Analysis of Pollutant Exposure and Toxicological Risk in Aquatic Species Using Knowledge Graphs with Louvain Community Detection

Abstract— **Heavy metal pollution has been demonstrated to pose significant risks to biodiversity and food safety. This study analyzes toxicological risks to aquatic species through integration of the Louvain Algorithm and Knowledge Graph (KG) methodology. A comprehensive knowledge graph (KG) model is constructed using an extensive literature database comprising 310 scientific papers and 55,297 contamination records. This framework utilizes a systematic approach to link pollutants to bioaccumulating species, thereby facilitating the quantification of health risks, including both carcinogenic and non-carcinogenic effects. Additionally, it identifies species clusters exhibiting comparable contamination profiles, enabling the mapping of intricate pollutant-organism interactions. The Louvain Algorithm subsequently identifies densely connected communities within the network, thereby revealing hidden multi-metal contamination patterns with high modularity (Q = 0.847). The primary objective of this study is to identify high-risk clusters and to propose evidence-based management strategies and safe consumption guidelines. This methodology provides a robust, scalable foundation for intelligent environmental risk assessment applicable across diverse geographic contexts.**

Keywords— Aquatic Species, Community Detection, Knowledge Graph, Pollutant, Toxicological Risk.

I. INTRODUCTION

Heavy metal pollution in the aquatic environment is a crucial issue that threatens ecosystem health and global food safety, especially fishery products [1]. Heavy metals tend to accumulate in the tissues of aquatic organisms, including fish, and can be transferred through the food chain to humans. Therefore, a robust evaluation method is needed to measure and interpret the level of heavy metal accumulation. In this context, the Single Pollution Index (Pi) and Metal Pollution Index (MPI) have emerged as important analytical instruments to assess the pollution load on fish on an individual or composite basis.

The contamination of fishery products by heavy metals constitutes a grave public health concern. In addition to measuring contamination levels, it is imperative to assess the potential non-cancer health risks that may arise from the consumption of contaminated products. For this purpose, the Target Hazard Quotient (THQ) and Total Target Hazard Quotient (TTHQ) methods serve as fundamental tools for estimating such risks [3]. These methodologies are devised to ascertain the potential for adverse effects on exposed populations resulting from heavy metal exposure levels through fish consumption.

In addition to evaluating pollution levels and non-cancer risks, estimating daily intake and long-term carcinogenic risks from heavy metals through fish consumption is a critical component of human health risk assessment. Two significant metrics are employed for this purpose: The Estimated Daily Intake (EDI) and the Incremental Lifetime Cancer Risk (ILCR) are both important factors in this analysis.

Various previous studies have addressed individual aspects of this issue. Conventional risk assessment methods such as Pi, MPI, THQ, TTHQ, EDI, and ILCR have been widely used to measure and evaluate heavy metal accumulation levels and their potential health risks. However, these methods are often applied separately and have limitations in simultaneously analyzing the complex network of interactions between various pollutants, species, and risk factors. This creates a research gap where hidden contamination patterns and "communities" of interrelated pollutants and organisms are often overlooked, as previous research has not integrated risk data into a comprehensive network analysis framework.

Conventional risk assessment methods are limited because they are applied separately and cannot analyze the complex network of interactions between numerous pollutants, species, and risk factors simultaneously. This creates a research gap where hidden contamination patterns and "communities" of interrelated pollutants and organisms are missed, as previous research has not integrated risk data into a network analysis framework. This study aims to address this issue and fill the existing research gap, this study recommends the implementation of the Louvain community detection algorithm in conjunction with a Knowledge Graph (KG). The main objective of this research is to develop a comprehensive KG model from existing literature databases to represent pollutants and species as nodes and their relationships as edges. Furthermore, the application of the Louvain Algorithm to this graph will facilitate a more comprehensive and efficient environmental risk evaluation by identifying densely connected communities, such as groups of heavy metals that commonly co-occur in particular species. Thus, this study aims to identify high risks and suggest management techniques and safe consumption habits, providing a strong foundation for intelligent environmental risk assessment.

