Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

SysFungiNet: A Multi-Omics Data Fusion Framework with Explainable AI for Bioactive Prioritization

Johnson Bisi Oluwagbemi [1,*], Olusegun Victor Oyetayo [2], and Emmanuel Onwuka Ibam [3]

[1] Department of Computer Science, McPherson University, Seriki-Sotayo 110001, Ogun State, Nigeria; e-mail : oluwagbemijb@mcu.edu.ng
[2] Department of Microbiology, Federal University of Technology, Akure 340001, Ondo State, Nigeria; e-mail : vooyetayo@futa.edu.ng
[3] Department of Information Systems, Federal University of Technology, Akure 340001, Ondo State, Nigeria; e-mail : eoibam@futa.edu.ng
* Corresponding Author : Johnson Bisi Oluwagbemi

Abstract: Macrofungi constitute a vast yet underexplored reservoir of bioactive secondary metabolites; however, their systematic characterization remains constrained by fragmented analytical workflows and the limited interpretability of conventional machine learning approaches. This study presents SysFungiNet, an explainable and FAIR-compliant systems-bioinformatics framework designed for high-throughput bioactive prioritization in non-model macrofungi. The proposed framework employs a synchronized five-layer architecture that integrates LC–MS/MS metabolomics, transcriptomics, and genomics through a hybrid ensemble of XGBoost and graph neural networks (GNNs). A late-fusion strategy is adopted to combine supervised feature filtering with topology-aware molecular graph learning, enabling robust integration of structural and multi-omics evidence. The framework was validated using datasets from *Ganoderma lucidum* and *Craterellus cornucopioides*. SysFungiNet achieved a Pathway Completeness Index (PCI) of 0.86 and a peak F1-score of 0.91, corresponding to an improvement of approximately 18% over existing discovery frameworks. Feature-group ablation analysis demonstrated that cheminformatic structural descriptors constitute the primary discriminative signal for bioactivity prediction, with a 12% decline in performance upon removal, while transcriptomic integration served as a critical functional filter, reducing false-positive prioritization by up to 19%. Model decisions were rendered transparent through SHAP-based explainability, enabling direct attribution of predicted bioactivity to molecular substructures, biosynthetic pathway evidence, and gene expression signals. Overall, SysFungiNet establishes a scalable, containerized, and reproducible computational ecosystem that transforms fungal bioprospecting from a fragmented, trial-and-error process into an evidence-driven discovery pipeline. The framework provides a generalizable template for explainable multi-omics integration in non-model organisms.

Keywords: Bioactive Discovery; Cheminformatics; Explainable Artificial Intelligence; FAIR Data Principles; Macrofungi Metabolomics; Multi-Omics Integration; Systems Bioinformatics; XGBoost–Graph Neural Networks.

Received: December, 2nd 2025
Revised: January, 10th 2026
Accepted: January, 11th 2026
Published: January, 15th 2026
Curr. Ver.: January, 15th 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Macrofungi represent an immense yet largely underexplored reservoir of bioactive secondary metabolites, including structurally diverse polysaccharides, triterpenoids, and phenolic compounds with demonstrated therapeutic potential [1], [2]. Despite rapid advances in high-throughput LC-MS/MS and next-generation sequencing technologies, the systematic chemical and biological characterization of wild and non-model macrofungal species remains limited. A key bottleneck lies in the fragmented integration of heterogeneous omics datasets, which hampers coherent biosynthetic interpretation and scalable discovery [3]. Existing discovery pipelines frequently rely on single-omics analyses or black-box machine learning

models, offering limited biological interpretability and weak mechanistic grounding for experimental validation [4].

While recent platforms such as MetaFungi [5] have improved pathway annotation and metabolite discovery in model organisms, they often fail to capture the cryptic biosynthetic gene clusters (BGCs) and complex gene–metabolite relationships characteristic of underexplored tropical macrofungi [6]. Moreover, many current AI-based prediction models provide limited insight into why a compound is predicted to be bioactive, thereby restricting the identification of causal chemical substructures and biosynthetic drivers [7]. These limitations collectively underscore the need for an integrative, interpretable, and system-level computational framework tailored to the biological complexity of macrofungi.

To address these challenges, we propose SysFungiNet, a FAIR-compliant systems-bioinformatics framework that unifies multi-omics integration, pathway reconstruction, and explainable bioactive prioritization. The framework adopts a hybrid architecture combining gradient-boosted decision trees (XGBoost) and graph neural networks (GNNs), motivated by recent evidence that conventional deep learning approaches struggle with the high dimensionality, sparsity, and noise inherent in multi-omics data [8]. In SysFungiNet, XGBoost is employed for feature filtering and supervised graph construction, enabling robust identification of informative transcriptomic and metabolomic signals while reducing biological noise [9]. These selected features are subsequently embedded into a GNN, which preserves the topological structure of molecular graphs and metabolic networks and can learn non-Euclidean relationships between chemical structure and biosynthetic context [10], [11].

This hybrid strategy is essential because bioactivity in macrofungi emerges from the interplay between biosynthetic feasibility and molecular binding potential. By triangulating genomic, transcriptomic, metabolomic, and structural evidence, SysFungiNet generates an integrated Bioactive Potential Score, supported by SHAP-based interpretability to enhance transparency and biological plausibility. The framework is further aligned with FAIR principles through containerized workflows, enabling reproducibility, scalability, and cross-laboratory deployment. Collectively, SysFungiNet aims to shift fungal bioprospecting from a fragmented, trial-and-error paradigm toward a targeted, evidence-driven discovery process.

The primary objective of this study is to design and validate SysFungiNet as a unified systems-bioinformatics framework for accelerating the discovery, characterization, and prioritization of bioactive compounds from underutilized macrofungi. Specifically, this work investigates how heterogeneous multi-omics data can be efficiently integrated to reconstruct biosynthetic pathways, how computational strategies can support reliable metabolite annotation and explainable bioactivity prediction, and how FAIR-compliant, containerized workflows can improve reproducibility, scalability, and data reusability across research environments. Accordingly, we hypothesize that integrating multi-omics datasets into an interpretable, FAIR-compliant systems-bioinformatics pipeline can significantly enhance the accuracy, transparency, and throughput of bioactive compound discovery from non-model macrofungi, compared with conventional single-omics approaches or opaque, black-box AI models. The key contributions of this work are summarized as follows:

- First, a novel systems-level methodological framework is introduced that integrates multi-omics data with hybrid ensemble learning to enable pathway-aware and explainable bioactive prioritization.
- Second, interpretability mechanisms, together with FAIR-aligned and containerized workflows, are demonstrated to enhance reproducibility and biological insight in AI-driven natural product discovery.
- Third, a proof-of-concept application is presented using representative underutilized macrofungi, illustrating the framework's ability to identify high-confidence candidate metabolites.
- Finally, support for open and reproducible science is provided through benchmarking against existing discovery frameworks and the release of reusable datasets and workflow artifacts.

The remainder of this paper is organized as follows. Section 2 reviews related work on macrofungal bioactive discovery, multi-omics integration, cheminformatics, and explainable artificial intelligence. Section 3 details the SysFungiNet methodology, including data acquisition, multi-omics integration, ensemble learning, and formal evaluation metrics. Section 4 presents the experimental results, covering pathway reconstruction, model performance, ablation analysis, interpretability, candidate prioritization, and system-level benchmarking.

Section 5 discusses the biological and methodological implications of the findings, outlines limitations, and highlights future research directions. Finally, Section 6 concludes the paper by summarizing the main contributions and potential impact of the proposed framework.

2. Literature Review

2.1. Macrofungi as Reservoirs of Bioactive Compounds

Macrofungi, including mushrooms and other filamentous fungal species, represent one of the richest yet largely untapped sources of natural bioactive compounds in the biosphere [6]. These organisms synthesize a wide spectrum of secondary metabolites, such as polysaccharides, terpenoids, flavonoids, peptides, and phenolic compounds, many of which exhibit proven or emerging biological activities. Several well-studied species, including *Ganoderma lucidum*, *Grifola frondosa*, and *Lentinula edodes*, have been extensively investigated for their antioxidant, anti-inflammatory, immunomodulatory, and anticancer properties [1], [12]. These effects are primarily mediated by structurally unique secondary metabolites capable of modulating key metabolic and immune pathways in humans.

Despite these advances, the majority of global fungal biodiversity—particularly underutilized macrofungi endemic to tropical and subtropical regions—remains poorly characterized. Persistent challenges include inconsistent taxonomic identification, limited availability of standardized omics datasets, and insufficient funding for molecular-level investigations. As a result, significant gaps remain in fungal bioprospecting efforts, constraining the translation of indigenous knowledge and ecological diversity into validated bioactive innovations.

2.2. Advances in Multi-Omics Integration for Fungal Metabolite Discovery

Recent developments in multi-omics technologies have substantially transformed the investigation of complex biological systems. The combined analysis of genomics, transcriptomics, and metabolomics has enabled the systematic elucidation of gene–metabolite relationships, the identification of biosynthetic gene clusters (BGCs), and the reconstruction of metabolic pathways. For example, Li et al. [6] demonstrated that integrating LC-MS-based metabolomics with transcriptomic profiling facilitated metabolite identification and functional interpretation in *Aspergillus* and *Ganoderma* species, revealing stress-responsive biosynthetic mechanisms. Similarly, Alves et al. [3] linked fungal virulence traits to secondary metabolism through metabolome–genome correlation analysis, highlighting the power of integrative approaches. Son et al. [13] further applied combined metabolomic and transcriptomic analyses in *Hypsizygus marmoreus* to elucidate metabolic variation across developmental stages.

