

Big Data in creating a clinical decision support system

Gaysler E.^{1*}, Demenkov P.^{1,2}, Ivanisenko V.^{1,2}, Ivanisenko T.²

¹ Zelman Institute of Medicine and Psychology, Novosibirsk State University, Novosibirsk, Russia

² Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* Evgeniy.gaysler@mail.ru

Key words: CDSS, association of symptom with disease, ternary logic

Motivation and Aim: CDSS (Clinical decision support system) – “... Link the results of clinical decisions with data available for a particular patient, influencing the choice of a medical decision for more effective medical care ...” (Robert Hayward, Center for Health Evidence). The relevance of the creation of CDSS is due to the general trend in the development of digital medicine, aimed at reducing the number of medical errors and optimizing the work of medical institutions. It is believed that the development and implementation of CDSS belong to the most important directions in the development of artificial intelligence in medicine [1]. There are two main varieties of CDSS - based on “scientific knowledge” and “non-knowledgeable”, based on the results of processing the collected statistical data, most often in the form of a neural network. There is an approach to the formation of reasoning by precedents (retrospective data on individual clinical cases are used) [2].

Methods and Algorithms: A hybrid approach is proposed: on the one hand, the knowledge contained in the largest library of peer-reviewed medical journals PubMed is used (28 million articles were used), on the other hand, these data (in our case, the number of mutual mentions of a particular symptom with various diseases) are statistically processed to build an information base (712 symptoms and 584 diseases, syndromes, and disease states). “Knowledge” CDSS use inference mechanisms that take the form of IF-THEN rules. In medical practice, situations often arise (the symptom is not manifested) that do not fit into the usual Boolean logic, so the ternary logic (YES; UNKNOWN; NO) is used in the inference mechanism. The directed selection algorithm is implemented as a question-answer system, and the results of the answers are used to assess the degree of proximity of diseases to the specified set of symptoms (their presence or absence). The solutions at the top of the list of diseases are recombined and the next question is found based on the proximity score. The process can be stopped at any time. The result is a list of diseases with an indication of the likelihood of matching the indicated symptoms. **Results:** Testing in a medical institution showed that the probability of a match between the diagnosis established by experienced doctors and the disease from the first 5 noted by the system is more than 80 %.

Conclusion: The chosen approach to the creation of CDSS has shown its potential effectiveness. Extended testing of the system by employees of the Zelman Institute of Medicine and Psychology NSU will clarify the operation of the inference mechanism. Further development of the system involves the creation of a base of precedents (based on the case histories of many patients), which will allow us to confirm or refute the hypotheses of choosing a diagnosis based on the work of the CDSS.

References

1. Moja L. et al. *Am J Public Health*. 2014;104(12):e12-22.
2. Tanveer Syeda-Mahmood plenary talk: The role of machine learning in clinical decision support. *SPIE Newsroom*. 2015.