BIO Web of Conferences **163**, 01007 (2025) https://doi.org/10.1051/bioconf/202516301007
ICBBB 2025

Anti-BioEn: an advanced framework for accurate bioactive agent classification based on hybrid models and graph feature encoding method.

Md. Shazzad Hossain Shaon[1], *Md. Fahim* Sultan[1], *Tasmin* Karim[1], *Mst Shapna* Akter[1]

[1]Department of Computer Science and Engineering, Oakland University, Rochester, MI 48309, USA

Abstract. Bioactive agents are compounds that have an influence on human beings, organs, or tissues. These agents, which might be found in both natural and synthetic chemicals, are able to interact with biological systems and produce a variety of therapeutic or biological responses. In this regard, this work proposes a stacking method for categorizing five important bioactive agents: antibacterial, anti-HIV, antioxidant, antiparasitic, and antiprotozoal. This study has been designed with a graph-based feature extraction approach that successfully captures intricate interactions between molecular structures of bioactive substances. These extracted characteristics were then put into a stacking strategy, which is a strong ensemble learning technique that leverages the capabilities of several machine learning models to improve classification accuracy. By utilizing this innovative technique, the model outperformed state-of-the-art methods across all assessment criteria with more than 85% in terms of accuracy. The findings demonstrate the efficacy of the graph-based technique in conjunction with the stacking model, making it a useful tool for detecting important bioactive chemicals such as antibacterial, anti-HIV, antioxidant, antiparasitic, and antiprotozoal compounds for drug development, hence facilitating the development of novel therapeutic agents for essential health applications.

1 Introduction

The raising resistance of pathogenic microbes to traditional chemical medications requires the development of revolutionary treatment techniques for infectious disorders [1, 2]. Antimicrobial peptides (AMPs) are the first series of defense against many infections. These peptides have a wide range of functions, including antibacterial (ABPs), anti-HIV (AHPs), antioxidant (AOPs), antiparasitic (APPs), and antiprotozoal (APZPs) properties, making them important in preventing infections and improving outcomes of treatment [3]. In general, these functionalities are small-molecule polypeptides that perform a variety of tasks against target species. These peptides serve an important role in innate immunity and are effective against a variety of pathogens, including bacteria, viruses, fungi, and parasites [4]. According

to the study, ABPs are reduced and entirely wipe out bacteria, making them crucial in the treatment of bacterial illnesses. They work by inhibiting critical bacterial activities such cell wall formation, protein production, and DNA replication [5-7]. AHPs were developed to prevent the reproduction and spread of the Human Immunodeficiency Virus (HIV). Anti-HIV medicines, such as reverse transcriptase and protease inhibitors, disrupt critical steps of the viral life cycle [8-10]. AOPs neutralize free radicals, which are unstable chemicals that may harm cells and lead to aging, cancer, and cardiovascular disease. These chemicals defend the organism from oxidative stress by stabilizing free radicals and reducing cellular damage [11-13]. APPs, these medications are used to treat parasitic infections, which can include protozoa, helminths, and ectoparasites. Antiparasitic medications kill parasites directly or prevent them from surviving and reproducing within the host [14-16]. These are a group of medication to combat parasite that target protozoan illnesses like malaria, leishmaniasis, and giardia. Traditional antiprotozoal medicines frequently target essential processes in protozoa, such as DNA synthesis or protein creation, however there is growing interest in natural chemicals and AMPs as potential strategies for eliminating protozoan infections [17-20]. Traditional procedures endeavor to determine the most optimum solutions; however, these methods are generally time-consuming and labor-intensive. As a result, there is a rising emphasis on artificial intelligence systems to detect functionality more effectively. Researchers could leverage AI to accelerate the identification process, saving time and money on medication development and functional analysis.

In this work, a sophisticated algorithm has been developed to detect five particular bioactive molecules using existing AMP databases [3]. Recent study has highlighted the significance of a generalized model, which provides a more efficient way for biologists and bioinformatics applications. Rather of using distinct classifiers for each bioactive agent, a unified predictor is ideal and attainable for detecting antibacterial, anti-HIV, antioxidant, antiparasitic, and antiprotozoal chemicals. This integrated strategy improves the detection method via reducing computational complexity and increasing the accuracy of bioactive agent prediction. As a result, the suggested approach offers considerable advantages in terms of enhancing drug development and bioinformatic research.