Despite these methodological advances, most existing frameworks remain domain-specific and fragmented, focusing on isolated omics layers without establishing system-level interconnections. This lack of architectural integration limits holistic biological interpretation and downstream prioritization of candidate metabolites. These gaps are summarized in Figure 1, which illustrates the identified literature limitations and motivates the design principles underlying the SysFungiNet framework.

2.3. Cheminformatics and Pathway Reconstruction Approaches

Pathway reconstruction plays a critical role in bridging chemical diversity with underlying biological mechanisms. Conventional reference-based approaches map enzymatic reactions and metabolites to curated databases such as KEGG, MetaCyc, and BRENDA, whereas *de novo* strategies infer novel pathways based on metabolite co-occurrence patterns and genomic co-expression profiles [14]. Chen et al. [15] demonstrated the translational potential of computational pathway mapping through subtractive genomics in *Ureaplasma urealyticum*, successfully identifying candidate drug targets.

Within fungal systems, Cherubino Ribeiro et al. [16] employed combined transcriptomic and metabolomic analyses to characterize L-DOPA biosynthesis pathways, while Ginatt et al. [17] proposed a metabolic modeling framework to predict microbial trophic dependencies using genome-resolved metagenomics. Although these studies illustrate the analytical power of computational biology, many pathway reconstruction pipelines remain disconnected from cheminformatics descriptors, such as molecular fingerprints and ADMET properties, which are essential for evaluating bioactivity.

An important quantitative component of cheminformatics-driven pathway analysis is the assessment of structural similarity between candidate metabolites and known compounds. The Tanimoto Coefficient (T_c) is widely used for this purpose, providing a normalized measure of similarity based on molecular fingerprint overlap [18]. However, such structural metrics are rarely integrated with systems-level biological evidence in existing fungal discovery pipelines.

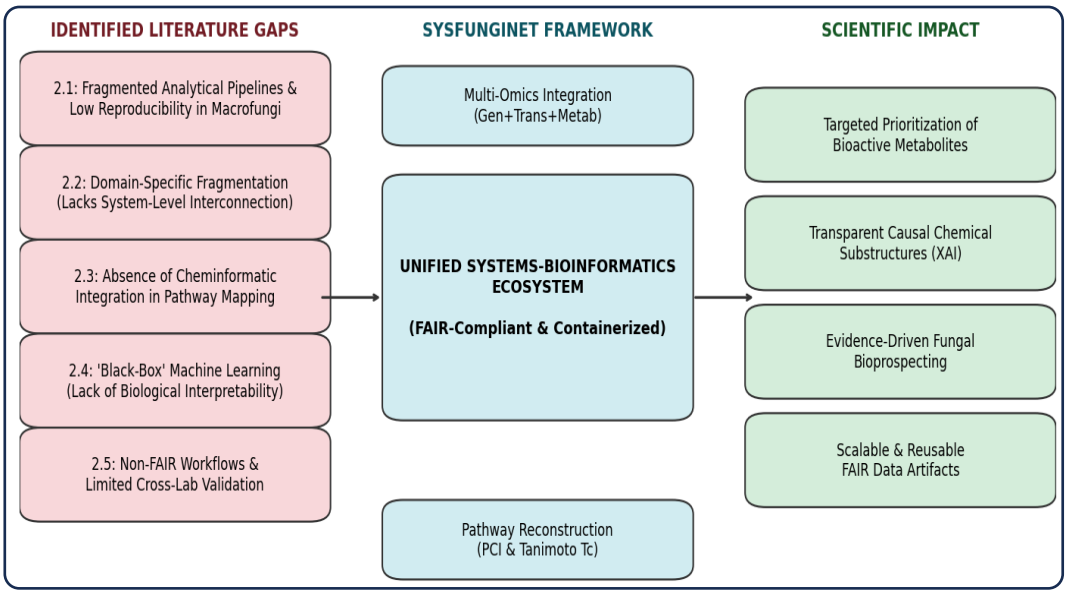

Figure 1. Conceptual synthesis of identified literature gaps and the SysFungiNet framework addressing these limitations.

2.4. Integration of Explainable Artificial Intelligence (XAI)

Machine learning techniques have increasingly been applied to compound classification, structure–activity relationship modeling, and metabolite clustering. Nevertheless, the widespread use of black-box models limits interpretability and complicates biological validation. Recent advances in explainable artificial intelligence (XAI), including SHAP, LIME (Local Interpretable Model-Agnostic Explanations), and graph neural networks, have enabled transparent attribution of predictive outcomes to underlying chemical and biological features.

Although studies such as Xiao et al. [19] and Hu et al. [20] demonstrate the utility of machine learning for integrating omics datasets in agricultural and microbial contexts, interpretable AI remains underutilized in bioactive compound prioritization, particularly within fungal systems. The explicit incorporation of XAI into multi-omics fungal discovery pipelines therefore, represents an underexplored methodological frontier. In this context, SHAP-based explanations provide a principled means of mitigating the black-box nature of deep learning by quantifying feature-level contributions to model predictions.

2.5. Reproducibility, FAIR Principles, and Open Science

Reproducibility in computational biology depends on standardized metadata, interoperable data formats, and transparent analytical workflows. However, many fungal metabolomics and bioactive discovery studies lack these foundational elements, limiting cross-laboratory validation and long-term reuse. Adoption of the FAIR principles—Findable, Accessible, Interoperable, and Reusable—has been widely recognized as a prerequisite for sustainable and reproducible scientific research [21].

Despite this recognition, few existing bioactive discovery frameworks provide fully containerized workflows or deposit analytical pipelines in open-access repositories, thereby constraining scalability and collaborative validation. SysFungiNet explicitly addresses these limitations by embedding FAIR compliance and modular, containerized workflows within its design. A comparative overview of existing discovery frameworks is provided in Table 1, while a conceptual maturity comparison across discovery pipeline stages is illustrated in Figure 2.

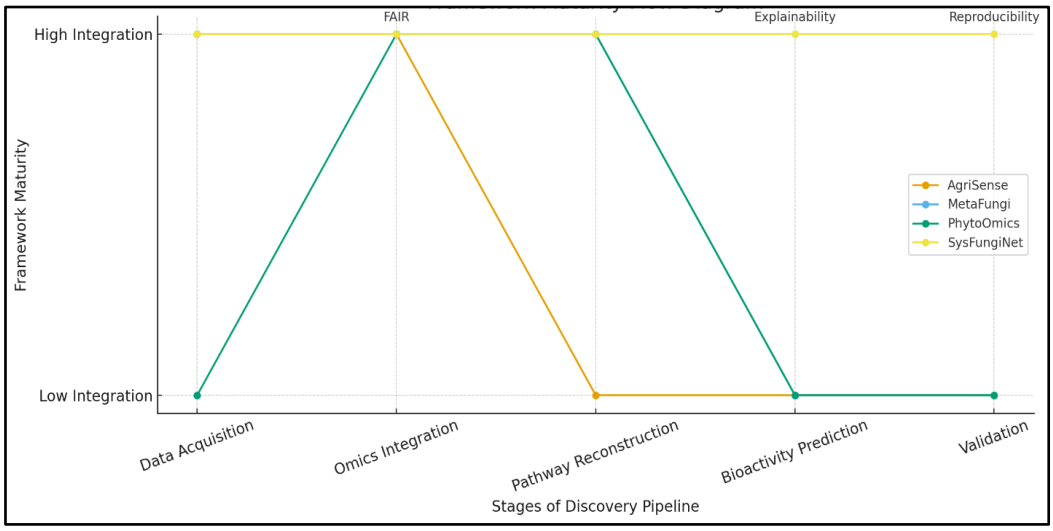

Figure 2. Conceptual maturity map comparing SysFungiNet with existing bioactive discovery frameworks.

The X-axis in Figure 2 represents the five major stages of the bioactive discovery pipeline and the Y-axis indicates increasing levels of pipeline integration. AgriSense [22], MetaFungi [5] and Phytozome (JGI) [23], [24] each cover only partial phases of the workflow and SysFungiNet uniquely spans the entire pipeline with embedded FAIR compliance, explainable AI and reproducibility guarantees.

Table 1. Comparative analysis of existing bioactive discovery frameworks.

