International Journal of Chemical and Biochemical Sciences
(ISSN 2226-9614)

Journal Home page: www.iscientific.org/Journal.html

Leveraging deep learning to create a sophisticated biomarker for bioremediation

Ramesh Chandra Tripathi[1], Anuradha Rohinkar[2], Kavina Ganapathy[3], Megha Pandeya[4], Prabhjot Kaur[5]

[1]*Professor, College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India*
[2]*Assistant Professor, Department of Biochemistry, Parul University, PO Limda, Vadodara, Gujarat, India*
[3]*Assistant Professor, Department of Biotechnology, School of Sciences, JAIN (Deemed-to-be University), Karnataka, India*
[4]*Assistant Professor, Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Uttar Pradesh, India*
[5]*Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India.*

Abstract

The ability of microorganisms to degrade and purify pollutants underlies bioremediation's potential for environmental restoration. Finding trustworthy biomarkers that capture the dynamic interactions between microbial populations and pollutants is essential for tracking the effectiveness of bioremediation processes. Marine settings naturally contain several trace elements (TE), which are crucial for maintaining human health. *Mytilus Chilensis*, a rich source of total ethylene (TE), becomes polluted with elevated levels of TE concentration because it is grown adjacent to industrial outfalls. We describe to anticipate lower concentrations for Cu and Ni bioremediation in microbial communities using the Stochastic Firefly Optimized Ensemble Convolution Neural Network (SFO-ECNN). The goal of this project is to develop a sophisticated biomarker for bioremediation assessment. A sustainable method of environmental restoration called bioremediation uses microorganisms' ability to break down and purify contaminants. The industry is growing, and new tools to monitor and assess the efficacy of bioremediation processes are becoming increasingly necessary. Consequently, it functions as a biomarker of pollution. Determinant inequalities are used to reduce the concentration of TE and Mutual Information (MI) is maximized without the addition of any external chemicals to accomplish in situ bioremediation in *Mytilus Chilensis*.

Keywords: M. Chilensis, Stochastic Firefly Optimized Ensemble Convolutional Neural Network, Trace elements, MI