2 Related works

In the modern era, advancements across all fields have led to an increased emphasis on modernization, including within the realm of bioinformatics. Experts in this discipline are currently focusing on computational techniques for addressing complexities. As a result, plenty of time has been spent determining the most optimum solutions using striking algorithmic and data-driven techniques. These computational approaches have proven significant in overcoming hurdles in biological research, helping to facilitate higher quality analysis and fast resolution of issues. Jukič et al., used the most recent machine learning algorithms to uncover new antibacterial agents and targets, including small compounds and antibacterial peptides [21]. Ivanenkov et al., demonstrated Support Vector Machine (SVM)-based technique for identifying ABPs with an accuracy of 75.5% [22]. However, there is still plenty of space for enhancement of functionality. Fjell et al, recommended another approach to detection of ABPs based on machine learning methods and obtained a better accuracy [23]. Yang et al. proposed quantitative structure activity relationship (QSAR) and SVM machine learning method to improve the accuracy [24]. Lata et al. designed SVM and Artificial Neural

Network (ANN) based model to detect the ABPs [25]. Rosa et al. developed anti-HIV therapies through machine learning [26]. Foglierini et al., proposed another machine learning method for HIV-1 bNAbs detection [27]. Ge et al. predicted antiviral peptide, which are correlated to HIV [28]. Hesamzadeh et al., fully explained the AOPs based on artificial intelligence methods [29]. Zhang et al. proposed bioactive agent identification based on machine learning [30]. Chen et al. developed Deep2Pep model based on deep learning approaches with 73.7% accuracy overall [31]. In another article, Zhang et al., proposed Matthew's correlation coefficient (MCC) metrics outcome with Esm4ao model for AOPs detection [32]. Periwal et al. designed a model for APPs detection, where the authors used Extremely Gradient Boosting (XGBoost) method [33]. Furthermore, machine learning algorithms for predicting the functionality of AMPs have been intensively studied, revealing their potential to improve accuracy and efficiency identification [3, 34-36].

However, this work presents another technique that helps advance the field in various ways:

a) The goal of this research is to identify five bioactive compounds to develop a strong classifier that could be used as a generalist framework. This single classifier can detect all five classifications, resulting in considerable time and cost savings.
b) This technique follows Chou's five-step requirements, incorporating multiple machine learning algorithms into our architecture before emerging on an effective choice known as the Anti-BioEn concept.
c) To improve categorization, we use a graph-based feature extraction strategy.

As a result, this technique could contribute to finding effective medicinal products while also aiding in larger advances in bioinformatics, pharmacology, and personalized medicine. Finally, this method demonstrates potential for developing innovations that will have a major influence on future public health outcomes.

3 Materials & Methods

3.1 Benchmark Dataset

The research study gathered the data through iAMPCN article [3]. The authors claimed that they
organised the samples from APD3, dbAMP, DRAMP, AnOxPePred, HAPPENN and so many repositories [37-41, 3]. The dataset used in this study includes a variety of bioactive peptides, each with a distinctive sample size. ABPs comprise 1,000 positive and 1,000 negative samples. AHPs are composed of 812 positive and 812 negative samples. 180 positive and 180 negative samples reflect AOPs. APPs have 457 positive samples and 457 negative samples, which are evenly divided into positive and negative data. In the end, APZPs are represented by 53 positive and 53 negative samples. However, we set the datasets as 20% for test size and 80% for train phase of each.

3.2 Methodological Overview

The study aims to determine the most successful model for identifying bioactive substances. Figure 1 illustrates every step of the methods, which begins with dataset obtaining. Following that, a feature extraction procedure was executed using a graph-based feature extraction

approach. The collected properties were incorporated into machine learning algorithms that classified the biochemical agents. Six machine learning models were implemented; therefore, an ensemble approach was constructed from these six baseline models, which produced Anti-BioEn, a stacking-based strategy. Numerous evaluation requirements were applied, and it was discovered that Anti-BioEn was the most advantageous approach to determine every biologic agent.

