International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 2 | ISSN : 2456-3307

Disease Predictive, Best Drug : Big Data Implementation of Drug Query with Disease Prediction, Side Effects & Feedback Analysis

Naveenraj I, Thurab Ahmed A, Dilly Ganesh N, Gokula Krishnan, Hema latha, Subedha V
Department of CSE, Panimalar Institute Of Technology, Chennai, Tamil Nadu, India

ABSTRACT

In the existing system, several precautions should be taken in using pharmaceutical drugs, for both healthcare experts, who prescribe and administer drugs, and for drug purchasers. In the proposed gadget, side effects and effectiveness, relies upon on traits of patients, which include age, gender, existence, and genetic profiles. Our intention is to offer a device to help experts and consumers in locating and selecting pills. To reap this intention, we broaden an approach that lets in a person to query for pills that satisfy a fixed of situations based on drug residences, inclusive of drug symptoms, facet outcomes, and drug interactions, and additionally takes into account affected person profiles. The amendment paintings is the combination of big information and android primarily based input user with the aid of any consumer for smooth facts evaluation method. We also analyze the disease and satisfactory drug cautioned to that precise affected person via large facts analysis. Use can put up the query thru gadget or via android application additionally. We also set up appointment to the fine physician for the consultation based on person feedbacks.

Keywords : Predictive and Personalized Query System

I. INTRODUCTION

In the existing system, several precautions should be taken in using pharmaceutical drugs, for both healthcare experts, who prescribe and administer drugs, and for drug purchasers. In the proposed gadget, side effects and effectiveness, relies upon on traits of patients, which include age, gender, existence, and genetic profiles. Our intention is to offer a device to help experts and consumers in locating and selecting pills. To reap this intention, we broaden an approach that lets in a person to query for pills that satisfy a fixed of situations based on drug residences, inclusive of drug symptoms, facet outcomes, and drug interactions, and additionally takes into account affected person profiles. The amendment paintings is the combination of big information and android primarily based input user with the aid of any consumer for smooth facts evaluation method. We also analyze the disease and satisfactory drug cautioned to that precise affected person via large facts analysis. Use can put up the query thru gadget or via android application additionally. We also set up appointment to the fine physician for the consultation based on person feedbacks.

II. METHODS AND MATERIAL

1. Supporting Drug Prescription via Predictive and Personalized Query System
   Samamon Khemmarat and Lixin Gao

In this paper, we advocate an technique for answering drug queries to guide drug prescription. To address incomplete and noisy records, we allow both precise and close matches when answering queries. The answers are ranked by using using the shape of a drug information network to quantify the chance of associations among drug and drug houses in the case that the institutions are lacking. We show how our approach could help practitioners to make informed selection while prescribing capsules via several examples.While this work addresses one of the principal problems in supporting prescription, which is how to obtain and rank the solutions, further paintings is needed to gain a complete prescription guide system. First, our modern device assumes a question graph is given by way of customers. For better usability, a module that assists customers in building the query graphs, which includes a shape-based totally user interface or a module that translates natural language queries to query graphs, have to be evolved. Second,as noted in our assessment, our machine is supposed for assisting specialists, now not changing them. The answers received need to be reviewed through professionals. Thus, a module that provides users with assisting proof for solutions can be useful. Additionally, there are many other elements aside from sideeffects and drug interactions which can be vital in prescribing capsules. For our future paintings, we would really like to reap evaluation and feedbacks from practitioners and improve our machine to better support their desires in keeping with real-international usage.

2. DrugBank 3.0: a comprehensive resource for ‘Omics’ research on drugs Craig Knox , Vivian Law…ET.All

DrugBank 3.0 carries a considerable range of improvements over its predecessor (DrugBank 2.Zero). As highlighted during this text, numerous improvements have been made in the amount, excellent, depth and organization of the information provided. These include the addition of latest drugs, new targets, new information fields, new links and new tools. DrugBank three.0 now carries illustrated drug-action pathways, drug transporter information, drug metabolite facts, pharmacogenomic statistics, damaging drug response statistics, ADMET statistics, pharmacokinetic information, extensive computed belongings information and chemical classification information. DrugBank three.Zero additionally offers increased database links, stepped forward search tools for drug–drug and meals–drug interplay, new equipment for searching and viewing drug pathways and loads of recent drug entries with designated patent, pricing and producer records. These additions were complemented by using improvements to the pleasant and quantity of existing data, especially with reference to drug goal, drug description and drug action facts. With those enhancements DrugBank three.0 ought to be a whole lot extra beneficial for a much broader range of ‘omics’ packages. It is was hoping that with extra user feedback, DrugBank will retain to expand to in shape the desires of its users and offer an an increasing number of beneficial, information-wealthy drug useful resource.

