Deciphering the Key Properties for Overcoming Bacterium-induced Drug Resistance and Potentiating Anti-cancer Immune Response

Bárbara Pinto Freitas, Maria Olívia Pereira, and Anália Lourenço

Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Abstract. Antimicrobial and cancer resistance represent two of the most significant global health challenges nowadays, with growing evidence suggesting interconnections between bacterial infections and cancer progression. There is an urgent need for innovative strategies that synthesically incorporate biological information and computational methodologies. This study aims to implement machine learning (ML) models to predict antimicrobial peptides (AMPs) with anticancer potential. This approach leverages positive and negative datasets to build comprehensive models, establishing a *in silico* framework to predict the anticancer activity of antimicrobial peptides as well as their antimicrobial properties, paving the way for novel investigations into the dual therapeutic potential of these peptides in infectious disease treatment and cancer therapy.

Keywords: Antimicrobial Resistance, Drug Discovery, Antimicrobial Peptides, Machine Learning

1 Introduction

1.1 Context and Motivation

The increase in antimicrobial and anticancer drug resistance poses a major public health threat, leading to a loss of effectiveness of current treatments [1]. Nowadays, the current global landscape is slightly different from the past because drug resistant pathogens and the connection between bacterial infections and cancer progression have significantly undermined traditional treatments and increased mortality rates. It is now more challenging than ever to introduce an effective antibiotic to the market considering the rapid emergence of antimicrobial resistance (AMR), which exceeds the rate of antibiotic drug discovery. As reported by the European Centre for Disease Prevention and Control [2], in the EU, 25000 people die annually due to infections caused by resistant bacteria. 10 million deaths per year are projected between the years 2015 and 2050 if current infection and resistance problems are not reversed. Hence, new approaches must be developed to accelerate the rate of drug discovery process and meet the current demand for new antibiotics [3]. Since recent advances in computational power and analysis of extensive biological datasets, ML protocols are significantly impacting the prediction of AMR patterns, serving as 'the bridge' between *in silico* predictions and clinical applications [4], creating unprecedented opportunities to develop treatments that can simultaneously combat bacterial infections while improving anticancer responses, offering hope for overcoming one of the most urgent public health challenges of our time.

1.2 Objectives

The main objective of this project is to develop computational tools to understand the key properties that make antimicrobial peptides overcome bacteria-induced drug resistance while boosting anticancer immune responses given a comprehensive dataset and, subsequently, a range of conventional ML models will be considered and the one with the best predictive performance will be selected. In short, finding a model capable of predicting the dual activity of the given candidates in the dataset.

2 Contextual Explanation

2.1 Antimicrobial Resistance

The overuse and misuse of antibiotics in clinical settings have accelerated resistance development in pathogenic microorganisms. Effective treatments are becoming increasingly obsolete, and common infections are more difficult and more expensive to treat. A key concern is the emergence of multidrug resistant pathogens, including members of the ESKAPE group (Enterococcus faecium, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.), frequently involved in hospital-acquired infections [5].

Bacteria can develop resistance to antimicrobial treatments through several mechanisms. At the molecular level, bacterial cells employ several strategies to evade antimicrobial activity [6]. A prevalent mechanism involves reducing the intracellular concentration of antimicrobial agents by promoting efflux pump activity [7]. In addition, bacteria can modify their molecular targets, preventing antimicrobial binding and neutralizing therapeutic efficacy. Another strategy involves enzymatic degradation or modification of antimicrobial agents, making them ineffective before they can reach their targets [8]. Bacteria can also form biofilms, a protective microenvironment in which they can persist despite antimicrobial exposure [9]. However, despite recent initiatives including global surveillance programs and stricter antimicrobial regulation, the development of new antimicrobial treatments has not kept pace with the evolution of resistance. Most newly approved antibiotics belong to existing classes, offering only incremental improvements rather than fundamentally new mechanisms of action [10]. Therefore, there is an urgent need to refocus on alternative therapeutic strategies to contain the spread of AMR. This includes investment in rapid diagnostics, better infection control practices, and novel research approaches to understand resistance evolution and transmission.