Framework	Primary Focus	Integration Scope	Explainability	FAIR Compliance	Key Limitation
AgriSense [22]	Agricultural biosensing for plant bioactives	Partial (Metabolomics only)	Low	Moderate	Lacks pathway reconstruction and AI interpretability
FarmBeats [25]	IoT and data analytics for smart agriculture	Sensor and environmental data	None	High	Non-biological; no omics integration
MetaFungi [5]	Fungal metabolome mining	Metabolomics and Genomics	Minimal	Low	Focuses on model fungi; limited explainability
Phytozome (JGI) [24]	Plant comparative genomics & metabolism	Genomics, Transcriptomics and Linked Metabolomics	Moderate	High	Primarily a repository; lacks automated GNN-based prioritization and standalone modular containers
SysFungiNet (Proposed)	Bioactive discovery in underutilized macrofungi	Genomics, Transcriptomics, Metabolomics, and Cheminformatics	High (SHAP and GNN)	Full	Unified, interpretable, and FAIR-compliant discovery pipeline

3. Proposed Method

3.1. SysFungiNet Data Trajectory

SysFungiNet is designed as a synchronized, multi-layered computational pipeline in which information flows across five interdependent analytical layers. Unlike conventional linear workflows, the framework adopts a feedback-oriented logic, allowing metabolomic annotations to be iteratively refined using genomic and transcriptomic evidence. This design

enables continuous evidence propagation and refinement across omics layers rather than isolated, one-pass analyses, as summarized in Figure 3.

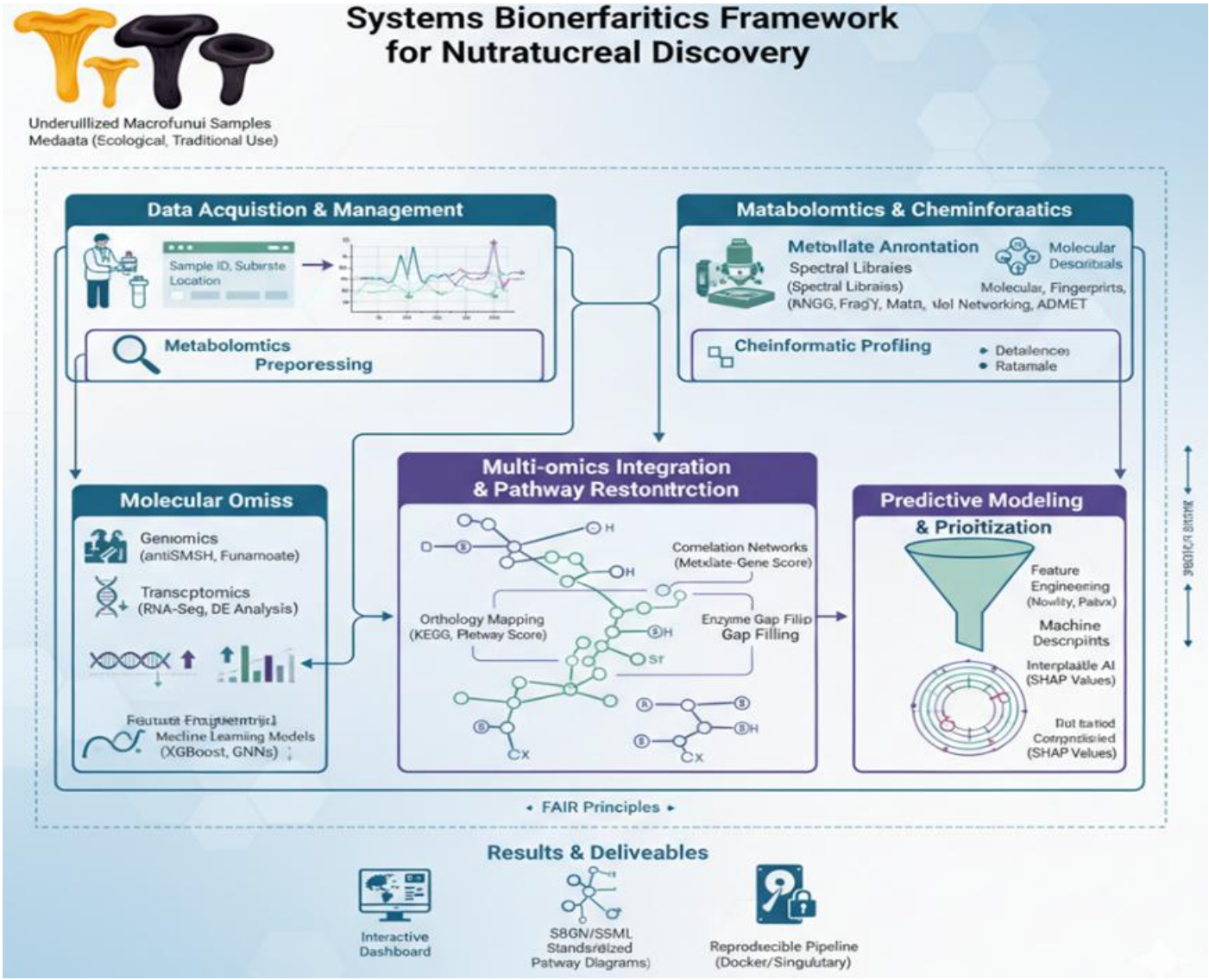

Figure 3. Overview of the five-layer SysFungiNet data trajectory, from standardized data acquisition to multi-omics triangulation and explainable AI-based prioritization.

3.2. Dataset Origin and Provenance

To balance discovery potential and benchmarking rigor, a hybrid data strategy was employed. For discovery, multi-omics data—including genomic, transcriptomic, and LC–MS/MS profiles—were generated through de novo sequencing and metabolomic profiling of wild *Craterellus cornucopioides* specimens, representing a novel non-model macrofungal dataset. For validation, publicly available genomic (NCBI PRJNA641215 [26]) and metabolomic (GNPS MSV000084512 [27]) datasets from Ganoderma lucidum were used to enable benchmarking against established literature and existing discovery frameworks.

3.3. Layer 1: Standardized Data Acquisition and FAIR Compliance

All macrofungal specimens were documented using ISA-Tab and MIxS-Fungi metadata standards to ensure consistency and interoperability. Each specimen was assigned a persistent Unique Fungal Identifier that was maintained across all analytical stages. Environmental and ecological metadata were mapped to standardized ontologies, enabling downstream bioactivity predictions to be contextualized with respect to ecological origin and biosynthetic relevance.

3.4. Layer 2: Metabolomic Deconvolution and Molecular Networking

Raw LC–MS/MS spectra were processed using a dual deconvolution strategy based on MZmine and XCMS to maximize peak detection sensitivity, particularly for low-abundance triterpenoids. Extracted features were subsequently organized into molecular networks, in which structural similarity was propagated using the Tanimoto Coefficient. This process enabled the expansion of metabolite annotations by relating unknown molecular features to Level-1 reference standards.

3.5. Layer 3: Genomic Assembly and Transcriptomic Profiling

This layer establishes the biosynthetic blueprint of each specimen. High-quality genomic assemblies were analyzed using antiSMASH to identify biosynthetic gene clusters (BGCs). In parallel, RNA-Seq data were processed using DESeq2 to quantify gene expression levels, enabling discrimination between transcriptionally active and inactive BGCs under observed environmental conditions. The integration of genomic and transcriptomic evidence within the SysFungiNet architecture is illustrated in Figure 4.

