algorithms

MDPI

Article

Multi-Scale Feature Mixing of Language Model Embeddings for Enhanced Prediction of Submitochondrial Protein Localization

Rong Wang [1], Menghua Wang [2], Yibo Wu [3], Lixiang Yang [2] and Xiao Wang [2,*]

[1] School of Electronics and Information, Zhengzhou University of Light Industry, Zhengzhou 450002, China; wangrong@zzuli.edu.cn
[2] School of Computer Science and Artificial Intelligence, Zhengzhou University of Light Industry, Zhengzhou 450002, China; 332405070486@zzuli.edu.cn (M.W.); yanglixiang@email.zzuli.edu.cn (L.Y.)
[3] School of Computational Economics, Henan University of Economics and Law, Zhengzhou 450046, China; 202334180336@stu.huel.edu.cn
* Correspondence: wangxiao@zzuli.edu.cn

Abstract

Accurate prediction of submitochondrial localization is fundamental to understanding mitochondrial biogenesis and cellular metabolic pathways. While deep representations from pre-trained protein language models (pLMs) have significantly advanced the field, traditional global average pooling methods often fail to capture critical, localized N-terminal targeting signals, particularly in long sequences where these motifs are mathematically diluted. To resolve this "signal dilution" bottleneck, we developed a multi-scale architecture that explicitly integrates high-resolution N-terminal features with global evolutionary context derived from ESM-2 embeddings. The proposed framework utilizes an orthogonal mixing strategy consisting of Token-mixing and Channel-mixing. Token-mixing is specifically designed to detect spatial rhythmic patterns across residue positions, while Channel-mixing refines the biochemical signatures within the latent feature space. Extensive benchmarking across diverse datasets demonstrates that our approach effectively maintains signal integrity. Compared to existing state-of-the-art methods, the model achieves a superior overall Generalized Correlation Coefficient (GCC) of 0.7443 on the SM424-18 dataset and 0.7878 on the SubMitoPred dataset, outperforming the latest benchmarks by 9.4% and 16.1%, respectively. Furthermore, on the independent M983 test set, our method maintained a high GCC of 0.6945, demonstrating a 9.9% improvement relative to the state-of-the-art methods. This robust and efficient framework provides a high-precision tool for large-scale mitochondrial proteomics.

Keywords: subcellular localization; submitochondrial localization; protein language model; deep learning

Academic Editors: Roberto Pagliarini and Carla Piazza

Received: 5 February 2026
Revised: 8 March 2026
Accepted: 10 March 2026
Published: 11 March 2026

1. Introduction

Mitochondria are double-membrane organelles that serve as the fundamental hub for cellular energy transduction, metabolic integration, and programmed cell death [1]. Beyond their primary role in ATP synthesis through oxidative phosphorylation, mitochondria are involved in critical processes such as calcium signaling, heme biosynthesis, and the regulation of innate immunity [2]. The mammalian mitochondrial proteome comprises approximately 1100 to 1500 distinct proteins, yet only a small fraction is encoded by the mitochondrial DNA (mtDNA). The vast majority of these proteins are encoded by the nuclear genome, synthesized on cytosolic ribosomes, and subsequently imported into the organelle

through sophisticated protein translocases [3,4]. Within the mitochondrion, proteins are strictly partitioned into four distinct sub-compartments: the outer mitochondrial membrane, the inner mitochondrial membrane, the intermembrane space, and the matrix. Each compartment possesses a unique microenvironment and protein composition tailored to its specific physiological functions; for instance, the inner mitochondrial membrane houses the respiratory chain complexes, while the matrix contains the enzymes for the tricarboxylic acid (TCA) cycle. Consequently, a protein's sub-organellar localization is a prerequisite for its functional maturity. Mislocalization is often associated with severe pathological conditions, including neurodegenerative diseases and metabolic syndromes [5]. Given the labor-intensive nature of experimental techniques such as immunofluorescence and sub-cellular fractionation, there is an urgent need for high-throughput computational tools to accurately predict submitochondrial localization from primary sequence data.

In recent years, the explosive growth of protein sequence data has enabled researchers to extract deeper structural information, significantly advancing research in computer-aided protein sub-subcellular prediction, including subnuclear [6–11], subchloroplast [12–16], and submitochondrial localization [17–21]. The computational prediction of protein submitochondrial localization has progressed through several distinct stages of development. Early research focused on the identification of N-terminal presequences, which are characterized by an abundance of positively charged residues and the ability to form amphiphilic α-helices. Claros and Vincens developed MitoProt II [22], which utilized these statistical properties to estimate the probability of mitochondrial import. Similarly, Emanuelsson and colleagues introduced TargetP [23,24], which employed artificial neural networks to discriminate between mitochondrial targeting signals and other sorting peptides based on N-terminal residue patterns.

As the field moved toward resolving internal sub-organellar partitioning, researchers integrated a wider array of biological descriptors. Du and Li developed SubMito [17], which hybridized pseudo-amino acid composition with various physicochemical features of segmented sequences using Support Vector Machines (SVMs). Du and Li further improved this by incorporating evolutionary information into the Pseudo-Amino Acid Composition (PseAAC) framework [25]. Building upon these traditional machine learning approaches, SubMito-XGBoost [20] was developed to leverage the power of ensemble learning. By fusing multiple types of feature information, including evolutionary profiles and physicochemical properties, and employing the eXtreme Gradient Boosting (XGBoost) algorithm, this method significantly improved the robustness and predictive accuracy of submitochondrial localization.

The introduction of deep learning allowed for the extraction of non-linear sequential dependencies. Savojardo et al. developed DeepMito [18], which used multi-branch convolutional neural networks (CNNs) to process evolutionary profiles. Subsequently, the same group released BUSCA [26] for integrative sub-cellular mapping. More recently, several specialized frameworks have further refined protein submitochondrial localization prediction. iDeepSubMito [19] implemented a deep learning framework optimized for hierarchical feature extraction from sequences. DeepPred-SubMito [27] introduced a multi-channel CNN combined with dataset balancing treatments to improve performance on minority classes. Furthermore, Jiang et al. proposed a framework focused on residue-level interpretation [28], allowing for the visualization of amino acids that contribute to localization decisions.

More recently, the emergence of protein language models (pLMs), such as ESM-2, has further pushed the performance boundaries by leveraging high-dimensional embeddings that encode deep evolutionary and structural information. Recent pLM-based methodologies typically involve transformer architecture fine-tuning or the integration of complex attention mechanisms on top of the sequence embeddings. While theoretically highly expressive, these transformer-heavy approaches face significant limitations in submitochondrial

localization. First, they are intrinsically data-hungry and heavily parameterized, making them highly susceptible to severe overfitting when applied to the small-scale and imbalanced datasets available in this domain (e.g., containing fewer than 600 sequences). Second, they frequently rely on global attention or mean pooling mechanisms across the entire sequence. In long mitochondrial proteins, this mathematically dilutes the high-resolution signal of short N-terminal motifs within the massive non-targeting sequence data, causing the model to miss critical 'sorting codes'.

