



# A Data-Intensive CDSS Platform Based on Knowledge Graph

Ming Sheng<sup>1</sup>(✉), Qingcheng Hu<sup>1</sup>, Yong Zhang<sup>1</sup>, Chunxiao Xing<sup>1</sup>,  
and Tingting Zhang<sup>2</sup>

<sup>1</sup> RIIT&BNRCIST&DCST, Tsinghua University, Beijing 100084, China  
msheng24@126.com,

{huqingcheng, zhangyong05, xingcx}@tsinghua.edu.cn

<sup>2</sup> Yellow River Engineering Consulting Co., Ltd., Zhengzhou 450000, China  
zhangtt@yrec.cn

**Abstract.** Clinical Decision Support Systems (CDSSs) are very important for doctors and hospitals to improve the medical service quality. There are two types of CDSSs – knowledge-based CDSSs and data-intensive CDSSs. This paper presents a framework based on knowledge graph to integrate the two methods and proposes a data-intensive clinical decision support platform. This platform provides a series of clinical decision support services, including inquiry, inspection, diagnosis, medication & treatment and prognosis. This platform has been used in a system for village doctors.

**Keywords:** CDSS · Knowledge graph · Medical knowledge base  
Data-intensive · Smart health

## 1 Introduction

Clinical Decision Support Systems (CDSSs) in Health Care have a long history going back to the 1970s. A CDSS is the computer software designed to contribute to clinical treatments and diagnoses [1]. More precisely, Clinical Decision Support (CDS) can be defined as “a process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health and healthcare delivery”, and a CDSS is an implementation of one or more CDS interventions [2]. CDSSs can support many different activities such as diagnosis, therapy, monitoring, or prevention and are used in all kinds of medical domains such as chronic illness, acute care, primary care, and patient advice lines. CDSSs may provide many different services such as access to knowledge, statistical calculations and individual adaptations, recommendations, reminders or alerts to different user groups.

There are two ways to facilitate the CDSSs: the knowledge-based systems and the data-intensive systems. The medical knowledge based systems perform a variety of functions, including knowledge acquisition, knowledge translation, logical reasoning based on evidence-based medical literature and the approved clinical practice guidelines. While the data intensive systems utilize the data from patients’ Electronic Medical Records (EMR). Compared to medical knowledge based way, data-intensive way yields a significant performance overhead.

Through collaborating with hospitals, we have already put many efforts to apply big data and machine learning technologies to deal with health-related data. In [3], a deep learning based information extraction framework is designed to extract medical entities and relations from Chinese EMR. In [4], for the problem of patient's cost profile estimation, a model based on a patient's historical visits in EMR is presented. In [5], several deep learning algorithms are proposed to classify the topics of documents in diabetics. In [6], we investigate the influences of different language models on the sentiment classifications for Chinese health forums. In [7], a QA prototype is developed to answer health related questions. Based on these efforts, we decided to implement a system to integrate them to help doctors and hospitals.

This paper provides a unified framework based on knowledge graph to integrate knowledge-based and data-intensive methods, and designs a data-intensive clinical decision support platform - IDS. It follows the process of a patient's visit. This platform provides a series of clinical decision support services including inquiry, inspection, diagnosis, medication & treatment and prognosis.

This paper is organized as follows. Section 2 introduces the related work. Section 3 provides the integration framework. Section 4 presents the architecture of IDS. Section 5 shows the process of a patient visit in IDS. Section 6 demonstrates the application of IDS. Finally we conclude the paper in Sect. 7.

## 2 Related Work

In this section, we will discuss the related work from three aspects: knowledge-based CDSSs, data-intensive CDSSs and the application of CDSSs.

**Knowledge-Based CDSSs:** Shen et al. [8] described the construction and optimization of the sensitivity and specificity of a decision support system named IDDAP, which is based on ontologies for infectious disease diagnosis and antibiotic therapy. Nazari et al. [9] utilized expert systems and fuzzy logic, and developed an intelligent system, which is capable of diagnosing occurrence of heart diseases. Ohno-Machado [10] considered the update of medical knowledge base and the contextual information.

**Data-Intensive CDSSs:** In [11], a Semantic Web-based, multi-strategy reasoning approach is presented and integrates deductive and plausible reasoning and exploits Semantic Web technology to solve complex clinical decision support queries. Beaudoin et al. [12] evaluated the machining learning for CDSS to enhance antimicrobial stewardship programs. Piri et al. [13] analyzed data from more than 1.4 million diabetics and developed a CDSS for predicting DR.

