IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 3, September 2024, pp. 3157~3163
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i3.pp3157-3163

Machine learning for potential anti-cancer discovery from black sea cucumbers

Muhammad Fahrury Romdendine[1], Rizka Fatriani[2], Wisnu Ananta Kusuma[1,2,3], Annisa[1], Mala Nurilmala[4]
[1]Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia
[2]Tropical Biopharmaca Research Center, IPB University, Bogor, Indonesia
[3]Indonesian Society of Bioinformatics and Biodiversity, Jakarta, Indonesia
[4]Department of Aquatic Product Technology, Faculty of Fisheries and Marine Science, IPB University, Bogor, Indonesia

Article Info

Article history:

Received Jan 19, 2024
Revised Feb 20, 2024
Accepted Feb 28, 2024

Keywords:

Anti-cancer
Cascade deep forest
Drug discovery
Drug-target interactions
Machine learning

ABSTRACT

Despite being an abundant marine organism in Indonesia, black sea cucumbers (*Holothuria atra*) is still underutilised due to its slightly bitter taste. This study aims to identify potential anti-cancer compounds from black sea cucumbers using machine learning (ML) to perform drug discovery. ML models were used to predict interactions between compounds from the organism with cancer-related proteins. Following prediction, all compounds were computationally validated through molecular docking. The validated compounds were then screened using absorption, distribution, metabolism, excretion, and toxicity (ADMET) Lab 2.0 to assess their drug-like properties. The results showed that ML predicted seven out of 86 compounds were interacted with cancer-related proteins. Computational validation from the results showed that four out of seven compounds demonstrated stable interaction with proteins where only one compound meet the criteria of drug-like compound. The framework of ML and computational validation highlighted in this study shows a great promise in the future of drug discovery specifically for marine organisms. Since computational method only works in prediction realms, wet lab validation and clinical trials are imperative before the drug candidate can be produced as actual anti-cancer drug.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Wisnu Ananta Kusuma
Department of Computer Science, IPB University, Bogor, Indonesia
Jl. Taman Kencana No. 3, Bogor, Jawa Barat, 16128, Indonesia
Email: ananta@apps.ipb.ac.id

1. INTRODUCTION

The black sea cucumber (*Holothuria atra*), a marine organism, has been traditionally used in medicine for various purposes with its potential medicinal properties [1]. Recent *in vitro* and *in vivo* studies have suggested that extracts from black sea cucumbers exhibit anti-cancer properties [2], [3]. However, it is crucial to conduct additional research to identify bioactive compounds in black sea cucumber extracts that specifically target cancer-related proteins.

The process of identifying new drugs has been transformed by introducing cutting-edge technologies, including machine learning (ML), molecular docking, and absorption, distribution, metabolism, excretion, and toxicity (ADMET) prediction. These cutting-edge techniques have greatly improved the discovery and development of prospective medicinal medicines [4]–[6].

The burden of cancer in Indonesia is considerable, with cervical cancer now holding the distressing position of being the second-leading cause of death [7]. Between 2014 and 2018, the Indonesian government allocated Rp3.5 trillion to cancer-related expenditures. The top five cancers in terms of prevalence during this period were cervical, breast, lung, colorectal, and liver cancer. Therefore, it is crucial to explore innovative strategies beyond conventional medicines to improve patient outcomes, overcome drug resistance, and address unmet medical needs associated with this devastating disease.

The process of identifying new drugs has been transformed by introducing cutting-edge technologies, including ML, molecular docking, and ADMET prediction. These cutting-edge techniques have greatly improved the discovery and development of prospective medicinal medicines [4]–[6].

In predicting drug-target interactions (DTI), one of most fundamental process in drug discovery, numerous computational approaches leverage the capabilities of ML algorithms. For example, Wang *et al.* [8] conducted a DTI study employing a newly developed algorithm based on chemogenomics feature space. In another study, Chu *et al*. [9] demonstrated that a general-purpose novel algorithm called cascade deep forest (CDF) outperformed other state-of-the-art DTI algorithms. Other DTI studies also incorporated ML algorithmic options [10]–[13]. However, none have been found to continuously perform the three computational approaches mentioned before to identify black sea cucumber's bioactive compounds as a potential alternative cancer medicine.