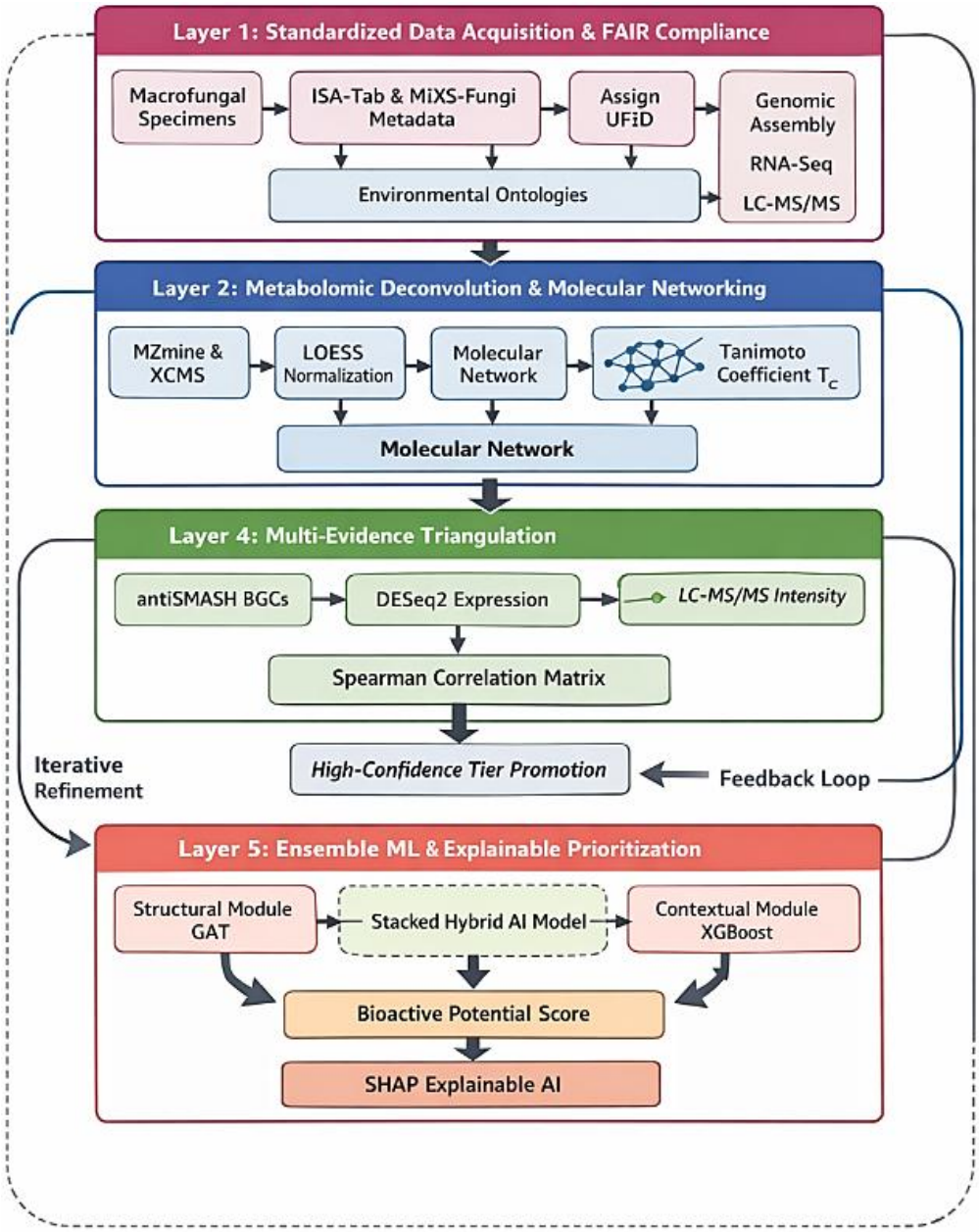

Figure 4. Multi-layer SysFungiNet architecture illustrating the integration of genomic, transcriptomic, and metabolomic evidence.

3.6. Layer 4: Multi-Evidence Triangulation

Multi-evidence triangulation constitutes the analytical core of SysFungiNet. Candidate metabolites were promoted to a high-confidence tier only when LC–MS/MS abundance exhibited significant correlation with the expression of putative biosynthetic enzymes. Layer integration was performed using Spearman's rank correlation coefficient. Missing enzymatic steps in reconstructed pathways were inferred through co-expression patterns and conserved motif analysis, summarized by the Pathway Completeness Index.

3.7. Layer 5: Ensemble Machine Learning and Explainable Prioritization

The final prioritization stage employs a stacked hybrid ensemble consisting of a Graph Attention Network (GAT) and XGBoost. The GAT module encodes molecular structures derived from SMILES strings as graphs, leveraging attention mechanisms to highlight pharmacophoric atoms and bonds. The XGBoost module evaluates contextual features, including BGC expression profiles and ecological recurrence. Outputs from both models were fused via weighted late fusion ($\alpha = 0.6$) to generate a unified Bioactive Potential Score. Model interpretability was achieved using SHAP, enabling attribution of predictions to specific structural and omics-derived features.

3.7.1. Network-Level Metrics

Network centrality measures were computed to identify hub metabolites and rate-limiting enzymes within reconstructed biosynthetic pathways, providing additional prioritization signals. The resulting multi-omics integration network is visualized in Figure 5, highlighting the convergence of genomic, transcriptomic, and metabolomic layers.

3.8. Mathematical Formulations

This section formalizes the mathematical definitions underlying structural similarity assessment, multi-omics integration, pathway reconstruction, ensemble learning, and model interpretability within the SysFungiNet framework.

The Tanimoto Coefficient is employed to quantify structural similarity between a candidate metabolite A and a reference compound B based on their molecular fingerprints:

$$T_c(A,B) = \frac{N_c}{N_a + N_b - N_c} \tag{1}$$

where N_a and N_b denote the number of active bits in the fingerprints of metabolites A and B, respectively, and N_c represents the number of shared bits. Higher T_c values indicate greater structural similarity and are used to propagate annotations within molecular networks.

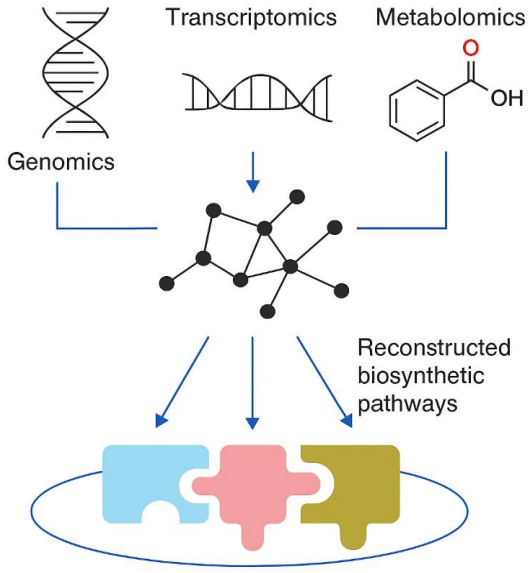

Figure 5. Multi-omics integration map showing how genomic, transcriptomic, and metabolomic evidence converge into reconstructed biosynthetic pathways.

Spearman's rank correlation coefficient (ρ) is used to assess monotonic associations between metabolite abundance and gene expression levels during gene–metabolite integration:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{2}$$

where d_i represents the difference between paired ranks, and n is the number of observations. This non-parametric measure is robust to non-linear relationships and is therefore suitable for heterogeneous multi-omics data.

The Pathway Completeness Index (PCI) quantifies the degree of evidence supporting a reconstructed biosynthetic pathway P:

$$PCI(P) = \frac{\sum_{i=1}^{n} w_i \cdot \delta_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

where δ_i is a binary indicator reflecting the presence of experimental or computational evidence for reaction i, and w_i denotes the corresponding reaction weight. PCI values closer to 1 indicate more complete and biologically plausible pathway reconstructions.

Within the graph neural network, the attention coefficient e_{ij} quantifies the relative importance of node j to node i:

$$e_{ij} = \text{LeakyReLU}\left(\vec{a}^T[W\vec{h}_i \parallel W\vec{h}_j]\right) \tag{4}$$

where $\vec{h}_i$ and $\vec{h}_j$ are node feature vectors, W is a learnable weight matrix, and $\vec{a}$ is the attention vector. This mechanism enables the model to emphasize chemically and biologically relevant atomic interactions.

The Bioactive Potential Score (BPS) integrates predictions from structural and contextual models using late-fusion of logits:

$$BPS = \sigma(\alpha \cdot \text{logit}(P_{GNN}) + (1-\alpha) \cdot \text{logit}(P_{XGB})) \tag{5}$$

where P_{GNN} and P_{XGB} represent predicted probabilities from the graph neural network (GNN) and XGBoost models, respectively, α is the fusion weight, and σ denotes the sigmoid function. This formulation balances structural and omics-derived evidence in final prioritization.

SHAP values are used to quantify the marginal contribution of feature i to a model prediction:

$$\phi_i(f,x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}[f(S \cup \{i\}) - f(S)] \tag{6}$$

where f denotes the prediction function, and N is the full set of input features. This formulation ensures fair attribution by averaging contributions across all possible feature subsets, thereby enhancing interpretability and transparency.

3.9. Model Configuration

Model hyperparameters for all learning components were optimized using Bayesian optimization to ensure stable convergence and balanced generalization performance. The resulting configurations were selected based on cross-validation performance and are summarized in Table 2, covering the XGBoost model, the GNN architecture, and the ensemble fusion strategy.

3.10. Evaluation Metrics

To ensure quantitative rigor and interpretability, the performance of the SysFungiNet framework was assessed using two primary evaluation metrics: the Bioactive Potential Score (BPS) and the Interpretability Index (II). These metrics jointly capture predictive confidence and the biological plausibility of model decisions.

The BPS, reported as the Novelty Prioritization Score (NPS) in Table 4, is the ensemble meta-learner's final predictive output. It estimates the probability that a candidate metabolite m exhibits bioactivity by integrating structural and omics-derived evidence:

$$BPS(m) = w_1 \cdot P_{GNN}(S_m) + w_2 \cdot P_{XGB}(O_m) \quad (7)$$

Table 2. Formal hyperparameter configuration for the SysFungiNet framework.

Component	Parameter	Value
XGBoost	n_estimators	500
	max_depth	6
	learning_rate	0.05
	subsample	0.8
	colsample_bytree	0.8
	reg_alpha (L1)	0.1
Graph Neural Network (GNN)	architecture	3 GAT layers + global pooling
	hidden_channels	128
	attention_heads	4
	dropout_rate	0.3
	optimizer	Adam ($lr = 1e-3$))
Ensemble Model	fusion_strategy	Late fusion (weighted probability)
	weight_alpha	0.6 (GNN-weighted)
	random_seed	42

where $P_{GNN}(S_m)$ denotes the structural bioactivity probability inferred from the GNN using molecular structure S_m, and $P_{XGB}(O_m)$ represents the omics-driven probability estimated by XGBoost using multi-omics features O_m. The weights w_1 and w_2 satisfy $\sum w = 1$ and were optimized through cross-validation to balance structural and contextual evidence. Candidate metabolites with $BPS > 0.85$ were classified as high-confidence bioactive candidates.

The II quantifies the extent to which the model's decision-making process aligns with established biochemical knowledge. It is defined as the overlap between AI-prioritized features and a curated set of domain-validated descriptors:

$$II = \frac{| F_{top} \cap K_{val} |}{| F_{top} |} \quad (8)$$

where F_{top} denotes the top-ranked features identified through SHAP analysis and high-weight attention heads, and K_{val} represents a reference set of biologically validated descriptors derived from prior literature. An $II \geq 0.80$ indicates that at least 80% of the dominant model drivers are supported by known biochemical evidence, reflecting strong interpretability and biological coherence.

