IDDAT: An Ontology-Driven Decision Support System for Infectious Disease Diagnosis and Therapy

Ying Shen*, Deng Yang*, Jin Zhang*, Yaliang Li†, Nan Du†, Wei Fan†, Min Yang‡, and ✉Kai Lei*
* *Shenzhen Key Lab for Information Centric Networking & Blockchain Technology (ICNLAB)*
School of Electronics and Computer Engineering
Peking University Shenzhen Graduate School
Shenzhen, China
Email:{shenying,leik}@pkusz.edu.cn; ydeng@pku.edu.cn; zhangjin@sz.pku.edu.cn
†*Tencent Medical AI Lab*
Palo Alto, USA
Email: {yaliangli,ndu,davidwfan}@tencent.com
‡*SIAT, University of the Chinese Academy of Sciences*
Shenzhen, China
Email:min.yang@siat.ac.cn

***Abstract*—Decision Support Systems (DSS) has become increasingly important due to its broad applications in various domains. Significant progresses have been made on ensuring more precise decision-making by leveraging appropriate data and knowledge from knowledge bases. However, the current DSSs related to antibiotics consider only therapy rather than diagnosis, and they were developed from a physician's perspective. Based on these two points, this study presents IDDAT, an ontology-driven decision support system for aiding Infectious Disease Diagnosis and Antibiotic Therapy. Based on patient-entered information, this freely accessible system aims to identify infectious disease, and provide an antibiotic therapy specifically adapted to the patient. We show the effectiveness of IDDAT by applying it to a diagnosis classification task. Experimental results reveal the system's advantages in term of the area under the curve (AUC) of receiver operating characteristic (ROC) (89.91%).**

***Keywords*-Decision Support System; diagnosis; therapy; ontology; infectious disease**

I. Introduction

Infections are disorders caused by disease-causing agents - such as bacteria, viruses, and fungi arthropods. Infectious disease is illness resulting from an infection and can be classified by the type of organism causing the infection [1] Infectious diseases resulted in more than 10 million deaths in 2015 (about 17% of all deaths) and were among the leading causes of death across all income groups [2].

In the field of medicine, many efforts have been made to design and develop a decision-support system (DSS). Health Evaluation through Logical Processing (HELP) [3]was the first hospital information system developed for antibiotic therapy to support the appropriate antibiotics selection in case of infections. Various DSSs for infectious disease diagnosis that have been designed inside the HELP environment are dedicated to improving antibiotic prescription and treatment for surgical prophylaxis [4] and preventing adverse drug events [5].

Another expert system, TREAT [6], was developed based on a causal probabilistic network to improve antibiotic theapy in hospitalized patients. The antibiotic decision support system (ADSS [7] proposed user-centered design techniques for the infectious disease diagnosis and antibiotic prescription in intensive care units. The DSS [8] developed in University of South Carolina adopted a multi-method intervention to assess the impact of a clinical decision support system, so as to prevent the overuse of antibiotics in primary care.

However, the current DSSs related to antibiotics have two potential limitations. 1) They mainly focus on the study of therapy rather than diagnosis [9]. 2) They were designed from a physician's perspective. Nevertheless, since infectious diseases are common and frequently occurring diseases, people desire to gain access to knowledge via Internet [10].

To alleviate these limitations, this paper presents the structure and usage mode for an integrated diagnosis and therapy Decision Support System for infectious diseases (IDDAT). This freely accessible system aims to identify infectious disease, based on patient-entered information and the ontology. Currently, ontologies are being used to organize biomedical knowledge and data in various studies, such as clinical knowledge management, biomedical knowledge representation and medical decision support [11]. In IDDAT, the ontology is developed from existing ontologies, by combining extra information from medical website. The decision-making process is completed automatically. If any problems arise, patient should consult a doctor or be referred to a hospital.

Based on our previous work [12], this study carries out

the ontology expansion with reliable clinical data resources, and reduces some redundant computations in the inference stage. Our results are reproducible. Our results are reproducible. Source code: https://github.com/shenyingpku/dssmanager. Ontology: www.iasokg.com. Video: https://youtu.be/F4tuqwi4rlE.

II. System Architecture

We design an IDDAT for the infectious disease and antibiotic prescription by automatically collecting behavior clusters and constructing an ontology. The research emphasis of this paper can be summarized as follows: 1) generating an ontology, 2) incorporating a patient's self-inspection to propose a patient-centered infectious diagnosis and therapy system, 3) identifying and classifying possible infectious diseases based on knowledge contained in the ontology, 4) assessing the ontology and IDDAT we proposed.