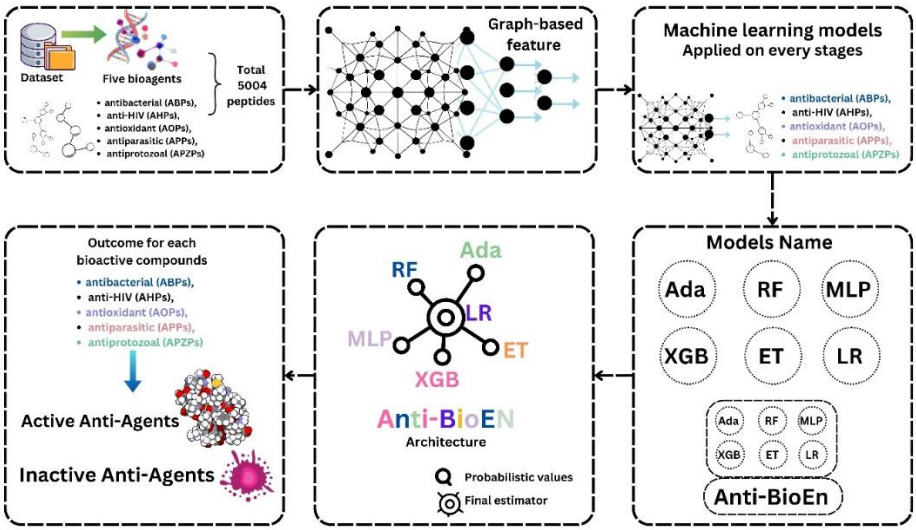

Fig 1. Methodology of the study, where first step collects the dataset, then feature extractions, model apply and finding the optimal model based on evaluation metrics.

3.3 Feature Extraction Procedure

This study used graph-based feature extraction to describe the interactions between bioactive compounds. GraphCodeBERT was used for feature extraction because of its ability to capture structural relationships as well as semantic content from graph representations [42, 43]. This technique allowed a deeper knowledge of bioagent interactions, allowing for more accurate categorization across many categories. The components bioactive agents: ABPs, AHPs, AOPs, APPs, and APZPs can be seen as vertices in a graph, while their interactions or dependencies are shown as edges. Each node receives an embedding that represents both its semantic and graphological properties. This delivers an extensive, multidimensional depiction of the bioactive chemicals or additional substances in analysis. Figure 2. represent the overall procedure of the feature extraction.

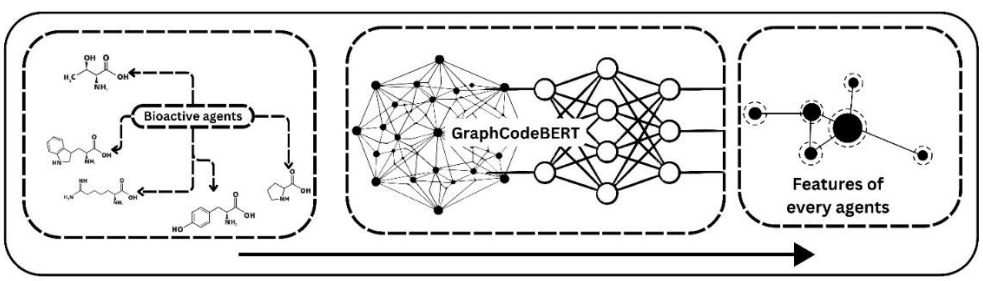

Fig 2. Graph-based Feature extraction procedure of every bioactive agent.

3.4 Proposed Approach Construction

Initially, six baseline models were used: Logistic Regression (LR), AdaBoost (ADA), Extreme Gradient Boosting (XGB), Multi-Layer Perceptron (MLP), Extra Trees (ET), and Random Forest [43-48]. These models were combined into a mounting technique [49], with ADA, XGB, MLP, ET, and RF serving as base learners and LR as the finalized predictor in the layering algorithm called Anti-BioEn. The accumulative approach incorporates results from fundamental models to increase overall predictive performance. The mathematical expression for the final forecast is as follows:

$$\begin{cases} a_t(x) = \sum_{t=1}^{N} \alpha_t . a_t(x_i) \\ f(x) = \sum_{i=1}^{T} \alpha_t a_t(x) \\ g(x) = \sum_{i=1}^{N} g_i(x) \\ y(x) = W^L a^{L-1} + b^L \\ e(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \\ L = \sigma(W^T \varphi(x) + b) \\ \varphi(x) = \begin{pmatrix} f(a_t(x)) \\ f(f(x)) \\ f(g(x)) \\ f(y(x)) \\ f(e(x)) \end{pmatrix} \\ \sigma(z) = \frac{1}{1+e^{-z}} \\ Anti - BioEn = \{\varphi(x) + \sigma(z)\} \end{cases} \quad (1)$$

Where, $a_t(x)$ is ADA, $f(x)$ is the prediction is a weighted sum of the ADA, $g(x)$ means the XGB, $y(x)$ is the MLP, $e(x)$ is the ET and L is the LR $\varphi(x)$ means the overall probabilistic values of all base line models. $\sigma(z)$ is the overall logistic function.