This study aims to identify anti-cancer agents from black sea cucumbers through an integrated approach of DTI predictions using ML and computational validation using molecular docking, and ADMET analysis. The significance of the research lies in its capacity to demonstrate the role of ML in discovering novel compounds and understanding their interactions with target proteins. The findings can potentially contribute to developing effective anti-cancer therapies, signifying the role of ML in modern drug discovery, and expanding treatment options in cancer research.

2. METHOD

This study used cancer-related protein data from three sources: the cancer genome atlas [14], ijah analytics (http://ijah.apps.cs.ipb.ac.id/), and the human protein atlas [15]. Only proteins from cervical, breast, lung, colorectal and liver cancers were taken. Protein interaction data with common known compounds were taken from BindingDB [16]. Finally, the compound data from black sea cucumbers was taken by liquid chromatography-mass spectrometry (LC-MS) procedure in the wet lab. All the data taken was then preprocessed so that it was suitable for the ML modeling process later. One of the crucial preprocessing steps is the generation of negative interaction data between common compounds and cancer proteins.

The interaction data acquired from BindingDB typically comprises a substantial volume, whereas the computational capacity available to researchers is constrained. The potential for failure during the ML model training process arises when attempting to employ the entire interaction dataset concurrently. To address this challenge, we adopted a strategy of dataset fragmentation through random sampling. Ten distinct data subsets were extracted for each feature combination, with each subset containing samples calculated based on the minimum sample size determined in (1). This calculation incorporated a margin of error of 0.01 and a 95% confidence interval. A visual depiction of this data sampling process is presented in Figure 1.

$$n' = \frac{n}{1+\frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2 N}} \quad (1)$$

with, z = Z score, ε = margin of error, N = population size, and $\hat{p}$ = population proportion. The Z-score was derived from the calculation of the confidence interval, while the population proportion used was set at 50% due to the binary nature of the case.

Data on common compounds and cancer proteins obtained are still in the form of strings, namely SMILES and FASTA. Feature engineering is first done before it can be modeled with ML. Feature extraction from FASTA strings for protein data is done using protein descriptors including: AAC [17], AAIndex1 [18], PAAC [19], and ATC [20]. Meanwhile, SMILES string extraction was performed using molecular fingerprints including: ECFP [21], Klekota-Roth [22], MACCS, Morgan [23], and PubChem [24]. The extraction results of both types of features are then combined so that it will produce twenty different types of feature space combinations. The illustration of the combined feature space is presented in Figure 2, where m, n, and i represent the dimensions of each feature space, and r represents the number of compound-protein pairs.

The ML algorithms used in this study included CDF, extreme gradient boosting machine (XGBoost), light gradient boosting machine (LightGBM), logistic regression (LR), multi-layer perceptron neural networks (MLPNN), random forest (RF), and k-nearest neighbors (KNN). The DF21 (package source

code available at: https://github.com/LAMDA-NJU/Deep-Forest) package was utilized to implement the CDF algorithm, while the Scikit-Learn [25], XGBoost [26], and LightGBM [27] packages were used for the other algorithms. All algorithms were executed with default hyperparameter configurations. The performance performance metrics used for each algorithm are accuracy, precision, recall, F1-score, and area under the curve (AUC) score. Algorithm with best performance were then used to perform DTI between cancer-related proteins with compounds from black sea cucumbers. Following the successful prediction of positive interactions through ML, the findings undergo computational validation through molecular docking and ADMET analysis.

