International Journal of Data and Network Science 6 (2022). 1013–1022

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Artificial intelligence for target symptoms of Thai herbal medicine by web scraping

Chairote Yaiprasert[a*] and Gorawit Yusakul[b]

[a]School of Science, Walailak University, Nakhon Si Thammarat, 80160, Thailand
[b]School of Pharmacy, Walailak University, Nakhon Si Thammarat, 80160, Thailand

C H R O N I C L E

Article history:
Received: October 27, 2021
Received in revised format: November 20, 2021
Accepted: January 21, 2022
Available online: January 24 2022

Keywords:
Artificial Intelligence (AI)
Machine Learning (ML)
Thai Herbal Medicine (THM)
Programming

A B S T R A C T

Machine learning (ML) is implementing artificial intelligence (AI) research within medicine that has made dramatic progress in recent years. In addition to standard treatments, the role of complementary and alternative medicine should be mentioned. Traditional Thai medicine has received growing acceptance as a complementary approach to modern medicine by using local herbs. A vast amount of Thai herbal knowledge and information is freely available on the Internet. The reader must evaluate each website and decide to use trustworthy and appropriate information. This study aimed to acquire Thai herbal knowledge recorded in the Thai language system on the Internet by scraping websites using programming techniques. The knowledge was extracted with programming, and the types of Thai herbs were classified corresponding to target symptoms by the machine learning algorithm. The ML method organized the process when sufficient achievement was reached in order to give reliable and high accuracy results from the training data set. The validation of extracted knowledge was achieved by using the part-of-speech tag patterns analysis. This study showed that the programming and machine learning system was appropriate for obtaining and classifying Thai herbal medicines knowledge.

1. Introduction

Most people use the Internet as their primary information source. However, the information from the Internet is only partially helpful and reliable. The Internet is the primary gateway for finding health information (Du et al., 2020; Gianfredi et al., 2021). The elderly is a vulnerable group in the use of Internet health information (Chang et al., 2020). Obtaining information about health on the Internet can lead to fears and misunderstandings (Gianfredi et al., 2021; Serçekuş et al., 2021). Health professionals should lead people to reliable sources (Serçekuş et al., 2021). Knowledge of Thai herbal medicine (THM) is another body of knowledge that can be accessed through the Internet. The application of herbal knowledge is sensitive to health. Therefore, information from reliable sources with relevant research is required. Filtering the proper knowledge to meet the needs is an exciting problem. The use of big data in scientific research is intended to grow and become more visible in daily healthcare practice (Murdoch & Detsky, 2013). Scientists use AI to analyze large amounts of data and help automate health tasks. It can define processes for extracting and interpreting data with a common set of automated methods. This has always been a growing inspiration to analyze data in regard to Thai herbal knowledge and information. Artificial intelligence (AI) technology plays a role in various medical fields, with the ability of the machine to mimic human analytical reasoning. Recent research has linked health care with AI to acquire the proper knowledge in herb usage. Some studies have focused on using new Chinese herbal medicine clustering algorithm to mimic the configuration of artificial bee colonies (Han et al., 2019), implementing knowledge graph for traditional Chinese medicine care design (Yu et al., 2017), learning the

* Corresponding author.
E-mail address: ychairot@mail.wu.ac.th (C. Yaiprasert)

doi: 10.5267/j.ijdns.2022.1.010

representation of different networks related to symptoms to predict herb selection (Wang et al., 2019), and facilitating the discovery of new drugs based on basic traditional Chinese medicine knowledge (Wang et al., 2019). AI may mean the development of traditional Chinese medicine data on integrated clinical practice and patient outcomes for medical knowledge discovery and decision support system design (Zhou et al., 2010). Chinese herbal medicine image recognition and artificial neural network retrieval simplify herbology classification, which reduces false memory by non-professional consumers (Sun & Qian, 2016). Interestingly, the development of a model for classifying action between herbs, clearing heat and herbal stasis, and activating blood can choose herbs for specific diseases based on the theory of Chinese medicine (Chen et al., 2018). Overall, the AI and learning tools facilitate research, drug discovery, and decision-making on unclear clinical evidence of herbal therapies.

