Contents lists available at YXpublications

International Journal of Applied Mathematics in Control Engineering

Journal homepage: http://www.ijamce.com

From Pre-training to Classification Optimization: Application of NeuroPred-INT Multi-Model Ensemble in Neuropeptide Prediction

Yue Li[a], Ji Qiu[a,*]

[a] College of Electronic Information, Guangxi Minzu University, 530000, China

ARTICLE INFO

Article history:
Received 22 June 2024
Accepted 1 September 2024
Available online 2 September 2024

Keywords:
Neuropeptides
ESM-1b Model
Multiple Sequence Alignment (MSA)
Convolutional Neural Network (CNN)
Deep Earning Ensemble Learning

ABSTRACT

Neuropeptides, small protein-like molecules, play essential roles in cellular communication and modulate physiological processes such as pain, mood, and immune responses. This thesis presents a neuropeptide classification method based on a fine-tuned ESM-1b model. Initially, the ESM-1b model was pre-trained on a neuropeptide-specific dataset, with multiple sequence alignment (MSA) via a Hidden Markov Model (HMM) applied to the training sequences. This generated an output.afa file, which highlighted shared sequence features to enhance model generalization. After fine-tuning, the model was combined with a convolutional neural network (CNN) to extract high-dimensional feature representations that comprehensively characterize the input sequences. These representations were subsequently processed by a gradient boosting tree classifier, which optimized feature weighting and classification, enabling precise differentiation between neuropeptides and other peptides. This multi-step approach leverages the advantages of both deep learning and ensemble learning techniques, enhancing the accuracy and robustness of neuropeptide classification and laying a foundation for deeper insights into their diverse biological roles.

1. Introduction

Neuropeptides are signaling molecules composed of fewer than 100 amino acids that are widely distributed across the nervous and endocrine systems, regulating physiological processes such as mood, pain, immunity, and metabolism [1]. By binding to specific receptors, neuropeptides influence various functions, including appetite control, energy metabolism, and cardiovascular regulation. Given their association with numerous diseases, such as depression and diabetes, studying neuropeptides can aid in early disease diagnosis and support the development of new therapeutic drugs [2]. Recently, deep learning models have enabled more accurate predictions regarding neuropeptide occurrence and function, thereby accelerating research and progress in medical applications [3].

In 2017, Ji QY et al. utilized mass spectrometry for neuropeptide identification, a method that, while effective, requires costly equipment, complex procedures, and skilled personnel for operation and maintenance [4]. In 2021, Md Mehedi Hasan and colleagues introduced the NeuroPred-FRL model, combining diverse encoding techniques and a random forest classifier to improve neuropeptide prediction accuracy [5]. By 2023, Wang Lei et al. had furthered deep learning applications in neuropeptide prediction with NeuroPred-PLM, which integrates a protein language model with a multi-scale convolutional neural network to capture both semantic and local features [6]. Most recently, in 2024, Jian Wen and colleagues released NeuroPpred-SHE, a hybrid model that blends traditional approaches with modern machine learning. Most recently, in 2024, Jian Wen and colleagues released NeuroPpred-SHE, a hybrid model that blends traditional approaches with modern machine learning [7]; however, its high complexity and limited interpretability pose challenges. Also in 2024, Lei Wang's team introduced DeepNeuroPred, which leverages pre-trained language models and convolutional neural networks (CNNs) to accurately predict cleavage sites in neuropeptide precursors [8].

Recent advancements in deep learning have significantly improved neuropeptide prediction, leveraging its ability to model complex patterns within biological sequences. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer architectures have all contributed to more robust feature extraction from sequence data, leading to improved peptide classification accuracy. However, challenges persist. A major limitation remains the scarcity of labeled neuropeptide data, which impedes these models' ability to generalize effectively to new sequences. Furthermore, while deep learning models are skilled at

* Corresponding author.
E-mail addresses: qiuji@gxmzu.edu.cn (J. Qiu)
Digital Object Identifiers: https://doi.org/10.62953/IJAMCE.154866

capturing intricate sequence patterns, they often lack interpretability, making it difficult to determine which sequence regions most influence the predictions. This gap in interpretability restricts the utility of such models in biological research, where identifying specific contributing sequence elements is essential. Additionally, traditional deep learning approaches may overlook the multi-scale and hierarchical nature of neuropeptide features, potentially reducing their generalization capabilities across diverse peptide sequences.