3. A side effect resource to capture phenotypic effects of drugs Michael Kuhn, Monica Campillos…Et.All

For fifty two% of all capsules, the sections describing unfavorable reactions at the labels contain facts approximately the frequencies of facet consequences. The FDA SPL labels have been amenable to a thorough analysis and actual frequency information can be extracted from the tables detailing the facet effect frequencies via analyzing the contents of the desk captions and cell(e.G., caption in the first row: ‘Percentage of Patients Reporting Event’, caption in the first column: ‘anemia’, mobile: ‘12’: 12% of the patients document anemia). Although it become no longer feasible to infer precise frequency records for PDF labels, standardized preferred frequency levels (‘rare’, ‘infrequent’ and ‘common’) will be extracted for each label sorts from sections that listed the frequency and a number of aspect outcomes (e.G., ‘ frequent: headache, dizziness, ⋯’) Furthermore, facet effects that took place in the put up - advertising and marketing section had been also extracted from all labels. For any given drug–aspect effect pair, multiple frequencies might be to be had, as an example, from medical trials in exclusive indication areas. On the website, we display all suggested frequencies to the person.

4. KEGG : Kyoto Encyclopedia of Genes and Genomes
   Minoru Kanehisa* and Susumu Goto

The KEGG pathway illustration makes a speciality of the network of gene products, mostly proteins but consisting of practical RNAs.As illustrated in Figure 2, the metabolic pathway is a community of indirect protein–protein interactions, that is truly a community of enzyme–enzyme relations. In assessment, the regulatory pathway often includes direct protein–protein interactions,which include binding and phosphorylation, and some other magnificence of oblique protein–protein interactions, that are family members of transcription factors and transcribed gene products via gene expressions. The generalized protein–protein interaction community that consists of those 3 kinds of interactions is an summary network, however it's miles specifically useful to hyperlink with genomic records because the nodes (gene merchandise) of this network can be without delay correlated with the nodes (genes) within the genome. With this concept of generalized protein–protein interaction network, we are increasing the collection of manually drawn reference pathway diagrams.

5. An introduction to ROC analysis
   Tom Fawcett

ROC graphs are a totally beneficial device for visualizing and comparing classifiers. They are able to offer a richer degree of type performance than scalar measures inclusive of accuracy, mistakes charge or errors fee. Because they decouple classifier performance from magnificence skew and blunders fees, they have blessings over different assessment measures inclusive of precision-recall graphs and lift curves. However, as with all evaluation metric, the use of them wisely calls for knowing their characteristics and limitations. It is was hoping that this article advances the overall know-how approximately ROC graphs and allows to promote better evaluation practices inside the pattern reputation network.

6. Exploring The Pharmacogenomics Knowledge Base (Pharmgkb) For Repositioning Breast Cancer Drugs By Leveraging Web Ontology Language (Owl) And Cheminformatics Approaches*
   Qian Zhu, Cui Tao,Feichen Shen, Christopher G. Chute

This report presents our preliminary work specializing in computational drug repositioning application development leveraging PGx information integration and Semantic Web era exploration for FDA accepted BCDs. We have correctly validated the application of this software to reposition current BCDs with new uses, and locate capacity negative results. Our work illustrates that PGx facts offers enough information to support drug repositioning and, furthermore, that Semantic Web era offers technical support for formal representation and semantic inference of facts. This is our first attempt to use a PGx resource and Semantic Web technology for address drug repositioning in a computational manner. With the promising results of this look at, we are able to amplify this investigation in several directions: 1) In the modern have a look at, we explored most effective PharmGKB as a PGx aid, which is not enough to perceive extra novel associations for BCDs. We will combine extra PGx-associated resources, along with an FDA biomarkers table, the DrugBank database, the Comparative Toxicogenomics Database, and the Kyoto Encyclopedia of Genes and Genomes. 2) Once extra PGx assets are included, one drug is probably inferred to a couple of PGx institutions. Then we are able to propose to define some “gold requirements” for prioritizing the relevance of those associations to particular drugs. The standards is probably constructed at the range of co-occurrences of the PGx associations, as supported by means of guides, etc. 3) We worked only on BCDs on this have a look at. In future studies, we are able to enlarge our attempt to other most cancers drug classes or different classes of medication, inclusive of antidepressants, the use of the identical strategy that we carried out in this study

