

MINING BIOMEDICAL LITERATURE TO DISCOVER NATURAL CURE FOR RECURRENT DISEASE

Farhi Marir¹, Hussein Fakhry¹ and Aida J. Azar²

¹College of Technological Innovation, Zayed University,
Academic City, Dubai, UAE

²College of Medicine, Mohammed Bin Rashid University of
Medicine and Health Sciences (MBRU), Dubai Health Care City,
Dubai, United Arab Emirates.

ABSTRACT

The advances in digital data collection and storage technology allows the storage of a huge amount of medical publications in MEDLINE. This database contains more than 25 million references to journal articles and abstracts in life sciences and biomedicine. This research work builds on Swanson use of mathematical association between A and C concepts/terms through a list of B concept/terms retrieved from large medical literature databases that contain either A&B or B&C terms links A to C. Swanson discovered evidence that fish oil (A) may cure vessel blood disorder (C) and that magnesium (A) may be effective against migraine headache (C), which were clinically proven two years later. We present a cooccurrence mining algorithm and an A&C pre-defined domain Knowledge Base (containing for instance Garlic Composition and Blood pressure causes) to filter and reduce the exponential number of shared B terms retrieved from MEDLINE articles using Swanson's Arrowsmith machine. The reduced number of relevant B terms makes it easier to build scientific evidence to validate publicly known remedies for recurrent diseases for instance establishing whether an important association exists between garlic and its impact on blood pressure.

KEYWORDS

Co-occurrence Text Mining, ABC Arrowsmith Discovery Machine, Dietary Aliments & Disease Knowledge Base, and MEDLINE medical database.

1. INTRODUCTION

During the eighties, Swanson [1] developed a new bibliographic-based approach that associates medical literature and articles from MEDLINE database for creating new scientific knowledge. Based on the mathematical associative relationship, he stated that if one publication states a relationship between two phenomena A and B while another publication reports on the relationship between B and C phenomena a number $B_1, B_2, B_3...B_n$ term connections can be made, and new scientific knowledge could be generated. In a decade, he identified seven examples of complementary non interactive structures in the biomedical literature. Figure 1 shows three A to C linkages: Blood viscosity (B_1), Platelet Aggregation (B_2), and Vascular Reactivity (B_3). The inference that fish oil may benefit Reynaud's disease was clinically proven two years later [2] In a similar way, he discovered several associative relationships like magnesium deficiency and the occurrence of migraine, which was also clinically validated [3]. These scientific knowledge discoveries are generated using Arrowsmith computer-aided tool [4]

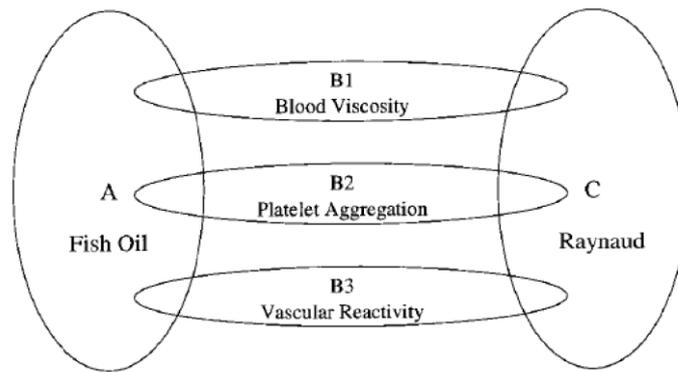


Figure 1. Diagram representing Swanson's first discovery: Fish oil cures Reynaud's diseases [5]

The potential of Swanson's research has been widely acknowledged by the scientific community, but likewise, its complexity concerns the vast information space and the possible exponential number of Bi connections given that MEDLINE contains more than twenty-five million medical articles. In addition to this complexity is that most information is represented in natural language, and this requires advanced natural language processing. For these reasons, it took Swanson ten years to discover seven natural cures to recurrent diseases.

This research aimed at enhancing Swanson's research work, particularly enhancing the Arrowsmith tool with advanced Text Mining techniques such as the co-occurrence Text Mining approach [6, 7] and proper Knowledge Base regarding the components of the dietary alimant and the sources of disease to undertake further bibliographic-based medical discoveries. These will be saved in a medical knowledge repository consisting of three entries the remedy/medicine, the disease, and scientific evidence that validate or refute the link between remedies and a given disease. Such a knowledge repository will motivate the medical research community to further investigate the clinical validity of the Text Mining induced findings of links between diseases and their cures.