The Anti-BioEn approach has various benefits that allow being implemented. To begin, this layering tactic integrates various models, using their distinct strengths while correcting for their flaws, resulting in more accurate overall forecasts than any single model. The selected baseline models use a variety of approaches, ranging from linear regression to ensemble methods, which improves the model's capacity to generalize to new data. Furthermore, by merging numerous models, the Anti-BioEn model decreases the danger of overfitting by averaging out the predictions, resulting in stronger performance on test data

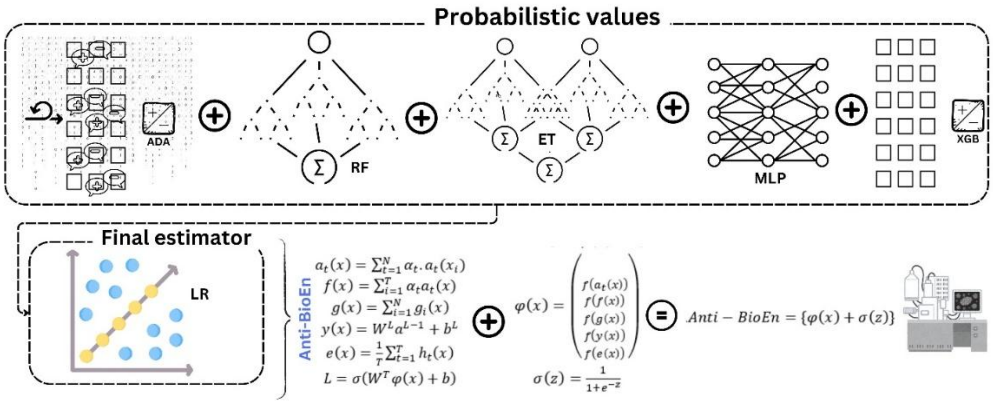

Fig 3. Overall procedure of the Anti-BioEn model based on hybrid models

4 Results

Several assessment measures were used for each bioactive molecule, including accuracy (Acc), sensitivity (Sn), specificity (Sp), precision (Pr), F1 score (F1), Kappa score (Kp), and Matthews Correlation Coefficient (MCC). These measures were found effective in solving the categorization challenge. Table 1 presents the study's overall results.

Table 1. Performance measurements of Anti-BioEn model with others applied model for 5 bioagents detection.

Bioactive agents	Model	Acc	Sn	Sp	Pr	F1	Kp	MCC
ABPs	LR	0.6868	0.5552	0.6590	0.5591	0.5774	0.4523	0.4531
AHPs		0.6060	0.4557	0.6280	0.6185	0.50	0.5724	0.5718
APPs		0.5885	0.6555	0.6244	0.6096	0.6571	0.7390	0.7391
AOPs		0.5512	0.6096	0.6571	0.6890	0.4541	0.7300	0.7330
APZPs		0.6525	0.6096	0.6571	0.7291	0.7219	0.6250	0.6299
ABPs	ADA	0.5268	0.6052	0.6673	0.6096	0.6571	0.7390	0.4400
AHPs		0.5564	0.5558	0.6460	0.5885	0.6555	0.6244	0.50
APPs		0.4593	0.7750	0.7444	0.5512	0.6096	0.6571	0.7010
AOPs		0.7151	0.8120	0.8133	0.8236	0.7046	0.6010	0.6530
APZPs		0.7022	0.8571	0.6577	0.8091	0.8010	0.6250	0.5214
ABPs	MLP	0.7268	0.7052	0.7573	0.7291	0.7070	0.4523	0.4531
AHPs		0.6460	0.5885	0.6555	0.6244	0.6460	0.6525	0.6096
APPs		0.7444	0.5512	0.6096	0.6571	0.7444	0.7390	0.6391
AOPs		0.7750	0.7444	0.5512	0.7750	0.6141	0.60	0.5612
APZPs		0.7912	0.8161	0.6574	0.7501	0.6510	0.7555	0.6100
ABPs	RF	0.7168	0.6051	0.5573	0.6571	0.5512	0.4523	0.4531
AHPs		0.7167	0.6559	0.6266	0.6010	0.6530	0.5724	0.5512
APPs		0.6495	0.6755	0.5644	0.6250	0.6299	0.4644	0.4644
AOPs		0.7550	0.7012	0.7335	0.7171	0.7142	0.7120	0.6330
APZPs		0.7171	0.7171	0.6776	0.7201	0.7010	0.6151	0.6101
ABPs	ET	0.6010	0.6530	0.5724	0.7291	0.7070	0.4523	0.4531
AHPs		0.6250	0.6299	0.7046	0.6010	0.6530	0.5724	0.5718