The proposed IDDAT, depicted in Fig. 1, consists of three components: the ontology model, the clinical decision support model and the user interaction model. In this section, we elaborate the many tasks inside each of these components.

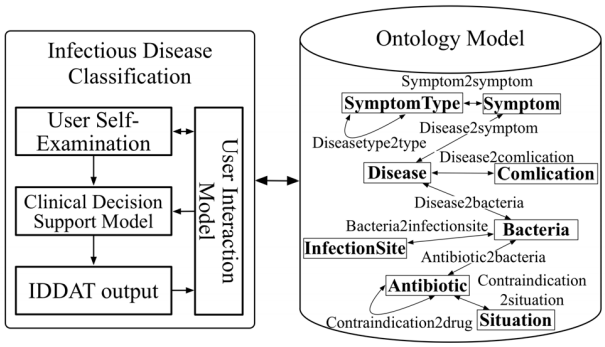

Figure 1. IDDAT architecture

A. Ontology Model

A MySQL database was created based on the ontology hierarchical conceptual schema, which covers the following nine dimensions: antibiotic, disease, complication, bacteria, animal, symptom type, symptom, infection site, and situation.

We generate an antibiotic ontology by reusing the existing ontologies such as Disease Ontology (DO), Infectious Disease Ontology (IDO), Human Phenotype Ontology (HPO), NCBI organismal classification ontology, and DrugBank. Through entity linking, these existing ontologies were linked to the MySQL database via the ontology components "class name" and "alias", which are tagged by the relation "hasExactSynonym" in Web Ontology Language (OWL). We compare the information between the new input class and the existing class in the database. Given the existing classes in the database, we merge the new input information into the corresponding existing classes. Classes not found in the database are regarded as new classes.

For websites such as Wikipedia, we only crawl a depth of two layers to ensure the relevance of the content. In the "list of antibiotics"[1] on the Wikipedia webpage where antibiotics are classified by class, we can extract the antibiotics name, common uses, mechanism of action (MOA) as well as the hierarchical relationships between antibiotics.

From the Wikipedia infobox concerning infectious disease and antibiotics (see Fig. 2), we can extract information related to speciality, symptoms, duration, complications, causes, diagnostic methods, treatments and so on. The extracted information was linked to the MySQL database via the class name through entity linking.

Cholera

Specialty	Infectious disease
Symptoms	Large amounts of watery diarrhea, vomiting, muscle cramps[1][2]
Complications	Dehydration, electrolyte imbalance[1]
Usual onset	2 hours to 5 days after exposure[2]
Duration	Few days[1]
Causes	*Vibrio cholerae* spread by fecal-oral route[3][1]
Risk factors	Poor sanitation, not enough clean drinking water, poverty[1]
Diagnostic method	Stool test[1]
Prevention	Improved sanitation, clean water, cholera vaccines[4][1]
Treatment	Oral rehydration therapy, zinc supplementation, intravenous fluids, antibiotics[1][5]
Frequency	3–5 million people a year[1]
Deaths	28,800 (2015)[6]

Figure 2. Wikipedia "Cholera" infobox

The Antibiotic Guidelines (2017) and Johns Hopkins ABX (Antibiotic) Guide are considered reliable guiding documents. Codes 001-139 (infectious and parasitic diseases) of International Classification of Diseases (ICD-9) [13] is adopted to recognize the names of infectious disease and complications, while the named entity recognition (NER) is employed to extract other knowledge related to infectious disease. Taking the extraction of symptom and clinical sign as an example. We first build a symptom corpus based on the Human Phenotype Ontology (HPO), which is a standardized vocabulary of phenotypic abnormalities encountered in human disease. Then we use NegEx [14]

[1]https://en.wikipedia.org/wiki/List_of_antibiotics

to find negation scopes in the patient-entered information, and identify symptom relevant entities with the help of the symptom corpus. Finally, the extracted symptom is matched in the MySQL database to determine whether to retain or remove it.

The Owlready[2] package is used to convert MySQL tables to OWL ontology. An ontology related to the infectious disease diagnosis and therapy is thereby generated.

B. Clinical Decision Support Model

The Clinical Decision Model, which consists of disease diagnosis module and disease therapy module, works based on the ontology. The ontology helps IDDAT to process the user's input, identify the infectious diseases and pathogenic bacteria, and decide the therapeutic plan.

