

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,000

Open access books available

148,000

International authors and editors

185M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



A Method for Plotting Disease Drug Analysis and Its Complications by Combining Sources of Scientific Documents Using Deep Learning Method with Drug Repurposing: Case Study Metformin

Zahra Rezaei and Behnaz Eslami

Abstract

Drugs for medical purposes aim at saving one's life and improving their life quality. Side effects or adverse drug reactions (ADRs) on patients are studied as an important issue in pharmacology. In order to prevent the adverse drug effects, clinical trials are conducted on the drug production process, but the process of these trials is very costly and time consuming. So, various text mining methods are used to identify ADRs on scientific documents and articles. Using existing articles in the reference websites such as PubMed to predict an effective drug in the disease is a vital way to declare the drug effective. However, the effective integration of biomedical literature and biological drug network information is one of the major challenges in diagnosing a new drug. In this study, we use medical text documents to train the BioBERT model so that we can use it to discover potential drugs for treating diseases. Then, we are able to create a graphical network of drugs and their side effects with this method as well as it provides us with an opportunity to identify effective drugs that have been used in many diseases so far while having the ability to be used effectively on other diseases.

Keywords: adverse drug reactions, drug repurposing, deep learning, natural language processing, social network

1. Introduction

What makes reusing old drugs worthwhile is the cost of developing a new drug, according to research [1]. The cost of developing a new drug reaches billions of dollars, which includes analysis, testing, validation costs, and so on. More

importantly, duration of developing new drugs may take long nearly 9 to 12 years to launch a new drug.

This practice, therefore, is deemed to be of importance in the pharmaceutical industry because it accelerates the development of drugs and reduces the cost of drug production, especially for pandemic diseases such as COVID-19, and the need to use this scientific process in the field of artificial intelligence is essential.

Due to the rapid growth of scientific articles in medical research, the analysis of medical textual documents using text mining methods has become very popular. The emergence of powerful deep learning methods and their maturity in text mining has created various development ways for different types of text analysis. The only drawback of deep learning models is their training using a large number of input data, which has made an unsurmountable challenge in medical topics. Fortunately, various medical sites such as PubMed allow the use of textual data and have seriously contributed to the development of deep learning models.

Improvements in healthcare and nutrition have generated remarkable increases in life expectancy worldwide. Although our understanding of the molecular basis of these morbidities has quickly advanced, effective novel treatments are still lacking. Today, the topic of reusing drugs based on text mining methods and based on valid scientific articles is important and vital, because based on the characteristics of pharmacokinetics and pharmacodynamics, the process of data generation has already been approved and validated by scientific communities and the study of side effects, and their impact on other diseases will significantly save the time and cost of the data generation process. Creating new drug profiles based on previously valid drugs is a way of bypassing the drug production cycle.

Metformin is one such drug currently being investigated for novel applications.

What is clear from the clinical evidence is that metformin is prescribed in the treatment of diabetes. The aim of this research is to investigate the effects of metformin on various diseases that are reflected in PubMed documents. What will be studied in this report are the results of the use of metformin in the prevention of various diseases.

This chapter aims to provide the reported results, available in medical literatures for potential of metformin to prevent or treat different kinds of disorders.

Furthermore, some of the previous researches in the field of drug reuse have been reviewed in the second chapter. In the third chapter, the proposed research model is discussed and in the fourth chapter, the explanation of the model architecture is discussed. The implementation results and final outputs of the proposed method are explained in the last section.

2. Related works

Drug reuse is used to treat diseases other than an approved disease (such as drug use in new drugs, development of indications, or change of indications), including the development of new medical applications for previously approved drugs, as well as the evolutionary cycle. A drug is defined for the use and development of drugs that are in the drug archive. This strategy is not very new, but it has gained significant momentum in the last decade as approved scientific sources on drug reuse have identified side effects.

About one-third of the approvals in recent years correspond to drug repurposing, and repurposed drugs currently generate around 25% of the annual revenue for the pharmaceutical industry [2].