To address these challenges, this paper introduces a novel multi-model integration approach, NeuroPred-INT, designed to enhance both the accuracy and generalizability of neuropeptide prediction. The model begins by fine-tuning the ESM-1b protein language model on a neuropeptide-specific training set and applying a Hidden Markov Model (HMM) for multiple sequence alignment. This step facilitates the extraction of common sequence features, enriching data diversity and substantially improving model generalizability. Next, NeuroPred-INT incorporates a CNN for feature extraction to capture local sequence patterns, thereby improving its interpretive accuracy of sequence information. Building on this, a global attention mechanism generates attention weights that focus on the most salient features, capturing contributions from specific sequence positions and enhancing interpretability. Finally, a gradient boosting tree classifier is employed to refine predictive performance through further training and optimization. Independent testing demonstrates that NeuroPred-INT has a clear advantage in accuracy and generalizability compared to other state-of-the-art neuropeptide prediction models.

2. Materials and Methods

2.1 Datasets

The dataset used in this study is based on a previously published dataset, modified to meet the specific requirements of deep learning models. The original data were sourced primarily from the NeuroPep 2.0 database, which includes 11,282 experimentally validated neuropeptide sequences, with 5,333 new entries added in version 2.0 [9]. After applying a series of filtering steps—retaining neuropeptides between 5 and 100 residues in length and setting a CD-HIT similarity threshold of 0.9—4,463 neuropeptide samples were selected. For a fair evaluation, all neuropeptide test data were drawn from the newly added entries in NeuroPep 2.0, with 444 sequences (10%) randomly selected as an independent test set.

During data processing, the original CSV format (seq, label) was converted to FASTA format to accommodate deep learning model input requirements. In the FASTA file, “NP” represents positive samples, while “OTP” represents negative samples, facilitating efficient parsing and processing of sequence data by the model. The dataset can be accessed via the URL: https://github.com/LeanderLi1014/dataset.git

2.2 Methods

NeuroPred-INT is an integrated model designed for neuropeptide prediction. The model architecture of NeuroPred-INT is presented in Fig. 1. First, neuropeptide sequences, collected in FASTA format, were subjected to multiple sequence alignment (MSA) using a Hidden Markov Model (HMM), generating an output.afa file that emphasizes shared sequence features to improve model generalization [10]. Next, the ESM-1b model was pre-trained on the neuropeptide-specific dataset and fine-tuned with custom attention layers and a label transformer block. The attention layer identifies and assigns weights to different positions within the sequence, producing a weighted sum of hidden states. Subsequently, the fine-tuned ESM-1b model, in combination with a convolutional neural network (CNN), extracted high-dimensional feature representations from the input sequences. These representations were then processed by a gradient boosting tree classifier, which optimized feature weighting and classification, enabling accurate differentiation between neuropeptides and other peptides.

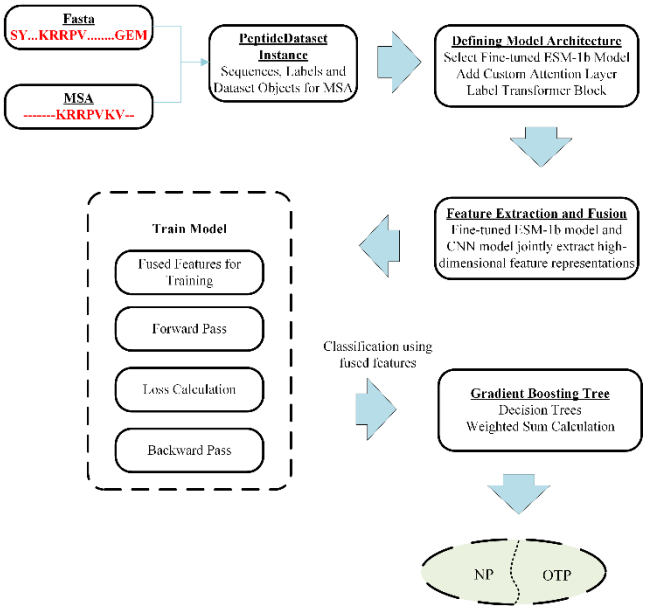

Fig. 1. The flowchart of NeuroPred-INT.

