Indian medicinal plants for diabetes: text data mining the literature of different electronic databases for future therapeutics.

Bhanumathi Selvaraj[1*], Sakthivel Periyasamy[2]

[1]Department of Computer Science and Engineering, Faculty of Computing, Sathyabama University, Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai, India

[2]Department of Electronics and Communication Engineering, Anna University, Guindy, Chennai, India

Abstract

Diabetes, a metabolic disorder, affects nearly 7% of world population and predicted that it would be the seventh leading cause of death by the year of 2030. The prevalence and morbidity of diabetes are increasing rapidly because of the lifestyle and diet changes occurring with urbanization. Medicinal plants and their derivatives have been proven to be an effective and safe therapy offering various benefits, for example, the moderate reduction in hypoglycaemia, in the treatment and prevention of diabetes. However, the identification of such valuable Indian medicinal plants for diabetes from biomedical literature is not comprehensively explored. In this study, we have investigated Indian medicinal plants for diabetes in the biomedical literature using text data mining technique. We discovered a total of 203 Indian medicinal plants for diabetes in 355 articles out of 15651 articles of text corpus in the dataset. In addition, we analysed the importance of Indian medicinal plants for the treatment of diabetes by means of the frequency of 203 plants in 355 articles, which identified 22 anti-diabetic Indian medicinal plants that showed ≥ 9 frequencies. *Momordica charantia*, also known as bitter melon, had the highest frequency (≥ 51 frequencies) among 203 Indian plants, indicating that it is the most important Indian medicinal plant for the treatment of diabetes. In addition, we compared the identified 203 plants with previously reported database of anti-diabetic Indian medicinal plants, which showed the identification of 100 new anti-diabetic Indian medicinal plants. The results from this study could provide helpful information for future experimental and clinical studies, and the development of future therapeutic for diabetes.

Introduction

Diabetes, also known as diabetes mellitus, is a progressive metabolic disorder, which is characterized by elevated levels of glucose or sugar in the blood. The disease is caused by inadequate secretion of insulin, body's poor response to insulin, or both, which severely damage the body's systems, including the heart, blood vessels, eyes, kidneys, and nerves. Type 2 diabetes is the most common in adults that occurs when the body becomes resistant to insulin or does not secrete adequate insulin while Type 1 diabetes is a juvenile or insulin-dependent that occurs when the pancreas produces little or no insulin by itself. Diabetes is rapidly growing worldwide, and affected 422 million people in 2014 and resulted in over 3 million deaths [1]. The World Health Organization (WHO) estimated that diabetes would be the seventh leading cause of death by the year of 2030, and suggested that healthy lifestyle and right medication and regular screening can prevent and avoid the consequence of diabetes, respectively [1].

For many decades, medicinal plants have been beneficial resources for the treatment of several diseases, including diabetes [2-5]. Some well-known drugs in current-use for diabetes have been developed from plants such as metformin drug derived from the *Galega officinalis* [6]. Many studies have also indicated the advantages of medicinal plants in the therapeutic development, for example availability and acceptable risk-benefit ratio. Though the ethnobotanical community has reported a list of anti-diabetic medicinal plants [7], in the search for new treatments and cures, yet more medicinal plants are being explored for their therapeutic development. At this juncture, efforts have been also made to compile information on the medicinal plants to exclusively develop databases for diabetes. For example, InDiaMed [8] and DIAB [9] databases which have been developed using manual curation from various sources i.e. PubMed [10], Scopus [11], Science Direct [12] and Wiley [13], and also from folklore medicinal usage. However, limited efforts have gone to comprehensively gather Indian medicinal plants for diabetes.

Text data mining is an approach, which is often used to extract and analyse and/or evaluate information in the area of biomedical discovery [14]. It is capable of producing significant results that would be helpful to answer particular research queries, for example, finding medicinal plants in biomedical literature for experimental and clinical research [15]. The text data mining methodology has been successfully employed to discover and analyse novel herbs or plants or formulas, from the traditional Chinese medicine or historical literature or electronic databases, for different diseases such as vascular dementia [16], dysmenorrhea [17], chronic cough [18], respiratory disease [19], diabetic nephropathy [20], age-related dementia and memory impairment [21]. However, there is no such a study on Indian medicinal plants for diabetes in electronic literature databases.

