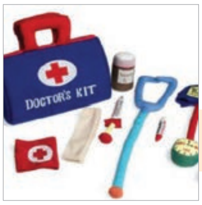


The Role of Bioinformatics in Personalized Medicine: Your Future Medical Treatment

Margareta Deidre Valeska, Gabriella Patricia Adisurja, Stefanus Bernard, Renadya Maulani Wijaya, Muhammad Aldino Hafidzhah, Arli Aditya Parikesit
Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta, Indonesia

ABSTRACT

Bioinformatics is beneficial in personalized medicine. Two methods stand out, the randomized algorithm and computer assisted drug design (CADD). This article will discuss application, pitfalls, and future of those two methods. Suggestion to improve the clarity of the bioinformatics research in the field of personalized medicine will also be reviewed.

Keywords: Bioinformatics, CADD, personalized medicine, randomized algorithm

ABSTRAK

Bioinformatika berperan sangat penting dalam personalized medicine. Dua metode penting dalam kajian ini adalah randomized algorithm dan computer assisted drug design (CADD). Kajian ini membahas aplikasi, kekurangan, dan masa depan kedua metode tersebut. Saran-saran untuk meningkatkan efek riset bioinformatika dalam kajian personalized medicine juga akan ditelaah. **Margareta Deidre Valeska, Gabriella Patricia Adisurja, Stefanus Bernard, Renadya Maulani Wijaya, Muhammad Aldino Hafidzhah, Arli Aditya Parikesit. Peran Bioinformatika dalam *Personalized Medicine* : Terapi Masa Depan**

Kata kunci: Bioinformatika, CADD, personalized medicine, randomized algorithm

INTRODUCTION

In recent decades, an accelerated pace of technology development, sparking the rise of fields which intersect classical fields had been seen. Coined in 1970 by Ben Hesper and Paulien Hogeweg, bioinformatics is an interdisciplinary field which merges biology, computer science, mathematics, and information technology.[1-3] Bioinformatics will eventually play part to compute data of human genomics diversity for creating a further breakthrough in medicine.

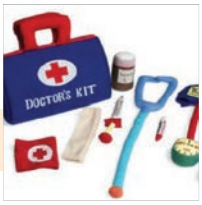
In 2019, approximately 7,7 billion humans live on earth with wide diversity of races and ethnicity.[4] In Indonesia there are approximately 300 different ethnics according to national census in 2010 by Statistics Indonesia (BPS).[5]

Why do people look so diverse? The answers lie down deep inside our genetic materials, which evolved from time to time due to various factors. Evolution of our genetic material creates different races and ethnics groups even though we share same amount of 99.9% genetic materials with each other.[6] Wide diversity among human populations is definitely a good thing. It will cause diverged immune-system regulations in case of pathogen attack.

Generraly, some of immune systems are probably compromised while others aren't, and vice versa. Thus, it prevents human extinction in case of global pandemic occurs. However, some problems are likely to be happened in medication, some will show positive results while others will not, even in the same case of disease. Compatibility between treatments and patient is one important goal to reach preferred results. In future, treatment will be focused to a concept called personalized medicine.[7]

Personalized medicine is a type of medical care in which treatment is customized individually for patient. It is possible because we are genetically different from each other.[8] Two important keys are involved in the concept. First, attempts conducted in medical research itself shows how personalized medicine is.[9] Various attempts include shifting emphasis in medicine from reaction to prevention, selection of optimal therapy, reducing time and cost of clinical trials, reducing overall estimated cost of healthcare and minimalizing risk of adverse drug reactions. The second key is rapid growth of information technology brought development of new tools to decode human genomes, large-scale studies of genetic variations and medical informatics.[10] The rapid development of new sequencing technologies also brought sequencing to a new level. Since human genome project was finished in 2003, new sequencing platforms keep emerging. Thus making the whole genome sequencing became cheaper, from $2.3 billion in 2003 until just $1000 in 2016 and might become cheaper and faster in the next years or decades.[11,12] Development of more advanced algorithms to produce learning machine and artificial intelligence (AI) also play role in handling datasets efficiently and predicting acurate outcome accurately.[13] AI, as a tool that could imitate brain-based