The paper is organized as follows Section 2 gives the literature review on Text Mining and knowledge discovery in biomedical literature. Section 3 presents the process of knowledge discovery through co-occurrence Text Mining and domain Knowledge Base, and its implementation on a specific case of "Garlic" and its impact on "Blood Pressure". The last two sections are devoted to the conclusion and references.

2. LITERATURE REVIEW

In the last two decades, the advances in digital data collection devices and data storage technology allowed the storage of huge amounts of academic publications e.g., MEDLINE which contains more than 25 million references to medical journal articles in life sciences with a concentration on biomedicine abstracts (<https://www.nlm.nih.gov/pubs/factsheets/medline.html>). Underneath these empirical data are great opportunities for knowledge discovery using artificial intelligence and machine learning particularly Data Mining and Knowledge Discovery (DMKD). DMKD utilizes methods, algorithms, and techniques from a variety of disciplines to extract useful information, emerged during the 1990s, and grew rapidly, and in 2007, there were 17 data mining conferences [10].

This research work initiated the use of Text Mining approach of knowledge repository to mine the empirical literature in MEDLINE database to build scientific evidence that consolidates

publicly known (but not clinically proven) remedies for given diseases e.g., building scientific evidence that for instance garlic reduces the blood pressure.

Text Mining in biomedical literature has grown in the past few years to be a major tool for bioinformatics. For many applications, numerous methods have been developed. To contribute to this still-growing field, it is important to systemize the methods that are already in use. Usually, the Text Mining approach is based on natural language processing, mathematical and statistical backgrounds, it can be a strategic source of evidence for literature-based discoveries in biomedical sciences. For this reason, researchers and practitioners from various fields are using Text Mining.

Three types of Text Mining approaches are used for new knowledge discovery in the biomedical domain: rule-based or knowledge-based approaches, statistical or machine-learning based approaches, and co-occurrence-based approaches [8]. The most famous work on theory-research using the basic associative relationship as an evidence tool is the research work undertaken by Dr. Swanson [1] who has generated scientific evidence for several literature-based hypotheses that have been corroborated experimentally and clinically. He used an associative relationship between two concepts A and C to retrieve articles that contain the third set of concepts $B_1, B_2, B_3 \dots B_n$ that connect A and C concepts. The major limitation of Swanson approach is that the number of B_i connectors could grow exponentially as the number of articles retrieved from MEDLINE could be millions which will complicate or even make impossible the generation of scientific evidence. Another research work reported by Hui [8] uses the rule-based approach of Text Mining to automatically identify the status of obesity and related co-morbidities based on the patient's clinical discharge summaries. Some of the authors of this paper have undertaken several research works combining Text Mining term co-occurrence and statistical techniques approach to mine social networks and the holy Quran for developing knowledge repository for diabetes and Islamic financial business processes respectively [6,7].

Building on Swanson's successful research work and tackling the limitations of Swanson's basic associative relationship approach and being able to deal with millions of retrieved articles, this research aims at enhancing this research area through the development and implementation of Text Mining approach that combines term co-occurrence and domain Knowledge Base to filter relevant B_i terms retrieved from articles in MEDLINE. This new Text Mining approach will reduce the number of B_i to those semantically relevant to the analysis making it easier to generate scientific evidence for publicly known disease remedies and help in producing more relevant and coherent outcomes.

3. KNOWLEDGE DISCOVERY PROCESS AND IMPLEMENTATION

This research falls within theory-building methodology as it aims at building scientific evidence (theory) that links natural or chemicals existing in a dietary ailment that could be a remedy to a recurrent disease using Text Mining techniques and domain Knowledge Base to mine MEDLINE database. Data Mining and knowledge discovery support the two common strategies of theory building or scientific knowledge discovery [10] either by validating an existing theory known as a theory-to-research strategy or by developing a new theory known as a research-to-theory strategy. Text Mining supports the two strategies common to theory building. The first is of a research-to-theory strategy, whereas the second is of a theory-to-research strategy [11]. The research work fits well with the research-to-theory strategy which aims at deriving the laws of nature from a careful examination of all the available data which in this case is the 25 million medical articles contained in the MEDLINE dataset. As described by Reynold [11], the essences of this research-to-theory strategy are as follows:

