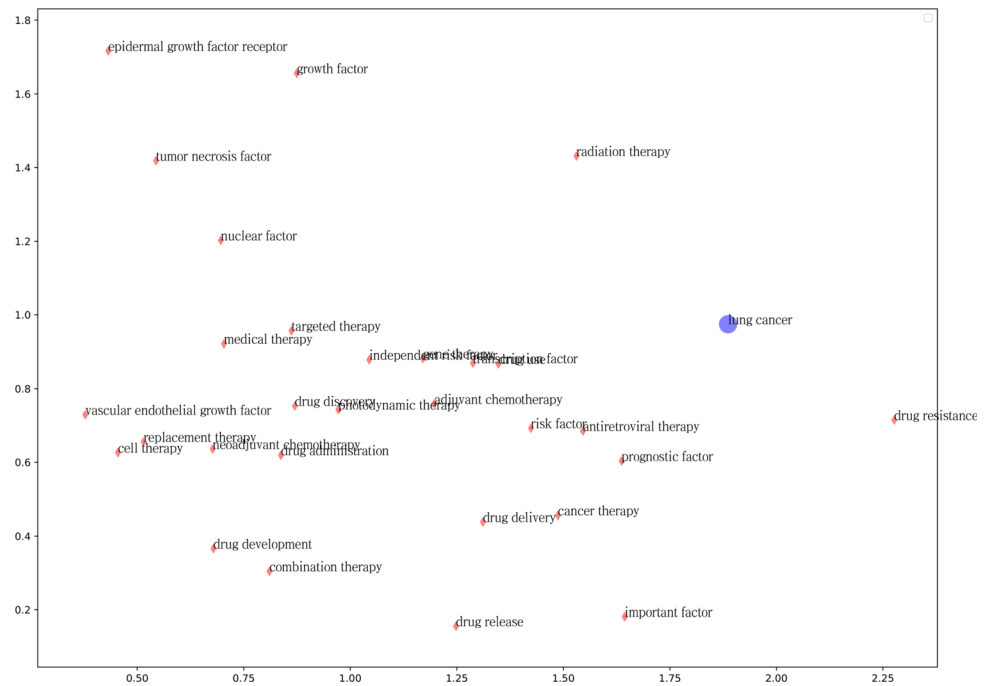
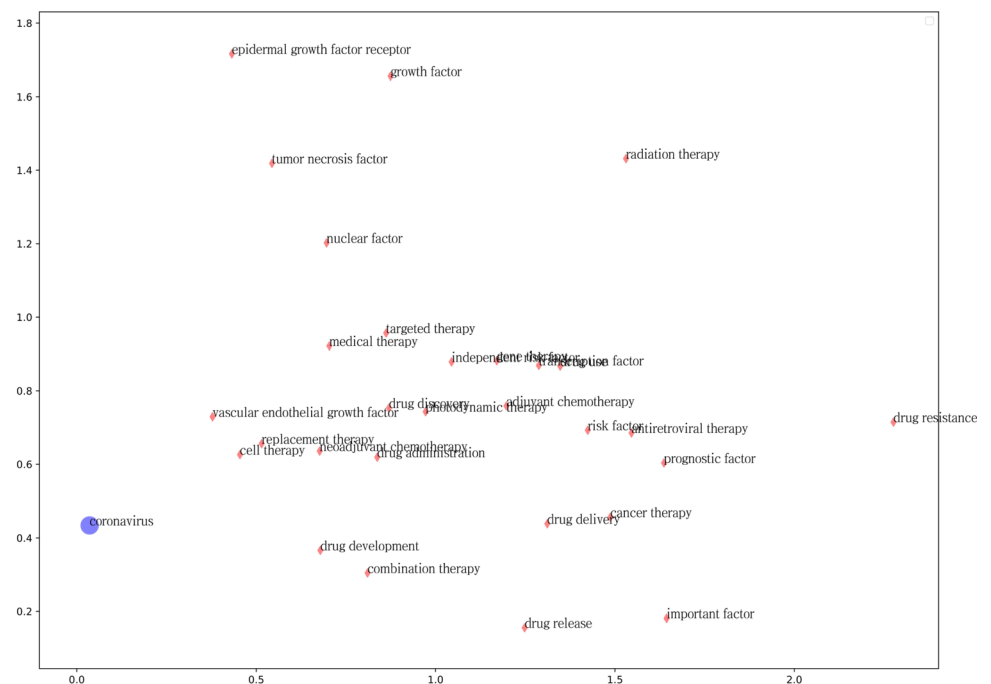