1) Disease Diagnosis Module: This module attempts to obtain an etiological diagnosis by analyzing the patient's self-examination in the diagnosis stages:

a) High risk options. Patients provide information regarding whether they suffer from kidney low function, whether they are elderly individuals or infants, and whether they are pregnant or breastfeeding. If the patient chooses one of these options, IDDAT will identify as risky case and warn the user to consult a doctor.

b) Body temperature. Fever can be divided into: no fever (36∼37.2°C), low fever (37.3∼38.0°C), fever (38.1∼39.0°C), high fever (39.1∼40.0°C), and ultra-high fever (40.1°C or above). The infectious diseases caused by different bacteria have different body temperature. For example, the pathogenic bacteria such as viruses, atypical, positive bacteria, negative bacteria, enterobacter, nonfermenters causes the high fever, fever, high or ultra-high fever, fever, low fever, and no fever respectively. This step reduces the possible range of bacteria to B_α.

c) Infection location. Infections can be classified by the location or organ system that is infected by different types of pathogenic bacteria B_β. The identification of infection locations can further confirm whether the previous step accurately identifies the type of bacteria. We infer the relevant bacteria and disease by taking the intersection of the steps 1.b and 1.c:

$$D_a = \text{map2disease}(B_\alpha \cap B_\alpha) \tag{1}$$

If the infection location cannot be confirmed, the patient may select "unknown" and proceed to the next step.

d) Symptom and clinical sign. Different symptoms noticed by a patient reflect the presence of different infectious diseases. In IDDAT, each symptom or clinical sign selected by the patient increases the probability of illness by 3%. IDDAT performs disease inference based on selected symptoms and signs, then provides a list of possible diseases D_b.

[2]pythonhosted.org/Owlready/

e) Complication. Complication indicates the unfavorable evolutions or consequences occur in certain diseases. The presence of complications can further distinguish or confirm the type of infectious disease. For each disease among the possible diseases obtained in the step 1.d, the presence of the corresponding complication will increase the likelihood of illness by 5%.

The diagnosis results is interactively produced by the aforementioned self-examination steps. A list of possible diseases will be determined based on the intersection of steps 1.c and 1.d.

$$D = D_a \cap D_b \tag{2}$$

IDDAT will rank the possible infectious diseases in descending order of probability according to the results obtained in Eq.(2).

C. Disease Therapy Module

The disease therapy module is used to evaluate the severity of a disease. After identifying the pathogenic bacteria and infectious disease, IDDAT attempts to select appropriate active antibiotics based on the antimicrobial spectrum [14]. For pathogenic bacteria, the antibiotics with a sufficient action rate (equal to 2 or 3) are considered useful for the therapy, where 3 is very good, 2 is good, 1 is poor and 0 is inactive. Fig. 3 illustrates the framework of the therapy module.

*a) Antibiotic and drug data compilation.*IDDAT identifies the contraindication between the antibiotics and drugs that the patient is currently taking, and determines the contraindication of combined antibiotic therapy. IDDAT provides the appropriate antibiotics along with a drug introduction from an authoritative document, including acceptable or unacceptable uses, frequency of administration, dosage, etc. IDDAT will highlight contraindicated antibiotics in red as a warning.

b) Compilation of patient situation and antibiotic. In Addition to drug-drug interactions, drug contraindications are also a noteworthy concern specific to certain population. For example, patients with a history of tendon disorder may have an adverse reaction to moxifloxacin molecules. The option selected by the patient will help IDDAT determines if the patient has contraindications to a certain antibiotic. IDDAT will highlight contraindicated antibiotics in orange.

c) Final treatment guidelines. IDDAT provides the following information as a reference for IDDAT users: possible infectious disease name, appropriate treatment plan, recommended antibiotics and corresponding introduction, and relevant clinical cases or medical records.