Thai herbal medicine has been an essential part of primary health care in Thailand for centuries thanks to notable efficacy with few side effects. A recent survey indicated that 32.6% of participants are THM users (Kanjanahattakij et al., 2019). An enormous amount of herbal knowledge has already been published in the form of text or documents for a long time (Kee et al., 2015). A primary factor of this investigation is the lack of experts to classify herbal property definitions and target symptoms analysis, leading to misinterpretations about herb usage. In THM, an herb is usually effective for a specific symptom, and the task involves organizing the herbs in a diagnostic or therapeutic class. Each type of herb has more specific medical potency and efficacy in combination with other herbs. The proven medically effective formula or decoction method has been reported in text, document, or media for future use or recommendations. Acquiring knowledge from document records is one of the most useful research projects in the medical field by extracting knowledge from semi-structured text (Cao et al., 2004), studying the algorithm to recognize natural language processing and to acquire ideas and links from documents (Wen-Xiang et al., 2019), and working with semantically tagged texts (Mikkelsen & Aasly, 2002). Recently, natural language processing and machine learning (ML) methods have shown great engagement in analyzing clinical records (Sheikhalishahi et al., 2019). There is a significant increase in the use of machine learning compared with rule-based methods (Sheikhalishahi et al., 2019). However, these ML implementations do not deal with big data.

This paper represents an efficient acquisition of THM knowledge from big data on the Internet using a web scraping technique. ML was used to specify herbal properties and target symptoms. Many symptoms can be mixed to indicate medicinal properties or vice versa. Web scraping was developed to implement the proposed design by using functional programming to import data from the URLs of the web provided. In addition, Unicode regular expression is a language structure that can be used for the keywords of the associated rule by using some patterns of target symptoms. The keywords related to herbal properties and symptoms can be used in the ML process to organize the THM knowledge recorded. The Unicode patterns of target symptoms were implemented to specify THM. The Wolfram programming language can solve all technical issues of ML techniques (Wolfram, 2017). Details of the research model framework is shown in Fig. 1.

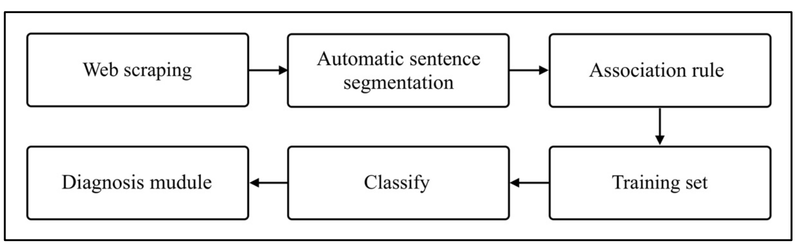


Fig. 1. Research model framework.

2. Materials and Methods

2.1 Web scraping

The Internet is the wellspring of the world's knowledge. The information provided by search engine sites or social media applications is sometimes overloaded. Reliable THM data is easy to access by users, but it is quite hard to analyze the vast amount and redundancy of the information. Many THM resources are freely available on the Internet. This study used information freely from MedThai, which provides THM's number 1 ranked website in Thailand (MedThai, 2021). This study acquired THM data from free and available sources on the Internet containing information about various herbal knowledge published over the years. This statement includes common names and scientific names of herbs, dosing methods, herbal properties, side effects, notices, references, etc. This study acquired data with web scraping or web harvesting methods. Web scraping is the act of obtaining data or information from websites to use for THM knowledge purposes that can be done automatically by using a programming processing method. This technique was applied to extract vast amounts of data from websites, collecting and saving it to a local computer or database. The computation was performed using a MacBook Pro, 2.3 GHz Intel Core i5 computer processor, with 8 GB 2133 MHz LPDDR3 memory. Mathematica version 12 of the Wolfram Language was used in this study for algorithmic development. Functional programming language was used for web scraping in this study. The programming provided the import of web data mechanisms for interfacing with the web. It can manipulate most web crawling-related processes efficiently. The "Import" function of the functional programming was performed to import all data from the web URLs, and the "Data" option was used for extracting data and content. The programming can download many pages per second while being robust spider traps against some crashes from web protection. The "Table" function generated a list of the web URLs values. The listing process chose the page that was the target of scraping. The number of pages scraped was in the range of one until the last number of pages. The simple function of web scraping programming is shown in Code Snippet 1.

```
1    Import[URL, "Data"]
2    Table[expression, {i, 1, imax }]
```

Code Snippet 1. The simple form of functions: (1) the “Import” function takes the file to be in the specified "Data" format, and (2) the “Table” function generates a values list of expressions when *i* operates from 1 to *i* maximum values. The operation of a loop function involves iterating over an object.

The listing process chose the page as the scraping target. The number of pages scraped was in the range from one until the last number of pages. The name of the herb determined the number of pages on the web. Code Snippet 2 is the snippet process code example for web scraping from the web URLs.

```
URL="url/"<>ToString[HerbalNames];
HerbalData = Table[Import[URL, "Data"] , {i, 1, Length[HerbalNames]}];
```

Code Snippet 2. Web scraping THM data from web URLs operates with the functional programming language as the “Import” function. The “Table” function is a looping of the importing process of the web page, which begins on page one until the last page of the herbal name.