1. Select a phenomenon and list all the characteristics of the phenomenon. The phenomenon is to identify all possible links between remedy-disease pairs giving a detailed list of their characteristics and the level of impact of remedy on the diseases.
2. Measure all the characteristics of the phenomenon in a variety of situations which means investigating the effect of the remedy on different persons e.g., female, male, old, and young person.
3. Analyze the resulting data carefully and determine any systematic patterns among the data “worthy” of further attention. This is the core of this research as it enhanced Swanson associative tool [4] with the advanced co-occurrence text mining tool [5,6] to semantically filter retrieved B_i concepts that have a medical relationship to both A and C.
4. Identify significant patterns during the process of mining the MEDLINE database, to formalize the discovered patterns as theoretical statements or scientific evidence that validate the hypothesis linking remedy-disease.

3.1. The Co-occurrence Algorithm and Domain Knowledge Base

First, we developed a Knowledge Base on the components of Dietary alimnt for instance “*The major bioactive compounds of garlic are its organosulfur compounds, such as diallyl thiosulfonate (allicin), diallyl sulfide (DAS), diallyl disulfide (DADS), diallyl trisulfide (DATS), E/Z-ajoene, S-allyl-cysteine (SAC), and S-allyl-cysteine sulfoxide (alliin)*” [6] In addition to statements on the impact of garlic on blood pressure for instance: “*Kyolic garlic has also shown promise in improving cardiovascular health by reducing arterial stiffness, elevated cholesterol levels and blood ‘stickiness’*” [8]. In addition, the prebiotic properties of garlic increase in increasing gut microbial richness and diversity” [9]

Once the Knowledge Base on known pairs of cure-disease was identified, we used R programming language to build a corpus of retrieved B_i terms to undertake the usual Text Mining process of eliminating stop words, and undertaking stemming and Part-of-Speech (PoS) Tagging to tag words according to the grammatical context of the word, hence dividing up the words into nouns, verbs, etc. This was important for the exact analysis of relations between words in the B_i list of terms itself and in relation to the domain Knowledge Base. In the same way, we produced B_i terms, and used a co-occurrence algorithm to filter B_i shared by “Garlic” and “Blood pressure” Knowledge Base as follow:

1. *Prepare the resulting B_i list of terms by eliminating stop words, stemming, and tagging special biomedical words,*
2. *Use co-occurrence algorithm to match terms in the B_i list of terms with the A (KB_Garlic) and C ($KB_BloodPressure$) Knowledge Base*
3. *Use the result of co-occurrence mining process to produce an ordered list of A components (shared by KB_Garlic and $KB_BloodPressure$) that could be used as evidence of links between “Garlic” and “Blood pressure”*

3.2. The Implementation of the Knowledge Discovery Process

The knowledge discovery process was used to find whether the dietary alimnt “Garlic” may help in reducing “Blood pressure”. First, we used the Arrowsmith tool to extract from the MEDLINE bio-medical database all medical articles and abstracts that contained the term “Garlic” (Figure 2).

The screenshot shows a web interface for literature retrieval. At the top, there are navigation tabs: Start, A-Literature, C-Literature (highlighted), B-list, Filter, and Literature. A search bar contains the text 'Blood pressure' and a search button. Below the search bar, there are links for Home, Two-Node Literature, Job ID: 16181, Show Options, and Hide. A confirmation message asks 'Use 'Blood pressure' for ARROWSMITH?' with a 'Yes' button. Below this, it says 'Items 1 - 20 of 25000 (Next)'. Three search results are listed:

- Cerebral blood flow in stroke patients with sleep apnea: any role of single-night positive airway pressure therapy?**
Neuro endocrinology letters. 2021 ;42(8);
Jurik M, Siarnik P, Valovieova K, Karapin P, Klobucnikova K, Tureani E, Kollar B
Despite the significant reduction of respiratory events, single-night PAP therapy does not improve overall cerebral blood flow, as defined by CFI.
PMID: 34969188
- Fluid Resuscitation in Tactical Combat Casualty Care; TCCC Guidelines Change 21-01. 4 November 2021.**
Journal of special operations medicine : a peer reviewed journal for SOF medical professionals. 2021 ;21(4):126-137
Deaton TG, Auten JD, Betzold R, Butler FK, Byrne T, Cap AP, Donham B, DuBose JJ, Fisher AD, Hancock J, Jourdain V, Knight RM, Littlejohn LF, Martin MJ, Toland K, Drew B
Hemorrhagic shock in combat trauma remains the greatest life threat to casualties with potentially survivable injuries. Advances in external hemorrhage control and the increasing use of damage control resuscitation have demonstrated significant success in decreasing mortality in combat casualties. Presently, an expanding body of literature suggests that fluid resuscitation strategies for casualties in hemorrhagic shock that include the prehospital use of cold-stored or fresh whole blood when available, or blood components when whole blood is not available, are superior to crystalloid and colloid fluids. On the basis of this recent evidence, the Committee on Tactical Combat Casualty Care (TCCC) has conducted a review of fluid resuscitation for the combat casualty who is in hemorrhagic shock and made the following new recommendations: (1) cold stored low-titer group O whole blood (CS-LTOWB) has been designated as the preferred resuscitation fluid, with fresh LTOWB identified as the first alternate if CS-LTOWB is not available; (2) crystalloids and Hextend are no longer recommended as fluid resuscitation options in hemorrhagic shock; (3) target systolic blood pressure (SBP) resuscitation goals have been redefined for casualties with and without traumatic brain injury (TBI) coexisting with their hemorrhagic shock; and (4) empiric prehospital calcium administration is now recommended whenever blood product resuscitation is required.
PMID: 34969143
- Elevated BNP and High Brachial Pulse Pressure in Patients with Diabetes.**
American journal of hypertension. 2021 ;:

Figure 2. A - Literature Retrieval

Figure 2 shows that 7,103 bio-medical papers and abstracts were retrieved that contained the term “Garlic”. The same process was used to retrieve from the literature containing the C term i.e., “Blood pressure” (Figure 3).

The screenshot shows a web interface for literature retrieval, similar to Figure 2. The search bar contains 'Blood pressure'. The confirmation message asks 'Use 'Blood pressure' for ARROWSMITH?' with a 'Yes' button. It says 'Items 1 - 20 of 25000 (Next)'. Three search results are listed:

- Cerebral blood flow in stroke patients with sleep apnea: any role of single-night positive airway pressure therapy?**
Neuro endocrinology letters. 2021 ;42(8);
Jurik M, Siarnik P, Valovieova K, Karapin P, Klobucnikova K, Tureani E, Kollar B
Despite the significant reduction of respiratory events, single-night PAP therapy does not improve overall cerebral blood flow, as defined by CFI.
PMID: 34969188
- Fluid Resuscitation in Tactical Combat Casualty Care; TCCC Guidelines Change 21-01. 4 November 2021.**
Journal of special operations medicine : a peer reviewed journal for SOF medical professionals. 2021 ;21(4):126-137
Deaton TG, Auten JD, Betzold R, Butler FK, Byrne T, Cap AP, Donham B, DuBose JJ, Fisher AD, Hancock J, Jourdain V, Knight RM, Littlejohn LF, Martin MJ, Toland K, Drew B
Hemorrhagic shock in combat trauma remains the greatest life threat to casualties with potentially survivable injuries. Advances in external hemorrhage control and the increasing use of damage control resuscitation have demonstrated significant success in decreasing mortality in combat casualties. Presently, an expanding body of literature suggests that fluid resuscitation strategies for casualties in hemorrhagic shock that include the prehospital use of cold-stored or fresh whole blood when available, or blood components when whole blood is not available, are superior to crystalloid and colloid fluids. On the basis of this recent evidence, the Committee on Tactical Combat Casualty Care (TCCC) has conducted a review of fluid resuscitation for the combat casualty who is in hemorrhagic shock and made the following new recommendations: (1) cold stored low-titer group O whole blood (CS-LTOWB) has been designated as the preferred resuscitation fluid, with fresh LTOWB identified as the first alternate if CS-LTOWB is not available; (2) crystalloids and Hextend are no longer recommended as fluid resuscitation options in hemorrhagic shock; (3) target systolic blood pressure (SBP) resuscitation goals have been redefined for casualties with and without traumatic brain injury (TBI) coexisting with their hemorrhagic shock; and (4) empiric prehospital calcium administration is now recommended whenever blood product resuscitation is required.
PMID: 34969143
- Elevated BNP and High Brachial Pulse Pressure in Patients with Diabetes.**
American journal of hypertension. 2021 ;:

Figure 3. C - Literature Retrieval

Figure 3 shows that 25,000 bio-medical papers and abstracts were found that contained the term “Blood pressure”. Next, we undertook the process of extracting B lists of terms that are shared by both A-Literature and C-Literature lists giving the results as shown in Figure 4.

