Mining Drug Properties for Decision Support in Dental Clinics

Abstract. With poly-pharmacy becoming more common, it is important for health providers to be aware of the drug profile of patients before prescribing. Although there are many methods on extracting information on drug interactions, they do not integrate with the patients' medical history. This paper describes state-of-the art approaches in extracting the term frequencies of drug properties, and using this knowledge to decide if a drug is suitable for prescription after considering if there is any drug allergy the patient may have and the drugs that the patient is currently taking. An experiment is conducted to evaluate the accuracy of associating the similarity ratio in terms of their term frequencies to the similarity between them. Experimental evaluation of our model yields an accuracy of over 80% which is superior to models that use other methods. Since a drug is to be avoided if it is similar to a drug that patient is allergic to, our model will help dentist decide if a drug is suitable for prescription to the patient. Hence such an approach, when integrated within the clinical workflow will reduce prescription errors thereby increasing the health outcome of the patients.

Keywords: adverse relationship; drug allergy; drug properties; knowledge-base; personalised prescription; similarity ratio; term frequency

1 Introduction

With the increased in volume and complexity of data encountered by dentists, the use of decision support systems to help in their decision making is becoming more of a necessity than a luxury. In order for decision support systems to be readily accepted by dentists, they should have, among other features, the ability to provide assistance on drug prescription based on individual patient profile [6].

In fact, predicting drug interactions at point-of-care to reduce prescription error is important as an adverse event can lead to serious health consequences for the patients and resulting in expensive legal suits for the practitioners. One of the cause for hospital admissions worldwide is adverse drug reaction and incidence of such hospitalised patients can be as high as 24% [12]. Naturally, many such

admissions could have been avoided if more care are taken in the prescription of medicine by considering possible drug allergy the patient may have.

So far many methods have been developed with the aim of extracting information on drug interactions [3,4]. A recent work by [2] has developed a literature-based system to find out the similarity of a drug pair. However, unfortunately these methods do not integrate with the patients medical history within the clinical workflow of the clinic.

With this motivation in mind, this paper describes state-of-the-art approaches in determining if a drug pair is similar as well as using such information to support the prescription decision of the dentist. In this study, we proposed a three-tier conceptual framework - knowledge layer, data layer and user layer. Data mining is performed within the data layer, with knowledge extracted from the knowledge layer and presented to the user layer for decision making. The unique approach in our model is in the way the relationship of a drug pair is associated with the properties of the drug pair.

As with many decision support systems that are developed based on knowledge discovered from data mining [15], this paper describes a model which computes the similarity within a drug pair for predicting its suitability before the dentist prescribes it to the patient.

By using neighborhood similarities and textual data from currently available open source datasets to predict the relationship of a drug pair, knowledge obtained from data mining is used for prescription support of health care professionals. Based on a novel data-driven text mining technique, clusters of drugs which have adverse interactions are collected, together with their properties, by identifying their field markers from the web content. This information allows a similarity ratio to be computed which indicates if a drug pair can be safely prescribed to the patient.

Our work performs well compared to other methods of prediction, with a F measure of 64% with drug properties gathered from textual data obtained through bio medical sources. The model in this paper can be easily utilised in predicting if a drug pair is suitable for prescription, by considering the existing drugs that the patient is taking to avoid adverse interactions and also the drug allergies the patient may have to avoid cross allergy.

This study will help provide strategies in research agenda and priorities such as methodologies for knowledge reasoning and inference in the context of a dental clinic. Research outcome of this project will help dentists reduce potential risk of issuing allergic drugs to a patient and as a result, improve the quality of treatment in clinics. The advice on adverse drug reaction is increasingly important especially with an increased probability of an adverse interaction given the increasing use of multiple medications among patients.

This system which delivers information on interacting drug pairs based on the patients drug profile will also benefit those who are involved in clinical education relating to drug dispensing, such as in medicine, nursing and pharmacy. The practical use of data mining techniques in supporting the dentists prescription of drugs will has great potential to extend to the wider medical domain,

since there is also a need for doctors to ensure the safe prescription of drugs to patients. Our work will make significant contribution to the transformation of current health care industry to evidence-based, personalised health care, since patients individual conditions are considered in decision-making support within the clinical workflow.