2.2 Data manipulation

Setting up the AI for THM in terms of the direct interpretation of the THM knowledge system (mostly recorded in the Thai language) is complicated in structured or unstructured web data. Information from the web scraping process is not yet usable. Cleaning data involves the challenge of acquiring and managing the vast amount of metadata. This study required data processing before entering the AI process. The success key to solving this problem was based on transforming all types of data forms to the string as a pattern of text format. This study used the “ToString” function to manipulate the Thai language of herbal information as a Unicode form to string format. Then, the “StringSplit” function was used to split the data series into a list of substrings separated by whitespace, combined with the “Flatten” function. The programming can implement structured or unstructured web data using the “Flatten” function that effectively deleted only the inner braces. The integrated action of the three functions is shown in Code Snippet 3.

```
DataSet = Flatten[StringSplit[ToString[HerbalData]]];
```

Code Snippet 3. The THM information from the web URLs was converted to string format and separated by whitespace. The new structure of the THM data set was consequently created with the “ToString”, “StringSplit” and “Flatten” functions.

The traditional language was a problem for the programming process since the program understood only machine language. The English language was used as a primary symbol to communicate with the program. This study solved this problem by converting the Thai language to Unicode format. Moreover, the Thai language style format was another problem for text extraction. Almost all THM data is recorded in the Thai language. Traditional Thai language problems include four related features of the Thai writing style. First, both single words and compound words have no clear boundaries. Second, long strings may be interpreted as one or more sentences because the sentence ending is unclear. Third, compound words and their sentences may be unpredictable because some Thai compound words resemble sentences. Finally, the vague meaning of the word is often used. Moreover, Thai sentences are often long and complicated. Both grammatical and object matters are often ignored. Space does not just act as a hint at the end of a sentence, but it is also used between proper noun phrases and particular words in the sentence. In addition, some words have the same form as the sentence structure. Therefore, there may be an ambiguity between words and sentences. However, this study solved the problem by defining the THM knowledge from the keywords.

To create automatic sentence segmentation, a process was necessary to specify the keyword scope of the symptoms and therapy. The symptom and therapy keywords set in this study were essential conditions to determine the extracted THM knowledge from the imported data. In this study, the Unicode of the Thai language was the symptom keywords set. The meanings of keywords included the following: alleviate, cure, fever, heal, help, hurt, mitigate, pain, prevent, sick, vomiting, etc. The list of symptoms and therapy keywords determined the pairing by observing the common phrase patterns of symptoms and treatment in the THM data set. This keyword set matched with the THM data set was used to create the new list of target symptom datasets by using the “StringCase” programming process. The THM string expressions can contain any objects specified in the records for substring expressions. The simple usage "StringCase" function is displayed in Code Snippet 4. Then, an association of ML format was automatically generated from the extracted pattern of target symptoms used to ascertain keys and herbal types defined as values. An example algorithm in Code Snippet 5 shows the programming process to produce an association format.

```
StringCases["string", patterns]
```

Code Snippet 4. The simple form of the "StringCase" function returns a list whose elements are the substrings of string matching any of the patterns.

```
SymptomKeywords = {"\\ :0e22\\:0e32"~~___~~x_, "\\:0e41\\:0e01\\:0e49"~~___~~x_,
"\\:0e15\\:0e49\\:0e32\\:0e19"~~___~~x_};
TargetSymptom = Flatten[StringCases[DataSet, SymptomKeywords]];
AssociationForm = Table[{TargetSymptom[[i]] -> HerbalNames}, {i, 1, Length[TargetSymptom]}];
```

Code Snippet 5. The association data set is created from the Unicode keywords of symptoms and therapy matching with web scraping data sets. The association format of the ML training set is the relationship between the THM target symptoms and herbal names.

ML algorithms are developed to automate learning and decisions based on patterns and inferences. The quantity and quality of the training package are essential in the development of an ML system. The "Classify" function has been programmed with various kinds of classification based on the association data set or the training examples. The ML possible settings are available on the functional programming, such as Decision Tree, Gradient Boosted Trees, Logistic Regression, Markov, Naïve Bayes, Nearest Neighbors, Neural Network, Random Forest, and Support Vector Machine. The Wolfram Language recommends the ML method by Markov to support text data processing (Wolfram, 2017). However, this study was combined on all ML methods available in the Wolfram Language. All ML methods were programmed to be the same.

The program designed the new structured training set using the “Flatten” function related to the required ML function. ML programming can classify the essential training set from the association data set by using the “Classify” function. The machine was assigned to learn the Unicode recognition from the association rule, which is a rule-based ML method for finding interesting relationships between variables in large databases. Each symptom and therapy imported to designate the primary key of the association rule is shown in Code Snippet 6. The training time of the model that defined a language model was computed for each class.

```
TrainingSet = Flatten[AssociationForm];
ModelClassification = Classify[TrainingSet, Method->"Method"];
```

Code Snippet 6. A training set is defined from the associated symptoms and therapy data set of the Unicode keywords.

