The Journal of Basic and Clinical Dentistry

Original Article

Hybrid Peptide Classifier Model for Predicting Periodontal Cell-Penetrating Peptides

Pradeep Kumar Yadalam[1,*]

1. Department of Periodontology, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, 600077, India.

***Corresponding Author:** Pradeep Kumar Yadalam, Department of Periodontology, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, 600077, India. Email: pradeepkumar.sdc@saveetha.com.

Citation: Yadalam, P.K. Hybrid Peptide Classifier Model for Predicting Periodontal-Cell Penetrating Peptides. J Basic Clin Dent, 2025;2(1), 1–13.

Received: 26th February 2025
Revised: 15th May 2025
Accepted: 16th May 2025
Published: 24th May 2025

Abstract.

***Background*:** Periodontitis, a global infectious disease-causing tooth loss and tissue destruction, is linked to diabetes and cardiovascular diseases. Conventional treatments fail to eliminate pathogens, necessitating alternative therapies. Cell-penetrating peptides (CPP) are promising for therapeutic applications like genetic defect correction and gene silencing, but face challenges like cytotoxicity and immune responses. They also manage periodontal disease by delivering agents directly to targeted tissues, improving drug penetration and treatment outcomes. CPP unique ability to traverse cellular membranes is key. Hybrid Peptide Classifier, a novel model using an LLM-based attention network, combines the strengths of multiple neural network layers to model peptide sequence structure and dependencies effectively. By improving medication delivery straight to infected periodontal sites, CPP provide a novel treatment option for periodontitis because of their antimicrobial activity and tissue-penetrating capacity. This model aims to predict periodontal cell-penetrating peptides, accelerating advancements in peptide-based therapies and drug delivery systems. ***Methods*:** The peptide classification dataset was sourced from thegleelab.org/MLCPP/MLCPPData.html, featuring sequences for both positive and negative sample classes. A custom PyTorch Dataset class was related to maintaining a consistent sequence length. The dataset was split into training and testing subsets and loaded into DataLoader objects for efficient batch processing. The hybrid peptide classifier class is a neural network model designed for peptide classification, initialized with vocabulary size, embedding dimension, hidden dimension, and maximum sequence length, and subjected to training with an epoch of 10 with early stopping. A hybrid architecture comprising convolutional and bidirectional LSTM layers was used to categorize peptide sequences. ***Results*:** The model exhibited strong classification performance with an accuracy of 85.2%, an F1-

score of 0.88, and an AUC of 0.93. ***Conclusion*:** CPP are promising tools for drug delivery and gene therapy, but challenges like data imbalances and experimental variability need to be addressed. Our study showed promising results in better classifying the peptide sequences. Future research should focus on refining machine learning techniques, integrating diverse datasets, and implementing rigorous validation protocols to improve peptide classification models' reliability and patient outcomes in peptide-based therapeutics. This model provides a basis for creating customized, targeted peptide treatments in periodontology.

Keywords: Peptide Sequences, BERT, Periodontal Disease.

1. Introduction

Periodontitis is a complex infectious disease caused by pathogenic microorganisms, destroying periodontal tissues and potentially leading to tooth loss, affecting a few million people worldwide. The disease is linked to systemic conditions like diabetes and cardiovascular diseases[1,2] Its onset is primarily initiated by plaque biofilm, which serves as a habitat for various microorganisms. Conventional treatments,[3] including scaling and surgical interventions, often fail to eliminate periodontal pathogens, necessitating the exploration of alternative therapies.

Cell-penetrating peptides (CPP)[4,5] have attracted significant interest since the early 1990s, particularly following studies on the HIV Tat protein that demonstrated their ability to deliver cargo molecules into cells. Advancements in life sciences have led to a surge in biologically active drugs, particularly biopharmaceuticals, which now make up a significant portion of FDA-approved drugs. Peptides and proteins, key biomacromolecules, offer low toxicity and high bioactivity but face challenges like receptor selectivity and oral bioavailability. CPP, which include various classifications based on their properties and origins, have shown promise in clinical applications, with over 25 CPP-based products currently in clinical evaluation.[6]

CPP transports various payloads, making it promising for therapeutic applications, such as correcting genetic defects or silencing disease-causing genes in gene and cancer therapies. However, clinical translation challenges include cytotoxicity, limited tissue specificity, and immune responses.[7] Additionally, CPPs play a role in managing periodontal disease by delivering therapeutic agents directly to targeted tissues, which enhances drug penetration and allows for localized therapy with fewer side effects, thereby improving patient compliance and treatment outcomes. These short amino acid sequences, typically ranging from 5 to 30 residues in length, have garnered attention in biochemistry, molecular biology, and therapeutic development due to their unique ability to traverse cellular membranes. With cationic properties and hydrophobic regions, CPP effectively interacts with negatively charged phospholipid membranes of the periodontal fibroblasts.[8]

