Identification of Antimicrobial Peptides and Their Functional Activities Using Ensemble Learning and Amino Acid Encoding Methods

Zhirui Liu *

The Webb Schools, LA, US

* Corresponding Author Email: rliuzhirui5151@gmail.com

Abstract. Antimicrobial resistance (AMR) has become a significant clinical crisis worldwide, resulting in substantial economic losses and posing a threat to human health. Because antimicrobial peptides (AMPs) can interact with membrane bilayers and inhibit microbial growth, they represent a promising therapeutic candidate to overcome AMR. Despite their potential, AMP development is strongly constrained due to the high costs and inefficiencies in identifying novel and effective AMPs, particularly those with specific functional activities against diverse microbial species. To address these limitations, I developed a comprehensive ensemble learning framework for both AMP identification and functional activity prediction. This approach integrates four amino acid encoding methods (one-hot encoding, BLOSUM62 substitution matrix, AAIndex physicochemical properties, and pseudo amino acid composition) with ensemble learning algorithms to predict 14 different antimicrobial activities, including antibacterial, antifungal, antiviral, anticancer, and other specialized functions. I also evaluated various advanced machine learning algorithms across different encoding approaches and built comprehensive model combinations for each task. The experimental results showed that my ensemble learning-based model achieved the highest performance, with an accuracy of 97.66% for AMR prediction. Additionally, the ensemble approach demonstrated superior performance compared to other individual models. It is able to achieve balanced accuracy across various functional activity tasks. In conclusion, my proposed method offers a practical approach for AMP discovery, helping to address the AMR crisis.

Keywords: Antimicrobial Peptides, Functional Activities Identification, Ensemble Learning, Machine Learning, Amino Acid Encoding.

1. Introduction

The discovery and commercialization of ampicillin in 1958 and 1961 marked a significant breakthrough in fighting against bacterial infections and saved countless lives worldwide. However, the rapid emergence of antimicrobial resistance (AMR) has become a global health crisis, resulting in significantly increased mortality rates and placing a substantial economic burden on healthcare systems [1–3]. It is predicted that the AMR will surpass cancer as the most serious public health challenge by 2050 [1,2,4]. In this context, antimicrobial peptides (AMPs) have emerged as a promising therapeutic strategy against AMR. AMPs are small peptides that exist primarily in nature, distributed across diverse organisms, such as bacteria, fungi, animals, and plants. They range from 10 to 50 amino acids in length and are characterized by their cationic and amphiphilic α-helical structure.

The unique combination of amino acid composition, amphipathicity, cationic charge, and size enables AMPs to interact with bacterial cell membranes through 'barrel-stave', 'carpet', or 'toroidal-pore' mechanisms. By targeting the negatively charged bacterial cell membranes, AMPs disrupt the integrity of these membranes. This allows larger molecules to penetrate the cell, ultimately leading to cell death through interactions with intracellular DNA and RNA [5–11].

Compared to traditional antibiotics, AMPs offer several advantages. First, AMPs **exhibit broad-spectrum antimicrobial activity**. Due to their diverse molecular structures and mechanisms, they effectively target not only Gram-positive bacteria and Gram-negative bacteria, but also fungi, as well as some viruses and parasites. **Second, the unique mechanism of AMPs, which primarily involves direct interaction with microbial cell membranes through electrostatic attraction, makes it**

difficult for microorganisms to develop resistance. Third, **many AMPs can enhance the host immune system** and provide additional therapeutic benefits beyond their direct antimicrobial effects. These comprehensive advantages position AMPs as valuable candidates for future biomedical applications, particularly in addressing AMR challenges.

Nonetheless, the development and clinical application of AMPs is strongly constrained by high **production costs** and **significant technical challenges.** The development process requires extensive experimental work to screen and evaluate candidate molecules, assessing their antimicrobial activity, toxicity, and stability. This complex screening process is both time-consuming and rarely successful, so it is highly challenging to develop reliable AMPs suitable for clinical application [9,11–17].

To address this issue, computational methods offer a cost-effective and accurate approach to identify promising AMPs and accelerate the development of antimicrobial drugs for clinical use. Up until now, several machine learning-based models have been developed for AMP identification, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Extremely Randomized Trees (ExtraTree), among others. They have demonstrated excellent performance and achieved significant success in AMP and their activity identification [10,11,14,16–18]. Apart from these, deep learning methods, such as ANN and CNN, have also been introduced into AMP prediction. Notable examples include AMPDiscover and multi-AMP [5,6,15,17–26]. These methods have provided significant assistance in accelerating the discovery of AMPs, especially for those with diverse activities. However, these methods also have some limitations. They primarily focused on enhancing AMP and activity prediction performance from an algorithmic perspective.

Given that AMPs are very short amino acid peptide chains, their feature extraction methods have a significant impact on model prediction accuracy. With this in mind, this study comprehensively compares four amino acid encoding methods for multi-task AMP prediction: one-hot encoding [26], BLOSUM62 substitution matrix [27], AAIndex physicochemical properties [28], and pseudo amino acid composition (PAAC) [29]. These encoding methods are evaluated using a range of advanced machine learning algorithms, from traditional methods to deep learning approaches.

Moreover, I developed a novel CNN-based ensemble learning framework that integrates these encoding methods through individual CNN models and combines their outputs via a sophisticated voting mechanism. This design ensures robust and reliable predictions across a wide range of antimicrobial activities.

Overall, my proposed multi-task ensemble learning framework serves as a comprehensive computational pipeline for both AMP identification and functional activity prediction. It offers significant advantages over single-task approaches and holds substantial value for accelerating antimicrobial drug discovery in the era of growing antimicrobial resistance.

2. Materials and Methods

2.1. AMP dataset collection

The AMP data sequences were downloaded from the public repository [26]. The dataset consists of a total of 49,115 positive AMP samples and 195,525 negative AMP samples. Expanding the scope of the study, I utilized a diverse range of AMP functional activity datasets, including antibacterial, antiviral, anti-gram-positive, anti-gram-negative, insecticidal, anti-mammalian cells, antibiofilm, antifungal, anticancer, antihIV, endotoxin, antiparasitic, and anti-MRSA. To reduce sequence redundancy, the CD-HIT tool was applied to eliminate sequences with greater than 40% pairwise identity between the positive and negative dataset samples. Then, the samples are divided into a training set, a test set, and an independent test set for validation of the computational models.