The rest of paper is organised as follows: Section 2 discusses the related work in data mining and how model differs in the way drug-relationship is detected and deployed for use. Section 3 introduces the framework of our model and Section 4 outlines the parameters used for evaluating our work. This is followed by Section 5 where the results are discussed and compared with other approaches and Section 6 presents the conclusions obtained.

2 Related Work

Many systems have been developed using data mining techniques to explore drug interactions. In fact, such techniques are evolving quickly to improve the accuracy of the experiments, though in most situations results may not be sufficient to derive Drug Drug Interactions (DDI) [17]. A recent work by [1] attempt to determine DDI by identifying neutral candidates, negation cues and scopes from bio-medical articles. Features extracted from these articles include linguistic definition of negation, the position of the drugs discussed in the sentence and the linguistic-based confident level of an interaction. By using data sets from drug bank, it is reported that the results achieved an F measure of 68.4%. Text mining techniques are also recently used to predict protein interaction from bio medical literature [11].

Another common way of examining drug interactions is to extract relevant information from text. For example, [16] has developed a method that combines text mining and automated reasoning to predict enzyme- specific DDIs. [18] also uses text mining techniques to create features based on relevant information such as genes and disease names extracted from drug databases to augment limited domain knowledge. These features are then used to build a logistic regression model to predict drug-drug interaction (DDI).

Another study to extract information on drug-drug interactions (DDI) from biomedical text was proposed by Bui et al. [3]. DDI pairs are mapped according to their syntactic structure followed by the generation of feature vectors for these DDI pairs. These feature vectors are then used for the generation of a predictive model which classify the drug pair as interacting or not interacting [3].

Though these studies use data mining methods to extract relevant information for the prediction of drug interaction, unfortunately, these works only confine to the knowledge layer and data computing layer as outline in the three-tier framework in this paper.

The crucial need of using the knowledge obtained from data mining motivates us to develop the three-tier conceptual framework proposed in this paper. Although our system is similar to that proposed by [4] in terms of using information from the patient, the unique approach adopted in this paper goes one step

further in using such information to support the decision making process for the dentist at point-of-care within the clinical workflow of the dental clinic. In this model, an additional user layer is introduced. This layer provides an important interface between the user and knowledge mined from bio medical data sources. Moreover, state-of-the-art approach adopted in the data layer allows the efficient extraction of features. These features relate the similarity of a drug pair in terms of the shared difference in their term frequencies. Experiment results show that this approach performs favorably compared to other existing models.

3 Proposed Method

The aim of this study is to propose an unique approach in supporting the dentist in drug prescription taking into account the drug that patient is currently taking and the drug that the patient is allergic to. In order to advise the dentist if the drug to be prescribed is suitable, a three-tier conceptual model is used. These layers are the knowledge layer, data computing layer and user layer (Figure 1). Essentially, each layer is defined by the task that they are responsible for. At each layer, data is transformed and processed which culminate as an advice to the dentist if the drug is safe for prescription and to suggest an alternative drug.

3.1 Knowledge Layer

Many decision support systems depend on information retrieved from the web for further processing and inference-making to arrive at a decision [5]. In this layer, a taxonomy of drugs will be generated using information gathered from the web with expert domain knowledge. The publicly available website *http://www.drugs.com* contains independent and accurate information on more than 24,000 prescription drugs pertaining to interactions, dosages and other important information for both patients and professionals. It is maintained in collaboration with the US Food and Drug Administration (FDA), which acknowledges that such partnerships "are part of FDAs effort to ensure the public has easy access to reliable, useful information that can help people protect and improve their health"[1]. Each drug is being described from a different perspective to suit both patients and health professionals. These are contained under the heading "Overview" and "Professional" for each drug. Information on side effects are also found under the heading "Side Effects". In this study, term frequencies from each of these descriptions are extracted so that similarity ratio can be computed. Essentially, this layer consists of the term frequencies of each drug within the drug taxonomy $\mathcal{T}$ for the respective drug properties described under the headings "Overview", "Professional" and "Side Effects".