Start A-Literature C-Literature **B-list** Filter Literature

A-query: Garlic
C-query: Blood pressure

The B-list contains title words and phrases (terms) that appeared in both the A and the C literature. **264** articles appeared in both literatures and were not included in the process of computing the B-list but can be viewed [here](#). The results of this search are saved under id # **16181** and can be accessed from the start page after you leave this session. There are **14120** terms on the current B-list (**1701** are predicted to be relevant), which is shown ranked according to predicted relevance. The list can be further trimmed down using the filters listed in the left margin.

To assess whether there appears to be a biologically significant relationship between the AB and BC literatures for specific B-terms, please select one or more B-terms and then click the button to view the corresponding AB and BC literatures. Use Ctrl to select multiple B-terms.

Rank	Prob	B-term
1	0.81	adiponectin
2	0.81	phase microextraction
3	0.81	solid phase microextraction
4	0.81	toll receptor
5	0.82	jnk
6	0.82	liquid chromatography tandem
7	0.82	carvedilol
8	0.82	chromatography tandem
9	0.82	wnt
10	0.82	carotid intima
11	0.82	--carotid intima media
12	0.82	akt
13	0.82	vegf
14	0.82	optical coherence
15	0.82	--optical coherence tomography
16	0.82	atorvastatin
17	0.82	mtor
18	0.82	stat3
7166	0.00	cell surface
7167	0.00	experimental data
7168	0.00	influence environmental
7169	0.00	heterologous
7170	0.00	cancellation
7171	0.00	desensitization
7172	0.00	precipitation
7173	0.00	tissue vivo
7174	0.00	propagation
7175	0.00	trade
7176	0.00	reservation
7177	0.00	activity evaluation
7178	0.00	year experience
7179	0.00	stimuli
7180	0.00	enlargement
7181	0.00	microorganism
7182	0.00	operation
7183	0.00	staining
7184	0.00	lacking
7185	0.00	connection

Figure 4. The retrieved B list of terms shared by A-Literature and C-Literature Lists

Figure 4 retrieved the huge number of 7,185 B_i terms and for this reason, there was a need to develop semantic filters to be able to build the evidence that “Garlic” reduces “Blood pressure”.

For this reason, after preparing the B list of terms the third step was to implement the cooccurrence algorithm using Python programming language as shown in Figure 5.

```

1  #-*- coding: utf-8 -*-
2  """
3  Created on Sat Jan  1 18:01:05 2022
4
5  @author: Z9701
6  """
7
8  B_List= open("D:\\a a RIFs\\2019 RIF Swanson\\The paper\\B_List_GarlicVSBloodPressure.txt", 'r', encoding = 'utf-8')
9  KB_Garlic= open("D:\\a a RIFs\\2019 RIF Swanson\\The paper\\KB_Garlic.txt", 'r', encoding = 'utf-8')
10 KB_BloodPressure= open("D:\\a a RIFs\\2019 RIF Swanson\\The paper\\KB_BloodPressure.txt", 'r', encoding = 'utf-8')
11
12 words1 = B_List.read().split()
13 words2 = KB_Garlic.read().split()
14 words3 = KB_BloodPressure.read().split()
15 words = set(words1) & set(words2) & set(words3)
16
17 with open('outfile.txt', 'w') as output:
18     for word in words:
19         output.write('{} appears {} times in B_List and {} times in KB_Garlic and {} times in KB_BloodPressure.\n'.forma

```

Figure 5. The co-occurrence Python code for filtering B List of terms

Figure 6 shows the filtered list of B terms that resulted from the co-occurrence Python code (Figure 5). The filtering process retrieved terms that occurred simultaneously in B terms lists, “Garlic” Knowledge Base, and “Blood pressure” Knowledge Base (Figure 6).

inhibiting appears 9 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 human appears 85 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 enzyme appears 11 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 reducing appears 5 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 hydrogen appears 4 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 linked appears 2 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 animal appears 8 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 production appears 7 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 in appears 2 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 pressure appears 7 times in B_List and 1 times in KB_Garlic and 5 times in KB_BloodPressure.
 blood appears 24 times in B_List and 2 times in KB_Garlic and 6 times in KB_BloodPressure.
 content appears 7 times in B_List and 1 times in KB_Garlic and 1 times in KB_BloodPressure.
 angiotensin appears 5 times in B_List and 1 times in KB_Garlic and 2 times in KB_BloodPressure.