2.2. Methods

2.2.1 Amino Acid Sequence Encoding Methods

The AMP sequences comprise multiple amino acids that can be represented using the 20 distinct letters of the genetic code. To enable effective analysis through machine learning and deep learning prediction models, these sequences must be converted into numerical formats through encoding approaches. In this study, I implemented four distinct encoding methods to capture different aspects of amino acid sequence information comprehensively. All four encoding methods are compared across different machine learning algorithms, and then integrated into an ensemble learning framework to evaluate their respective performance in predicting AMP functional activity.

(1) One-hot encoding: One-hot encoding represents each amino acid as a **20-dimensional binary vector** with only one element set to 1, generating an L × 20 matrix for sequences of length L. This method preserves **exact positional information and treats each amino acid as an independent categorical variable**, making it **particularly suitable for convolutional neural networks**. However, it does not capture biological similarities between chemically related amino acids, treating all residues as completely distinct entities [26].

(2) BLOSUM62 matrix encoding: BLOSUM62 encoding replaces **each amino acid with its corresponding 20-dimensional substitution vector from the BLOSUM62 matrix that reflects evolutionary relationships and substitution probabilities between amino acids.** This approach captures functional similarities and conservation patterns, where chemically similar residues have higher substitution scores. Unlike one-hot encoding, **BLOSUM62 provides a continuous numerical representation that embeds evolutionary knowledge crucial for antimicrobial peptide prediction [27].**

(3) AAIndex physicochemical property encoding: AAIndex encoding represents each **amino acid using five key physicochemical properties: hydrophobicity, molecular volume, charge, polarity, and accessible surface area.** These properties directly relate to antimicrobial mechanisms, including membrane interaction, penetration ability, and electrostatic binding to bacterial surfaces. This encoding reduces dimensionality while capturing biologically relevant features that govern AMP activity based on underlying biophysical principles [28].

(4) Pseudo Amino Acid Composition (PAAC) encoding: **PAAC combines normalized amino acid frequencies with sequence-order correlation features. It results in a fixed 40-dimensional vector (20 compositional + 20 correlation features with $\lambda = 10$).** This method captures both global compositional characteristics and local sequence patterns while maintaining consistent representation regardless of sequence length. PAAC is particularly effective for shorter peptides where compositional features may be more informative than detailed positional information [29].

2.2.2 CNN Ensemble learning framework

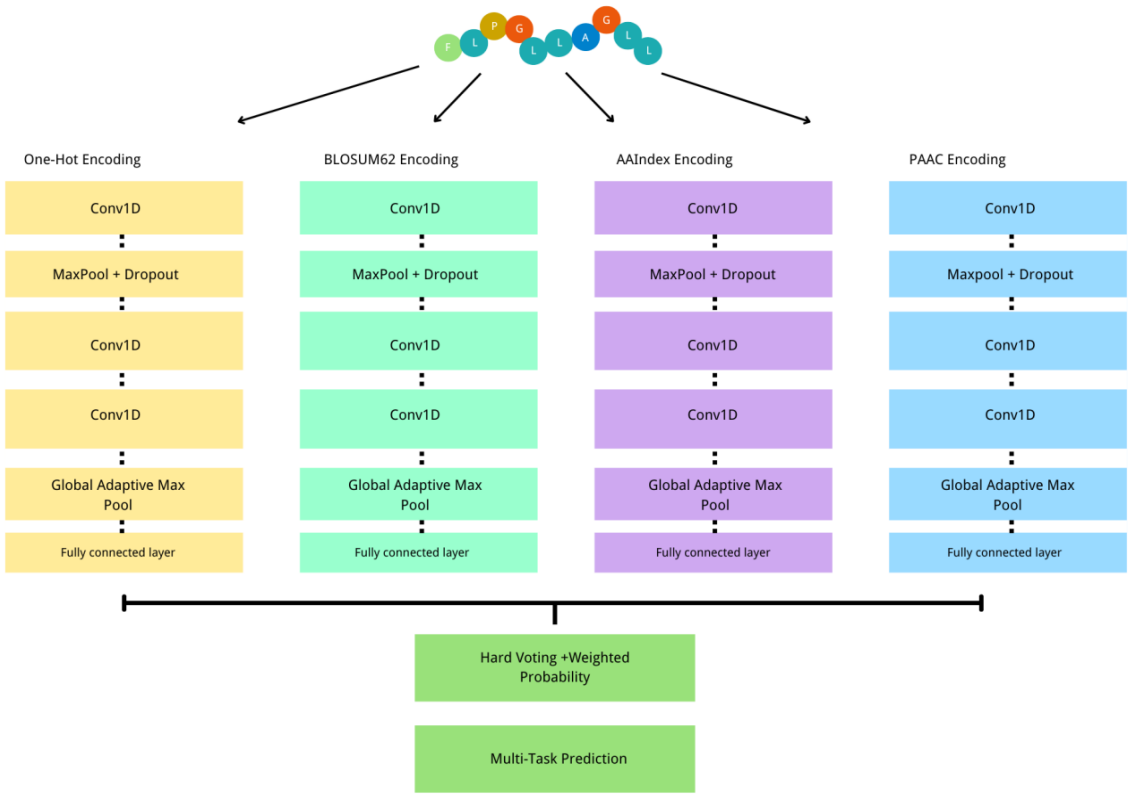

Figure. 1 Figure1

The CNN ensemble learning framework is designed to leverage the complementary strengths of different amino acid encoding methods through specialized convolutional neural networks. The ensemble consists of four encoding-specific CNN models, each optimized for a particular sequence representation: *CNN_OneHot for one-hot encoding, CNN_BLOSUM62 for evolutionary substitution matrices, CNN_AAIndex for physicochemical properties, and CNN_PAAC for pseudo amino acid composition.*

Each CNN architecture employs a multi-scale convolutional design with three parallel convolutional branches, utilizing kernel sizes of 3, 5, and 7 to capture local sequence patterns at different scales. Next, the model applies batch normalization, LeakyReLU activation, and max pooling operations. The multi-scale features are concatenated and processed through additional convolutional layers (64→128→256 channels), followed by global adaptive max pooling and a series of fully connected layers (256→64→32→1). Dropout regularization (p=0.5) is applied to prevent overfitting. To address the extreme class imbalance inherent in antimicrobial peptide datasets, the framework incorporates Focal Loss ($\alpha = 1$, $\gamma = 2$) instead of the standard binary cross-entropy, which focuses learning on hard-to-classify minority samples while downweighting easy negatives.