Definition 1. – **[Drug Taxonomy]**
Let $\mathcal{T}$ be the conceptual taxonomy of all drugs. $\mathcal{T}$ consists of the domain of drugs linked by their semantic relations of advantageous and adverse, and is defined as a 3-tuple $\mathcal{T} := \langle \mathbb{D}, \mathbb{R}, \mathcal{H}_{\mathbb{D}}^{\mathbb{R}} \rangle$, where

[1] http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm212844.htm

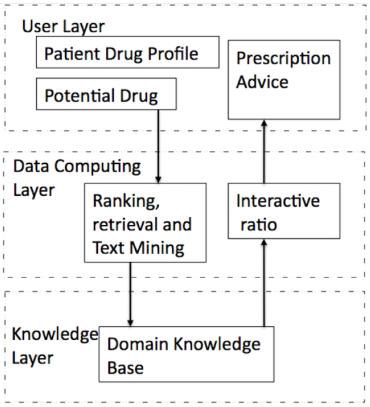

Fig. 1. Three-tier framework

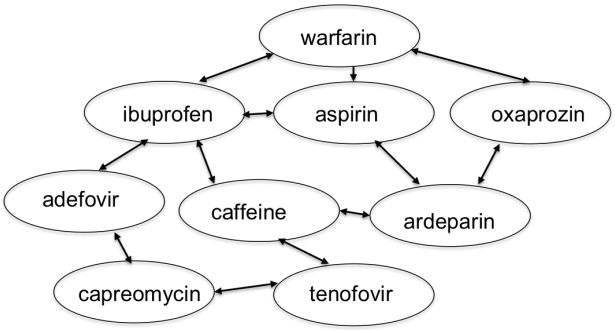

Fig. 2. Subset of Drug Taxonomy

- $\mathbb{D} = \{d_1, d_2, ..., d_{|\mathbb{D}|}\}$ *is the domain set of drugs;*
- $\mathbb{R} = \{r^-, r^0\}$ *is a set of semantic relations, where* $r^-(d_i, d_j)$ *means that the effects of drugs* d_i *and* d_j *are adverse;* $r^0(d_i, d_j)$ *means that the effects of drugs* d_i *and* d_j *are not defined.*
- $\mathcal{H}_{\mathbb{D}}^{\mathbb{R}}$ *is the taxonomical structure constructed by all* $d \in \mathbb{D}$ *linked by* $r \in \mathbb{R}$.

Definition 2. – **[Drug Properties]**
The three properties associated with each drug will have their set of term frequencies. Let D be the set of drugs in drugbank, t_v *be the set of terms in "Overview",* t_p *be the set of terms in "Professional" and* t_s *be the set of terms in "Side Effects"*
Then
term frequency in Overview $tf_v = \{d_i, \{t_{v_1}, t_{v_2}, ...t_{v_n}\}\}$ *where* $d_i \in D$
term frequency in Professional $tf_p = \{d_i, \{t_{p_1}, t_{p_2}, ...t_{p_n}\}\}$ *where* $d_i \in D$
term frequency in Side Effect $tf_s = \{d_i, \{t_{s_1}, t_{s_2}, ...t_{s_n}\}\}$ *where* $d_i \in D$

The ability to store interactive drug pairs within a network of nodes and edges allows the knowledge layer to be represented by a directed acyclic graph (DAG). Each drug is represented as a vertex on the graph. The edges that connect a pair of vertices show the interactions between the drug pair. The cluster of drugs that has adverse interactions with a given drug can be known from the drug taxonomy. Since this project aims to find the similarity between a drug pair based on the term frequencies of each drug, the DAG allows such information to be obtained efficiently by having an algorithm to traverse through the network for the existence of the drug pair. Moreover, features of the drugs stored within each node of the DAG contributes to the Data Layer for mining the properties in terms of their term frequencies (see next section).

Figure 2 shows a subset of the major drug interactions in the drug taxonomy. Note that nodes in the taxonomy are connected to one another through arrows which shows that they have an adverse relationship, as the drug pair has adverse

interaction. Each node on the drug taxonomy will consists of a drug with its associated properties. Such a chain of drug interactions will form the backbone of the drug DAG. As shown in Figure 2, *capreomycin* has adverse interactions with *adefvoir* and *tenofovir*, which in turn interacts respectively with *ibuprofen* and *caffeine*. However, *ibuprofen* and *caffeine* are also in an adverse relationship, which shows that a given drug may interact adversely with more than one drug.

