Development of Jamu formula prediction system module of Ijah analytics based on pharmacology activity and particular efficacy target

N S Ramadhanti1, W A Kusuma1,3, and R Heryanto2

1Department of Computer Science, Bogor Agricultural University, Bogor, West Java, Indonesia
2Department of Chemistry, Bogor Agricultural University, Bogor, West Java, Indonesia e-mail: ananta@apps.ipb.ac.id
3Tropical Biopharmaca Research Center, Bogor Agricultural University, Bogor, West Java, Indonesia

Abstract. Indonesia is one of countries with abundant amount of herbal plant species. These plants have potential to be used as an ingredient of alternative medication, which is known as Jamu. There have been already many previous research related to the topic of formulating jamu efficacy using network pharmacology that construct link among plants and disease. Moreover, some researchers were more interested in predicting the interaction between chemical compound and protein to formulate Jamu. Commonly, Jamu formula contains four basics of pharmacology activities including analgesic, anti-inflammatory, antibacterial, and target pharmacology activity. Ijah has not implemented these four essential characteristics yet for predicting jamu formula. This research implemented new molecular Jamu formulation composed by some active compounds from plant with analgesic, anti-inflammatory, antibacterial and target pharmacology activity characteristic. The result of this study showed a list of plants which consisted of herbal compounds that could be used to formulate jamu as an alternative to synthetic medicine.

Keywords: Ijah, Jamu formula, pharmacology

1. Introduction
Indonesia is one of the countries in the world which has the highest number of herbal plant species. Parts of herbal plants, such as root and leave, can be processed into a concoction so that it has efficacy to maintain health and to cure many kind of illness. The concoctions are often used by people in some countries like China and Indonesia. In Indonesia this concoction is called ‘Jamu’. According to Purwaningsih (2013) [1], in accordance with the policy of Indonesian Ministry of Health in 2010 [2], scientific analysis of Jamu is required to support the truth about the efficacy and safety of Jamu.

In 2010, there was a statistical research about the correlation of Jamu formula and its efficacy using bi-plot based on Robust Principal Component Analysis (PCA) by Afendi et. al. (2010) [3]. The result showed the direct effect on disease from plants that usually become the main ingredient of certain Jamu. Afendi et. al. (2012) [4] then developed a classification model of jamu formula with Partial Least Square Discriminant Analysis (PLS-DA) method with accuracy of 71.6%. Furthermore, Fitriawan (2013) [5] compared the accuracy of Afendi’s work to the one that used Support Vector Machine (SVM) method. Based on his research, implementing data cleaning process in advance, the accuracy was improved to...
95.43%. The undersampling and oversampling process can also improve the accuracy since the dataset is unbalanced. The studies conducted by Afendiet al. (2012) [4] and Fitriawan (2013) [5] then were integrated in a Jamu Formula Prediction System Ijah (Indonesia Jamu Herb) with a hypothesis that generally jamu is composed by plants with four particular characteristics, such as analgesic, antibacterial, anti-inflammatory, and efficacy. Next in 2016, Amir [6] conducted research about jamu formula prediction for diabetic disease with graph mining.

Conventional medicine commonly uses single component - single target approach aiming to find the most effective compound for the certain protein disease target. Furthermore, currently new diseases are exists and become more complex [7]. Therefore, we required an approach to formulate drugs that can handle the complex diseases. Gu et. al. (2011) [8] conducted research about relation of compound in drugs and protein in type 2 diabetic disease and formulate Traditional Chinese Medicine (TCM) in the molecular level with multi-component – multi-target approach using compound-protein network. Further research was conducted to find Jamu formula for the same disease. In this research, Jamu formula was formulated from the classification of plant compound and significant protein related to type 2 diabetic diseases. Characteristic for this classification was obtained from molecular docking simulation of herbal plant with target protein [9].