To address the dual bottlenecks of signal dilution and model overfitting, we propose Mito-Mixer, a multi-scale feature mixing framework. We introduce a Mixer-based architecture [29] that employs two orthogonal operations: Token-mixing (analogous to learning the 'spatial melody' or positional arrangement of amino acids within targeting motifs) and Channel-mixing (which refines the 'biochemical vocabulary' or specific physical properties at each position). By utilizing a parameter-efficient MLP-Mixer architecture on top of frozen ESM-2 embeddings, Mito-Mixer bypasses the computational overhead and overfitting risks associated with transformer fine-tuning while explicitly preserving localized N-terminal features. This multi-scale approach ensures that short, critical motifs remain prominent even in exceptionally long proteins, bridging the gap between global evolutionary context and local targeting signals.

2. Materials and Methods

2.1. Datasets

In this study, we utilize three benchmark datasets to evaluate the performance and generalizability of Mito-Mixer. These datasets are curated from high-quality experimental annotations and represent different scales of mitochondrial proteomic data. The characteristics of these datasets are summarized in Table 1.

Table 1. Statistical distribution of the three datasets (SM424-18, SubMitoPred and M983).

Dataset	Matrix	Intermembrane Space	Inner Membrane Proteins	Outer Membrane	Total
SM424-18	135	25	190	74	424
SubMitoPred	174	32	282	82	570
M983	177	- *	661	145	983

* Note: The M983 dataset inherently lacks experimental data for intermembrane space proteins, hence the omission in this column.

The SM424-18 dataset was curated from UniProtKB/Swiss-Prot (release 2018_02) [18]. It contains non-fragmented mitochondrial proteins with experimental evidence. To minimize sequence redundancy, the CD-HIT program [30] was employed to establish a 40% identity threshold. The final dataset consists of 424 proteins partitioned into four subcompartments: 135 matrix proteins, 25 intermembrane space proteins, 190 inner membrane proteins, and 74 outer membrane proteins.

The SubMitoPred dataset was constructed by Kumar et al. [31]. The selection criteria included full-length protein sequences longer than 50 residues. Sequence homology was limited to 40% using the CD-HIT program. This dataset contains 570 protein sequences, including 174 matrix proteins, 32 intermembrane space proteins, 282 inner membrane proteins, and 82 outer membrane proteins.

To demonstrate the generalizability of Mito-Mixer, we utilize the M983 dataset as an independent testing set. Constructed by Du et al. [25], M983 is a large-scale dataset composed of 983 mitochondrial proteins. It includes 177 matrix proteins, 661 inner membrane proteins, and 145 outer membrane proteins. Testing on M983 ensures that the model maintains high predictive accuracy on large-scale data that was not encountered during the initial training phase.

2.2. The Mito-Mixer Framework

The core architectural innovation of this study is the Mito-Mixer, a framework designed to bridge the gap between high-dimensional evolutionary semantics and the fine-grained, localized targeting motifs required for protein submitochondrial localization. Mito-Mixer is motivated by a critical biological observation: while mitochondrial proteins exhibit vast diversity in total length, their localization is predominantly dictated by short N-terminal motifs, such as Mitochondrial Targeting Signals (MTS). Standard protein language model (pLM) pipelines often rely on Global Mean Pooling, which collapses the entire sequence into a single vector. This process mathematically dilutes the high-resolution N-terminal signal within the noise of the much longer non-targeting sequence. To address this, Mito-Mixer treats residue-level embeddings as a structured feature map and utilizes an MLP-based mixing strategy to capture long-range dependencies and biochemical patterns without the heavy computational overhead of Transformers. The overall architecture of the Mito-Mixer is illustrated in Figure 1. Before detailing the mathematical formulations, Figure 1 visually maps the data flow. Conceptually, the model first extracts a raw multi-scale feature map from the protein, and then passes this map through alternating blocks that separately refine the spatial arrangement of the amino acids and their intrinsic chemical properties.

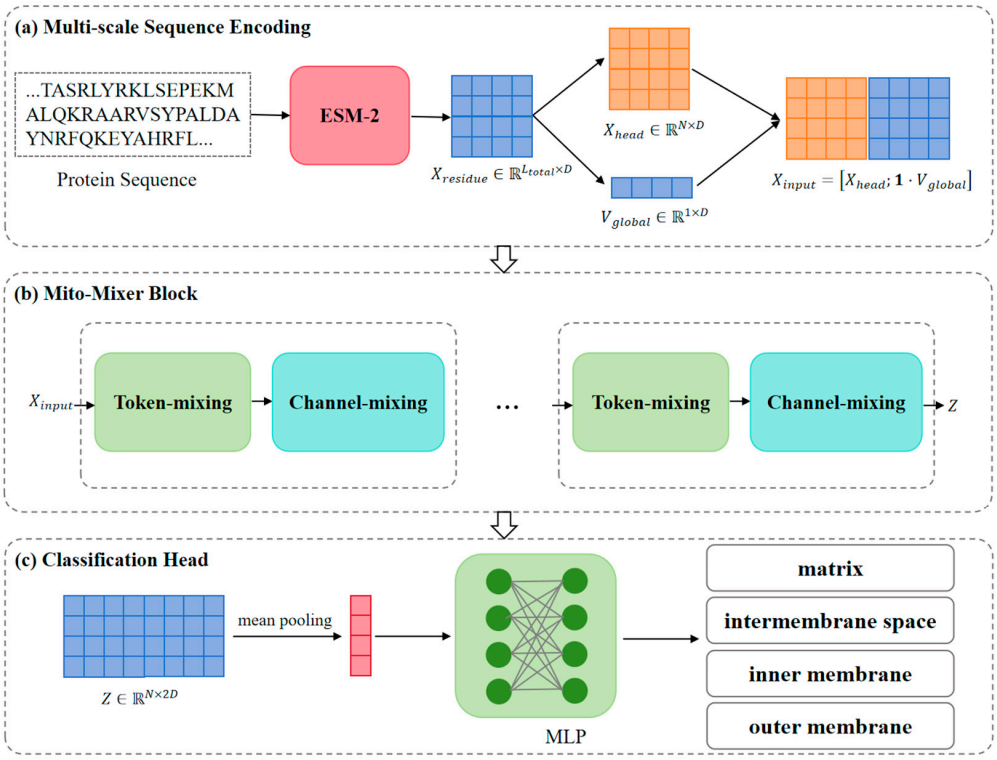

Figure 1. The Mito-Mixer Framework, a multi-scale feature mixing framework for protein sub-mitochondrial localization prediction. The hierarchical structure consists of three primary stages: (**a**) Multi-scale Sequence Encoding. The framework extracts deep residue-level representations ($X_{residue} \in \mathbb{R}^{L_{total} \times D}$) using pre-trained ESM-2 embeddings to preserve positional integrity. To combat signal dilution, a multi-scale input (X_{input}) is constructed by concatenating the localized N-terminal head-features (X_{head}, covering the first N residues) with a globally pooled context vector (V_{global}) that captures the protein's overall evolutionary identity. (**b**) The Mito-Mixer Block. This core module alternates between two MLP-based mixing strategies. The Token-mixing layer transposes the input matrix to scan across the positional dimension (N), effectively capturing the "spatial melody" and long-range dependencies of targeting motifs. Subsequently, the Channel-mixing layer operates on the embedding dimension (D) to disentangle and refine biochemical features sensitive to specific sub-mitochondrial environments. (**c**) Classification Head. The refined features (Z) are aggregated via global pooling and passed through an MLP classifier.