Finally, the ensemble prediction strategy employs a consensus-based voting mechanism where positive classification requires agreement from at least two models (MIN_POSITIVE_PREDICTIONS=2). This approach effectively reduces false positives while maintaining sensitivity for detecting antimicrobial activities. Model weights are calculated based on validation performance using a composite score ($0.4 \times F1 + 0.3 \times Recall + 0.2 \times Balanced\ Accuracy + 0.1 \times AUC$), which prioritizes the detection of the positive class. Final predictions combine both hard voting for binary decisions and weighted averaging for probability estimates.

2.3. Other machine learning methods

To comprehensively evaluate my CNN ensemble approach, I implemented seven additional machine learning algorithms as baseline comparisons across all four encoding methods.

Traditional models included Logistic Regression with L2 regularization, Lasso regression with L1 regularization for feature selection, K-Nearest Neighbors (k=5) for instance-based learning, Decision Tree (max depth=10) for rule-based classification, Random Forest (100 estimators) for ensemble learning, Support Vector Machine with RBF kernel for non-linear classification, and Multi-Layer Perceptron with two hidden layers (64, 32 neurons) as a deep learning baseline. All models utilized **standardized features through StandardScaler preprocessing and default parameters without hyperparameter optimization to ensure fair comparisons**. **This systematic evaluation across 32 combinations (4 encodings × 8 algorithms) per task enabled the identification of optimal approaches.** It demonstrated the superior performance of my CNN ensemble framework compared to conventional machine learning methods.

3. Results

3.1. AMP prediction analysis

Amino acid compositional patterns between AMPs and non-AMPs. As shown in Figure 2, it demonstrates distinct amino acid compositional patterns between AMPs and non-AMPs. AMPs exhibit significantly higher frequencies of cationic amino acids (lysine and arginine) and hydrophobic residues, while showing substantially lower frequencies of acidic amino acids (aspartate and glutamate). The difference analysis reveals that lysine, cysteine, and arginine are the most enriched amino acids in AMPs, with frequency differences exceeding 2%. The scatter plot confirms systematic compositional divergence as the amino acids in the graph fall below the equal frequency line. These compositional biases reflect the fundamental physicochemical requirements for membrane interaction and antimicrobial activity. It validates the importance of amino acid encoding methods that capture these distinctive chemical signatures for effective AMP prediction.