Full length article **Corresponding Author*, e-mail: rctripathig@gmail.com

1. Introduction

The vast diversity of environmental microbial communities is responsible for controlling human health and several biogeochemical cycles. Biotechnology uses these natural communities for a variety of purposes. To investigate microbial diversity, high throughput sequencing (HTS) technology is becoming more widely available [1]. Pesticide-contaminated soils can be remediated by combining microbial co-metabolism, exo-enzyme synthesis, microorganism sensitivity to pollutants and bioavailability of the contaminants. To increase the efficiency of bioremediation and encourage soil microbial activity, a range of chemical and biological strategies is used. The methods that are most in line with nature are those involving plants and earthworms, as they have clear advantages over any potential drawbacks [2]. The industrialization and population explosion are causing an alarming increase in the pollution load. Industrialization's use of chemicals in the production of high-tech advancements has led to the production of pollutants that are not biodegradable, including heavy metals, xenobiotics and hydrocarbons [3]. The main reason for the extreme strain on the Earth's ecology is the growing human population. Developing nations are under tremendous pressure to meet their economic demands and their industrialization is further harming the environment as rich nations continue to consume their resources and pollute the environment [4]. The use of living things, such as plants, algae and microorganisms, in place of physicochemical and mechanical methods to remove pollutants from contaminated environments and restore them to their pre-pollution state is known as bioremediation [5]. In freshwater, marine and terrestrial ecosystems, environmental contaminants (ECs) are building up because of increased human activity. They contaminate the environment, which lowers human welfare because they can enter the body through food or inhalation. In addition to people, aquatic and terrestrial organisms are negatively impacted by ECs [6]. Toxic chemicals can leak into the environment accidentally or on purpose, which is one way that industrial operations contribute to environmental contamination. Without a strict regulatory framework to prevent environmental pollution, industry and human activities grow quickly in developing nations. There are a considerable number of polluted locations around the globe, which could pose a risk to public health [7]. Numerous contaminants, such as hydrocarbons, industrial effluents, heavy metals, xenobiotics, pharmaceuticals, radionuclides and polychlorinated biphenyls, Poly-aromatic hydrocarbons (PAHs), have increased due to the water's rapid industrialization and population growth, soil and environment. The ecosystem, flora and human health are all harmed by these hazardous compounds [8]. Environmental pollutants that can cause several diseases in humans and animals are known to be chlorinated aromatic and aliphatic chemicals. Of these, the most difficult categories of pollutants to deal with are the polychlorinated dibenzo-p-dioxins and furans (PCDD/Fs), because of their toxicity and stubbornness [9]. One frequent, widespread and ubiquitous anthropogenic pollutant is petroleum hydrocarbon (HC) contamination. The exploitation of natural resources is among the human activities that are linked to an increase in the incidence of HC spills in these environments. These activities are the consequence of unintentional spills or previous waste management errors [10]. The study [11] listed the main problems with the traditional bioremediation method. The advantages of merging artificial intelligence with nano bioremediation are examined to address the shortcomings of a conventional strategy for cleaning up regions affected by crude oil. The study [12] provided an overview of the potential of bacterial bio-films for bioremediation against various metal ions. There has been discussion of the composition, mechanism and interspecies communication of bacteria that produce bio-films. There exist multiple methods via which bio film-associated microorganisms interact with heavy metals. The article [13] summarized the knowledge currently known regarding blood biomarkers utilized in fish stress responses produced by climate change. Fish populations as a whole or individual fish are monitored for physiological fitness using changes in informative blood-based indicators. The study [14] showed improved hydrocarbon breakdown capabilities when recovered from Equator water. The degradation investigations revealed broad growth, eliminating 93% and 83% of the phenanthrene at 0.1 and 20 MPa in 72 hours, respectively. The contamination of land, water and air by dangerous compounds brought by numerous human activities is a significant problem facing the world today. They exploit resources in a way that seems they are utilizing them recklessly [15]. The study [16] examined the function of plant growth-promoting bacteria (PGPB) in the environmentally safe, low-cost bioremediation of heavy metals. The study [17] provided an overview of the types, effects and properties of commercially available Organo-chlorine pesticides (OCP) in soil and water sources during the bioremediation process. The procedures outlined in this paper are thought to be an efficient and sustainable solution since they completely convert organic compounds into a non-toxic final product. The research [18] offered once-in-a-lifetime chance to improve the welfare of all life in planetary ecosystems and broaden knowledge of intricate biological systems. The study [19] provided a detailed description of the dynamics of microbial ecology in laboratory microcosms polluted with solvents chlorinated and 1, 4-dioxane. Local microorganisms prevailed despite bio-augmented Pseudonocardia dioxanivorans CB1190's improved 1, 4-dioxane elimination and rapid metabolism of carbohydrates along with amino acids. Gaining expertise in the quickly developing fields of generative AI and protein engineering is necessary to utilize machine-learning models as a navigational aid for protein fitness landscapes [20]. This part of the paper follows an identical format as the rest: In part 2, the methods are discussed in detail. Part 3 discusses the results and the conclusions coupled with recommendations for additional research are covered in part 4. Ecosystems and human health are threatened by environmental contamination, which calls for creative and long-lasting remediation strategies. Utilizing biological organisms to break down contaminants is a promising environmentally acceptable approach known as bioremediation. Because of the complexity of biological systems and the multitude of factors influencing the efficacy of bioremediation, it can be challenging to identify and optimize the crucial biomarkers that facilitate efficient pollutant degradation. A systematic and data-driven approach for finding and characterizing advanced biomarkers in bioremediation is lacking, which is the main obstacle to the development of targeted and optimal treatments. A more sustainable and efficient method of environmental cleanup might be fostered by the successful application of such a strategy, which has the potential to revolutionize bioremediation procedures.

2. Materials and Methods

This section covers the topic of using deep learning to develop an advanced biomarker for bioremediation. Fig.1 shows the methodological design.

2.1. Dataset

Water samples were filtered using a 0.45 µm filter membrane before nutritional analysis. A centrifuge tube was used to gather the filtered water samples, which were kept at -20°C. The bio-film equipment collected microbial water samples on days 1, 9, 17 and 25. Bio-film samples were taken from the polyethylene brush surface and stored at -80°C using sterile EP tubes. The alkaline potassium permanganate technique was used to calculate CODMn. The automated discontinuous chemical analyzer SmartChem was used to measure ammonia nitrogen, nitrate, nitrite, orthophosphate, total phosphorus (TP). These consisted of the following: The hypobromite oxidation method was used to estimate the concentration of NH_4 + -N and the potassium per-sulfate oxidation method was used to determine the amounts of TN and TP; The cadmium column reduction method was used to estimate the concentration of NO_3--N and naphthalene ethylenediamine spectrophotometry was used to assess the concentration of NO_2--N and the phosphoro-molybdenum blue spectrophotometric concentration [21].