Recent studies investigate the antibacterial effects of a small peptide, RR9, derived from penetratin, which targets oral bacteria linked to periodontitis, specifically *Streptococci oralis*, *Streptococci gordonii*, and *Streptococci*. This peptide shows promise in addressing bacterial infections associated with periodontal disease..[9] Various assays assess RR9's capability to inhibit

bacterial growth in planktonic and biofilm states, revealing low cytotoxicity to human gingival fibroblasts. Findings indicate that RR9 impedes streptococcal growth and decreases inflammation, highlighting its potential as an antimicrobial agent for managing periodontal disease. The overuse of antibiotics in clinical settings has led to decreased effectiveness and increased drug-resistant bacteria.[10] Antimicrobial peptides (AMPs) and CPP are being explored as alternatives. CPP, such as penetratin, can efficiently deliver therapeutic agents into cells and exhibit strong bactericidal properties while sparing mammalian cells. A derivative of penetratin, RR9, is simpler to synthesize but needs further evaluation for its effectiveness against periodontal bacteria; however, there is a lack of classification and prediction of CPP for optimizing therapies.

Machine learning techniques like BERT Bidirectional Encoder Representations from Transformers and LLMs-large are crucial for classifying CPP for various applications. Predictive modeling helps in screening new CPP for biomedical applications. Automated classification systems reduce experimental trial time and costs, making drug development more efficient and effective.[11] Machine learning is essential for theoretical understanding and practical healthcare and drug innovation applications. Researchers can identify promising candidates for drug delivery by collecting a comprehensive dataset of known CPP, tokenizing peptide sequences, training the model, extracting relevant features, and classifying new or unexplored peptides. The model's performance is assessed using metrics such as accuracy, precision, recall, and F1-score, and it is continuously improved through feedback and new experimental data.[12] Leveraging the representational capabilities of BERT and similar models enhances the ability to classify CPP, accelerating advancements in peptide-based therapies and drug delivery systems. Architectures that capture subtle sequence features are necessary because existing models, such as Graph CPP, often misclassify short sequences and result in high false positives. We acknowledge the need for a novel peptide classifier using an LLM-based attention network. Accordingly, we developed a hybrid peptide classifier, a new model that combines the strengths of multiple neural network layers to effectively model the structure and dependencies in peptide sequences for improved periodontal treatment management. Therefore, we aim to create and evaluate a hybrid neural network model for highly accurate predictions of periodontal cell-penetrating peptides.

2. Materials and Methods

2.1. Dataset Preparation

We obtained the dataset from http://www.thegleelab.org/MLCPP/MLCPPData.html.[13] The dataset retrieved for GT-positive.txt includes sequences representing the positive class, while GT-negative.txt contains sequences representing the negative class. This information pertains to cell-penetrating peptides. The FASTA format often includes headers starting with ">"; we eliminated these lines and extra whitespaces to capture each sequence accurately. The `load sequences` function reads a file, removes header lines, and returns the cleaned sequences as a list. It contains

427 positive sequences and 854 negative sequences. After loading, we combined the sequences and assigned labels (1 for positive and 0 for negative) (Figure 1).

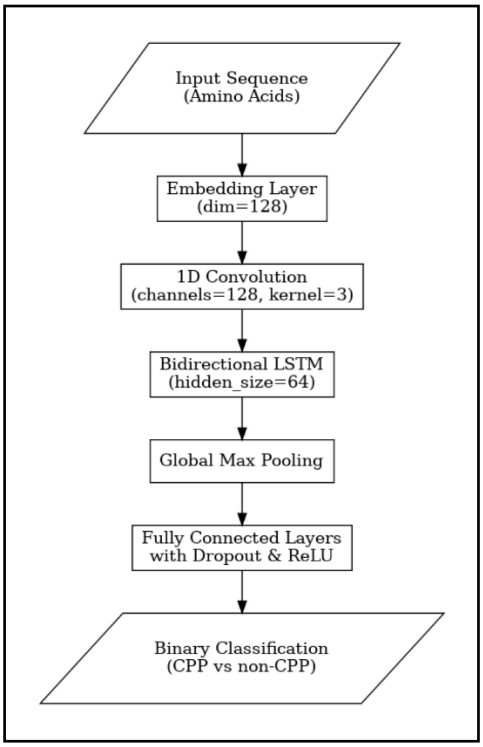

Figure 1 shows the workflow of the model architecture

2.2. Dataset Class

Peptide sequences consist of 20 standard amino acids. A custom PyTorch Dataset class was then created. This class transforms each amino acid in the peptide into its respective index from the vocabulary mapping and padding to maintain a consistent sequence length set to 100. The dataset was divided into training and testing subsets in an 80:20 ratio, using stratified sampling to preserve class distribution. These subsets were then loaded into DataLoader objects for efficient batch processing during training.