3.2 Mining Properties

In this paper, we are only interested in the content in terms of properties of each drug within the drug taxonomy $\mathcal{T}$. This will enable us to compute the similarity of their term frequencies. Such properties of the drugs are obtained from the content provided under "Professional", "Side Effects" and "Overview". To determine the similarity within a drug pair, information on the term frequencies of each drug are collected.

Given the numerous techniques and algorithms available in data mining [13], the approach adopted in this layer is focused on speed and accuracy with easy interfacing between the knowledge layer and the user layer. Thus, at this layer, the properties associated with the drugs residing on the drug taxonomy backbone are mined, namely the term frequencies tf_v, tf_p and tf_s respectively for "Overview", "Professional" and "Side Effects". Before term frequencies and document frequencies are computed from the properties of the drug, data is preprocessed. This is also a common practice where pre-processing is applied before any mining techniques are used on the text [10] which includes stop word removal and stemming. The idea behind our novel approach is to use the similarity ratio between the feature vectors of term frequencies to decide if the drug pair is similar, and using this knowledge to advise dentist if it is suitable for prescription, taking into consideration the individual medical status of the patient.

Thus the data layer should aim to reduce the content of the relevant web page by excluding stop words and do stemming on the documents.

3.3 User Layer

In order to transform patient profiles and data in a knowledge base into usable and useful knowledge for the dentist, the importance of an user-friendly user layer cannot be over emphasised. A good interface is also crucial in the technology diffusion process to enable high acceptance and absorption rates by the users. In fact, a poorly designed user interface can only reduce the performance and benefits to clinicians [9], resulting in a barrier against system adoption.

Besides providing the range of drugs whose properties are to be retrieved and processed by the underlying layers, the other function of this layer is to present the results back to the dentist on whether the drug in question is safe for prescription. In this system, the data layer essentially consists of the drug profile of the patients and the drug which the dentist is going to prescribe. The user layer also consists of the results after computation of similarity ratio is completed at the data layer. The result acts as a supporting tool to the dentist

User Layer	Data Layer	Knowledge Layer
• Efficient mapping of user requirements	• Efficient choice of programming approach	• Bio medical data sources, drug taxonomy
• User friendly interface	• Implementation of data mining • Algorithm design	• Drug properties

Table 1. Features of Conceptual Framework

for decision support. If the drug to be prescribed is found to adversely interact with the drug that patient is currently taking or is similar to the drug that patient is allergic to, then it is in this layer that an alternative drug is presented to the dentist.

As highlighted in Table 1 for the three layers in this framework, user requirements in the user layer need to be efficiently mapped onto the data layer to enable useful and relevant information to be extracted for further computing of the similarity ratio.

3.4 Feature Extraction

To arrive at a decision as to whether a drug is suitable for prescription, a similarity ratio is required as a measure of how similar they are in terms of their term frequencies. These term frequencies are obtained after performing data mining on each of the properties of the drugs associated with "Professional", "Overview" and "Side Effects". Once the properties of drugs are known, their term frequencies are computed. Feature vectors of *tf*idf* are then constructed based on the term frequencies. By determining the difference between feature vector of a pair of drugs, the similarity ratio can be calculated. If p and q are the feature vectors, then the cosine similarity is given by:

Algorithm 1: Predicting Drug Similarity

input :

1. Let D be the set of drugs in Drugbank and $d_i \in D$
2. Let W_i^c be content of a web page
3. Let t be the set of terms in W_i^c
4. term frequency in Overview $tf_v = \{d_i, \{t_{v_1}, t_{v_2}, ...t_{v_n}\}\}$
5. term frequency in Professional $tf_p = \{d_i, \{t_{p_1}, t_{p_2}, ...t_{p_n}\}\}$
6. term frequency in Side Effects $tf_s = \{d_i, \{t_{s_1}, t_{s_2}, ...t_{s_n}\}\}$

output: true if similarity ratio S above threshold θ, false otherwise

```
while ( for a given drug pair d1, d2) do
    //get term frequency from document
    while t ∈ W_i^c do
        tf_v(d_i, {t}) ← tf_v(d_i, {t}) ∪ t
        tf_p(d_i, {t}) ← tf_p(d_i, {t}) ∪ t
        tf_s(d_i, {t}) ← tf_s(d_i, {t}) ∪ t
    end
    feature vector f_v = {d_1, d_2, .....d_n} where d_i ⊂ D
    compute tf * idf from feature vector fv
    compute similarity ratio S
    if (S > θ) then
        return false;
    end
    return true;
end
```

$$s(p,q) = \frac{\sum_{i=1}^{k} p_i.q_i}{\sqrt{a.b}} \quad where \quad a = \sum_{i=1}^{k} {p_i}^2 \quad and \quad b = \sum_{i=1}^{k} {q_i}^2$$