In this study, we have analysed the Indian medicinal plants for diabetes from different electronic literature databases using text data mining approach. Our text data mining strategy successfully extracted 355 unique articles, which contain only Indian medicinal plants for diabetes, from the 15651 articles of the text corpus, which resulted in the identification of 203 Indian medicinal plants for diabetes. In addition, the frequencies of plants were analysed from the 355 articles, and it was found that 22 Indian medicinal plants were ≥ 9 frequencies, emphasizing their importance for the treatment of diabetes. In addition to this, the identified 203 plants were compared to the exclusive database previously developed for the Indian medicinal plants for diabetes. The results showed that the identification of several new Indian medicinal plants for diabetes. The overall results of the study could provide fruitful information for future experimental and clinical studies, and the development of future therapeutic for diabetes.

Materials and Methods

The biomedical research has produced the abundance of knowledge in the published literature, which stress the need of novel methods for knowledge extraction, visualization, and analysis to uncover new and meaningful hypotheses. A text mining technique, a subfield of data mining, that seeks to extract useful new information from unstructured or semi-structured sources to address the most crucial questions. This mining approach is designed and implemented extensively to extract and analyse comprehensive information in the biomedical literature of diseases, natural products, herbs, etc.

In this study, the text data mining approach was adopted for analysing Indian medicinal plants for diabetes in the biomedical literatures, and our mining strategy comprises the following steps: data collection, data synthesis, data extraction and data pre-processing, and data analysis, as schematically shown in Figure 1. These steps have been explained below.

Data collection

Literatures were searched with different electronic databases including PubMed/MEDLINE [10], Scopus [11], Science Direct (SciDir) [12] and Wiley [13] using the following

different search query terms: Indian medicinal plants for diabetes, Indian traditional medicinal plants for diabetes, treatment of diabetes with Indian medicinal plants, treatment of diabetes with Indian medicinal herbs, Indian herbal plants for diabetes, Indian herbs for anti-diabetic, etc. These query terms were derived from the different biomedical literature and manuals. Each search query term was typed in the search box of the electronic databases to separately retrieve information relevant to the Indian medicinal plants for diabetes. Only journal articles were considered and retrieved. Different search query terms to the electronic databases produced search results that were subsequently combined to form a single text corpus for each database.

Data synthesis

Data synthesis is a process of collecting essential data from the different sources and this is an important step in the text mining. In the above collected text corpus, each article contains much information such as article title, abstract, journal name, author's name, URL or PubMed ID, DOI, etc. Articles, which only contain both the title and abstract, were considered and all other information were removed as our aim was to specifically focus on the extraction of Indian medicinal plants. As several search query terms were used to retrieve the same information from the databases, there is a possibility of duplicate entries in the text corpus. The duplicate entries were therefore removed using Microsoft Excel as described in [22]. In addition, the titles with no abstract articles were also excluded. Finally, the title and abstract of the each article in the text corpus were split into separate text documents for the text data mining process.
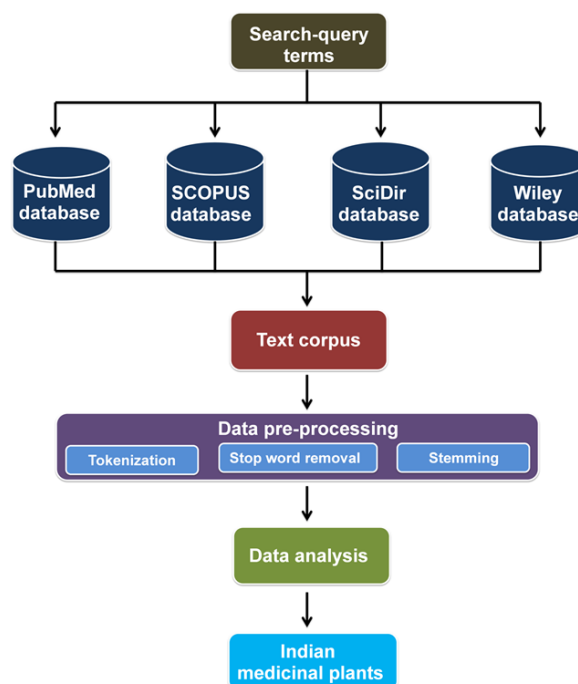

