AI Medicine
https://www.sciltp.com/journals/aim

Article

Hierarchical Enzyme Function Prediction Based on Structural Confidence and Active-Site-Aware Attention

Qiuhao Wang [1], Yuexiao Wang [1], Aohan Mei [2], Tian Zhang [3] and Nan Li [4,*]

[1] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
[2] College of Software, Northeastern University, Shenyang 110819, China
[3] Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518052, China
[4] School of Artificial Intelligence, Shanxi University, Taiyuan 030006, China
* Correspondence: linan10@sxu.edu.cn

How To Cite: Wang, Q.; Wang, Y.; Mei, A.; et al. Hierarchical Enzyme Function Prediction Based on Structural Confidence and Active-Site-Aware Attention. *AI Medicine* **2026**, *3*(2), 6. https://doi.org/10.53941/aim.2026.100006

Received: 16 June 2026
Revised: 19 July 2026
Accepted: 24 July 2026
Published: 11 August 2026

Abstract: Accurate enzyme function prediction is important for protein functional annotation, drug discovery, metabolic-pathway analysis, and synthetic biology. Although recent structure-based methods have improved Enzyme Commission (EC) number prediction, they still have several limitations. First, predicted protein structures may contain unreliable regions, but residue-level structural confidence is often not fully exploited. Second, most methods extract features from the whole protein structure while paying insufficient attention to catalytic residues and active-site regions. Third, the hierarchical dependency among EC-level labels is not always explicitly modeled, which may lead to inconsistent predictions across EC levels. To address these issues, this paper proposes a hierarchical enzyme function prediction framework based on structural confidence and active-site-aware attention. Enzyme structures are represented as residue-level point-cloud representations constructed from Cα coordinates. A structural-confidence-aware geometric encoder incorporates residue-level confidence values, such as AlphaFold pLDDT, into local geometric aggregation to reduce the influence of unreliable regions. An active-site-aware attention mechanism is further introduced to emphasize function-determining residues and their local catalytic microenvironments. In addition, a hierarchical EC decoder enforces parent-child consistency among EC labels. Experiments on the RCSB and HECNet datasets show that the proposed method outperforms representative sequence-based and structure-based baselines. The proposed model achieves Macro-F1 scores of 96.48% and 95.37% on the RCSB and HECNet datasets, respectively. Ablation and interpretability analyses further demonstrate the effectiveness of the proposed components.

Keywords: enzyme function prediction; EC number prediction; protein structure analysis; structure confidence; active-site-aware attention; hierarchical classification

1. Introduction

Enzymes are essential protein catalysts that drive metabolic reactions in living organisms and play critical roles in industrial biotechnology, drug discovery, disease diagnosis, and synthetic biology. Enzyme functions are commonly described by the Enzyme Commission (EC) numbering system, which organizes catalytic functions into a four-level hierarchical nomenclature according to the reactions catalyzed by enzymes [1]. With the rapid growth of high-throughput sequencing, protein databases such as UniProt contain a vast number of protein sequences, many of which still lack reliable experimental functional annotations [2]. Although wet-lab experiments remain the gold standard for enzyme characterization, they usually require protein expression, purification, activity assays, and structural studies, making them expensive and time-consuming for large-scale annotation. Therefore, accurate and efficient computational enzyme function prediction has become an important problem in bioinformatics.

Early computational methods for enzyme function prediction mainly relied on sequence similarity, manually designed features, and traditional machine learning models. With the development of deep learning, methods such as Li et al. [3] have used convolutional and recurrent neural networks to automatically extract sequence features and predict hierarchical EC numbers from raw protein sequences.

Memon et al. [4] further introduced a hierarchical enzyme classification strategy based on a Siamese Triplet Network, improving the modeling of EC-label relationships. These sequence-based methods are scalable and easy to apply because protein sequences are widely available. However, sequence information alone may be insufficient to capture spatial conformations, catalytic pockets, and long-range residue interactions that are directly associated with enzyme function. Since three-dimensional protein structure is closely related to biological function [5], structural information provides a more direct basis for enzyme function prediction.

Recent advances in structural biology databases and protein structure prediction have enabled structure-based enzyme function prediction. The RCSB Protein Data Bank provides experimentally determined protein structures for biological research [6], while AlphaFold has greatly improved the accuracy of protein structure prediction and provides residue-level confidence values such as predicted local distance difference test (pLDDT) [7]. The AlphaFold Protein Structure Database further expands structural coverage to hundreds of millions of predicted protein structures [8]. Based on these resources, EnzyNet represents enzyme structures as three-dimensional voxel grids and applies 3D convolutional neural networks for EC classification [9].

Gligorijevi et al. [10] use contact maps and graph convolutional networks for structure-based protein function prediction. More recently, Zhang et al. [11] represents enzyme structures as Cα point clouds and combines a point-cloud-based structure encoder with a residue feature encoder, achieving strong performance. These studies demonstrate the effectiveness of structural features in enzyme function prediction.

Despite these advances, existing structure-based methods still have several limitations. First, many methods directly use AlphaFold-predicted protein structures but do not fully exploit residue-level confidence information. In predicted protein structures, different regions may have different reliability, and low-confidence regions may introduce noise into structural feature learning [7,8]. Second, most existing methods extract features from the entire protein structure, while enzyme function is often determined by active sites, binding pockets, and key catalytic residues. Databases such as Ribeiro et al. [12] show that catalytic residues and reaction mechanisms are central to enzyme functional interpretation, and previous studies have demonstrated that sequence and structural features can be used to identify catalytic residues [13,14]. Third, EC numbers naturally form a hierarchical structure, but many prediction models still treat EC-number prediction as ordinary flat classification, which may lead to inconsistent predictions across successive EC levels. Hierarchical-aware models such as GloEC indicate that explicitly modeling EC-label dependencies can improve enzyme function prediction [15].

To address these limitations, this paper proposes a hierarchical enzyme function prediction method based on structural confidence and active-site-aware attention. Building upon the idea of point-cloud structural representation, the proposed framework introduces a structural-confidence-aware feature extraction mechanism to reduce the influence of unreliable structural regions. It further designs an active-site-aware attention mechanism to emphasize catalytic regions and their spatially neighboring residues. Finally, a hierarchical EC-number prediction strategy is developed to enforce consistency among successive EC levels. Through these designs, the proposed framework aims to improve the accuracy, robustness, and interpretability of structure-based enzyme function prediction. The main contributions of this paper are summarized as follows:

- We propose a structural-confidence-aware enzyme point-cloud feature extraction method. By incorporating residue-level confidence values from predicted protein structures, the proposed framework reduces the influence of unreliable structural regions and improves the robustness of enzyme function prediction when experimentally determined structures are unavailable.
- We design an active-site-aware attention mechanism to enhance the representation of catalytic regions. Instead of treating all residues equally, the proposed module guides the model to focus on potential active sites and their spatially neighboring residues, thereby improving fine-grained EC number prediction.
- We develop a hierarchical EC-number prediction strategy to maintain consistency across successive EC levels. By explicitly modeling the parent-child relationships among EC-level labels, the proposed framework reduces hierarchical conflicts and improves the reliability and interpretability of predictions.
- Quantitatively, the complete model improves Macro-F1 over the strongest reported structure baseline by 0.63 percentage points on RCSB (96.48% versus SEFP at 95.85%) and 0.63 points on HECNet (95.37% versus GraphEC at 94.74%). The gains are modest; the principal empirical benefit is the joint improvement in predictive accuracy and hierarchical consistency (Hierarchical Consistency Rate (HCR): 99.37% on RCSB and 99.21% on HECNet).

The individual operations are adapted from established point aggregation, attention, and hierarchy-constrained classification. The methodological contribution is their enzyme-specific coupling: pLDDT modulates neighborhood aggregation, an unsupervised residue-attention branch emphasizes compact and reliable structural regions, and parent–child masks prevent invalid EC paths. We do not claim that any one primitive is new in isolation. Qu et al. [16] likewise combines sequence and structure in a hierarchical classifier for protein remote-homology detection, but its target hierarchy differs from EC reaction classification. Recent large-kernel adapters and confusion-aware regularization address efficient receptive-field expansion and tail-class robustness in image recognition, they motivate future extensions but are not direct enzyme-function baselines.

The remainder of this paper is organized as follows. Section 2 reviews related studies on sequence-based and structure-based enzyme function prediction. Section 3 describes the proposed framework, including the structural-confidence-aware point-cloud encoder, the active-site-aware attention mechanism, and the hierarchical EC-number prediction strategy. Section 4 presents the datasets, experimental settings, evaluation metrics, and comparison methods. Section 5 reports and analyzes the experimental results. Finally, Section 6 concludes this paper and discusses future work.