Figure 6. Filtered List of B terms

Finally, after further analysis we found that the term “angiotensin” (Figure 7) could be used as part of the evidence that garlic could reduce the impact of blood pressure as it fits with the following statement:” Blood pressure reducing properties of garlic have been linked to its hydrogen sulphide production [4,5] and allicin content – liberated from alliin and the enzyme alliinase [6,10] – which has angiotensin II inhibiting and vasodilating effects (the dilatation of blood vessels, which decreases blood pressure), as shown in animal and human”. Other terms like enzymes, hydrogen, and blood could be further analyzed to further consolidate the evidence.

Rank	Prob	B-term
1514	0.22	anatoxin b1
1515	0.22	endophytic fungi
1516	0.22	antimicrobial potential
1517	0.22	anti proliferative effect
1518	0.22	soy protein
1519	0.22	experimental autoimmune encephalomyelitis
1520	0.22	matrix metalloproteinase-1
1521	0.22	signaling inflammatory
1522	0.22	identification a
1523	0.22	lipogenesis
1524	0.22	androgen receptor expression
1525	0.22	evaluation herbal
1526	0.22	intravenous iron
1527	0.22	segmentation
1528	0.22	macromolecular crowding
1529	0.22	nfkappab
1530	0.22	angiotensin converting
1531	0.22	peritoneal adhesion
1532	0.22	angiotensin converting enzyme
1533	0.22	data mining
1534	0.22	blinded trial

Figure 7. the cocccurence of the the530th term “angiotensen” in the original B terms

4. CONCLUSIONS AND DISCUSSION

This research aimed at building scientific evidence that linked natural or chemicals existing in a dietary ailment that could be a remedy to a recurrent disease using Text Mining techniques and domain Knowledge Base to mine MEDLINE database knowledge discovery as a research-to-theory strategy. It was successfully able to discover and establish an important association on the specific case of “Garlic” and its impact on “Blood pressure”. The results of this research work will be added to the results of other experimental studies to ascertain the efficacy of garlic in reducing blood pressure. As this research project is ongoing, we will be investigating other important associations reported by practicing leading physicians in the region.

Non-pharmacological treatment options for blood pressure or hypertension have the potential to reduce the risk of cardiovascular diseases at a population level. A systematic study on the impact of garlic on blood pressure was undertaken [15] and concurs with the results of our study on mining biomedical literature. They undertook a systematic clinical study on animals and found evidence that garlic reduced blood pressure. [15] Also primary studies were conducted in humans and in a meta-analysis of 12 studies in subjects with uncontrolled hypertension, garlic was found to be effective in reducing the blood pressure. [16] The results stated that garlic lowers blood pressure in hypertensive subjects, improved arterial stiffness and gut microbiota. [16] However, as the number of subjects per trial was too small this warrants further research to be able to ascertain a causal inference between the positive effect of garlic in reducing blood pressure. Randomized controlled clinical trials with adequate sample size in subjects with elevated blood pressure is the gold standard to test the efficacy of this new therapy of the use of garlic in lowering the patients’ blood pressure. However, conducting a large randomized controlled clinical trial is extremely costly and time-consuming. For this reason, using Swanson’s Arrowsmith machine will be a first step to find associations as it is quick and an inexpensive tool where data mining and knowledge discovery can be used to seek any possible connections between garlic and lowering the blood pressure. Furthermore, as interest in complementary medicine for blood pressure is increasing, we conducted this approach of mining biomedical literature to find in the B list the term “angiotensin” shared by “Garlic” literature and by “Blood pressure”. In the literature garlic is a natural angiotensin converting enzyme inhibitor which has an effect on reducing blood pressure. [17] To consolidate our findings, we also ran the mining process on some of a couple of cure-disease discovered by Swanson like for instance Fish oil (A)

and Blood viscosity (C), Magnesium Deficiency (A), Migraine headache (C), and we reached at the same conclusion as reported in the paper [10]. The findings of our literature-based mining on the impact of garlic on blood pressure can be validated only with further research such as in double blind controlled clinical trials which is the golden standard to prove a causal association.