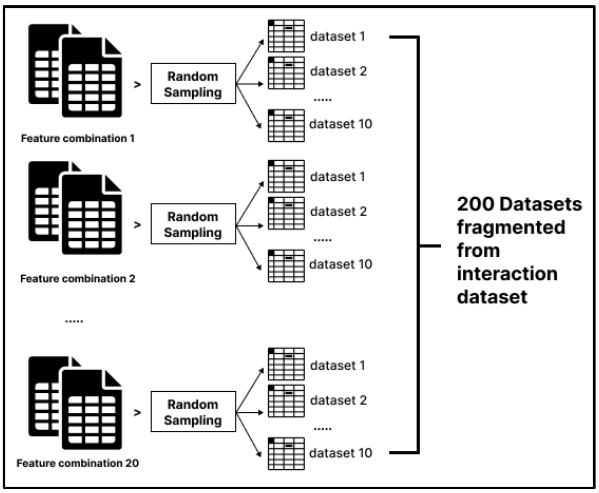

Figure 1. Illustration of data sampling process

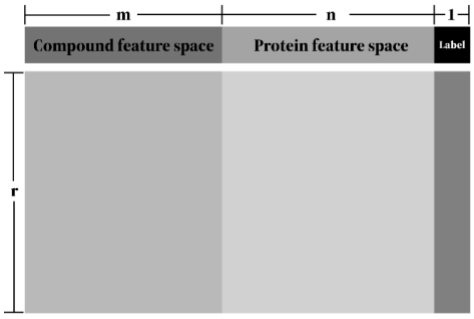

Figure 2. Illustration of combined feature space in final dataset for modeling

3. RESULTS AND DISCUSSION

3.1. Data acquisitions and preprocessing

A total of 550 unique cancer-related proteins were acquired. Querying the interactions of these 550 proteins in the BindingDB database resulted in 139,881 interaction data. Meanwhile, 86 bioactive compounds were obtained from wet lab identification using LC-MS. The interaction dataset from BindingDB only contained known interaction between small molecules and proteins of interest (positive instances). To meet the ML modeling criteria, negative interaction instances were generated with a 1:1 ratio, resulting in 269,235 interaction instances. Illustration of negative sample generation is presented in Figure 3.

Figure 3. Illustration of negative sample generation that generates negative interaction samples from acquired compound-protein interactions dataset

3.2. Performance evaluation of machine learning models

The performance metrics indicated that CDF was outperformed in almost every metric, except for the recall value, which was surpassed by RF. The highest overall performance was achieved by CDF, with an accuracy of 81.5%, F1-score of 81.4%, an AUC score of 93.7%, a precision of 91.8%, and Cohen's Kappa of 74.9%. On the other hand, the highest recall value was obtained by KNN (87%), followed by LightGBM (86.2%) and CDF (86.1%). The recall value difference was not appreciably siginifcant, so CDF was still preferred. The overall performance metrics calculation (averaged from all feature space combination) is presented in Table 1. The selection of the ML algorithm for the prediction stage was based on comparing the AUC score, as it was deemed to represent the overall effectiveness of an algorithm [28]. The AUC scores are calculated from ROC curves, which are presented in Figure 4.

Table 1. Overall performance metrics of each ML algorithm.

	CDF	RF	XGBoost	LightGBM	KNN	LR	NN
overall accuracy	0.815	0.813	0.814	0.807	0.757	0.771	0.762
overall F1-score	0.814	0.807	0.779	0.759	0.685	0.543	0.437
overall ROC-AUC	0.900	0.889	0.862	0.846	0.750	0.655	0.503
overall precision	0.858	0.834	0.779	0.768	0.661	0.570	0.469
overall recall	0.780	0.784	0.778	0.752	0.716	0.566	0.448
overall kappa	0.652	0.633	0.570	0.537	0.372	0.236	0.003