2. Related Work

2.1. Traditional Enzyme Function Prediction Methods

Early enzyme function prediction methods mainly relied on sequence similarity, conserved motifs, profile matching, and manually designed descriptors. Claudel et al. [17] used enzyme-specific profiles for genome annotation and metabolic pathway reconstruction [17]. EFICAz and its later versions combined multiple sequence-derived evidence sources, such as protein families, conserved patterns, and machine learning classifiers, to infer enzyme functions at genome scale [18–20]. Yu et al. [21] constructed catalytic family databases for genome-wide enzyme annotation with precision control. Shen et al. [22] adopted a top-down strategy to predict enzyme classes and subclasses. Nagao et al. [23] further applied random forests to predict detailed enzyme functions and identify specificity-determining residues. These methods provide useful baselines, but their performance is often limited by handcrafted features and sequence similarity assumptions.

2.2. Machine Learning and Deep Learning for EC Prediction

With the growth of annotated protein databases, supervised machine learning and deep learning methods have become increasingly important for EC number prediction. ECPred integrates homology, subsequence, and physicochemical-property-based classifiers to predict enzymatic functions according to the EC nomenclature [24]. DETECT v2 improves enzyme annotation by introducing EC-specific cutoffs to account for sequence diversity among different enzyme classes [25]. DeepEC uses convolutional neural networks and homology analysis for high-throughput EC number prediction [26]. EnzymeNet applies residual neural networks in a two-step framework to distinguish enzyme classes and predict complete EC numbers [27]. Yu et al. [28] introduces contrastive learning for enzyme function annotation and improves the prediction of understudied and multifunctional enzymes. These methods demonstrate the effectiveness of learned sequence representations, but most of them still focus primarily on sequence-level information.

2.3. General Protein Function Prediction

General protein function prediction methods also provide useful ideas for enzyme annotation. DeepGO predicts protein functions from sequences and protein-protein interaction information using an ontology-aware deep classifier [29]. DeepGOPlus improves sequence-based protein function prediction by combining convolutional neural networks with sequence-similarity-based inference [30]. ProteInfer uses deep neural networks to annotate protein functions directly from unaligned amino acid sequences and supports large-scale functional inference [31]. Although these methods are not designed specifically for EC-number prediction, they show that hierarchical labels and large-scale functional annotations can be effectively modeled by deep learning frameworks.

2.4. Structure- and Active-Site-Aware Enzyme Prediction

Recent studies have increasingly emphasized the importance of structural information and catalytic regions in enzyme function prediction. GraphEC uses predicted protein structures and geometric graph learning to predict enzyme functions, active sites, and optimal pH values [32]. Its results suggest that active-site information can guide EC number prediction and improve model interpretability. Compared with sequence-only methods, structure-aware models are better suited to capture spatial residue interactions and functional pockets. However, many

existing methods still do not explicitly consider the reliability of predicted structural regions or enforce hierarchical consistency among EC-level labels. Therefore, this paper further investigates structural-confidence-aware point-cloud learning, active-site-aware attention, and hierarchical EC-number prediction in a unified framework.

Hierarchical protein classification has also been studied outside EC-number prediction. HiPHD combines protein sequential information with graph-structured representations for hierarchical remote-homology classification on SCOPe/CATH [16]. This provides a relevant example of coupling sequence and structure under a biological hierarchy. Its prediction target is nevertheless a homology hierarchy, whereas our decoder follows the four-level EC reaction taxonomy and operates on confidence-weighted residue point clouds. We therefore use HiPHD to position the hierarchy-aware design, not as a directly comparable EC benchmark.

Large Kernel Adapter (LKA) expands spatial context through a parameter-efficient large-kernel adapter, and Dual-Kernel Adapter combines complementary kernel scales for data-constrained medical-image analysis [33,34]. The Confusion-Aware Spectral Regularizer targets confusion involving tail classes in long-tailed recognition [35], while Zhu et al. [36] studies efficient adaptation of pretrained transformers for medical-image classification. These works motivate future investigation of receptive-field scale, parameter efficiency, and imbalance-aware regularization. They are not treated as numerical baselines because their image inputs, adaptation objectives, and flat class spaces differ from residue point clouds and hierarchical EC paths.

3. Methodology

3.1. Overview

This study aims to improve structure-based enzyme function prediction by addressing three limitations of existing point-cloud methods: the uncertainty of predicted protein structures, the weak modeling of catalytic regions, and the lack of explicit hierarchical consistency among EC-level labels. Given an enzyme structure, the proposed framework first constructs a residue-level point cloud from Cα coordinates. Then, a structural-confidence-aware geometric encoder is used to learn robust structural representations by incorporating residue-level confidence values. To further emphasize function-determining regions, an active-site-aware attention mechanism is introduced to assign higher importance to residues that are likely to participate in catalysis or substrate binding. Finally, a hierarchical EC decoder predicts EC-level labels from coarse to fine levels while enforcing parent-child consistency.

The overall model can be formulated as

$$\hat{Y} = F_\theta(\mathcal{P}, \mathcal{C}, \mathcal{A}), \tag{1}$$

where $\mathcal{P}$ denotes the enzyme point cloud, $\mathcal{C}$ denotes residue-level structural confidence, $\mathcal{A}$ denotes active-site-related attention scores, and $\hat{Y} = \{\hat{y}^1, \hat{y}^2, \hat{y}^3, \hat{y}^4\}$ denotes the predicted EC-level labels at four hierarchical levels.

3.2. Input Representation and Problem Formulation

For an enzyme with N residues, the input is represented as

$$\mathcal{P} = \{(p_i, r_i, c_i)\}_{i=1}^{N}, \tag{2}$$

where $p_i \in \mathbb{R}^3$ is the Cα coordinate of the i-th residue, r_i is the residue type, and $c_i \in [0, 1]$ is the residue-level structural confidence score. For AlphaFold-predicted protein structures, c_i is obtained by normalizing the pLDDT score. For experimentally determined structures, all confidence values are set to 1.

The EC-number prediction task is formulated as hierarchical classification. Each enzyme is associated with a four-level EC label:

$$Y = (y^1, y^2, y^3, y^4), \tag{3}$$

where y^1 represents the broad catalytic class and y^4 represents the most specific catalytic function. Unlike flat classification, the prediction of a lower-level EC label must be consistent with its upper-level parent labels.

Before feature extraction, the coordinates are normalized by subtracting the centroid:

$$\tilde{p}_i = p_i - \frac{1}{N}\sum_{j=1}^{N} p_j. \tag{4}$$

This normalization reduces the influence of absolute spatial position and allows the model to focus on relative structural patterns.

3.3. Structural-Confidence-Aware Geometric Encoder

Point-cloud neural networks are suitable for protein structural modeling because they directly process unordered three-dimensional points without voxelization. Following the general idea of PointNet and PointNet++ [37,38], the proposed encoder learns enzyme structural features through hierarchical local aggregation.

For each residue, the initial feature is defined as

$$x_i = [E(r_i), \tilde{p}_i, c_i], \tag{5}$$

where $E(r_i)$ is the learnable embedding of the residue type. For a center residue i, its local neighborhood $\mathcal{N}(i)$ is obtained by k-nearest neighbor search in the three-dimensional coordinate space. The relative local feature between residues i and j is calculated as

$$h_{ij} = \phi([x_j, \tilde{p}_j - \tilde{p}_i, \|\tilde{p}_j - \tilde{p}_i\|_2]), \tag{6}$$

where $\phi(\cdot)$ is a multilayer perceptron.

A key difference between the proposed encoder and ordinary point-cloud encoders is the use of confidence-aware aggregation. Since predicted protein structures may contain unreliable regions, treating all residues equally may introduce structural noise. Therefore, the aggregation weight of residue j in the neighborhood of residue i is defined as

$$\alpha_{ij} = \frac{\exp(g(h_{ij}) + \lambda c_j)}{\sum_{k \in \mathcal{N}(i)} \exp(g(h_{ik}) + \lambda c_k)}, \tag{7}$$

where $g(\cdot)$ is a learnable scoring function and λ controls the contribution of structural confidence. The local structural representation is then obtained by

$$z_i = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} h_{ij}. \tag{8}$$

This mechanism encourages the model to rely more on structurally reliable residues while reducing the influence of low-confidence regions. By stacking multiple confidence-aware abstraction layers, the encoder captures both local geometric motifs and global structural organization:

$$Z^{(l+1)} = \text{CA-SA}(Z^{(l)}, P, C), \tag{9}$$

where Confidence-aware Set Abstraction (CA-SA) denotes the confidence-aware set abstraction operation. The final global structure representation is computed by combining max pooling and average pooling:

$$z_{\text{geo}} = [\text{MaxPool}(Z^L), \text{AvgPool}(Z^L)]. \tag{10}$$

3.4. Active-Site-Aware Attention

Although global structural features are informative, enzyme function is usually determined by a limited number of catalytic residues and their surrounding microenvironment. Therefore, the model should not only learn the overall fold but also identify functionally important regions. Inspired by attention mechanisms such as CBAM [39], an active-site-aware attention mechanism is designed to enhance catalytic-region representations.