2.3. Hybrid Model Design

The `Hybrid Peptide Classifier` class is a neural network model for peptide classification. It initializes with vocabulary size, embedding dimension, hidden dimension, and maximum sequence length. The model includes an embedding layer, a 1D convolutional layer, a bidirectional LSTM, and a fully connected output layer. During the forward pass, input sequences are embedded, reshaped, processed through the convolution and LSTM layers, and pooled for classification. The model is then instantiated and moved to the appropriate device (CPU or GPU), demonstrating flexibility in handling computational resources.

2.4. Model Architecture

Our hybrid architecture combines an embedding layer that converts discrete characters (amino acids) into continuous representations, a convolutional layer to capture local patterns, and a bidirectional LSTM for long-range dependencies. The output passes through fully connected layers for binary classification. The model captures both local and sequential information from peptide sequences.

Embedding Layer: This layer transforms each amino acid index into a dense vector representation (dimension 128) to continuously represent the discrete symbols.

Convolutional Layer (Conv1d): The embedded sequence is permuted for convolution. The 1D convolution (128 output channels, kernel size 3) captures local patterns and short motifs significant for classifying peptide activity.

Bidirectional LSTM: Features are permuted back as a sequence, and a bidirectional LSTM captures long-range dependencies in both directions. The hidden state size is set to 64 for each direction, yielding an output dimension of 128.

Global Max Pooling: A global max pooling operation reduces the sequence into a fixed-length feature vector representing salient features across the sequence.

Fully Connected Layers: The resulting vector goes through dense layers with dropout and ReLU non-linearity, producing an output with a binary cross-entropy loss function to classify the peptide.

Loss Function and Optimization: The training used binary cross-entropy loss with logits, integrating sigmoid activation with binary cross-entropy. The Adam optimizer was selected with a learning rate of 0.001.

Epoch-wise Training and Evaluation: The model was trained with batch sequences and gradient descent, monitored for progress, and evaluated using a sigmoid function. It recorded training losses and test accuracies across ten epochs. Accuracy, precision, recall, F1-score, and area under the curve (AUC) were among the evaluation metrics. Additionally, we used 5-fold cross-validation to test the model's robustness.

3. Results

The average training loss of 0.0877 indicates that the model learned effectively during training, minimizing errors on the dataset. The test accuracy of 0.8521, or approximately 85.21%, reflects the model's ability to generalize well to new, unseen data. In peptide classification, this performance suggests that the model can reliably identify and categorize peptide sequences. These results imply that the model is well-suited for peptide classification tasks. The results indicated an F1-score of 0.88, with a precision of 0.85 and a recall of 0.84.

Table 1. Analysis of the predicted sequences

Sequence	True label	Predicted label	Probability
DDLRINLWNFEITNQSFNIADIKPANMEELTE	0	1	0.69144577
PKIACSQGWYDLW	0	0	0.011187438
QSLPSLRHLQLLPSP	0	1	0.9997904
VTVNISSPNTKNLRQL	0	1	0.9987011
SMAMGRLGLRPG	0	1	0.9998616
TPLEAIASSLTELFAPNLHE	0	0	0.020890346
VIRVHFRLPVRTV	1	0	0.19742046
LNDALSQLVGQQV	0	0	0.006681363
DPKGDPKGVTVTVTVTVTGKGDPKPD	1	1	0.9990503

Table 1 summarizes model predictions and the true and predicted labels for various sequences. The model accurately predicted sequences with a high probability of 1 (0.9990503) but misclassified some sequences with a true label of 0 (0.0067). The model also struggled with correctly classifying negative sequences, with many false positives. The results suggest further refinement to improve predictive accuracy, particularly distinguishing between negative and positive classes.