As seen in Algorithm 1, this cosine similarity s is computed after gathering the features within the properties of each drug. Based on the cosine similarity of each drug pair, average values are computed. With these values as a guide, a threshold value θ is set to maximise the F measure for each of the drug property. Section 5 explains the approach in deciding on the threshold value.

For a given drug pair, the similarity ratio taken from each of the three properties can also be combined and use as a prescription support to check if the drug to be prescribed is similar to the drug that patient is allergic to. In our work, equal weights are used to compute the overall similarity ratio. The average value of θ is also used as a threshold against the overall similarity ratio.
Thus the overall similarity ratio for drug p and drug q is given by:

$sim(p,q) = 0.33 * S_v + 0.33 * S_p + 0.33 * S_s$ where
S_v is the similarity ratio for the drug property from "Overview"
S_p is the similarity ratio for the drug property from "Professional"
S_s is the similarity ratio for the drug property from "Side Effects"

If the similarity ratio exceeds the average threshold value θ, the model will return a false, indicating that the drug to be prescribed is similar to the drug that patient is allergic to.

3.5 Preparing Drug Properties

With the URL links to clusters of interactive drugs readily available, the content of drugs.com website is scanned for information on the cluster of drugs that are interactive with each entry in the drug taxonomy.

Each web page contains field markers to delimit the relevant content on interactive drugs and these markers are used to extract the cluster of drugs that adversely interact with each drug in the dataset. Information on the cluster of adversely interactive drugs is crucial as it provides the ground truth in deciding whether a drug pair is in an adverse relationship. Such information are contained in the drug taxonomy $\mathcal{T}$ within the knowledge layer of the conceptual framework. Besides mining the information related to clusters of drug interactions for each drug in the drug taxonomy, the underlying properties for each drug is also obtained to provide information on the similarity of a drug pair. These properties are term frequencies tf_v, tf_p and tf_s mined respectively from "Overview", "Side Effects" and "Professional" tab of each drug in the drug taxonomy. These knowledge will then be used to compute the cosine similarity of the drug pair - refer to Figure 3 for the flow of the experimental design.

4 Experimental Evaluation

We choose an experimental approach to assess the accuracy and efficiency of the proposed method. It tests the hypothesis that similar drug pairs have a higher similarity ratio compared to those dissimilar pairs.

To evaluate the performance of our model, precision, recall and F measure are adopted in this experiment. Precision indicates how accurately the model predicts drug pairs as similar, while recall indicates how accurately similar drug pairs are predicted.

$$precision = \frac{TP}{TP + FP} \quad recall = \frac{TP}{TP + FN} \tag{1}$$

where TP refers to those non-similar pair that are predicted correctly(True Positive), FN refers to similar pairs predicted incorrectly (False Negative) and FP refers to non similar pairs predicted incorrectly (False Positive).

$$Fmeasure = \frac{2 * precision * recall}{precision + recall} \tag{2}$$

Accuracy is also used to measure the percentage of correct predictions combining both the similar and dissimilar predictions.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

5 Results and Discussion

With the unique three-tier conceptual framework where knowledge is extracted from the knowledge base and delivered to the data layer, the result from this model demonstrates the efficiency and robustness of our model. Not only does the algorithm is able to compute the similarity of the drug pair, such information can be adopted as a decision support to health professional in drug prescription. The prediction is based on the hypothesis that a drug-pair is similar if the cosine similarity ratio between the frequent patterns are high.