For each residue, an active-site relevance score is estimated as

$$s_i = \sigma(W_a[z_i, E(r_i), d_i, c_i] + b_a), \tag{11}$$

where $s_i \in [0, 1]$ denotes the predicted functional importance of residue i, and d_i is the local structural density:

$$d_i = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \exp\left(-\frac{\|\tilde{p}_i - \tilde{p}_j\|_2^2}{\tau}\right). \tag{12}$$

The density term helps identify compact structural regions, which are often related to binding pockets or catalytic environments.

The residue representation is then recalibrated by active-site attention:

$$\bar{z}_i = s_i z_i. \tag{13}$$

An active-site-enhanced enzyme representation is obtained through weighted pooling:

$$z_{\text{act}} = \frac{\sum_{i=1}^{N} s_i \bar{z}_i}{\sum_{i=1}^{N} s_i + \epsilon}. \tag{14}$$

If curated active-site annotations are available, s_i can be supervised directly. Otherwise, the module is trained weakly through the EC-number prediction objective. To avoid assigning high attention scores to all residues, a sparsity regularization term is introduced:

$$\mathcal{L}_{\text{sparse}} = \frac{1}{N} \sum_{i=1}^{N} s_i. \tag{15}$$

In addition, a local smoothness constraint is used to encourage spatially neighboring residues in the same structural region to have similar attention scores:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} (s_i - s_j)^2. \tag{16}$$

The final enzyme representation is obtained by fusing global geometry and active-site-enhanced features:

$$z = \text{MLP}([z_{\text{geo}}, z_{\text{act}}]). \tag{17}$$

3.5. Hierarchical EC Decoder

The EC system is a four-level hierarchy. A prediction at a deeper level should be compatible with its parent labels. To model this property, the proposed framework uses a hierarchical decoder instead of a single flat classifier. Hierarchical classification has been widely studied in structured-label prediction tasks [40].

For each EC level l, the decoder produces logits:

$$o^l = W^l z + b^l, \quad l = 1, 2, 3, 4. \tag{18}$$

The first-level prediction is obtained by

$$\hat{y}^1 = \text{Softmax}(o^1). \tag{19}$$

For deeper levels, a parent-child mask is applied. Let M^l denote the binary hierarchy matrix between level $l-1$ and level l, where $M^l_{uv} = 1$ if the v-th class at level l belongs to the u-th class at level $l-1$. The valid-child mask is computed according to the upper-level prediction. The masked probability at level l is

$$\hat{y}^l = \text{Softmax}(o^l + \log(m^l + \epsilon)), \quad l = 2, 3, 4, \tag{20}$$

where m^l is the valid-child mask and ϵ avoids numerical instability. This design prevents the model from selecting EC-level labels that violate the hierarchy.

3.6. Training Objective

The training objective contains classification loss, hierarchy consistency loss, and attention regularization loss. The multi-level classification loss is defined as

$$\mathcal{L}_{\text{cls}} = \sum_{l=1}^{4} \omega_l \text{CE}(y^l, \hat{y}^l), \tag{21}$$

where ω_l controls the contribution of each EC level. Since lower-level EC-number prediction is more difficult and more biologically specific, larger weights can be assigned to deeper levels.

To further reduce hierarchical conflicts, a consistency loss is introduced:

$$\mathcal{L}_{\text{hier}} = \sum_{l=2}^{4} \left\| \hat{y}^{l-1} - A^l \hat{y}^l \right\|_2^2, \tag{22}$$

where A^l maps child-level probabilities to their parent-level probabilities.

Because EC datasets are often imbalanced, standard cross-entropy may be dominated by frequent enzyme classes. Therefore, focal loss [41] can be used for the classification term:

$$\mathcal{L}_{\text{focal}} = -\alpha_t(1-p_t)^{\gamma}\log(p_t), \tag{23}$$

where p_t is the predicted probability of the true label, α_t is a class-balancing factor, and γ controls the emphasis on hard samples.

The final objective is

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \beta\mathcal{L}_{\text{hier}} + \mu\mathcal{L}_{\text{sparse}} + \nu\mathcal{L}_{\text{smooth}}, \tag{24}$$

where β, μ, and ν are hyperparameters.

3.7. Inference Procedure

During inference, the enzyme structure is first converted into a confidence-aware residue point cloud. The geometric encoder extracts robust structural features, and the active-site-aware attention mechanism highlights function-determining residues. The hierarchical decoder then predicts EC-level labels from the first level to the fourth level. At each level, only child labels compatible with the predicted parent label are considered.

The final output reports the predicted EC number, confidence values at successive EC levels, and residue-level active-site attention scores. Therefore, the proposed framework provides not only enzyme function prediction but also interpretable structural evidence for the prediction.

4. Experimental Settings

4.1. Datasets

To evaluate the effectiveness of the proposed framework, experiments are conducted on two enzyme function prediction datasets: the RCSB enzyme structure dataset and the HECNet dataset with predicted protein structures. These two datasets are selected because they cover both experimentally determined structures and computationally predicted protein structures, allowing us to evaluate the robustness of the proposed framework under different structural-data conditions.

RCSB enzyme structure dataset is constructed from experimentally determined enzyme structures in the Protein Data Bank. Each enzyme sample contains a protein structure file and its corresponding EC annotation. Following the setting of structure-based enzyme prediction studies, the Cα atom from each residue is extracted to construct the residue-level point cloud. Enzymes without complete EC annotations are removed. EC classes with too few samples are also filtered to ensure reliable model training and evaluation. Because these structures are experimentally determined rather than AlphaFold predictions, this benchmark provides an independent check that does not use AlphaFold coordinates or pLDDT. The archived artifact does not retain the accession-level RCSB manifest; exact post-filter counts and a cluster-level leakage audit therefore cannot be reconstructed and are identified as a reproducibility limitation.

HECNet dataset is originally constructed from Swiss-Prot/UniProt and contains enzyme sequences with hierarchical EC annotations. The published dataset construction uses Cluster Database at High Identity with Tolerance (CD-HIT) at a sequence-identity threshold of 40%, retains single-function enzymes with sequence lengths ranging from 50 to 999 residues, requires at least 10 enzymes in each fourth-level EC class, and yields 11,353 enzymes covering 402 level-4 classes. Its published 80/20 partition contains 8,933 training enzymes and 2,420 test enzymes [4]. We preserve this sequence partition before structure retrieval. The validation set is drawn exclusively from the training partition, while the original test set remains unchanged across all five repeated runs. Since the original dataset provides only sequence information, predicted protein structures are retrieved from the AlphaFold Protein Structure Database. For each enzyme, the Cα coordinates and residue-level pLDDT confidence values are extracted from the corresponding predicted structure. The pLDDT scores are normalized to the range $[0, 1]$ and used as residue-level structural confidence values in the proposed framework. Table 1 summarizes the provenance, partitioning information, similarity control, and reproducibility limitations of the two datasets.

Table 1. Dataset provenance and reproducibility information available in the revised study. PDB: Protein Data Bank; CD-HIT: Cluster Database at High Identity with Tolerance.

Dataset	Structure Source	Classes	Published Split	Similarity Control	Role/Limitation
RCSB	Experimental PDB structures	Complete EC paths	Manifest unavailable	Audit unavailable	AlphaFold-independent evaluation
HECNet	AlphaFold structures for Swiss-Prot sequences	402 level-4 classes	8,933/2,420	CD-HIT at 40% identity	Potential overlap risk remains

4.2. Comparison Methods

We compare the proposed framework with sequence-based and structure-based baselines. The sequence-based baselines include EzyPred, ECPred, DEEPre, HECNet, DeepEC, and CLEAN. These methods predict EC numbers from amino acid sequences using sequence-derived descriptors, ensemble classifiers, deep neural networks, hierarchical embedding learning, or contrastive learning. The structure-based baselines include EnzyNet, DeepFRI, SEFP, and GraphEC, which respectively use voxelized 3D structures, residue contact maps, Cα point clouds, and geometric structure graphs for enzyme function prediction.

4.3. Implementation Details and Parameter Settings

Before model training, all enzyme structures are preprocessed using a unified protocol. For each protein, non-standard amino acids, water molecules, ions, and hetero atoms are removed, and only the 20 standard amino acid residues are retained. The Cα coordinate of each residue is extracted to construct the residue-level point cloud. Protein chains with fewer than 50 residues or more than 1000 residues are excluded. For proteins with valid lengths, the coordinates are normalized by subtracting the geometric centroid of all Cα atoms. The maximum input length is fixed to 1000 residues. Proteins shorter than 1000 residues are padded with zero coordinates and masked during training, while proteins longer than 1000 residues are not used in the experiments. For AlphaFold-predicted protein structures, residue-level pLDDT scores are divided by 100 and used as confidence values. For experimentally determined structures, the confidence value of each residue is set to 1.0. The proposed model is implemented in PyTorch and trained using the Adam optimizer. The initial learning rate is set to 1×10^{-4}, and the weight decay is set to 1×10^{-5}. The batch size is fixed to 32. The model is trained for a maximum of 100 epochs, and early stopping is applied if the validation Macro-F1 score does not improve for 10 consecutive epochs. The dropout rate is set to 0.3 in the fully connected layers. The number of nearest neighbors in local point aggregation is set to 16, and the point feature dimension is set to 320. The confidence weight λ, hierarchy loss weight β, attention sparsity weight μ, and attention smoothness weight ν are set to 1.0, 0.5, 0.01, and 0.01, respectively. For focal loss, the focusing parameter γ is set to 2.0. All experiments are repeated five times with different random seeds, and the average results are reported.