3.1. Training Curves

Training Loss Curves

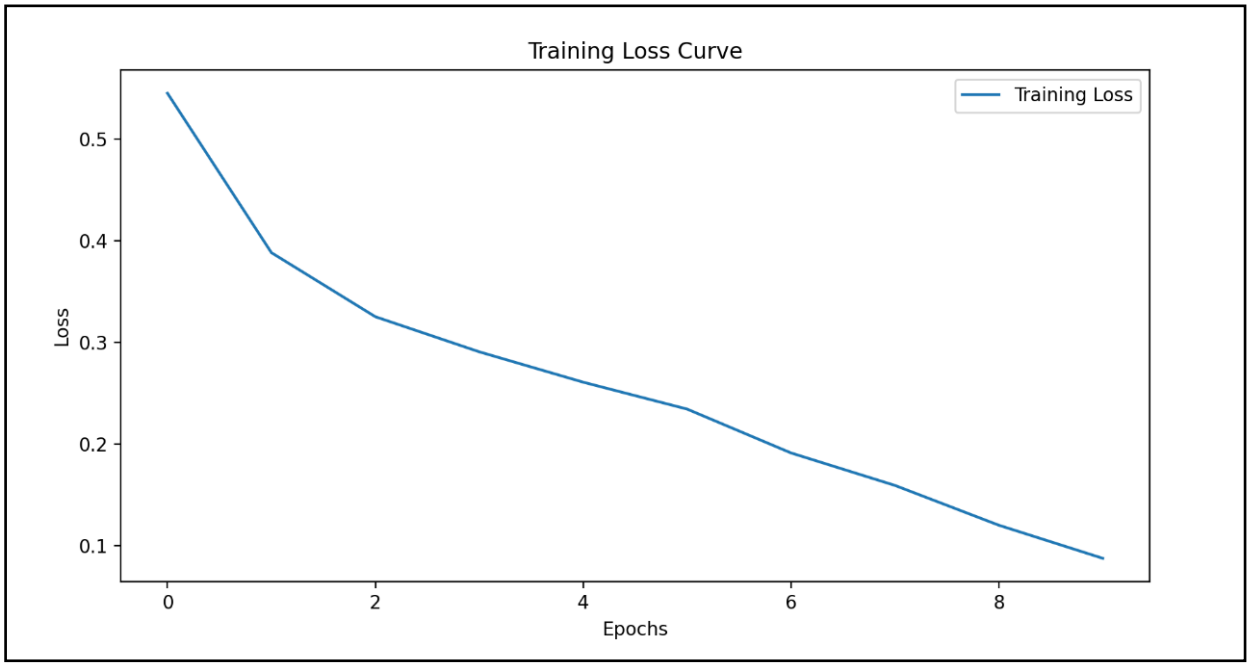

Figure 2. shows the training loss curve and how the loss decreased over the epochs. A steadily declining loss curve indicates effective training. The model's training loss significantly decreased as it learned to identify peptide sequence patterns, starting at a higher value and stabilizing around 0.0877.

Test Accuracy Curve

The test accuracy curve (Figure 2) depicts the model's performance on the unseen test data. With accuracy improving over epochs, the curve indicates that the model generalizes well without overfitting the training data.