**Application of CDSSs:** Greenes et al. [14] discussed how to integrate workflow-oriented models with a user-interactive mode of CDSS. Huberts et al. [15] introduced an innovative structured modeling approach in which model personalization is guided by sensitivity analysis and in which the effect of input uncertainties and model assumptions are considered during model corroboration. Humm et al. [16] described a personalised CDSS for cancer care. Anya et al. [17] presented the design of an awareness environment for cross-boundary clinical decision support in e-health that

takes account of the concept of work practice as a design requirement. Bennett et al. [18] provided a substantive review of international literature evaluating the impact of CDSSs on the care of emergency department (ED) patients. Wajid et al. [19] discussed some of the current challenges in designing an efficient CDSS as well as some of the latest techniques that have been proposed to meet these challenges.

Our method combines these two strategies - knowledge-based method and data-intensive method, and will be discussed detailedly in the next section

### 3 Integration Framework

Current CDSSs use either the knowledge-based method or the data-intensive method. There are few works to consider them together. This paper designs a new integration framework which could combine these two methods together.

The simplest way is to use them separately in one system. For a decision problem, if there are enough features in the input data, the knowledge-based method will be used. Otherwise, the data-intensive method will be used. These two methods can also be used at the same time, thus the final output will be the weighted average of the outputs from them.

In the knowledge-based method, the foundation is the medical knowledge base. While in the data-intensive method, the foundation is the medical sample base. The key to combine these two methods is how to connect the medical knowledge base and the medical sample base.

We propose a unified framework to connect the medical knowledge base and the medical sample base. Base on this framework, the medical knowledge base can be expanded and the medical sample base can be improved as well. So that the decision support efficiency can be improved. Specifically, it includes two aspects:

- The medical sample base supports the medical knowledge base: Based on a large number of samples, the framework can discover new rules and even discover problems within the original rules, so that the medical knowledge base can be constantly expanded and verified.
- The medical knowledge base supports the medical sample base: The relationships between the features of one sample are very important, which can contribute to both the selection of learning models and the adjustment of their parameters. Those relationships between features may be revealed by the medical knowledge base.

In order to support the above two aspects, we use the knowledge graph as a link between the medical sample base and the medical knowledge base. Specifically, we extract the triples from the medical sample base to form a knowledge graph. We perform rule detection in the knowledge graph, abstract those rules into knowledge and add them to the medical knowledge base. Meanwhile, we could apply the medical knowledge base to the knowledge graph, i.e., we can give more semantics to the nodes and edges as well as create more connections between the nodes. Moreover, those semantics and connections can be mapped to the relationships among the features of the data from the medical sample base.

### 4 Platform Architecture

Guided by the integration framework, we developed IDS - a data-intensive clinical decision support platform, whose architecture is shown in Fig. 1. It has eight modules: data integration (①), knowledge fusion (②), medical sample base (③), medical knowledge graph (④), medical knowledge base (⑤), model integration (⑥), rule reasoning (⑦) and IDS services (⑧).