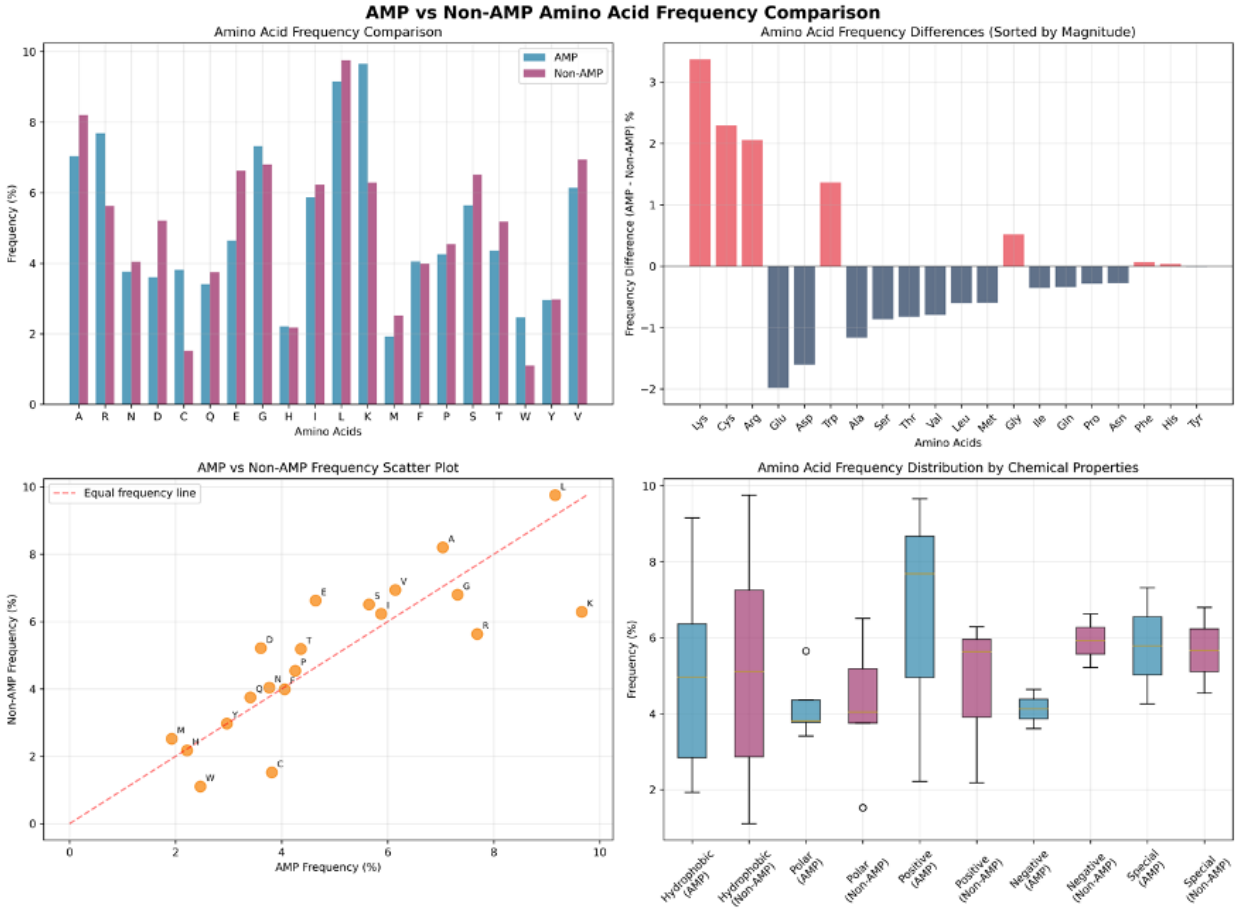

Figure. 2 Amino acid composition comparison between AMPs and non-AMPs

T-SNE visualizations of different amino acid encoding methods between AMPs and non-AMPs. Figure 3 presents t-SNE visualizations comparing the discriminative power of four amino acid encoding methods using 2000 randomly selected sequences. The results demonstrate varying degrees of class separability. With AAIndex encoding achieving the highest separation score (32.35) and superior clustering quality (0.06), the result indicates optimal discrimination between AMP and non-AMP sequences based on physicochemical properties. PAAC encoding shows moderate separation (17.25) with clear but overlapping clusters, while One-hot and BLOSUM62 encodings exhibit lower separation scores (10.88 and 4.94, respectively) with substantial class overlap. The visualization confirms that encoding methods incorporating biological and chemical knowledge (AAIndex and PAAC) provide better feature representations for AMP classification compared to purely structural encodings (One-hot and BLOSUM62), supporting the rationale for ensemble approaches that leverage multiple encoding perspectives.

Figure. 3 t-SNE visualization comparison of different amino acid encoding methods for AMP and non-AMP classification

AMP prediction model performance analysis. I summarize the classification performance of our CNN ensemble model on the binary AMP prediction task using 14,902 test sequences. The model demonstrates excellent overall performance, achieving 97.66% accuracy, which effectively distinguishes between AMPs and non-AMPs. For non-AMP classification, the model achieves high precision (98.52%) and recall (98.21%), indicating robust identification of negative samples. In AMP classification, the model also exhibits strong performance, with 95.50% precision and 96.25% recall, demonstrating effective identification of antimicrobial peptides while keeping the false-positive rate low.

Table. 1 Performance evaluation of the CNN ensemble model for binary AMP classification

	Precision	Recall	f1-score	support
Non-AMP	0.9852	0.9821	0.9837	10691
AMP	0.9550	0.9625	0.9587	4211
Accuracy			0.9766	14902
Macro Avg	0.9701	0.9723	0.9712	14902
Weighted Avg	0.9766	0.9766	0.9766	14902

3.2. AMP functional activities performance.

CNN ensemble framework-based AMP prediction analysis. The comprehensive evaluation of the CNN ensemble framework against machine learning approaches reveals significant performance advantages across multiple antimicrobial activity prediction tasks. Figure 4 presents a direct comparison between the ensemble model and the best-performing individual algorithm for each of the 14 functional activities. This demonstrates the superior and consistent performance of the ensemble approach. **The results show that the CNN ensemble model achieves higher balanced accuracy in 11 out of 14 tasks, with particularly notable improvements in antiviral (0.764 vs 0.734), antibiofilm (0.730 vs 0.696), and antihiv (0.695 vs 0.598) prediction tasks. The ensemble model demonstrates remarkable consistency across different activity types, maintaining balanced accuracy above 0.55 for all tasks. In contrast, individual algorithms show considerable variability in performance.**

Specifically, the antiviral prediction task represents the most successful application of our ensemble approach. It achieved a balanced accuracy of 0.764 compared to 0.734 for the best individual algorithm. The 3.0% improvement is particularly significant given the complexity of antiviral peptide mechanisms. The antibiofilm task shows a substantial improvement (0.730 vs 0.696), reflecting the ensemble's ability to capture the complex physicochemical requirements for biofilm disruption through multiple mechanisms. This includes membrane disruption and interference with bacterial communication. Most notably, the antihiv task demonstrates the largest absolute improvement, with a balanced accuracy of 0.695 compared to 0.589. This 16.7% increase highlights the particular value of ensemble learning for highly specialized antimicrobial activities. The superior performance in antihiv prediction likely stems from the ensemble's ability to leverage multiple feature types simultaneously.

By combining positional information (one-hot encoding), evolutionary conservation (BLOSUM62), physicochemical properties (AAIndex), and compositional features (PAAC), the model captures a more complete picture of the sequence-function relationship. Each encoding method contributes complementary insights, enabling more accurate identification of features essential for HIV inhibition.

Finally, the results also reveal tasks where the ensemble improvement is minimal. For example, the anticancer, endotoxin, and chemotactic tasks show a slight increase or decrease (0.623 vs 0.618, 0.619 vs 0.562, 0.585 vs 0.636, respectively). This suggests that for some activities, the best individual algorithm-encoding combination may already capture most of the relevant sequence information.

The performance analysis of encoding methods provides a comprehensive evaluation of all 32 encoding-algorithm combinations across tasks, offering insight into which approaches are best suited for specific antimicrobial activities. As shown in Figure 5, this analysis reveals essential patterns in performance that highlight the strengths and limitations of different encoding strategies.

Specifically, it demonstrates clear task-specific preferences for certain encoding-algorithm combinations, with darker green regions indicating superior performance. For example, antibacterial prediction consistently demonstrates strong performance across multiple combinations, particularly with AAIndex-based encodings and tree-based algorithms, such as Random Forest or Decision Tree.

Furthermore, AAIndex encoding consistently yields competitive results across most algorithms and tasks, reinforcing our hypothesis that physicochemical properties provide a robust foundation for antimicrobial activity prediction. These advantages suggest that physicochemical properties are crucial for modeling general antibacterial activity.

Conversely, PAAC encoding shows more variable performance. While PAAC excels in certain tasks (particularly those with shorter sequences or limited training data), it underperforms in others. This variability aligns with the theoretical foundation of PAAC, which emphasizes compositional features that may be more informative for certain types of antimicrobial mechanisms.

Regarding the machine learning algorithms, Random Forest consistently achieves competitive performance across most encoding types. This reflects its robustness in handling high-dimensional feature spaces and capturing complex nonlinear relationships.