Figure 1. *The work flow of text data mining used in this study.*

Indian medicinal plants for diabetes: text data mining the literature of different electronic databases for future therapeutics

Data extraction and pre-processing

Data extraction: Data extraction is a process of retrieving relevant data from unstructured data sources for further data processing. Documents that are relevant to Indian medicinal plants for diabetes were extracted from the corpus, which contains several text documents with title and abstract. This was performed using all possible keywords such as Indian, medicinal, diabetes, anti-diabetic, diabetic, etc.

Data pre-processing: This is an important step in the text data mining, which is used to exclude unwanted information in the corpus. It provides more relevant information during the data analysis process. Before performing the following steps of the pre-processing, the upper case alphabets in the each document were converted into lower case alphabets [23].

Tokenization: This is a process of breaking a stream of text into words or terms. Title with its abstract of each article's document in the corpus was split into separate tokens which are useful to identify punctuation marks, special symbols, numbers, etc.

Stop word removal: Stop words are most frequently occurred words and do not have any significant meaning in the document. Therefore, they were excluded to reduce the dimensionality of the text documents. The punctuation marks, numeric values and special symbols were removed from the tokenized documents. In addition, the stop words such as at, in, to, we, the, etc., were collected and stored as a list. This list of stop words was matched with the tokens one by one and then matched tokens were excluded from the text document.

Stemming: This is also used to further reduce the dimensionality of the text documents by identifying and removing the words with common roots, for example, stemmer, stemmed, stems, stemming, which are based on the root word 'stem'. For this, the most widely used Porter's stemming algorithm [24,25] was employed.

Data analysis

It involves the extraction of actual information, validation and analysis of them for the identification of useful knowledge. Here, the Indian medicinal plants for diabetes were extracted, validated with their synonyms, frequency count, and analysed with the existing database [8]. For the data synthesis and data extraction and pre-processing, Python 2.7.10 [26] and Natural Language Tool Kit (NLTK) [27] were used.

Results

Information retrieval and extraction for Indian medicinal plants for diabetes

Different search query terms for Indian medicinal plants for diabetes yielded a total of 15651 articles, which covered articles published until April 2016. PubMed, Scopus, Science Direct and Wiley were considered in this study. The search queries on these databases resulted in the text corpus of 7157 in PubMed, 2573 in Scopus, 3745 in Science Direct, and 2186