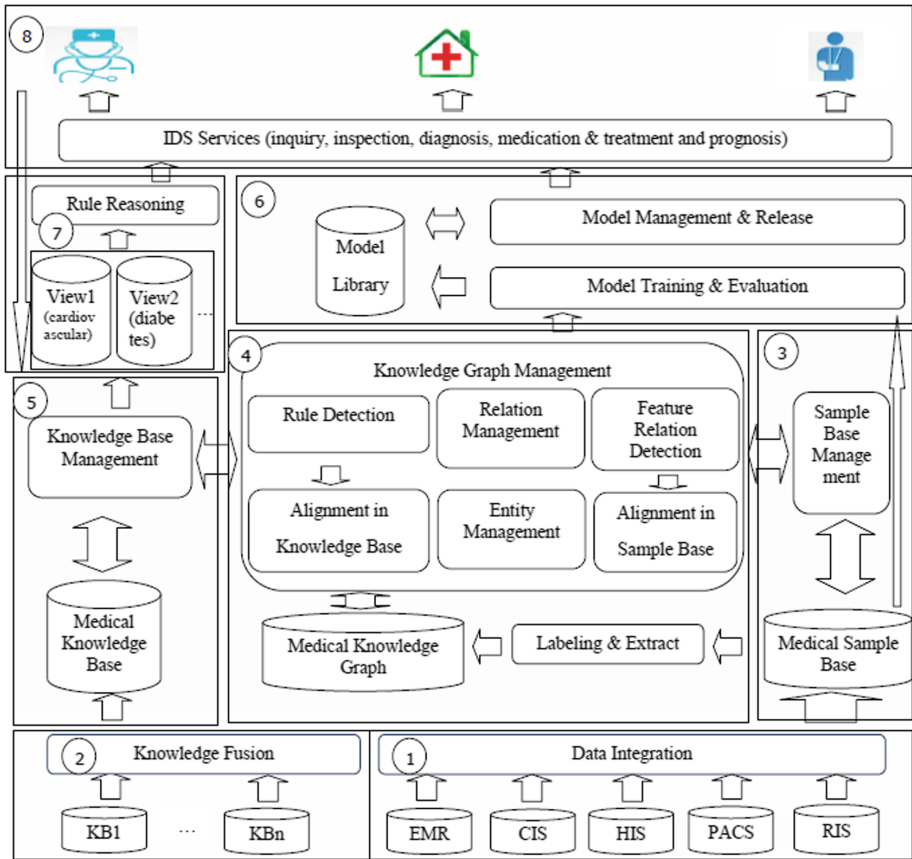


Fig. 1. IDS Architecture

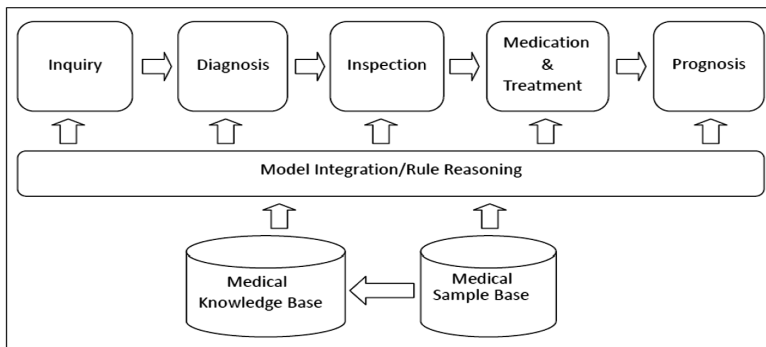
The main functions of IDS are described below:

- Data integration: It includes four parts: the raw data storage, the data standard management, the data cleaning and the data management.
- Knowledge fusion: Based on the open knowledge bases in various fields and the international mainstream standards such as ICD10, this module performs knowledge fusion on various knowledge repositories.

- Medical sample base: It standardizes the records from the data integration and models those records based on the knowledge graph.
- Medical knowledge graph: It includes rule detection, relation management, feature relation detection, alignment in medical knowledge base, entity management, alignment in medical sample base and labeling & extract.
- Medical knowledge base: It includes code management, metathesaurus management, semantic network management, specialist lexicon management, and new concept/relation review.
- Model integration: It includes model training, model management and model release. The whole process includes model selection, model training, model evaluation, parameter tuning and model persistence. All of the models are stored in the model library.
- Rule reasoning: For certain specific area such as cardiovascular and diabetes, it performs a rule reasoning based on customized views from the medical knowledge base.
- IDS Services: It packs the selected models from model management & release module. With the packed models, the platform will provide web services that follow restful standards. It includes chief complaint input formalization, intelligent diagnosis, disease-based intelligent recommendation, similar case matching, re-hospitalization risk prediction.

## 5 The Process of a Patient Visit in IDS

Around the process of a patient visit, this platform provides a series of clinical decision support services - inquiry, inspection, diagnosis, medication & treatment and prognosis. This process is based on the models which are trained by the data from the medical sample base, or the rules in the medical knowledge base (see Fig. 2). Due to page limit, here we only discuss the implementation based on model integration.



**Fig. 2.** The process of IDS

## 5.1 Inquiry

This step includes two functions. The first one is to help doctors regulate their chief complaints and remind doctors of key information that is easy to ignore (see Table 1). The second one is to match the most similar case according to the chief complaint.