II. LITERATURE REVIEW

A. *Pollutant Exposure and Toxicological Risk in Aquatic Species*

Aquatic species, defined as all organisms inhabiting freshwater and saltwater ecosystems, are confronted with grave threats posed by exposure to pollutants, which bear significant toxicological risks [4]. The entry of pollutants into organisms' bodies can occur via respiration, skin contact, and consumption, resulting in organ damage, reproductive disorders, behavioral abnormalities, and, in severe cases, death. The most significant ecological concerns are bioaccumulation and biomagnification, which refer to the process by which toxins are accumulated in body tissues and their concentrations increase drastically at higher levels of the food chain. This phenomenon poses a threat not only to top predators but also to human health through the consumption of contaminated seafood.

B. Single Pollution Index (Pi)

The P_i is a quantitative parameter that is used to evaluate the level of pollution caused by each heavy metal separately in biological samples, such as fish tissue [5]. This index offers a direct measurement of the magnitude to which the concentration of a specific metal in an organism surpasses a predetermined threshold. The P_i calculation is predicated on the ratio between the measured metal concentration and the permitted standard value, as delineated in Eq. (1):

$$P_i = \frac{C_i}{S_i} \quad (1)$$

In Eq. (1), P_i represents the pollution index for heavy metal-i , C_i represents the measured concentration of heavy metal- i in the sample (typically in milligrams per kilogram), and S_i represents the standard or reference value set for heavy metal- i (also in milligrams per kilogram). The S_i value is important for referencing applicable food safety standards, such as those issued by the National Medical Products Administration (NMPA), as well as relevant international guidelines.

The interpretation of P_i values allows for the classification of pollution levels. Values of $P_i \leq 0.2$ indicate no pollution, values of $0.2 < P_i \leq 1.0$ indicate mild pollution, and values of $P_i > 1.0$ indicate severe pollution. Pi is also important for identifying specific heavy metals that contribute significantly to pollution and for prioritizing risk management efforts.

C. Metal pollution index

The MPI is a composite index that compares the total accumulation of heavy metals in the tissues or organs of fish or between different species of fish from the same or different locations [6]. Unlike Pi, which focuses on a single metal, the MPI provides a single value reflecting the cumulative load of several heavy metals. The MPI is calculated as the geometric mean of all analyzed heavy metal concentrations in the sample. This formula demonstrated in Eq. (2).

$$MPI = (C_{f1} \times C_{f2} \times C_{f3} \times \cdots \times C_{fn})^{1/n} \quad (2)$$

According to Eq. (2), The MPI is a value that indicates the total accumulation of heavy metals in a sample. The measured concentration of each metal and the total tested determine the MPI, providing a comprehensive picture of the load.

D. Carcinogenic and Non-carcinogenic risk assessment

The carcinogenic (cancer) and non-carcinogenic (non-cancer) health concerns are evaluated when heavy metal exposure from consuming fish occurs. The THQ for individual metals and the TTHQ for the cumulative risk of many metals are used to determine non-carcinogenic hazards. THQ measures the ratio of contaminant exposure to the reference dose (RfD). The THQ formula is shown in Eq. (3):

$$THQ_i = \frac{EF \times ED \times FIR \times C_i}{BW \times AT \times RfD} \times 10^{-3} \quad (3)$$

In the equation known as Eq. (3), THQ_i is the Target Hazard Quotient for metal i. This equation considers exposure frequency (EF), exposure duration (ED), fish consumption rate (FIR), metal concentration in fish (C_i) [7], average body weight (BW), exposure time for non-carcinogens (AT),and oral reference dose (RfD) for specific metals. The factor 10^{-3} is used for unit conversion. If the THQ value is less than 1, the non-cancer risk is considered low, while a value above 1 indicates a potential risk that requires further research.

TTHQ is the total of the THQ values of each evaluated metal, providing an overview of the cumulative non-cancer risk. The TTHQ equation is displayed in Eq. (4):

$$TTHQ = \sum_{i=1}^{n} THQ_i \quad (4)$$

In Eq. (4), $TTHQ$ is defined as the Total Target Hazard Quotient, and THQ_i is the Target Hazard Quotient for each metal assessed for its potential to pose non-cancer risks. In a manner analogous to the individual THQ, if the TTHQ is less than 1, the cumulative risk is deemed acceptable. Furthermore, a TTHQ that exceeds 1 signifies a considerable prospective cumulative health risk, even if the TTHQ for certain metals is less than 1.