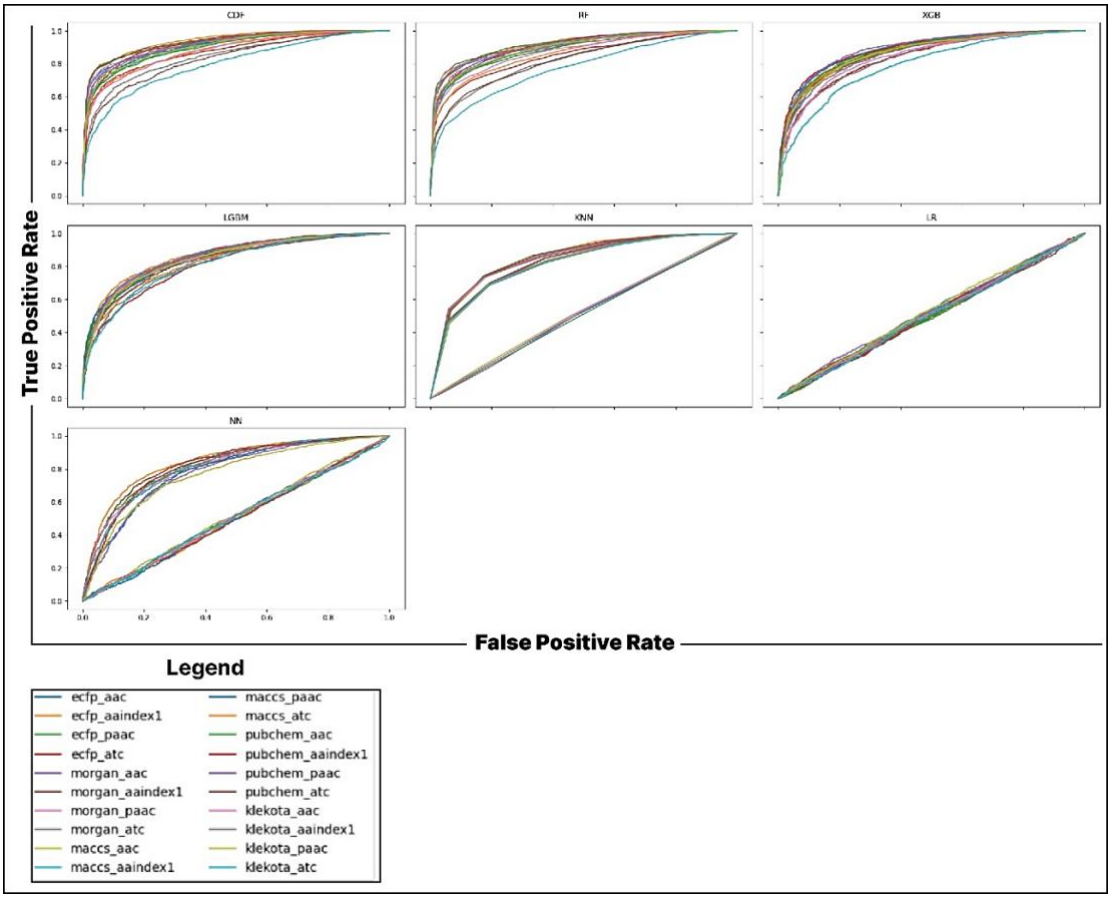

Figure 4. ROC Curve of each feature combination in multiple algorithms

3.3. Drug-target interactions predictions and validations

The prediction was carried out using CDF algorithm since it gained best AUC score. This process identified seven compound-protein pairs with positive interaction. The predicted results are presented in Table 2. Additionally, provided is the confidence score value, representing the probability of the prediction falling into the positive class. This value is directly extracted from the prediction results using CDF. The PubChem ID and Uniprot ID of the predicted interacting compound and protein are also included.

Validation through molecular docking revealed that only four out of the seven compound-protein pairs exhibited favorable binding affinity values below -6.0 kcal/mol. These four pairs are afimoxifene-PIK3CB (-12.7 kcal/mol), danazol-CYSLTR2 (-12.3 kcal/mol), taxifolin-PIK3CB (-10.0 kcal/mol), and terfenadine-UBE2F (-6.6 kcal/mol). Subsequently, validation via ADMET analysis indicated that only the taxifolin compound demonstrated optimal drug properties, successfully meeting most of the tested parameters in the ADMET analysis.

Table 2. DTI prediction results that are labeled positive yielded by CDF

Bioactive compound name (PubChem ID)	Protein/gene name (Uniprot ID)	Confidence score
Meclizine (4034)	UBE2F (Q969M7)	0.8445
Taxifolin (439533)	PIK3CB (P42338)	0.7906
Terfenadine (5405)	UBE2F (Q969M7)	0.8560
Afimoxifene (449459)	PIK3CB (P42338)	0.8921
Selegiline (26757)	UBE2F (Q969M7)	0.8752
Phencyclidine (6468)	UBE2F (Q969M7)	0.8962
Danazol (28417)	CYSLTR2 (Q9NS75)	0.8027

3.4. Discussions

The superiority of CDF algorithm in this study is due to its nature of addressing limitations of neural network-based algorithms. This algorithm leverages the properties of neural networks, such as layer-by-layer learning, simultaneous feature transformation, and complex structure, to achieve comparable performance. With similar reliability, this algorithm aims to overcome the dependence of neural networks on hyperparameter tuning, which is often done through trial and error and is inefficient. CDF is built using a layered structure like neural networks, but each node is replaced with ensemble learning techniques, such as RF. This design choice reduces the number of hyperparameters required for CDF.