Support Vector Machines show particular strength with standardized features from AAIndex and PAAC encodings, but perform less favorably with the high-dimensional representations from one-hot and BLOSUM62 encodings. Neural network approaches (both ANN and CNN) demonstrate superior performance with raw sequence-based encodings (one-hot and BLOSUM62). This highlights their ability to learn hierarchical feature representations from low-level sequence information.

The task-specific performance patterns suggest that **different antimicrobial activities require different computational approaches. With some benefiting from the pattern recognition capabilities of deep learning methods, others respond better to the explicit feature engineering of traditional machine learning approaches combined with biologically informed encodings.**

Finally, these comparative analyses demonstrate that our CNN ensemble framework achieves superior and more consistent performance compared to individual algorithms across the majority of antimicrobial activity prediction tasks. The systematic evaluation of 32 algorithm-encoding combinations per task reveals essential insights into the suitability of different computational approaches for specific antimicrobial activities, and it confirms the value of ensemble learning for capturing the complex and diverse requirements of antimicrobial peptide function prediction.

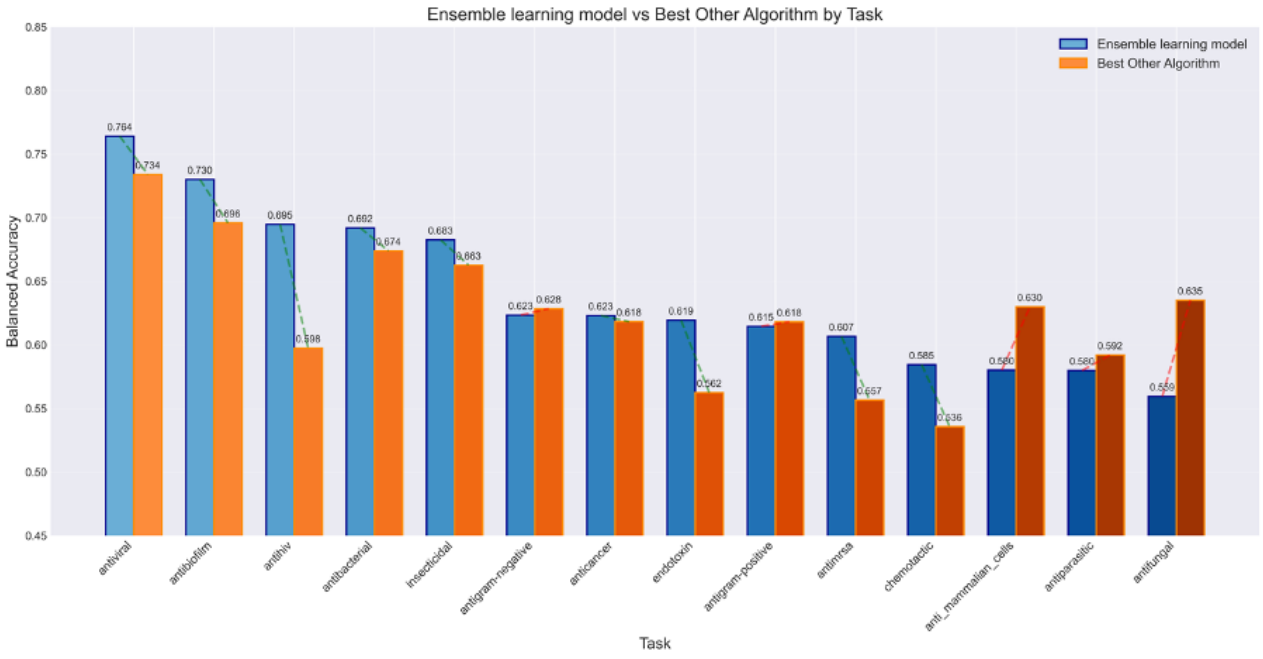

Figure. 4 Performance comparison between the CNN ensemble model and the best individual algorithms across 14 antimicrobial activity prediction tasks