Drug reuse involves identifying new uses for existing drugs. Prominent examples of the use of these methods include sildenafil and thalidomide as a result of serendipity [3].

Graphs of drugs, genes, and diseases are created and clustering methods are developed to predict new edges between drugs and diseases [4].

Disease genes and drug genes are modeled. Relationships from Medical Scientific Documents and Induction of Indirect Relationships Between Drugs and Diseases Proposed a ranking method based on the similarity of the drug target to rank these relationships [5].

When predicting a new drug-target interactions (DTI), drug-drug interaction (DDI) [6], there are three levels of prediction using machine learning techniques. First, it preprocesses input data such as drug side effects, drug chemical structure, and disease genes and provides training data through feature extraction. The appropriate machine learning algorithm is then used for training. Third, we apply a predictive model to get the results of drug repositioning in the test dataset. The data is transformed into a consistent, normalized format, such as computer-readable vectors and matrices, before being entered into the machine learning model to train the representation. Representation learning [7] (or feature learning) is a set of techniques for transforming raw data into something that can be effectively used through machine learning. Representation learning is mainly divided into a supervised learning approach and an unsupervised learning approach and extracts the properties of the input data of the downstream.

3. Material and methods

We look for relevant publication in PubMed through using metformin as key word. The data of this research are documents and scientific articles written in English between 1994 and 2020. In this direction, we applied named entity recognition (NER) BioBERT method.

The used NER method includes three main phases (**Figure 1**); it is started with textual documents from PubMed which are entered as input data, followed by preprocessing phase to improve data, and eventually, in the third phase, grouping data into train and test categories is done, and NER via deep learning algorithm – BioBERT method – runs to extract patterns.

3.1 Data sources

We looked at 18000 publications in PubMed using metformin as a keyword. The abstracts of 16,000 out of them were analyzed by NER BioBERT. This search covered studies which have been done between 1994 and 2020.



Figure 1.
The workflow of the proposed model-based strategy.

3.2 Preprocessing

The preprocessing of comments in both datasets was performed as follows:

1. Data shuffling
2. Converting all uppercase words into lowercase
3. Elimination of special characters such as @,!, /, *, \$.
4. Remove stop words such as at, of, the.
5. Convert acronym or abbreviation to complete
6. Lemmatization

3.3 Deep classification

Bidirectional Encoder Representations from Transformers for Biomedical Text Mining can be considered a particular language pretrained model on a large-scale biomedical corpus. According to the mentioned architecture, the knowledge from a large number of biomedical documents by BioBERT [8] is transferred to biomedical text mining models with the least amount of modification in the architecture. Whereas competitive performances with previous novel models appeared by BERT and BIOBERT essentially have better performance on the following three biomedical text mining functions: biomedical named entity recognition and biomedical clustering based on the effect of Metformin.

Different diseases and the trend of metformin impact in publication during several years, based on drug effect on various diseases.

BioBERT effectively moved the data from a part of biomedical textual documents to biomedical text mining models by some alterations in a particular structure. Whereas BERT had outlined excellent function with previous models, BioBERT discernibly overwhelmed them on entity recognition and clustering concerning metformin effect on individuals' wellbeing.

We investigated publications based on the association between metformin and type 1 and 2 diabetes. And separately, we explored them in accordance with the effect of metformin on other disease.

PubTator [9] and BEST [10] are two of the potential sources that automatically can extract compounds and proteins from PubMed or PubMed Central (PMC). However, these two sources are not able to extract the combined and interactive relationships between the drug and the disease. To address these issues, we began building a pipeline using NER to identify studies containing DTI and extract related data. We, then, trained the BioBERT model on known studies containing DTIs and used this model to predict new drug studies.

Indeed, given an input sentence $X = \{x_1, x_2, \dots, x_N\}$ where x_i is the i -th word/token and N speaks to the length of the sentence. The objective of NER is to categorize each word/token in X and allot it to corresponding name $y \in Y$, where Y may be a pre-defined list of all conceivable name sorts (e.g. CHEMICAL as Drugs, Infection).