articles in Wiley. As our main aim was to discover new Indian medicinal plants for diabetes and to understand which Indian plants have been frequently used for diabetes from the electronic databases, the title and abstract of each article were considered. Articles that contain only title were removed. The duplicate articles were deleted from each of the text corpus, which produced 3644 in PubMed, 630 in Scopus, 1493 in Science Direct and 643 articles in Wiley. Since the specific focus was to discover Indian medicinal plants for diabetes, Indian or India relevant articles were extracted from the each of the text corpus, which left 1675 in PubMed, 547 in Scopus, 710 in Science Direct and 94 articles in Wiley. Subsequently, diabetes relevant articles were also extracted using terms like 'diabetes', 'diabetic', 'anti-diabetic', etc. The outcome of this extraction was 251, 282, 63 and 33 articles in PubMed, Scopus, Science Direct and Wiley, respectively. After performing the pre-processing steps of the text data mining such as tokenization, stop word removal, stemming on the text corpus, and articles of the Indian medicinal plant for diabetes were obtained: 180 in PubMed, 206 in Scopus, 40 in Science Direct, 28 in Wiley. These articles were combined and duplicates were removed, which resulted in a total of 355 unique Indian medicinal plant articles were used for data analysis. A schematic representation of the text data mining of Indian medicinal plants for diabetes is shown in Figure 2.
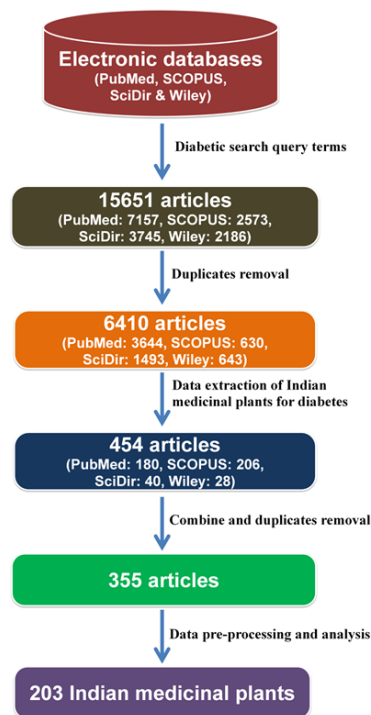

Figure 2. *Schematic representation of text data mining strategy adopted in this study.*

Indian medicinal plants for diabetes

The text data mining process identified 355 unique articles, which showed Indian medicinal plants for diabetes, from the 15651 articles obtained from different electronic databases including PubMed, Science Direct, Scopus and Wiley. 203 unique Indian medicinal plants for diabetes were successfully

identified from the 355 unique articles, which are shown in Supplementary Table 1. It is observed that multiple plants were found in some articles but in most of the articles, a single plant was found. In addition, although most of the articles dealt with the Indian medicinal plants for diabetes, few articles rarely dealt with the reviews of diabetes of Indian medicinal plants. However, such articles should be included in the identification of Indian medicinal plants for diabetes as they would be helpful for the contemporary conceptions of diabetes and interpretation of the data for the same.

Most frequently used Indian medicinal plants for diabetes

Since several plants have been reported for the treatment of diabetes, the identification of frequently used Indian medicinal plants for this metabolic disorder would be useful in short listing plants for further clinical research. The frequencies of Indian medicinal plants were therefore computed by counting plants in each of the 355 unique articles, and the results are shown in Supplementary Table 1. Top 22 repeatedly used Indian medicinal plants were identified (Table 1), which showed ≥ 9 frequencies (Figure 3). Of the top 22 ranked medicinal plants, the first ten plants had more than 19 frequencies, indicating their importance for the treatment of diabetes. The five most frequent Indian medicinal plants (Table 1) are *Momordica charantia* (Bitter Gourd), *Syzygium cumini* (Jamun), *Gymnema sylvestre* (Gurmar), *Trigonella foenum-graecum* (Fenugreek) and *Tinospora cordifolia* (Guduchi), which are used for the type 2 diabetes mellitus, which all had $\geq$ 28 frequencies in 355 unique articles. In addition, some of the

lowest frequent Indian medicinal plants (Table 1) are *Aloe vera* (*Aloe vera*), *Swertia chirata* (Chirayita), *Mangifera indica* (Mango) and *Ficus racemosa* (Cluster fig), indicating that they have been also explored for the treatment of diabetes but not effectively.
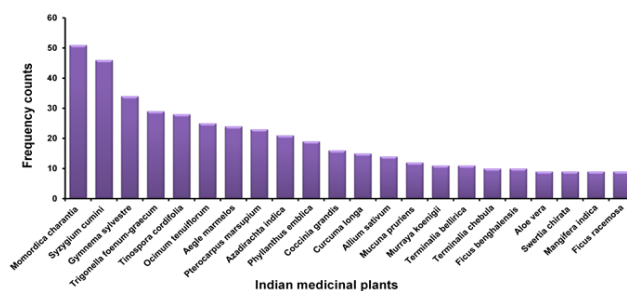