Barkan (2017) [10] developed Ijah Webserver that can predict relationship between plant - compound and protein - disease. Next, Kurnia (2017) [11] conducted further research which predicts link between compound and protein in IJAH using Bipartite Local Model Network Interaction-Profile Inferring (BLMNII) for compound - protein interaction network. The first version of Ijah, which developed before Ijah Webserver, is able to formulate jamu using plants with analgesic, anti-inflammatory, antibacterial, and desired efficacy characteristics while the research by Barkan (2017) [10] formulated jamu with considering the link between plant-compound and protein-diseases. This research would combine the two versions of Ijah to formulate jamu with four compounds as input. These four compounds would be adjusted with four characteristics including analgesic, anti-inflammatory, antibacterial, and desired efficacy.

2. Materials and methods

2.1 Materials

Dataset used in this research is compound data which has four characteristics, including analgesic, anti-inflammatory, antibacterial, and desired efficacy. This dataset was obtained from Ijah database. Compound and protein interaction data was obtained from Ijah database which was predicted with BLMNII based on research by Kurnia (2017) [11]. Numbers of data used in this research can be seen on table 1. Compound with certain efficacy data consists of 87 activities which were grouped into 20 sub-groups on table 2.

| Table 1. Numbers of data in Ijah Database. |
|-------------------------------------------|
| Data                                      | Total          |
| Plant                                     | 4984           |
| Compound                                  | 17277          |
| Protein                                   | 3334           |
| Disease                                   | 4401           |
| Plant and compound interaction            | 32461          |
| Compound and protein interaction          | 7945           |
| Protein and disease interaction           | 5066           |
| Compound similarity                       | 6425730        |
| Protein similarity                        | 2556150        |
Table 2. Efficacy and its activities numbers.

| Sub-group                                          | Activities |
|----------------------------------------------------|------------|
| Bacteria, Viruses, Fungi and Parasites              | 7          |
| Bites and stings                                   | 1          |
| Bones                                              | 3          |
| Bowles and Bladder                                 | 6          |
| Cancer                                             | 4          |
| Chest and Lungs                                    | 6          |
| Ear, Nose and Throat and Eyes                      | 3          |
| Fevers                                             | 2          |
| General Well Being                                 | 2          |
| Heart and Blood                                    | 7          |
| Infectious Diseases                                | 1          |
| Liver, Kidneys, and other internal bits            | 6          |
| Nerves and Muscles                                 | 2          |
| Other medicinal uses                                | 13         |
| Pain Relief                                        | 3          |
| Sex and Reproduction                               | 7          |
| Skin, Hands and Feet                               | 4          |
| Stomach                                            | 5          |
| Treatment of Wounds and Bruises                    | 3          |
| Whole Body                                         | 2          |

2.2 Methods

This research was conducted according to the sequence of processes including data collection, system design, implementations, system testing, and deployment. Firstly, data was collected by scrapping and manually taken from various sources of public domain databases, such as PubChem, DrugBank, and KEGG to get compound data which was grouped into some particular activities. Next, this dataset was combined with that of stored in Ijah database showed in table 1. The interaction data between compound and protein was also taken from Ijah as prediction results of BLMNII algorithm (figure 1).

Then, to ensure that every chosen compound has at least one connection with plants, searching was done to look for herbal compounds with high similarity score. This similarity compound data was obtained from Ijah database using similarity measure of SIMCOMP [12].

This system design includes designing input and output interface which referred to the existing of Ijah. The interface was designed with Ijah webservice template and the input component was adjusted according to the first version of Ijah which providing the option menu for choosing four plants with certain characteristics.

In the implementation, the new module was developed based on the first version of Ijah. This program code of Ijah is available on page github.com/tttor/csipb-jamu-prj. The database should also be downloaded and ran in a local server. After the program ran locally, a new page should be made to be developed as a new module.

To validate the functionality of a new module, system testing was conducted based on the black box testing technique. Afterwards, there was another test such as comparing the results to the literatures to validate the result of herbal plants prediction as candidate of Jamu formula. The list of plants and disease generated from the prediction system was compared to those of on articles, paper or any other written research results.