2.2.1. Multi-Scale Sequence Encoding

The input to our framework is a primary amino acid sequence of length L_{total}. We utilize the pre-trained ESM-2 (Evolutionary Scale Modeling) [32] to extract deep residue-level representations. For a given sequence, we extract the hidden states from the final layer of ESM-2, resulting in an embedding matrix:

$$X_{residue} \in \mathbb{R}^{L_{total} \times D} \tag{1}$$

where D denotes the embedding dimension (e.g., $D = 1280$ for ESM2-650M). Unlike traditional methods that immediately compress this matrix, we maintain the positional resolution to preserve the integrity of the "sorting code."

To specifically combat signal dilution, we construct a multi-scale input $X_{input} \in \mathbb{R}^{N \times 2D}$ by combining localized N-terminal features with a global sequence context:

1. Head-Feature ($X_{head} \in \mathbb{R}^{N \times D}$): We truncate the first N residues (e.g., $N = 80$) of the sequence to isolate the MTS. For sequences shorter than N, zero-padding is applied.

2. Global-Context ($V_{global} \in \mathbb{R}^{1 \times D}$): We perform global average pooling on the full-length $X_{residue}$ to capture the protein's overall evolutionary and structural identity.

3. Integration: The global vector is expanded and concatenated with the local features to form the final mixing input:

$$X_{input} = \left[X_{head}; \mathbf{1} \cdot V_{global}\right] \tag{2}$$

To fuse multi-scale information, we broadcast the global semantic vector $V_{global} \in \mathbb{R}^{1 \times D}$ by multiplying it with a column vector of ones $\mathbf{1} \in \mathbb{R}^{N \times 1}$, yielding a contextual matrix $\mathbf{1}V_{global} \in \mathbb{R}^{N \times D}$. This matrix is concatenated with the N-terminal features X_{head} to ensure that every local residue embedding is augmented with the global evolutionary context of the full sequence. This ensures that the model evaluates local motifs against the backdrop of the protein's global characteristics.

2.2.2. The Mito-Mixer Block

The heart of the framework consists of a series of Mito-Mixer Blocks. Each block employs two specialized, alternating MLP layers, Token-mixing and Channel-mixing, to refine the input features.

Token-Mixing: Spatial Rhythmic Extraction

The Token-mixing layer is designed to capture the spatial patterns of the amino acids. In mitochondria, targeting signals are often defined not by a specific residue, but by a "spatial melody," such as the amphipathic α-helix where positively charged and hydrophobic residues appear at specific intervals. By transposing the input matrix, we allow the MLP to operate across the position dimension. This enables the model to learn how a residue at position i interacts with a residue at position j to form a functional motif. The operation includes a residual connection to prevent gradient degradation:

$$Y = X_{input} + (W_2 \sigma\left(W_1 \cdot LayerNorm\left(X_{input}\right)^T\right))^T \tag{3}$$

where W_1 and W_2 are learnable weights that "scan" the sequence for targeting patterns. This allows the model to capture global dependencies across the entire N-terminal region, surpassing the local-view limitations of traditional CNNs.

Channel-Mixing: Biochemical Refinement

While Token-mixing addresses "where" the signal is, Channel-mixing addresses "what" the signal represents. The D-dimensional features provided by ESM-2 contain a mixture of structural, evolutionary, and physicochemical information. Channel-mixing

re-organizes these features to filter out noise and highlight dimensions sensitive to submitochondrial environments (e.g., high hydrophobicity for the inner membrane). The operation acts independently on each residue position:

$$Z = Y + W_4(\sigma(W_3 \cdot LayerNorm(Y))) \tag{4}$$

where W_3 projects the features into a higher-dimensional space (determined by the Channel MLP expansion ratio, as optimized in Section 3.1) to disentangle overlapping biological signals, and W_4 compresses the refined "biochemical essence" back to the original dimension.

2.2.3. Classification Head

To optimize the final classification and enhance model robustness against data imbalance and biological ambiguity, the refined representation Z is aggregated via a global mean pooling layer and processed by an MLP classification head. We employ two optimization approaches: Weighted Cross-Entropy is integrated to assign higher penalty coefficients to minority classes, such as the intermembrane space, thereby preventing the model from developing a bias toward numerically dominant compartments like the matrix or inner membrane. Simultaneously, Label Smoothing is applied to soften the target distributions; this acknowledges the biological fluidity of proteins transitioning between sub-organellar compartments and prevents the model from over-fitting to "hard" labels, which ultimately improves its generalization capabilities on independent, large-scale proteomic data.

2.3. Performance Evaluation Metrics

To rigorously evaluate the classification performance of Mito-Mixer, we utilize a multi-class confusion matrix M, where each element $M_{i,j}$ represents the number of proteins belonging to actual class i that were predicted to be in class j. Based on this matrix, we employ two primary metrics to assess both class-specific and global model performance.

The Matthews' Correlation Coefficient (MCC) is used to score single-class predictions. For each submitochondrial category k, the $\mathrm{MCC}(k)$ is calculated to account for the balance between true positives, true negatives, false positives, and false negatives. It is defined as:

$$\mathrm{MCC}(k) = \frac{M_{k,k}n_k - o_k u_k}{\sqrt{(M_{k,k} + o_k)(M_{k,k} + u_k)(n_k + o_k)(n_k + u_k)}} \tag{5}$$

where the variables are defined as follows: $o_k = \sum_{i \neq k} M_{i,k}$ represents the number of over-predictions (false positives) for class k; $u_k = \sum_{i \neq k} M_{k,i}$ represents the number of under-predictions (false negatives) for class k; $n_k = \sum_{i \neq k} \sum_{j \neq k} M_{i,j}$ represents the number of proteins correctly predicted as not belonging to class k (true negatives).

To provide a unified single measure for multi-class classification, we adopt the Generalized Correlation Coefficient (GCC) [33]. First, we determine the number of proteins truly in class $k(a_k)$ and the number of proteins predicted to be in class $k(b_k)$:

$$a_k = \sum_{i=1}^{K} M_{k,i}, \; b_k = \sum_{i=1}^{K} M_{i,k} \tag{6}$$

We then define an expected matrix e, where each element $e_{i,j}$ represents the expected number of proteins in a random distribution:

$$e_{i,j} = \frac{a_i b_j}{N} \tag{7}$$

where $N = \sum_{i=1}^{K} \sum_{j=1}^{K} M_{i,j}$ is the total number of proteins in the dataset. The global GCC is then defined as:

$$\mathrm{GCC} = \sqrt{\frac{\sum_{i=1}^{K} \sum_{j=1}^{K} \frac{(M_{i,j} - e_{i,j})^2}{e_{i,j}}}{N(K-1)}} \tag{8}$$

The GCC value ranges from −1 to 1. A GCC of 1 indicates perfect prediction, while a value of 0 corresponds to performance no better than random guessing.