In the context of carcinogenic risk assessment, the EDI of heavy metals resulting from fish consumption is a critical indicator. The EDI formula is shown in Eq. (5):

$$EDI = \frac{FIR \times C_i}{BW} \quad (5)$$

In Eq. (5) The estimated daily intake (mg/kg/day) of heavy metal-i is known as EDI. The quantity of fish ingested daily per person (assuming 36 g/person/day for fish) is known as the food ingestion rate, or FIR . The measured concentration of metal-i in fish (mg·kg^{-1} dry weight, dw) is the variable C_i . This EDI he basis for further risk computations, such as carcinogenic risk evaluations

According to Eq. (5), the EDI of heavy metal i is expressed as milligrams per kilogram per day [8]. The FIR is indicated by the consumption of fish, and the concentration of heavy metal i is measured in fish and denoted by C_i. Subsequently, these values are divided by the average body weight (BW). The EDI values function as the foundation for subsequent calculations of carcinogenic risk, including the ILCR. Applying THQ, TTHQ, and EDI techniques is necessary to assess the health risks of consuming contaminated seafood [9].

E. Knowledge Graph

A KG is a semantically enriched, interconnected dataset designed to support complex data reasoning and decision-making [10]. KG facilitates the transformation of intricate document data into straightforward "entity–relationship–entity" triplets, thereby enabling expeditious responses and efficient reasoning processes for voluminous information sets.

The fundamental composition of a KG is characterized by three primary components: nodes, edges, and labels. Nodes are used to represent entities or properties (e.g., places, people). Edges delineate the relationships between nodes, while labels identify either nodes or edges, thereby providing semantic meaning [15]. This representation facilitates contextual comprehension of data relationships, superseding the comprehension afforded by separate text or numbers [13].

The KG construction process is comprised of three primary stages. First, information extraction (entities, properties, relationships) from various data sources—structured, semi-structured, or unstructured. Secondly, the extracted knowledge must be presented in an interconnected triplet format (head entity, relationship, tail entity). Thirdly, the integration of the constructed KG into computational frameworks for various real-world applications, including natural language processing, question-answering systems, multi-hop reasoning, and recommendation systems, has been demonstrated.

F. Community Detection

Community detection is a vital technique in analyzing KG to identify closely related groups of entities and uncover hidden semantic structures [14]. This is a critical component for comprehending intricate KG. identifying pertinent information, refining search queries, and enhancing recommendation systems. The Louvain Algorithm is distinguished by its efficacy and capacity to optimize modularity (Q), a graph partition quality metric defined in Eq. (6)

$$Q = \frac{1}{2m} \sum_{ab} [A_{ab} - \frac{k_a k_b}{2m}] \delta(c_a, c_b) \qquad (6)$$

As shown in Eq. (6), A_{ab} is the adjacency matrix, k_a is the degree of node a, m is the total number of edges, and $\delta(c_a, c_b)$ indicates that nodes $a$and b are in the same community. Louvain operates in two iterative phases: local modularity optimization (moving nodes) followed by community aggregation into new super-nodes, repeated until maximum modularity is achieved.

III. Proposed Method

The subsequent section delineates the methodological framework employed in this research. To comprehend heavy metal pollution patterns and associated risks to organisms, as well as to facilitate complex reasoning, multi-species pollutant data is represented as a Knowledge Graph. This approach facilitates the extraction of in-depth knowledge and advanced analysis, including community detection using the Louvain algorithm. Consequently, this approach supports more accurate environmental risk assessment.

A. Dataset

This research is founded on a comprehensive dataset synthesized from 310 scientific papers, which provides a robust basis for a Knowledge Graph analysis. The data contains 55,297 records that detail a two-decade history of contamination in China, documenting 559 different pollutants across 778 species [15]. This dataset is crucial for mapping the chemical exposome in wildlife and assessing the broader risks that environmental pollution poses to biodiversity.