ACKNOWLEDGEMENTS

The authors would like to thank Zayed University for their financial support through the scheme of Research Incentive Funding (RIF code: R19048 and PRFA code: R20092).

COMPETING INTERESTS

The authors declare that there is no conflict of interest

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper. Dr. Farhi Marir, Dr. Hussein Fakhry and Aida J. Azar contributed to the study conception and design, draft manuscript, revising the manuscript critically for important intellectual content, and final approval. All authors reviewed the results and approved the final version of the manuscript.

TRANSPARENCY

The authors affirm that all information submitted in this manuscript is accurate and true. All data has been reported and no data has been omitted.

REFERENCES

- [1] Swanson, D. R. (1986). Fish Oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine*, 30, 7-18.
- [2] Ralph A. Digiacomio, Joel M. Kremer, Dhiraj M. Shah. Fish-oil dietary supplementation in patients with Raynaud's phenomenon: A double-blind, controlled, prospective study, *The American Journal of Medicine*, Volume 86, Issue 2, 1989, Pages 158-164
- [3] Swanson, D. R. and Smalheiser N. R. (1996). Undiscovered Public Knowledge: A Ten-Year Update. *KDD-96 Proceedings*. AAAI (www.aaai.org).
- [4] Arrowsmith tool located in http://arrowsmith.psych.uic.edu/arrowsmith_uic/index.html, accessed on the 25th of March 2022
- [5] Marc Weeber et al. (2001). Using Concepts in Literature-Based Discovery: Simulating Swanson's Raynaud-Fish Oil and Migraine-Magnesium Discoveries *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY*, 52(7):548-557, 2001
- [6] Marir, F., and Huwida, S. (2016). Text Mining Social Networks to Discover New Symptoms and Treatment for Diabetes Patients, 3rd International Workshop on Machine Learning and Data Mining for Sensor Networks (MLDM-SN), May 23-26, 2016 in Madrid, Spain
- [7] Marir, F. et al. (2018). A Recursive Co-Occurrence Text Mining of the Quran to Build Corpora for Islamic Banking Business Processes. *The International Conference on Intelligent Human Systems Integration* (<http://ihsint.net/info.html>).
- [8] Shang, A., Cao, S. Y., Xu, X. Y., Gan, R. Y., Tang, G. Y., Corke, H., Mavumengwana, V., & Li, H. B. (2019). Bioactive Compounds and Biological Functions of Garlic (*Allium sativum* L.). *Foods* (Basel, Switzerland), 8(7), 246. <https://doi.org/10.3390/foods8070246>
- [7] Ried K. (2020). Garlic lowers blood pressure in hypertensive subjects, improves arterial stiffness and gut microbiota: A review and meta-analysis. *Experimental and therapeutic medicine*, 19(2), 1472-1478. <https://doi.org/10.3892/etm.2019.8374>

- [8] Ried, K., Travica, N., & Sali, A. (2018). The Effect of Kyolic Aged Garlic Extract on Gut Microbiota, Inflammation, and Cardiovascular Markers in Hypertensives: The GarGIC Trial. *Frontiers in nutrition*, 5, 122. <https://doi.org/10.3389/fnut.2018.00122>
- [9] Bretonnel K. C. and Lawrence Hunter: Getting Started in Text Mining. *PLoS Comput Biol.* Vol. 4(1). (2008)
- [10] Hui Y., Irena S., John A. K., Goran N.: A Text Mining Approach to the Prediction of Dis-ease Status from Clinical Discharge Summaries. *Journal of the American Medical Informatics Association* Vol.16, No. 4. (2009)
- [11] Kdnuggets, Meetings/Conf. in Data Mining, Knowledge Discovery, and Web Mining (2007). Available online at <http://www.kdnuggets.com/meetings/index.html>.
- [12] Reynolds, P. D. (1971). *A primer in theory construction*. New York: Macmillan.
- [13] Lynham, S. A. (2000b). Theory building in the human resource development profession. *Human Resource Development Quarterly*, 11(2), 159-178.
- [14] Karin Ried, Oliver R Frank, Nigel P Stocks, Peter Fakler & Thomas Sullivan (2008). Effect of garlic on blood pressure: A systematic review and meta-analysis, *BMC Cardiovascular Disorders* 8, Article number: 13
- [15] Ried K. Garlic lowers blood pressure in hypertensive subjects, improves arterial stiffness and gut microbiota: A review and meta-analysis. *Exp Ther. Med.* 2020; 19(2):1472-1478. doi:10.3892/etm.2019.8374).
- [16] Sharifi AM, Darabi R, Akbarloo N. Investigation of antihypertensive mechanism of garlic in 2K1C hypertensive rat. *J Ethnopharmacol* 2003;86:219-24