2.4. Training and Validation Paradigm

To rigorously train the Mito-Mixer framework and prevent overfitting, we employed a stratified 5-fold cross-validation strategy for the benchmark datasets. This stratification is crucial to ensure that the highly imbalanced class proportions, particularly the sparse intermembrane space proteins, are uniformly distributed across all folds. Within each fold, the dataset was split into 80% for training and 20% for validation. The model was trained using the AdamW optimizer. To ensure optimal convergence and computational efficiency, an early stopping mechanism was implemented: training was halted if the validation loss failed to improve for 20 consecutive epochs, up to a maximum of 50 epochs.

3. Results and Discussion

3.1. Hyperparameter Optimization

To identify the optimal architectural and training configuration for Mito-Mixer, we conducted an automated hyperparameter search using Optuna version 4.7.0 [34], a state-of-the-art optimization framework. The search process utilized Bayesian Optimization via the Tree-structured Parzen Estimator (TPE) algorithm, which intelligently explores the search space by balancing exploration of new parameter combinations with exploitation of previously identified high-performance regions. The Hyperparameter Optimization (HPO) space was carefully defined to cover structural components of the Mixer blocks, the classification head, and various optimization settings. The specific search ranges are detailed in Table 2.

Table 2. Search space and configuration of hyperparameters for the Mito-Mixer optimization.

Category	Hyperparameter	Search Range/Values	Type
Mixer Structure	Mixer Blocks	(2, 6)	Integer
	Token MLP Ratio *	(2.0, 6.0)	Float
	Channel MLP Ratio *	(2.0, 6.0)	Float
	Dropout	(0.05, 0.35)	Float
Classifier Head	Hidden Units	(256, 1024)	Int (Step 64)
	Head Dropout	(0.10, 0.55)	Float
Optimization	Learning Rate (lr)	$(1 \times 10^{-5}, 5 \times 10^{-4})$	Log-uniform
	Weight Decay	$(1 \times 10^{-6}, 1 \times 10^{-2})$	Log-uniform
	Label Smoothing	(0.0, 0.15)	Float
	Batch Size	{4, 8, 16}	Categorical

* Note: The Token and Channel MLP Ratios represent the expansion factor of the hidden layer dimensions within their respective MLP blocks relative to the input feature dimension (D). For example, a Channel MLP Ratio of 4.0 projects the D-dimensional input to a hidden size of $4D$ before compressing it back to D.

During the HPO process, the N-terminal truncation length (N) was fixed at 80 residues. We explicitly excluded N from the Optuna search space to prevent conflating the optimization of network capacity with the biological boundaries of targeting signals. A dedicated parametric sweep of N was subsequently performed (detailed in Section 3.2).

The optimization was implemented using PyTorch version 2.1.2 and executed on an NVIDIA GeForce RTX 4090 GPU. To ensure the robustness and generalizability of the selected parameters, we employed a stratified 5-fold cross-validation strategy. This stratification guarantees that the highly imbalanced class proportions (e.g., the sparse intermembrane space proteins) are preserved across all folds. During each trial, the model was trained for a maximum of 50 epochs. To prevent over-fitting and reduce unneces-

sary computational expenditure, an early stopping mechanism was implemented, which terminated the training if the validation loss failed to improve for 20 consecutive epochs. The objective function for Optuna was defined as the maximization of the average Generalized Correlation Coefficient (GCC) across all five folds, ensuring that the final model configuration excels in global multi-class classification consistency.

Following the completion of 50 optimization trials, the best-performing configurations were identified. While the search space was shared, the optimal parameters converged to slightly different values for the SM424-18 and SubMitoPred datasets, reflecting their differing scales and sequence diversities. The final selected parameters, which were used to generate the benchmark results in the following sections, are presented in Table 3.

Table 3. Final optimized hyperparameters for the Mito-Mixer model. The optimal values were determined using Bayesian Optimization (TPE algorithm) via the Optuna framework, as detailed in Section 3.1.

Category	Hyperparameter	SM424-18 Value	SubMitoPred Value
Mixer Structure	Mixer Blocks	2	2
	Token MLP Ratio	5.9498	4.3691
	Channel MLP Ratio	2.7094	4.2464
	Dropout	0.2477	0.3448
Classifier Head	Hidden Units	384	448
	Head Dropout	0.4049	0.3539
Optimization	Learning Rate (lr)	2.7502×10^{-5}	6.3089×10^{-5}
	Weight Decay	0.0016	9.0343×10^{-5}
	Label Smoothing	0.0683	0.0855
	Batch Size	4	16

3.2. Analysis of N-Terminal Truncation Length (N)

Since Mitochondrial Targeting Signals (MTS) are typically located at the N-terminus, the choice of N-terminal Truncation Length (N) represents a trade-off between capturing sufficient signaling information and avoiding the inclusion of excessive "noise" from the mature protein region. To investigate this relationship, we conducted a sensitivity analysis by varying N from 20 to 200 residues while keeping all other hyperparameters constant. This parameter sweep for N was systematically evaluated using the same stratified 5-fold cross-validation protocol employed during the main hyperparameter optimization, ensuring that the selected lengths are robust and not artifacts of a specific data split. The results across both benchmark datasets exhibit a distinct non-linear relationship between sequence length and predictive performance.

In the SM424-18 dataset, the model performance exhibits significant sensitivity to the truncation length. As illustrated in Figure 2, starting from a baseline GCC of approximately 0.685 at $N = 20$, the performance climbs steadily as more sequence context is included. The model reaches its global maximum performance at $N = 80$, achieving a GCC of 0.7443. This suggests that for this specific dataset, 80 residues are sufficient to encompass the majority of mitochondrial targeting presequences (MTS) while maintaining a high signal-to-noise ratio. Beyond $N = 80$, the GCC fluctuates and generally trends downward, dipping as low as 0.696 at $N = 190$. This decline validates our hypothesis that including excessive residues from the mature protein can dilute the specialized targeting features, making it harder for the Mixer blocks to extract relevant localization cues.

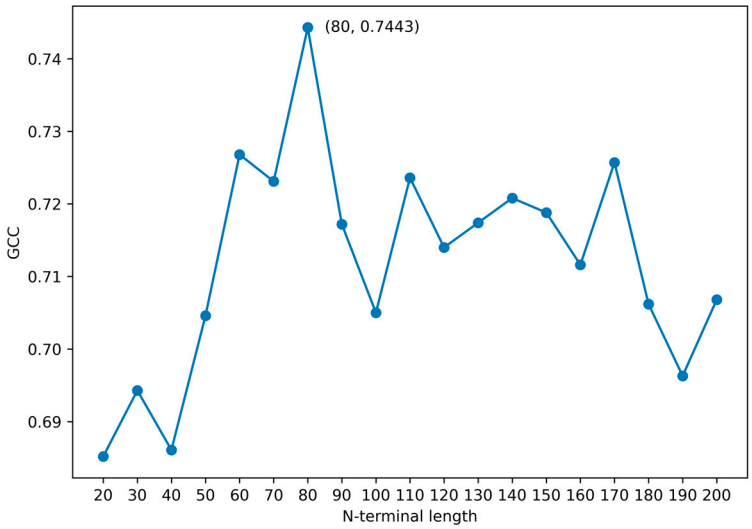

Figure 2. Sensitivity Analysis of the Mito-Mixer Framework to N-terminal Truncation Length (N) on the SM424-18 Benchmark Dataset. Demonstrating the Correlation Between Truncation Length (N) and Model Performance with a Global Maximum GCC of 0.7443 Achieved at $N = 80$.