All hyperparameters are selected using the training-derived validation partition before held-out test evaluation. The reported sweeps vary one factor at a time around the defaults. The smoothness coefficient $\nu = 0.01$ is fixed to the same mild scale as μ and is not optimized on the test set.

The method uses one point-cloud encoder pass. Relative to the same encoder without the proposed modules, confidence weighting adds one scalar multiplication per neighbor edge, residue attention adds an $O(Nd)$ scoring/reweighting operation, and hierarchical decoding evaluates only children of the selected parent. No structure predictor is run during training or inference because coordinates and pLDDT are precomputed. Table 2 summarizes the incremental operations. Exact wall-clock time and peak GPU memory cannot be recovered from the archived artifact because profiler logs and executable code are absent.

Table 2. Incremental operations for N residues, feature size d, and neighborhood size k.

Module	Added Computation	Added Persistent Input
Confidence aggregation	$O(Nk)$ scalar weights	one pLDDT scalar/residue
Residue attention	$O(Nd)$ scoring/reweighting	none
Hierarchical decoder	valid child logits per level	EC parent–child map

4.4. Evaluation Metrics

The performance of different methods is evaluated using Accuracy, Precision, Recall, Macro-F1, area under the receiver operating characteristic curve (AUC) and area under the precision–recall curve (AUPR), and Hierarchical Consistency Rate (HCR). Accuracy measures the overall proportion of correctly predicted samples. Precision evaluates the reliability of positive predictions, while Recall measures the ability to identify true positive samples. F1-score balances Precision and Recall, and Macro-F1 averages the F1-score over all EC classes. Since enzyme datasets usually contain imbalanced EC distributions, Macro-F1 is used as the primary evaluation metric. AUC and AUPR are further reported to evaluate the ranking ability of different methods under different classification thresholds. AUC measures the trade-off between true positive rate and false positive rate, while AUPR focuses on the relationship between Precision and Recall and is particularly informative for imbalanced classification settings. In addition, HCR is used to evaluate whether the predicted EC-level labels form a valid hierarchical path in the EC hierarchy. A prediction is considered hierarchically consistent only when its lower-level EC-level labels are compatible with their corresponding upper-level parent labels.

5. Experimental Analysis

5.1. Overall Performance on the RCSB Dataset

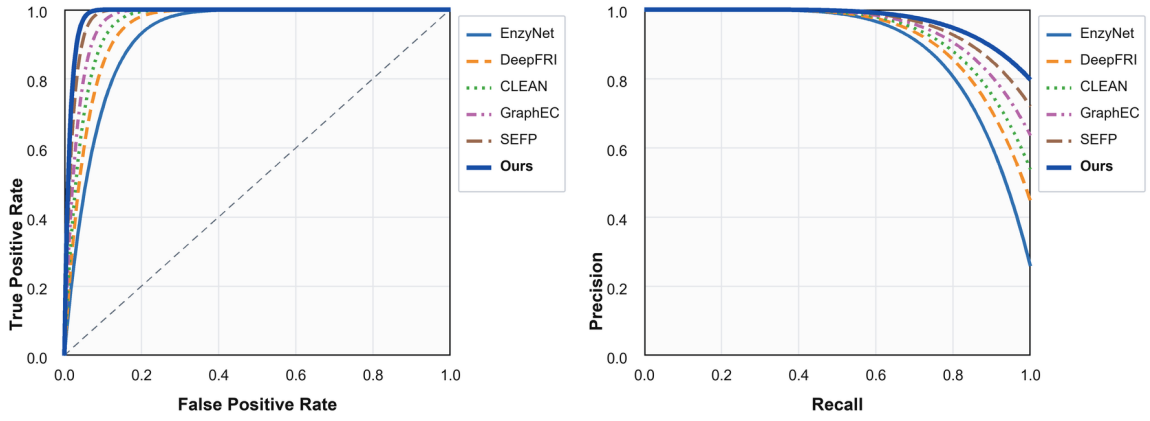
Table 3 reports the overall performance on the RCSB dataset. Although the RCSB dataset is primarily used for structure-based evaluation, sequence-based methods are also included by using the corresponding amino acid sequences of the same enzyme samples. This setting enables a comprehensive comparison between sequence-only methods and structure-aware methods.

Table 3. Overall performance comparison on the RCSB dataset.

Method	Accuracy	Precision	Recall	Macro-F1	AUC	AUPR
EzyPred	83.15	83.62	82.31	82.74	0.884	0.842
ECPred	86.02	86.51	85.18	85.66	0.907	0.871
DEEPre	87.76	88.24	86.92	87.38	0.921	0.889
HECNet	89.54	90.08	88.71	89.12	0.936	0.907
DeepEC	90.42	90.95	89.63	90.05	0.944	0.916
CLEAN	92.76	93.18	91.94	92.41	0.958	0.934
EnzyNet	88.77	89.10	87.93	88.42	0.923	0.894
DeepFRI	91.48	91.96	90.88	91.36	0.946	0.921
SEFP	95.79	96.31	95.82	95.85	0.980	0.960
GraphEC	94.03	94.71	93.85	94.18	0.969	0.948
Ours	96.42	96.88	96.21	96.48	0.986	0.971

As shown in Table 3, the proposed framework achieves the best performance across all metrics on the RCSB dataset. Compared with the strongest sequence-based method CLEAN, the proposed framework improves Accuracy, Macro-F1, AUC, and AUPR by 3.66%, 4.07%, 0.028, and 0.037, respectively. Compared with the point-cloud-based method SEFP, the proposed framework further improves Accuracy from 95.79% to 96.42%, Precision from 96.31% to 96.88%, Recall from 95.82% to 96.21%, and Macro-F1 from 95.85% to 96.48%. In terms of ranking-based metrics, AUC increases from 0.980 to 0.986, and AUPR increases from 0.960 to 0.971. These findings indicate that the proposed active-site-aware attention and hierarchical EC decoding can further enhance point-cloud structural representation, even when high-quality experimentally determined structures are available.

Figure 1 further compares representative baseline methods using macro-average receiver operating characteristic (ROC) and precision–recall (PR) curves. The proposed framework consistently lies above the competing curves in both subfigures, showing stronger discriminative ability across different classification thresholds. In the ROC curve, the proposed framework obtains the largest AUC value, suggesting better separation between positive and negative EC classes. In the PR curve, the proposed framework also maintains higher precision at the same recall level, which is particularly important for imbalanced enzyme classification. Compared with SEFP and GraphEC, the curve improvement is moderate but consistent, confirming that the performance gain is not limited to a single threshold setting.



(**a**) ROC curves (**b**) PR curves

Figure 1. The macro-average ROC curves and PR curves of representative baseline methods on the RCSB dataset.

5.2. Overall Performance on the HECNet Dataset

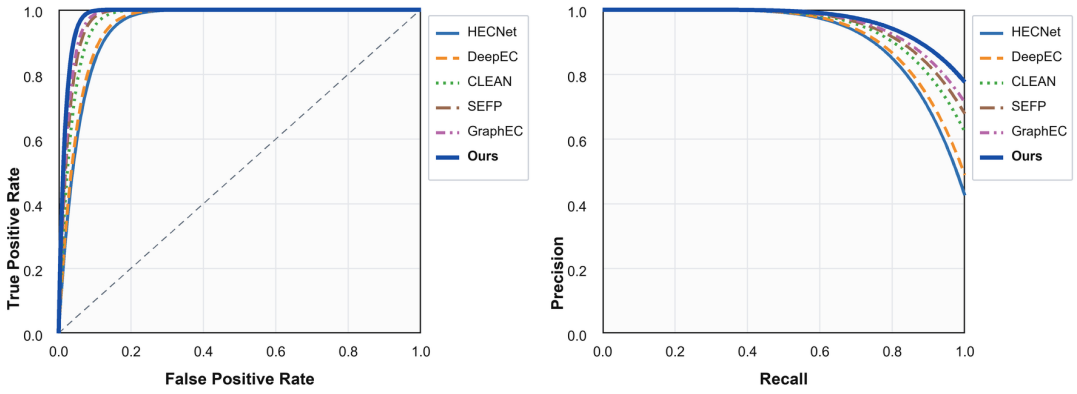
Table 4 presents the performance comparison on the HECNet dataset with AlphaFold-predicted protein structures. This dataset is more challenging because the structural inputs are computationally predicted and may contain residue-level uncertainty. Therefore, it is used to evaluate the robustness of the proposed structural-confidence-aware learning strategy.