Figure 3. *The 22 most frequent Indian medicinal plants for diabetes and their frequencies.*

Comparison with previous study

Anti-diabetic Indian medicinal plants identified using text data mining in this study were compared to the previously reported database. IndiaMed is a comprehensive database of Indian medicinal plants for diabetes, which has been developed by the manual curation of literature and by a collection of folklore medicinal usage [8]. A list of all Indian medicinal plants was collected from the database and compared with the 203 Indian medicinal plants, which showed that the identification of 100 new anti-diabetic Indian medicinal plants. The new plants identified in this study are highlighted in the yellow color in the Supplementary Table 1.

Table 1. *Frequently used 22 Indian medicinal plants for the treatment of diabetes from the biomedical literature, and their frequency, common name, Indian names and family.*

S. no.	Scientific name	Frequency	Common name	Other Indian names	Family
1	*Momordica charantia*	51	Bitter Gourd, Bitter melon	Karela, Ambalem, Karali, Kareti, Changkha	Cucurbitaceae
2	*Syzygium cumini*	46	Jamun, Java plum	Jam, Nagai	Myrtaceae
3	*Gymnema sylvestre*	34	Gurmar	Gudmar, Medhashingi, Kavali, Bedaki, Chakkarakkolli	Asclepiadaceae
4	*Trigonella foenum-graecum*	29	Fenugreek, Greek-clover, Greek hay	Methi, Mente, Mentepalle, Mentesoffu, Uluva	Fabaceae
5	*Tinospora cordifolia*	28	Guduchi, Gulbel, Indian tinospora	Giloy, Gulancha, Gulbel, Ningthoukhongli	Menispermaceae
6	*Ocimum tenuiflorum*	25	Holy basil	Tulsi, Trittavu, Tulshi	Lamiaceae
7	*Aegle marmelos*	24	Bel, Beli fruit, Bengal quince, Stone apple, Wood apple	Heirikhagok, Maredu, Vilvam, Vilvam, Sandiliyamu	Rutaceae
8	*Pterocarpus marsupium*	23	Indian kino tree, Malabar kino tree	Vijayasara, Bijasal, Bila, baengamara, bange, bendaga	Fabaceae
9	*Azadirachta indica*	21	Neem	Nimbay,Veppai, Sengumaru, Ariyaveppu, Vepa	Meliaceae
10	*Phyllanthus emblica*	19	Amla, Indian gooseberry	Aonla, Nellikka, Usirikaya, Betta nelli, Amalaka, Aonla	Phyllanthaceae

11	*Coccinia grandis*	16	Ivy gourd	Kunduru, Tondli, Kovai, Tondikay, Telakucha, Bimbika	Cucurbitaceae
12	*Curcuma longa*	15	Turmeric	Halodhi, Halud, Haldar, Arishina, Arisina, Manjal	Zingiberaceae
13	*Allium sativum*	14	Cultivated garlic	Naharu, Lahsun, Lahsan, Lassan, Belluli, Vellulli	Amaryllidaceae
14	*Mucuna pruriens*	12	Velvet bean, Cowitch, Cowhage, Nescafe	Kiwach, Khaj-kuiri, Naicorna, Kauso, Pilliadugu	Fabaceae
15	*Murraya koenigii*	11	Curry leafs-tree	Kari patta, Kudianim, Karivepillai, Kareapela, Girinimba	Rutaceae
16	*Terminalia bellirica*	11	Baheda, Belliric Myrobalan, Bastard	Akshah, Bahuvirya, Bibhitakah, Karshah, Vibhitakah, Barro	Combretaceae
17	*Terminalia chebula*	10	Chebulic Myrobalan, Myrobalan	Harra, Manahi, Hirad, kaDukkaay, Katukka, Nallakaraka	Combretaceae
18	*Ficus benghalensis*	10	Indian banyan tree	Barh, Khongnangtaru, Bargad, Vat, Alai, Marri chettu	Moraceae
19	*Aloe vera*	9	Aloe vera, aloe, Burn plant	Gheekumari, Khorpad, Kathalai, Chotthukathalai	Asphodelaceae
20	*Swertia chirata*	9	Chirayita	Chirayata, Chirata, Charayatah, Nilavembu, Shirattakuchi	Gentianaceae
21	*Mangifera indica*	9	Mango	Am, Heinou, Ma, Mamidi, Mangga, Mavinamara, Amba	Anacardiaceae
22	*Ficus racemosa*	9	Cluster fig	Goolar, Heibong, Paidi, Udumbara, Umber, Atti	Moraceae