A programming procedure created the diagnosis module. The module acquired input from the user as phrases of symptoms or therapy in one or more phrases that they needed. The module contained the main ML system classification obtained from the data of the user. The module used the “Table” function as a technique to obtain one or more input phrases from the user. The diagnosis module is shown in Code Snippet 7.

```
Diagnosis[inputPhrases_] := Module[{diag, ML, results},
diag = StringSplit[ToString[inputPhrases]];
ML = ModelClassification[diag, "TopProbabilities"];
results = Table[ML[[i]][[{1}]], {i, 1, Length[diag]}]
```

Code Snippet 7. The diagnosis module receives inputs from the user and uses the ML programming to return the result of herbal names corresponding with input phrases.

In addition, target symptom phrases to extract THM knowledge through the ML process had been formulated. This study also created a system for utilizing the learning data set. The program designed a single-word search of the target symptom to obtain all kinds of THM knowledge. On the other hand, identifying the herbal name can manifest all kinds of relevant target symptoms. The herbal properties and symptom search module are displayed in Code Snippet 8.

```
keys = Flatten[AssociationForm /. {x_ -> y_} -> {y -> x}];
HerbalProperties[input_] := Module[{},
TableForm[Flatten[DeleteCases[Table[Which[keys[[i]][[1]] ===ToString[input]<> ".",{keys[[i]]}], {i, 1,
Length[keys]}], Null]]]]
```

```
SymptomSearch[input_] := Module[{},
TableForm[Flatten[DeleteCases[Table[Which[StringContainsQ[ToString[keys[[i]][[2]]], ToString[input]],{keys[[i]]/.{x_-> y_}->{y->x}}], {i, 1, Length[keys]}], Null]]]]
```

Code Snippet 8. The herbal properties and symptom search module is created from keywords of the training data set.