The SubMitoPred dataset demonstrates a similar overall trend but favors a significantly longer truncation length to reach peak accuracy. As shown in Figure 3, unlike SM424-18, the performance on the SubMitoPred dataset shows a more prolonged upward trajectory, maintaining a GCC above 0.730 for most intervals. The optimal performance is achieved at $N = 170$, with a GCC of 0.7878. The requirement for a longer sequence length here likely reflects the presence of more complex or distal targeting signals within the SubMitoPred protein distribution, such as those found in certain inner membrane or intermembrane space proteins that rely on internal motifs rather than just N-terminal presequences. Similar to the first dataset, performance begins to drop once the length exceeds the optimal threshold, falling toward 0.749 at $N = 200$.

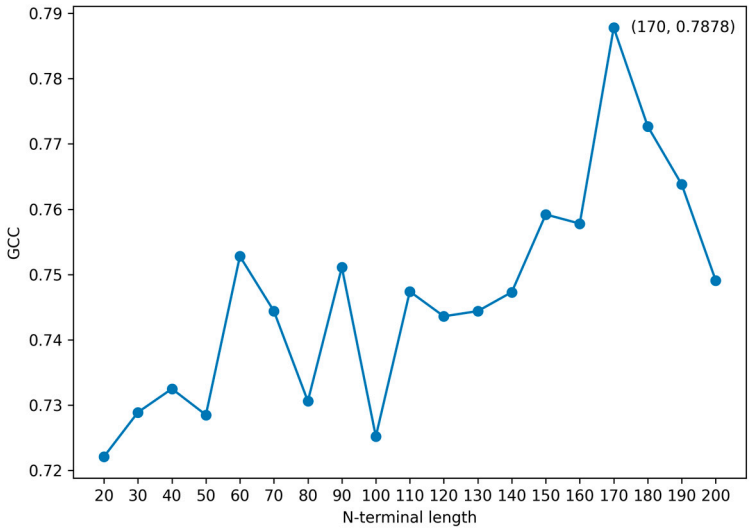

Figure 3. Sensitivity Analysis of the Mito-Mixer Framework to N-terminal Truncation Length (N) on the SubMitoPred Benchmark Dataset. Demonstrating the Correlation Between Truncation Length (N) and Model Performance with a Global Maximum GCC of 0.7878 Achieved at $N = 170$.

3.3. Performance on Benchmark Datasets

We evaluated the performance of Mito-Mixer on two widely used benchmark datasets: SM424-18 and SubMitoPred. To provide a rigorous assessment, we compared our model against several state-of-the-art (SOTA) predictors, including iDeepSubMito, DeepMito,

and SubMito. The performance was quantified using the class-specific Matthews' Correlation Coefficient (MCC) for the four sub-compartments (matrix, inner membrane, outer membrane, and intermembrane space) and the Generalized Correlation Coefficient (GCC) for overall classification consistency. The performance metrics reported for Mito-Mixer in Tables 4 and 5 represent the mean evaluation scores obtained from the stratified 5-fold cross-validation process. This methodology aligns with standard practices for evaluating models on these specific benchmarks, ensuring fair comparison against SOTA predictors.

Table 4. Performance comparison on the SM424-18 dataset.

Method	MCC(M) [a]	MCC(I) [a]	MCC(O) [a]	MCC(T) [a]	GCC [b]
DeepMito	0.65	0.47	0.46	0.53	0.54
iDeepSubMito	0.6825	0.5821	0.6583	0.6498	0.6803
Mito-Mixer (Ours)	0.8212	0.7011	0.7479	0.6870	0.7443

[a] MCC (M, I, O, T): Matthews Correlation Coefficient of Matrix, Inner membrane, Outer membrane and Intermembrane space localization, respectively. [b] GCC: Generalized Correlation Coefficient (Equation (8)).

Table 5. Performance comparison on the SubMitoPred dataset.

Method	MCC(M) [a]	MCC(I) [a]	MCC(O) [a]	MCC(T) [a]	GCC [b]
DeepMito	0.76	0.60	0.42	0.46	-
iDeepSubMito	0.7335	0.6477	0.6995	0.5494	0.6783
Mito-Mixer (Ours)	0.7886	0.7682	0.8404	0.7298	0.7878

[a] MCC (M, I, O, T): Matthews Correlation Coefficient of Matrix, Inner membrane, Outer membrane and Intermembrane space localization, respectively. [b] GCC: Generalized Correlation Coefficient (Equation (8)).

The SM424-18 dataset serves as a rigorous benchmark due to its high sequence diversity and the inherent difficulty in distinguishing between the four submitochondrial compartments. As demonstrated in Table 4, Mito-Mixer achieves a superior performance profile across all evaluation metrics compared to established deep learning architectures such as DeepMito and iDeepSubMito.

Specifically, our model achieves a GCC of 0.7443, representing a significant improvement over iDeepSubMito (0.6803) and DeepMito (0.54). This global metric underscores Mito-Mixer's ability to maintain high classification consistency across the entire dataset. Mito-Mixer reaches an MCC(M) of 0.8212 and an MCC(I) of 0.7011, outperforming iDeepSubMito by approximately 14% and 12%, respectively. This gain is largely attributed to the Channel-mixing module's ability to refine biochemical features, such as the high hydrophobicity characteristic of the inner membrane proteins. Our method records an MCC(O) of 0.7479, a marked increase over the 0.6583 achieved by the previous state-of-the-art iDeepSubMito. For the historically challenging intermembrane space category, Mito-Mixer achieves an MCC(T) of 0.6870. This is a substantial leap from DeepMito's 0.53, validating our strategy of using Token-mixing to preserve and amplify short, high-resolution N-terminal signals (such as cysteine-rich motifs) that are otherwise lost in global pooling methods. Specifically, Token-mixing is highly adept at capturing the strict spatial spacing of cysteine residues, such as the twin CX9C or CX3C motifs, which are the essential hallmarks recognized by the MIA (Mitochondrial Intermembrane space Assembly) pathway for IMS import.

To further validate the robustness and generalization of the Mito-Mixer framework, we evaluated its performance on the SubMitoPred dataset. This dataset provides an alternative distribution of mitochondrial proteins, allowing for a comprehensive comparison against existing high-performance predictors. As summarized in Table 5, Mito-Mixer consistently exceeds the performance of both DeepMito and iDeepSubMito across all metrics.

Mito-Mixer achieved an overall GCC of 0.7878, a substantial improvement over the 0.6783 reported for iDeepSubMito. This gain demonstrates that our model's multi-scale

feature integration is highly effective across different data sources. The class-specific analysis via MCC reveals several critical insights. The model shows exceptional proficiency in identifying membrane-associated proteins, achieving an MCC(I) of 0.7682 and a remarkable MCC(O) of 0.8404. This represents an improvement of approximately 12% and 14% over iDeepSubMito, respectively. The significantly higher scores in these categories reinforce our hypothesis that the Channel-mixing module successfully distills biochemical features like hydrophobicity and transmembrane helix propensity from the ESM-2 embeddings. For matrix proteins, Mito-Mixer recorded an MCC(M) of 0.7886. Compared to DeepMito (0.76) and iDeepSubMito (0.7335), our model more accurately captures the traditional N-terminal presequences that guide proteins to the mitochondrial interior. Notably, Mito-Mixer achieved an MCC(T) of 0.7298, which is the highest recorded for this difficult-to-predict compartment (Intermembrane Space). This is a 33% relative improvement over iDeepSubMito (0.5494) and significantly higher than DeepMito (0.46). This result confirms that our Token-mixing approach is uniquely capable of identifying the subtle, non-canonical signals, such as cysteine-rich motifs, that characterize intermembrane space-bound proteins.