7. A text based drug query system for mobile phones
   Akhil Langer et.all..,

SVMs has been correctly used in modelling complex, real-international problems such as text and photograph categorisation (Joachims, 1998), pattern recognition (Burges, 1998),bioinformatics (Byvatov and Schneider, 2003), and many others. SVM has robust regularization houses i.E., it can generalise the version to new data. SVM has been observed to carry out well on statistics sets with many features even though there are only a few facts times to train the model. The accuracy of the SVM version for our drug queries utility complies with outcomes received in diverse different utility domains. Some similar experiences can be observed in Osuna et al. (1997), Zhang and Lee (2003), and so forth.In this paper, we supplied our novel method of providing drug related facts to humans thru natural language queries in a manner that meets their immediately need in a value-powerful manner. Through various assets (on-line drug reference assets inAppendix A) we identified the classes into which a drug query may be categorised in order that simplest the required information is extracted and presented to the data seeker. Using classical SVMs we finished type accuracy of eighty one% in this venture. The biggest benefit of the system is that no installation of any kind is required on the person’s mobile phone and the QA machine can be speedy incorporated by means of a network provider issuer.Moreover, the system meets emergency wishes in an inexpensive manner which in any other case would have required consulting a doctor The machine as defined in this paper calls for minimal human involvement. The computerized method reduces response time for queries and at the identical time gives a diploma of flexibleness to the user in terms of the way they choose to formulate their query.This is actually an advantage over DICs which require manned personnel to reply touser queries. The flexibility to invite herbal language queries is tremendous to much less privileged folks that can't afford to have access to net enabled mobile telephones. More importantly this gadget architecture can be extended to different domain names as properly. Even while cell internet does gain recognition a gadget that translates herbal language queries probably sporting spelling errors will still be beneficial. This machine also can be used as an up to date reference tool through less skilled docs working towards in villages where net isn't always to be had.Extensive person trying out is necessary to eventually make the system to be had for customers.We have already conducted preliminary person trying out of associated structures which utilise NLP strategies for processing SMS queries targeted on the healthcare area (Langer and Banga, 2010; Banga et al., 2010; Langer et al., 2010). Find a Doctor (Langer and Banga,2010) processed unfastened shape SMS queries to extract medical doctor name, address and the health trouble to discover a closest physician for the fitness hassle. We advanced some other system (Langer et al., 2010) which completed a seek over Yahoo! Answers in fitness area to locate near matching answers to related questions for an SMS query with capacity brief paperwork commonly utilized in SMS domain. We acquired encouraging feedback from the folks that tested those systems due to their flexibility. We learnt that incorporating flexibility in a SMS primarily based device makes it very thrilling each from the consumer and developer factors of view. We agree with ours is a preliminary device that leverages properly installed algorithms from NLP and gadget learning to reduce the workload of DICs and additionally sets an instance for other factoid based systems. In this paper we've got not taken into consideration very well the intricacies of SMS texting language (besides coping with common spelling errors). Incorporating language modelsand noisy query evaluation can similarly improve the user enjoy in the sort of gadget via relaxing the regulations at the varieties of questions dealt with. Larger training set that can be acquired as soon as the system is taken into pilot level with a network provider issuer willhelp us better recognize the person behaviour and the type of queries anticipated through SMS. Based in this evaluation the gadget may be adapted to handle wider magnificence of question.

Current available drug databases offer us with greater general facts e.G., it'll supply all the aspect-effects of a drug even supposing information on a selected aspect-effect is asked. Either developing a database having targeted description may be finished or looking the net drug reference resources for matching FAQs may be carried out to retrieve unique records. This will permit the machine to reply queries like What is the aspect-impact of <drug_name> on <organ>?, Does <drug_name> reason <symptom>?, and many others. A practical machine will require efforts in the right course to make sure adoption of carrier by means of most people of population. In their study, Lee et al. (2012) talk the outcomes of perceived leisure for offerings adoption.