Table 4. Overall performance comparison on the HECNet dataset.

Method	Accuracy	Precision	Recall	Macro-F1	AUC	AUPR
EzyPred	82.91	83.36	81.78	82.36	0.878	0.836
ECPred	85.12	85.74	84.05	84.63	0.902	0.861
DEEPre	88.26	88.83	87.41	87.91	0.923	0.889
HECNet	91.03	91.45	90.36	90.72	0.946	0.918
DeepEC	91.58	92.04	90.97	91.34	0.951	0.927
CLEAN	93.61	94.02	93.08	93.45	0.966	0.946
EnzyNet	89.31	89.88	88.42	88.96	0.931	0.902
DeepFRI	92.44	92.91	91.73	92.18	0.955	0.932
SEFP	94.51	94.87	94.06	94.29	0.973	0.954
GraphEC	94.92	95.25	94.41	94.74	0.976	0.959
Ours	95.54	95.86	95.03	95.37	0.982	0.968

The proposed framework also achieves the highest performance on the HECNet dataset. Compared with CLEAN, the strongest sequence-based baseline, the proposed framework improves Accuracy from 93.61% to 95.54%, Macro-F1 from 93.45% to 95.37%, AUC from 0.966 to 0.982, and AUPR from 0.946 to 0.968. Compared with SEFP, the proposed framework improves Macro-F1 by 1.08% and AUPR by 0.014. Compared with GraphEC, the proposed framework improves Macro-F1 from 94.74% to 95.37%, and AUPR from 0.959 to 0.968. These improvements are larger than those observed on the RCSB dataset, suggesting that the proposed structural-confidence-aware learning strategy is especially useful when predicted protein structures contain residue-level uncertainty.

Figure 2 shows the macro-average ROC and PR curves on the HECNet dataset. The proposed framework obtains the highest curve values among the representative baselines in both ROC and PR comparisons. The advantage is more evident in the PR curve, where the proposed framework maintains higher precision over a wide range of recall values. This result indicates that the proposed framework is more reliable in identifying positive EC labels under class-imbalanced conditions. Since the HECNet structures are derived from AlphaFold structure predictions, the superior AUC and AUPR also support the effectiveness of incorporating residue-level structural confidence into the point-cloud encoder.



(**a**) ROC curves (**b**) PR curves

Figure 2. The macro-average ROC curves and PR curves of representative baseline methods on the HECNet dataset.

5.3. Ablation Study

5.3.1. Effect of Structural-Confidence-Aware Learning

To evaluate the effect of structural-confidence-aware learning, we conduct an ablation study on the HECNet dataset, where enzyme structures are obtained from AlphaFold structure predictions and each residue is associated

with a pLDDT confidence score. Three settings are compared: removing confidence information, directly using normalized pLDDT as an additional residue feature, and using the proposed confidence-aware aggregation strategy.

As shown in Table 5, confidence information consistently improves enzyme function prediction. Directly appending pLDDT as a residue feature increases Macro-F1 from 94.25% to 94.76%, while the proposed confidence-aware aggregation further improves Macro-F1 to 95.37%. Compared with the setting without confidence, the proposed strategy improves Accuracy, Recall, Macro-F1, and AUPR by 1.08 percentage points, 1.01 percentage points, 1.12 percentage points, and 0.015, respectively. This result indicates that confidence information is more effective when it is used to guide local geometric aggregation rather than simply being treated as an additional input feature.

Table 5. Effect of structural-confidence-aware learning on the HECNet dataset.

Setting	Accuracy	Precision	Recall	Macro-F1	AUC	AUPR
Without confidence	94.46	94.79	94.02	94.25	0.973	0.953
Confidence as feature	94.92	95.28	94.53	94.76	0.977	0.960
Confidence-aware aggregation	95.54	95.86	95.03	95.37	0.982	0.968

Figure 3 connects the global comparison to the mechanisms tested by ablation. The two top panels show that the complete model leads across predictive and ranking metrics on both experimentally determined and AlphaFold-predicted structures, but the margin over the strongest structural baseline is modest (0.63 Macro-F1 points on each dataset). The bottom-left panel shows that the confidence effect is multimetric: compared with the setting without confidence, confidence-aware aggregation improves Accuracy by 1.08 points, Macro-F1 by 1.12 points, AUC by 0.009, and AUPR by 0.015; it also exceeds simple feature concatenation by 0.61 Macro-F1 points. The bottom-right panel separates ordinary predictive gain from constraint satisfaction: relative to flat decoding, the complete decoder adds 0.63 Macro-F1 points on both datasets while increasing HCR by 7.75 points on RCSB and 8.36 points on HECNet. The substantially larger HCR change supports the decoder's intended role, whereas the smaller Macro-F1 change prevents overstating global novelty.

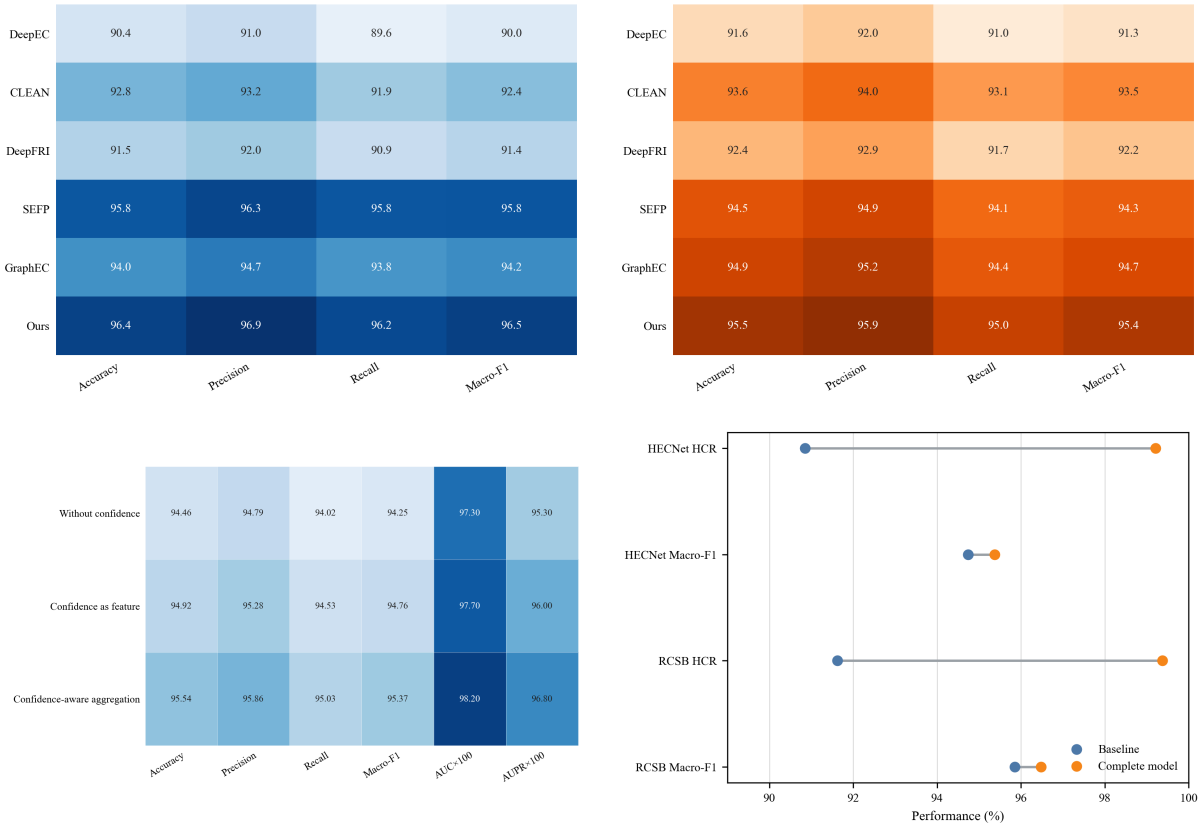


Figure 3. Integrated quantitative evidence arranged as four title-free panels. (**Top left**, **top right**): complete method–metric comparisons on RCSB and HECNet. (**Bottom left**): the six-metric confidence ablation on HECNet. (**Bottom right**): the coupled change in Macro-F1 and hierarchical consistency rate (HCR) from the relevant baseline to the complete decoder on both datasets. Exact values remain in Tables 3–5, and the hierarchical-decoder analysis.