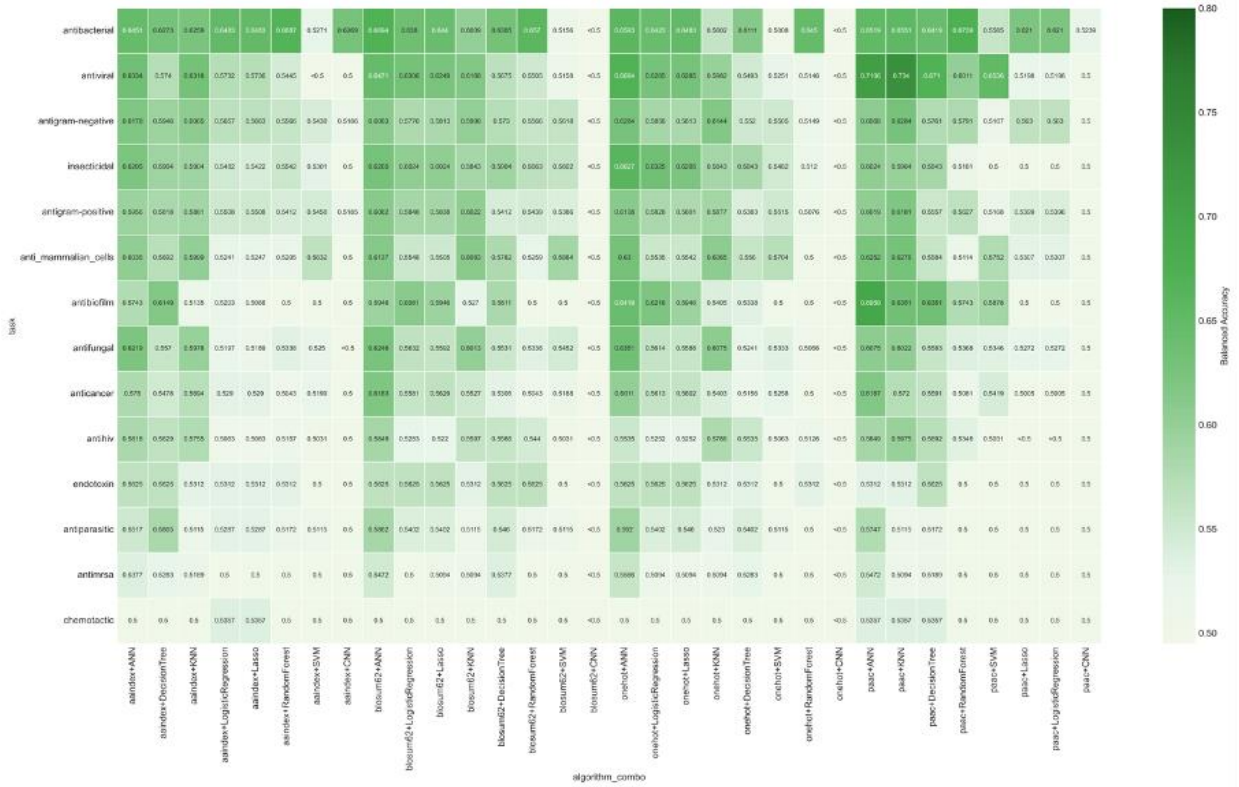

Figure. 5 Comprehensive performance heatmap of encoding-algorithm combinations across all antimicrobial activity prediction tasks

4. Conclusion

This study presents a comprehensive CNN ensemble learning framework-based model for AMP identification and functional activity prediction. It addresses the critical need for computational tools in combating the AMR phenomenon.

Firstly, the amino acid composition analysis confirms diverse biological features between AMPs and non-AMPs. AMPs contain significant enrichment in cationic residues (lysine, arginine) and depletion in acidic amino acids (aspartate, glutamate). Then, the t-SNE visualization demonstrates that four distinct amino acid encoding methods (one-hot, BLOSUM62, AAIndex, and PAAC) can extract the significant AMP-associated features. This supports our multi-encoding ensemble design.

Secondly, I integrate these four amino acid encoding methods within a voting-based ensemble learning strategy utilizing CNN. Each network is explicitly designed for one encoding embedding. This ensemble approach not only predicts AMPs but also their 14 different antimicrobial activities. The model demonstrates outstanding predictive performance, achieving an accuracy of 97.66%, which indicates a robust ability to discriminate between AMPs and non-AMPs. Notably, the comparative analysis across 32 algorithm-encoding combinations demonstrates the superior performance of my ensemble approach. It achieves a higher balanced accuracy in 11 out of 14 tasks compared to the best individual machine learning and deep learning methods.

While the approach demonstrates significant improvements, certain limitations warrant consideration. Future work could explore advanced data augmentation techniques, transfer learning approaches for data-limited tasks, and integration of additional biological information, such as structural features or experimental activity data.

In conclusion, our CNN ensemble learning framework provides a comprehensive, accurate, and practically applicable solution for multi-task antimicrobial peptide prediction. The systematic integration of multiple encoding perspectives through ensemble learning establishes a new benchmark for computational AMP discovery. This approach holds strong potential to accelerate the development of novel antimicrobial therapeutics in the fight against antimicrobial resistance. Overall,

this computational pipeline represents an essential step toward more efficient and cost-effective antimicrobial peptide discovery, addressing one of the most pressing challenges in modern medicine.

Acknowledgements

I express my fullest gratitude to Dr. Garrett Pendergraft and Dr. Tian for their time and guidance that made this paper possible.

References

[1] C.J.L. Murray, K.S. Ikuta, F. Sharara, L. Swetschinski, G. Robles Aguilar, A. Gray, C. Han, C. Bisignano, P. Rao, E. Wool, S.C. Johnson, A.J. Browne, M.G. Chipeta, F. Fell, S. Hackett, G. Haines-Woodhouse, B.H. Kashef Hamadani, E.A.P. Kumaran, B. McManigal, S. Achalapong, R. Agarwal, S. Akech, S. Albertson, J. Amuasi, J. Andrews, A. Aravkin, E. Ashley, F.-X. Babin, F. Bailey, S. Baker, B. Basnyat, A. Bekker, R. Bender, J.A. Berkley, A. Bethou, J. Bielicki, S. Boonkasidecha, J. Bukosia, C. Carvalheiro, C. Castañeda-Orjuela, V. Chansamouth, S. Chaurasia, S. Chiurchiù, F. Chowdhury, R. Clotaire Donatien, A.J. Cook, B. Cooper, T.R. Cressey, E. Criollo-Mora, M. Cunningham, S. Darboe, N.P.J. Day, M. De Luca, K. Dokova, A. Dramowski, S.J. Dunachie, T. Duong Bich, T. Eckmanns, D. Eibach, A. Emami, N. Feasey, N. Fisher-Pearson, K. Forrest, C. Garcia, D. Garrett, P. Gastmeier, A.Z. Giref, R.C. Greer, V. Gupta, S. Haller, A. Haselbeck, S.I. Hay, M. Holm, S. Hopkins, Y. Hsia, K.C. Iregbu, J. Jacobs, D. Jarovsky, F. Javanmardi, A.W.J. Jenney, M. Khorana, S. Khusuwan, N. Kissoon, E. Kobeissi, T. Kostyanev, F. Krapp, R. Krumkamp, A. Kumar, H.H. Kyu, C. Lim, K. Lim, D. Limmathurotsakul, M.J. Loftus, M. Lunn, J. Ma, A. Manoharan, F. Marks, J. May, M. Mayxay, N. Mturi, T. Munera-Huertas, P. Musicha, L.A. Musila, M.M. Mussi-Pinhata, R.N. Naidu, T. Nakamura, R. Nanavati, S. Nangia, P. Newton, C. Ngoun, A. Novotney, D. Nwakanma, C.W. Obiero, T.J. Ochoa, A. Olivas-Martinez, P. Olliaro, E. Ooko, E. Ortiz-Brizuela, P. Ounchanum, G.D. Pak, J.L. Paredes, A.Y. Peleg, C. Perrone, T. Phe, K. Phommasone, N. Plakkal, A. Ponce-de-Leon, M. Raad, T. Ramdin, S. Rattanavong, A. Riddell, T. Roberts, J.V. Robotham, A. Roca, V.D. Rosenthal, K.E. Rudd, N. Russell, H.S. Sader, W. Saengchan, J. Schnall, J.A.G. Scott, S. Seekaew, M. Sharland, M. Shivamallappa, J. Sifuentes-Osornio, A.J. Simpson, N. Steenkeste, A.J. Stewardson, T. Stoeva, N. Tasak, A. Thaiprakong, G. Thwaites, C. Tigoi, C. Turner, P. Turner, H.R. van Doorn, S. Velaphi, A. Vongpradith, M. Vongsouvath, H. Vu, T. Walsh, J.L. Walson, S. Waner, T. Wangrangsimakul, P. Wannapinij, T. Wozniak, T.E.M.W. Young Sharma, K.C. Yu, P. Zheng, B. Sartorius, A.D. Lopez, A. Stergachis, C. Moore, C. Dolecek, M. Naghavi, Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis, The Lancet 399 (2022) 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0.