Additionally, this structure was used after preprocessing to identify the relationship between the drug and the disease. In future research, we are going to create this

graph of relationships and use the number of drug references to a chemical structure as a weight to discover drug relationships.

4. Result

There have been a few detailed examinations into the relationship between metformin and the results of cures in different diseases. Moreover, these preclinical reports and dependable biological pathways have been known which clarify the atomic component of metformin and addressed in our research work. Nevertheless, the vital reply to this issue is the level of metformin adequacy against nondiabetic disarranges.

Metformin is the generic name of the drug which is produced and supplied under different brand names such as Metformex, Glucophage, and so on. As shown in **Figure 2**, the drugs extracted from the authoritative scientific articles are in the drug groups related to diabetes and some other drug groups. There are several classes of drugs used to control diabetes, and members belonging to each group have similar

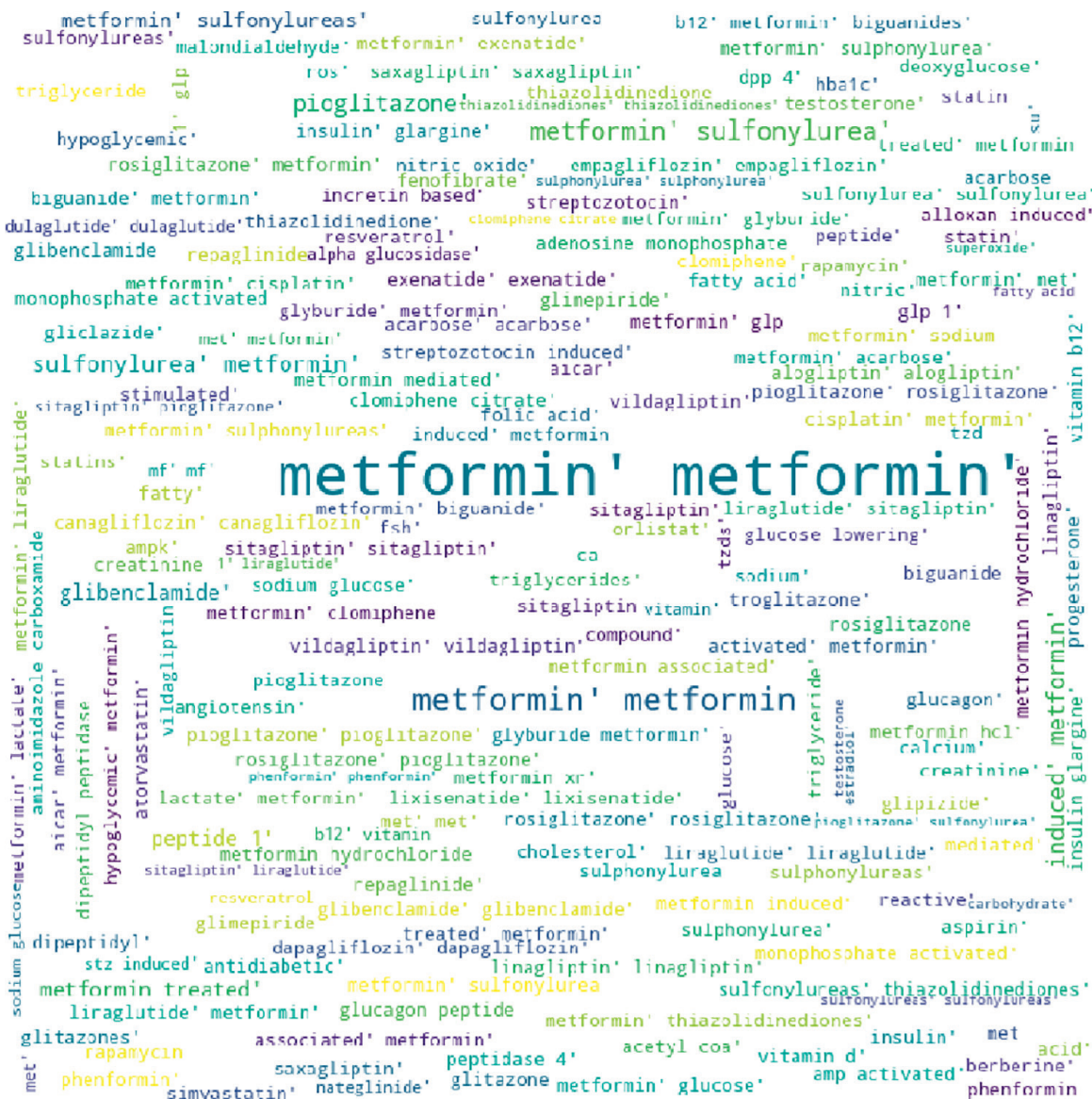


Figure 2.
The word-cloud of the BioBERT model in drugs.