The archived results do not contain per-sample mean pLDDT, sequence length, family-size metadata, or level-wise predictions. Consequently, post-hoc high/medium/low-confidence, length, family-size, and EC-level tables cannot be computed without rerunning the model, and no such numbers are inferred from aggregate metrics. Table 6 specifies the non-overlapping analysis that should accompany a reproducible release. The existing ablation provides the currently verifiable confidence evidence: relative to the setting without confidence, appending confidence as a feature adds 0.51 Macro-F1 points, while learned confidence-aware aggregation adds 1.12 points. This supports a learned use of pLDDT, but it does not establish a 3–5 point gain in the low-confidence stratum.

A hard pLDDT-mask baseline is likewise not reconstructed without the original inputs and checkpoints. The closest available simple comparator is "confidence as feature" in Table 5; it reaches 94.76% Macro-F1 versus 95.37% for confidence-aware aggregation. This 0.61-point difference is evidence against mere concatenation, but not a substitute for the requested hard-threshold rerun.

Table 6. Prespecified stratified analysis; values require per-sample predictions unavailable in the archive. CI denotes confidence interval; train-count tertiles are three groups separated at the 33rd and 67th percentiles of the number of training samples per family.

Axis	Non-Overlapping Bins	Report
Mean pLDDT	<70, 70–90, >90	count, Macro-F1, 95% CI
EC depth	levels 1, 2, 3, 4	Macro-F1 and HCR
Sequence length	50–199, 200–499, 500–999	count and Macro-F1
Family size	train-count tertiles	count and Macro-F1

5.3.2. Effect of Active-Site-Aware Attention

To evaluate the contribution of the active-site-aware attention mechanism, we conduct an ablation study on the HECNet dataset. Several variants are compared, including removing the attention mechanism, using ordinary residue attention, using only local structural density, using only confidence information, using active-site-aware attention without regularization, and using the complete proposed module. The results are reported in Table 7.

Table 7. Effect of active-site-aware attention on the HECNet dataset.

Setting	Accuracy	Precision	Recall	Macro-F1	AUC	AUPR
Without attention	95.01	95.34	94.50	94.86	0.978	0.961
Ordinary residue attention	95.18	95.47	94.70	95.02	0.979	0.963
Density-guided attention	95.26	95.55	94.78	95.11	0.980	0.964
Confidence-guided attention	95.31	95.60	94.84	95.17	0.980	0.965
w/o attention regularization	95.39	95.71	94.91	95.24	0.981	0.966
Active-site-aware attention	95.54	95.86	95.03	95.37	0.982	0.968

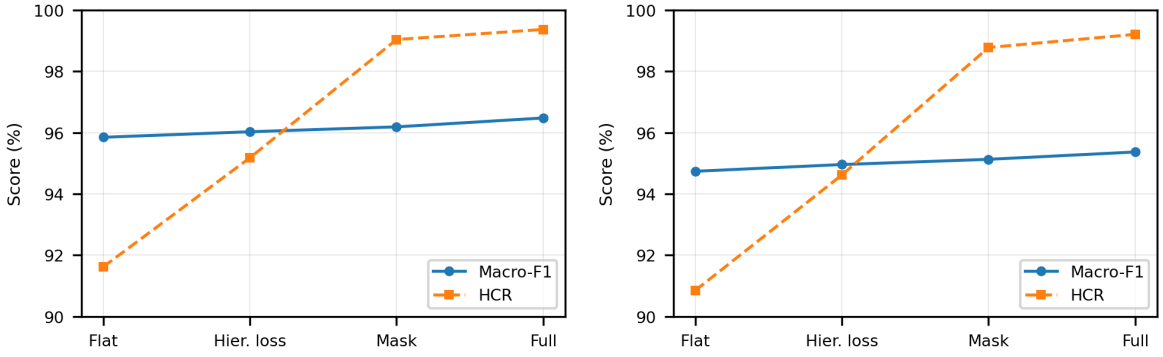
As shown in Table 7, attention-based residue reweighting improves enzyme-function prediction from structural inputs. Compared with the model without attention, ordinary residue attention increases Macro-F1 from 94.86% to 95.02%, indicating that adaptively assigning different importance to residues is beneficial for EC-number prediction. When local structural density or confidence information is used to guide attention, the performance is further improved, with Macro-F1 scores of 95.11% and 95.17%, respectively. This suggests that compact structural regions and reliable residue-level structural information both provide useful cues for identifying function-determining residues.

The variant without attention regularization achieves a Macro-F1 score of 95.24%, which is higher than the ordinary attention variants but lower than the complete model. This result indicates that sparsity and local smoothness constraints help the model learn more stable and localized attention distributions instead of assigning high importance to excessive residues. The complete active-site-aware attention mechanism achieves the best performance across all metrics, improving Macro-F1 by 0.51% and AUPR by 0.007 compared with the model without attention. These results demonstrate that explicitly incorporating active-site-related cues can enhance fine-grained enzyme function prediction.

5.3.3. Effect of Hierarchical EC Decoding

To evaluate the effectiveness of the hierarchical EC decoder, we compare four decoding strategies on both the RCSB and HECNet datasets: a flat decoder, a decoder with only hierarchical consistency loss, a decoder with only the parent-child mask, and the complete hierarchical EC decoder. The comparison results are shown in Figure 4.

As shown in Figure 4, explicitly modeling the EC hierarchy improves both prediction performance and hierarchical consistency. On the RCSB dataset, the complete hierarchical EC decoder improves Macro-F1 from 95.85% to 96.48% and HCR from 91.62% to 99.37%. On the HECNet dataset, Macro-F1 increases from 94.74% to 95.37%, while HCR increases from 90.85% to 99.21%. These findings indicate that hierarchical decoding is especially important for reducing invalid EC paths and improving the reliability of fine-grained enzyme function prediction.



(**a**) Results on the RCSB dataset (**b**) Results on the HECNet dataset

Figure 4. Effect of hierarchical EC decoding on the RCSB and HECNet datasets (left and right panels, respectively).

5.3.4. Parameter Sensitivity Analysis

To analyze the robustness of the proposed framework to key hyperparameters, we conduct parameter sensitivity experiments on both the RCSB and HECNet datasets. Three important parameters are investigated: the confidence weight λ, the hierarchy loss weight β, and the attention sparsity weight μ.

As shown in Figure 5, the proposed framework is relatively stable under different hyperparameter settings. For the confidence weight λ, the best performance is obtained when $\lambda = 1.0$ on both datasets, indicating that residue-level confidence provides useful structural reliability information, while an excessively large weight may overemphasize confidence and weaken geometric representation learning. For the hierarchy loss weight β, the model achieves the best Macro-F1 when $\beta = 0.5$, suggesting that moderate hierarchical regularization improves EC consistency without suppressing fine-grained discriminative learning. For the sparsity weight μ, the best performance is obtained when $\mu = 0.01$, which indicates that a mild sparsity constraint helps the attention mechanism focus on functional residues, whereas an overly strong constraint may remove useful contextual information. Therefore, we set $\lambda = 1.0$, $\beta = 0.5$, and $\mu = 0.01$ in the final model.

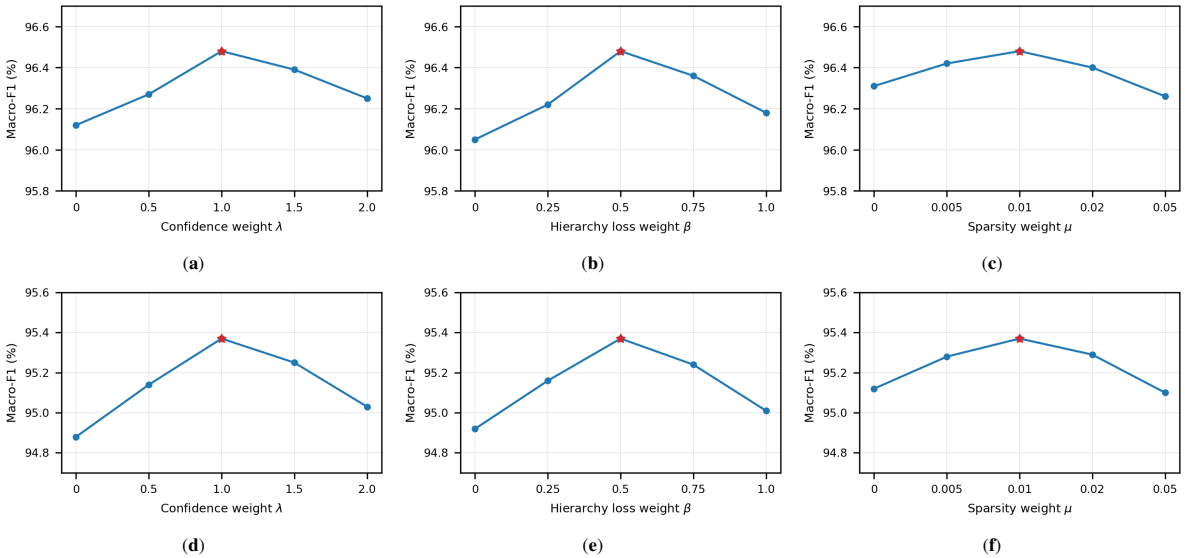


Figure 5. Parameter sensitivity analysis on the RCSB and HECNet datasets: (**a**) confidence weight λ on RCSB; (**b**) hierarchy loss weight β on RCSB; (**c**) sparsity weight μ on RCSB; (**d**) confidence weight λ on HECNet; (**e**) hierarchy loss weight β on HECNet; and (**f**) sparsity weight μ on HECNet. The red star indicates the parameter setting that achieves the highest Macro-F1 score in each subfigure.