III. Results and Discussion

A. Data set

In the ontology we generated, there are 1,317,018 classes, 7,731,914 axioms, and 1,269,340 inheritance relations involving infectious diseases, antibiotics, syndromes, and other

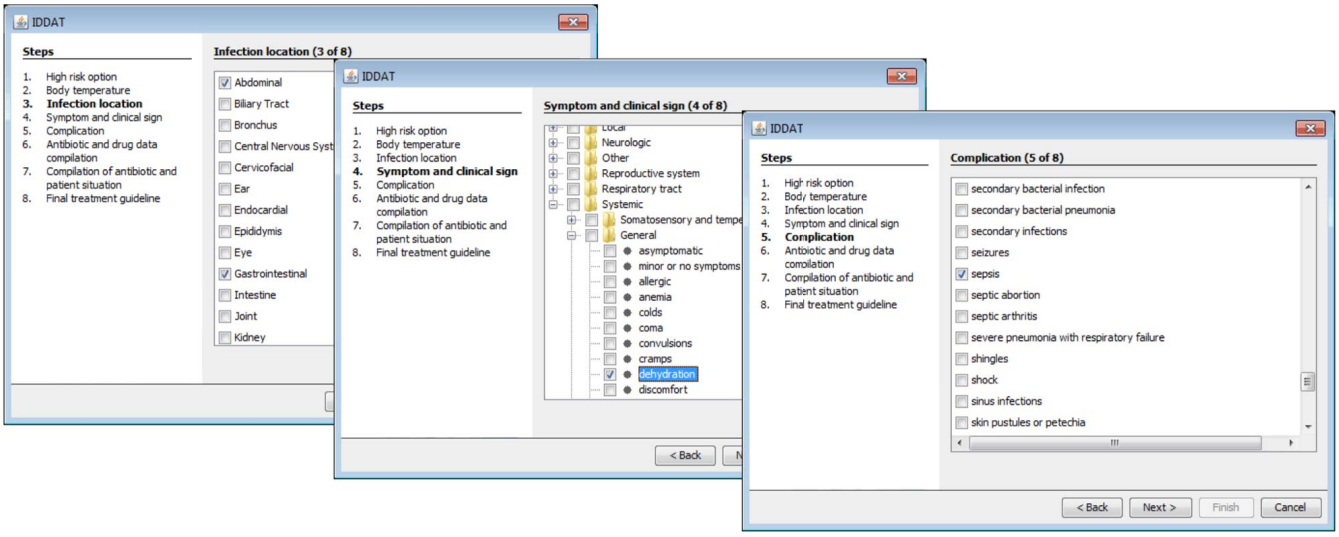

Figure 3. Self-selection of infection location, symptoms and complications

relevant knowledge. IDDAT contains 510 infectious diseases and corresponding treatment methods in combination with 336 different infection locations, 381 types of complications, 942 relevant symptoms of the circulatory, urinary and other systems, 838,651 types of bacteria, 366 types of antibiotics, 1,511 pairs of antibiotic resistance between bacteria and antibiotics, 445 pairs of drug interaction relationships and 87 pairs of contraindication for antibiotic-specific population.

B. Performance of IDDAT for Infectious Disease Identification

The following is an example of babesiosis diagnosis and therapy that can be used as a reference for IDDAT users.

Assume that the test user did not specify special pathological or physiological conditions, thereby no alarm of urgent therapy is triggered. The user first selected his/her body temperature as "fever". IDDAT thus excluded diseases caused by viruses and positive bacteria. The user then chose any infected locations, including the urinary tract, abdominal, and gastrointestinal (see Fig. 4 left panel).

Accordingly, possible pathogens included Bacteroides, Campylobacter, E. coli, Chlamydia, Yersinia, Salmonella, Candida, Fusobacterium, Pseudomonas, and 10 other bacteria.

Based on the symptoms and clinical signs selected by the user, i.e., abdominal pain and dehydration (see Fig. 4 middle panel), IDDAT identified the following diseases corresponding with the symptoms:

- Campylobacteriosis: has_symptom (fever, headache, diarrhea, dehydration, dysentery, cramps, abdominal pain...)
- Chlamydia: has_symptom (mucopurulent cervical discharge, dehydration, abdominal pain, pain with sexual intercourse, fever, pain with urination, urinary urgency, testicular pain or swelling...)
- Salmonellosis: has_symptom (enteritis, diarrhea, fever, abdominal pain, dehydration, vomiting, hypovolemic shock, septic shock, oliguria,...)
- Shigellosis (Bacillary dysentery): has_symptom (abdominal pain, dysentery, dehydration, diarrhea, fever, reactive arthritis, seizures...)
- Typhoid fever: has_symptom (exhausted and emaciated, fever, headache, chills, cough, coma, rose spots appear on the lower chest and abdomen, dehydration, diarrhea, fever, abdominal pain...)
- Yersiniosis: has_symptom (abdominal pain, fever, right-sided abdominal pain, joint pains, dehydration, diarrhea, rashes...)