Discussion

Several studies have previously reported the use of medicinal plants for the treatment of different ailments including diabetes, wherein they have indicated that the medicinal plants are served as precursors for the development of drugs and have the acceptable pharmacological properties. Text data mining has been a valuable tool for various applications in the areas of molecular biology, toxicogenomics, and medicine. For example, the technique has been widely used to correctly identify medicinally important plants for numerous diseases in the biomedical literature, i.e. vascular dementia [16], dysmenorrhea [17], chronic cough [18], diabetic nephropathy [20] and age-related dementia and memory impairment [21]. In this study, we report the systematic analysis of Indian medicinal plants for diabetes in the biomedical literature using text data mining technique. Articles that only mentioned Indian medicinal plants or herbs for the treatment and prevention of diabetes were included in the dataset, and articles that have no abstracts were removed from the dataset. Our findings suggest that medicinal plants are important for the treatment of diabetes and many plants from India have been reported for the disease.

We successfully identified 203 Indian medicinal plants for diabetes from 355 articles which were extracted from 15631 articles using text data mining approach. The results of identified 203 Indian medicinal plants and frequently used 22 plants are shown in Supplementary Table 1 and Table 1, respectively. We have evaluated the frequency of Indian medicinal plants for diabetes to understand which specific plants are mostly preferred among Indian medicinal plants for diabetes. It was found that six plants (*Momordica charantia*, *Syzygium cumini*, *Gymnema sylvestre*, *Trigonella foenum-graecum*, *Tinospora cordifolia* and *Ocimum tenuiflorum*) highly explored for the metabolic disorder ($\geq$ 25 frequencies). This result suggests that these plants are important for the treatment and prevention of diabetes.

Momordica charantia (*M. charantia*), also known as bitter melon, bitter gourd, bitter squash, or balsam-pear, which belongs to Curcubitaceae family, had the highest frequency ($\geq$ 51 frequencies) among any other Indian medicinal plants reported for diabetes. This result is in line with previous studies that *M. charantia*is most widely used and popular plant for its anti-diabetic properties in India [28,29]. Although *M. charantia* has been a versatile plant for the treatment of many diseases, it has been particularly explored extensively for its anti-diabetic properties [30]. Fruits, leaves or seeds of bitter melon possess more than 225 medicinal constituents [31], which may exert their medicinal effects either separately or together. However, charantin, insulin-like peptide and alkaloid and steroidal glycoside, from the extracts of fruits of bitter melon have been shown to possess anti-diabetic and hypoglycaemic activity [32,33], and the mechanism of action of these chemicals for the treatment and prevention of diabetes are unclear.

The 203 anti-diabetic Indian medicinal plants identified in this study (Supplementary Table 1) were compared to known Indian medicinal plants for diabetes in InDiaMed database [8]. This database contains plants that are compiled based on the

manual curation of the literature and a collection of folklore medicinal usage. In addition, since the database is used a mix of scientific, synonyms and common names of plants, they were verified with Wikipedia [34] and flowers of India [35]. The comparison analysis with InDiaMed showed the identification of 100 new Indian medicinal plants (Supplementary Table 1) and 103 plants were common with the database. The analysis further indicates that the employed text data mining strategy is capable of identifying new Indian medicinal plants for diabetes in biomedical literature.