APPs		0.7046	0.6010	0.8010	0.6250	0.6299	0.7390	0.6391
AOPs		0.8010	0.6250	0.6500	0.6590	0.6641	0.6408	0.6031
APZPs		0.7100	0.5123	0.5535	0.6095	0.6011	0.6150	0.5881
ABPs	XGB	0.7161	0.6852	0.7171	0.7158	0.6970	0.3121	0.4031
AHPs		0.7261	0.7158	0.7161	0.7050	0.6535	0.4222	0.5112
APPs		0.8064	0.7050	0.8045	0.80	0.7158	0.5393	0.6493
AOPs		0.8100	0.80	0.8036	0.8170	0.7050	0.6404	0.6834
APZPs		0.7913	0.8170	0.7171	0.7905	0.7911	0.5555	0.5895
ABPs	**Anti-BioEn**	**0.7268**	**0.7052**	**0.7573**	**0.7291**	**0.7070**	**0.4523**	**0.4531**
AHPs		**0.7867**	**0.7558**	**0.8260**	**0.8285**	**0.7435**	**0.5724**	**0.5718**
APPs		**0.8695**	**0.8750**	**0.8644**	**0.8596**	**0.8672**	**0.7390**	**0.7391**
AOPs		**0.8750**	**0.90**	**0.8333**	**0.8590**	**0.7142**	**0.7300**	**0.7330**
APZPs		**0.8125**	**0.8571**	**0.7777**	**0.8091**	**0.8010**	**0.6250**	**0.6299**

The Anti-BioEn model was tested on a variety of bioactive chemicals, giving noteworthy results in terms of numerous classification criteria. The model's accuracy forABPs was 72.68%, with a sensitivity of 70.52% and a specificity of 75.73%. In the instance of AHPs, the model performed better, with an accuracy of 78.67%, sensitivity of 75.58%, and specificity of 82.60%. The accuracy increased further for APPs, where the model achieved an excellent 86.95% accuracy, with sensitivity and specificity values of 87.50% and 86.44%, respectively. The model achieved 87.50% accuracy for AOPs, as well as 90% sensitivity and 83.33% specificity. Finally, the performance of APZPs was excellent, with 81.25% accuracy, 85.71% sensitivity, and 77.77% specificity. These findings show that the Anti-BioEn model is an efficient approach for classifying bioactive substances, with strong performance across various kinds of peptides. The model's ability to continuously attain excellent accuracy and sensitivity across a wide range of bioactive substances highlights its promise as a dependable tool in bioinformatics research. As shown in Figure 4, Anti-BioEn achieved superior performance compared to the other methods, which is demonstrated more clearly in the results.

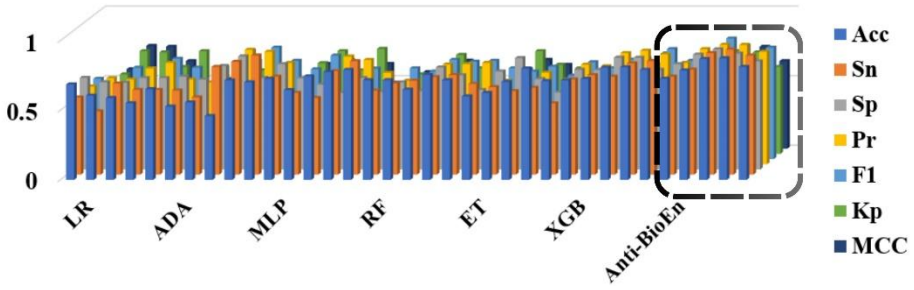

Fig 4. Demonstrates all the performance results and comparison with Anti-BioEn

5 Conclusion

This study effectively proposed a specific structure for the categorization of bioactive compounds based on the Anti-BioEn model, which uses a layering strategy to improve prediction performance. The use of a graph-based feature extraction technique has proven useful in capturing the complex molecular interactions seen in bioactive compounds. The Anti-BioEn model has greatly outperformed standard classification approaches by using a

stacking strategy to combine the strengths of numerous machine learning algorithms achieved an accuracy of 72.68% for antibacterial peptides (ABPs), 78.67% for anti-HIV peptides (AHPs), 86.95% for antioxidant peptides (APPs), 87.50% for antiparasitic peptides (AOPs), and 81.25% for antiprotozoal peptides (APZPs). Overall, the findings underscore the Anti-BioEn model's potential as a powerful tool for discovering essential bioactive molecules, enabling drug development and leading to the discovery of new therapeutic agents for crucial medical needs. Future research should focus on expanding the dataset to include a broader range of bioactive agents, which would enhance the model's robustness and applicability. Additionally, exploring alternative feature extraction methods and machine learning algorithms may provide insights into improving classification accuracy and efficiency.

