A Conceptual Model for Classification of Biomedicine Research

R Hardiyati¹*, M Silalahi¹, M Amelia¹, I M. Nadhiroh¹, R Rahmaida¹, and T Handayani¹

¹Research Center for Development of Science and Technology, Indonesian Institute of Sciences, Jakarta 12710, Indonesia

*E-mail: rhardiyati@gmail.com

Abstract. Many researches in biomedicine had been done in Indonesia, but only a few research results go downstream or applied in the industry. Mapping the potency of research downstream in biomedicine is carried out by classifying research outputs in biomedicine. One of the research outputs that can be used and easy to get is research publications. An effective and efficient classification of biomedicine research can be done using a computerized predictive model based on text mining of research publications. The objective of this study is to introduce a conceptual model in classification of biomedicine research based on research publications. The model of classification is defined based on findings in in-depth interview with researchers and industries about biomedicine discovery process, and a literature study. The relation of the classification model with the technology readiness level (TRL) model is analyzed by mapping. This conceptual model of classification of publication in biomedicine research to research downstream potency, developed as part of a research framework to build a computerized predictive model based on research publications.

Keywords: biomedicine, conceptual model, predictive model, research downstream, research publications

1. Introduction

Indonesian resources, as one of megabiodiversity country, has potency that can be developed as a source of biomedicine/herbal medicine/phytopharmaca [1,2]. Many researches about biomedicine had been done in Indonesia, but only few result of the research that go downstream or applied in industry. In the other hand, developing medicines to find and produce a new medicine is a long and complex process. It takes at least 10-12 years complete journey from initial research to the marketplace for a new medicine, and it comes from thousands or million compounds from the beginning [3,4].

To mapping the potency of research downstream in biomedicine through classifying research output in biomedicine research. Research output that easy to get and will give the pictures of research result is research publications. Classification of research publications in biomedicine will show the potency of research downstream that had been done in Indonesia. And rich and detail information stored in biomedicine research publications have never been analyzed for research mapping. An effective and efficient classification of biomedicine research can be done using a computerized predictive model based on text mining of research publications. The objective of this study is to introduce a conceptual model in classification of biomedicine research based on research publications. This conceptual model of classification of publication in biomedicine research to research downstream
potency is part of a research framework developed to build a computerized predictive model based on research publications.

2. Materials and methods

2.1. Data collecting
This study is based on research publication from journal indexed by Scopus and Google Scholar that published until March 2017. Keywords that been used in research publications from journal indexed by Scopus and Google Scholar are combination between “biomed*”, “drug”, “herbal medicine”, and “medical biology” and keyword “Indonesia”, in journal title with type article. But from Google Scholar, this study only uses articles written in English. Articles about biomedicine also obtained by filtering from data in research about biodiversity Indonesia [5]. There are 209 articles from Scopus, 662 articles from Google Scholar, and 387 articles from biodiversity research.

2.2. Literature study and in-depth interview
Literature study and in-depth interview is done to get as much as possible knowledge about R&D new medicine. Literatures about R&D new medicine are not only gets from research publications that written by researchers, but also popular publications that provided by biopharmaceutical industries or drug agency. This study also doing in-depth interview with biomedicine researchers from academia and industry to get information about stages in new medicine process discovery. We collected information from academia and industry so this study gets different point of views in new medicine process discovery. Researchers from academia good in breakthrough in finding new compound, meanwhile researchers from industry good in standard in every stage of new medicine process discovery.

2.3. Screening
Screening in this study has a purpose to give a label in each article, whether an article has topic in new biomedicine process discovery or not. If include, then in which stage. In screening step also to gather keywords that demonstrated the stage of article in new biomedicine process discovery. A set of keywords is used to build corpus that contains a distinctive entity for every stage in new biomedicine process discovery. Screening step is done based on knowledge and informations from literature study and in-depth interview.

2.4. Built the conceptual model for classification
Based on knowledge and information from literature study and in-depth interview and also result from screening step, conceptual model for classification of biomedicine research is built. In this step, we propose the generalization in stage of new medicine process discovery.

3. Results
Biomedicine research in Indonesia have been done decade ago. At least from collecting data, the oldest article about biomedicine is from 1982. And also, Indonesia as one of megabiodiversity country has so many potential resources that can be developed as biomedicine source. But there are only few biomedicine researches that can be applied in industries or reach to research downstream.