8. Panacea, a semantic enabled drug recommendations discovery framework Charalampos Doulaverakis et.all..,

The paper supplied Panacea, a framework for semanticenabled drug guidelines discovery. The framework makes use of a layered reasoning technique were the medical oncology and the patient statistics instances are fed to an extended RDFS reasoner so one can infer implicit knowledge. Drug suggestions are generated the use of the second reasoning layer in which any common rule engine can be used. As a proof of concept implementation, Jena reasoner and the Drools rule engine has been included.Two distinctive opinions have been conducted. One performance assessment concerning necessities and efficiency of the proposed technique, and a fine evaluation regarding the system’s outcome in terms of actual scientific records. The best assessment gave insights concerning viable extensions that would make the gadget extra in keeping with modern clinical practice. Future work on Panacea will consciousness on providing ways to deal with the issues exposed in the course of the nice assessment and offer effects that greater carefully suit a medical doctor’s decision. These ought to include improvements including the weighting of interactions and contraindications in keeping with a severity observation and probabilistic inference primarily based on those weights. To this stop, Drools is being prolonged with a fuzzy reasoning engine, which even as it’s still in development,it’s actively supported and it's miles mature enough to be able to use it as a testing framework. Finally, the addition of dosage guidelines inside the rules is an ongoing work.

9. Towards pharmacogenomics knowledge discovery with the semantic web
   Michel Dumontier and Natalia Villanueva-Rosales

In this newsletter, we display that the semantic internet framework enables records integration and semantic annotation, two key additives of expertise discovery. While there exists considerable debate in terms of first-rate practices for the design of ontologies,we consider that easy, intuitive conceptualizations are stimulated by using clean requirements stemming from diverse units of use cases and offer a persuasive means to evaluate the application of ensuing ontologies.In flip, using standardized formal expertise representation languages inclusive of OWL makes viable a reliable illustration of information,whose syntax and semantics cause a steady interpretation. With a populated ontology in hand,new approaches to mine the knowledge base found on the semantic web will find new styles in the facts, which in turn may be captured as expressive ontologies. Our works to extend PharmGKB using the OWL framework with the curated information on the pharmacogenomics of despair demonstrates how 1/3 events can make contributions and increase to a growing established know-how base. Clearly, the semantic internet paperwork the basis for a promising research location in pharmacogenomics know-how discovery..

10. Medical Question Answering:Translating Medical Questions into SPARQL Queries
    Asma Ben Abacha et.all.., Contributions

In this paper, we tackled automatic question evaluation inside the clinical domain and supplied an unique method having main characteristics:

- ✓ The proposed technique allows handling distinctive types of questions, such as questions with more than one expected solution type and a couple of cognizance.
- ✓ It lets in a deep analysis the use of special facts extraction strategies primarily based on

(i) area expertise and
(ii) natural language processing techniques (e.G. Use of patterns, device getting to know) which permit extracting medical entities but also semantic members of the family or even additional data about the affected person (age,sex, etc.).

Three Our approach is based on Semantic Web technologies,which give greater expressivity, wellknown formalization languages and makes our corpus and question annotations sharable via the web.

Limitations and Future Work

Although our method turned into aimed to be frequent (i.E. W.R.T. The target query sorts), greater specific processes are still required to deal with (i) complex questions (e.G. Why, whilst)and (ii) questions with new semantic family members that are not described in our reference ontology.

To address scalability in this final trouble, we plan to carry out a syntactic analysis of the natural language question and test the contribution of syntactic dependencies on elements:

i. Confirmation of previously extracted semantic family members
ii. Dtection of unknown members of the family: syntactic dependencies (Subject-Verb-Object) can update triplets (Entity1-Relation-Entity2) if an abstract RDF assets is defined within the reference ontology.

Our very last aim is the theory of a question answering system for the clinical area. Figure three presents the planned architecture for the whole query-answering gadget (QAS).