Alamat Korespondensi *email: arli.parikesit@i3l.ac.id*



cognitive function, already contribute medical technology development, ranging from drug design to the development of assisted-system for clinics.[14,15] Based on this statement, one notable application of deep learning is in the breast cancer diagnostics.[16]

For a clear picture of what personalized medicine is, let's assume in the future, there is a male person affected by a certain chronic disease. He went to the doctor and the doctor gave him common medical treatment as he gave to others, but the same as others. It turns out the treatment didn't work for him even though it worked for others. The doctor then fully sequenced his genomes, compared his genomes with reference and looked for specific reasons that might cause that treatment didn't work for him. The doctor also used AI to simulate treatment according to the genome sequenced, genomes reference provided and disease type.[17] AI provides several treatment methods based on databases, complete with success probability rate.[18] In the end, doctors can take the best decision in treating him. Hence, based on one explanation below, personalized medicine is very useful and promising in future medical treatment. In this respect, there are two important bioinformatic methods utilized for computing personalized medicine data, which are randomized algorithm and Computer-Aided Drug Design (CADD). These two methods are beneficial in building AI agents for more precision prediction.

Randomized Algorithm in Life Sciences and Medical Research

Randomized algorithm is a type of algorithm that utilizes random numbers, to solve problems and make decisions.[19] It is commonly used to reduce time and space complexity compared to other standard algorithms because it calculated results using probability or chances.[20] Randomized algorithms are commonly divided into two categories, Las Vegas and Monte Carlo algorithms. Las Vegas is an algorithm that produces correct results with a time complexity based on random value. Monte Carlo algorithm produces results with probabilities; easier to find its worst-case time complexity.

With advanced technology and science in recent decades, data and algorithms have found their way into integration with life sciences, as well as with medical research. Both medical and life sciences in particularly have huge amounts growing which have to be of data that's being updated every day and require advanced computational power to analyze. From these needs and demands in life sciences and medical research, the implementation of algorithms to be able to compute and analyze scientific data in fast, accurate, and efficient way is expected. Randomized algorithm, one of those algorithms can be usefully applicated in almost all aspects of life scienced medical research, such as evolution, cancer, genomic study, and else.

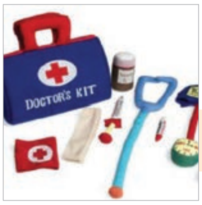
The simplest application of this algorithm is built to perform statistical analysis using probability therapy theory Monte Carlo to analyze large amount of biological data (genomic, transcriptomic, proteomic, and others omics-based studies).[21] In medical chemistry, this type of algorithm can be used to generate random chemical compounds for high throughput scanning of drugs. In biological cluster analysis, random projection algorithm is used to maintain quality of clustering within a small factor.[22] On top of that; analysis for selection, crossover, or even mutation in evolutionary biology can be done using random algorithm by taking out the most probable score. The Monte Carlo method also can be used to simulate CT scan radiation allowed dose without prior knowledge of scanning specifications.[23] In a similar topic, it can be used to determine dose distributions in dental implants and radiotherapy.[24] Randomized algorithm is still a pretty new algorithm used in life science research; however, it seems to have big potential in the future and this seems to only be scratching the surface of what the algorithm is truly capable of. Although some of those applications could be generated for personalized medicine, the current application of the randomized algorithm is still limited to provide randomized controlled clinical trial in statistical perspective, in regards of the medical experimental design.[25] Utilizing randomized algorithm to design drug based on existing library or database could be feasible as there are currently millions of leading compounds possibly to be chosen, but some of them have unidentified functions.[26,27]

Computer-Aided Drug Design (CADD) in Life Sciences and Medical Research

Thanks to the vast amount of data obtained from developed technologies, such as high-throughput screening, the emergence of bioinformatics as a field has helped further progress in medical field. Drug discovery with CADD is one of them, which has benefited significantly from bioinformatics by reducing the time and effort of identifying and screening drug candidates. One successful example of CADD application is in designing HIV/AIDS drug by specifically targeting its protease enzymed in treating glaucoma patients by targeting carbonic anhydrase enzyme.[28] Some of those drugs are already sold in the market.[29-31]