2.2. Data Preprocessing using Z-Score Normalization

Z-score normalization, commonly referred as standardization, is a statistical method for changing a dataset's mean and standard deviation to zero and one, respectively. Each data point is taken, the dataset's mean is subtracted and the result is divided by the dataset's standard deviation. Z-score normalization helps the customers understand how a particular rating can fit into a regular, usual set of facts. Z-score is carried out to manage outliers in a collection. When the modified dataset will have a mean of zero and a standard deviation of one after Z-score normalization is applied. This normalization method is widely applied in a range of fields, such as statistics, data analysis along with machine learning, to compare and evaluate data that might have varying sizes or distributions. All variables are helped to be placed on a similar scale, which makes them easier to compare directly and more appropriate for some modeling or statistical tasks.

$$\bar{z} = \frac{z-\tau}{\varsigma} \tag{1}$$

Z stands for the quantitative component $\bar{z}$is the freshly assumed data point, τ denotes the data points' average and ς indicates the variance of the data points.

2.3. Bioremediation using stochastic firefly optimized ensemble Convolutional neural network (SFO-ECNN)

Stochastic Firefly Optimized Ensemble Convolutional Neural Network used to solve a bioremediation-related issue utilizing image data. Considering a situation where the stochastic firefly approach is applied to enhance the optimization process and ECNNs are used for image analysis in bioremediation initiatives could be envisioned. To increase accuracy and resilience, multiple ECNNs' outputs can be combined using ensemble learning.

2.3.1. Ensemble Convolutional Neural Network (ECNN)

Popular deep learning models for tasks like object detection, classification and image segmentation are Ensemble convolutional neural networks (ECNN). Because of its ability to record geographic patterns and feature hierarchies, it is helpful for tasks requiring grid-like data, like images.

Operation of Convolution

A tiny filter, referred as a kernel, is dragged across the input image during the convolution operation to extract local information. The procedure is carried out element by element and the output feature map has a single value because of summing the results. The convolution operation at a given position can be expressed using the following equation:

$$Q(j,i) = \sum_n \sum_m J(j+n, i+m) \times R(n,m) \tag{2}$$

Where

$Q(j,i)$ is the value at (j,i) in the output feature map.

$J(j+n,i+m)$ Represents the positional value $J(j+n,i+m)$ in the source image.

The positional value (n,m) in the convolution kernel is $R(n,m)$.

Function of Activation

An activation function (often known as ReLU, or Rectified Linear Unit), which introduces non-linearity, is applied element-by-element following the convolution operation. ReLU activation function is described as follows:

$$ReLU(y) = \max(0,y) \tag{3}$$

$$Q(j,i) = {}^{max}_{n,m} J(j \times g + n, i \times g + m) \tag{4}$$

Operating a pool

The spatial dimensions of the feature maps are reduced through pooling while essential information is kept. Max pooling is a common pooling technique and its equation for a specific location is:

$Q(j,i)$ is the value at location (j,i) in the output pooled map.

$J(j \times g + n, i \times g + m)$ is the value at location $(j \times g + n, i \times g + m)$ in the input feature map.
g is the stride of the pooling operation.

Layer with full connectivity

To generate final predictions, the attributes are flattened and run through many convolutional and pooling layers, then one or more fully connected layers. The equations for completely connected layers include bias addition and matrix multiplication. These are the basic formulas and elements of an ECNN.

2.3.2. Stochastic Firefly Optimization (SFO)

Fireflies are scarabs or insects with wings that give off light at night and squint. Bioluminescence is the process of making light from the lower middle area that is neither infrared nor ultraviolet. This optimization has two important steps: increasing the amount of light and making the SFO look good. For the best settings in Bilateral Filter (BF), the changed idea is used. Most of the time, it has some suspicions, which are:

i) One SFO is drawn to another SFO, no matter what sex they are. Because of this, all SFF are thought to be unisex.

ii) Bright SFO can be attracted to low-bright SFO if the amount of attraction is related to the amount of light.

iii) The setting of the target function has a direct effect on how bright the firefly is.

iv) If the two fireflies are the same color, they will move around randomly.