Fig. 14 Visualization of factors for coronavirus (blue dot) (color figure online)



Therefore, our method can be further used to find drugs to treat lung diseases.

We visualize factors and treatments closely related to lung cancer, breast cancer, and coronavirus, as shown in Figs. 12, 13, and 14, respectively. We observe the lung cancer is closely related to “antiretroviral therapy, radiation therapy, gene therapy, adjuvant chemotherapy, prognostic factor, nuclear factor, targeted therapy, photodynamic therapy”. The breast cancer is closely

related to “prognostic factor, antiretroviral therapy, adjuvant chemotherapy, radiation therapy, gene therapy, photodynamic therapy”. “tumour necrosis factor, epidermal growth factor receptor, nuclear factor, targeted therapy, combination therapy” has more inspiration for the treatment of breast cancer. The coronavirus is closely related to “vascular endothelial growth factor”, “cell therapy”, “replacement therapy”, “neoadjuvant chemotherapy”, “drug development”, “combination therapy”, “photodynamic

therapy”, etc. These results show that our method can learn term embeddings from a large-scale corpus to generate inspiration for diseases’ treatment.

5 Conclusion

In this paper, an unsupervised term mining method has been proposed for mining terms from a biomedical corpus. We have combined the term mining method with existing word vector training methods to learn term embeddings to capture the semantic similarity between terms. The proposed method can identify term variations and improve term representations. It is to be noted that the proposed method can be applied across domains without the need for external resources. We also analyzed the distribution of diseases and treatments based on the learned term embeddings, which can be used to explore treatment schemes for some challenging diseases. Extensive computer simulations were conducted to determine the effectiveness of the proposed method. PubMed—10K and PubMed—100K datasets were used for experimentation (see Table 1). A comprehensive evaluation was carried out through visualization of term embeddings, which used cosine similarity to determine similar terms to each term. The visualization helped to demonstrate the performance of the proposed method and serve as a way to explore treatments for novel diseases. The application of the proposed model can be applicable in several domains [31–54] for the future work.

Acknowledgements Prayag Tiwari received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 721321. M. Shamim Hossain extends his appreciation to the Researchers Supporting Project number (RSP-2020/32), King Saud University, Riyadh, Saudi Arabia for funding this work.