When the user selected "sepsis" (see Fig. 4 right panel) to describe complications, IDDAT can almost conclude that the patient suffers from Campylobacteriosis, since the sepsis was included in the obvious complications of Campylobacteriosis while it was not included in the complications of other possible diseases. Because the user selected the "Terfenadine" as his/her drug currently in use, the antibiotic Erythromycin was high-lighted in red as a warning (see Fig. 5). Based on the identified disease and recommended antibiotics, IDDAT read the infectious disease knowledge structured by the ontology then output the treatment guidelines (e.g., corresponding treatment plan, relevant medical records and antibiotic introduction) via the **User Interaction Model** (see Fig. 5).

C. Evaluating IDDAT with ROC curves

A total of 510 infectious diseases are considered in this study. More than 90% of infectious disease-related medical

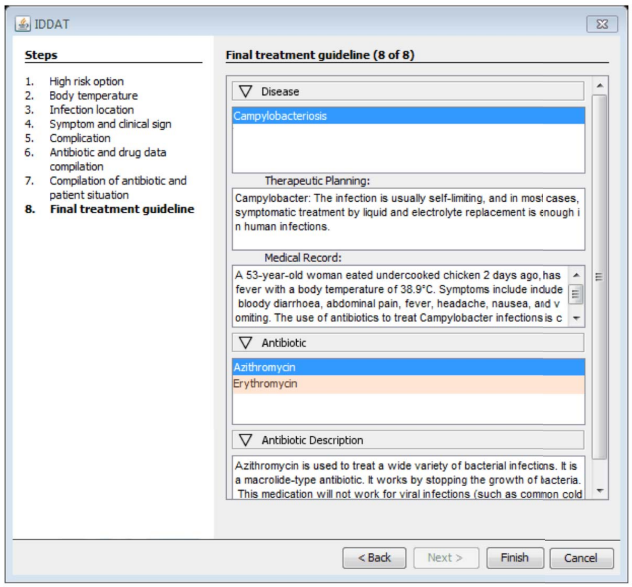

Figure 4. IDDAT therapy information

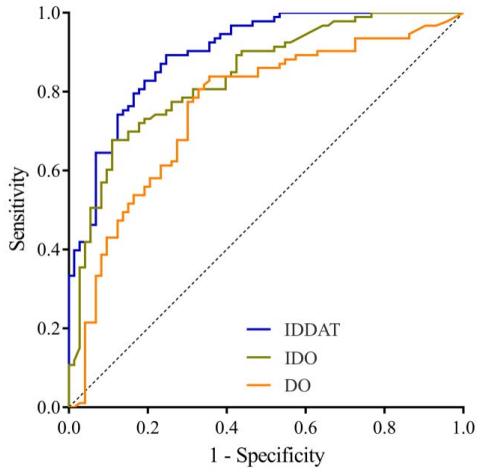

Figure 5. ROC chart and AUC for the evaluation of classifiers

records are utilized as training set, while 137 randomly selected medical records involving N infectious diseases (N = 1 to 137) are used as test sets. The doctors' diagnoses of these 137 medical records are compared with the results output by IDDAT to assess their accuracy.

Rather than evaluating the diagnosis results obtained from patient-centered self descriptions, we test the diagnostic classification of IDDAT based on our ontology, IDO, and DO (infectious disease-relevant knowledge) for the following reasons. 1) The patient-centered self-description is related to patient knowledge and subjectivity, the evaluation of which is difficult to quantify. 2) To our knowledge, none of the existing antibiotic treatment systems release their source codes, which makes us difficult to conduct the experimentally comparison.

Receiver Operating Characteristic (ROC) is used to reflect the accuracy of the diagnostic classification. We adopt different thresholds to determine whether the test data provides accurate diagnoses based on different ontology. Evaluation of the diagnosis classifier with the test data presents probability pairs [P1, P2] that specify a probability of 0 or 1.

Fig. 6 shows the results in terms of ROC curves. The experiments demonstrate that combined application of IDDAT and our ontology has robust superiority over competitors. The Area Under the ROC Curve (AUC) is 0.8991, revealing the effectiveness of IDDAT.

IV. Conclusion

This study proposes the IDDAT system for infectious diagnosis and therapy, that takes into account the characteristics of individual patients. IDDAT's operation depends on various ontologies. Technically, a reproducible schema of DSSs is proposed to identify infectious diseases using the knowledge structured in an ontology. Practically, the IDDAT can provide possible references for doctors or ontologists.