2.2 Antimicrobial Peptides

In contrast to some traditional treatments, AMPs have emerged as promising candidates to face the AMR crisis, as a result of their broad-range activity and mechanisms of action, bypassing conventional resistance strategies [11] [12]. These peptides are synthesized from prokaryotic entities to mammals and play a vital role as integral components of their innate immune defense mechanisms [13]. Several studies have revelead that certain cationic AMPs, in addition to their antimicrobial action, can be toxic to cancer cells with no adverse effect on normal cells [14] [15] [16], making them candidates as new tumor treatment options. Some AMPs properties, such as amphipathicity, moderate hydrophobicity, and positive charge, cause therapeutic activity in cancers [17]. These AMPs, that also induce toxicity in cancer cells, are called anticancer peptides. They were classified

by Boohaker et al. [18] into three classes based on their mechanisms of action: 1) Peptides that form pores, 2) Peptides than can cells, and 3) Peptides that target tumors.

Pore-forming peptides, such as Magainin II, are commonly a part of the primary immune defense system [19]. They induce necrosis or apoptosis in cancer cells. If negatively charged peptides attack cancer cells, they eventually lead to cellular lysis, then necrosis. Other ones disrupt the mitochondrial membrane and act through the induction of apoptosis [20]. The tumor targeting ones detect the integrin receptors on the cancer cell membrane in breast and lung cancer cells, brain tumors, and ovarian carcinoma cells [20]. Finally, cell penetrating peptides, such as BR2, developed by Lim et al. [21], have toxic effects in several in vitro cancer lines. These peptides have been used because they are involved in the delivery of anticancer drugs to cells [22].

In this way, AMPs represent a class of molecules with considerable therapeutic potential due to their dual ability to target both infectious agents and cancer cells. Their membrane disruption mechanisms, structural versatility, and selective cytotoxicity make them possible alternatives to conventional antibiotics and chemotherapeutics. This study builds on that potential by using computational tools to identify and predict AMPs with anticancer activity, with the aim of supporting the rational design of next-generation therapeutics.

2.3 Literature Review

Following PRISMA guidelines [23], a systematic review of the literature was performed, using keywords such as *antimicrobial*, *anticancer* and *AMP* to locate relevant studies published in the last few years. Table 1 lists the antimicrobial peptides with experimentally validated anticancer properties identified in the review. The peptides originate from various sources, with the alfa helix being the predominant structure because of its amphipathic nature.

Table 1. List of the 18 anticancer peptides (AMPs with anticancer activity), experimentally validated in the last few years, identified through a systematic literature review.