Mapping the potency of research downstream in biomedicine can be done by classifying biomedicine research. Biomedicine research not only resulting research publication, but also patent. But from research publications, there are rich and detail information that can shows the picture of biomedicine research that have been done in Indonesia. Because of the large amount of research publications need to be analysed, so classification of biomedicine research can be done using a computerized predictive model based on text mining of research publications. This conceptual model of classification of biomedicine research, developed as part of a research framework to build a computerized predictive model based on research publications.
Process to produce a new medicine is takes a long time and big cost. A new medicine comes from thousands of compounds that had been resulted in initial research. Based on literature study in process for research and developing (R&D) new medicine and also based on in-depth interview with expert from biomedicine researchers and biomedicine industries, this study propose the conceptual model for classification of biomedicine research.

R&D process of new medicine is generally divided into 4 big steps:

1. Basic Research
2. Drug Discovery
3. Pre-clinical
   a. in vitro
   b. in vivo
4. Clinical Trials
   a. Phase I
   b. Phase II
   c. Phase III
   d. Phase IV

3.1. Basic research
This stage is initial stage of R&D of new medicine. In this stage researchers usually learning more about diseases and identifying potential targets. In this stage, through increasing the knowledge about disease at molecular level so does discovering and developing the potential medicines. Also in this stage [3]. At this stage in, researchers may be found thousands of potential compounds candidates for development as a medical treatment.

After early testing, such as examining mRNA/protein, genetic associations, and phenotypic screening, only a small number of compounds look promising and proceed for further study [4]. Drugs fail in the clinic for two main reasons; the first is that they do not work and the second is that they are not safe. As such, one of the most important steps in developing a new drug is target identification and validation. Validation techniques range from in vitro tools through the use of whole animal models [4].

3.2. Drug discovery
A small number of promising compounds, in this stage researcher seek one lead compound or a promising molecule. They do in some ways, including try to creating a molecule in living or synthetic material, using high-throughput screening techniques to select a few promising possibilities from thousands potential candidates, identifying compounds found in nature, and using biotechnology to genetically engineer living systems to produce disease-fighting molecules [3,6].

After one lead compound or a promising molecule identified, researcher conduct experiments to gather information such as how it is absorbed, distributed, metabolized, and excreted, best dosage, best way to give the drug, side effects, effectiveness as compared with similar drugs, and how easily it can be produced and manufactured [3,6]. In this stage researches are at level 2 or 3 of TRL.

3.3. Pre-clinical
In vitro tests are experiments conducted in the lab and in vivo studies are those in living cell and tissue cultures and animal models. Through these techniques, scientists work to understand how the drug works and what the potential side effects on humans might be [3]. In this stage, efficacy and potential risks are evaluated before human trials. So in pre-clinical studies must be demonstrated quality, safety, and efficacy. In vitro test demonstrated the efficacy against the identified target and toxicity. Meanwhile in vivo test demonstrated efficacy, toxicity, and also safety.

In this stage researchers must work out how to scale up the quantity for clinical trials. Major tests that undertaken during pre-clinical trials are pharmacokinetic profile (absorption, distribution,
metabolism, excretion), pharmacodynamics profile (physiological, drug action, relationship between drug concentration and effect), bioequivalence and bioavailability, acute toxicity, chronic toxicity, reproductive toxicity and teratogenicity, mutagenicity, carcinogenicity, and immunotoxicity, administration of medicine at three different dosage level. In this stage researches are at level 4-5 of TRL. In this stage researchers should collaborate with industry.

3.4. Clinical trials

Required standards before allowed to proceed to clinical trials in human, candidate medicine must be done test for toxicity at least 2 mammals [3,6].

- Phase I (tens volunteers, initial safety testing in a small group of healthy volunteers)
- Phase II (hundreds volunteers, assess safety and efficacy in a small group of patients)
- Phase III (thousands volunteers, demonstrate safety and efficacy in a large group of patients)
- Phase IV (evaluate the longterm safety or effects in specific patient subgroups)

In this stage, researches are at level 6-9 of TRL. and the research can passed through industry in level 6.

4. Discussion

4.1 Screening and conceptual model of classification

In this study we collect 1,258 articles about biomedicine and only 546 articles that include new medicine R&D. Furthermore, with the help of experts, 546 articles are then classified into 5 stages of new medicine R&D. And experts also provide keywords in the abstract as features that can indicate the level of each article in new drug R&D (see table 1).