The first step in drug discovery is diagnosing a disease by its well-characterized symptoms, then identifying potential chemicals that can alleviate these symptoms or even to find and fix the root of the problems. An important approach in drug discovery is high-throughput screening (HTS), which starts from assessing thousands of compounds againts a set of defined targets to find useful active compounds.[32] Those large datasets require processing and further analysis to be concluded; thus, this is where bioinformatics takes place. Bioinformatics pipelines had been created to process data from HTS with CADD methods, especially to assess natural compound of products.[33]

Another aid of bioinformatics in drug discovery is identification of potential drug targets based on genomics, epigenetics, proteomics, and transcriptomics studies. Mutations associated with genetic diseases have been discovered by genomics and whole exome sequencing of the patients. Some of these mutations can cause disease and may be used as drug targets, while others are associated to the genes that induced associated with the disease. Bioinformatics tools such as position weight matrix can be used to scan genome patterns to determine if its mutation has a major impact on gene function.[34] In addition to mutations, epigenetic modifications may also affect drug targets. Epigenetics is the study of changes caused by the modification of gene expression rather than those caused by changes directly on the genome. Bioinformatics can be used to identify these changes and determine design of a drug



which can correct these changes.[35]

Scientists constantly find it challenging to compile and analyze data obtained from experiments to conclude meaningful results. The advent of bioinformatics has revolutionized biological research and allowed researchers to cut down costs and time spent on their studies.36 Advances in bioinformatics subfields, genomics and epigenetics greatly assists drug discovery, and various other medical fields.35 CADD has been applied in employment of artificial intelligence, including studies of drug delivery system, among for antibody-drug conjugation, and ligand-targeted conjugates as well.[37]

Outlook

Although the application of both randomized algorithm and CADD in personalized medicine is promising, several issues should be resolved. First, both methods could be extensive and expensive in computational cost. The growth of big data will accelerate submission of population-based genomics data abundantly and this situation can lead to "bottleneck effect" if there is insufficient computational power. Due to high demand of vast computational efforts are necessary, outsourcing to external High Performance Computing provider would be a feasible option.[38] Second, currently only there are few certified experts in bioinformatics have capability to analyze personalized medicine data, especially with particular methods which are already explained above. Both randomized algorithm and CADD are computational methods that could be operated comprehensively by bioinformatics experts. As bioinformatics experts are still limited in practice, there will be possible delay to meet the analysis request. The easiest solution is to outsource this demand to foreign suppliers.[39] Third, there are only few centralized bioinformatics database similar to NCBI/GENBANK, DDBJ, and EBI-EMBL in Indonesia. Although several databases are available, they are not centralized and designed for specific needs.[40,41] Based on those facts, most Indonesian bioinformaticians still rely on foreign centralized database as sources to retrieve data in their computational efforts.

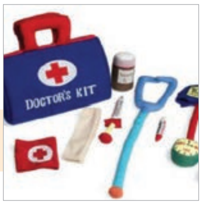
Although it will be difficult to reach a quick breakthrough in this issue, bioinformatic research is still useful to develop personalized medicine. A bioinformatic research group in medical school can help analyzing data that needs specialized computational efforts. The main challenge is to push forward bioinformatics research from algorithm to the software development. By supporting and focusing on computer science study, innovation will be achieved and it will be continuously upgrading exsisting application of randomized algorithm, while CADD will be devised in long term research.