Peptide	Source	Sequence	Structure	Type of Cancer	Mechanism of Action
LTX-315 [24]	Synthetic	KKWWKWIRW	α-helix	Melanoma, Sarcoma	Induces tumor cell death and acts as an in situ vaccine
Lactoferricin [25]	Milk	FKCRRWQWRMKKLGAPSITCVRRAF	β-sheet	Breast carcinoma, Leukemia	Induction of apoptosis via membrane interaction
Tachyplesin I, II, III [26]	*Tachypleus tridentatus*	KWCFRVCYRGICYRRCR	β-hairpin	Melanoma, Cervical cancer	Membrane disruption
Cecropin B [14]	*Hyalophora cecropia*	WKVFKKIEKMGRNIRNGIVKAGPAIAVLGEAKALSYGLRPG	α-helix	Ovarian, Endometrial cancers	Inhibition of cell viability
Cathelicidin-BF [27]	*Bungarus fasciatus*	KFFRKLKKSVKKRAKEFFKKPRVIGVSIPF	α-helix	Cervical, Liver, Breast cancer	Membrane disruption, cationic targeting
TP4 [28]	*Oreochromis niloticus*	FIHHIIGGLFSAGKAIHRLIRRRRR	α-helix	Sarcoma	Calcium overload, ROS production, mitochondrial hyperpolarization
Magainin II [29]	*Xenopus laevis*	GIGKFLHSAKKFGKAFVGEIMNS	α-helix	Bladder cancer	Pore formation in membranes
Pleurodicin [30]	*Pleuronectes americanus*	GWGSFFKKAAHVGKHVGKAALTHYL	α-helix	Hepatocellular carcinoma, NSCLC	Targeting microbial membranes
Epinecidin-1 [31]	*Epinephelus coioides*	GFIFHIIKGLFHAGKMIHGLVTRRRH	α-helix	Fibrosarcoma	Membrane attack and immune regulation
Periplanetasin-5 [32]	*Periplaneta americana*	GNNRGRVRVGRVGRVAGVGRGG	α-helix	Leukemia	Induces apoptotic and necrotic death
D-LAK-120A [33]	Synthetic	d-KLKLKLALKALKAALKLA	α-helix	Glioblastoma, NSCLC, Melanoma	Pore formation and ROS increase
Scyreprocin [34]	*Scylla paramamosain*	LPSVGLKALKTGNTGVRVQVCVVTAA	α-helix	Lung cancer	Membrane disruption and apoptosis induction
BR2 [35]	Derived from Buforin IIb	RAGLQFPVG $[RLLR]_3$	α-helix	Colon, Cervical carcinoma	Apoptosis induction
Figainin 1 [36]	*Boana raniceps*	FIGTLIPLALGALTKLFK	α-helix	Breast, Colorectal, Lung cancer	Membrane disruption, apoptosis
palustrin-Ca [37]	*Lithobates catesbeianus*	GFLDIIKDTGKEFAVKILNNLKCKLAGGCPP	Turn/α-helix	Cervical cancer	Membrane disruption
A4K14-Citropin 1.1 [38]	*Phyllomedusa bicolor*	GLWSKIKEVGKEAAKAAAKAAGKAALGAVSEAV	α-helix	Prostate adenocarcinoma, Glioblastoma	Membrane disruption
GA-K4 [39]	Synthetic	FLKWLFKWAKK	α-helix	Not specified	Membrane disruption
Pardaxin [40] [41]	*Pardachirus marmoratus*	H-GFFALIPKIISSPLFKTLLSAVGSALSSSGGQE-OH	Turn/α-helix	Prostate cancer, Glioblastoma	ROS production, membrane depolarization

3 Methodology

To systematically predict AMPs with antimicrobial and anticancer activity, we designed a multistep bioinformatics workflow. The pipeline includes three main steps: (1) Homology Analysis, including pairwise alignments and BLAST to validate dual activity, (2) Dataset Construction, including sequence collection, feature extraction, and data labeling and pre-processing (3) Supervised Learning, involving model training, hyperparameter optimization, and evaluation. A schematic overview of the complete workflow is presented in Fig. 1.

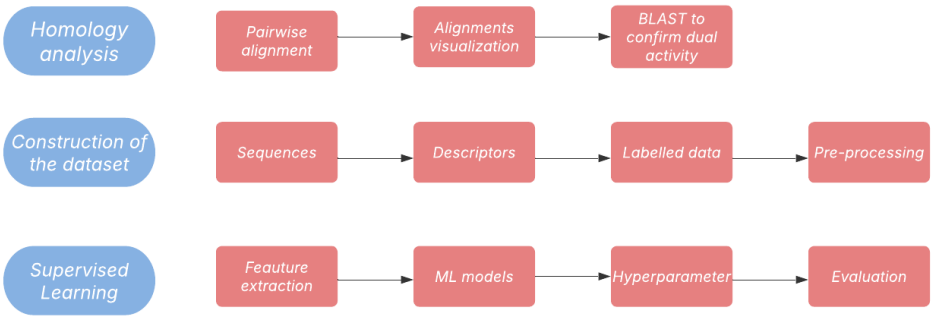

Fig. 1. Schematic representation of the general workflow.