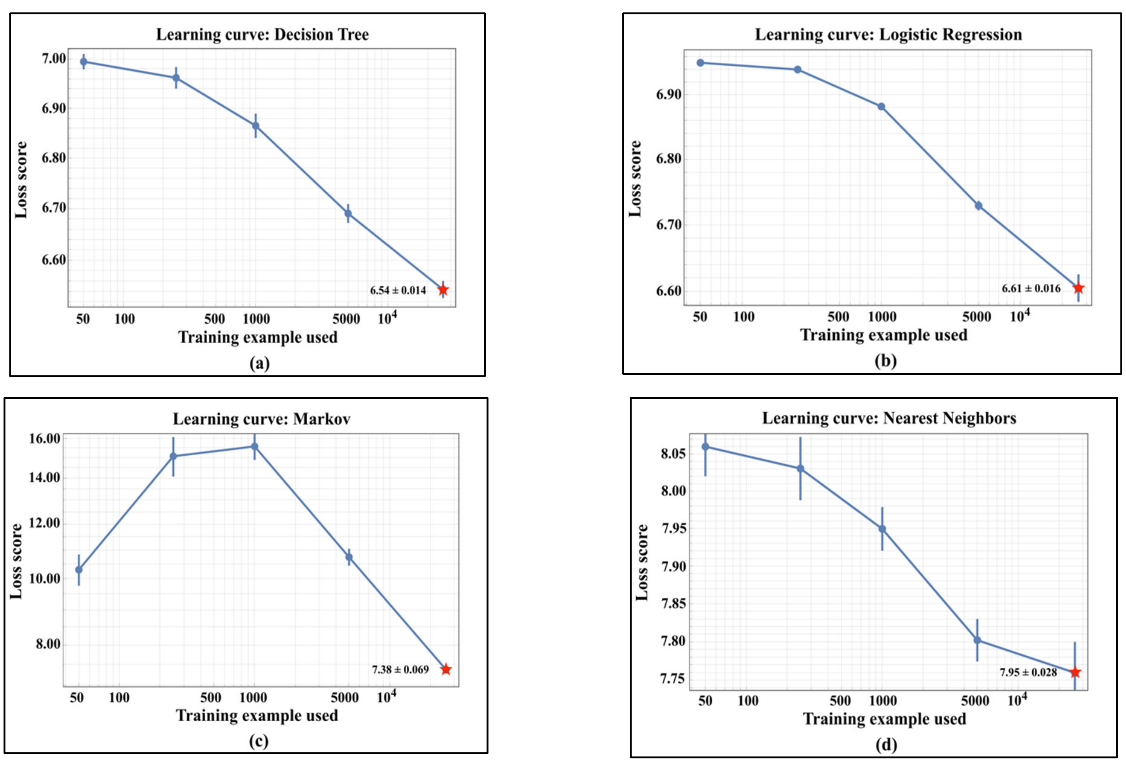
3. Results

This study used the THM data available on the Internet containing information about various herbal knowledge published over the years. This statement included common names and scientific names of herbs, dosing methods, herbal properties, side effects, notices, and references. The web still does not have any helpful information in the algorithm. This study must be done in advance to retrieve data used only in the tutorial algorithm. In the data pre-processing, the programming obtained all the herbal names from web pages. The repeat of the herbal name or the variable affected the program operation. The operation of a loop function involved iterating over the herbal name. This study had an algorithm to review duplicate listings and automatically eliminated them before the web scraping process. The algorithm found that the number of herbal names was 1,242, and the number of repeat cases was 182. Therefore, this study received a list of 1,060 herbal names. Web scraping algorithm was used with the web-based data sets and discovered many exciting results. It included 6,784,074 words of herbal knowledge information, which included the web scraping algorithm that converted the Thai language data into Unicode for communication with the computer. A sample of the Unicode data result is shown in Fig. 2.

\:0e0b\:0e36\:0e48\:0e07\:0e2d\:0e07\:0e04\:0e4c\:0e1b\:0e23\:0e30\:0e01\:0e2d\:0e1a\:0e2a\:0e48\:0e27\:0e19\:0e43\:0
e2b\:0e0d\:0e48\:0e43\:0e19\:0e19\:0e49\:0e33\:0e21\:0e31\:0e19\:0e08\:0e30\:0e1b\:0e23\:0e30\:0e01\:0e2d\:0e1a\:0e4
4\:0e1b\:0e14\:0e49\:0e27\:0e22 1,8 cineol (20-60%), ?-caryophyllene, ?-pinene, ?-terpineol, ?-terpinyl acetate (20-53%)

Fig. 2. THM information has been converted to Unicode for communication and processing with computers. The relevant English information was not converted to Unicode format.

The results of the keyword definition algorithm were able to determine the target symptom for 32,646 pairs of herbs. The pairing of the target symptom and the herb type was defined as the association rule for the ML training data set. In this study, nine ML methods were used to test learning from the training set. The Decision Tree, Logistic Regression, and Support Vector Machine method did not support the processing of Unicode character data. It consumed both times on high machine resources, resulting in the suppressed calculation. However, other ML methods can be trained, as displayed in the learning results in Fig. 3.



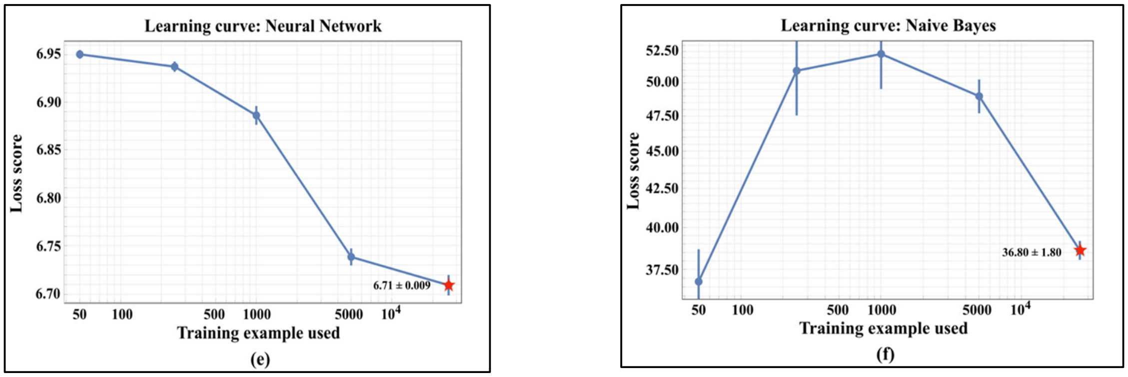


Fig. 3. The learning curve of ML methods shows the number of training examples of Unicode text data versus cross-entropy loss by algorithm processing. (a)-(e) display excellent learning as a low loss score. (f) shows poor learning results.

The ML needed to minimize errors. The objective function could determine the error measurement and be often called loss score or mean cross-entropy loss. The loss score provided the best performances in terms of classification. The success of the model will almost always be higher on the training data set. It will return a lower mean cross-entropy loss of the model, including a lower error between the training set and validation set of learning curves. The ML learning text Unicode results found a low loss score below 7.95 ± 0.029 from Decision Tree, Logistic Regression, Markov, Naïve Bayes, Nearest Neighbors, and Neural Network method.

The ability to receive targeted symptoms of the application was intended to check the THM knowledge. After the collection was complete, the knowledge extraction process would be entered. With the message, the agent would be invoked according to the training set boundary. The agent worked with input text from the boundary onwards until it reached the boundary and the syntax. The context would approach the combined part of the message. If the search was successful, the agent would perform pre-designed actions one-by-one in the order specified in advance.