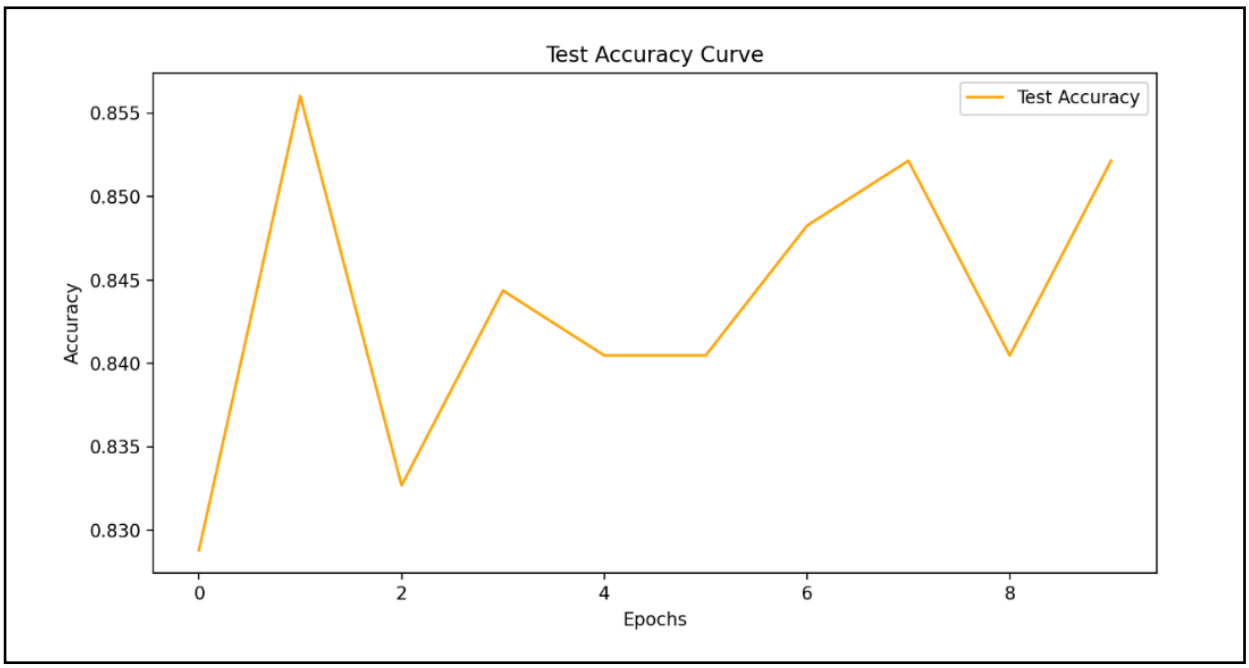

Figure 3. Test Accuracy Curve (Epoch vs. Test Accuracy).

Figure 3 shows that the accuracy improved steadily and reached approximately 85.21%, showing that the hybrid architecture effectively distinguishes the two classes of peptide sequences of the test accuracy curve.

Confusion Matrix

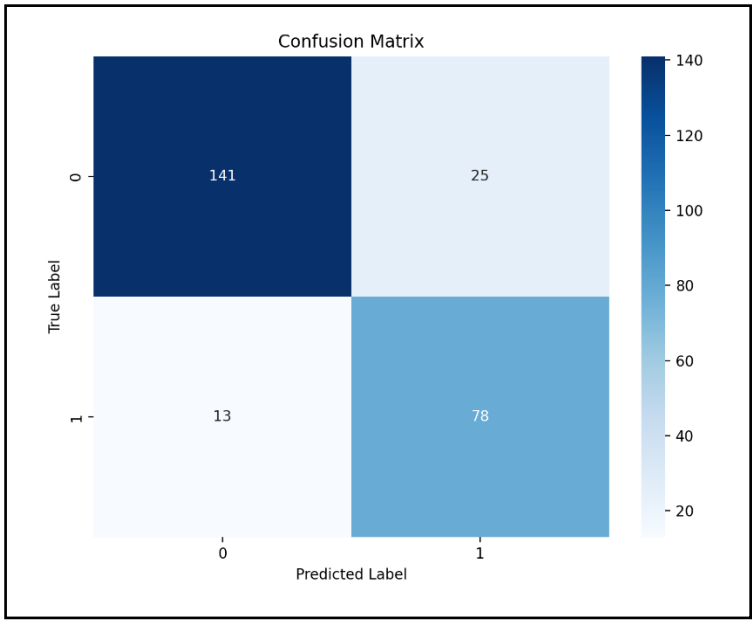

Figure 4. displays a confusion matrix that analyzes classification performance by identifying true positives, false positives, and false negatives.

The confusion matrix demonstrated high prediction accuracy, with a significant proportion of positive samples correctly classified. There are more false positives than false negatives, suggesting that the model may be biased toward predicting the positive class. This implies that threshold tuning or class balancing techniques are required to increase specificity (Figure 4).

3.2. ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve (Figure 5) further validates model performance. This curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various thresholds. The A(AUC) provides a single scalar defining the model's ability to distinguish between classes. A high AUC value indicates that the model effectively separates the positive and negative classes.