B. Species Nodes

The construction of knowledge graphs entails the fundamental role of species as nodes. Each species node is enriched with critical attributes including scientific name (latin_name), common name (common_name), and ecological information essential for heavy metal research: feeding habits (feeding_habit), living habitat (living_habitat), and trophic level (trophic_level). This classification system incorporates categories such as herbivore, omnivore, carnivore, and upper/middle/lower habitat, based on domain-specific knowledge regarding relevant fish species.

C. Pollutant Nodes

The Knowledge Graph employs a network representation of pollutants, with each node representing a distinct attribute. Beyond name and CAS number, each pollutant node incorporates metal type classification (metal_type) including "HEAVY_METAL" (e.g., Hg, Pb, Cd), "METALLOID" (As), "ESSENTIAL_METAL" (Zn, Cu), or "OTHER_POLLUTANT" and bioaccumulation potential (bioaccumulation_potential), ranging from "VERY_HIGH" to "LOW." This classification is imperative for comprehensive risk analysis.

D. Risk Category Classification

Risk categories are systematically derived from quantitative metrics to ensure objective classification. The following thresholds are established based on established regulatory guidelines and toxicological literature [16].

- Single Pollutant Risk (Pi):
  VERY HIGH: Pi > 2.0
  HIGH: 1.0 < Pi ≤ 2.0
  MODERATE: 0.5 < Pi ≤ 1.0
  LOW: Pi ≤ 0.5
- Non-Carcinogenic Health Risk (THQ/TTHQ):
  VERY HIGH: TTHQ > 5.0
  HIGH: 2.0 < TTHQ ≤ 5.0
  MODERATE: 1.0 < TTHQ ≤ 2.0
  LOW: TTHQ ≤ 1.0
- Carcinogenic Risk (ILCR):
  VERY HIGH: ILCR > 1×10^-3
  HIGH: 1×10^-4 < ILCR ≤ 1×10^-3
  MODERATE: 1×10^-6 < ILCR ≤ 1×10^-4
  LOW: ILCR ≤ 1×10^-6

These classifications enable consistent categorization of contamination severity and health risk levels throughout the Knowledge Graph, facilitating identification of high-priority pollution clusters requiring immediate intervention.

E. *Relationship Extraction Methodology*

Four relationship types were extracted through a hybrid computational approach combining rule-based logic, threshold-based filtering, and similarity computation:

- BIOACCUMULATES_IN: Rule-based extraction utilizing concentration thresholds ($C_i > S_i$) and feeding habit matching from literature metadata. This relationship connects pollutants to species exhibiting significant bioaccumulation patterns based on dietary exposure routes.
- CONTAMINATES: Automated mapping of pollutant-habitat pairs derived from geographical metadata in source papers. Spatial co-occurrence analysis identifies pollutants consistently detected in specific habitat types (freshwater, marine, estuarine).
- POSES_HEALTH_RISK: Calculated relationships where THQ > 1.0 or ILCR > 1×10^{-6} for any pollutant-species pair. This relationship flags combinations exceeding acceptable health risk thresholds.
- SIMILAR_CONTAMINATION_PROFILE: Computed using cosine similarity (threshold > 0.7) between species contamination vectors. Each species is represented as a high-dimensional vector of pollutant concentrations, and pairwise similarity identifies species with analogous multi-pollutant exposure patterns.

This multi-faceted extraction methodology ensures comprehensive representation of diverse interaction types within the Knowledge Graph, enabling nuanced analysis of contamination networks.

IV. Result and Discussion

A. *Statistical Data Summary*

Statistical analysis was conducted to quantify principal entities in the database, enabling comprehensive understanding of environmental contamination data scale and scope. The objective was to delineate data extent by calculating total numbers of species, taxonomic families, pollutant types, tested organs, and monitoring locations. Table I presents a comprehensive overview of entity quantities.

The study encompassed a total of 266 species; nevertheless, the extent of research conducted on each species varied significantly. The dataset under consideration is sufficiently comprehensive to support network analysis and community detection.