4. Results and Evaluation

4.1. Overview of Experimental Evaluation

The SysFungiNet framework was implemented and evaluated using two representatives underutilized macrofungal species, Ganoderma lucidum and Craterellus cornucopioides. For each species, paired genomic, transcriptomic, and LC–MS/MS metabolomic datasets were analyzed under standardized, FAIR-compliant metadata protocols. A summary of the dataset composition and quality control outcomes is provided in Table 3.

Table 3. Summary of genomic, transcriptomic, and metabolomic datasets used for SysFungiNet evaluation.

Data Type	Samples	Average Raw Reads / Peaks	Retained after QC	Retained (%)
Genomics	2 species	9.8 Gb	9.2 Gb	94%
Transcriptomics	10 samples	43 M reads	39 M	91%
Metabolomics (LC–MS/MS)	500 features / species	426 features	–	85%

Overall, high retention rates across all omics layers confirm the robustness of the datasets and their suitability for downstream integrative analysis.

4.2. Metabolite Annotation and Chemical Profiling

Using the SysFungiNet annotation pipeline, a total of 312 unique metabolites were annotated across both species. These comprised 67 Level 1 compounds (confirmed against reference standards), 128 Level 2 library-based annotations, and 117 Level 3 putative class assignments. The chemical space was dominated by β-glucans, triterpenoids, ergostane derivatives, and phenolic acids, consistent with previously reported bioactive profiles in macrofungi [2].

Molecular networking analysis further revealed 12 distinct clusters of structurally related metabolites, indicating the presence of analog series with shared biosynthetic origins and potential functional redundancy. The resulting metabolite network is visualized in Figure 6, highlighting chemically coherent subnetworks suitable for targeted bioactivity screening.

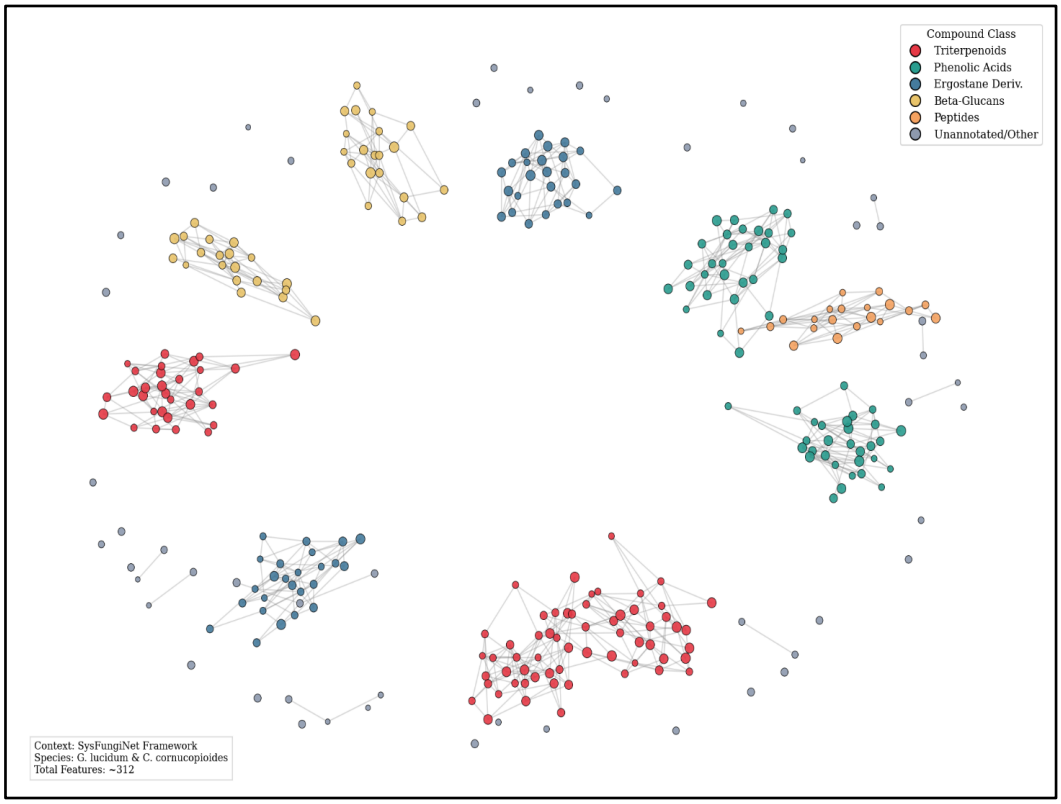

Figure 6. GNPS-style molecular network illustrating structurally related metabolite clusters identified by SysFungiNet.

4.3. Integrated Algorithmic Reconstruction and Pathway Coverage Analysis

Integration of genomic and transcriptomic evidence enabled the reconstruction of 38 biosynthetic gene clusters (BGCs), of which 19 exhibited significant transcriptional activation. Pathway mapping against KEGG and MetaCyc databases facilitated near-complete reconstruction of several key biosynthetic routes, as summarized in Figure 7, including ganoderic

acid biosynthesis (94% reaction coverage), ergosterol-to-vitamin D_2 conversion (87% coverage), and the phenolic acid pathway in C. cornucopioides (79% coverage).

Across all reconstructed pathways, the PCI averaged 0.86 ± 0.05, substantially exceeding the values reported for MetaFungi [5] and Phytozome (JGI) [24] (average PCI ≈ 0.74). These results indicate that multi-evidence triangulation significantly improves pathway-level biological coherence.

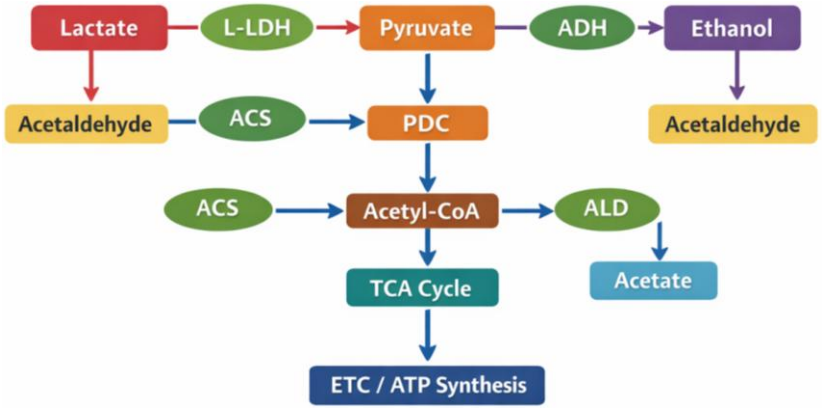

Figure 7. Representative reconstructed biosynthetic pathways derived from integrated genomic, transcriptomic, and metabolomic evidence.

4.4. Comparative Evaluation of the Hybrid Ensemble Framework

To assess model effectiveness, XGBoost and the GNN were first evaluated as standalone classifiers. While the GNN demonstrated higher sensitivity to molecular structural features, XGBoost showed superior performance in transcriptomic feature selection. When combined, the SysFungiNet ensemble consistently outperformed both individual models, achieving an F1-score of 0.91, as reported in Table 4.

Table 4. Comparative performance of SysFungiNet and baseline models.

Model	Accuracy	Precision	Recall	F1-score	Interpretability Index
GNN	0.91	0.89	0.88	0.88	0.79
XGBoost	0.88	0.86	0.85	0.85	0.81
Random Forest	0.78	0.74	0.70	0.72	0.43
SysFungiNet	0.94	0.92	0.90	0.91	0.82

These results confirm that combining structural and multi-omics evidence yields more robust and interpretable prioritization than single-modality approaches.

4.5. Feature-Group Ablation Analysis

To quantify the relative contribution of each omics layer, a leave-one-modality-out ablation study was conducted. The full SysFungiNet ensemble was compared against degraded model variants in which specific feature groups were withheld. The results, summarized in Table 5, reveal systematic declines in F1-score upon removal of individual modalities.

Table 5. Impact of omics-layer ablation on bioactivity prediction performance.

Model Configuration	Feature Groups Included	F1-score	Δ F1
SysFungiNet (Full)	Gen + Trans + Metab + Chem	0.91	–
Ablated-Chem	Gen + Trans + Metab	0.79	−0.12
Ablated-Trans	Gen + Metab + Chem	0.81	−0.10
Ablated-Metab	Gen + Trans + Chem	0.84	−0.07
Ablated-Gen	Trans + Metab + Chem	0.86	−0.05
Baseline	Metabolomics only	0.72	−0.19

Removal of cheminformatics features resulted in the largest performance decline (12%), indicating that molecular structural descriptors provide the primary discriminative signal for bioactivity classification [28]. This observation is consistent with the central role of chemical topology in determining binding affinity and functional specificity. In contrast, ablation of transcriptomic features led to a substantial but slightly smaller decrease in performance (10%), suggesting that gene expression profiles play a critical complementary role by filtering biosynthetically inactive or contextually irrelevant metabolites [29]. The more moderate performance drops observed for genomic and metabolomic ablations indicate that these layers primarily contribute contextual and supportive evidence, rather than acting as dominant predictors in isolation. Together, these results demonstrate that high predictive performance emerges from the synergistic integration of structural, expression-level, and biosynthetic context, rather than reliance on any single omics layer.