Conclusions

Diabetes has become the most common disorder in both developing and developed countries, and the metabolic disorder is rapidly widespread in most parts of the world. In addition, the WHO has predicted that diabetes would be the seventh leading cause of death by the year of 2030 [1]. Recently, it has also been reported that around 30% of diabetic patients use some form of complementary and alternative medicine [30,36]. Complementary and alternative medicines are the use of medicinal plants and other dietary supplements, which are used as alternatives to mainstream Western medical treatment. Various medicinal plants have been explored for the treatment and control of diabetes. However, there is still room for the identification of a list of medicinal plants for the treatment and prevention of diabetes from the biomedical literature. To address this, we have identified new Indian medicinal plants from biomedical literature, obtained from the different electronic databases, using text data mining techniques. The text data mining techniques resulted in the identification of 203 anti-diabetic Indian medicinal plants and 22 of which had ≥ 9 frequencies. Of 203 anti-diabetic plants, *M. charantia* had the highest frequency (≥ 51 frequencies), indicating its anti-diabetic potentials and it also suggests for conducting clinical trials to further clarify its therapeutic benefits. Moreover, the comparison analysis indicated that the identification of 100 new anti-diabetic Indian medicinal plants which have not been reported, to the best of our knowledge, in InDiaMed, an exclusive database of Indian medicinal plants for diabetes.

This study shows that the adopted text data mining strategy is capable of extracting Indian medicinal plants for diabetes from biomedical literature. Therefore, the presented text data mining approach may be used in the identification of medicinal plants for other diseases. Moreover, it could be expected that identified plants in the study may be useful for future experimental and clinical studies, and the development of future therapeutic for diabetes.

Acknowledgements

References

1. http://www.who.int/mediacentre/factsheets/fs312/en/
2. Grover JK, Yadav S, Vats V. Medicinal plants of India with anti-diabetic potential. J Ethnopharmacol 2002; 81: 81-100.
3. Modak M, Dixit P, Londhe J, Ghaskadbi S, Devasagayam TP. Indian herbs and herbal drugs used for the treatment of diabetes. J Clin Biochem Nutr 2007; 40: 163-173.
4. Rizvi SI, Mishra N. Traditional Indian medicines used for the management of diabetes mellitus. J Diabetes Res 2013; 2013: 712092.
5. Wang Z, Wang J, Chan P. Treating type 2 diabetes mellitus with traditional chinese and Indian medicinal herbs. Evid Based Complement Alternat Med 2013; 2013: 343594.
6. Oubre AY, Carlson TJ, King SR, Reaven GM. From plant to patient: an ethnomedical approach to the identification of new drugs for the treatment of NIDDM. Diabetologia 1997; 40: 614-617.
7. Alarcon-Aguilara FJ, Roman-Ramos R, Perez-Gutierrez S, Aguilar-Contreras A, Contreras-Weber CC. Study of the anti-hyperglycemic effect of plants used as antidiabetics. J Ethnopharmacol 1998; 61: 101-110.
8. Tota K, Rayabarapu N, Moosa S, Talla V, Bhyravbhatla B. InDiaMed: A Comprehensive Database of Indian Medicinal plants for Diabetes. Bioinformation 2013; 9: 378-380.
9. Arulrayan N, Rangasamy S, James E, Pitchai D. A database for medicinal plants used in the treatment of diabetes and its secondary complications. Bioinformation 2007; 2: 22-23.
10. http://www.ncbi.nlm.nih.gov/pubmed
11. https://www.scopus.com/
12. http://www.sciencedirect.com
13. http://www3.interscience.wiley.com/cgibin/home
14. Gonzalez GH, Tahsin T, Goodale BC, Greene AC, Greene CS. Recent advances and emerging applications in text and data mining for biomedical discovery. Brief Bioinform 2016; 17: 33-42.
15. May BH, Lu C, Lu Y, Zhang AL, Xue CC. Chinese herbs for memory disorders: a review and systematic analysis of classical herbal literature. J Acupunct Meridian Stud 2013; 6: 2-11.
16. Feng S, Ren Y, Fan S, Wang M, Sun T. Discovery of acupoints and combinations with potential to treat vascular dementia: a data mining analysis. Evid Based Complement Alternat Med 2015; 2015: 310591.
17. Yu S, Yang J, Yang M, Gao Y, Chen J, Ren Y, Zhang L, Chen L, Liang F, Hu Y. Application of acupoints and meridians for the treatment of primary dysmenorrhea: a data mining-based literature study. EvidBased Complement Alternat Med 2015; 2015.
18. Shergis JL, Wu L, May BH, Zhang AL, Guo X. Natural products for chronic cough: Text mining the East Asian historical literature for future therapeutics. Chron Respir Dis 2015; 12: 204-211.