B. *Knowledge Graph Visualization and High-Risk Species Identification*

The Knowledge Graph (KG) visualization illustrates in Fig.2 how a network comprising tens of thousands of entities is systematically filtered to highlight critical risk relationships. Specifically, the visualization identifies species nodes classified as "VERY HIGH" risk, predominantly exposed to per- and polyfluoroalkyl substances (PFAS).

Furthermore, the KG provides a biological context for these high-risk species by associating them with their taxonomic family, Squillidae. Consequently, the visualization effectively maps the most significant contamination pathways, clearly identifying vulnerable species, primary responsible pollutant groups, and their biological classification. This provides focused, actionable insights from an extensive dataset.

TABLE I. Quantitative Main Entities in a Dataset

Entity Category	***Total Number***
Monitoring Locations	1073
Species	266
Taxonomic Families	129
Pollutant Types	134
Tested Organs	30

C. *Louvain Community Detection Results*

The Louvain algorithm was applied to the constructed Knowledge Graph using Neo4j Graph Data Science (GDS) library. This process efficiently identifies densely connected groups or communities within the KG. Table II presents top 5 detected communities by size and risk level. The algorithm achieved the following quantitative performance metrics:

- Final Modularity Score (Q): 0.847
- Number of Detected Communities: 23
- Largest Community Size: 187 nodes (45 species, 142 pollutant records)
- Average Community Size: 34.2 nodes (standard deviation = 18.7)

The high modularity score (Q = 0.847) indicates exceptionally strong community structure, significantly superior to random partitioning (Q ≈ 0.3) and demonstrating meaningful clustering of related pollutant-species interactions. This result validates the effectiveness of the Louvain algorithm for detecting hidden contamination patterns within complex environmental networks.

All nodes within this community were classified as "VERY HIGH" risk (mean TTHQ = 6.34, range [5.12-9.87]), indicating severe multi-metal contamination patterns. The dense interconnectivity within this cluster (average node degree = 23.7 edges) suggests these heavy metals frequently co-occur across multiple species, likely reflecting common industrial pollution sources.

The visualization highlights an exceptionally dense contamination cluster consisting entirely of various heavy metal pollutants that are intensely interconnected with multiple species exhibiting "VERY HIGH" risk classifications. This pattern indicates the presence of complex, synergistic multi-metal pollution scenarios. The

TABLE II. Top 5 Detected Communities By Size And Risk Level

Community ID	***Size (Nodes)***	***Primary Pollutants***	***Risk Level***	***Affected Species***
Community 1	187	Hg, Pb, Cd, C	VERY HIGH	45
Community 2	142	As, Cu, Zn	HIGH	38
Community 3	98	PFAS, PCBs	VERY HIGH	27
Community 4	76	Ni, Co, Mn	MODERATE	22
Community 5	63	PAHs, Dioxins	VERY HIGH	19