3.1 Case Test: common flu symptoms

A randomized test to identify common flu symptoms (Family et al., 2012) using symptom phrases in the Thai language was conducted. The amount of THM data and the validity classified from the target symptom indication are shown in Table 1.

Table 1
The number of THM classifications obtained from the identification of flu symptoms phrases.

	Fever	Fever, Headache	Fever, Headache, Cough	Fever, Headache, Cough, Sore throat	Fever, Headache, Cough, Sore throat, Blocked	Fever, Headache, Cough, Sore throat, Phlegm, Runny nose
Decision Tree	482	612	482	482	482	482
Logistic Regression	24	293	138	201	187	192
Markov	442	6	4	5	13	8
Naïve Bayes	4	7	4	1	1	2
Nearest Neighbors	1	1	1	1	1	1
Neural Network	1056	1056	1056	1056	1056	1056

Identifying one of the targeted symptoms could be found to be a remedy from a variety of herbal remedies. The results of the classification of the ML might produce both the same and different THM data. Combining target symptom phrases might result in classification accuracy in some cases. The algorithm using a phrase pattern selected the words closest to the learning data and returned the closest THM data. In addition, phrase sequencing had the effect of probabilities in identifying the THM data. Nevertheless, a single phrase of the target symptom produced the most accurate results. The results of the sample studies on the identification of the target symptoms of the ML approach were broadly consistent with the favorable learning outcomes. Only the Naïve Bayes method corresponded to the effect of learning with high error values or high learning loss scores.

Moreover, considering the false results of the case study, the ML classification had an error in some parts of the Thai language. Thai characters repeating in phrases referred to different targeted symptoms. For example, the target symptom in the term “xxx” meant Symptom Type I. If added to “xxxyyy”, it meant Symptom Type II. Then, specifying a target symptom of these algorithms as “xxx” would result in Type I and Type II. However, the studies could resolve the error using the herbal properties system to double-check that Type I and Type II results met the target symptom “xxx”.

This ML system was implemented in the module of functional programming that supported all computer operating systems. The algorithm had a wide range of applications in a variety of forms of THM knowledge, as shown in Code Snippet 9. The user interface screen was displayed in one-line for any application. Users can determine any requirement in the bracket.

```
Diagnosis[Symptom Phrases]
HerbalProperties[Herbal Name]
SymptomSearch[Symptom]
```

Code Snippet 9. The input interface displays the one-line short form module of the diagnosis symptom system, herbal properties system, and symptom search system.

Diagnosis symptom system: This system would call the top probability herbal name of a target symptom. The user could identify symptoms or therapy in one or more phrases that they needed. The retrieval of the target symptom on each phrase-wise granularity could be given as an optional input.

Herbal properties system: This system would return the herbal properties. The clustering algorithm was produced for searching the herbal properties by using an herbal name. This module was an instrument for cross-checking and comparing with the diagnosis symptom system.

Symptom search system: Targeted symptoms and treatments were provided as ML training data. Therefore, it could take advantage of the training data set. The system would only identify one target symptom from the training data so that all types of THM knowledge can be displayed.

4. Discussion and conclusions

The functional programming always efficiently imported data and interpreted the textual input form as an actual expression to evaluate. The import of data processing was the first fundamental importance of this study. The developer had to learn the inspect tool (development mode) of the web browser. In addition, the inspect tool may be called by other names such as developer tool, inspect element tool, and web inspector in some cases. The inspect tool was a valuable tool for web developers, as it let the programmer check and edit the HTML and CSS of the page displayed in the browser. The web URL values were embedded under the web page. The inspect tool can therefore help to acquire the web URL values of the THM web data for this successful study.

The disadvantage of this study was that web scraping produced a lot of unnecessary data. Data on cleanliness data was the second essential aspect of this study. Data cleaning included recognizing and fixing contaminated or inaccurate records of a data set. This refers to classifying remote parts of the data and then deleting the dirty data. Usually, cleansing data is challenging and time consuming, even when this study had a strategy in place by using keywords. Symptom and therapy keywords were essential to extract the THM knowledge from imported data sets. The process selected only relevant information and ignored irrelevant information. It performed the job of cleansing and tuning data much more comfortably, faster, and more precisely than searching for irrelevant data and removing the dirty data.

THM has become successful and has been viewed as a beneficial medical treatment. Reporting operational decisions through in-depth information from past data is an example of evidence-based medicine. This study demonstrated an easy way to extract THM knowledge on the web by using AI. The algorithm can distinguish the different types of Internet data forms from therapy (e.g. treat, relieve, cure, neutralize, etc.) and the keywords of symptoms (e.g. cough, mucus, scald, stomachache, etc.). The key success of this ML programming depended on the training set from the THM data set. The programming included highly automated functions like organizing functions based on an association data set and ML classification methods. ML techniques returned the value of the herbal name that covered complex information associations on the website. Programming allowed the ML system to access complex solutions similar to a doctor who carefully prepares evidence to reach a reasonable conclusion. This study recommends that the effects of valid ML analysis should be considered together. Defining decision-making processes together resulted in higher accuracy. However, these systems can observe and process almost limitless information, unlike a single doctor.