Acknowledgement

The authors would like to thank to Institute for Research and Community Services of Indonesia International Institute for Life Sciences (i3l) for their heartfelt support. Thank also goes to Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia for providing Hibah Penelitian Dasar DIKTI/LLDIKTI III 2019 No. 1/AKM/PNT/2019

REFERENCES

1. Hesper B, Hogeweg P. Bioinformatica: Een werkconcept. Kameleon. 1970;1(6):28–9.
2. Hogeweg P. The roots of bioinformatics in theoretical biology. PLoS Comput Biol [Internet]. 2011 Mar [cited 2019 Jun 13];7(3):e1002021. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21483479
3. Ouzounis CA. Rise and demise of bioinformatics? Promise and progress. PLoS Comput Biol [Internet]. 2012. Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002487
4. Nation U. World population prospects - Population division - United Nations [Internet]. 2019 [cited 2019 Jul 19]. Available from: https://population.un.org/wpp/
5. Badan Pusat Statistik. Sensus penduduk tahun 2010 [Internet]. 2010. Available from: http://sp2010.bps.go.id/index.php/site/index
6. Huang T, Shu Y, Cai YD. Genetic differences among ethnic groups. BMC Genomics [Internet]. 2015 Dec 21 [cited 2019 Jun 13];16:1093. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26690364
7. Vogenberg FR, Isaacson BC, Pursel M. Personalized medicine: Part 1: Evolution and development into theranostics. P T [Internet]. 2010 Oct [cited 2019 Jun 13];35(10):560–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21037908
8. Verma M. Personalized medicine and cancer. J Pers Med [Internet]. 2012 Jan 30 [cited 2019 Jun 13];2(1):1–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25562699
9. Gyeman AA, Ofori-Asenso R. Perspective: Does personalized medicine hold the future for medicine? J Pharm Bioallied Sci [Internet]. 2015 [cited 2019 Jun 13];7(3):239–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26229361
10. Pasipoularides A. Genomic translational research: Paving the way to individualized cardiac functional analyses and personalized cardiology. Int J Cardiol [Internet]. 2017 Mar 1 [cited 2019 Jun 13];230:384–401. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28057368
11. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet [Internet]. 2011 Nov [cited 2019 Jun 13];52(4):413–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21698376
12. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. J Genet Genomics [Internet]. 2011 Mar 20 [cited 2019 Jun 13];38(3):95–109. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21477781
13. Parikesit AA. Kontribusi aplikasi medis dari ilmu bioinformatika berdasarkan perkembangan pembelajaran mesin (machine learning) terbaru. Cermin Dunia Kedokt [Internet]. 2018;45(9):700–3. Available from: http://www.kalbemed.com/DesktopModules/EasyDNNNews/DocumentDownload.ashx?portalid=0&moduleid=471&articleid=225&documentid=65
14. Yu MJ. Natural product-like virtual libraries: Recursive atom-based enumeration. J Chem Inf Model [Internet]. 2011 Mar 28 [cited 2014 Aug 8];51(3):541–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21388152
15. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Invest Radiol. 2017;52(7):434-40.
16. Ramanto KN, Parikesit AA. The usage of deep learning algorithm in medical diagnostic of breast cancer. Malaysian J Fundam Appl Sci [Internet]. 2019 Apr 18 [cited 2019 Apr 18];15(2):274–81. Available from: https://mjfas.utm.my/index.php/mjfas/article/view/1231