By walking around randomly and being drawn to each other, the fireflies create new arrangements that last for generations. The fireflies' appeal should have something to do with how well they do their job. Because they are so beautiful, they can split into smaller groups and each smaller group swarms around the nearby models. During the development of the SFO algorithm, the rate of change was rather random. The main reason for this method of adjustment is to improve the PSNR rate of the demising process. The objective function explains brightness, but other kinds of brightness could be explained in the same way. Attractiveness is measured by how bright a firefly is, so one firefly is used to figure out how far away another one is. The amount of light is written as:

$$Att_j = xjf^{-\propto r^2} \tag{5}$$

Where xj is the level of appeal and r = 0 is an absorption rate that controls how much the light power goes down. Because of this ability, they can figure out how attractive each SFF is to get the best values for the spatial and intensity kernels.

$$Light\ Att = Actual\ light\ {}_0 f^{-\propto r^2} \tag{6}$$

The actual light is the beauty of the original light when attr = 0. The distance between any two fireflies at positions N_n and N_m can be described as a Cartesian distance as follows:

$$rmm = \sqrt{\sum_{p=1}^{t}(N_n^s - N_m^s)^2} \tag{7}$$

From the above equation, we can see that the parameter is the s^{th} part of the n^{th} Firefly's spatial coordinate and t is the number of dimensions.

The push toward the Attractive SFO

The firefly is moved to the new place, but in general, it will stay in its current spot. It's required by:

$$N_n^{s+1} = N_m^s + \propto (N_n^s - N_N^s) + \sigma\left(rand - \tfrac{1}{2}\right) \tag{8}$$

Here, N_n^{s+1} is the place of the next generation's firefly σ is the parameter for the irregularity and a rand is a random number that has been safely passed around somewhere using the mutated rate.

Process changed for "rand."

The mutation rate is used to choose the random values of the growth updating process. This is based on the probability values. To reach the goal, the new chances of change at generation t+1 are made better for each site.

$$rand = \{K_2\,(e_{max} - e_w)/\left(e_{max} - e_{avg}\right), e_w \geq e_{avg}\ and\ K_{4,}e_w \leq e_{avg} \tag{9}$$

Where $max, min, and\ avg$ are the maximum, minimum and average objective functions of the current iterations as well as L4 and L2 are control factors in the range [0, 1].

Process of termination

$Iter = iter + 1$ depicts the SFF process will keep going until the best values, which are σ^s and σ^r, are found.

3. Results and Discussion

MATLAB version 7.0.0 was used in this study. Table 1 shows the concentration correlation matrix for seven trace elements. Copper and nickel contamination appears to be common near the ocean and it is typical of industrialized coastal areas. Six elements that are associated with nickel and copper had their concentrations discovered to have lower bounds. The lower bound for each of these components that maximizes the MI is displayed in Tables 2 and 3. The development of a low-cost method to reduce the TE toxicity in marine mussels before they are consumed by humans has been the focus of intensive study in recent years. The best technique for bioremediation of pollutant concentration is natural attenuation since it reduces toxicity without introducing more chemicals externally. Conversely, every single SFO-ECNN expected value surpasses the lower bound values projected by the MI, hence requiring the incorporation of external chemicals. Since natural attenuation, bioremediation does not require the addition of chemicals. To bio-remediate, the poisoning of a particular TE, such as Cu & Ni within the biomarker *Mytilus Chilensis*, the elemental concentration was lowered in this study by applying the maximizing of MI. To address the drawbacks, a sophisticated biomarker for bioremediation has been created using the SFO-ECNN technique.

Table 1. Matrix of correlation for trace element concentrations

Elements	Cr	Al	Mn	Fe	CO	Ni	Cu
Cr	1	0.85	0.88	0.86	0.30	0.19	0.22
Al	-	1	0.80	0.82	0.31	0.27	0.08
Mn	-	-	1	0.81	0.26	0.24	0.14
Fe	-	-	-	1	0.45	0.25	0.17
CO	-	-	-	-	1	0.44	0.29
Ni	-	-	-	-	-	1	0.61
Cu	-	-	-	-	-	-	1

Table 2. Comparing Ni acquired by maximizing Mutual Information (MI)

Elements	The existing value of Ni	Lower bound values by MI	SFO-ECNN
Al	0.19	0.71	0.92
Cr	0.27	0.62	0.91
Mn	0.24	0.58	0.88
Fe	0.25	0.02	0.32
CO	0.44	0.51	0.10
Cu	0.61	0.52	0.81

Table 3. Comparison of Cu derived by Mutual Information (MI) maximization

Elements	The existing value of Cu	Lower bound values by MI	SFO-ECNN
Al	0.22	0.46	0.92
Cr	0.08	0.52	0.91
Mn	0.14	0.49	0.88
Fe	0.17	0.12	0.32
CO	0.29	0.39	0.10
Ni	0.61	0.48	0.91