11. Querying Web-Scale Information Networks Throug Bounding Matching Scores Jiahui Jinyz, Samamon Khemmaratz, Lixin Gaoz, Junzhou Luoy

In this paper, we address the pinnacle-okay famous person question hassle for large informative networks. We gift an set of rules for locating the pinnacle-okay high-quality solutions efficiently. The set of rules utilizes the sure-based top-okay detection mechanism to attain the solutions speedy without indexing. A distributed system for the algorithm is evolved to allow scalability. Our experiments validated that the proposed approach can provide meaningful solutions and might efficiently compute solutions for a query on massive records networks with billions of nodes. Ciently solution queries on billion-node in-formation networks.

12. A Distributed Approach for Top-k Star Queries on Massive Information Networks
    Jiahui Jinyz, Samamon Khemmaratz, Lixin Gaoz, Junzhou Luoy

In this paper, we address the pinnacle-k superstar question problem for massive informative networks. We present an algorithm for locating the pinnacle-okay first-rate solutions efficiently. The algorithm makes use of the bound-based totally pinnacle-k detection mechanism to reap the answers quick with out indexing. A allotted system for the algorithm is advanced to permit scalability. Our experiments tested that the proposed technique can provide significant solutions and can effectively computeanswers for a query on big records networks with billions of nodes.

13. Neighborhood Based Fast Graph Search in Large Networks
    Arijit Khan et.all..,

In this paper, we described a brand new graph similarity measure, community primarily based graph similarity, and proposed an statistics propagation version to transform a big network into a hard and fast of multidimensional vectors, where state-of-the-art indexing and similarity seek algorithms are available. We proved, beneath this degree, that subgraph similarity seek is NP tough, whilst graph similarity fit is polynomial. We delivered a criterion to choose the best propagation price with respect to extraordinary node labels in a graph. We in addition investigated the strategies to index the community vectors and to compress them with the aid of deleting non-discriminative labels, as a consequence optimizing the query processing time. The proposed method, called Ness, isn't simplest green, but additionally robust in opposition to shape adjustments and data loss. Empirical effects display that it can quickly and as it should be discover incredible fits in massive networks, with negligible time value. In destiny paintings, it is going to be thrilling to remember the graph alignment trouble, whilst the node labels in graphs are not exactly same, i.E the same person may have slightly extraordinary usernames in Facebook and Twitter

14. NeMa: Fast Graph Search with Label Similarity
    Arijit Khan.. et.all..,

In this paper, we have introduced NeMa , a novel graph querying framework through subgraph matching that lets in for ambiguity in both structure and node labels. We convert the neighborhood of every node into a multi-dimensional vector, after which practice an inference set of rules to pick out the most desirable graph suits. We in addition check out how NeMa may be extended to various graph question-processing packages, together with RDF query answering, graph matching with facet labels, and locating pinnacle-k approximate fits. Our experimental consequences over real-life datasets show that NeMa correctly reveals exquisite suits, in comparison to ultra-modern graph querying techniques. In future paintings, one may remember approximate subgraph matching over graph streams, and additionally more sophisticated label similarity metrics,e.G.,ontology and semantic similarity

15. Characterization of Schizophrenia Adverse Drug Interactions through a Network Approach and Drug Classification
    Jingchun Sun,Min Zhao Ayman H. Fanous and Zhongming Zhao

In this examine, we supplied a complete research of damaging drug-drug interactions of the antipsychotics used to treat schizophrenia. We integrated the network analysis with ATC drug classifications, which furnished the destructive drug interplay traits of SCZ pills as well as normal and atypical pills. However, a great deal greater work is wanted to gather extra detrimental drug interaction information and develop advanced pharmacogenomics network methods. Potential findings could be used to predict unfavorable drug-drug interactions and enhance the coadministration of more than one drugs, which in turn may lead to the avoidance of the drugdrug interplay negative outcomes.

III. RESULTS AND DISCUSSION

A. Existing System

In the existing system, several precautions have to be taken in the use of pharmaceutical drugs, for each healthcare experts, who prescribe and administer pills, and for drug consumers.

DISADVANTGES

- Does not analyze the disease
- There is no automatic drugs
- Less security
- There is no feedback system

B. Proposed System

In the proposed system, aspect effects and effectiveness, depends on characteristics of patients, consisting of age, gender, life, and genetic profiles. Our intention is to provide a tool to assist specialists and consumers in finding and deciding on tablets. To gain this purpose, we broaden an method that allows a user to query for capsules that fulfill a hard and fast of conditions based on drug residences, which includes drug indicators, side consequences, and drug interactions, and additionally takes into account affected person profiles.