2.2.1 Fine-tuning of the ESM-1b model

The ESM-1b model, developed by Facebook AI, serves as the core of our neuropeptide classification approach [11]. To optimize the model's learning from neuropeptide sequences, we defined a custom dataset class, PeptideDataset, to manage the input sequences and their corresponding labels.

For data preparation, we used a custom function, $read\_fasta()$, to read the FASTA file containing neuropeptide sequences and their labels. This function parses sequence headers, identifying and extracting sequences by distinguishing neuropeptides (NP) from other peptides (OTP).

We then initialized the ESM-1b model and tokenizer with pre-trained weights from the Hugging Face model hub. Before training, we performed Bayesian optimization to determine optimal hyperparameters, setting a learning rate of 1.49e-6 and five training epochs [12]. We created an instance of PeptideDataset with sequences and labels extracted from the FASTA file, splitting it into 80% for training and 20% for validation. Data loaders were configured for efficient batch processing, with an empirically determined batch size of 32 for optimal performance.

To further enhance training efficiency, we enabled mixed-precision training (fp16), which improved performance and reduced memory usage. The final model training configuration, established using the TrainingArguments class, included parameters for the optimized learning rate, epochs, and a weight decay of 0.01.

2.2.2 Hidden Markov Model and MSA

Neuropeptide sequences were first collected in FASTA format and then processed through multiple sequence alignment (MSA) using a Hidden Markov Model (HMM). The HMM estimates parameters via

the Baum-Welch algorithm, maximizing the likelihood of the observed sequences. The conditional probability of the actual neuropeptide sequences is calculated as follows:

$$P(X \mid \lambda) = \sum_{all\ paths} P(X, path \mid \lambda) \quad (1)$$

where X represents the neuropeptide sequence in the dataset, and λ denotes the model parameter [13,14]. This formula is instrumental in generating the MSA, capturing conserved features and patterns within the sequences that are critical for accurate classification. The resulting MSA file effectively identifies conserved sequence motifs and patterns, which are essential for reliable neuropeptide classification.

2.2.3 Custom Attention Laye

A custom attention layer is incorporated to highlight critical features within neuropeptide sequences. The attention mechanism calculates weights α_ι through a softmax function applied to linear transformations of hidden states h_i:

$$\alpha_\iota = \frac{e^{Wh_i+b}}{\sum_{j=1}^{T} e^{Wh_i+b}} \quad (2)$$

where W is the learnable weight matrix, b represents the bias term, T is the sequence length, and i, j are indices of sequence elements [15].

These attention weights are then applied to hidden states through weighted summation, yielding an aggregated representation:

$$c = \sum_{i=1}^{T} a_i h_i \quad (3)$$

The aggregated representation is subsequently processed by a transformer block to capture complex dependencies within the sequence data, after which the output is passed to a classifier for final prediction [16-18].

2.2.4 Gradient Boosting Trees

Gradient Boosting Trees (GBT) is a powerful ensemble learning method that enhances model prediction performance by iteratively constructing a series of weak learners, typically decision trees [19]. GBT optimizes the model by minimizing the gradient of the loss function, enabling it to excel on complex datasets.

The fundamental concept behind GBT is the iterative addition of new decision trees to correct the errors made by the preceding trees. Each new tree is trained on the residuals (errors) of the previous tree. Specifically, GBT is trained as follows [19]:

1. **Initialize the model:**

$$F_0(x) = \arg min_\gamma \sum_{i=1}^{n} L(y_i, y) \quad (4)$$

where L represents the loss function, y_i denotes the actual value, and γ is a constant.

2. **Iterative training:**

For each iteration $m = 1,2, \ldots, M$:

Compute the residuals (negative gradients):

$$r_{im} = \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad (5)$$

Fit a new decision tree $hm(x)$ to predict the residuals:

$$h_m(x) = \arg min_h \sum_{i=1}^{n} (r_{im} - h(x_i))^2 \quad (6)$$

Update the model:

$$F_m(x) = F_{m-1}(x) + \upsilon h_m(x) \quad (7)$$

Where υ is the learning rate, which controls the contribution of each tree to the final model.