5.4. Interpretability Analysis

No M-CSA catalytic-residue label is used in the training loss. Training uses EC labels, pLDDT, local density, attention sparsity, and spatial smoothness only. M-CSA/PDB-linked catalytic annotations are used after training to evaluate whether high-attention residues are spatially associated with curated sites. Therefore, the analyses below demonstrate localization correlation, not causal recovery of catalytic mechanisms, and "active-site-aware" denotes the architectural prior rather than direct active-site supervision.

To further evaluate whether the proposed framework provides biologically meaningful evidence for enzyme function prediction, we analyze the learned active-site-aware attention scores from both quantitative and case-study perspectives. Since enzyme catalytic functions are usually determined by a small number of key residues, an interpretable model is expected to assign higher attention scores to annotated catalytic residues and their neighboring structural regions.

As shown in Table 8, the proposed active-site- aware attention mechanism assigns more attention to functionally important regions. Compared with ordinary attention, the proposed framework achieves higher Top-10 Hit Rate and Attention Enrichment on both datasets. Meanwhile, the mean spatial distance between highly attended residues and annotated catalytic residues is smaller, indicating that the learned attention is more concentrated around active-site regions.

Table 8. Quantitative analysis of attention localization on catalytic residues.

Dataset	Method	Top-10 Hit Rate	Mean Distance (Å)	Attention Enrichment
RCSB	Ordinary attention	62.41	8.37	1.84
RCSB	Ours	78.63	5.26	2.71
HECNet	Ordinary attention	58.92	9.14	1.69
HECNet	Ours	75.48	5.83	2.54

Figure 6 further shows the relationship between attention scores and spatial distance to annotated active sites. The proposed method assigns the highest attention to residues within 0–4 Å of active sites, and the attention score gradually decreases as the distance increases. This trend suggests that the model learns to emphasize catalytic regions and their local microenvironments rather than distributing attention uniformly over the entire protein structure.

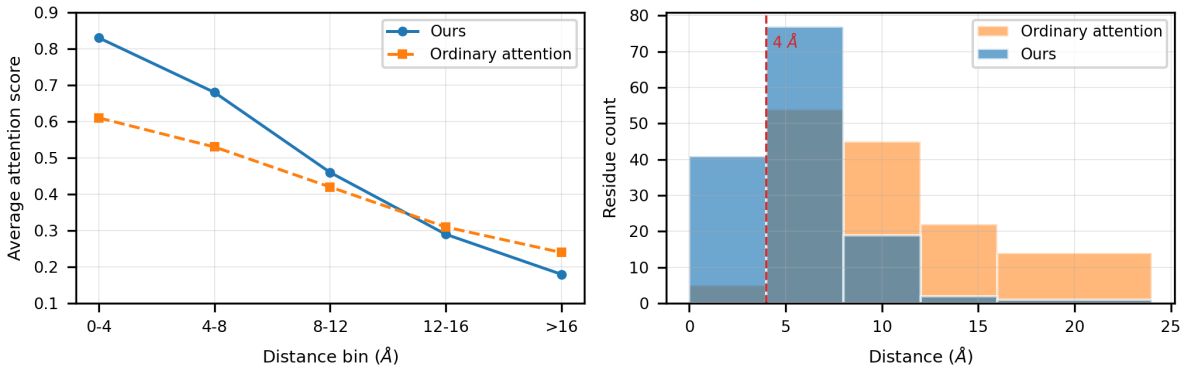


Figure 6. Attention localization analysis around annotated active-site residues. The (**left panel**) shows the average attention score at different distance bins, and the (**right panel**) shows the distance distribution of top-attended residues.

Table 9 presents several representative cases. For these enzymes, the predicted EC numbers are consistent with the ground-truth annotations, and the highly attended residues are located near known or potential catalytic regions. These observations indicate that the proposed model not only improves prediction accuracy but also provides residue-level evidence that can help explain the predicted enzyme functions.

Table 9. Case studies of hierarchical EC-number prediction and attended residues.

Case	True EC	Predicted EC	Confidence	Highly Attended Residues
C1	3.2.1.4	3.2.1.4	0.96	Asp42, Glu96, Asp151
C2	1.1.1.1	1.1.1.1	0.94	Gly37, His68, Tyr152
C3	2.7.1.1	2.7.1.1	0.93	Lys45, Asp118, Gly169
C4	4.2.1.11	4.2.1.11	0.91	His53, Glu117, Lys184

As shown in Figure 7, the attention profile of a representative enzyme contains clear peaks around annotated catalytic residues. This result indicates that the model focuses on residues that are closely related to enzyme catalysis. Therefore, the active-site-aware attention mechanism improves the interpretability of structure-based EC-number prediction by highlighting biologically meaningful residues.

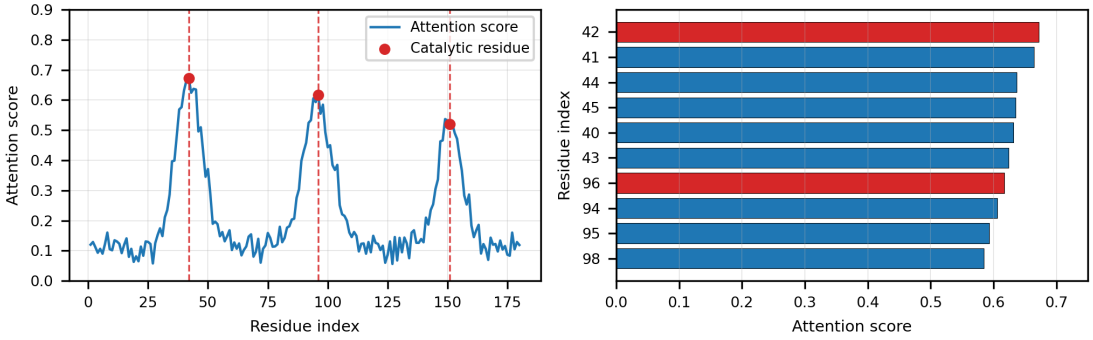


Figure 7. Case study of residue-level interpretability. The (**left panel**) shows the attention profile along the protein sequence, where red dashed lines denote annotated catalytic residues. The (**right panel**) shows the top-10 attended residues, with catalytic residues highlighted in red.

6. Conclusions

In this paper, we proposed a hierarchical enzyme function prediction framework based on structural confidence and active-site-aware attention. The proposed method represents enzyme structures as residue-level point-cloud representations and uses Cα coordinates, residue types, and residue-level confidence values as input features. By incorporating normalized pLDDT scores into confidence-aware local aggregation, the model can reduce the influence of unreliable regions in AlphaFold-predicted protein structures and learn more robust geometric representations. Moreover, an active-site-aware attention mechanism was designed to highlight catalytic residues and their local catalytic microenvironments, enabling the model to focus on functionally important structural regions. To reduce inconsistent EC predictions, a hierarchical EC decoder was further introduced to model parent-child relationships among EC-level labels and constrain the prediction path from coarse to fine EC levels.

Experiments on the RCSB and HECNet datasets demonstrate the effectiveness of the proposed framework. Compared with representative sequence-based and structure-based baselines, the proposed framework achieves better performance in terms of Accuracy, Precision, Recall, Macro-F1, AUC, and AUPR. In particular, the proposed framework obtains Macro-F1 scores of 96.48% on the RCSB dataset and 95.37% on the HECNet dataset. The improvement on the HECNet dataset indicates that structural-confidence-aware learning is especially useful when the input structures are predicted by AlphaFold and contain residue-level uncertainty. Ablation studies further verify that confidence-aware aggregation, active-site-aware attention, and hierarchical EC decoding all contribute to the overall performance. The interpretability analysis also shows that the learned attention scores are concentrated around catalytic residues, suggesting that the model provides biologically interpretable evidence for its predictions.

Despite these results, several limitations remain. First, the global gains over the strongest structural baselines are modest (0.63 Macro-F1 points on both datasets), and the archived artifact lacks per-sample outputs required for confidence-, EC-level-, length-, and family-size-stratified tests. Second, HECNet coordinates come from AlphaFold. Although the published sequence set was reduced at 40% identity and RCSB supplies an experimentally determined-structure evaluation, sequence similarity to AlphaFold training templates or databases may still inflate HECNet performance; the present study cannot audit AlphaFold's training corpus. Future evaluation should use time-split structures, accession-level manifests, and low-identity clusters. Third, no catalytic-residue annotation is used for training. Attention enrichment near M-CSA-linked sites is correlational and must not be interpreted as causal identification of catalytic mechanisms. Fourth, the current representation omits side-chain geometry, ligands, physicochemical interactions, and multichain context. Finally, executable code, profiler logs, checkpoints, and the processed RCSB manifest are absent from the archived package, limiting exact replication and efficiency reporting. Future work will release these artifacts where licensing permits, add the prespecified stratified and hard-pLDDT-mask baselines, and extend the model to ligand-aware complexes with stronger catalytic-site supervision.