3.4. Generalization on Independent Test Set (M983)

To evaluate the practical utility and robustness of Mito-Mixer, we conducted an independent generalization test using the M983 dataset. For this generalization experiment, we utilized the Mito-Mixer model that was fully trained and optimized on the SubMitoPred benchmark dataset. This model was subsequently deployed to predict the localization of the entirely unseen sequences in the M983 dataset. This is a critical measure of whether the model has learned general biological principles of protein sorting rather than merely over-fitting to specific dataset biases.

The results, detailed in Table 6, demonstrate that Mito-Mixer maintains a high level of predictive accuracy even when confronted with diverse sequences from the M983 dataset. Mito-Mixer achieved a GCC of 0.6945, significantly outperforming iDeepSubMito (0.6321) and DeepMito (0.5558). This result confirms that the hierarchical mixing architecture effectively captures evolutionary and structural signatures that remain consistent across different proteomic sources.

Table 6. Generalization comparison on the independent test set M983.

Method	MCC(M) [a]	MCC(I) [a]	MCC(O) [a]	MCC(T) [a]	GCC [b]
DeepMito	0.6297	0.6024	0.7366	0.0000	0.5558
iDeepSubMito	0.7358	0.7181	0.8146	0.0000	0.6321
Mito-Mixer (Ours)	0.7656	0.7713	0.9302	0.0000	0.6945

[a] MCC (M, I, O, T): Matthews Correlation Coefficient of Matrix, Inner membrane, Outer membrane and Intermembrane space localization, respectively. [b] GCC: Generalized Correlation Coefficient (Equation (8)).

Specifically, our model achieved a near-perfect MCC(O) of 0.9302, a substantial leap over previous methods. This suggests that the Channel-mixing component is highly reliable at identifying the specific biochemical signatures of outer membrane-resident proteins, such as beta-barrel structures or unique hydrophobic anchors. The model maintained strong performance in the matrix and inner membrane categories, with MCC(M) of 0.7656 and MCC(I) of 0.7713, respectively. These scores indicate that the multi-scale integration of N-terminal "Head" features and global context allows the model to generalize the complex "MTS plus transmembrane domain" logic required for these compartments. A notable observation in Table 6 is the 0.0000 MCC(T) recorded across all models, including Mito-Mixer. This result is not a reflection of model failure, but rather a characteristic of the M983 independent test set. Unlike the benchmark datasets used for training, the M983 dataset does not contain any proteins categorized under the intermembrane space. Despite the

reduced number of target classes, Mito-Mixer's superior GCC of 0.6945, compared to 0.6321 for iDeepSubMito, demonstrates that our model remains the most robust architecture for identifying the primary mitochondrial sub-compartments in large-scale, real-world proteomic data. We explicitly acknowledge this as a limitation of the current generalization test, as the model's robust capacity to identify novel IMS proteins cannot be externally validated on this specific independent dataset.

To provide deeper visual insight into the performance metrics, Figure 4 presents the 4-class confusion matrix generated from the independent M983 test set. A detailed analysis of the diagonal elements (true positives) and off-diagonal elements (misclassifications) reveals where the model excels and where it encounters biological ambiguity. The model demonstrates exceptional accuracy in identifying Matrix proteins (with 173 out of 177 correctly classified) and Outer Membrane proteins (133 out of 145). The most prominent source of misclassification originates from the Inner Membrane category, where 77 proteins were incorrectly predicted as Matrix residents and 31 as Intermembrane Space proteins. This specific misclassification pattern is highly consistent with mitochondrial biology: many Inner Membrane proteins share classical N-terminal targeting presequences with Matrix proteins (as both are initially imported via the TOM/TIM23 translocase complexes) and rely on secondary, downstream hydrophobic "stop-transfer" signals for membrane anchoring. It is inherently challenging for purely sequence-based models to perfectly disambiguate these compartments when secondary signals are subtle or atypical. Furthermore, as previously noted, the true label row for the Intermembrane Space is entirely zero due to the M983 dataset's specific composition. Nevertheless, the matrix confirms that the model maintained stable control over false positive predictions for this class (only 35 total false assignments out of 983 samples), which was crucial for preserving the high overall Generalized Correlation Coefficient (GCC).

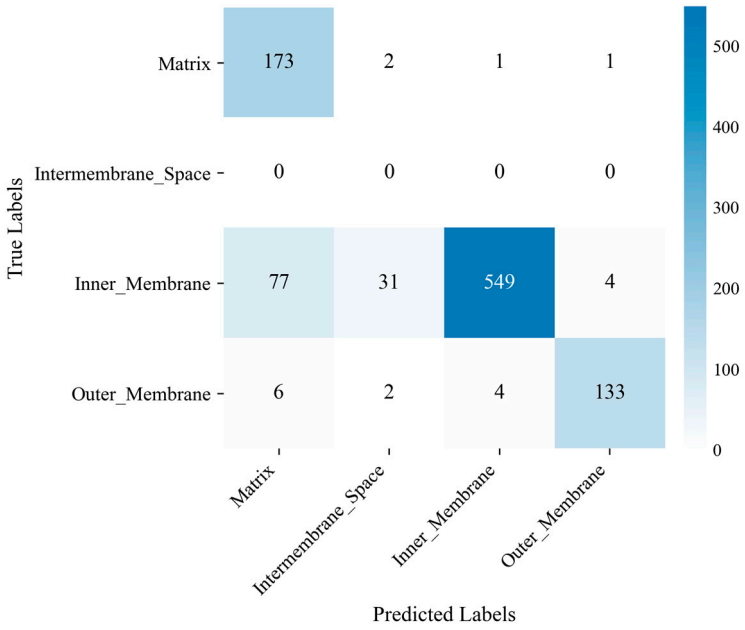

Figure 4. 4-class Confusion Matrix for the Mito-Mixer model evaluated on the independent M983 test set.

3.5. Ablation Study

To systematically evaluate the contribution of each core component in Mito-Mixer, we conducted a series of ablation experiments across two benchmark datasets: SM424-18 and SubMitoPred. These experiments were designed to disentangle the benefits of our multi-scale input strategy from the architectural innovations of the Mixer block. To ensure

statistical consistency, all ablation variants were evaluated using the same stratified 5-fold cross-validation protocol employed for the primary model evaluation. The results are summarized in Table 7.

Table 7. Ablation study results (GCC) on the SM424-18 dataset and the SubMitoPred dataset.