The complexity of the CDF structure can adapt to the complexity of the training data. Unlike neural network, whose complexity is determined upfront, the number of layers in CDF depends on the data. The addition of layers in CDF is based on the evaluation of the previous layers, and the process will be stopped if there is no significant improvement in performance. Additionally, unlike neural networks, CDF does not require backpropagation, which means it does not rely on training with a large amount of labelled data to achieve good performance. Considering the characteristics of this algorithm, the superior performance of CDF in this study is not surprising.

Discussing Taxifolin as a compound from black sea cucumbers selected in this study, a literature search shows that Taxifolin compounds indeed exhibit promising anti-cancer activities [29]–[33]. This further demonstrates the significance of the ML approach implemented in this study as an effort in modern drug discovery.

4. CONCLUSION

This research successfully utilized ML, molecular docking, and ADMET analyses, to identify novel anti-cancer agent from the black sea cucumber. The results highlighted the effectiveness of the CDF algorithm in determining DTI. Through molecular docking validation, four promising compounds were identified: afimoxifene, danazol, taxifolin, and terfenadine. Subsequent ADMET analysis provided valuable insights into these compound‘s absorption, distribution, metabolism, excretion, and toxicity characteristics. Among them, Taxifolin exhibited the most favourable results, passing the highest number of ADMET parameters. These findings underscore the significance of ML in discovering novel compounds and comprehending their interactions with target proteins, contributing to modern drug discovery efforts. Taxifolin has shown promise as a lead compound, warranting further development for anti-cancer drugs. However, additional experimental validation is necessary to ascertain the efficacy and safety of these compounds, ultimately paving the way for potential therapeutic interventions against cancer.

ACKNOWLEDGEMENTS

The authors would like to express our deep appreciation for the financial support provided by the Directorate of Higher Education, Research and Technology, Ministry of Education, Culture, Research and Technology. This support was instrumental in the successful execution of the research project in accordance with the contract for the Implementation of the Research Programme Year 2023 (No: 102/E5/PG.02.00.PL/2023), dated 19 June 2023. This financial support has played a crucial role in advancing our scientific pursuits and is gratefully acknowledged.

REFERENCES

[1] W.-T. Lee *et al.*, "Black sea cucumber (Holothuria atra Jaeger, 1833) rescues pseudomonas aeruginosa -infected caenorhabditis elegans via reduction of pathogen virulence factors and enhancement of host immunity," *Food & Function*, vol. 10, no. 9, pp. 5759–5767, 2019, doi: 10.1039/C9FO01357A.

[2] R. Pangestuti and Z. Arifin, "Medicinal and health benefit effects of functional sea cucumbers," *Journal of Traditional and Complementary Medicine*, vol. 8, no. 3, pp. 341–351, Jul. 2018, doi: 10.1016/j.jtcme.2017.06.007.

[3] S. Bordbar, F. Anwar, and N. Saari, "High-value components and bioactives from sea cucumbers for functional foods-a review," *Marine Drugs*, vol. 9, no. 10, pp. 1761–1805, Oct. 2011, doi: 10.3390/md9101761.

[4] A. S. Rifaioglu, H. Atas, M. J. Martin, R. C. -Atalay, V. Atalay, and T. Doğan, "Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases," *Briefings in Bioinformatics*, vol. 20, no. 5, pp. 1878–1912, Sep. 2019, doi: 10.1093/bib/bby061.

[5] G. Bocci, E. Carosati, P. Vayer, A. Arrault, S. Lozano, and G. Cruciani, "ADME-space: a new tool for medicinal chemists to explore ADME properties," *Scientific Reports*, vol. 7, no. 1, Jul. 2017, doi: 10.1038/s41598-017-06692-0.

[6] C. Cava and I. Castiglioni, "Integration of molecular docking and in vitro studies: a powerful approach for drug discovery in breast cancer," *Applied Sciences*, vol. 10, no. 19, pp. 1–18, Oct. 2020, doi: 10.3390/app10196981.

[7] S. A. Kristina, D. Endarti, and H. Aditama, "Prediction of productivity costs related to cervical cancer mortality in Indonesia 2018," *Malaysian Journal of Medical Sciences*, vol. 29, no. 1, pp. 138–144, Feb. 2022, doi: 10.21315/mjms2022.29.1.13.