Code Availability <https://github.com/prayagtiwari/TermInformer>.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Li J, Hu R, Liu X, Pandey HM, Chen W, Wang B, Jin Y, Yang K (2019) A distant supervision method based on paradigmatic relations for learning word embeddings. *Neural Comput Appl* 2019:1–10
- Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) pp 1532–1543
- Settles B (2004) Biomedical named entity recognition using conditional random fields and rich feature sets. In: Proceedings of the international joint workshop on natural language processing in biomedicine and its applications (NLPBA/BioNLP), pp 107–110
- Wallach HM (2004) Conditional random fields: an introduction. Technical Reports (CIS), p 22
- Leaman R, Gonzalez G (2008) BANNER: an executable survey of advances in biomedical named entity recognition. In: Bio-computing 2008, pp 652–663
- Habibi M, Weber L, Neves M, Wiegandt DL, Leser U (2017) Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics* 33(14):i37–i48
- Tang B, Cao H, Wang X, Chen Q, Xu H (2014) Evaluating word representation features in biomedical named entity recognition tasks. *BioMed Res Int* 2014:240403. <https://doi.org/10.1155/2014/240403>
- Yao L, Liu H, Liu Y, Li X, Anwar MW (2015) Biomedical named entity recognition based on deep neural network. *Int J Hybrid Inf Technol* 8(8):279–288
- Wang X, Zhang Y, Ren X, Zhang Y, Zitnik M, Shang J, Langlotz C, Han J (2019) Cross-type biomedical named entity recognition with deep multi-task learning. *Bioinformatics* 35(10):1745–1752
- Wang X, Lyu J, Dong L, Xu K (2019) Multitask learning for biomedical named entity recognition with cross-sharing structure. *BMC Bioinform* 20(1):427
- Yoon W, So CH, Lee J, Kang J (2019) Collabonet: collaboration of deep neural networks for biomedical named entity recognition. *BMC Bioinform* 20(10):249
- Cho H, Lee H (2019) Biomedical named entity recognition using deep neural networks with contextual information. *BMC Bioinform* 20(1):735
- Lafferty JD, McCallum A, Pereira FC (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the eighteenth international conference on machine learning, pp. 282–289
- Nadeau D, Sekine S (2007) A survey of named entity recognition and classification. *Lingvist Investig* 30(1):3–26
- Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P (2011) Natural language processing (almost) from scratch. *J Mach Learn Res* 12:2493–2537
- Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C (2016) Neural architectures for named entity recognition. arXiv preprint [arXiv:1603.01360](https://arxiv.org/abs/1603.01360)
- Chiu JP, Nichols E (2016) Named entity recognition with bidirectional LSTM-CNNs. *Trans Assoc Comput Linguist* 4:357–370
- Ma X, Hovy E (2016) End-to-end sequence labeling via bi-directional lstm-cnns-crf. arXiv preprint [arXiv:1603.01354](https://arxiv.org/abs/1603.01354)
- Akbik A, Blythe D, Vollgraf R (2018) Contextual string embeddings for sequence labeling. In: Proceedings of the 27th international conference on computational linguistics, pp 1638–1649
- Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. arXiv preprint [arXiv:1802.05365](https://arxiv.org/abs/1802.05365)
- Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
- Beltagy I, Lo K, Cohan A (2019) SciBERT: a pretrained language model for scientific text. arXiv preprint [arXiv:1903.10676](https://arxiv.org/abs/1903.10676)
- Alsentzer E, Murphy JR, Boag W, Weng WH, Jin D, Naumann T, McDermott M (2019) Publicly available clinical BERT embeddings. arXiv preprint [arXiv:1904.03323](https://arxiv.org/abs/1904.03323)
- Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J (2020) BioBERT: a pre-trained biomedical language representation