Acknowledgment

This work was financially supported by the National Natural Science Foundation of China (No.61602013), and the Shenzhen Key Fundamental Research Projects (Grant No. JCYJ20170818091546869).

References

[1] K. E. Nelson and C. M. Williams, Infectious disease epidemiology. Burlington, MA: Jones & Bartlett Publishers, 2013.

[2] J. Giesecke. Modern infectious disease epidemiology. Florida, USA: CRC Press, 2017.

[3] G. J. Kuperman, R. M. Gardner, and T. A. Pryor. HELP: a dynamic hospital information system. Berlin, Germany: Springer Science & Business Media, 2013.

[4] J. T. Johnson, R. L. Wagner, D. E. Schuller, J. Gluckman, J. Y. Suen, and N. L. Snyderman, "Prophylactic antibiotics for head and neck surgery with flap reconstruction," Archives of Otolaryngology-Head & Neck Surgery, vol. 118, pp. 488-490, 1992.

[5] N. Singh, P. Rogers, C. W. Atwood, M. M. Wagener, and V. L Yu, "Short-course empiric antibiotic therapy for patients with pulmonary infiltrates in the intensive care unit: a proposed solution for indiscriminate antibiotic prescription," American journal of respiratory and critical care medicine, vol. 162, pp. 505-511, 2000.2.

[6] M. Paul, S. Andreassen, E. Tacconelli, A. D. Nielsen, N. Almanasreh, U. Frank, R. Cauda, and L. Leibovici, "Improving empirical antibiotic treatment using TREAT, a computerized decision support system: cluster randomized trial," Journal of Antimicrobial Chemotherapy, vol. 58, pp. 1238-1245, 2006.

[7] K. A. Thursky, and M. Mahemoff, “User-centered design techniques for a computerised antibiotic decision support system in an intensive care unit,” International Journal of medical informatics, vol. 76, pp. 760-768, 2007.

[8] C. B. Litvin, S. M. Ornstein, A. M. Wessell, L. S. Nemeth, and P. J. Nietert, “Adoption of a clinical decision support system to promote judicious use of antibiotics for acute respiratory infections in primary care,” International journal of medical informatics, vol. 81, pp. 521-526, 2012.

[9] V. Sintchenko, J. R. Iredell, G. L. Gilbert, E. Coiera, “Hand-held computer-based decision support reduces patient length of stay and antibiotic prescribing in critical care,” Journal of the American Medical Informatics Association, vol. 12, pp. 398-402, 2005.

[10] F. Griffiths, J. Cave, F. Boardman, J. Ren, T. Pawlikowska, R. Ball, A. Clarke, and A Cohen, “Social networks-the future for health care delivery,” Social science & medicine, vol.75, pp. 2233-2241, 2012.

[11] S. Khler, M. H. Schulz, P. Krawitz, S. Bauer, S. Dolken, C. E. Ott, C. Mundlos, D. Horn, S. Mundlos, and P. N. Robinson, “Clinical diagnostics in human genetics with semantic similarity searches in ontologies,” American Journal of Human Genetics, vol. 85, pp. 457-464, 2009.

[12] Y. Shen, K. Yuan, D. Chen, J. Colloc, M. Yang, Y. Li, and K. Lei, “An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription,”Artificial intelligence in medicine, pp. 20-32, 2018.

[13] H. Quan, V. Sundararajan, P. Halfon, A. Fong, B. Burn, J. C. Luthi, D. Saunders, C. A. Beck, and W. A. Ghali, “Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data,” Medical care, pp. 1130-1139, 2005.

[14] W. W. Chapman, D. Hilert, S. Velupillai, M. Kvist, M. Skeppstedt, B. E. Chapman, M. Conway, M. T. Mowery, and L. Delegerd, “Extending the NegEx lexicon for multiple languages,” Studies in health technology and informatics, vol. 192, pp. 677-681, 2013.

[15] S. B. Singh, H. Jayasuriya, J. G. Ondeyka, K. B. Herath, C. Zhang, D. L. Zink, N. N. Tsou, R. G. Ball, A. Basilio, O. Genilloud, “Isolation, structure, and absolute stereochemistry of platensimycin, a broad spectrum antibiotic discovered using an antisense differential sensitivity strategy,” Journal of the American Chemical Society, vol. 128, pp. 11916-11920, 2006.