3.1 Dataset loading and preparation

Construction of the Positive Dataset: A robust dataset of dual-function AMPs, characterized by antimicrobial and anticancer activity, was constructed by integrating two complementary sources. First, from the literature review conducted following PRISMA guidelines. Secondly, AMPs were retrieved from the DRAMP [42] public database by filtering entries annotated with both 'antimicrobial' and 'anticancer' activity in the Activity column of the dataset.

DRAMP is a regularly updated public database, and notably, all peptides collected from the literature review were already represented, reinforcing the consistency and reliability of the final dataset. After merging and removing redundant entries based on the sequence column, a non-redundant and biologically validated dataset was obtained.

3.2 Homology analysis

To evaluate sequence distinctiveness, we performed pairwise alignment against two reference sets: general AMPs (annotated with antimicrobial activity) and the dataset with confirmed dual-activity.

Initially, a general AMP dataset was filtered to retain only peptides annotated with antimicrobial activity. Pairwise alignment was performed between the dual-activity peptides and this reference AMP dataset. Each alignment was classified according to its percentage identity: sequences with the

identity 90% were classified as highly similar, and those with 50 to 89.99% identity were classified as moderately similar.

As a follow-up step, the dataset was compared to a established negative peptide dataset lacking bioactivity avaliable online [43], using a local BLASTp search. A identity threshold of 50% was applied to detect any high similarity matches that could compromise the discriminative value of the positive set.

Construction of the Negative Dataset: To ensure the development of robust and generalizable models, a negative dataset was obtained by selecting the peptide sequences with less than 50% sequence homology from the initial pairwise aligment between dual-function AMPs and conventional AMPs. This homology threshold ensures that the negative examples are sufficiently distinct from active peptides while maintaining biological relevance. This approach improves the discriminative power of the predictive algorithm by focusing on subtle sequence differences and not on the obvious dissimilarities.

3.3 Supervised Learning

To ensure robust model evaluation, we employed a 10-fold cross-validation strategy in the training dataset. In this approach, the dataset was partitioned into ten equal subsets; In each iteration, one part was designated for validation, and the other nine were used to train the model. This process was repeated ten times and each subset was used once as a validation set.

We applied stratified sampling to maintain consistent proportions of positive and negative peptide classes in all folds and chose to divide our dataset into 90% for training (including cross-validation) and 10% for the test.

Machine Learning Models: A prediction model was developed using some ML algorithms, including, Support Vector Machines (SVM), Random Forest(RF), Naive Bayes (NB) and Decision Tree (DT). The aim was to determine the most effective algorithm in predicting antimicrobial and anticancer activity on unseen data. All models were implemented and evaluated using the "Scikit-learn" library [44].

Hyperparameter otimization: To enhance the performance of the classification models, a hyperparameter optimization step was conducted using a grid search approach via GridSearchCV function (Scikit-learn library). The method in question, systematically explores a predefined set of hyperparameter combinations for each algorithm through cross-validation.

As a result of the large number of parameter combinations, the grid search generated multiple candidate models. Each was evaluated using 10-fold cross-validation, with performance assessed according to the F1-score.

Grid search allows the selection of the most effective hyperparameter configuration for each algorithm, resulting in models with optimized predictive performance.

Evaluation of model performance: To assess the predictive performance of each model, a comprehensive set of evaluation metrics was considered. These included F1-score, as the main one and, the Matthews Correlation Coefficient (MCC), accuracy, precision, recall, and the area under the receiver operating characteristic curve (ROC-AUC) for the best performing model. The evaluation was conducted on the 10% of the dataset reserved for independent testing, not involved in

model training. Most of these metrics are derived from the confusion matrix, which incorporates the following components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), ensuring a reliable assessment of the predictive capabilities of each considered model.