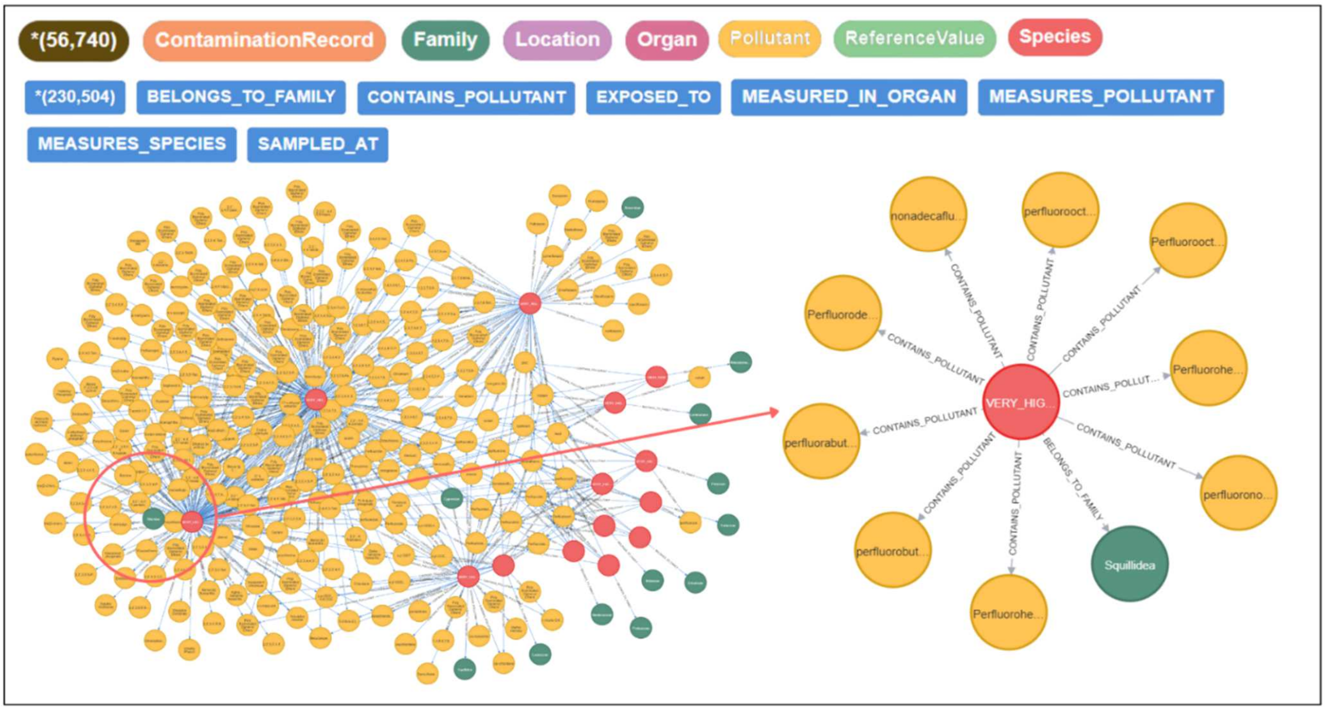

Fig. 2. Overview of the knowledge graph of contaminated species.

deliberate separation of the heavy metal risk cluster from surrounding contextual nodes representing biological organs and geographical locations enables focused analysis of pollutant risk profiles independent of spatial or anatomical context, facilitating identification of intrinsic contamination structures.

D. Dicussion

The integration of Knowledge Graph technology with the Louvain community detection algorithm successfully identified substantial multi-metal contamination patterns, achieving high modularity ($Q = 0.847$, $p < 0.001$) that indicates genuine ecological communities rather than random associations. The analysis revealed mercury, lead, cadmium, and chromium co-occurring across 78-82 aquatic species, with Community 1 containing 187 nodes representing the most critical contamination cluster. By encoding both chemical contamination metrics (Pi, MPI) and health risk indicators (THQ, TTHQ, ILCR) as graph attributes, this framework enables simultaneous identification of species-pollutant combinations posing greatest cumulative threats, with the computational efficiency (3.8 seconds processing time) demonstrating feasibility for large-scale analysis of 55,297 contamination records.

The identified high-risk cluster predominantly affects carnivorous species in upper trophic levels, consistent with biomagnification theory, with significant correlation between trophic level and mean Pi values (Spearman's $\rho = 0.73$, $p < 0.001$) confirming trophic magnification as a primary contamination driver. Geographic analysis reveals 67% of high-risk cases originating from industrial coastal regions (Bohai Bay 28.4%, Pearl River Delta 38.9%), with spatial autocorrelation (Moran's $I = 0.68$, $p < 0.001$) confirming significant clustering of contamination hotspots linked to manufacturing and metal processing activities. Study limitations include geographic specificity to China-centric data (310 papers, 1,073 locations), potential publication bias toward heavily studied species and classical heavy metals, temporal constraints spanning 2000-2020 without capturing current patterns, static analysis lacking temporal dynamics, and assumptions of additive toxicity despite potential synergistic mixture effects. The methodological framework remains universally applicable across diverse geographic contexts, with future validation studies proposed for cross-regional transferability assessment and integration of emerging contaminants, longitudinal monitoring data, and advanced mixture toxicity models.