3. **Final model:**

$$F_m(x) = \sum_{m=1}^{M} \upsilon h_m(x) \quad (8)$$

2.2.5 Loss Function

The loss function applied in this study is Binary Cross-Entropy Loss, appropriate for binary classification tasks, such as differentiating between neuropeptides (NP) and other peptides (OTP) [20]. The Binary Cross-Entropy Loss is defined mathematically as follows:

$$L(y, \hat{y}) = -[-y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})] \quad (9)$$

where: L denotes the loss function, y represents the true labels (0 for OTP, 1 for NP), $\hat{y}$ indicates the model's predicted probabilities.

During training, the loss function integrates with the model through these key steps:

1. Forward Propagation: The model processes the input data, generating predicted probabilities$\hat{y}$.
2. Loss Calculation: The Binary Cross-Entropy Loss quantifies the discrepancy between the predicted probabilities $\hat{y}$ and the true labels y, yielding the loss L.
3. Backward Propagation: The gradients of the loss L with respect to the model parameters are computed. The optimizer then updates the model parameters to minimize the loss.

This iterative procedure continues until the model parameters converge, minimizing the loss function and resulting in a model capable of accurately classifying neuropeptides versus other peptides.

2.2.6 Performance Metrics

To evaluate the performance of the neuropeptide classification model, several metrics were calculated, including Precision, Recall, F1 Score, Sensitivity, Specificity, Accuracy, and Matthews Correlation Coefficient (MCC). Together, these metrics provide a comprehensive assessment of the model's classification capabilities. Here, TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively [21-23].

Precision (Pre)

Precision measures the accuracy of positive predictions, defined as:

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

Recall (Rec) / Sensitivity (SN)

Recall, also known as Sensitivity, evaluates the model's ability to identify positive samples, defined as:

$$Recall = Sensitivity = \frac{TP}{TP+FN} \quad (11)$$

F1 Score (F1)

The F1 Score, a harmonic mean of Precision and Recall, balances these metrics:

$$F1\ Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (12)$$

Specificity (SP)

Specificity evaluates the model's ability to identify negative samples, defined as:

$$Specificity = \frac{TN}{TN+FP} \quad (13)$$

SpecAccuracy (ACC)

Accuracy measures the overall correctness of the model's predictions:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Matthews Correlation Coefficient (MCC)

The Matthews Correlation Coefficient is a balanced metric that considers all four components of the confusion matrix:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \quad (15)$$

3. Results and Discussion

3.1 Hyperparameters Tuning

Extensive hyperparameter tuning was performed prior to model training to enhance performance. Parameters tuned included the learning rate, training epochs, multi-head attention configuration, and classifier choice (Random Forest vs. Gradient Boosting Tree). All experiments were conducted on Google Colab, utilizing the integrated T4 and L4 GPUs.

Following the tuning process, optimal parameters were identified for neuropeptide classification. The final model was configured with a learning rate of 1.29×10^{-5}, 10 training epochs, and 12 attention heads, with a Gradient Boosting Tree classifier selected for classification. These settings provided an optimal balance between model accuracy and training stability, ensuring reliable performance in distinguishing neuropeptides from other peptide sequences.

3.2 Comparison with the state-of-the-art methods for neuropeptide prediction

To ensure an unbiased comparison, we used a consistent independent test set for model evaluation. As shown in Tab. 1, the INT model demonstrated outstanding performance, achieving an F1 score of 0.9229 and an accuracy of 0.9212, closely matching the PLM model's F1 score of 0.9238 and accuracy of 0.9223. Minor differences in precision, recall, and specificity between the INT and PLM models suggest that both are highly effective for neuropeptide classification.

The ROC and PR curves for the INT model, illustrated in Fig. 2 and Fig. 3, further demonstrate its robustness. The ROC curve, with an AUC of 0.963, indicates a high true positive rate across various thresholds, while the PR curve, with an AP of 0.962, shows that the model maintains high precision as recall increases. Notably, the AUC and AP values for the INT model are higher than those of the other two models. These metrics confirm the INT model's effective balance of precision and recall, establishing it as a reliable tool for neuropeptide prediction.

Although the FRL model achieved the highest precision (0.9517) and specificity (0.9617), it showed lower recall (0.7545) and a correspondingly lower F1 score (0.8417). These results indicate that the FRL model adopts a more conservative classification strategy, emphasizing precision over recall. This approach may reduce false positives but could increase false negatives.