**Table 1.** Chief complaint input formalization.

<p><b>Input (Voice input) :</b></p> <p>My name is Zhang San, male, 85 years old. Suffering from continuous chest tightness at about 9:00 pm three days ago and feeling dyspnea and arrhythmia. At the same time I also feel nauseous and sweating, vomiting, and numbness in the upper extremities. There has been no sign of remission.</p>
<p><b>Output :</b></p> <p>[Sex]: ( male )</p> <p>[Age]: ( 85 )</p> <p>[Weight]: ( <b>unknown</b> )</p> <p>[First symptoms]: ( persistent chest tightness ) ( stomach ache ) ( difficulty breathing )</p> <p>[Concomitant symptoms]: ( nausea ) ( vomiting ) ( upper limb numbness )</p> <p>[Onset nature]: ( persistent )</p> <p>[Incentives]: ( <b>unknown</b> )</p> <p>[The longest duration of chest pain]: ( <b>unknown</b> )</p> <p>[Mitigation mode]: ( can't ease )</p> <p>[Past history]: ( <b>unknown</b> )</p> <p><b>The "unknown" part is a reminder to the doctor so that the doctor can further inquire.</b></p>

## 5.2 Diagnosis

- (1) Intelligent diagnosis. The input of this function is the structured chief complaint data, and the output is disease risk ranking, e.g.:
1. [ACS cardiogenic chest pain 80%]:
    - 1.1 (STEMI 60%)
    - 1.2 (Aortic dissection 20%)
  2. [Non-ACS cardiogenic chest pain 15%]:
    - 2.1 (Arrhythmias 10%)
    - 2.2 (Ischemic cardiomyopathy 5%)
    - 2.3 (Coronary heart disease xx%)
  3. [Non-cardiogenic diseases 5%]:
    - 3.1 (Respiratory diseases 3%)
    - 3.2 (Digestive system diseases 1%)
    - 3.3 (Nervous system disease 1%)

- (2) Cardiovascular diseases risks ranking. The input of this function is the dynamic or static image files and the output is the disease risk ranking.
- (3) Similar case matching. The input of this function is the structured chief complaint, inspection result and diagnosis. The output is based on the EMR from hospitals that is stored in the medical sample base, as described above.

### 5.3 Inspection

- (1) Intelligent inspection items recommendation. The input of this function is the disease risk ranking and the output is the suggested inspection e.g.:

[Sign detection]:

(blood oxygen xx%) (blood pressure xxx/xxx mmHg) (ECG)

[Blood test]:

(creatinine xx umol/L) (BNP xx pg/ml ) (PT-INR x) (creatinine protein xx ng/ml)

- (2) Similar case matching. The input of this function includes the structured chief complaint and inspection result. The output is based on the EMR from hospitals that is stored in the medical sample base.

### 5.4 Medication and Treatment

- (1) Intelligent medication & treatment recommendation. The input of this function is the structured chief complaint, diagnostic result and inspection result. The output is the medication and medical treatment advice, e.g.:

Medication advice:

[Ant platelet]: (Aspirin)

[Anti-angina]: (Nitroglycerin)

[Sedative]: (Morphine)

[Heart rate, blood pressure]: (Atropine) (Dopamine)

[Surgical treatment]: (Interventional surgery)

Medical treatment advice:

[Conventional disposal]: (CPR) (Oxygen absorption) (Defibrillation)

[Surgical treatment]: (Interventional surgery)

[Unable to dispose of]: (Enter a high level of specialist treatment)

- (2) Similar case matching. The input of this function contains the structured chief complaint, diagnostic result and inspection result. The output is based on the EMR from hospitals that is stored in the medical sample base, as described above.

### 5.5 Prognosis

After going through the whole process of a patient visit (inquiry, inspection, diagnosis, medication & treatment), this platform provides the prognosis prediction services. The purpose is to assess the effectiveness of treatment from the perspective of health

economics. Specifically this platform uses re-hospitalization risk prediction model to achieve the above purpose. Meanwhile the input of the service is just like the EMR from hospitals which is stored in the medical sample base. The output of the service is the days of re-hospitalization.

## 6 Application of IDS

### 6.1 Motivation

At present, the quality of primary health care in China is at a relatively low level. Many village doctors need professional guidance from experienced doctors who are very busy themselves. CDSS could help village doctors to reduce their misdiagnosis rate. We applied IDS to support the village doctors.