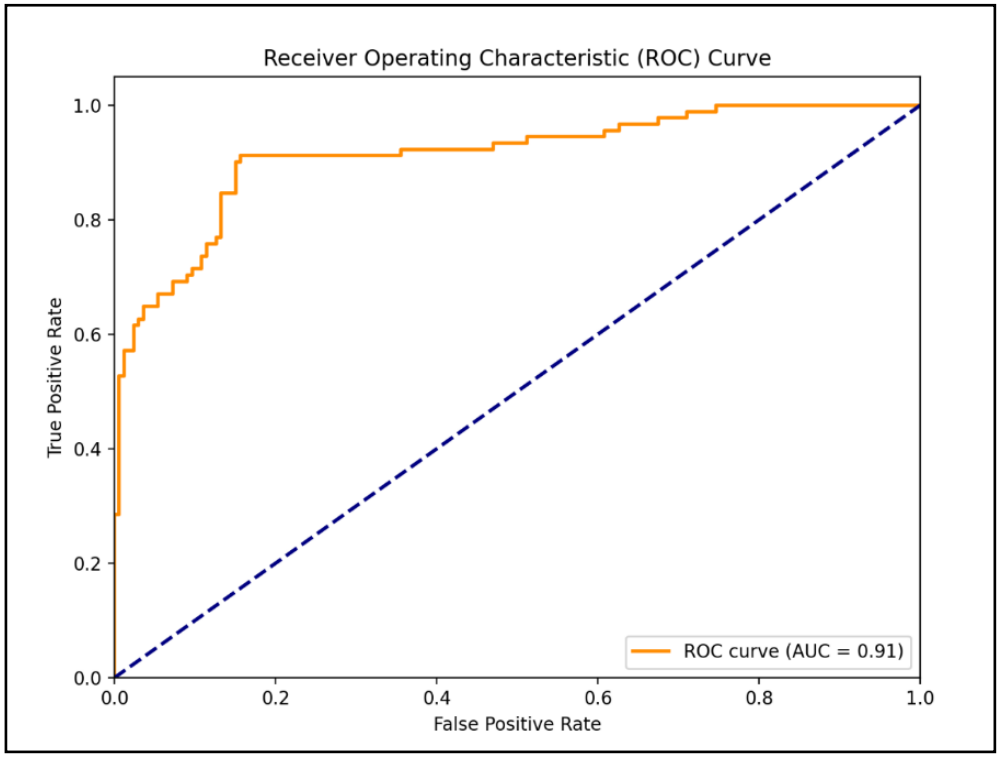

Figure 5. ROC Curve- The ROC plot shows that the obtained curve deviates well above the diagonal line (which represents random guessing), indicating a strong discriminatory power by the model.

4. Discussion

CPP are short, cationic peptides that traverse cellular membranes, making them valuable drug delivery tools and gene therapy tools. Recent advancements in natural language processing, especially models like BERT, offer innovative methods to automate this classification.[13,14] Accurately classifying CPP reveals the relationship between structure and function, aiding in developing improved drug delivery systems. This helps researchers identify candidates for targeting diseases that benefit from enhanced cellular uptake of therapeutics. One previous study on nanoparticles functionalized with lactoferrin-derived cell-penetrating peptide (hLF) improved intracellular drug delivery in nanomedicine. The modified NPs showed superior storage stability, uniform size, and increased cellular uptake. They also examined synthetic peptides like RR9, which show antimicrobial activity against early-stage oral colonizers and their potential for combating periodontal diseases.[11,15] CPP has demonstrated significant potential in revolutionizing the treatment of periodontal infections. Our objective was to predict peptide sequences using a hybrid classifier method to enhance the accuracy of these predictions. This approach aims to facilitate the synthesis of new peptides and provide deeper insights for future researchers working on novel CPP within this model.

Our study demonstrated that the model achieved a training loss of 0.0877, indicative of effective learning with minimal errors. The test accuracy was 0.8521, reflecting the model's capability to generalize to new data. Performance in peptide classification indicates reliability in identifying and categorizing peptide sequences. However, some sequences with a true label of 0 were misclassified, and the model exhibited difficulties in correctly classifying negative sequences (see Figures 2, 3, 4, and 5). Like a recent study, CPP facilitates cellular delivery, with CPP1708 representing the largest dependable CPP database utilized. Graph CPP surpasses earlier approaches, attaining substantial gains in prediction accuracy, illustrated by a Matthews correlation coefficient of 0.5787 and a AUC of 0.8459. Furthermore, the model adeptly learns peptide representations and shows high prediction confidence for peptides that are fewer than 40 amino acids long.[16] In parallel, a study implementing 10-fold cross-validation of the binary classification model AMP-BERT used a dataset of 1778 a AMPs and 1778 non-AMPs, repeated five times to evaluate generalization to external data.[12] Results revealed strong predictive performance, with average accuracy at 0.9280, an area under the receiver operating characteristic curve (AUROC) of 0.9665, an area under the precision-recall curve (AUPR) of 0.9653, sensitivity of 0.9262, specificity of 0.9303, and an F1 score of 0.9278. Additionally, a previous study demonstrated the efficacy of BertAIP, a model for protein feature extraction utilizing a fully connected feed-forward network for AIP classification, which achieved high accuracy alongside a Matthews correlation coefficient of 0.451, surpassing other methods. The model also enhanced interpretability by examining and visualizing crucial amino acids.[11,17] These metrics demonstrate the model's effectiveness in predicting unseen data after training. Future directions in peptide classification of CPP include enhancing machine learning models, integrating multi-omics data, improving feature extraction, conducting cross-species analysis, developing experimental validation protocols,

focusing on therapeutic potential, creating comprehensive databases for known and predicted CPP, and developing personalized medicine approaches. These advancements aim to capture complex patterns in peptide sequences, differentiate novel CPP from non-penetrating peptides, and develop personalized strategies for designing CPP tailored to specific therapeutic targets. These advancements will help researchers better understand the biological systems involved in CPP classification and develop more effective treatments.