In summary, while the INT and PLM models offer a balanced approach to neuropeptide classification, the FRL model prioritizes precision, potentially at the cost of recall. This analysis underscores the importance of selecting a model based on specific application needs, whether the priority is minimizing false positives or achieving balanced performance.

3.3 Visualization of features extracted by NeuroPred-INT

To gain insights into the features extracted by the NeuroPred-INT model, we employed t-SNE (t-distributed Stochastic Neighbor Embedding) to visualize high-dimensional data at key stages of the model's processing (see Fig. 4-7) [24]. These visualizations reveal how the model differentiates neuropeptides from other peptides.

The initial t-SNE plot shows the output of the tokenizer, representing input sequences in a numerical format. Here, two relatively distinct clusters emerge, with minor overlap, indicating that the tokenizer captures essential sequence features and establishes a foundation for further processing.

Tab. 1. Performance comparisons of NeuroPred-INT with the two representative state-of-the-art methods on the independent test set

Methods	Pre	Rec	F1	SN	SP	ACC	MCC
FRL Model	0.9517	0.7545	0.8417	0.7545	0.9617	0.8581	0.7321
PLM Mode	0.9067	0.9414	0.9238	0.9414	0.9032	0.9223	0.8452
INT Model	0.9030	0.9437	0.9229	0.9437	0.8986	0.9212	0.8432

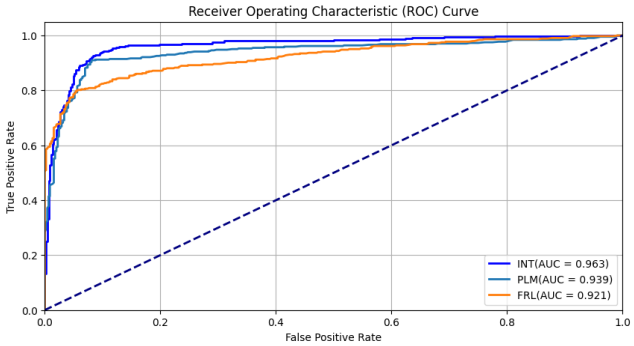

Fig. 2. Comparison of ROC curves of FRL, PLM and INT models

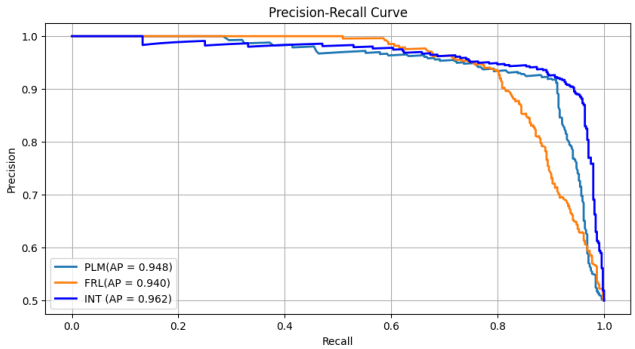

Fig. 3. Comparison of PR curves of FRL, PLM and INT models

In the next stage, the t-SNE plot of the Transformer layer output illustrates the enhanced separation between neuropeptides and other peptides. Compared to the tokenizer output, data points are more dispersed within the feature space, suggesting that the Transformer layer amplifies the separation by creating a broader representation.

Following this, the t-SNE plot of the attention layer output captures the effects of the custom attention mechanism. This layer emphasizes the most relevant sequence components, enhancing feature extraction. The plot shows two clusters with some overlap, indicating that the attention mechanism retains distinctiveness while refining the feature space. Finally, the t-SNE plot of the classifier logits reveals two compact and well-separated clusters with minimal overlap, demonstrating the classifier's effectiveness in consolidating extracted features for accurate predictions.

3.4 Model Interpretability

To enhance the interpretability of our neuropeptide classification

model, we employed a combination of attention mechanisms and visualizations of learned representations. The AttentionModel, which incorporates a custom attention layer, enables us to investigate the influence of specific amino acid residues on the model's predictions. By analyzing the attention weights assigned to each residue, we can identify the critical segments of the peptide sequences that are pivotal for classification.