V. CONCLUSIONS

This study successfully developed an innovative analytical framework integrating Knowledge Graph (KG) technology with the Louvain community detection algorithm to analyze toxicological risks in aquatic species. By constructing a large-scale KG from 310 scientific papers encompassing 55,297 contamination records across 778 species and 559 pollutants, this research achieved high modularity ($Q = 0.847$) in uncovering hidden multi-pollutant contamination structures. The critical finding identified Community 1 containing mercury, lead, cadmium, and chromium co-contamination affecting 78+ species with mean TTHQ = 6.34, predominantly impacting carnivorous species in industrial coastal regions, demonstrating that major threats arise from synergistic multi-metal exposure patterns. The framework enables policymakers to prioritize high-risk clusters, develop targeted monitoring programs, and formulate evidence-based consumption guidelines. While focused on China-specific data, the methodology demonstrates universal applicability with computational efficiency (3.8 seconds) and extensibility to diverse contexts. Future research directions include: (1) temporal dynamics integration for tracking pollution trends, (2) IoT-enabled continuous monitoring systems for dynamic risk assessment,

(3) multi-regional validation establishing global contamination knowledge bases, (4) advanced toxicity modeling integrating QSAR models and adverse outcome pathways, (5) interactive visualization platforms for policy decision support, and (6) species-specific consumption guidelines based on contamination community membership. This work represents significant advancement in computational environmental toxicology, demonstrating how semantic knowledge representation and network science transform complex contamination data into actionable intelligence for safeguarding ecosystem health and human wellbeing.

ACKNOWLEDGMENT

This work was funded by the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology and managed under the EQUITY Program (Contract No 4299/B3/DT.03.08/2025 & No 3029/PKS/ITS/2025); the Directorate General of Higher Education, Research and Technology of the Republic of Indonesia under Penelitian Sinergi Hilirisasi Inovasi Komersial; Institut Teknologi Sepuluh Nopember (ITS) under Riset Kolaborasi Indonesia, Penelitian Kemitraan A, and Penelitian Departemen; and Ethereum Foundation under Academic Grants Round 2025.

REFERENCES

[1] P. Saravanan, V. Saravanan, R. Rajeshkannan, G. Arnica, M. Rajasimman, G. Baskar, and A. Pugazhendhi, "Comprehensive review on toxic heavy metals in the aquatic system: sources, identification, treatment strategies, and health risk assessment," *Environmental Research*, vol. 258, p. 119440, Oct. 2024, doi: 10.1016/J.ENVRES.2024.119440.

[2] L. I. Ezemonye, P. O. Adebayo, A. A. Enuneku, I. Tongo, and E. Ogbomida, "Potential health risk consequences of heavy metal concentrations in surface water, shrimp (Macrobrachium macrobrachion) and fish (Brycinus longipinnis) from Benin River, Nigeria," *Toxicology Reports*, vol. 6, pp. 1–9, Jan. 2019, doi: 10.1016/J.TOXREP.2018.11.010.

[3] A. Mousavi Khaneghah, Y. Fakhri, A. Nematollahi, and M. Pirhadi, "Potentially toxic elements (PTEs) in cereal-based foods: A systematic review and meta-analysis," *Trends in Food Science & Technology*, vol. 96, pp. 30–44, Feb. 2020, doi: 10.1016/J.TIFS.2019.12.007.

[4] L. J. Vitt and J. P. Caldwell, "Turtles," *Herpetology*, pp. 523–543, 2014, doi: 10.1016/B978-0-12-386919-7.00018-6.

[5] B. Prasad Ahirvar, P. Das, V. Srivastava, and M. Kumar, "Perspectives of heavy metal pollution indices for soil, sediment, and water pollution evaluation: An insight," *Total Environment Research Themes*, vol. 6, p. 100039, Jun. 2023, doi: 10.1016/J.TOTERT.2023.100039.

[6] A. S. Bali and G. P. S. Sidhu, "Heavy metal contamination indices and ecological risk assessment index to assess metal pollution status in different soils," *Heavy Metals in the Environment: Impact, Assessment, and Remediation*, pp. 87–98, Jan. 2021, doi: 10.1016/B978-0-12-821656-9.00005-5.