BERT outperforms traditional peptide classification methods due to its transformer-based architecture, effectively capturing contextual and long-range dependencies in peptide sequences. This enables better feature extraction compared to models relying on local motifs. Its ability to distinguish between non-penetrating peptides and functional CPPs enhances understanding in CPP research, aiding predictions of high transduction efficiency and improving drug delivery systems. Incorporating BERT into the Hybrid Peptide Classifier boosts classification accuracy and relevance for individualized treatment plans. Misclassifications resulting from overlapping sequence motifs with CPP call for consideration of negative class augmentation and threshold adjustment. This model faces several limitations in peptide classification, including data imbalance, limited understanding of mechanisms, variability in experimental conditions, peptide stability and degradation, complexity of signaling pathways, generalizability across different contexts, and reliance on quality data[18–20], and ethical and regulatory challenges. The model showed excellent generalization abilities with high AUC and F1 scores, suggesting it learned sequence features linked to CPPs. A significant drawback is its propensity to incorrectly classify negative sequences as positives, most likely due to class imbalance or overlapping sequence motifs. This emphasizes the need for additional improvement, including threshold tuning, increasing the diversity of negative samples, and implementing sophisticated loss functions like focal loss to boost discrimination.

Data imbalance can lead to biased models, while the underlying mechanisms of CPP penetration are not fully understood. Experimental conditions can introduce variability, and the interaction of CPP with cellular signaling pathways can influence their effectiveness. Addressing these limitations can lead to improved periodontal therapeutic applications.

5. Conclusion

Our hybrid deep learning model proved effective in peptide classification tasks by demonstrating strong predictive performance in classifying periodontal cell-penetrating peptides, with an F1-score of 0.88 and an AUC of 0.93. These findings illustrate how the model can serve as a reliable means of expediting the development of peptide-based therapies by functioning as a robust pre-screening platform that reduces the cost and duration of experiments. Future research should explore ensemble strategies and attention-enhanced architectures to further enhance its specificity, particularly in identifying negative sequences. Combining multi-omics data and employing advanced transformer-based models like ProteinBERT may open new avenues in

periodontal precision medicine, allowing for more biologically informed and tailored peptide design.

Abbreviation	Full Form
CPP	Cell-Penetrating Peptide
LLM	Large Language Model
BERT	Bidirectional Encoder Representations from Transformers
LSTM	Long Short-Term Memory
AMP	Antimicrobial Peptide
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operating Characteristic Curve
AUPR	Area Under the Precision-Recall Curve
NPs	Nanoparticles
hLF	Human Lactoferrin-derived Peptide
ROC	Receiver Operating Characteristic
FASTA	(Not an acronym, but refers to the text-based format for representing sequences)
ReLU	Rectified Linear Unit (activation function)
MLCPP	Machine Learning Cell-Penetrating Peptides (Dataset source)
GraphCPP	Graph-based Cell-Penetrating Peptide classifier

Declarations:

Supplementary Materials: Not applicable.

Author Contributions: Conceptualization, P.K.Y; methodology P.K.Y; software, P.K.Y; validation, P.K.Y; formal analysis, P.K.Y; investigation, P.K.Y; resources, P.K.Y; data curation, P.K.Y; writing-original draft preparation, P.K.Y; writing—review and editing, P.K.Y; visualization, P.K.Y; supervision, P.K.Y; project administration, P.K.Y; funding acquisition. The author has read and agreed to the published version of the manuscript.

Funding: There is no funding for this article.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Consent for publication: Not applicable.

Data Availability Statement: All data was included in the article.

Acknowledgments: None.

Conflict of Interests: The authors declare no conflict of interest.