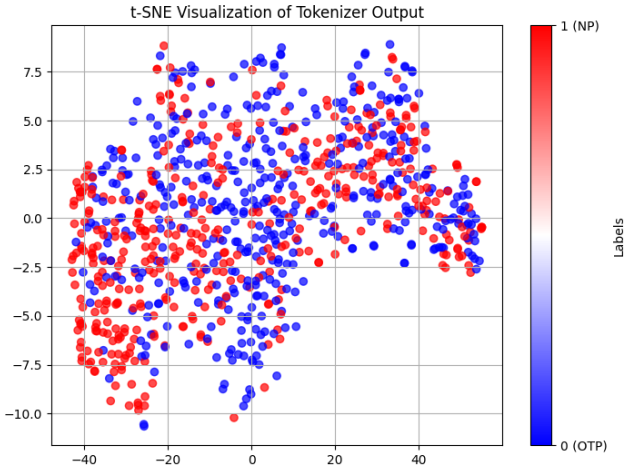

Fig. 4. Visualization of Tokenizer Output

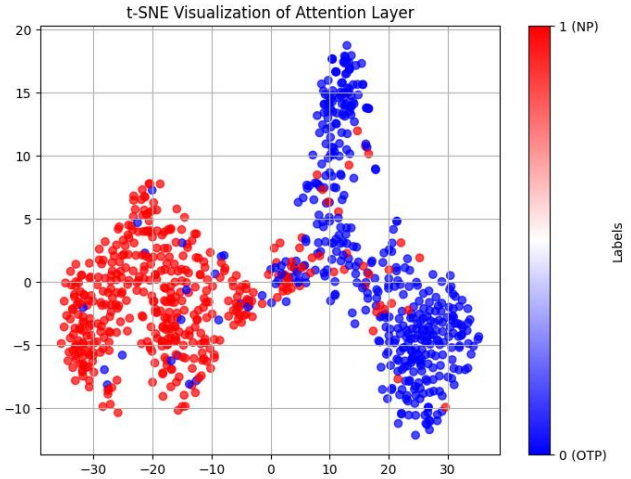

Fig. 5. Visualization of Transformer Layer

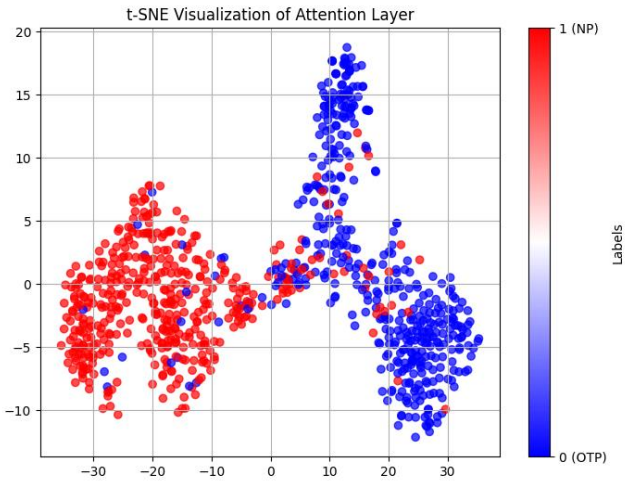

Fig. 6. Visualization of Attention Layer

For each input sequence, we computed the attention weights generated by the model during the forward pass. These attention weights, derived from the last hidden states, reflect the significance of each amino acid in determining whether a sequence belongs to the neuropeptide (NP) or non-neuropeptide (OTP) class.

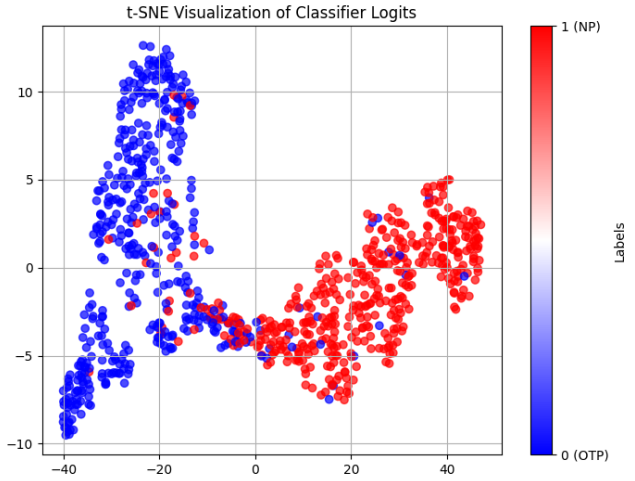

Fig. 7. Visualization of Classifier Logits

The visualization results indicate that the model predominantly focuses on key residues characteristic of each class. For instance, in NP sequences, specific residues such as F and G exhibited significantly higher attention weights, suggesting their crucial roles in neuropeptide functionality. Conversely, OTP sequences displayed distinct patterns, with different residues gaining attention, aiding in their differentiation from NPs.