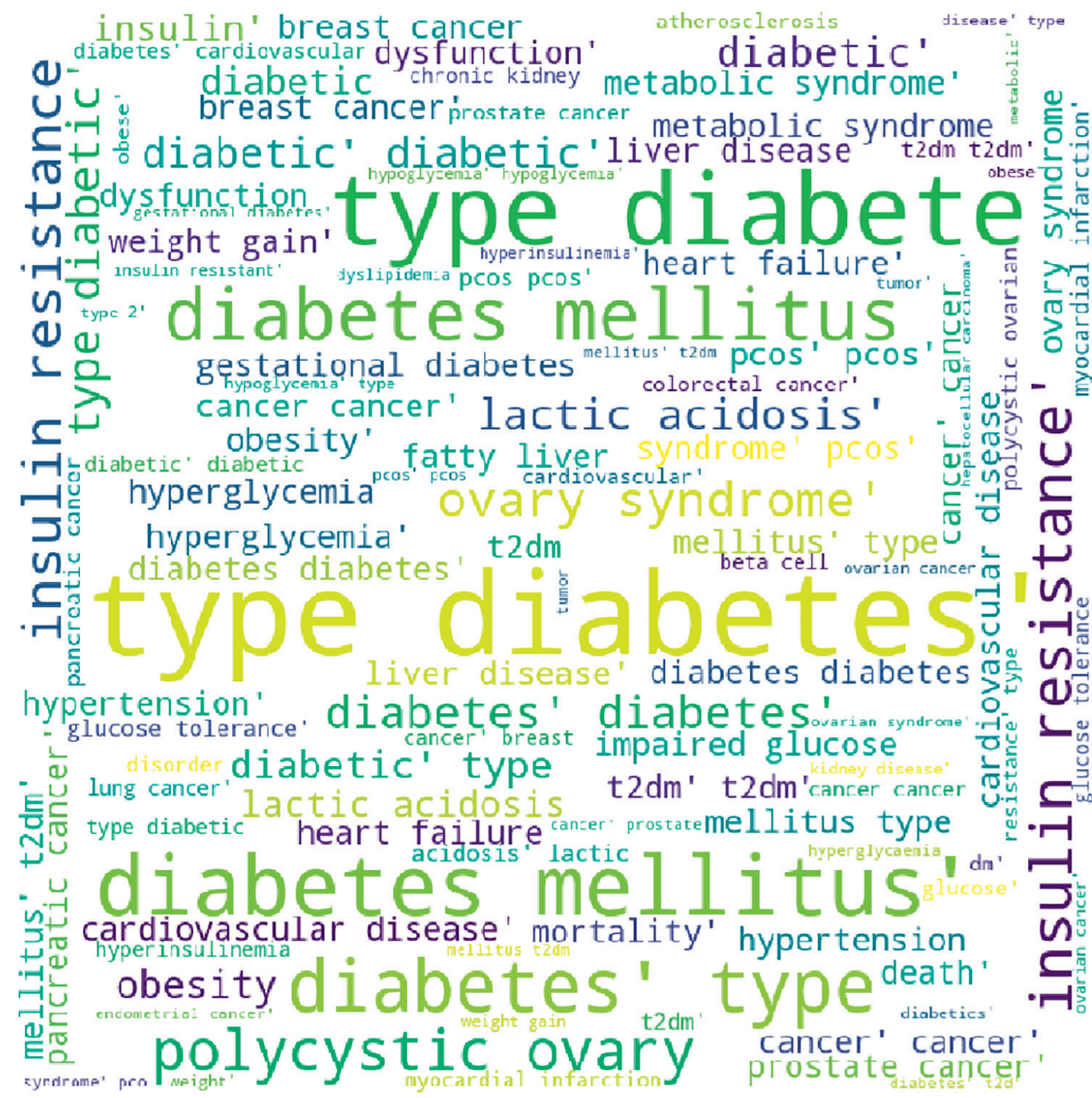


Figure 3.
The word-cloud of the BioBERT model in disease.

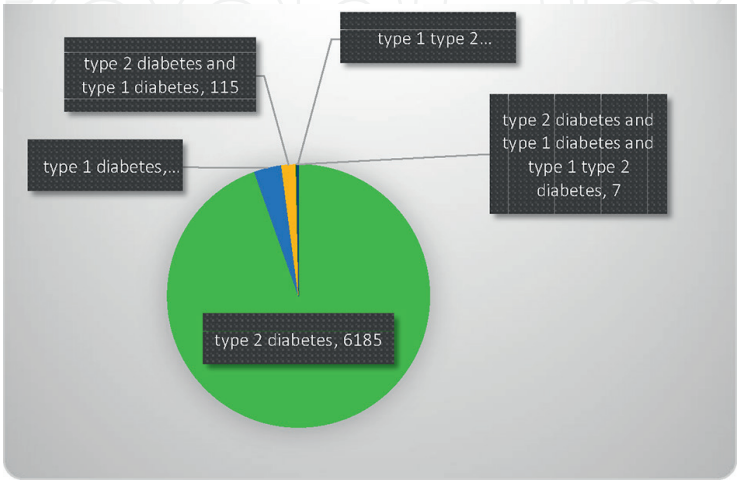


Figure 4.
The word-cloud of the BioBERT model in disease.

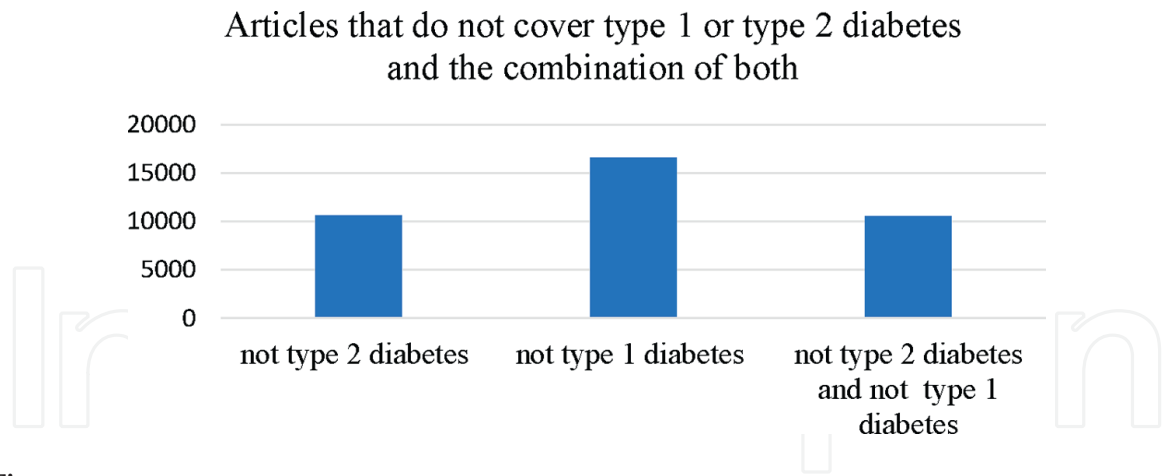


Figure 5.
The word-cloud of the BioBert model in disease.

functions. One of these drug classes is biguanides. Metformin, the only member of this drug group, works in three ways:

- Decreased amount of glucose produced in the liver
- Decreased absorption: the amount of glucose that the body absorbs
- Increased effect of insulin on the body

Diabetes medications are generally prescribed to lower blood glucose. For example, in articles, it refers to synthetic alternatives and antidiabetic drugs to reduce perfusion or kidney function, exacerbate the antihypertensive effects, exacerbate metabolic acidosis, and so on (Figures 3–5).