4 Results

Sequence Similarity Distributions from Pairwise Alignment

To investigate potential redundancy between dual-activity AMPs and antimicrobial-only peptides, pairwise global alignment was performed. The results revealed a total of 32,665 high-identity alignments, in contrast to 322,842 moderate-similarity alignments, suggesting that although some dual-function peptides share conserved motifs with canonical AMPs, the majority fall below the strict identity threshold.

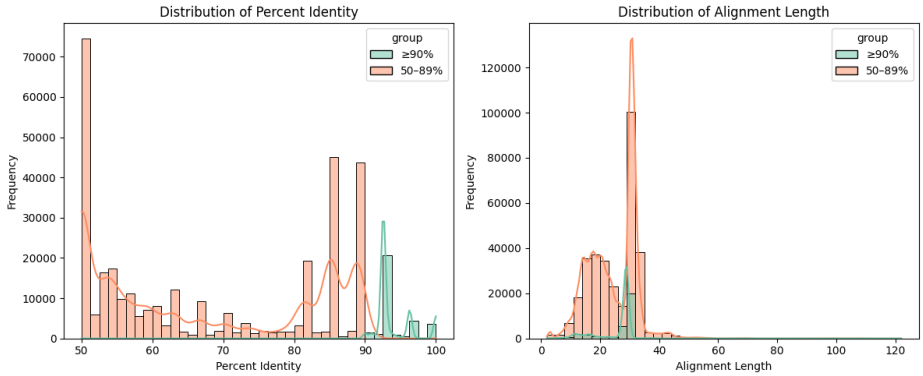

Fig. 2. Sequence similarity distributions from pairwise global alignments of dual-activity AMPs and antimicrobial-only peptides. The left panel shows the percentage identity distribution and the right panel the alignment lengths.

The percentage identity distribution (left panel in Figure 2) revealed sequence diversity, with a predominance of alignments with moderate similarity. This suggests that different sequences may perform similar functions through conserved local motifs, even in the absence of high overall identity. The presence of alignments with identity 90% indicates variants within the same peptide families, with potential for rational optimization. These findings reinforce the idea that antimicrobial activity is more associated with conserved structural features, while anticancer activity may rely on other properties that are less dependent on primary sequence or structural conservation.

The distribution of alignment lengths (right panel in Figure 2) further confirms that alignments in the moderate similarity group span a broader range, while the high-identity matches tend to involve more localized alignments, with the majority of alignment lengths being short. This is consistent with the characteristic size of AMPs.

The results suggests that although some dual-function peptides share conserved motifs with canonical AMPs, most exhibit distinct sequence features, minimizing redundancy.

Charge-Based Descriptor Distribution

To investigate charge impact in AMPs biologically motivated sequence descriptors were extracted to quantify the distribution of charged residues in each peptide. Specifically, the proportion of positively charged (Arg, Lys, His), negatively charged (Asp, Glu), and neutral residues within the first 50% of each sequence were computed, resulting in three charge descriptors: positive_50%, negative_50%, and neutral_50%.

To further explore these features, we focused on visualizing the spatial distribution of the positively charged residues across all dual-function AMPs, that can be visualized in Figure 3. There is a clear accumulation near the N-terminal region, consistent with the role of cationic residues in initiating membrane interactions, a well established mechanism for antimicrobial and anticancer activities.