- model for biomedical text mining. *Bioinformatics* 36(4):1234–1240
25. Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. *J Mach Learn Res* 3:1137–1155
 26. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*
 27. Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. *Trans Assoc Comput Linguist* 5:135–146
 28. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
 29. Forney GD (1973) The viterbi algorithm. *Proc IEEE* 61(3):268–278
 30. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemom Intell Lab Syst* 2(1–3):37–52
 31. Sodhro AH, Pirbhulal S, Sangaiah AK (2018) Convergence of IoT and product lifecycle management in medical health care. *Future Gen Comput Syst* 86:380–391
 32. Sodhro AH, Luo Z, Sangaiah AK, Baik SW (2019) Mobile edge computing based QoS optimization in medical healthcare applications. *Int J Inf Manag* 45:308–318
 33. Sodhro AH, Pirbhulal S, Qaraq M, Lohano S, Sodhro GH, Junejo NUR, Luo Z (2018) Power control algorithms for media transmission in remote healthcare systems. *IEEE Access* 6:42384–42393
 34. Sodhro AH, Malokani AS, Sodhro GH, Muzammal M, Zongwei L (2020) An adaptive QoS computation for medical data processing in intelligent healthcare applications. *Neural Comput Appl* 32(3):723–734
 35. Tiwari P, Qian J, Li Q, Wang B, Gupta D, Khanna A, Rodrigues JJ, de Albuquerque VHC (2018) Detection of subtype blood cells using deep learning. *Cogn Syst Res* 52:1036–1044
 36. Qian J, Tiwari P, Gochhayat SP, Pandey HM (2020) A noble double dictionary based ECG compression technique for IoTH. *IEEE Internet Things J*. <https://doi.org/10.1109/JIOT.2020.2974678>
 37. Jaiswal AK, Kumar S, Gupta D, Khanna A, Rodrigues JJ (2019) Identifying pneumonia in chest X-rays: a deep learning approach. *Measurement* 145:511–518
 38. Rodrigues MB, Da Nóbrega RVM, Alves SSA, Rebouças Filho PP, Duarte JBF, Sangaiah AK, De Albuquerque VHC (2018) Health of things algorithms for malignancy level classification of lung nodules. *IEEE Access* 6:18592–18601
 39. Piccialli F, Casolla G, Cuomo S, Giampaolo F, Di Cola VS (2019) Decision making in IoT environment through unsupervised learning. *IEEE Intell Syst* 35(1):27–35
 40. Casolla G, Cuomo S, Di Cola VS, Piccialli F (2019) Exploring unsupervised learning techniques for the Internet of Things. *IEEE Trans Ind Inform* 16(4):2621–2628
 41. Piccialli F, Cuomo S, di Cola VS, Casolla G (2019) A machine learning approach for IoT cultural data. *J Ambient Intell Hum Comput* 2019:1–12
 42. Ahmad M, Jabbar S, Ahmad A, Piccialli F, Jeon G (2020) A sustainable solution to support data security in high bandwidth healthcare remote locations by using TCP CUBIC mechanism. *IEEE Trans Sustain Comput* 5(2):249–259. <https://doi.org/10.1109/TSUSC.2018.2841998>
 43. Wang J, Han K, Alexandridis A, Chen Z, Zilic Z, Pang Y, Jeon G, Piccialli F (2020) A blockchain-based eHealthcare system inter-operating with WBANs. *Future Gen Comput Syst* 110:675–685
 44. Qureshi KN, Din S, Jeon G, Piccialli F (2020) An accurate and dynamic predictive model for a smart M-Health system using machine learning. *Inf Sci* 538:486–502
 45. Tiwari P, Melucci M (2019) Towards a quantum-inspired binary classifier. *IEEE Access* 7:42354–42372
 46. Wang D, Tiwari P, Garg S, Zhu H, Bruza P (2020) Structural block driven enhanced convolutional neural representation for relation extraction. *Appl Soft Comput* 86:105913
 47. Tiwari P, Melucci M (2018) Towards a quantum-inspired framework for binary classification. In: *Proceedings of the 27th ACM international conference on information and knowledge management*, pp 1815–1818
 48. Tiwari P, Melucci M (2019) Binary classifier inspired by quantum theory. In: *Proceedings of the AAAI conference on artificial intelligence*, vol 33, pp 10051–10052
 49. Tiwari P, Melucci M (2018) Multi-class classification model inspired by quantum detection theory. *arXiv preprint arXiv:1810.04491*
 50. Aujla GS, Chaudhary R, Kaur K, Garg S, Kumar N, Ranjan R (2018) SAFE: SDN-assisted framework for edge-cloud interplay in secure healthcare ecosystem. *IEEE Trans Ind Inform* 15(1):469–480
 51. Rathee G, Garg S, Kaddoum G, Choi BJ (2020) A decision-making model for securing IoT devices in smart industries. *IEEE Trans Ind Inform*. <https://doi.org/10.1109/TII.2020.3005252>
 52. Garg S, Kaur K, Kumar N, Rodrigues JJ (2019) Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: a social multimedia perspective. *IEEE Trans Multimed* 21(3):566–578
 53. Garg S, Kaur K, Kumar N, Kaddoum G, Zomaya AY, Ranjan R (2019) A hybrid deep learning-based model for anomaly detection in cloud datacenter networks. *IEEE Trans Netw Serv Manag* 16(3):924–935
 54. Fang Q, Sang J, Xu C, Hossain MS (2015) Relational user attribute inference in social media. *IEEE Trans Multimed* 17(7):1031–1044

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.