By computing the similarity ratio between drug pairs, their average values are obtained as a guide to set the threshold θ in order to maximise the F measure. As shown in Table 2, a range of values for θ are applied for each of the drug properties "Overview", "Professional" and "Side Effects". For example, θ of 0.94 is used as a threshold to compute the recall, precision and F measure for features gathered from the drug property "Professional" as the maximum value of F score occurs at this value. Figure 4 shows the recall, precision and F scores achieved with drug properties gathered from "Overview", "Professional" and "Side Effects".

As indicated in Figure 4, the recall rate of 70% is achieved from drug properties obtained from "Overview", showing that our model performed much better than other methods of prediction. In contrast, the work by [16] achieved 45%

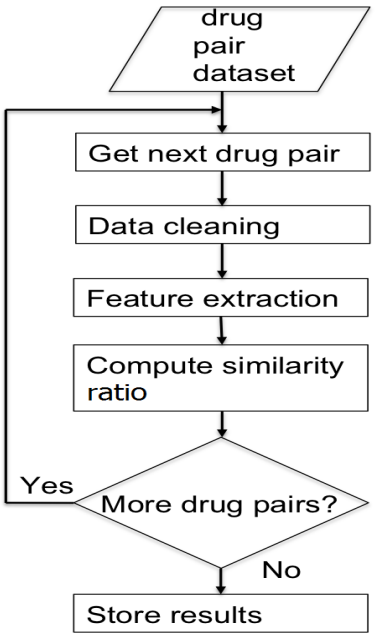

Fig. 3. Experimental Design

θ levels	Overview	Professional	SideEffect
0.80	0.50	0.47	0.50
0.82	0.50	0.43	0.51
0.84	0.51	0.45	0.84
0.86	0.52	0.42	0.84
0.88	0.50	0.46	0.84
0.90	0.50	0.48	0.85
0.92	0.54	0.51	0.87
0.94	0.59	**0.60**	0.87
0.96	**0.63**	0.53	**0.90**
0.98	0.55	0.45	0.85
0.99	0.51	0.40	0.73

Table 2. F measure at different theta

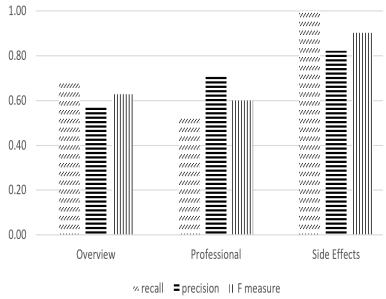

Fig. 4. Performance Comparison against Different Drug Properties

with predictions based on drug metabolism. In terms of accuracy, which measures the percentage of correct predictions combining both the similar and dissimilar predictions, our system comes out at over 80% compared to 69% where drug predictions are based on the relationship between drug targets [18].

Drug currently taking	*Drug Allergy*
Warfarin	Amoxicillin
Drug prescribed	*Drug recommended*
Methotrexate	Cladribine
Acetaminophen	Acetaminophen
Metronidazole	-nill-

Table 3. Sample Result of Recommender

To illustrate the conceptual framework of this study, the same model can be used to decide if the drug is suitable for prescription. Based on the overall similarity (as explained in Section 3.4) from the three properties of the drug pair, the system can detect if the drug is similar to the drug that patient has allergy. This approach highlights the usefulness of our framework where knowledge generated from the data layer can be applied to the user layer and becomes useful to the user, in this case, as a decision support tool to the health professional. This novel

way allows us to support the dentist with the right prescription, ensuring the drug is not in adverse relationship with the drug patient is taking while at the same time it is not similar to the drug that patient is allergic to, which is the aim of this study.

For example, for a patient with heart problem consuming Warfarin and is allergic to Amoxicillin, our model is able to recommend an alternative drug. In other words, the drug should not be in adverse interaction with Warfarin and should also not be similar to Amoxicillin. From Table 3, as Methotrexate is not safe, our model suggests Cladribine as an alternative. However, since Acetaminophen is safe, nothing new is suggested and there is also the scenario where the model is unable to recommend any drug from the existing knowledge base.

6 Conclusions

This paper has presented a novel approach in advising the suitability of a drug prescription by predicting the similarity of a drug pair. This is different from conventional prediction of drug interactions. We have applied our model by integrating the prediction of the drug interaction with the personal medical status of the patient by taking into consideration the drug that the patient is taking to avoid adverse interaction and the drug that patient has allergy to avoid cross allergy. It has also been demonstrated in this research through the experiment that the three-tier approach adopted in our research design performs well and thus can readily be implemented within the clinical workflow of a dental clinic.