### 6.2 Implementation

The application was developed using Hadoop, Spark, Scale and Java. Figure 3 shows an example of decision support for diagnosis. The input of this interface contains three parts:

1. The upper left part of the figure is the chief complaint input area, such as sex, age and weight.
2. The lower left part of the figure is the current medical history input area which is text input.

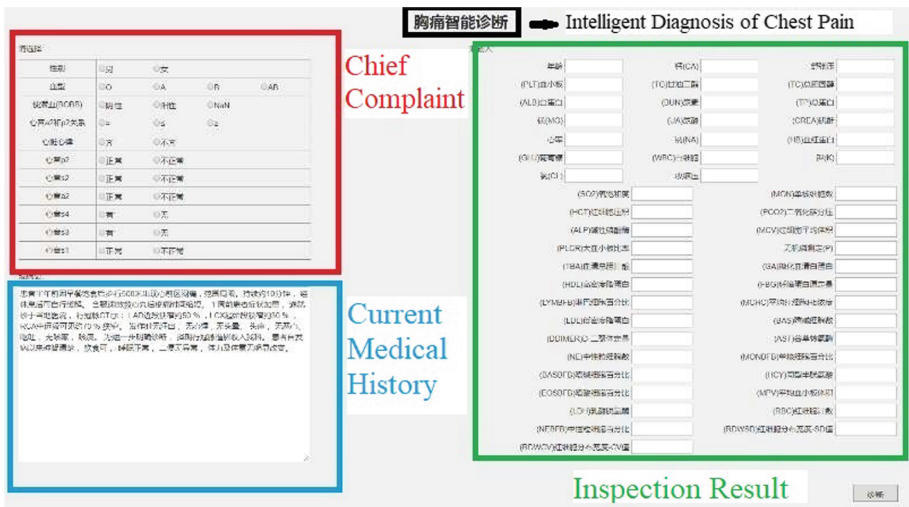


Fig. 3. Example of User Input Interface



3. The right part of the figure is the inspection result input area, such as blood oxygen, blood pressure, creatinine, BNP, PT-INR and creatinine protein.

The output is disease risk ranking (see Fig. 4).

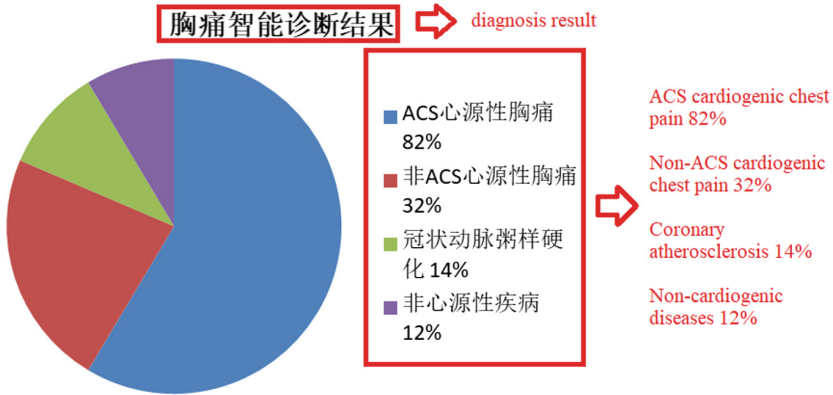


Fig. 4. Example of Result Interface

## 7 Conclusion

This paper proposes IDS, a data-intensive clinical decision support platform based on knowledge graph. This paper designs a new integration framework which connects the medical knowledge base and the medical sample base. Under this framework, the medical knowledge base can be expanded and the medical sample base can be improved as well. Therefore, the decision support efficiency can be improved. Following the process of a patient visit, this platform provides a series of clinical decision support services. The platform has been integrated in an application for village doctors.

## References

1. Berner, E.S.: Clinical Decision Support Systems: Theory and Practice, vol. 233. Springer, New York (2007). <https://doi.org/10.1007/978-0-387-38319-4>
2. Osheroff, A., Teich, J.M., Levick, D.: Improving Outcomes with Clinical Decision Support: An Implementer's Guide, 2nd edn. Healthcare Information and Management Systems Society (HIMSS), Chicago (2012)
3. Tian, B., Zhang, Y., Liu, K.X., Xing, C.X.: Deep learning based information extraction framework on Chinese electronic health records. In: SEKE, pp. 86–91 (2018)
4. Zhao, K., et al.: Modeling patient visit using electronic medical records for cost profile estimation. In: Pei, J., Manolopoulos, Y., Sadiq, S., Li, J. (eds.) DASFAA 2018. LNCS, vol. 10828, pp. 20–36. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-91458-9\\_2](https://doi.org/10.1007/978-3-319-91458-9_2)