19. Fu XJ, Song XX, Wei LB, Wang ZG. Study of the distribution patterns of the constituent herbs in classical Chinese medicine prescriptions treating respiratory disease by data mining methods. Chin J Integr Med 2013; 19: 621-628.

20. Zhang L, Li Y, Guo X, May BH, Xue CC. Text mining of the classical medical literature for medicines that show potential in diabetic nephropathy. Evid Based Complement Alternat Med 2014; 2014: 189125.

21. May BH, Lu C, Bennett L, Hugel HM, Xue CC. Evaluating the traditional Chinese literature for herbal formulae and individual herbs used for age-related dementia and memory impairment. Biogerontology 2012; 13: 299-312.

22. May BH, Zhang A, Lu Y, Lu C, Xue CC. The systematic assessment of traditional evidence from the premodern chinese medical literature: a text-mining approach. J Altern Complement Med 2014; 20: 937-942.

23. Bhanumathi S, Sakthivel P. A framework of protein-drug association for Malaria by text data mining of biomedical literature. RJPBCS 2016; 7: 1493-1499.

24. Porter MF. Snowball: A Language for Stemming Algorithms 2001.

25. Porter MF. An Algorithm for Suffix Stripping. Program 1980; 14: 130-137.

26. https://www.python.org/downloads/release/python-2710/

27. http://www.nltk.org/

28. Paul A, Raychaudhuri SS. Medicinal uses and molecular identification of two Momordica charantia varieties - a review. E J Bio 2010; 6: 43-51.

29. Joseph B, Jini D. Antidiabetic effects of Momordica charantia (bitter melon) and its medicinal potency. Asian Pac J Trop Dis 2013; 3: 93-102.

30. Leung L, Birtwhistle R, Kotecha J, Hannah S, Cuthbertson S. Anti-diabetic and hypoglycaemic effects of Momordica charantia (bitter melon): a mini review. Br J Nutr 2009; 102: 1703-1708.

31. Taylor L. Herbal secrets of the rainforest: Bitter melon (Momordica charantia). USA Sage Press (2nd edn.) 2002; 1-100.

32. Pitiphanpong J, Chitprasert S, Goto M, Jiratchariyakul W, Sasaki M, Shotipruk A. New approach for extraction of charantin from Momordica charantia with pressurized liquid extraction. Sep Purif Technol 2007; 52: 416-422.

33. Harinantenaina L, Tanaka M, Takaoka S, Oda M, Mogami O. Momordica charantia constituents and antidiabetic screening of the isolated major compounds. Chem Pharm Bull (Tokyo) 2006; 54: 1017-1021.

34. Melloni GE, Ogier AG, de Pretis S, Mazzarella L, Pelizzola M. DOTS-Finder: a comprehensive tool for assessing driver genes in cancer genomes. Genome Med 2014; 6: 44.

35. http://www.flowersofindia.net/botanical.html

36. Ryan EA, Pick ME, Marceau C. Use of alternative medicines in diabetes mellitus. Diabet Med 2001; 18: 242-245.

*Correspondence to

Bhanumathi Selvaraj

Department of Computer Science and Engineering

Faculty of Computing

Sathyabama University

India