Acknowledgements

N/A

Fundings

N/A

Data availability statement

Based on request

Author contribution statement

Data curation, Methodology, Data Analysis, Visualization, Conceptualization, Original Writing, Draft and Review writing, Model Applied: Md. Shazzad Hossain Shaon, Md. Fahim Sultan, Tasmin Karim; Supervision, Review: Mst Shapna Akter.

References

[1] The raising resistance of pathogenic microbes to traditional chemical medications requires the development of revolutionary treatment techniques for infectious disorders.

[2] Blair, J.M., Webber, M.A., Baylay, A.J., Ogbolu, D.O. and Piddock, L.J., 2015. Molecular mechanisms of antibiotic resistance. Nature reviews microbiology, 13(1), pp.42-51.

[3] Xu, J., Li, F., Li, C., Guo, X., Landersdorfer, C., Shen, H.H., Peleg, A.Y., Li, J., Imoto, S., Yao, J. and Akutsu, T., 2023. iAMPCN: a deep-learning approach for identifying antimicrobial peptides and their functional activities. Briefings in Bioinformatics, 24(4), p.bbad240.

[4] Magana, M., Pushpanathan, M., Santos, A.L., Leanse, L., Fernandez, M., Ioannidis, A., Giulianotti, M.A., Apidianakis, Y., Bradfute, S., Ferguson, A.L. and Cherkasov, A., 2020. The value of antimicrobial peptides in the age of resistance. The lancet infectious diseases, 20(9), pp.e216-e230.

[5] Osazee, F.O., Mokobia, K.E. and Ifijen, I.H., 2024. The urgent need for tungsten-based nanoparticles as antibacterial agents. Biomedical Materials & Devices, 2(2), pp.614-629.

[6] Ferrara, F., Capuozzo, M., Pasquinucci, R., Langella, R., Trama, U., Nava, E. and Zovi, A., 2024, May. Antibacterial agents and the fight against antibiotic resistance: A real-world evidence analysis of consumption and spending by an Italian healthcare company. In Annales Pharmaceutiques Françaises (Vol. 82, No. 3, pp. 545-552). Elsevier Masson.

[7] Brüssow, H., 2024. The antibiotic resistance crisis and the development of new antibiotics. Microbial Biotechnology, 17(7), p.e14510.

[8] Mulukutla, A., Shreshtha, R., Deb, V.K., Chatterjee, P., Jain, U. and Chauhan, N., 2024. Recent advances in antimicrobial peptide-based therapy. Bioorganic Chemistry, p.107151.

[9] Jimoh, M.A., Jimoh, M.O., Bello, M., Raimi, I.O., Okunlola, G.O., Mkhwanazi, N. and Laubscher, C.P., 2024. In vitro anti-HIV, cytotoxicity and nutritional analysis of Trianthema portulacastrum L.(Aizoaceae). BMC Complementary Medicine and Therapies, 24(1), p.35.

[10] Kubheka, M.X., Ndlovu, S.I. and Mkhwanazi, N.P., 2024. Anti-HIV Activity and Immunomodulatory Properties of Fractionated Crude Extracts of Alternaria alternata. Microorganisms, 12(6), p.1150.

[11] Halliwell, B., 2024. Understanding mechanisms of antioxidant action in health and disease. Nature Reviews Molecular Cell Biology, 25(1), pp.13-33.

[12] Pinilla-González, V., Rojas-Solé, C., Gómez-Hevia, F., González-Fernández, T., Cereceda-Cornejo, A., Chichiarelli, S., Saso, L. and Rodrigo, R., 2024. Tapping into Nature's Arsenal: Harnessing the Potential of Natural Antioxidants for Human Health and Disease Prevention. Foods, 13(13), p.1999.

[13] Bajaj, S., Singh, S. and Sharma, P., 2024. Role of antioxidants in neutralizing oxidative stress. In Nutraceutical Fruits and Foods for Neurodegenerative Disorders (pp. 353-378). Academic Press.