Dataset	Experiment	Configuration	GCC [a]
SM424-18	Mito-Mixer	Full Model	0.7443
	Exp A	No Mixer	0.7143
	Exp B	No N-term Slicing	0.6911
	Exp C	Token-only	0.7251
	Exp C	Channel-only	0.7121
SubMitoPred	Mito-Mixer	Full Model	0.7878
	Exp A	No Mixer	0.7380
	Exp B	No N-term Slicing	0.7536
	Exp C	Token-only	0.7403
	Exp C	Channel-only	0.7386

[a] GCC: Generalized Correlation Coefficient (Equation (8)).

The first two experiments address the "signal dilution" problem by testing the processing mechanism and the input scope respectively.

Exp A (No Mixer—Architecture Ablation): We replaced the Mixer blocks with a standard Global Average Pooling layer applied to the N-terminal slice. While this model still utilizes the N-terminal information, it lacks the non-linear interaction capacity of the Mixer blocks. In SM424-18, the GCC dropped from 0.7443 to 0.7143, and in SubMitoPred, it fell from 0.7878 to 0.7380. This indicates that simply isolating the N-terminus is insufficient; the model requires the Mixer's ability to decode complex spatial interdependencies to achieve optimal accuracy.

Exp B (No N-terminal Slicing—Input Ablation): We tested the necessity of explicit N-terminal focus by feeding the full protein sequence directly into the Mixer blocks. Despite the powerful modeling capability of the Mixer architecture, the GCC declined to 0.6911 on SM424-18 and 0.7536 on SubMitoPred. This confirms that in long sequences, localized targeting signals are mathematically "diluted" by the massive amount of mature protein data, proving that the specialized $\left[X_{head}; \mathbf{1} \cdot V_{global}\right]$ strategy is essential to preserve high-resolution targeting codes.

We further investigated the individual contributions of the two fundamental mixing mechanisms within the Mixer block. The results reveal that neither dimension alone is sufficient to maintain the model's predictive power.

Token-only Variant: This configuration focuses purely on spatial correlations and positional motifs. The GCC showed a clear decrease to 0.7251 on SM424-18 and 0.7403 on SubMitoPred. While this variant can still recognize some "rhythmic" arrangements like amphipathic α-helices, the performance gap compared to the full model highlights that spatial patterns alone cannot accurately resolve sub-compartment identities without the refined biochemical context provided by channel interaction.

Channel-only Variant: This version treats each residue as an independent biochemical entity, ignoring its position. This led to a further decline in performance, with GCC values falling to 0.7121 and 0.7386, respectively. The drop in consistency proves that treating residues as isolated chemical units, ignoring their sequential context, severely handicaps the model's ability to recognize structural motifs, such as the cysteine-rich patterns essential for the intermembrane space pathway.

Overall, the full Mito-Mixer architecture consistently outperforms all ablated versions across both datasets. This proves that state-of-the-art accuracy in protein submitochondrial localization requires a synergistic integration of multi-scale spatial focus and dual-

dimensional feature mixing, as the "sorting code" is composed of both spatial arrangements and specific biochemical features.

3.6. Computational Case Study: Disentangling Bipartite Targeting Signals

To biologically validate our hypothesis regarding signal dilution, we analyzed the model's processing of proteins with bipartite targeting signals, such as Cytochrome c1 (a resident of the inner mitochondrial membrane). Biologically, these proteins rely on a complex N-terminal code: an initial positively charged amphipathic α-helix that directs the protein to the Matrix via the TIM23 complex, immediately followed by a hydrophobic "stop-transfer" sequence that arrests translocation and anchors it in the Inner Membrane.

In our ablation studies, models relying solely on global mean pooling frequently misclassified such proteins as Matrix residents. Mathematically, the broader sequence data diluted the subtle, localized hydrophobic stop-transfer signal. However, Mito-Mixer successfully classifies these proteins into the Inner Membrane. By explicitly truncating the input to the high-resolution N-terminal head and processing it through the Channel-mixing module, the framework effectively isolates and amplifies the biochemical signature (hydrophobicity) of the stop-transfer signal against the background of the initial matrix-targeting sequence. This case study confirms that Mito-Mixer aligns with known biological sorting mechanisms, successfully decoding complex, multi-part targeting motifs that are easily lost in global embeddings.

4. Conclusions

In this study, we introduced Mito-Mixer, a novel multi-scale MLP-Mixer framework specifically engineered to address the "signal dilution" challenge in protein submitochondrial localization. By integrating high-resolution N-terminal features with global evolutionary semantics from the ESM-2 protein language model, Mito-Mixer provides a robust solution for annotating protein positions within the complex mitochondrial structure. Our comprehensive evaluation demonstrates that the model effectively mitigates the dilution of essential sorting codes, a common failure point for traditional global pooling methods. Through the explicit slicing of N-terminal sequences and the specialized use of the Token-mixing module, the framework successfully preserves critical, short-range targeting signals that are often overwhelmed by the larger mature protein body. The experimental results across the SM424-18 and SubMitoPred datasets establish Mito-Mixer as a state-of-the-art tool in the field. Notably, the model achieved a significant leap in identifying the intermembrane space proteins and exhibited high precision in distinguishing between the inner and outer membranes. These improvements in MCC and GCC metrics are further reinforced by the high performance on the M983 independent test set, which confirms that Mito-Mixer has captured fundamental biological principles of protein translocation rather than dataset-specific biases. This robust generalization capability underscores its potential for large-scale proteomic annotation and its utility in advancing our understanding of mitochondrial function and related molecular diseases. Furthermore, our systematic ablation studies revealed that the synergy between Token-mixing for spatial motifs and Channel-mixing for biochemical refinement is the primary driver of the model's accuracy.

Furthermore, while attention-based architectures are theoretically highly expressive, they are intrinsically data-hungry and prone to severe overfitting when applied to small-scale, highly imbalanced datasets like the currently available submitochondrial benchmarks (e.g., comprising only 424 to 570 sequences). Mito-Mixer explicitly mitigates this risk by utilizing a parameter-efficient MLP-Mixer architecture to decode the deep evolutionary semantics already captured by the pre-trained ESM-2 model. This design strikes an optimal

balance between feature expressiveness and model generalization, avoiding the heavy computational overhead and overfitting risks associated with additional Transformer blocks.

While Mito-Mixer demonstrates robust predictive capabilities, the current framework remains constrained by its reliance on purely sequence-based features and the inherent biases of small-scale experimental datasets. Future developments should address these constraints by incorporating 3D structural information (e.g., predicted via AlphaFold) to further refine the model's precision, particularly for minority class proteins.

Author Contributions: Conceptualization, X.W.; Methodology, R.W. and X.W.; Supervision, X.W.; Writing—original draft, R.W. and M.W.; Writing—review and editing, X.W. and R.W.; Software, Y.W. and L.Y.; Validation, Y.W. and L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by funds from the Key Research Project of Colleges and Universities of Henan Province (No. 22A520013, No. 23B520004), the Key Science and Technology Development Program of Henan Province (No. 232102210020, No. 252102210137).

Data Availability Statement: The source codes and data for Mito-Mixer are available at https://github.com/xwanggroup/Mito-Mixer (accessed on 5 February 2026).

Conflicts of Interest: The authors declare there are no conflicts of interest.