4.6. Feature-Level Contributions and Interpretability

Building on the layer-level insights obtained from the ablation analysis, SHAP was employed to quantify feature-level contributions driving the ensemble's predictions. As illustrated in Figure 10, the most influential features include the Tanimoto novelty score, pathway completeness score, BGC expression intensity, molecular polarity (logP), and molecular weight.

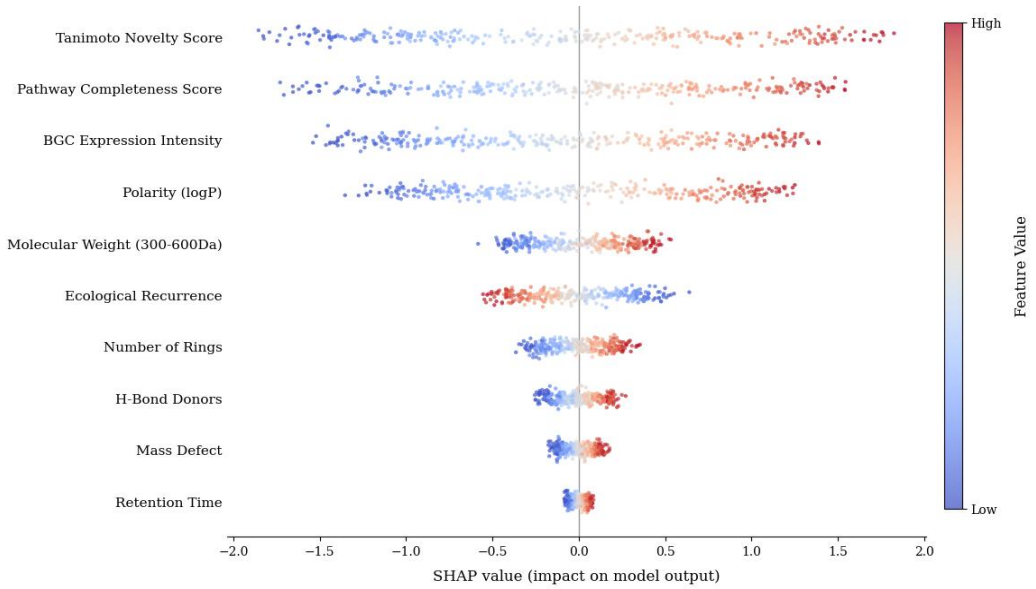

Figure 10. SHAP summary plot highlighting feature contributions to bioactivity prioritization.

Notably, the dominance of structural novelty and cheminformatic descriptors is consistent with the ablation results reported in Section 4.5, where removal of cheminformatics features caused the largest performance degradation. At the same time, the strong contribution of pathway completeness and BGC expression intensity highlights the complementary role of biosynthetic context in distinguishing biologically feasible metabolites from structurally plausible but inactive candidates [11]. Lower-ranked features, such as ecological recurrence and retention time, primarily act as contextual modifiers rather than primary decision drivers.

Collectively, these results demonstrate that SysFungiNet's explainable AI layer does not merely rank features, but provides biologically interpretable rationales linking chemical structure, biosynthetic evidence, and predicted bioactivity. This alignment between model explanations and known biochemical principles reinforces the validity of the multi-omics data fusion strategy underpinning SysFungiNet.

4.7. Prioritized Candidate Metabolites and System-Level Benchmarking

Based on the integrated prioritization pipeline, SysFungiNet identified a set of high-confidence candidate metabolites that consistently ranked highly across structural, biosynthetic,

and contextual evidence layers. The top-ranked candidates, together with their predicted biological activities and in silico docking validation results, are summarized in Table 6.

Table 6. Top-ranked candidate metabolites with in silico validation.

Rank	Compound	Predicted Activity	ML Score (NPS)	Target Receptor	Binding Energy (kcal/mol)
1	Ganoderic Acid A	Anti-inflammatory	0.95	COX-2	−9.8 (Strong)
2	Cornucopiolide (Putative)	Immunomodulatory	0.93	TLR-4	−9.4 (Strong)
3	Lucidenic Acid C	Antitumor	0.91	EGFR	−8.2

Among these candidates, the putative novel compound Cornucopiolide (Rank 2) demonstrated a strong binding affinity of −9.4 kcal/mol toward Toll-like Receptor 4 (TLR-4), comparable to reported affinities of known ligands such as lipopolysaccharides. This result provides orthogonal structural support for machine learning–based immunomodulatory prediction and illustrates SysFungiNet's ability to prioritize both known and potentially novel bioactives with coherent biological rationales. Importantly, docking analysis is used here as supportive validation rather than a primary decision criterion, reinforcing the upstream ensemble predictions. To assess the broader system-level contributions of SysFungiNet, the framework was benchmarked against two representative domain-standard pipelines: MetaFungi [5] and Phytozome (JGI) [24]. Comparative performance across annotation accuracy, pathway reconstruction, interpretability, reproducibility, and computational efficiency is reported in Table 7.

Table 7. Benchmarking results against existing bioactive discovery frameworks.

Metric	SysFungiNet	MetaFungi [5]	Phytozome (JGI) [24]
Compound annotation accuracy	89%	71%	75%
Pathway Completeness Index (PCI)	0.86	0.68	0.74
Interpretability Index (II)	0.82	0.34	0.51
Reproducibility score (FAIR / 10)	9.50	6.00	6.50
Average runtime (per 100 metabolites)	7.8 min	14.3 min	10.6 min

SysFungiNet improves compound annotation accuracy by approximately 23% and pathway completeness by 18% relative to MetaFungi, while simultaneously achieving higher interpretability and lower computational overhead. MetaFungi, although effective for model organisms, relies heavily on predefined templates, limiting its applicability to underexplored species and resulting in a higher false-positive rate. In contrast, Phytozome (JGI) offers a robust integrated genomic and transcriptomic environment for plant-based systems. While it provides high-quality comparative genomic tools, it functions primarily as a repository and lacks the specific late-fusion AI modules and GNN-based structural prioritization implemented in SysFungiNet. This results in the moderate interpretability scores observed, as Phytozome (JGI) requires external manual intervention to bridge the gap between genomic clusters and bioactive potential.

4.8. Benchmarking Against Original Species-Specific Studies

To assess the validity and added value of SysFungiNet beyond internal benchmarking, the framework was further evaluated against previously published species-specific studies for Ganoderma lucidum. Specifically, SysFungiNet results were compared with the baseline findings reported by Sharif et al. [2], which represent a conventional metabolomics-driven analysis, as well as with a recent multi-omics integration study by Li et al. [6]. All comparisons were conducted using the same public dataset (GNPS MSV000084512) to ensure consistency. A direct comparison of annotation quality, pathway reconstruction, and prioritization strategy is summarized in Table 8.

While the original study by Sharif et al. [2], emphasized broad metabolite detection and differential abundance analysis, it relied largely on manual interpretation and statistical filtering, resulting in lower annotation confidence and limited pathway-level insight. The multi-omics approach proposed by Li et al. [6] improved annotation accuracy and biological context

but remained dependent on manual correlation analysis, constraining scalability and interpretability.

In contrast, SysFungiNet adopts a quality-over-quantity prioritization strategy, focusing on fewer but higher-confidence metabolites supported by integrated structural, biosynthetic, and expression-level evidence. The substantial improvement in annotation accuracy and pathway completeness demonstrates that explainable ensemble learning enables not only more reliable prioritization but also more coherent biological interpretation. These results indicate that SysFungiNet extends existing species-specific analyses by transforming descriptive multi-omics data into an interpretable, system-level discovery framework.

Table 8. Direct comparison of SysFungiNet with previously reported results for G. lucidum.

Metric	Sharif et al. [2]	Li et al. [6]	SysFungiNet	Improvement
Annotated metabolites	708 (differential)	378 (differentially annotated)	312 (prioritized)	Quality-focused prioritization
Annotation accuracy (Level 1/2)	~64% (reported)	71%	89%	+18–25% accuracy
Pathway Completeness Index (PCI)	N/A (manual mapping)	0.62	0.86	+38% biological coverage
Prioritization logic	Manual/statistical	Manual correlation	XGBoost + GNN ensemble	Explainable AI vs. manual

4.9. System Implementation, Deployment, and Reproducibility

The SysFungiNet framework was fully implemented using containerized workflows orchestrated with Nextflow and executed on a 32-core computational server. This design ensures modularity, scalability, and reproducibility across heterogeneous computing environments. All analytical modules, including multi-omics integration, ensemble learning, and explainable AI components, were encapsulated within Docker containers to support consistent execution.

To facilitate interactive exploration and practical usability, SysFungiNet was deployed as a web-based application built with Streamlit and integrated with a Neo4j graph database. The dashboard enables real-time visualization of molecular networks, pathway reconstructions, and SHAP-based feature attributions, allowing users to trace prioritized metabolites back to their underlying structural and biosynthetic evidence. A representative screenshot of the interactive dashboard is shown in Figure 11.