[2] C.S. Ho, C.T.H. Wong, T.T. Aung, R. Lakshminarayanan, J.S. Mehta, S. Rauz, A. McNally, B. Kintses, S.J. Peacock, C. de la Fuente-Nunez, R.E.W. Hancock, D.S.J. Ting, Antimicrobial resistance: a concise update, Lancet Microbe 6 (2025) 100947. https://doi.org/10.1016/j.lanmic.2024.07.010.

[3] I.N. Okeke, M.E.A. de Kraker, T.P. Van Boeckel, C.K. Kumar, H. Schmitt, A.C. Gales, S. Bertagnolio, M. Sharland, R. Laxminarayan, The scope of the antimicrobial resistance challenge, The Lancet 403 (2024) 2426–2438. https://doi.org/10.1016/S0140-6736(24)00876-6.

[4] Md.A. Salam, Md.Y. Al-Amin, M.T. Salam, J.S. Pawar, N. Akhter, A.A. Rabaan, M.A.A. Alqumber, Antimicrobial Resistance: A Growing Serious Threat for Global Public Health, Healthcare 11 (2023) 1946. https://doi.org/10.3390/healthcare11131946.

[5] S. Panwar, M. Thapliyal, V. Kuriyal, V. Tripathi, A. Thapliyal, GEU-AMP50: Enhanced antimicrobial peptide prediction using a machine learning approach, Mater Today Proc 73 (2023) 81–87. https://doi.org/10.1016/j.matpr.2022.09.326.

[6] Y. Lin, Y. Cai, J. Liu, C. Lin, X. Liu, an advanced approach to identify antimicrobial peptides and their function types for penaeus through machine learning strategies, BMC Bioinformatics 20 (2019) 291. https://doi.org/10.1186/s12859-019-2766-9.

[7] C.D. Santos-Júnior, M.D.T. Torres, Y. Duan, Á. Rodríguez del Río, T.S.B. Schmidt, H. Chong, A. Fullam, M. Kuhn, C. Zhu, A. Houseman, J. Somborski, A. Vines, X.-M. Zhao, P. Bork, J. Huerta-Cepas, C. de la Fuente-Nunez, L.P. Coelho, Discovery of antimicrobial peptides in the global microbiome with machine learning, Cell 187 (2024) 3761-3778.e16. https://doi.org/10.1016/j.cell.2024.05.013.

[8] S. Zhang, Y. Sun, K. Yin, J. Zhang, L. Du, S. Wang, D. Zheng, R. Li, ML-AMPs designed through machine learning show antifungal activity against C. albicans and therapeutic potential on mice model with candidiasis, Life Sci 366–367 (2025) 123485. https://doi.org/10.1016/j.lfs.2025.123485.

[9] F. Wan, F. Wong, J.J. Collins, C. de la Fuente-Nunez, Machine learning for antimicrobial peptide identification and design, Nature Reviews Bioengineering 2 (2024) 392–407. https://doi.org/10.1038/s44222-024-00152-x.

[10] C. Bournez, M. Riool, L. de Boer, R.A. Cordfunke, L. de Best, R. van Leeuwen, J.W. Drijfhout, S.A.J. Zaat, G.J.P. van Westen, CalcAMP: A New Machine Learning Model for the Accurate Prediction of Antimicrobial Activity of Peptides, Antibiotics 12 (2023) 725. https://doi.org/10.3390/antibiotics12040725.

[11] K. Musin, E. Asyanova, How Machine Learning Helps in Combating Antimicrobial Resistance: A Review of AMP Analysis and Generation Methods, Int J Pept Res Ther 31 (2025) 59. https://doi.org/10.1007/s10989-025-10716-z.

[12] A. Cesaro, M. Bagheri, M. Torres, F. Wan, C. de la Fuente-Nunez, Deep learning tools to accelerate antibiotic discovery, Expert Opin Drug Discov 18 (2023) 1245–1257. https://doi.org/10.1080/17460441.2023.2250721.

[13] J.M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N.M. Donghia, C.R. MacNair, S. French, L.A. Carfrae, Z. Bloom-Ackermann, V.M. Tran, A. Chiappino-Pepe, A.H. Badran, I.W. Andrews, E.J. Chory, G.M. Church, E.D. Brown, T.S. Jaakkola, R. Barzilay, J.J. Collins, A Deep Learning Approach to Antibiotic Discovery, Cell 180 (2020) 688-702.e13. https://doi.org/10.1016/j.cell.2020.01.021.

[14] J. Yan, J. Cai, B. Zhang, Y. Wang, D.F. Wong, S.W.I. Siu, Recent Progress in the Discovery and Design of Antimicrobial Peptides Using Traditional Machine Learning and Deep Learning, Antibiotics 11 (2022) 1451. https://doi.org/10.3390/antibiotics11101451.

[15] T.-T. Lin, L.-Y. Yang, I.-H. Lu, W.-C. Cheng, Z.-R. Hsu, S.-H. Chen, C.-Y. Lin, AI4AMP: an Antimicrobial Peptide Predictor Using Physicochemical Property-Based Encoding Method and Deep Learning, MSystems 6 (2021). https://doi.org/10.1128/mSystems.00299-21.

[16] G. Wang, I.I. Vaisman, M.L. van Hoek, Machine Learning Prediction of Antimicrobial Peptides, in: 2022: pp. 1–37. https://doi.org/10.1007/978-1-0716-1855-4_1.