References

1. Ramamurthy J, Nedumaran N. Evaluation of C-reactive Protein and Interleukin-6 Levels in Periodontitis Patients: An Experimental Study. World Journal of Dentistry. 2025; 15:772–6.
2. Burra Anand D, Ramamurthy J, Kannan B, Jayaseelan VP, Arumugam P. N6-methyladenosine-mediated overexpression of TREM-1 is associated with periodontal disease. Odontology. 2025;113(2):834-843. doi: 10.1007/s10266-024-01009-w.
3. Ramamurthy J. Evaluation of Antimicrobial Activity of Nanoformulated Grape Seed Oil against Oral Microbes: An In Vitro Study. World Journal of Dentistry. 2024; 15:44–7.
4. Du JJ, Zhang RY, Jiang S, Xiao S, Liu Y, Niu Y, Zhao WX, Wang D, Ma X. Applications of cell penetrating peptide-based drug delivery system in immunotherapy. Front Immunol. 2025;16:1540192. doi: 10.3389/fimmu.2025.1540192.
5. Miwa A, Kamiya K. Cell-Penetrating Peptide-Mediated Biomolecule Transportation in Artificial Lipid Vesicles and Living Cells. Molecules. 2024;29(14):3339. doi: 10.3390/molecules29143339.
6. Kalmouni M, Oh Y, Alata W, Magzoub M. Designed Cell-Penetrating Peptide Constructs for Inhibition of Pathogenic Protein Self-Assembly. Pharmaceutics. 2024;16(11):1443. doi: 10.3390/pharmaceutics16111443.
7. Gong X, Han Y, Wang T, Song G, Chen H, Tang H, Huang X, Deng K, Wang S, Wang Y. Cell-Penetrating Peptide Induced Superstructures Triggering Highly Efficient Antibacterial Activity. Adv Mater. 2025;37(4):e2414357. doi: 10.1002/adma.202414357.
8. Hardan L, Chedid JCA, Bourgi R, Cuevas-Suárez CE, Lukomska-Szymanska M, Tosco V, Monjarás-Ávila AJ, Jabra M, Salloum-Yared F, Kharouf N, Mancino D, Haikel Y. Peptides in Dentistry: A Scoping Review. Bioengineering (Basel). 2023;10(2):214. doi: 10.3390/bioengineering10020214.
9. Galindo-Camacho RM, Haro I, Gómara MJ, Espina M, Fonseca J, Martins-Gomes C, et al. Cell penetrating peptides-functionalized Licochalcone-A-loaded PLGA nanoparticles for ocular inflammatory diseases: Evaluation of in vitro anti-proliferative effects, stabilization by freeze-drying and characterization of an in-situ forming gel. Int J Pharm. 2023;639:122982. doi: 10.1016/j.ijpharm.2023.122982.
10. BermúBermúdez M, Hoz L, Montoya G, Nidome M, Pérez-Soria A, Romo E, et al. Bioactive synthetic peptides for oral tissues regeneration. Front Mater. 2021;8:728213. doi:10.3389/fmats.2021.728213.
11. Gao W, Zhao J, Gui J, Wang Z, Chen J, Yue Z. Comprehensive Assessment of BERT-Based Methods for Predicting Antimicrobial Peptides. J Chem Inf Model. 2024;64(19):7772-7785. doi: 10.1021/acs.jcim.4c00507.
12. Lee H, Lee S, Lee I, Nam H. AMP-BERT: Prediction of antimicrobial peptide function based on a BERT model. Protein Sci. 2023;32(1):e4529. doi: 10.1002/pro.4529.

13. Manavalan B, Subramaniyam S, Shin TH, Kim MO, Lee G. Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy. J Proteome Res. 2018;17(8):2715-2726. doi: 10.1021/acs.jproteome.8b00148.
14. Zhu L, Chen Z, Yang S. EnDM-CPP: A Multi-view Explainable Framework Based on Deep Learning and Machine Learning for Identifying Cell-Penetrating Peptides with Transformers and Analyzing Sequence Information. Interdiscip Sci. 2024. doi: 10.1007/s12539-024-00673-4.
15. Zhang H, Zhang Y, Zhang C, Yu H, Ma Y, Li Z, Shi N. Recent Advances of Cell-Penetrating Peptides and Their Application as Vectors for Delivery of Peptide and Protein-Based Cargo Molecules. Pharmaceutics. 2023;15(8):2093. doi: 10.3390/pharmaceutics15082093.
16. Imre A, Balogh B, Mándity I. GraphCPP: The new state-of-the-art method for cell-penetrating peptide prediction via graph neural networks. Br J Pharmacol. 2025;182(3):495-509. doi: 10.1111/bph.17388.
17. Xu T, Wang Q, Yang Z, Ying J. A BERT-based approach for identifying anti-inflammatory peptides using sequence information. Heliyon. 2024;10(12):e32951. doi: 10.1016/j.heliyon.2024.e32951.
18. Long Y, Wu M, Liu Y, Fang Y, Kwoh CK, Chen J, Luo J, Li X. Pre-training graph neural networks for link prediction in biomedical networks. Bioinformatics. 2022;38(8):2254-2262. doi: 10.1093/bioinformatics/btac100.
19. Qiao Y, Zhu X, Gong H. BERT-Kcr: prediction of lysine crotonylation sites by a transfer learning method with pre-trained BERT models. Bioinformatics. 2022;38(3):648-654. doi: 10.1093/bioinformatics/btab712.
20. Qasim R, Bangyal WH, Alqarni MA, Ali Almazroi A. A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification. J Healthc Eng. 2022;2022:3498123. doi: 10.1155/2022/3498123.