Finally, after testing, the application that has already run locally was placed into the server. The new data that would be used also added to the server database. Then, the server was configured so that the program run and could be accessed through Internet.
Algorithm part 1: BLMNII
Input: Interaction matrix (A), Compound similarity matrix (S_s), Protein similarity matrix (S_p), i, j
Output: Pre_ij
1 pre_i = NII_Prediction (A, S_s, S_p, i, j)
2 pre_j = NII_Prediction (A, S_s, S_p, i, j)
3 pre_ij = max(pre_i, pre_j)

Algorithm part 2: NII_Prediction
1 if compound-i has an interaction:
2 I = compound-i label according to value in A
3 else
4 I = compound-i label by adding the S_s multiplication to A for each compound-i
5 Smith-Waterman, Gaussian and SIMCOMP kernel calculation.
6 Calculate kernel combination
7 Model = SVM_Train (Kernel, I)
8 if protein-j has an interaction:
9 Pred = SVM_Test(Model, Kernel[j])
10 else:
11 Pred = SVM_Test(Model, S_p[j])
12 Return Pred

Figure 1. BLMNII algorithm [13].

3. Results and discussion
This research yielded an additional module to improve the first version of Ijah by adding the option menu for choosing four plants with certain characteristics such as analgesic, anti-inflammatory, antibacterial, and desired efficacy activity. This development referred to a hypothesis that every Jamu formula should be obtained based on their pharmacology activities pattern. Furthermore, as mentioned before, the dataset was collected both by scraping and searching manually.

Scraping was done using Python programming language with Beautifulsoup4 library. Firstly, the HTML code from the pages that contain compound with particular activity was analyzed. The scraping script extracted id and name of the compound. To know the exact id and name on the HTML code, the results were inspected by conducting ‘Inspect Element’. The scraping started with page parsing using lxml parser. This procedure was done to parse the document for further processing. Moreover, this information with certain characteristics is the name of compound stored in DrugBank database that obtained by analyzing the HTML code. Those lists of compound then were transformed into a data frame using pandas package and saved into CSV format.

In addition, for the uncategorized activity on DrugBank, the compounds were searched and collected manually. As an example, one of the uncategorized activities is diuretic. Collecting data manually started using keyword ‘diuretic’ in DrugBank search box. The id and name of the compounds listed then were copied one by one to a document.

Next, the data collected from scraping and manual process was combined into one document that consisted of activity id and compound id in CSV format. After collecting data, to synchronize the new dataset with the existing data on Ijah, compounds that did not exist on Ijah were removed. The numbers of collected data before and after removal can be seen on table 3.
System design was initialized by designing interface for input and output of the web page. The input interface was improved by combining the interface of the first version of Ijah and that of the Ijah webserver showed in figure 2. The input interface was designed to accommodate user to choose the combination of four compounds which representing different characteristics. However, because there are 87 activities, thus, activities are divided into 20 sub-groups as showed in table 2. The design of input interface could be seen on figure 3.

The formulation is represented as a network that connects plant-compound-protein-disease as showed in figure 4. On the network, there were plants which connected to herbal compounds, represented by Herbal c. Moreover, the Herbal c was connected to four selected compounds, represented by Compound 1, 2, 3, 4 which respectively have the characteristics of analgesic, antibacterial, anti-inflammatory, and activity of a particular target.
The implementation was conducted by modifying code of Ijah on ‘webserver/src/app/ijahv1’. The first modification was conducted to change the input interface and to remove the use-case example function. The input interface was modified by adjusting according to the design shown in Figure 4. The modification result could be seen at figure 5. Input box was used to choose the compounds that belong to the analgesic, antibacterial and anti-inflammatory compound, respectively. The compound with target efficacy could be filled after selecting activity which could be selected after selecting the efficacy. Before continuing the modification, the new data of efficacy, activity, and compound-activity data were imported to PostgreSQL. The typescript (.ts) and PHP file were modified to adjust the HTML file with the new added dataset. The adjustment was conducted to renew the selection on input box. For instance, on analgesic input box, the selection is a list of analgesic compound obtained from the collecting data.