References

1. Wiedemann, N.; Pfanner, N. Mitochondrial machineries for protein import and assembly. *Annu. Rev. Biochem.* **2017**, *86*, 685–714. [CrossRef]
2. West, A.P.; Shadel, G.S.; Ghosh, S. Mitochondria in innate immune responses. *Nat. Rev. Immunol.* **2011**, *11*, 389–402. [CrossRef]
3. Chacinska, A.; Koehler, C.M.; Milenkovic, D.; Lithgow, T.; Pfanner, N. Importing mitochondrial proteins: Machineries and mechanisms. *Cell* **2009**, *138*, 628–644. [CrossRef]
4. Pfanner, N.; Warscheid, B.; Wiedemann, N. Mitochondrial proteins: From biogenesis to functional networks. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 267–284. [CrossRef] [PubMed]
5. Suomalainen, A.; Battersby, B.J. Mitochondrial diseases: The contribution of organelle stress responses to pathology. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 77–92. [CrossRef] [PubMed]
6. Huang, W.-L.; Tung, C.-W.; Ho, S.-W.; Hwang, S.-F.; Ho, S.-Y. ProLoc-GO: Utilizing informative Gene Ontology terms for sequence-based prediction of protein subcellular localization. *BMC Bioinform.* **2008**, *9*, 80. [CrossRef]
7. Huang, W.-L.; Tung, C.-W.; Huang, H.-L.; Ho, S.-Y. Predicting protein subnuclear localization using GO-amino-acid composition features. *Biosystems* **2009**, *98*, 73–79. [CrossRef] [PubMed]
8. Kumar, R.; Jain, S.; Kumari, B.; Kumar, M. Protein sub-nuclear localization prediction using svm and pfam domain information. *PLoS ONE* **2014**, *9*, e98345. [CrossRef]
9. Wang, S.; Liu, S. Protein Sub-Nuclear Localization Based on Effective Fusion Representations and Dimension Reduction Algorithm LDA. *Int. J. Mol. Sci.* **2015**, *16*, 30343–30361. [CrossRef]
10. Wang, S.; Yue, Y. Protein subnuclear localization based on a new effective representation and intelligent kernel linear discriminant analysis by dichotomous greedy genetic algorithm. *PLoS ONE* **2018**, *13*, e0195636. [CrossRef]
11. Littmann, M.; Goldberg, T.; Seitz, S.; Bodén, M.; Rost, B. Detailed prediction of protein sub-nuclear localization. *BMC Bioinform.* **2019**, *20*, 205.
12. Wang, X.; Zhang, W.; Zhang, Q.; Li, G.-Z. MultiP-SChlo: Multi-label protein subchloroplast localization prediction with Chou's pseudo amino acid composition and a novel multi-label classifier. *Bioinformatics* **2015**, *31*, 2639–2645. [CrossRef] [PubMed]
13. Wan, S.; Mak, M.-W.; Kung, S.-Y. Ensemble Linear Neighborhood Propagation for Predicting Subchloroplast Localization of Multi-Location Proteins. *J. Proteome Res.* **2016**, *15*, 4755–4762. [CrossRef] [PubMed]
14. Wan, S.; Mak, M.-W.; Kung, S.-Y. Transductive Learning for Multi-Label Protein Subchloroplast Localization Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *14*, 212–224. [CrossRef]
15. Bankapur, S.; Patil, N. An Effective Multi-Label Protein Sub-Chloroplast Localization Prediction by Skipped-Grams of Evolutionary Profiles Using Deep Neural Network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 1449–1458. [CrossRef]
16. Wang, X.; Han, L.; Wang, R.; Chen, H. DaDL-SChlo: Protein subchloroplast localization prediction based on generative adversarial networks and pre-trained protein language model. *Brief. Bioinform.* **2023**, *24*, bbad083. [CrossRef]

17. Du, P.; Li, Y. Prediction of protein submitochondria locations by hybridizing pseudo-amino acid composition with various physicochemical features of segmented sequence. *BMC Bioinform.* **2006**, *7*, 518. [CrossRef]
18. Savojardo, C.; Bruciaferri, N.; Tartari, G.; Martelli, P.L.; Casadio, R. DeepMito: Accurate prediction of protein sub-mitochondrial localization using convolutional neural networks. *Bioinformatics* **2020**, *36*, 56–64. [CrossRef]
19. Hou, Z.; Yang, Y.; Li, H.; Wong, K.-C.; Li, X. iDeepSubMito: Identification of protein submitochondrial localization with deep learning. *Brief. Bioinform.* **2021**, *22*, bbab288. [CrossRef] [PubMed]
20. Yu, B.; Qiu, W.; Chen, C.; Ma, A.; Jiang, J.; Zhou, H.; Ma, Q. SubMito-XGBoost: Predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting. *Bioinformatics* **2020**, *36*, 1074–1081. [CrossRef]
21. Mei, S. Multi-kernel transfer learning based on Chou's PseAAC formulation for protein submitochondria localization. *J. Theor. Biol.* **2012**, *21*, 121–130. [CrossRef] [PubMed]
22. Deng, J.; Xu, W.; Jie, Y.; Chong, Y. Subcellular localization and relevant mechanisms of human cancer-related micropeptides. *FASEB J.* **2023**, *37*, e23270. [CrossRef]
23. Emanuelsson, O.; Brunak, S.; Von Heijne, G.; Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2007**, *2*, 953–971. [CrossRef]
24. Armenteros, J.J.A.; Salvatore, M.; Emanuelsson, O.; Winther, O.; von Heijne, G.; Elofsson, A.; Nielsen, H. Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2019**, *2*, e201900429. [CrossRef]
25. Du, P.; Yu, Y. SubMito-PSPCP: Predicting Protein Submitochondrial Locations by Hybridizing Positional Specific Physicochemical Properties with Pseudoamino Acid Compositions. *BioMed Res. Int.* **2013**, *2013*, 263829. [CrossRef]
26. Savojardo, C.; Martelli, P.L.; Fariselli, P.; Profiti, G.; Casadio, R. BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* **2018**, *46*, W459–W466. [CrossRef]
27. Wang, X.; Jin, Y.; Zhang, Q. Deeppred-submito: A novel submitochondrial localization predictor based on multi-channel convolutional neural network and dataset balancing treatment. *Int. J. Mol. Sci.* **2020**, *21*, 5710. [CrossRef]
28. Jiang, Y.; Wang, D.; Yao, Y.; Eubel, H.; Künzler, P.; Møller, I.M.; Xu, D. MULocDeep: A deep-learning framework for protein subcellular and suborganellar localization prediction with residue-level interpretation. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4825–4839. [CrossRef]
29. Tolstikhin, I.O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. MLP-Mixer: An all-mlp architecture for vision. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24261–24272.
30. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]
31. Kumar, R.; Kumari, B.; Kumar, M. Proteome-wide prediction and annotation of mitochondrial and sub-mitochondrial proteins by incorporating domain information. *Mitochondrion* **2018**, *42*, 11–22. [CrossRef] [PubMed]
32. Lin, Z.; Akin, H.; Rao, R. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379*, 1123–1130. [CrossRef] [PubMed]
33. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C.A.F.; Nielsen, H. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics* **2000**, *16*, 412–424. [CrossRef] [PubMed]
34. Akiba, T.; Sano, S.; Yanase, T. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.