5. Chen, X., Zhang, Y., Zhao, K., Hu, Q., Xing, C.: Domain supervised deep learning framework for detecting Chinese diabetes-related topics. In: Pei, J., Manolopoulos, Y., Sadiq, S., Li, J. (eds.) DASFAA 2018. LNCS, vol. 10828, pp. 53–71. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-91458-9\\_4](https://doi.org/10.1007/978-3-319-91458-9_4)
6. Zhang, Y., Zhang, Y., Xu, J., Xing, C., Chen, H.: Sentiment analysis on Chinese health forums: a preliminary study of different language models. In: Zheng, X., Zeng, D.D., Chen, H., Leischow, Scott J. (eds.) ICSH 2015. LNCS, vol. 9545, pp. 68–81. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-29175-8\\_7](https://doi.org/10.1007/978-3-319-29175-8_7)
7. Yin, Y., Zhang, Y., Liu, X., Zhang, Y., Xing, C., Chen, H.: HealthQA: a Chinese QA summary system for smart health. In: Zheng, X., Zeng, D., Chen, H., Zhang, Y., Xing, C., Neill, Daniel B. (eds.) ICSH 2014. LNCS, vol. 8549, pp. 51–62. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-08416-9\\_6](https://doi.org/10.1007/978-3-319-08416-9_6)
8. Shen, Y., Yuan, K.: An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. *Artif. Intell. Med.* **86**, 20–32 (2018)
9. Nazari, S., Fallah, M., Kazemipour, H., Salehipour, A.: A fuzzy inference- fuzzy analytic hierarchy process-based clinical decision support system for diagnosis of heart diseases. *Expert Syst. Appl.* **95**, 261–271 (2018)
10. Ohno-Machado, L.: Clinical decision support: informatics interventions for better patient care. *JAMIA* **25**(5), 457 (2018)
11. Mohammadhassanzadeh, M., Woensel, W.V., Abidi, R.S.: Semantics-based plausible reasoning to extend the knowledge coverage of medical knowledge bases for improved clinical decision support. *BioData Min.* **10**(1), 7:1–7:31 (2017)
12. Beaudoin, M., Kabanza, F., Nault, V.: Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs. *Artif. Intell. Med.* **68**, 29–36 (2016)
13. Piri, S., Delen, D., Liu, T.: A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble. *Decis. Support Syst.* **101**, 12–27 (2017)
14. Greenes, R.A., Bates, D.W., Kawamoto, K., Middleton, B.: Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. *J. Biomed. Inform.* **78**, 134–143 (2018)
15. Huberts, W., Heinen, S.G.H., Zonnebeld, N.: What is needed to make cardiovascular models suitable for clinical decision support? A viewpoint paper. *J. Comput. Sci.* **24**, 68–84 (2018)
16. Humm, B.G., Walsh, P.: Personalised clinical decision support for cancer care. In: Hoppe, T., Humm, B., Reibold, A. (eds.) *Semantic Applications*, pp. 125–143. Springer, Heidelberg (2018). [https://doi.org/10.1007/978-3-662-55433-3\\_10](https://doi.org/10.1007/978-3-662-55433-3_10)
17. Anya, O., Tawfik, H.: An “Awareness” environment for clinical decision support in e-health. In: Rojas, I., Ortuño, F. (eds.) *IWBBO 2018*. LNCS, vol. 10814, pp. 456–467. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-78759-6\\_41](https://doi.org/10.1007/978-3-319-78759-6_41)
18. Bennett, P., Hardiker, N.R.: The use of computerized clinical decision support systems in emergency care: a substantive review of the literature. *JAMIA* **24**(3), 655–668 (2017)
19. Wajid, S.K., Hussain, A., Luo, B., Huang, K.: An investigation of machine learning and neural computation paradigms in the design of clinical decision support systems (CDSSs). In: Liu, C.-L., Hussain, A., Luo, B., Tan, K.C., Zeng, Y., Zhang, Z. (eds.) *BICS 2016*. LNCS (LNAI), vol. 10023, pp. 58–67. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-49685-6\\_6](https://doi.org/10.1007/978-3-319-49685-6_6)