[14] Scalese, G., Mosquillo, M.F., Pérez-Díaz, L. and Gambino, D., 2024. Biosynthesis of ergosterol as a relevant molecular target of metal-based antiparasitic and antifungal compounds. Coordination Chemistry Reviews, 503, p.215608.

[15] Chakraborty, S., Ghosh, S., Dalui, S. and Dey, A., 2024. A review on the anti-parasitic activity of ruthenium compounds. The Journal of Basic and Applied Zoology, 85(1), p.17.

[16] Orosco, D., Mendoza, A.R. and Meléndez, C.M., 2024. Exploring the Potential of Natural Products as Antiparasitic Agents for Neglected Tropical Diseases. Current Topics in Medicinal Chemistry, 24(2), pp.89-108.

[17] Mendonça-Junior, F.J.B., 2024. Special Issue "Drug Discovery of Antiprotozoal Agents". Pharmaceuticals, 17(2), p.176.

[18] Jooste, J., Legoabe, L.J., Ilbeigi, K., Caljon, G. and Beteck, R.M., 2024. Hydrazinated geraniol derivatives as potential broad-spectrum antiprotozoal agents. Archiv der Pharmazie, p.e2400430.

[19] Abbas, H., Younus, M., Fareed, Z., Saleem, M.M., Jones, M.K., Raheemi, H., Ijaz, A. and Saleem, M.N., 2024. Antiprotozoal Resistance. In Antiparasitic Drug Resistance in Veterinary Practice (pp. 19-40). GB: CABI.

[20] Hagemann, C.L., Macedo, A.J. and Tasca, T., 2024. Therapeutic potential of antimicrobial peptides against pathogenic protozoa. Parasitology Research, 123(2), p.122.

[21] Jukič, M. and Bren, U., 2022. Machine learning in antibacterial drug design. Frontiers in Pharmacology, 13, p.864412.

[22] Ivanenkov, Y.A., Zhavoronkov, A., Yamidanov, R.S., Osterman, I.A., Sergiev, P.V., Aladinskiy, V.A., Aladinskaya, A.V., Terentiev, V.A., Veselov, M.S., Ayginin, A.A. and Kartsev, V.G., 2019. Identification of novel antibacterials using machine learning techniques. Frontiers in pharmacology, 10, p.913.

[23] Fjell, C.D., Jenssen, H., Hilpert, K., Cheung, W.A., Panté, N., Hancock, R.E. and Cherkasov, A., 2009. Identification of novel antibacterial peptides by chemoinformatics and machine learning. Journal of medicinal chemistry, 52(7), pp.2006-2015.

[24] Yang, X.G., Chen, D., Wang, M., Xue, Y. and Chen, Y.Z., 2009. Prediction of antibacterial compounds by machine learning approaches. Journal of computational chemistry, 30(8), pp.1202-1211.

[25] Lata, S., Sharma, B.K. and Raghava, G.P., 2007. Analysis and prediction of antibacterial peptides. BMC bioinformatics, 8, pp.1-10.

[26] Rosa, R.S., Santos, R.H., Brito, Á.Y. and Guimarães, K.S., 2014, July. Insights on prediction of patients' response to anti-HIV therapies through machine learning. In 2014 International Joint Conference on Neural Networks (IJCNN) (pp. 3697-3704). IEEE.

[27] Foglierini, M., Nortier, P., Schelling, R., Winiger, R.R., Jacquet, P., O'Dell, S., Demurtas, D., Mpina, M., Lweno, O., Muller, Y.D. and Petrovas, C., 2024. RAIN: machine learning-based identification for HIV-1 bNAbs. Nature Communications, 15(1), p.5339.

[28] Ge, R., Xia, Y., Jiang, M., Jia, G., Jing, X., Li, Y. and Cai, Y., 2024. HybAVPnet: A novel hybrid network architecture for antiviral peptides prediction. IEEE/ACM Transactions on Computational Biology and Bioinformatics.

[29] Hesamzadeh, P., Seif, A., Mahmoudzadeh, K., Ganjali Koli, M., Mostafazadeh, A., Nayeri, K., Mirjafary, Z. and Saeidian, H., 2024. De novo antioxidant peptide design via machine learning and DFT studies. Scientific Reports, 14(1), p.6473.

[30] Zhang, Y., Bao, X., Zhu, Y., Dai, Z., Shen, Q. and Xue, Y., 2024. Advances in Machine Learning Screening of Food Bioactive Compounds. Trends in Food Science & Technology, p.104578.

[31] Chen, L., Hu, Z., Rong, Y. and Lou, B., 2024. Deep2Pep: A deep learning method in multi-label classification of bioactive peptide. Computational Biology and Chemistry, 109, p.108021.