According to the NERBIOBERT model, out of 16,781 articles reviewed by the PubMed site and analyzed in the article, 6185 papers refer to type 2 diabetes and 221 papers refer to type 1 diabetes as we know it. Type 2 diabetes is a chronic disease. It is characterized by high levels of sugar in the blood. Type 2 diabetes is also called type 2 diabetes mellitus and adult-onset diabetes. Although, 2388 papers used metformin in type 2 diabetes mellitus, and 1178 articles did not mention any disease at all. Therefore, based on the type of articles, if the adverse use of metformin for the treatment of another disease has been identified, it can be used in the treatment of that disease. What is important in this analysis is a demonstration of the disease and the drug so that by analyzing a large volume of authoritative articles, the use of the approved drug can be used in the treatment of other diseases.

5. Conclusion

Experimental results on the drugs and disease with using advanced deep learning models like Bert show that integrating pretrained biomedical language representation models (i.e. BERT and BioBERT) into a pipe of information extraction methods with multitask learning can improve the ability to collect drug repurposing knowledge from PubMed.

Hitherto, there has not been any clear answer for that in clinical trial, and also, the role of metformin on treatment or prevention of disease remains hypothetical on next

step, and we will extract the association between diabetes and other relevant disease with respect to administration of metformin as treatment.

Acknowledgements

The authors have no proprietary, financial, professional, or other personal interest of any nature in any product, service, or company. There is no conflict of interest in this study.

Author details


Zahra Rezaei^{1*} and Behnaz Eslami²

1 Department of Statistics and Information Technology, Institute of Judiciary, Tehran, Iran

2 Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

*Address all correspondence to: z.rezaei2010@gmail.com

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Dickson M, Gagnon JP. The cost of new drug discovery and development. *Discovery Medicine*. 2009;**22**(4):172-179
- [2] Naylor S, Kauppi MJ, Schonfeld JM. Therapeutic drug repurposing, repositioning and rescue part II: Business review. *Drug Discovery World*. 2015;**16**:57-72
- [3] Liu Z, Fang H, Reagan K, Xu X, Mendrick DL, William Slikker WT Jr. In silico drug repositioning: What we need to know. *Drug Discovery Today*. 2013;**18**:110-115
- [4] Sun P, Guo J, Winnenburg R, Baumbach J. Drug repurposing by integrated literature mining and drug-gene-disease triangulation. *Drug Discovery Today*. 2017;**22**:615-619
- [5] Yang H-T, Ju J-H, Wong Y-T, Shmulevich I, Chiang J-H. Literature-based discovery of new candidates for drug repurposing. *Briefings in Bioinformatics*. 2017;**18**:488-497
- [6] Zhu S, Bai Q, Li L, Xu T. Drug repositioning in drug discovery of T2DM and repositioning potential of antidiabetic agents. *Computational and Structural Biotechnology Journal*. 2022;**20**:2839-2847
- [7] Bengio ACPY. Vincent representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013;**35**(8):1798-1828
- [8] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;**36**(4):1234-1240
- [9] Wei C-H, Kao H-Y, Lu Z. PubTator. A web-based text mining tool for assisting biocuration. *Nucleic Acids Research*. 2013;**41**:W518-W522
- [10] Lee S, Kim D, Lee K, Choi J, Kim S, Jeon M, et al. BEST: Next-generation biomedical entity search tool for knowledge discovery from biomedical literature. *PLoS One*. 2016;**11**(10):e0164680