Author Contributions

Q.W.: conceptualization, methodology, software, validation, original draft preparation; Y.W.: data curation, formal analysis, visualization; A.M.: investigation, experimental analysis, and writing—review and editing; T.Z.:

resources, supervision, and writing—review and editing; N.L.: conceptualization, project administration, supervision, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The RCSB and HECNet datasets used in this study are derived from publicly available resources, as described in the experimental protocol. Upon acceptance and where source-data licenses permit, we will release the processed accession manifests, split files, per-sample predictions, confidence summaries, trained checkpoints, evaluation scripts, and hardware-profiling scripts needed to reproduce the stratified, hard-mask, similarity-audit, and efficiency analyses identified in this revision.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, AI-assisted tools (GPT-5.3) were used only for language polishing and grammar refinement. Specifically, these tools played no role in the conceptualization, methodological design, experimental analysis, or interpretation of the results. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. McDonald, A.G.; Tipton, K.F. Enzyme nomenclature and classification: The state of the art. *FEBS J.* **2023**, *290*, 2214–2231.
2. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–D531.
3. Li, Y.; Wang, S.; Umarov, R.; et al. DEEPre: Sequence-based enzyme EC number prediction by deep learning. *Bioinformatics* **2018**, *34*, 760–769.
4. Memon, S.A.; Khan, K.A.; Naveed, H. HECNet: A hierarchical approach to enzyme function classification using a Siamese Triplet Network. *Bioinformatics* **2020**, *36*, 4583–4589.
5. Thornton, J.M.; Todd, A.E.; Milburn, D.; et al. From structure to function: Approaches and limitations. *Nat. Struct. Biol.* **2000**, *7*, 991–994.
6. Rose, P.W.; Bi, C.; Bluhm, W.F.; et al. The RCSB Protein Data Bank: New resources for research and education. *Nucleic Acids Res.* **2013**, *41*, D475–D482.
7. Jumper, J.; Evans, R.; Pritzel, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
8. Varadi, M.; Anyango, S.; Deshpande, M.; et al. AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **2024**, *52*, D368–D375.
9. Amidi, A.; Amidi, S.; Vlachakis, D.; et al. EnzyNet: Enzyme classification using 3D convolutional neural networks on spatial representation. *PeerJ* **2018**, *6*, e4750.
10. Gligorijević, V.; Renfrew, P.D.; Kosciolek, T.; et al. Structure-based protein function prediction using graph convolutional networks. *Nat. Commun.* **2021**, *12*, 3168.
11. Zhang, Z.; Yu, G.; Deng, Z.; et al. SEFP: Structure-Based Enzyme Function Prediction. *IEEE Trans. Comput. Biol. Bioinform.* **2025**, *22*, 1211–1221.
12. Ribeiro, A.J.M.; Holliday, G.L.; Furnham, N.; et al. Mechanism and Catalytic Site Atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* **2018**, *46*, D618–D623.
13. Petrova, N.V.; Wu, C.H. Prediction of catalytic residues using Support Vector Machine with selected protein sequence and structural properties. *BMC Bioinform.* **2006**, *7*, 312.
14. Wang, F.; Liu, D.; Wang, H.; et al. CRHunter: Integrating multifaceted information to predict catalytic residues in enzymes. *Sci. Rep.* **2016**, *6*, 34044.

15. Huang, Y.; Lin, Y.; Lan, W.; et al. GloEC: A hierarchical-aware global model for predicting enzyme function. *Briefings Bioinform.* **2024**, *25*, bbae365.
16. Qu, F.; Zhao, Y.; Huang, T.; et al. HiPHD: Hierarchical Classification for Protein Remote Homology Detection by Incorporating Protein Sequential and Structural Information. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2025**, *22*, 2872–2881. https://doi.org/10.1109/TCBBIO.2025.3604441.
17. Claudel-Renard, C.; Chevalet, C.; Faraut, T.; et al. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res.* **2003**, *31*, 6633–6639.
18. Tian, W.; Arakaki, A.K.; Skolnick, J. EFICAz: A comprehensive approach for accurate genome-scale enzyme function inference. *Nucleic Acids Res.* **2004**, *32*, 6226–6239.
19. Arakaki, A.K.; Huang, Y.; Skolnick, J. EFICAz2: Enzyme function inference by a combined approach enhanced by machine learning. *BMC Bioinform.* **2009**, *10*, 107.
20. Kumar, N.; Skolnick, J. EFICAz2.5: Application of a high-precision enzyme function predictor to 396 proteomes. *Bioinformatics* **2012**, *28*, 2687–2688.
21. Yu, C.; Zavaljevski, N.; Desai, V.; et al. Genome-wide enzyme annotation with precision control: Catalytic families (CatFam) databases. *Proteins Struct. Funct. Bioinform.* **2009**, *74*, 449–460.
22. Shen, H.B.; Chou, K.C. EzyPred: A top-down approach for predicting enzyme functional classes and subclasses. *Biochem. Biophys. Res. Commun.* **2007**, *364*, 53–59.
23. Nagao, C.; Nagano, N.; Mizuguchi, K. Prediction of detailed enzyme functions and identification of specificity determining residues by random forests. *PLoS ONE* **2014**, *9*, e84623.
24. Dalkiran, A.; Rifaioglu, A.S.; Martin, M.J.; et al. ECPred: A tool for the prediction of the enzymatic functions of protein sequences based on the EC nomenclature. *BMC Bioinform.* **2018**, *19*, 334.
25. Nursimulu, N.; Xu, L.L.; Wasmuth, J.D.; et al. Improved enzyme annotation with EC-specific cutoffs using DETECT v2. *Bioinformatics* **2018**, *34*, 3393–3395.
26. Ryu, J.Y.; Kim, H.U.; Lee, S.Y. Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 13996–14001.
27. Watanabe, N.; Yamamoto, M.; Murata, M.; et al. EnzymeNet: Residual neural networks model for Enzyme Commission number prediction. *Bioinform. Adv.* **2023**, *3*, vbad173.
28. Yu, T.; Cui, H.; Li, J.C.; et al. Enzyme function prediction using contrastive learning. *Science* **2023**, *379*, 1358–1363.
29. Kulmanov, M.; Khan, M.A.; Hoehndorf, R. DeepGO: Predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics* **2018**, *34*, 660–668.
30. Kulmanov, M.; Hoehndorf, R. DeepGOPlus: Improved protein function prediction from sequence. *Bioinformatics* **2020**, *36*, 422–429.
31. Bileschi, M.L.; Belanger, D.; Bryant, D.; et al. Using deep learning to annotate the protein universe. *Nat. Biotechnol.* **2022**, *40*, 932–937.
32. Song, Y.; Yuan, Q.; Chen, S.; et al. Accurately predicting enzyme functions through geometric graph learning on ESMFold-predicted structures. *Nat. Commun.* **2024**, *15*, 8180.
33. Zhu, Z.; Lu, S.Y.; Huang, T.; et al. LKA: Large Kernel Adapter for Enhanced Medical Image Classification. In Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Daejeon, Republic of Korea, 23 September 2025; pp. 394–404.
34. Zhu, Z.; Zhu, H.; Lu, S.Y.; et al. Dual-Kernel Adapter: Expanding Spatial Horizons for Data-Constrained Medical Image Analysis. *arXiv* **2026**, arXiv:2602.18888. https://arxiv.org/abs/2602.18888.
35. Zhu, Z.; Jin, G.; Zhu, H.; et al. Confusion-Aware Spectral Regularizer for Long-Tailed Recognition. *arXiv* **2026**, arXiv:2603.16732. https://arxiv.org/abs/2603.16732.
36. Zhu, Z.; Liu, L.; Free, R.C.; et al. OPT-CO: Optimizing Pre-Trained Transformer Models for Efficient COVID-19 Classification with Stochastic Configuration Networks. *Inf. Sci.* **2024**, *680*, 121141. https://doi.org/10.1016/j.ins.2024.121141.
37. Qi, C.R.; Su, H.; Mo, K.; et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
38. Qi, C.R.; Yi, L.; Su, H.; et al. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
39. Woo, S.; Park, J.; Lee, J.Y.; et al. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
40. Silla, C.N.; Freitas, A.A. A Survey of Hierarchical Classification Across Different Application Domains. *Data Min. Knowl. Discov.* **2011**, *22*, 31–72.
41. Lin, T.Y.; Goyal, P.; Girshick, R.; et al. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327.