[32] Zhang, R., Li, Y., Li, Y. and Zhang, H., Esm4ao: A Confident Learning and Protein Language Model Based Predictor for Antioxidative Peptides Screening. Available at SSRN 4825353.

[33] Periwal, N., Arora, P., Thakur, A., Agrawal, L., Goyal, Y., Rathod, A.S., Anand, H.S., Kaur, B. and Sood, V., 2024. Antiprotozoal peptide prediction using machine learning with effective feature selection techniques. Heliyon.

[34] Xu, J., Li, F., Leier, A., Xiang, D., Shen, H.H., Marquez Lago, T.T., Li, J., Yu, D.J. and Song, J., 2021. Comprehensive assessment of machine learning-based methods for predicting antimicrobial peptides. Briefings in bioinformatics, 22(5), p.bbab083.

[35] Capecchi, A., Cai, X., Personne, H., Köhler, T., van Delden, C. and Reymond, J.L., 2021. Machine learning designs non-hemolytic antimicrobial peptides. Chemical science, 12(26), pp.9221-9232.

[36] Van Oort, C.M., Ferrell, J.B., Remington, J.M., Wshah, S. and Li, J., 2021. AMPGAN v2: machine learning-guided design of antimicrobial peptides. Journal of chemical information and modeling, 61(5), pp.2198-2207.

[37] Wang, G., Li, X. and Wang, Z., 2016. APD3: the antimicrobial peptide database as a tool for research and education. Nucleic acids research, 44(D1), pp.D1087-D1093.

[38] Jhong, J.H., Yao, L., Pang, Y., Li, Z., Chung, C.R., Wang, R., Li, S., Li, W., Luo, M., Ma, R. and Huang, Y., 2022. dbAMP 2.0: updated resource for antimicrobial peptides with an enhanced scanning method for genomic and proteomic data. Nucleic acids research, 50(D1), pp.D460-D470.

[39] Gupta, S., Sharma, A.K., Jaiswal, S.K. and Sharma, V.K., 2016. Prediction of biofilm inhibiting peptides: an in silico approach. Frontiers in Microbiology, 7, p.949.

[40] Olsen, T.H., Yesiltas, B., Marin, F.I., Pertseva, M., García-Moreno, P.J., Gregersen, S., Overgaard, M.T., Jacobsen, C., Lund, O., Hansen, E.B. and Marcatili, P., 2020. AnOxPePred: using deep learning for the prediction of antioxidative properties of peptides. Scientific reports, 10(1), p.21471.

[41] Timmons, P.B. and Hewage, C.M., 2020. HAPPENN is a novel tool for hemolytic activity prediction for therapeutic peptides which employs neural networks. Scientific reports, 10(1), p.10869.

[42] Guo, D., Ren, S., Lu, S., Feng, Z., Tang, D., Liu, S., Zhou, L., Duan, N., Svyatkovskiy, A., Fu, S. and Tufano, M., 2020. Graphcodebert: Pre-training code representations with data flow. arXiv preprint arXiv:2009.08366.

[43] Jiang, J., Ke, L., Chen, L., Dou, B., Zhu, Y., Liu, J., Zhang, B., Zhou, T. and Wei, G.W., 2024. Transformer technology in molecular science. Wiley Interdisciplinary Reviews: Computational Molecular Science, 14(4), p.e1725.

[44] Ying, C., Qi-Guang, M., Jia-Chen, L. and Lin, G., 2013. Advance and prospects of AdaBoost algorithm. Acta Automatica Sinica, 39(6), pp.745-758.

[45] Rigatti, S.J., 2017. Random forest. Journal of Insurance Medicine, 47(1), pp.31-39.

[46] Taud, H. and Mas, J.F., 2018. Multilayer perceptron (MLP). Geomatic approaches for modeling land change scenarios, pp.451-455.

[47] LaValley, Michael P. "Logistic regression." Circulation 117, no. 18 (2008): 2395-2399.

[48] Sharaff, A. and Gupta, H., 2019. Extra-tree classifier with metaheuristics approach for email classification. In Advances in Computer Communication and Computational Sciences: Proceedings of IC4S 2018 (pp. 189-197). Springer Singapore.

[49] Alexandropoulos, S.A.N., Aridas, C.K., Kotsiantis, S.B. and Vrahatis, M.N., 2019. Stacking strong ensembles of classifiers. In Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings 15 (pp. 545-556). Springer International Publishing.