AUTHORS

Dr. Farhi Marir is Fellow of British Computer Society. He joined ZU in 2013 and he is director of the CTI Research Centre on Big Data & Analytics. He was Acting Director of the Institution Research Office in ZU for two years (2014-2016) where he initiated ZU project for the development of ZU Data warehouse and Business Intelligent Dashboards. He is currently teaching Data analytics he developed, Knowledge Management, Emerging Technologies, and Data Warehousing. As for research in ZU, he is either a PI or Co-PI in six university RIFs grants worth more than 1 million AED applying data analytics in the domain of Islamic Finance, Health (Diabetics & Autism), Education, and Security. He is also leading an 175K AED Research Project on Big Data Analytics for Dubai Tourism which is funded by the UAE National Research Funding. Prof Marir received his PhD in Deductive Databases in UK in 1993 and worked for around twenty five years (1988-2013) in UK university where he taught a large number of courses related to Data science, artificial intelligence and knowledge management. He was also the Director of the Knowledge Management Research Center (KMRC) at London Metropolitan University (2000-2013). As the director of KMRC, Prof. Marir led around thirty-five UK and EU funded research projects worth more than £7 Million, supervised thirty PhD students to completion and published more than hundreds research papers. Well before studying for PhD in the UK, Prof Marir was head of a computer science department for six years at the University of Batna in Algeria (1982-1988). During this time, he set up one of the first university computer center in Algerian Universities.



Dr Hussein Fakhry received his PhD in October 1994 in Intelligent Control Systems & Robotics from the University of Waterloo, Canada. Since then, he assumed different academic and administrative positions at University of Windsor, Cairo University, University of Dubai and Zayed University.



Dr. Hussein Fakhry has a rich profile working at academia in different roles. Since September 2001 he worked at University of Dubai in different roles as faculty, then Assistant Dean and later as Dean of the College of Information Technology. After moving to Zayed University in September 2014, he assumed different roles as Program Coordinator, Assistant Chair and later as Assistant Dean for Students at the College of Technological Innovation. He is currently an Associate Professor in the College of Technological Innovation at Zayed University. Dr Fakhry has extensive professional experience in teaching and training particularly in Systems Analysis & Design, Database Design, Data warehousing, Expert Systems, Decision Support Systems, Strategic Management

Modelling using System Dynamics, Modelling and Simulation of the Supply Chains, and Operations Research Models. Dr Fakhry's research interests are in the areas Applications of Artificial Intelligence, Information Systems Research using System Dynamics, Information Systems Security, E-Commerce and E-Business, Decision Support Systems, and Assessment of Academic Programs. His research has appeared in several international journals and conferences such as Review of Business Information Systems, Communications of the International Information Management Association CIIMA, An International Journal of Information & Security, The Journal of Mathematical and Computer Modelling, Proceedings of the IEEE International Conference on Information and Communication Technologies, and the IADIS International Conference on Information Systems.

Dr. Aida J. Azar joined Mohammed Bin Rashid University of Medicine and Health Sciences (MBRU) since its inception in 2016. In her current position, she is Associate Professor Epidemiology and is one of the founding members. She is responsible for developing and coordinating the epidemiology, biostatistics, and research curriculum for the undergraduate medical students. In her previous positions, Dr. Aida was part of several collaborative groups on writing cardiovascular guidelines for Dutch cardiologists. Also, she was involved in the conduct and analysis of numerous multicenter clinical trials. Her dissertation on 3404 myocardial infarction patients was used as a reference for the Food and Drug Administration (FDA) to adjust the optimal intensity of long-term anticoagulant therapy in post-myocardial infarction patients, and to define the optimal anticoagulation level. She also took part in the FDA report for the approval of coronary stenting for selective placement in selected patients to prevent restenosis after angioplasty. The results were a milestone in interventional cardiology.