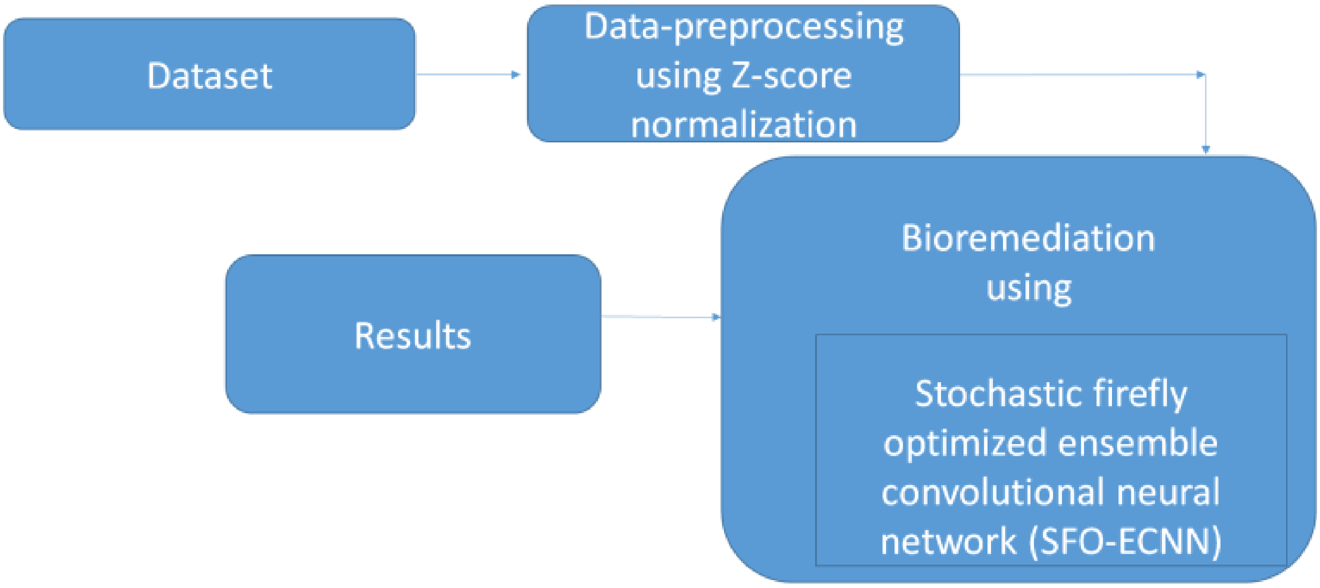

Figure 1. Methodological Design

4. Conclusions

This study shows that using SFO-ECNN to develop an advanced biomarker for bioremediation is a viable way to further environmental monitoring and remediation initiatives. The created biomarker provides a thorough and dynamic method of evaluating the efficacy of bioremediation procedures. Because of its capacity to adapt to shifting environmental factors and identify minute changes in microbial communities, this is an effective method for early intervention and the improvement of bioremediation techniques. Even though Cu and Ni's toxicity has been lowered in this study in a selected manner, we applied SFO-ECNN to expand its applicability to the bioremediation of other elements' toxicity. The advancement and improvement of sophisticated biomarkers helps to make better decisions, maximize remediation tactics that promote a more resilient and healthy environment. The combination of bioremediation and deep learning shows how innovative technologies can revolutionize environmental research and engineering by solving intricate ecological problems.

References

[1] R.B. Ghannam, S.M. Techtmann. (2021). Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. Computational and Structural Biotechnology Journal. 19 1092-1107.

[2] J.C. Sanchez-Hernandez, C. Narváez, X.A. Cares, P. Sabat, R. Naidu. (2023). Predicting the bioremediation potential of earthworms of different ecotypes through a multi-biomarker approach. Science of The Total Environment. 862 160547.

[3] H. Chandran, M. Meena, K. Sharma. (2020). Microbial biodiversity and bioremediation assessment through omics approaches. Frontiers in Environmental Chemistry. 1 570326.

[4] A. Pandey, P.H. Tripathi, A.H. Tripathi, S.C. Pandey, S. Gangola. (2019). Omics technology to study bioremediation and respective enzymes. In Smart bioremediation technologies. 23-43.

[5] Y.R. Yunusa, Z.D. Umar. (2021). Effective microbial bioremediation via the multi-omics approach: An overview of trends, problems and prospects. UMYU J. Microbiol. Res. 6 (1) 127-145.