The growing body of healthcare information has been released using AI and ML in health care (Kilic, 2019). AI can create sophisticated diagnostic insights similar to a health professional. Some researchers have concluded their article by wondering whether medical professionals are willing to trust AI technology that aims to increase the efficiency of clinical communication in their practices (Ryan et al., 2019). Medical professionals may trust AI if they can automate and improve the efficiency and effectiveness of the diagnosis (Jha & Topol, 2016; Yu et al., 2016; Choi et al., 2016). Nonetheless, using AI and human interaction comes with new challenges. Researchers working on developing AI technologies may not be aware of the extent of critical personal data available in behavioral research settings. AI has many possible uses in medicine and science. It remains to be seen where these things will occur. However, certainty lies in the inevitable change in technology. Therefore, all physicians need to be aware of current advances in AI as they tend to influence future healthcare offerings. They need to know and be confident in which cases to use AI-based recommendations to optimize clinical analysis. Essential concerns for physicians tend to include attention to the effects of AI on patient confidentiality. The key to success is tailored multidisciplinary collaboration and good engagement with relevant stakeholders. Nevertheless, this study did not affect personal and sensitive data. The ML learning process used publicly available THM data only.

Historically, clinical experts manually extracted most of the free-text clinical records data for archiving, retrieval, and analysis. This is particularly relevant for chronic diseases because clinical records are predominant over structured data. Several previous studies have looked at free-text or combined clinical records with structured data to predict and model patient trajectories. Such results may involve using data analysis methods and algorithms (such as shallow classifiers and rule-based approaches) that cannot capture transient relationships and long-term factors between clinical variables. This may be due to insufficient data to train the algorithm, and embedding methods have only recently been developed. SVM and Naïve Bayes algorithms are often used for machine learning-based tasks or in combination with rule-based methods. This may be due to the popularity of these algorithms. The same is true because Naïve Bayes is a relatively simple algorithm, where a small amount of training data is required. These tasks are also limited to processing in English only (M & Sagar, 2019). However, in this study, the ML approach used a key phrase system to respond to multiple symptoms. It no longer needed to take into account structured or unstructured big data. The study can solve language constraints. The functional programming used Thai Unicode as a medium for interpreting the meaning of data. This study was an alternative way to develop an application that supports every language in the world.

Regarding herb consumers and patients, AI and ML could be utilized to support their decision making on herbal supplements and medicines. Usually, most people rely on herbs for common disease treatment, health support, and disease prevention. Therefore, the precise and correct information using an integrated approach to herbal medicine could effectively increase appropriate drug use and consumer safety. Besides, the information combined and organized via AI and ML might help health science researchers discover new herbal candidates for specific diseases. The manipulation of diverse data and recategorized data should shorten the drug discovery time and minimize the random searching of new drug candidates from natural resources.

For future work, adopting a Unicode algorithm to develop ML can be used with all written languages, including the development of ML algorithms into mobile applications which will make the boundless world of knowledge more accurate. The use of AI and ML for health communication is an emerging area and has excellent potential for further development. The possibility of analyzing large amounts of data in cost-effective practices is enticing, while the potential to offer individual opinions, objectives, and repetitive feedback to many learners is exciting. Often, individuals may feel skeptical about requesting personal trainer assistance due to schedule restrictions or the stigma involved in finding practice time. An AI-driven training system can be used anytime and anywhere without limitations, and it can eliminate impurities related to finding practical methods or helping to communicate more effectively. However, to this day, most applications have been limited to targeting only a few concepts and have not examined the complex relationships between variables. It is still important to establish the evidence of this method's feasibility, reliability, acceptance, and effectiveness before it is put into practice.

Acknowledgement

The authors acknowledge with gratitude the financial support by Walailak University. An individual research grant was provided under the contract No. WU63207.

References

A Kilic, A. (2019). Artificial Intelligence and Machine Learning in Cardiovascular Healthcare. *The Annals of Thoracic Surgery*, *5*, 1323–1329.

Cao, C., Wang, H., & Sui, Y. (2004). Knowledge modeling and acquisition of traditional Chinese herbal drugs and formulae from text. *Artificial Intelligence in Medicine*, *32*(1), 3–13.

Chang, S. J., Yang, E., Lee, K.-E., & Ryu, H. (2020). Internet health information education for older adults: A pilot study. *Geriatric Nursing*, *42*(2), 533–539.