[8] Y. Wang *et al.*, "RoFDT: identification of drug–target iteractions from protein sequence and drug molecular structure using rotation forest," *Biology*, vol. 11, no. 5, May 2022, doi: 10.3390/biology11050741.

[9] Y. Chu *et al.*, "DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features," *Briefings in Bioinformatics*, vol. 22, no. 1, pp. 451–462, Jan. 2021, doi: 10.1093/bib/bbz152.

[10] A. K. Nasution, S. H. Wijaya, and W. A. Kusuma, "Prediction of drug-target interaction on Jamu formulas using machine learning approaches," in *2019 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2019*, IEEE, Oct. 2019, pp. 169–174. doi: 10.1109/ICACSIS47736.2019.8979795.

[11] A. Fadli, A. Annisa, and W. A. Kusuma, "Prediction of drug-target interaction using random forest in coronavirus disease 2019 case," *Bioinformatics and Biomedical Research Journal*, vol. 4, no. 1, pp. 1–7, Nov. 2021, doi: 10.11594/bbrj.04.01.01.

[12] Y. Ding, J. Tang, F. Guo, and Q. Zou, "Identification of drug-target interactions via multiple kernel-based triple collaborative matrix factorization," *Briefings in Bioinformatics*, vol. 23, no. 2, Mar. 2022, doi: 10.1093/bib/bbab582.

[13] Z. Cheng, C. Yan, F.-X. Wu, and J. Wang, "Drug-target interaction prediction using multi-head self-attention and graph attention network," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 4, pp. 2208–2218, Jul. 2022, doi: 10.1109/TCBB.2021.3077905.

[14] H. J. Lee, J. Palm, S. M. Grimes, and H. P. Ji, "The cancer genome atlas clinical explorer: a web and mobile interface for identifying clinical-genomic driver associations," *Genome Medicine*, vol. 7, no. 1, Dec. 2015, doi: 10.1186/s13073-015-0226-3.

[15] M. Uhlén *et al.*, "A human protein atlas for normal and cancer tissues based on antibody proteomics," *Molecular and Cellular Proteomics*, vol. 4, no. 12, pp. 1920–1932, Dec. 2005, doi: 10.1074/mcp.M500279-MCP200.

[16] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson, "BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities," *Nucleic Acids Research*, vol. 35, pp. 198-201, Jan. 2007, doi: 10.1093/nar/gkl999.

[17] B. Zhao *et al.*, "DescribePROT: database of amino acid-level protein structure and function predictions," *Nucleic Acids Research*, vol. 49, no. D1, pp. D298–D308, Jan. 2021, doi: 10.1093/nar/gkaa931.

[18] S. Kawashima, P. Pokarowski, M. Pokarowska, A. Kolinski, T. Katayama, and M. Kanehisa, "AAindex: amino acid index database, progress report 2008," *Nucleic Acids Research*, vol. 36, pp. D202-D205, Jan. 2008, doi: 10.1093/nar/gkm998.

[19] K. C. Chou, "Prediction of protein cellular attributes using pseudo-amino acid composition," *Proteins: Structure, Function and Genetics*, vol. 43, no. 3, pp. 246–255, May 2001, doi: 10.1002/prot.1035.

[20] R. Kumar *et al.*, "An in silico platform for predicting, screening and designing of antihypertensive peptides," *Scientific Reports*, vol. 5, no. 1, Jul. 2015, doi: 10.1038/srep12512.

[21] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *Journal of Chemical Information and Modeling*, vol. 50, no. 5, pp. 742–754, May 2010, doi: 10.1021/ci100050t.

[22] J. Klekota and F. P. Roth, "Chemical substructures that enrich for biological activity," *Bioinformatics*, vol. 24, no. 21, pp. 2518–2525, Nov. 2008, doi: 10.1093/bioinformatics/btn479.

[23] H. L. Morgan, "The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service," *Journal of Chemical Documentation*, vol. 5, no. 2, pp. 107–113, May 1965, doi: 10.1021/c160017a018.

[24] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant, "PubChem: a public information system for analyzing bioactivities of small molecules," *Nucleic Acids Research*, vol. 37, pp. W623–W633, Jul. 2009. doi: 10.1093/nar/gkp456.