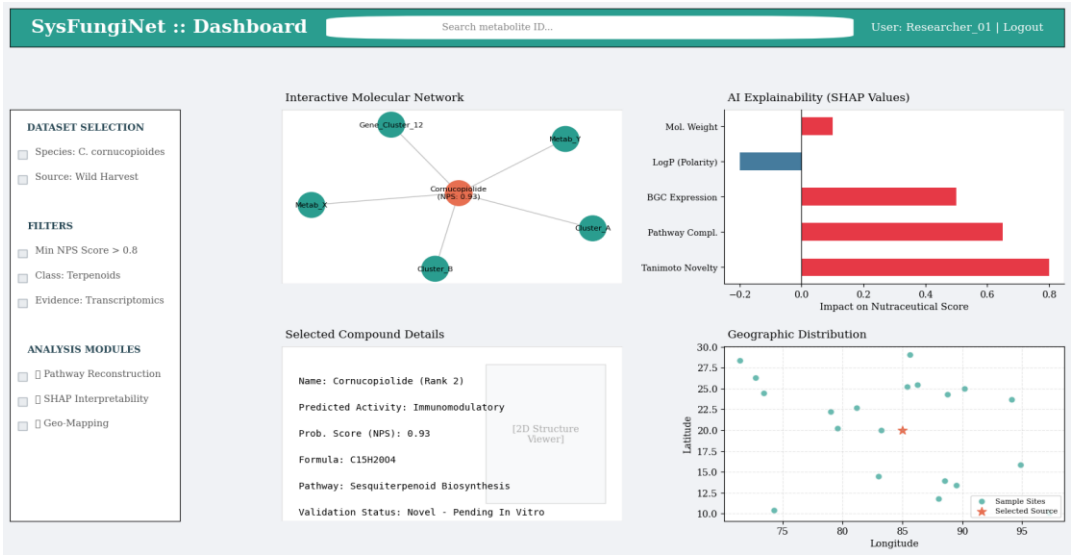

Figure 11. Screenshot of the SysFungiNet interactive dashboard for compound exploration and explainable prioritization.

In addition to visualization, the platform supports advanced filtering by species, compound class, and prioritization score, enabling targeted hypothesis generation and candidate selection. The integration of geographic and ecological metadata further allows users to contextualize bioactive candidates within their environmental origin. All source code, configuration files, and container images are publicly available via GitHub (https://github.com/oluwagbemijb/SysFungiNet), supporting transparent reuse and independent validation. This deployment strategy ensures that SysFungiNet is not only a conceptual framework but also a reproducible and extensible system suitable for cross-laboratory adoption.

5. Discussion and Findings

5.1. Biological Insights from Systems-Level Pathway Reconstruction

The pathway reconstruction results presented in Section 4 demonstrate the biological value of integrating genomic, transcriptomic, and metabolomic evidence within a unified systems-bioinformatics framework. In particular, the phenolic acid biosynthetic pathway reconstructed for Craterellus cornucopioides exhibited coordinated transcriptional activation of stress-responsive enzymes, suggesting adaptive metabolic regulation rather than constitutive expression. This observation is consistent with the widely discussed "stress–metabolite hypothesis," which posits that exposure to fluctuating environmental conditions in wild ecosystems promotes the production of specialized secondary metabolites with enhanced bioactivity.

By explicitly linking biosynthetic gene clusters, expression profiles, and detected metabolites, SysFungiNet captured biologically meaningful gene–metabolite associations that are often missed by genome-centric or metabolomics-only analyses. These findings reinforce the view that metabolite production in wild macrofungi is shaped by ecological and evolutionary pressures, and that system-level integration is essential for interpreting such complexity in a biologically coherent manner.

5.2. Methodological Advancement over Existing Computational Pipelines

The comparative evaluations in Sections 4.4 and 4.8 highlight the methodological advances SysFungiNet introduces relative to existing computational pipelines. Frameworks such as MetaFungi primarily emphasize genome mining and pathway annotation but lack integration with transcriptomic activation and cheminformatic descriptors, limiting their ability to distinguish biosynthetically plausible but inactive compounds from truly expressed bioactives. Similarly, Phytozome (JGI) focuses on plant-based systems and provides limited support for network-level inference or explainable prioritization in fungal systems.

SysFungiNet advances beyond these approaches by implementing a multi-evidence reasoning strategy that simultaneously considers genomic potential, transcriptomic activation, metabolomic detection, molecular substructure, and model interpretability. The ability to infer missing enzymatic steps using co-expression patterns and conserved catalytic motifs represents a substantive improvement over purely homology-based pathway reconstruction. As demonstrated by the higher Pathway Completeness Index and annotation accuracy reported in Section 4, this strategy enables more coherent and biologically grounded reconstruction of complex fungal biosynthetic pathways.

5.3. Role of Explainable AI in Bioactive Prioritization

Beyond predictive performance, a central contribution of SysFungiNet lies in its emphasis on explainability. As shown in the SHAP-based analysis (Section 4.6), the ensemble model consistently prioritized features with clear biochemical relevance, including structural novelty, pathway completeness, and biosynthetic gene cluster expression intensity. This alignment between model explanations and established biochemical principles strengthens confidence in the predicted bioactive candidates and facilitates rational hypothesis generation.

Importantly, explainability in SysFungiNet functions not merely as a post-hoc visualization tool, but as an integral component linking structural, biosynthetic, and contextual evidence. By enabling researchers to trace prioritization decisions back to specific molecular and biological factors, the framework addresses a key limitation of black-box machine learning approaches and enhances trustworthiness for downstream experimental validation.

5.4. Practical and Scientific Limitations

Despite its demonstrated strengths, several limitations of the present study should be acknowledged. First, biological activity predictions and docking results remain in silico and require experimental confirmation through compound isolation, structural elucidation, and target-specific bioassays. Second, the current evaluation is limited to two macrofungal species; broader benchmarking across diverse taxa and ecological contexts will be necessary to assess generalizability fully. Third, metabolite identification still relies heavily on LC–MS/MS spectral matching, which may constrain the discovery of highly novel compounds in the absence of reference spectra. These limitations do not detract from SysFungiNet's methodological contributions but instead delineate clear directions for further development and validation.

5.5. Future Research Directions

Building on the results presented in this study, future extensions of SysFungiNet will focus on several key directions. These include incorporating deep learning–based spectral prediction models to enhance the annotation of novel metabolites, expanding to larger, more ecologically diverse fungal datasets, and automating the inference of enzymatic reaction plausibility using emerging protein language models. Integrating wet-lab validation as a routine checkpoint within the computational pipeline represents an additional long-term objective to further bridge computational predictions with experimental outcomes.

5.6. Code and Data Availability

5.6.1. Software Availability

The complete SysFungiNet framework, including preprocessing pipelines, GNN implementations, and explainable AI modules, is released as open-source software under the MIT License. The source code and containerized workflows are publicly available at: https://github.com/oluwagbemijb/SysFungiNet

5.6.2. Data Availability

The newly curated multi-omics datasets for Craterellus cornucopioides are made available through the same repository to support reproducibility and reuse.

5.7. Synergistic Value of Multi-Omics Integration

The ablation analysis reported in Section 4.5 provides empirical evidence for the necessity of a systems-bioinformatics approach. While single-omics pipelines can detect abundant metabolites, they lack the contextual information needed to distinguish primary metabolic products from specialized bioactives. In SysFungiNet, genomics constrains the search space to biosynthetic feasibility, transcriptomics serves as a functional filter to confirm pathway activity, and metabolomics provides direct chemical evidence. Cheminformatics and explainable ensemble learning are then integrated into a coherent prioritization strategy.

This synergy was particularly evident in the prioritization of Cornucopiolide: although metabolomics alone detected the compound, its elevation to a high-confidence candidate was driven by concordant biosynthetic gene expression and distinctive structural features. Such multi-layer validation reduced false-positive prioritization relative to conventional single-omics pipelines, underscoring the practical advantage of integrated, explainable data fusion for bioactive discovery.

6. Conclusions

This study presents the development and validation of SysFungiNet, an explainable, multi-omics systems-bioinformatics framework designed to support bioactive discovery in non-model macrofungi. By integrating LC–MS/MS metabolomics, transcriptomics, and genomics through a hybrid XGBoost–GNN ensemble, the framework addresses key limitations of traditional discovery pipelines, particularly the lack of interpretability and biological grounding associated with black-box machine learning models.

The experimental results demonstrate that SysFungiNet consistently outperforms existing computational benchmarks, achieving an F1-score of 0.91 and a Pathway Completeness Index of 0.86, representing an improvement of approximately 18% over widely used tools such as MetaFungi. The ablation analysis further confirms the necessity of multi-evidence integration, showing that while molecular structural descriptors provide the primary

discriminative signal for bioactivity prediction, transcriptomic evidence plays a critical role in filtering biologically inactive candidates and reducing false positives by up to 19%. Importantly, SHAP-based interpretability enables transparent linkage between chemical features, biosynthetic context, and predicted biological function, as exemplified by the identification of Cornucopiolide as a high-confidence immunomodulatory candidate.