**Figure 4.** Connectivity among plant-compound-protein-diseases.

**Figure 5.** Implemented input interface.
Next, the output interface was designed and implemented. This interface, consist of relational network of plant, herbal compound, compound, protein, disease, was modified by modifying ‘searchFromDrugSide’ function and other functions related to it. The modification of this interface could be seen at figure 6.

**Figure 6. The output interface.**

System testing was performed by using black-box testing technique to validate the functionality of the new module. System testing began with choosing compounds to be tested. Based on analysis of the collected data, the total number of compounds related to plants, including direct or indirect relationship, is 58 compounds.

This testing used the compounds of morphine, vancomycin, indomethacin, quinine as samples which have analgesic, antibacterial, anti-inflammatory, and certain characteristic, respectively. The certain efficacy activity selected was analgesic activity which belonging to the pain relief sub-group. A snip of
the generated graph could be seen on figure 6. These four compounds were chosen for the reason of that they are used as ingredients of some medicine.

System searching results showed that all four compounds were herbal compounds. The results of the connectivity data showed that there were twelve types of plants that contain herbal compounds that were similar to the four compounds selected (see table 4). Table 4 showed that all plants containing quinine compounds were only different in the species level, however belonged to the one genus, namely cinchona or commonly known as quinine, except one plant that belong to the Olea europaea species. Then the four selected compounds were also associated with 20 proteins and 34 diseases.

Two diseases were chosen in order to validate the results. Both diseases are Inflammatory Bowel Disease (IBD) and Rheumatoid Arthritis. Both of them come from proteins that are connected to the selected target compound, namely quinine. IBD is a disease with symptoms of intestine that feels hot. Therefore, quinine can be used as the main ingredient of jamu to relieve IBD. This is proofed by Sharma’s (2017) [15] study which stated that Cinchona officinalis (quine) was an ideal homeopathic drug that was natural for overcoming symptoms of excessive fluid loss in IBD. Moreover, Rheumatoid arthritis (RA) is an autoimmune disease that causes inflammation of the joints. Based on Ma and Jiang’s research (2016) [16], one of the effective compounds to treat RA is quinine which can be found in quinine plants. The making of jamu with the four components of compounds could use the Amycolatopsis orientalis, Alpinia officinarum (galangal), Stephania cepharantha and Cinchona (quine) plants showed in table 4.

![Table 4](image)

Further analysis on the compounds was obtained from PubChem site [14]. On PubChem [14], it is explained that beside its analgesic characteristic, morphine also has narcotic property which used to relief a severe pain. Due to its narcotic property, the usage of morphine in medicine usually is very little. On PubChem [14], vancomycin is explained as antibacterial compound with very little dose of usage. Indomethacin compound has anti-inflammatory and analgesic property. On PubChem [14], it is explained that Indomethacin is often used as cardiovascular agents, cyclooxygenase inhibitors, gout suppressants, and tocolytic agents. Aside from being a non-narcotic analgesic compound, quinine is better known as an agent used for antimalarial. In PubChem[14] it is explained that Quinine is also commonly used to overcome muscle spasms.
Finally, the new module that has been tested successfully was installed to the server of Biofarmaka with some adjustment and dependency installations to make sure that the system is running in local server as well. All dataset used in this study is also been added to the server database. The system could be accessed through IPB network on http://172.24.10.200:4200.

4. Conclusion
This study applied network pharmacology approach to the plants-compound-protein-diseases network. The compounds were characterized into four groups such as analgesic, anti-inflamatory, antibacterial, and compound with specific target activity characteristic that are generally contained in existing jamu formulas. In this research, the amount of data collected and available data was very different. This research successfully developed a new module to accommodate the four characteristics of group of compounds and integrate this module to the Ihjah webserver. The black box testing showed that the functionalities of module could be implemented and could be run successfully. However, there is still drawback in this study that could be overcome in the next study. Only 0.1% of the compounds data which were collected that have a relationship with plants. Further research could add more dataset, such as data on compounds, plants, proteins, diseases, and interaction data, so that the future research on Jamu formulations could be more comprehensive.