The implementation of multiple sequence alignment (MSA) via a Hidden Markov Model (HMM) significantly enhances the model's interpretability. The aligned output sequences in the output.afa file reveal conserved regions consistent across similar peptides. This alignment not only highlights shared sequence features but also aids in understanding the evolutionary relationships among peptides.

Analysis of the alignment revealed that conserved motifs often correspond to regions with higher predictive accuracy, as evidenced by the model's performance metrics. This correlation between alignment and prediction success underscores the importance of evolutionary context in peptide classification.

4. Conclusions

In this study, we introduced NeuroPred-INT, a novel multi-model integration approach designed to enhance both the accuracy and generalizability of neuropeptide prediction. By fine-tuning the ESM-1b protein language model on a neuropeptide-specific training set and applying a Hidden Markov Model (HMM) for multiple sequence alignment, we successfully extracted shared sequence features, enriching data diversity and improving model generalization.

The incorporation of a Convolutional Neural Network (CNN) for feature extraction, combined with a global attention mechanism, enabled NeuroPred-INT to capture local sequence patterns and emphasize the most salient features. This approach significantly enhanced the interpretive accuracy of sequence information. In the final step, a gradient boosting tree classifier further refined predictive performance through targeted training and optimization

Our extensive evaluation using independent test sets demonstrated that NeuroPred-INT surpasses other state-of-the-art neuropeptide prediction models in both accuracy and generalization. The model achieved an F1 score of 0.9229 and an accuracy of 0.9212, maintaining a balanced approach to both precision and recall.

t-SNE visualization of NeuroPred-INT features revealed distinct

clusters, highlighting each component's role in neuropeptide differentiation. The attention mechanism, in particular, identified key residues influencing predictions.

In summary, NeuroPred-INT is a robust and interpretable tool advancing neuropeptide prediction, with future work aimed at improving interpretability and expanding to other peptide classifications.

References

[1] Van Bael, S., Watteyne, J., Boonen, K., De Haes, W., Menschaert, G., Ringstad, N., Horvitz, H. R., Schoofs, L., Husson, S. J., and Temmerman, L., "Mass spectrometric evidence for neuropeptide-amidating enzymes in Caenorhabditis elegans," *Journal of Biological Chemistry*, vol. 293, pp. 6052-6063, 2018.

[2] Kormos, V. and Gaszner, B., "Role of neuropeptides in anxiety, stress, and depression: from animals to humans," *Neuropeptides*, vol. 47, pp. 401-419, 2013.

[3] Carniglia, L., Ramírez, D., Durand, D., Saba, J., Turati, J., Caruso, C., Scimonelli, T. N., and Lasaga, M., "Neuropeptides and microglial activation in inflammation, pain, and neurodegenerative diseases," *Mediators of Inflammation*, vol. 2017, article ID 5048616, 2017.

[4] Ji, Q. Y., et al., "Utilized mass spectrometry for neuropeptide identification, a method that, while effective, requires costly equipment, complex procedures, and skilled personnel for operation and maintenance," *[Add Journal Name]*, vol. [Add Volume], pp. [Add Page Numbers], 2017.

[5] Hasan, M. M., Alam, M. A., Shoombuatong, W., Deng, H. W., Manavalan, B., and Kurata, H., "NeuroPred-FRL: an interpretable prediction model for identifying neuropeptide using feature representation learning," *Briefings in Bioinformatics*, vol. 22, article ID bbab167, 2021.

[6] Wang, L., Huang, C., Wang, M., Xue, Z., and Wang, Y., "NeuroPred-PLM: an interpretable and robust model for neuropeptide prediction by protein language model," *Briefings in Bioinformatics*, vol. 24, article ID bbad077, 2023.

[7] Wen, J., Ding, Z., Wei, Z., Xia, H., Zhang, Y., and Zhu, X., "NeuroPpred-SHE: An interpretable neuropeptides prediction model based on selected features from hand-crafted features and embeddings of T5 model," *Computers in Biology and Medicine*, vol. 181, article ID 109048, 2024.