[17] J. Xu, F. Li, A. Leier, D. Xiang, H.-H. Shen, T.T. Marquez Lago, J. Li, D.-J. Yu, J. Song, Comprehensive assessment of machine learning-based methods for predicting antimicrobial peptides, Brief Bioinform 22 (2021). https://doi.org/10.1093/bib/bbab083.

[18] B. Vishnepolsky, M. Grigolava, G. Managadze, A. Gabrielian, A. Rosenthal, D.E. Hurt, M. Tartakovsky, M. Pirtskhalava, Comparative analysis of machine learning algorithms on the microbial strain-specific AMP prediction, Brief Bioinform 23 (2022). https://doi.org/10.1093/bib/bbac233.

[19] B. Olcay, G.D. Ozdemir, M.A. Ozdemir, U.K. Ercan, O. Guren, O. Karaman, Prediction of the synergistic effect of antimicrobial peptides and antimicrobial agents via supervised machine learning, BMC Biomed Eng 6 (2024) 1. https://doi.org/10.1186/s42490-024-00075-z.

[20] Y. Wang, L. Wang, C. Li, Y. Pei, X. Liu, Y. Tian, AMP-EBiLSTM: employing novel deep learning strategies for the accurate prediction of antimicrobial peptides, Front Genet 14 (2023). https://doi.org/10.3389/fgene.2023.1232117.

[21] C.M. Van Oort, J.B. Ferrell, J.M. Remington, S. Wshah, J. Li, AMPGAN v2: Machine Learning-Guided Design of Antimicrobial Peptides, J Chem Inf Model 61 (2021) 2198–2207. https://doi.org/10.1021/acs.jcim.0c01441.

[22] S. Lertampaiporn, T. Vorapreeda, A. Hongsthong, C. Thammarongtham, Ensemble-AMPPred: Robust AMP Prediction and Recognition Using the Ensemble Learning Method with a New Hybrid Feature for Differentiating AMPs, Genes (Basel) 12 (2021) 137. https://doi.org/10.3390/genes12020137.

[23] J. Yan, P. Bhadra, A. Li, P. Sethiya, L. Qin, H.K. Tai, K.H. Wong, S.W.I. Siu, Deep-AmPEP30: Improve Short Antimicrobial Peptides Prediction with Deep Learning, Mol Ther Nucleic Acids 20 (2020) 882–894. https://doi.org/10.1016/j.omtn.2020.05.006.

[24] A. Jan, M. Hayat, M. Wedyan, R. Alturki, F. Gazzawe, H. Ali, F.K. Alarfaj, Target-AMP: Computational prediction of antimicrobial peptides by coupling sequential information with evolutionary profile, Comput Biol Med 151 (2022) 106311. https://doi.org/10.1016/j.compbiomed.2022.106311.

[25] A. Khorshidi, A.A. Peterson, Amp: A modular approach to machine learning in atomistic simulations, Comput Phys Commun 207 (2016) 310–324. https://doi.org/10.1016/j.cpc.2016.05.010.

[26] J. Xu, F. Li, C. Li, X. Guo, C. Landersdorfer, H.-H. Shen, A.Y. Peleg, J. Li, S. Imoto, J. Yao, T. Akutsu, J. Song, iAMPCN: a deep-learning approach for identifying antimicrobial peptides and their functional activities, Brief Bioinform 24 (2023). https://doi.org/10.1093/bib/bbad240.

[27] M.P. Styczynski, K.L. Jensen, I. Rigoutsos, G. Stephanopoulos, BLOSUM62 miscalculations improve search performance, Nat Biotechnol 26 (2008) 274–275. https://doi.org/10.1038/nbt0308-274.

[28] S. Kawashima, AAindex: Amino Acid index database, Nucleic Acids Res 28 (2000) 374–374. https://doi.org/10.1093/nar/28.1.374.

[29] K.-C. Chou, Pseudo Amino Acid Composition and its Applications in Bioinformatics, Proteomics and System Biology, Curr Proteomics 6 (2009) 262–274. https://doi.org/10.2174/157016409789973707.

[30] David, Marian. Analyticity, Carnap, Quine, and Truth. 2022.

[31] Dewey, John. Reconstruction in Philosophy. Enl. ed., 15 pr, Beacon Press, 1972.

[32] Ebbs, G. "Carnap and Quine on Truth by Convention." Mind, vol. 120, no. 478, Apr. 2011, pp. 193–237. DOI.org (Crossref), https://doi.org/10.1093/mind/fzr020.

[33] Friedman, Michael. Dynamics of Reason: The 1999 Kant Lectures at Stanford University. CSLI Publications, 2001.

[34] Habermas, Jürgen, et al. Reason and the Rationalization of Society. Nachdr., Beacon, 2007.

[35] Husserl, Edmund. The Phenomenology of Internal Time-Consciousness. Edited by Martin Heidegger, Translated by James S. Churchill, Indiana University Press, 2019.

[36] Kant, Immanuel. Critique of Pure Reason. Edited by Marcus Weigelt, Translated by Friedrich Max Müller, Penguin Books, 2007.

[37] Quine, W. V. "Main Trends in Recent Philosophy: Two Dogmas of Empiricism." The Philosophical Review, vol. 60, no. 1, Jan. 1951, p. 20. DOI.org (Crossref), https://doi.org/10.2307/2181906.

[38] Quine, W. V. Pursuit of Truth. Rev. ed, Harvard University Press, 1992.

[39] Rosser, Barkley. "W. V. Quine. Truth by Convention. Philosophical Essays for Alfred North Whitehead, Longmans, Green and Co., New York1936, Pp. 90–124." Journal of Symbolic Logic, vol. 1, no. 1, Mar. 1936, pp. 42–42. DOI.org (Crossref), https://doi.org/10.2307/2269329.

[40] Soames, Scott. Analytic Philosophy in America: And Other Historical and Contemporary Essays. Princeton University Press, 2014. DOI.org (Crossref), https://doi.org/10.1515/9781400850464.

[41] Uebel, Thomas E., and Alan W. Richardson, editors. The Cambridge Companion to Logical Empiricism. Cambridge University Press, 2007.