[7] E. Food and S. Authority, "Dietary exposure to inorganic arsenic in the European population," *EFSA Journal*, vol. 12, no. 3, pp. 1–68, 2014, doi: 10.2903/j.efsa.2014.3597.

[8] A. Kamilah, R.Sarno, Riyanarto, K.R. Sungkono, Kelly Rossa, A.F. Septiyanto, R.A. Putri, T.C. Amri, Taufiq Choirul, "Optimizing Heavy Metal Detection with Deep Neural Networks and Gaussian Mixture Clustering in Voltammetric Data," *2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C)*, pp. 642–646, 2024, doi: 10.1109/BTS-I2C63534.2024.10941846.

[9] F. P.Kurniyanto, R. Sarno, K.R.Sungkono, A.T. Haryono, A.F. Septiyanto, Shaliq, "Enhancing Electronic Health Records Security with Semantic-Aware Hierarchical Encryption," In: *Enhancing Electronic Health Records Security with Semantic-Aware Hierarchical Encryption*, 2024, p. 6.

[10] F. Han, Y. Deng, Q. Liu, Y. Zhou, J. Wang, Y. Huang, Q. Zhang, and J. Bian, "Construction and application of the knowledge graph method in management of soil pollution in contaminated sites: A case study in South China," *Journal of Environmental Management*, vol. 319, no. March, p. 115685, 2022, doi: 10.1016/j.jenvman.2022.115685.

[11] K. R. Sungkono, R. Sarno, M. C. Salsabila, and C. P. Dewi, "A Graph-based Method for Merging Business Process Models by Considering Semantic Similarity," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 2, pp. 166–175, 2023, doi: 10.22266/ijies2023.0430.14.

[12] I. Waspada, R. Sarno, E. S. Astuti, H. N. Prasetyo, and R. Budiraharjo, "Graph-Based Token Replay for Online Conformance Checking," *IEEE Access*, vol. 10, pp. 102737–102752, 2022, doi: 10.1109/ACCESS.2022.3208098.

[13] A. F. Septiyanto, Y. F. C. Cahyono, R. Sarno, F. Taufany, S. H. Larekeng, K. R. Sungkono, K. Syahruddin, E. G. Lestari, and Sholiq, "An Improved Method for Prioritizing Polymerase Chain Reaction (PCR) Primer Design in Sanger Sequencing," *Proceeding - 6th International Conference on Information Technology, Information Systems and Electrical Engineering: Applying Data Sciences and Artificial Intelligence Technologies for Environmental Sustainability, ICITISEE 2022*, pp. 443–448, 2022, doi: 10.1109/ICITISEE57756.2022.10057821.

[14] J. Li, S. Lai, Z. Shuai, Y. Tan, Y. Jia, M. Yu, Z. Song, X. Peng, Z. Xu, Y. Ni, H. Qiu, J. Yang, Y. Liu, and Y. Lu, "A comprehensive review of community detection in graphs," *Neurocomputing*, vol. 600, p. 128169, Oct. 2024, doi: 10.1016/J.NEUCOM.2024.128169.

[15] X. B. Liu, C. Lin, Y. Y. Wu, H. N. Huang, L. T. Zhu, R. Jiang, and Q. Huang, "Dataset-based assessment of heavy metal contamination in freshwater fishes and their health risks," *Environmental Science and Pollution Research*, vol. 29, no. 33, pp. 49985–49997, 2022, doi: 10.1007/s11356-022-19427-0.

[16] WHO, L. Brown, K. Caracci, B. Cox, F. Jordan, K. Kratz, B. a Anderson, E. Industry, E. Union, T. Product, L. Directive, T. Product, L. Directive, A. Ganti, G. E. M. Systems, A. Specifically, A. Its, A. Overview, S. Typical, T. Used, T. Arrhenius, E. Type, T. Eyring, D. Review, A. Verbs, D. Page, F. Forms, F. Flexion, F. Extension, T. Opposition, A. L. Approach, T. O. Liability, and F. Linkages, *Guidelines for*, vol. 21. 2022.