Chen, Z., Cao, Y., He, S., & Qiao, Y. (2018). Development of models for classification of action between heat-clearing herbs and blood-activating stasis-resolving herbs based on theory of traditional Chinese medicine. *Chinese Medicine*, *13*(1), 1–11.

Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F. & Sun, J. (2016). Doctor AI: Predicting clinical events via recurrent neural networks. in *Proceedings of Machine Learning for Healthcare*, *56*, 1–18.

Du, H., Yang, J., King, R. B., Yang, L., & Chi, P. (2020). COVID-19 Increases Online Searches for Emotional and Health-Related Terms. *Applied Psychology: Health and Well-Being*, *12*(4), 1039–1053.

Family, A., Fashner, J., Ericson, K. & Werner, S. (2012). Treatment of the common cold in children and adults. *American Family Physician*, *86(2)*, 153–159.

Gianfredi, V., Sandro, P., & Santangelo, O. E. (2021). What can internet users' behaviours reveal about the mental health impacts of the COVID-19 pandemic? A systematic review. *Public Health*, *198*, 44–52.

Han, N., Qiao, S., Yuan, G., Huang, P., Liu, D., & Yue, K. (2019). A novel Chinese herbal medicine clustering algorithm via artificial bee colony optimization. *Artificial Intelligence in Medicine*, *101*, 101760.

Jha, S., & Topol, E. J. (2016). Adapting to Artificial Intelligence. *JAMA*, *316*(22), 2353.

Kanjanahattakij, N., Kwankhao, P., Vathesatogkit, P., Thongmung, N., Gleebbua, Y., Sritara, P., & Kitiyakara, C. (2019). Herbal or traditional medicine consumption in a Thai worker population: pattern of use and therapeutic control in chronic diseases. *BMC Complementary and Alternative Medicine*, *19*(1), 1–9.

Kee, J. L., Hayes, E. R., & Mccuistion, L. E. (2015) *Pharmacology: A patient-centered nursing process approach*. Elsevier.

M, H., & Sagar, Dr. B. M. (2019). Smart Health Care Implementation Using Naïve Bayes Algorithm. *International Journal of Innovative Research in Computer Science & Technology, 7*(3), 90–93.

MedThai, (2021). *Med pedia.* Medthai.

Mikkelsen, G., & Aasly, J. (2002). Manual semantic tagging to improve access to information in narrative electronic medical records. *International Journal of Medical Informatics*, *65*(1), 17–29.

Murdoch, T. B., & Detsky, A. S. (2013). The Inevitable Application of Big Data to Health Care. *JAMA*, *309*(13), 1351.

Ryan, P., Luz, S., Albert, P., Vogel, C., Normand, C., & Elwyn, G. (2019). Using artificial intelligence to assess clinicians' communication skills. *BMJ*, *364*, l161.

Serçekuş, P., Değirmenciler, B., & Özkan, S. (2021). Internet use by pregnant women seeking childbirth information. *Journal of Gynecology Obstetrics and Human Reproduction*, *50*(8), 102144.

Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. *JMIR Medical Informatics*, *7*(2), e12239.

Sun, X., & Qian, H. (2016). Chinese Herbal Medicine Image Recognition and Retrieval by Convolutional Neural Network. *PLOS ONE*, *11*(6), e0156327.

Wang, N., Li, P., Hu, X., Yang, K., Peng, Y., Zhu, Q., Zhang, R., Gao, Z., Xu, H., Liu, B., Chen, J., & Zhou, X. (2019). Herb Target Prediction Based on Representation Learning of Symptom related Heterogeneous Network. *Computational and Structural Biotechnology Journal*, *17*, 282–290.

Wen-Xiang, D., Jian-Ping, Z., Jing, L., Zhi-Ying, Y., Hua-Ying, W., Zhong- Hua, Y., Yi-Ge, Z., Wen-An, Z., & Hui-Yong, H. (2019). Research on Text Mining of Syndrome Element Syndrome Differentiation by Natural Language Processing. *Digital Chinese Medicine*, *2*(2), 61–71.

Wolfram, S. (2017). *An elementary introduction to the Wolfram langauge*. Wolfram Media, Inc.

Yu, K.-H., Zhang, C., Berry, G. J., Altman, R. B., Ré, C., Rubin, D. L., & Snyder, M. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature Communications*, *7*(1).

Yu, T., Li, J., Yu, Q., Tian, Y., Shun, X., Xu, L., Zhu, L., & Gao, H. (2017). Knowledge graph for TCM health preservation: Design, construction, and applications. *Artificial Intelligence in Medicine*, *77*, 48–52.

Zhou, X., Chen, S., Liu, B., Zhang, R., Wang, Y., Li, P., Guo, Y., Zhang, H., Gao, Z., & Yan, X. (2010). Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. *Artificial Intelligence in Medicine*, *48*(2-3), 139–152.