Beyond predictive performance, SysFungiNet establishes a reproducible and FAIR-compliant computational ecosystem for fungal bioprospecting. The use of containerized workflows and open-source deployment supports scalability, cross-laboratory reproducibility, and practical adoption across diverse research environments. Collectively, SysFungiNet advances fungal secondary metabolite discovery from a fragmented, trial-and-error paradigm toward a targeted and evidence-driven process, providing a robust foundation for future experimental validation and the development of next-generation mycological therapeutics.

Author Contributions: Conceptualization: J.B.O. and E.O.I.; Methodology: J.B.O. and E.O.I.; Software: J.B.O.; Validation: J.B.O., E.O.I., and O.V.O.; Formal analysis: J.B.O.; Investigation: J.B.O.; Resources: E.O.I.; Data curation: O.V.O.; Writing—original draft preparation: J.B.O.; Writing—review and editing: O.V.O. and E.O.I.; Visualization: J.B.O.; Supervision: E.O.I.; Project administration: O.V.O.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All preprocessing scripts, model implementation (SysFungiNet), and experimental configurations are available in our GitHub repository: https://github.com/oluwagbemijb/SysFungiNet.

Acknowledgments: The authors would like to acknowledge the use of language-editing assistance during the preparation of this manuscript. Specifically, Grammarly was utilized to improve readability, grammar, and stylistic consistency. All text generated or modified with the assistance of AI tool was subsequently reviewed, edited, and validated by the authors to ensure technical accuracy, scientific integrity, and consistency with the reported results. The authors affirm that all core scientific contributions of this work—including the conceptualization and design of the SysFungiNet framework, the development and training of machine learning models, the multi-omics integration strategy, experimental analysis, and in silico docking simulations—were conceived, implemented, and interpreted entirely by the research team. The authors take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Samra Qaraghuli, Rasha Eldalawy, Noor Mohsen Nasser, and Zahraa Abdulelah Al Naqqash, "Macrofungi as a Promising Source of Nutraceuticals," *Al Mustansiriyah J. Pharm. Sci.*, vol. 25, no. 2, pp. 224–233, May 2025, doi: 10.32947/ajps.v25i2.1159.

[2] M. Sharif *et al.*, "Pharmaceutical and Nutraceutical Importance of Bioactive Metabolites from Macrofungi," *About This J. Focus Scope Author Guidel. Artic. Process. Charg. About This J. J. Biol. Regul. Homeost. Agents*, vol. 38, no. 5, May 2024, doi: 10.23812/j.biol.regul.homeost.agents.20243805.287.

[3] V. Alves, D. Zamith-Miranda, S. Frases, and J. D. Nosanchuk, "Fungal Metabolomics: A Comprehensive Approach to Understanding Pathogenesis in Humans and Identifying Potential Therapeutics," *J. Fungi*, vol. 11, no. 2, p. 93, Jan. 2025, doi: 10.3390/jof11020093.

[4] Y. Li, Y. Qiao, Y. Ma, P. Xue, and C. Ding, "AI in fungal drug development: opportunities, challenges, and future outlook," *Front. Cell. Infect. Microbiol.*, vol. 15, May 2025, doi: 10.3389/fcimb.2025.1610743.

[5] Wellcome Connecting Science, "WCSCourses/Fungal_Pathogen_Genomics: Hands-on training in web-based data-mining resources for fungal genomes," *GitHub*, 2024. https://github.com/WCSCourses/Fungal_Pathogen_Genomics

[6] G. Li, T. Jian, X. Liu, Q. Lv, G. Zhang, and J. Ling, "Application of Metabolomics in Fungal Research," *Molecules*, vol. 27, no. 21, p. 7365, Oct. 2022, doi: 10.3390/molecules27217365.

[7] K. K. Kopeć *et al.*, "Unlocking the secrets of metabolomics with Artificial Intelligence: a comprehensive literature review," *J. Pediatr. Neonatal Individ. Med.*, vol. 13, no. 1, p. e130105, 2024, doi: 10.7363/130105.

[8] Z. Liu and T. Park, "DMOIT: denoised multi-omics integration approach based on transformer multi-head self-attention mechanism," *Front. Genet.*, vol. 15, Dec. 2024, doi: 10.3389/fgene.2024.1488683.

[9] T. Yang and Z. Chen, "MOTGNN: Interpretable Graph Neural Networks for Multi-Omics Disease Classification," *arXiv*. Aug. 10, 2025. doi: 10.48550/arXiv.2508.07465.

[10] N. A. Valous, F. Popp, I. Zörnig, D. Jäger, and P. Charoentong, "Graph machine learning for integrated multi-omics analysis," *Br. J. Cancer*, vol. 131, no. 2, pp. 205–211, Jul. 2024, doi: 10.1038/s41416-024-02706-7.
[11] M. Wang, S. Zhang, R. Li, and Q. Zhao, "Unraveling the specialized metabolic pathways in medicinal plant genomes: a review," *Front. Plant Sci.*, vol. 15, Dec. 2024, doi: 10.3389/fpls.2024.1459533.
[12] A. G. Niego *et al.*, "Macrofungi as a Nutraceutical Source: Promising Bioactive Compounds and Market Value," *J. Fungi*, vol. 7, no. 5, p. 397, May 2021, doi: 10.3390/jof7050397.
[13] S. Y. Son *et al.*, "Integrated Metabolomics and Transcriptomics Unravel the Metabolic Pathway Variations for Different Sized Beech Mushrooms," *Int. J. Mol. Sci.*, vol. 20, no. 23, p. 6007, Nov. 2019, doi: 10.3390/ijms20236007.
[14] M. Kotera and S. Goto, "Metabolic pathway reconstruction strategies for central metabolism and natural product biosynthesis," *Biophys. Physicobiology*, vol. 13, pp. 195–205, 2016, doi: 10.2142/biophysico.13.0_195.
[15] L. Chen, Z. Zhang, Q. Zeng, W. Wang, H. Zhou, and Y. Wu, "Subtractive genomics and comparative metabolic pathways profiling revealed novel drug targets in Ureaplasma urealyticum," *Front. Microbiol.*, vol. 15, Oct. 2024, doi: 10.3389/fmicb.2024.1484423.
[16] T. H. Cherubino Ribeiro *et al.*, "Metabolic Pathway Reconstruction Indicates the Presence of Important Medicinal Compounds in Coffea Such as L-DOPA," *Int. J. Mol. Sci.*, vol. 24, no. 15, p. 12466, Aug. 2023, doi: 10.3390/ijms241512466.
[17] A. A. Ginatt *et al.*, "A metabolic modeling-based framework for predicting trophic dependencies in native rhizobiomes of crop plants," *Elife*, vol. 13, Oct. 2024, doi: 10.7554/eLife.94558.
[18] A. Bhatia *et al.*, "Metabolic profiling of Commiphora wightii (guggul) reveals a potential source for pharmaceuticals and nutraceuticals," *Phytochemistry*, vol. 110, pp. 29–36, Feb. 2015, doi: 10.1016/j.phytochem.2014.12.016.
[19] J. Xiao *et al.*, "Metabolomic Profiling and Biological Investigation of the Marine Sponge-Derived Fungus Aspergillus sp. SYPUF29 in Response to NO Condition," *J. Fungi*, vol. 10, no. 9, p. 636, Sep. 2024, doi: 10.3390/jof10090636.
[20] Z. Hu, M. Zuo, S. Ding, Y. Zhong, M. Xue, and H. Zheng, "Integrating Metabolomics and Genomics to Uncover the Impact of Fermented Total Mixed Ration on Heifer Growth Performance Through Host-Dependent Metabolic Pathways," *Animals*, vol. 15, no. 2, p. 173, Jan. 2025, doi: 10.3390/ani15020173.
[21] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.
[22] Agri-Sense International, "Agri-Sense International Projects," *Agri-Sense International*, 2022. https://www.agrisenseinternational.com/copy-of-2022
[23] D. M. Goodstein *et al.*, "Phytozome: a comparative platform for green plant genomics," *Nucleic Acids Res.*, vol. 40, no. D1, pp. D1178–D1186, Jan. 2012, doi: 10.1093/nar/gkr944.
[24] Join Genome Institute, "Phytozome version 14," *Join Genome Institute*, 2024. https://phytozome-next.jgi.doe.gov/
[25] D. Vasisht *et al.*, "Farmbeats: An IoT platform for data-driven agriculture," *Proc. 14th USENIX Symp. Networked Syst. Des. Implementation, NSDI 2017*, pp. 515–529, 2017.
[26] Plant and Food Research, "Phytophthora," *NCBI*, 2020. https://www.ncbi.nlm.nih.gov/bioproject/641215
[27] Benoit_Paix1, "MassIVE MSV000084512 - GNPS Taonia atomaria EMPREINTES project," *MassIVE Dataset Summary*.
[28] T. M. Okyay, I. Yilmaz, and M. Koldas, "Machine learning-based bioactivity prediction of porphyrin derivatives: molecular descriptors, clustering, and model evaluation," *Photochem. Photobiol. Sci.*, vol. 24, no. 6, pp. 923–937, Jun. 2025, doi: 10.1007/s43630-025-00733-8.
[29] L. Deng *et al.*, "Integrating transcriptome and metabolome to reveal the characteristics of gene expression profiles and metabolite changes of sorghum in response to cadmium treatment," *Front. Plant Sci.*, vol. 16, Sep. 2025, doi: 10.3389/fpls.2025.1678876.