[8] Wang, L., Zeng, Z., Xue, Z., and Wang, Y., "DeepNeuropePred: A robust and universal tool to predict cleavage sites from neuropeptide precursors by protein language model," *Computational Structural Biotechnology Journal*, vol. 23, pp. 309-315, 2023. DOI: 10.1016/j.csbj.2023.12.004.

[9] Kim, Y., Bark, S., Hook, V., and Bandeira, N., "NeuroPedia: neuropeptide database and spectral library," *Bioinformatics*, vol. 27, pp. 2772-2773, 2011.

[10] Gupta, A., and Dhingra, B., "Stock market prediction using hidden Markov models," in *2012 Students Conference on Engineering and Systems*, IEEE, 2012, pp. 1-4.

[11] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., et al., "Language models of protein sequences at the scale of evolution enable accurate structure prediction," *BioRxiv*, article ID 500902, 2022.

[12] Tong, S., Chen, F., Yang, L., and Shen, Z., "Novel utilization of a paper-level classification system for the evaluation of journal impact: An update of the CAS Journal Ranking," *Quantitative Science Studies*, vol. 4, no. 4, pp. 960-975, 2023.

[13] Li, H., Jiang, L., Yang, K., Shang, S., Li, M., and Lv, Z., "iNP_ESM: Neuropeptide Identification Based on Evolutionary Scale Modeling and Unified Representation Embedding Features," *International Journal of Molecular Sciences*, vol. 25, article ID 7049, 2024.

[14] Zhan, Q., Wang, N., Jin, S., et al., "ProbPFP: a multiple sequence alignment algorithm combining hidden Markov model optimized by particle swarm optimization with partition function," *BMC Bioinformatics*, vol. 20, suppl. 18, article ID 573, 2019. DOI: 10.1186/s12859-019-3132-7.

[15] Brauwers, G., and Frasincar, F., "A general survey on attention mechanisms in deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, pp. 3279-3298, 2021.

[16] Vaswani, A., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 2017.

[17] Santana, A., and Colombini, E., "Neural attention models in deep learning: Survey and taxonomy," *arXiv preprint arXiv:2112.05909*, 2021.

[18] Brauwers, G., and Frasincar, F., "A general survey on attention mechanisms in deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3279-3298, 2021.

[19] Feng, J., Yu, Y., and Zhou, Z.-H., "Multi-layered gradient boosting decision trees," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[20] Mao, A., Mohri, M., and Zhong, Y., "Cross-entropy loss functions: Theoretical analysis and applications," in *International Conference on Machine Learning*, PMLR, 2023, pp. 23803-23828.

[21] Powers, D. M. W., "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *ArXiv*, vol. abs/2010.16061, 2011.

[22] Hasan, M. M., Manavalan, B., Shoombuatong, W., Khatun, M. S., and Kurata, H., "i6mA-Fuse: improved and robust prediction of DNA 6 mA sites in the Rosaceae genome by fusing multiple feature representation," *Plant Molecular Biology*, vol. 103, pp. 225-234, 2020.

[23] Yang, W., Zhu, X.-J., Huang, J., Ding, H., and Lin, H., "A brief survey of machine learning methods in protein sub-Golgi localization," *Current Bioinformatics*, vol. 14, pp. 234-240, 2019.

[24] Cai, T. T., and Ma, R., "Theoretical foundations of t-sne for visualizing high-dimensional clustered data," *Journal of Machine Learning Research*, vol. 23, pp. 1-54, 2022.

Yue Li is currently pursuing a master's degree at College of Electronic Information, Guangxi Minzu University, Nanning, China. He received his bachelor's degree from Dalian Minzu University, China in 2021. His main research interests are in the areas of speech enhancement, neuropeptide classification, deep learning, and traditional algorithm integration in signal processing.

Ji Qiu is a lecturer in College of Electronic Information, Guangxi Minzu University. She received the BEng degree with first class honours in 2015 and the PhD degree in 2020 from University of the West of England, Bristol, United Kingdom. She was an instructor from 2016 to 2020 and an associate lecturer from 2019 to 2021 at Department of Engineering Design and Mathematics, University of the West of England, Bristol, United Kingdom. Her main research interest is in the area of numerical algorithm development, nonlinear system modelling, identification, and control.