[6] A.K. Wani, N. Akhtar, N. Naqash, C. Chopra, R. Singh, V. Kumar, J.H.P. Américo-Pinheiro. (2022). Bioprospecting culturable and unculturable microbial consortia through metagenomics for bioremediation. Cleaner Chemical Engineering. 2 100017.

[7] N.M. Jabbar, S.M. Alardhi, A.K. Mohammed, I.K. Salih, T.M. Albayati. (2022). Challenges in the implementation of bioremediation processes in petroleum-contaminated soils: A review. Environmental Nanotechnology, Monitoring & Management. 18 100694.

[8] P. Sharma, S.P. Singh, H.M. Iqbal, Y.W. Tong. (2022). Omics approaches in bioremediation of environmental contaminants: An integrated approach for environmental safety and sustainability. Environmental Research. 211 113102.

[9] H. Waseem, J. Ali, J.H. Syed, K.C. Jones. (2020). Establishing the relationship between molecular biomarkers and biotransformation rates: Extension of knowledge for dechlorination of polychlorinated dibenzo-p-dioxins and furans (PCDD/Fs). Environmental Pollution. 263 114676.

[10] K. Abdullah, D. Wilkins, B.C. Ferrari. (2023). Utilization of—Omic technologies in cold climate hydrocarbon bioremediation: a text-mining approach. Frontiers in Microbiology. 14 1113102.

[11] R. Patowary, A. Devi, A.K. Mukherjee. (2023). Advanced bioremediation by an amalgamation of nanotechnology and modern artificial intelligence for efficient restoration of crude petroleum oil-contaminated sites: a prospective study. Environmental Science and Pollution Research. 1-26.

[12] Z. Syed, M. Sogani, J. Rajvanshi, K. Sonu. (2023). Microbial biofilms for environmental bioremediation of heavy metals: a review. Applied Biochemistry and Biotechnology. 195 (9) 5693-5711.

[13] M. Shahjahan, M.J. Islam, M.T. Hossain, M.A. Mishu, J. Hasan, C. Brown. (2022). Blood biomarkers as diagnostic tools: An overview of climate-driven stress responses in fish. Science of the Total Environment. 843 156910.

[14] A.G. Kumar, D. Manisha, N.N. Rajan, K. Sujitha, D.M. Peter, R. Kirubagaran, G. Dharani. (2023). Biodegradation of phenanthrene by piezotolerant Bacillus subtilis EB1 and genomic insights for bioremediation. Marine Pollution Bulletin. 194 115151.

[15] H.I. Mohamed, M.H.A. Aal, O.M. El-Mahdy. (2024). Fungal metabolites and their role in remediation of environmental pollution. In Fungal Secondary Metabolites. Elsevier. 283-315.

[16] B.S. Saharan, T. Chaudhary, B.S. Mandal, D. Kumar, R. Kumar, P.K. Sadh, J.S. Duhan. (2023). Microbe-Plant Interactions Targeting Metal Stress: New Dimensions for Bioremediation Applications. Journal of Xenobiotics. 13 (2) 252-269.

[17] T. Kamalesh, P.S. Kumar, G. Rangasamy. (2023). An insights of organochlorine pesticides categories, properties, eco-toxicity and new developments in bioremediation process. Environmental Pollution. 122114.

[18] A. Biswas, A. Kumari, D.S. Gaikwad, D.K. Pandey. (2023). Revolutionizing biological science: The synergy of genomics in health, bioinformatics, agriculture, and artificial intelligence. OMICS: A Journal of Integrative Biology. 27 (12) 550-569.

[19] Y. Miao, T. Zhou, X. Zheng, S. Mahendra. (2023). Investigating Biodegradation of 1, 4-Dioxane by Groundwater and Soil Microbiomes: Insights into Microbial Ecology and Process Prediction. ACS ES&T Water.

[20] M. Mardikoraem, Z. Wang, N. Pascual, D. Woldring. (2023). Generative models for protein

sequence modeling: recent advances and future directions. Briefings in Bioinformatics. 24 (6) bbad358.

[21] Y. Wei, D. Shen, B. Lukwambe, Y. Wang, W. Yang, J. Zhu, Z. Zheng. (2022). The exogenous compound bacteria alter microbial community and nutrients removal performance in the biofilm unit of the integrated aquaculture wastewater bioremediation systems. Aquaculture Reports. 27 101414.