References
[1] Purwaningsih E H 2013 Jamu, obat tradisional asli Indonesia pasang surut pemanfaatannya di Indonesia eJKI [internet] Available on: http://journal.ui.ac.id/index.php/eJKI/article/view/2065
[2] [Kemenkes] Kementerian Kesehatan Republik Indonesia 2010 Riset Kesehatan Dasar 2010 [internet]. Available on: http://www.diskes.baliprov.go.id/files/subdomain/content/diskes/Januari%202015/%20RISKESDAS%202010.pdf
[3] Afendi FM, Darusman L K, Hirai A, Altaf-Ul-Amin M, Takahashi H, Nakamura K and Kanaya S 2010 System biology approach for elucidating the relationship between Indonesian herbal plants and the efficacy of jamu IEEE Int. Conf. on Data Mining Workshops 661–668
[4] Afendi F M, Darusman L K, Morita A H, Altaf-Ul-Amin M, Takahashi H, Nakamura K, Tanaka K and Kanaya S 2012 Efficacy prediction of jamu formulations by PLS modeling Curr. Comput. Aided Drug Des. 9(1) 46–59
[5] Fitriawan A 2013 Sistem klasifikasi khasiat formula jamu dengan metode support vector machine [Thesis] (Bogor: Institut Pertanian Bogor)
[6] Amir F 2016 Prediksi formula jamu untuk penyembuhan penyakit diabetes dengan teknik graph mining [Thesis] (Bogor: Institut Pertanian Bogor)
[7] Chen X, Yan C C, Zhang X, Zhang X, Dai F, Yin J and Zhang Y 2015 Drug–target interaction prediction: databases, webservers, and computational model Briefings in Bioinformatics 17(4) 696–712
[8] Gu J, Zhang H, Chen L, Xu X, Yuan G and Xu X 2011 Drug–target network and polypharmacology studies of a Traditional Chinese Medicine for type II diabetes mellitus Comp. Bio. Chem. 35(5) 293–297
[9] Wulandya S A, Afendi F M and Sumaryada T I 2017 Classification of active compounds of type 2 antidiabetic medicinal plants based on molecular docking Int. J. Sci. Eng. Res. 8(9) 583–588
[10] Barkan H Z 2017 Pengembangan sistem informasi prediksi konektivitas drug-target “Ijah Webserver” dengan metode prototyping [Thesis] (Bogor: Institut Pertanian Bogor)
[11] Kurnia A 2017 Prediksi formula jamu berkhasiat menggunakan teknik link prediction dari jejaring bipartite penyaya aktif dan protein [Thesis] (Bogor: Institut Pertanian Bogor)
[12] Hattori M, Okuno Y, Goto S and Kanehisa M 2003 Heuristics for chemical compound matching. Gen. Inf. 14 144–153
[13] Mei J P, Kwoh C K, Yang P, Li X L and Zheng J 2013 Drug–target interaction prediction by learning from local information and neighbors Bioinformatics 29(2) 238–245
[14]  [Pubchem] Pubchem - Open Chemistry Database 2018 Compound Summary for CID 3034034 [internet] Available on: https://pubchem.ncbi.nlm.nih.gov/compound/Quinine
[15]  Sharma M P 2017 A homeopathic approach to inflammatory bowel disease (crohn’s and colitis) Int. J. Rec. Adv. Sci. Tech. 4(3) 1-7
[16]  Ma Q and Jiang J G 2016 Functional components from nature-derived drugs for the treatment of rheumatoid arthritis [review] Curr. Drug Targets 17(14) 1673-1686

Acknowledgements
This research is supported by Bogor Agricultural University and Ministry of Research, Technology and Higher Education, Indonesia, under Competitive Research Grant from Directorate of Higher Education, Indonesia, 2017, contract no. 1718/TT3.11/PN/2018.