Acknowledgments

This research is partially supported by Glory Dental Surgery Pte Ltd, Singapore (*http://glory.sg*) and undertaken collaboratively with their panel of dentists.

References

1. Bokharaeian B, Diaz A, Chitsaz H (2016) Enhancing Extraction of Drug-Drug Interaction from Literature Using Neutral Candidates, Negation, and Clause Dependency. PLoS ONE 11(10):1-20
2. Brown AS, Patel CJ (2016) Meshdd: Literature-based drug-drug similarity for drug repositioning. Journal of the American Medical Informatics Association 27:1-5, DOI 10.1093/jamia/ocw142
3. Bui Q, Sloot P, vanMulligen E, Kors J (2014) A novel feature-based approach to extract drug-drug interactions from biomedical text. BioInformatics 30(23):3365-3371
4. Casillas A, Prez A, Oronoz M, Gojenola K, Santiso S (2016) Learning to extract adverse drug reaction events from electronic health records in spanish. Expert Systems with Applications 61:235 - 245
5. Domingues M, Jorge A, Soares C, Rezende S (2015) Web mining for the integration of data mining with business intelligence in web-based decision support systems. In: Azevedo A, Santos M (eds) Integration of Data Mining in Business Intelligence Systems, IGI Global, Hershey, USA, pp 120-142

6. Goh WP, Tao X, Zhang J, Yong J (2016) Decision support systems for adoption in dental clinics: A survey. Knowledge-Based Systems 104:195 - 206
7. Han J, Kamber M, Pei J (2011) Data Mining Concepts and Techniques. Morgan Kaufmann Publishers
8. Holzinger A, Jurisica I (2014) Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions. In: Holzinger A, Jurisica I (eds) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics, Springer-Verlag, Chap 1, pp 1-18
9. Horsky J, Phansalkar S, Desai A, Bell D, Middleton B (2013) Design of decision support interventions for medication prescribing. International Journal of Medical Informatics 82(6):492-503, DOI http://dx.doi.org/10.1016/j.ijmedinf.2013.02.003
10. Kriegel HP, Ntoutsi E (2014) Clustering high dimensional data: Examining differences and commonalities between subspace clustering and text clustering - a position paper. SIGKDD Explor News 15(2):1-8
11. Li Ao, Zang Qiguang,Sun Dongdong, Wang Minghui (2016) A text feature-based approach for literature mining of lncRNA-protein interactions. Neurocomputing 206:73-80
12. Lieber Nicolina, Ribeiro Eliane (2012) Adverse drug reactions leading children to the Emergency Department. Revista Brasileira de Epidemiologia, 15:265-274
13. Lu J, Hales A, Rew D, Keech M, Frohlingsdorf C, Mills-Mullett A, Wette C (2015) Data mining techniques in health informatics: A case study from breast cancer research. In: Renda EM, Bursa M, Holzinger A, Khuri S (eds) Information Technology in Bio- and Medical Informatics: 6th International Conference, ITBAM 2015, Valencia, Spain, September 3-4, 2015, Proceedings, Springer International Publishing, Cham, pp 56-70
14. Mishra SN, Jaiswal A, Ambhaikar A (2012) An effective algorithm for web mining based on topic sensitive link analysis. International Journal of Advanced Research in Computer Science and Software Engineering 2(4):278-282
15. Mukherjee S, Varde AS, Javidi G, Sheyban E (2014) Predictive analysis of engine health for decision support. SIGKDD Explor Newsl 15(2):39-49, DOI 10.1145/2641190.2641197, URL http://doi.acm.org/10.1145/2641190.264119720
16. Tari L, Anwar S, Liang S, Cai J, Baral C (2010) Discovering drugdrug interactions: a text-mining and reasoning approach based on properties of drug metabolism. Bioinformatics 26(18):547
17. Wu HY, Chiang CW, Li L (2014) Text mining for drug-drug interaction. Methods in molecular biology 1159:47 - 75
18. Yan S, Jiang X, Chen Y (2013) Text mining driven drug-drug interaction detection. 2013 IEEE International Conference on Bioinformatics and Biomedicine pp 349-354