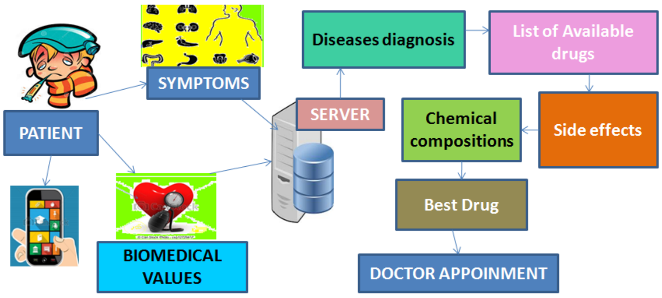

Figure 1. Proposed Architecture Diagram

C. Modification

The MODIFICATION work is the integration of Big Data and Android based totally input consumer by any user for smooth statistics evaluation procedure. We also examine the disease and great drug counseled to that unique affected person via Big Data analysis. Use can publish the query via device or via Android Application also. We also set up appointment to the Best Doctor for the session based totally on user feedbacks.

ADVANTAGES

- Easily analyze the disease
- Also provides the drugs based on disease
- Suggest best doctor
- Arrange the doctor arrangement

D. Algorithm / Methodology

- ❖ Machine Learning, Email

IV.CONCLUSION

In this paper, we propose an method for answering drug queries to support drug prescription. Our awareness is on the way to obtain and rank answers based on incomplete data and provide personalization. To cope with incomplete and noisy records, we permit both genuine and close suits whilst answering queries. We also gift an intuitive technique to display answers to users, which pursuits to help users to recognize the ranked outcomes and possibly refine their queries.

V. REFERENCES

[1]. S. Khemmarat and L. Gao, "Supporting drug prescription via predictive and personalized query system," in PervasiveHealth. IEEE, 2015.

[2]. C. Knox et al., "Drugbank 3.0: a comprehensive resource for omics research on drugs," Nucleic acids research, vol. 39, no. suppl 1, pp. D1035–D1041, 2011.

[3]. M. Kuhn et al., "A side effect resource to capture phenotypic effects of drugs," Molecular systems biology, vol. 6, no. 1, p. 343, 2010.

[4]. M. Kanehisa and S. Goto, "Kegg: kyoto encyclopedia of genes and genomes," Nucleic acids research, vol. 28, no. 1, pp. 27–30, 2000.

[5]. T. Fawcett, "An introduction to roc analysis," Pattern recognition letters, vol. 27, no. 8, pp. 861–874, 2006.

[6]. K. Sangkuhl et al., "Pharmgkb: understanding the effects of individual genetic variants," Drug Metab. Rev., vol. 40, no. 4, pp. 539–551, 2008.

[7]. A. Langer et al., "A text based drug query system for mobile phones," Int. J. Mob. Commun., vol. 12, no. 4, pp. 411–429, Jul. 2014.

[8]. C. Doulaverakis et al., "Panacea, a semantic-enabled drug recommendations discovery framework," J. Biomed. Semant., vol. 5, p. 13, 2014.

[9]. M. Dumontier and N. Villanueva-Rosales, "Towards pharmacogenomics knowledge discovery with the semantic web," Briefings in bioinformatics, vol. 10, no. 2, pp. 153–163, 2009.

[10]. A. Ben Abacha and P. Zweigenbaum, "Medical question answering: translating medical

questions into sparql queries," in Proceedings of the 2nd ACM SIGHIT. ACM, 2012, pp. 41–50.

[11]. J. Jin et al., "Querying web-scale information networks through bounding matching scores," in WWW 2015, 2015, pp. 527–537.

[12]. J. Jin et al., "A distributed approach for top-k star queries on massive information networks," in ICPADS 2014, 2014, pp. 9–16.

[13]. A. Khan et al., "Neighborhood based fast graph search in large networks," in SIGMOD 2011. ACM, 2011, pp. 901–912.

[14]. A. Khan et al., "Nema: Fast graph search with label similarity," in Proc. of the VLDB Endowment, vol. 6, no. 3, 2013, pp. 181–192.

[15]. J. Sun, H. Xu, and Z. Zhao, "Network-assisted investigation of antipsychotic drugs and their targets," Chemistry & biodiversity, vol. 9, no. 5, pp. 900–910, 2012.