17. Aworunse OS, Adeniji O, Oyesola OL, Isewon I, Oyelade J, Obembe OO. Genomic interventions in medicine. Bioinform Biol Insights [Internet]. 2018 [cited 2019 Jun 13];12:1177932218816100. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30546257
18. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: Past, present and future. Stroke Vasc Neurol [Internet]. 2017 Dec [cited 2019 Jun 13];2(4):230–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29507784
19. Sharma A. Randomized algorithms - GeeksforGeeks [Internet]. 2019 [cited 2019 Jun 13]. Available from: https://www.geeksforgeeks.org/randomized-algorithms/
20. Patel J, Sanjay S, Salanke G. A survey on randomized algorithms and its applications. Int J Adv Trends Comput Appl [Internet]. 2017;4(3):1–6. Available from: https://www.academia.edu/34988262/A_Survey_on_Randomized_Algorithms_and_its_Applications
21. Anaraki FP. Randomized algorithms for large-scale data analysis. Electr Comput Energy Eng Grad Theses Diss [Internet]. 2017 Apr 1 [cited 2019 Jun 13]. Available from: https://scholar.colorado.edu/ecen_gradetds/141
22. Liu W, Ye M, Wei J, Hu X. Fast constrained spectral clustering and cluster ensemble with random projection. Comput Intell Neurosci [Internet]. 2017 Sep 25 [cited 2019 Jun 13];2017:1–14. Available from: https://www.hindawi.com/journals/cin/2017/2658707/
23. Ming X, Feng Y, Liu R, Yang C, Zhou L, Zhai H, et al. A measurement-based generalized source model for Monte Carlo dose simulations of CT scans. Phys Med Biol. 2017;62(5):1759-76.
24. Çatli S. High-density dental implants and radiotherapy planning: Evaluation of effects on dose distribution using pencil beam convolution algorithm and Monte Carlo method. J Appl Clin Med Phys. 2015;16(5):46–52.
25. Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. BMC Med [Internet]. 2018 [cited 2019 Jun 13];16(1):150. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30145981
26. Chou KC. Structural bioinformatics and its impact to biomedical science. Curr Med Chem [Internet]. 2004 Aug [cited 2012 Nov 20];11(16):2105–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15279552
27. Muegge I. Selection criteria for drug-like compounds. Vol. 23, Medicinal Research Reviews. 2003. p. 302–21.
28. Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein J Org Chem [Internet]. 2016 [cited 2019 Jun 13];12:2694–718. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28144341
29. Drugbank. Saquinavir [Internet]. 2019. Available from: https://www.drugbank.ca/drugs/DB01232
30. Drugbank. Amprenavir [Internet]. 2019. Available from: https://www.drugbank.ca/drugs/DB00701
31. Drugbank. Dorzolamide [Internet]. 2019. Available from: https://www.drugbank.ca/drugs/DB00869
32. Scholler P, Zwier JM, Trinquet E, Rondard P, Pin JP, Prézeau L, et al. Time-resolved förster resonance energy transfer-based technologies to investigate G protein-coupled receptor machinery: High-throughput screening assays and future development. Prog Mol Biol Transl Sci. 2013;113:275-312.
33. Koehn FE, Carter GT. The evolving role of natural products in drug discovery. Nat Rev Drug Discov [Internet]. 2005 Mar [cited 2014 Jul 11];4(3):206–20. Available from: http://dx.doi.org/10.1038/nrd1657
34. Xia X. Position weight matrix, Gibbs sampler, and the associated significance tests in motif characterization and prediction. Scientifica (Cairo) [Internet]. 2012 [cited 2019 Jun 13];2012:1–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24278755
35. Xia X. Bioinformatics and drug discovery. Curr Top Med Chem. 2017;17(15):1709–26.
36. Dibyajyoti S, Talha Bin E, Swati P. Bioinformatics: The effects on the cost of drug discovery. Gall Med J [Internet]. 2013 May 8 [cited 2019 Jul 19];18(1):44. Available from: https://gmj.sljol.info/article/10.4038/gmj.v18i1.5511/
37. Wang C, Xu P, Zhang L, Huang J, Zhu K, Luo C. Current strategies and applications for precision drug design. Front Pharmacol [Internet]. 2018 [cited 2019 Jun 13];9:787. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30072901
38. Akbar Z, Handoko LT. GRID architecture through a public cluster. In: Proceedings of the International Conference on Computer and Communication Engineering 2008, ICCCE08: Global links for human development. IEEE [Internet]. 2008 [cited 2016 Aug 12]:1016–8. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4580761
39. Witarto AB. Bioinformatika: Mengawinkan teknologi informasi dengan bioteknologi trendnya di dunia dan prospeknya di Indonesia [Internet]. 2003. Available from: http://www.komputasi.lipi.go.id/data/1014224403/data/1110939469.pdf
40. Yanuar A, Mun'im A, Lagho ABA, Syahdi RR, Rahmat M, Suhartanto H. Medicinal plants database and three dimensional structure of the chemical compounds from medicinal plants in Indonesia. Int J Comput Sci [Internet]. 2011 Nov 28 [cited 2014 Mar 23];8(5):180–3. Available from: http://arxiv.org/abs/1111.7183
41. Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, et al. KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol. 2012;53(2).