[25] Fabian P. *et al.*, "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[26] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[27] G. Ke *et al.*, "LightGBM: a highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017, pp. 3147–3155.

[28] V. Premachandran and R. Kakarala, "Measuring the effectiveness of bad pixel detection algorithms using the ROC curve," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2511–2519, Nov. 2010, doi: 10.1109/TCE.2010.5681135.

[29] V. M. Gohil *et al.*, "Meclizine inhibits mitochondrial respiration through direct targeting of cytosolic phosphoethanolamine metabolism," *Journal of Biological Chemistry*, vol. 288, no. 49, pp. 35489–35499, Dec. 2013, doi: 10.1074/jbc.M113.489237.

[30] Z. Dostal, M. Sebera, J. Srovnal, K. Staffova, and M. Modriansky, "Dual effect of taxifolin on ZEB2 cancer signaling in HepG2 Cells," *Molecules*, vol. 26, no. 5, Mar. 2021, doi: 10.3390/molecules26051476.

[31] Y. J. Wang *et al.*, "Ketoconazole potentiates terfenadine-induced apoptosis in human Hep G2 cells through inhibition of cytochrome p450 3A4 activity," *Journal of Cellular Biochemistry*, vol. 87, no. 2, pp. 147–159, 2002, doi: 10.1002/jcb.10282.

[32] B. Blaya *et al.*, "Histamine and histamine receptor antagonists in cancer biology," *Inflammation and Allergy - Drug Targets*, vol. 9, no. 3, pp. 146–157, Jul. 2010, doi: 10.2174/187152810792231869.

[33] T. Tábi, L. Vécsei, M. B. Youdim, P. Riederer, and É. Szökő, "Selegiline: a molecule with innovative potential," *Journal of Neural Transmission*, vol. 127, no. 5. pp. 831–842, May 2020. doi: 10.1007/s00702-019-02082-0.

BIOGRAPHIES OF AUTHORS

Muhammad Fahrury Romdendine received his master's degree in computer science from IPB University with focus in bioinformatics research. He also received a Bachelor of Computer Science from IPB University. He is currently working in as a researcher in the field of artifical intelligence and bioinformatics. He can be contacted at email: romfahrury@apps.ipb.ac.id or mfromdendine@gmail.com.

Rizka Fatriani received a bachelor's degree from Andalas University and a master's degree from the University of Indonesia. She is currently working at the Tropical Biopharmaca Research Center. Her research focuses on bioinformatics, molecular docking, molecular biology, and cryopreservation. She can be contacted at email: fatrianirizka@gmail.com.

Wisnu Ananta Kusuma earned his bachelor's and master's degrees from the Bandung Institute of Technology and obtained his Ph.D. from the Tokyo Institute of Technology in 2012. He is an Associate Professor in the Department of Computer Science at IPB University. He serves as the Executive Secretary of the Institute for International Research on Advanced Technology at IPB University. Additionally, he coordinates the Bioinformatics Working Group at the Faculty of Mathematics and Natural Science, IPB University. He leads the Bioinformatics and High-Performance Computing Research Group at the Advanced Research Laboratory, IPB University. His research focuses on machine learning, high-performance computing, and bioinformatics, with over 60 published articles and extensive experience reviewing international journals. He also serves as the Chairperson of the Indonesian Society of Bioinformatics and Biodiversity and holds a position as an ExCo Member of the Asia Pacific Bioinformatics Network (APBioNet). He can be contacted at email: ananta@apps.ipb.ac.id.

Annisa is a senior lecturer in the Department of Computer Science at IPB University. Her research interests include data mining, skyline query, and recommender system. Her current research focusing on data mining and application of skyline query for recommender system in various fields. She can be contacted at email: annisa@apps.ipb.ac.id.

Mala Nurilmala received the Dr. degree in Aquatic Bioscience from The University of Tokyo, Japan. She is currently a Professor at Aquatic Product Technology Department, Faculty of Fisheries and Marine Science, IPB University-Indonesia since 2021. In addition, she is serving as a vice dean for resources, collaboration, and development at Faculty of Fisheries and Marine Science, IPB University. Her research interests are exploration bioactive compound from marine organisms as well as proteomic and genomic studies. She has published paper both national and international journals. She can be contacted at email: mnurilmala@apps.ipb.ac.id.