| No | Article Title | Stage | Features for corpus |
|----|---------------|-------|---------------------|
| 1  | The Effects of Mangiferin (Mangiferaindica L) in Doxorubicin-induced Cardiotoxicity in Rats | pre-clinical; in vivo | Mangiferafoetida; bacang ; mango; against DOX-induced cardiotoxicity; against doxorubicin (DOX); antioxidants; rat; dose; in comparison with; animals; adult, male; |
| 2  | Bioactive protein fraction DLBS1033 containing lumbrokinase isolated from Lumbricusrubellus: Ex vivo, in vivo, and pharmaceutic studies | pre-clinical; in vitro; ex vivo; in vivo | DLBS1033; bioactive protein fraction; Lumbricusrubellus; thrombosis-related disease; the stability of the bioactive protein fraction in gastric conditions; enteric coated tablet; mix; centrifuged; male; rats; fasted; obtained from the rat; incubated; surgery |
| 3  | The effect of a unique propolis compound (propoelix™) on clinical outcomes in patients with dengue hemorrhagic fever | clinical traits: phase 2 | unique propolis extract; Propoelix™; PopulusNigra L.; poplar bud exudates; effectiveness; tumor necrosis factor-α (TNF-α) levels; dengue hemorrhagic fever (DHF); double-blind; randomized; placebo-controlled trial; patients; treatment group; placebo group; significant |
| 4  | Toxicity studies of a bioactive protein with antithrombotic-thrombolytic activity, DLBS1033 | pre-clinical; in vivo | DLBS1033; bioactive protein fraction; Lumbricusrubellus; antithrombotic; thrombolytic agents; toxicity study; dose; mice; male; female; fasted; observation period; major organs; rat; body weight; pregnant; prenatal developmental study; |
| 5  | Cellular mechanism of the cytotoxic effect of extracts from syzygiumpolyanthum leaves | pre-clinical: in vitro | Syzygiumpolyanthum; antioxidant; assay; anti-proliferation effect; cytotoxic effect; cultured cell |
4.2 Features for Corpus
Collection of features used to build corpus and extracted from abstract or full paper. The classification model in biomedicine research is constructed from a specific corpus at each stage. Corpus was then used to build a computerized model to predict the position of a study in stages of new medicine process discovery based on its publication. Table 2 below is an example of a collection of features that become corpus in each stage.

| No | Stage in new medicine process discovery | Features for corpus |
|----|----------------------------------------|---------------------|
| 1  | Basic research                          | Potential compound; inhibition; activation |
| 2  | Drug discovery                          | inhibition; activation; therapeutic effect; promising compound; lead compound; sequence homology studies; DNA microarray; structural genomics; High-throughput Screening |
| 3  | Pre-clinical studies in vitro           | Syzygium polyanthum; antioxidant; assay; anti-proliferation effect; cytotoxic effect; cultured cell; efficacy |
| 4  | Pre-clinical studies in vivo            | Bioactive protein fraction; antithrombotic; thrombolytic agents; toxicity study; dose; mice; male; female; fasted; observation period; major organs; rat; male; female; adult; body weight; pregnant; prenatal developmental study; efficacy; safety; absorption; distribution; metabolism; excretion; bioequivalence; bioavailability; dosage level |
| 5  | Clinical trails                         | Effectiveness; tumor necrosis factor-α (TNF-α) levels; dengue hemorrhagic fever (DHF); double-blind; randomized; placebo-controlled trial; patients; treatment group; placebo group; significant; volunteers; safety; efficacy |

Based on the indicators set by the Ministry of Research and Higher Education, each stage in biomedicine research and development can be linked to TRL. The relations between the stages and the TRL can be illustrated in the following figure 1.

**Figure 1.** The relation between stages of biomedicine R&D and TRL.

5. Future Work
This conceptual model of classification of biomedicine research, developed as part of a research framework to build a computerized predictive model based on research publications (figure 2). Automatic classification uses text mining and machine learning. In parallel this study will also take sampling for the classification of articles based on expert opinion. Then validation of conceptual model will be analyzed using kappa consistency test. So that this study build a computerized predictive model based on research publications using model that have been validated.
Figure 2. Research framework to build a computerized predictive model based on research publications.

Acknowledgments
Authors acknowledge for the financial support of Pusat Penelitian Perkembangan Iptek (Pappiptek) LIPI to research.

References
[1] Riani E, Gumbira-said E, Syamsu K and Cordova M R 2013 Aphrodisiacs 7 340–6
[2] Elya B, Basah K, Mun’Im A, Yuliastuti W, Bangun A and Septiana E K 2012 J. Biomed. Biotechnol
[3] PHRMA 2015 Biopharmaceutical Research & Development: The Process Behind New Medicines 1–21
[4] Hughes J P, Rees S S, Kalindjian S B and Philpott K L 2011 Br. J. Pharmacol 162 1239–49
[5] Handayani T, Hardiyati R, Nadhiroh I M, Amelia M and Rahmaida R 2016 Kajian Saintometrika Perkembangan Publikasi Ilmiah Keanekaragaman Hayati Indonesia Sebagai Bahan Rekomendasi Kebijakan Arah Penelitian Keanekaragaman Hayati Nasional Lap. Kegiat. Penelit. DIPA Pappiptek LIPI
[6] FDA 2017 The Drug Development Process