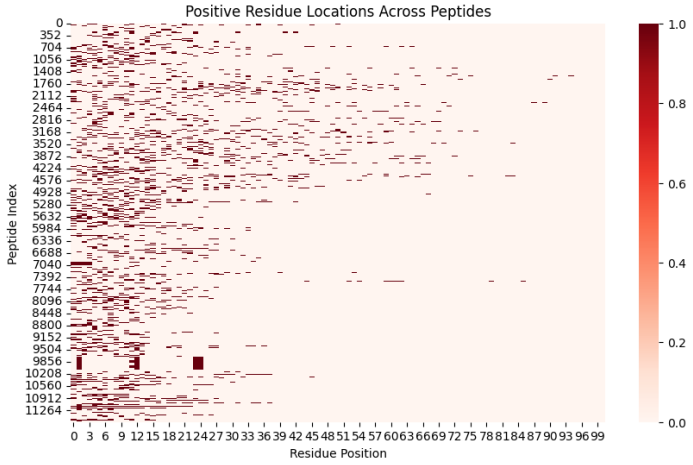

Fig. 3. Spatial distribution of positively charged residues (Arg, Lys, His) across dual-function peptides. There is a clear accumulation near the N-terminal region, supporting the biological role of cationic residues in initiating membrane interactions, which are essential for both antimicrobial and anticancer activities.

To evaluate potential redundancy among charge characteristics and their relevance, we performed a Pearson correlation analysis on standardized values. The resulting correlation matrix showed negligible correlations between descriptors, indicating that each feature contributes independently to the sequence representation.

These charge-based descriptors serve as a visualization tool to emphasize the role of positive charge in AMPs and were not included in the final dataset because they were already captured by the extracted features, avoiding redundancy in the feature set.

Validation Against a Negative Peptide Dataset

This step was crucial to ensure that none of the selected positive sequences exhibited sequence similarity to biologically inactive peptides, which could otherwise lead to misclassification or reduced model generalizability.

Using an identity threshold of 50%, no significant alignments were found between the AMPs with confirmed dual activity and the negative dataset. This absence of overlap reinforces its integrity and suitability for use in supervised learning applications.

Predictive Modeling of Dual-Function AMPs Using Interpretable Descriptors

The supervised learning phase in this study aimed to explore peptides with dual antimicrobial and anticancer activity through a robust machine learning pipeline. Two distinct datasets were used: a positive class composed of 2362 experimentally validated dual-function peptides, and a negative class containing 2713 sequences selected based on low sequence homology (less than 50).%) to antimicrobial peptides, derived from the pairwise alignment with dual-function AMPs.

For feature extraction, the peptide sequences were initially filtered to retain only those composed of standard amino acids. A comprehensive set of 403 descriptors was calculated for each sequence, integrating custom motif-based characteristics generated via "compute_descriptors" function, amino acid composition profiles, physicochemical properties, including aromaticity, instability index, isoelectric point, and hydropathy, calculated using BioPython's ProteinAnalysis module, as well as bigram frequencies capturing local sequence dependencies. After removing features with low variance, a matrix of 403 informative descriptors was retained.

For model development, we used a stratified 10-fold cross-validation strategy along with grid search to optimize hyperparameters for the four well established classification algorithms: SVM, RF, NB and DT. The best hyperparameter configuration for each model was used to retrain the classifiers on the training data and assess their performance on the data reserved for testing.

The results, including the best hyperparameters and the corresponding F1-Score, accuracy and MCC metrics are summarized in Table 2.

Table 2. Best hyperparameters, F1-score, Accuracy, and MCC for each ML model after 10-fold cross-validation.

ML Models	**Hyperparameters**	**F1-Score**	**Accuracy**	**MCC**
Random Forest	• max_depth = 20 • min_samples_split = 10 • n_estimators = 200	0.9366	0.9422	0.8690
Decision Tree	• max_depth = 10 • min_samples_split = 2	0.9242	0.9332	0.8560
SVM	• C = 0.1 • kernel = linear	0.9142	0.9200	0.8280
Naive Bayes	• var_smoothing = 1e-10	0.8707	0.8807	0.7445

As can be seen from the results presented above, all models achieved competitive performance, with Random Forest standing out as the best-performing classifier based on F1-score. Random Forest is one of the most used ensemble learning methods, particularly for peptide classification tasks. It is well suited for binary classification problems. RF constructs several decision trees during training and aggregates the outputs to improve generalization. Each tree is built using a random

subset of features and data samples, that helps reduce overfitting. Random Forest models are capable of handling high-dimensional data and noisy features, making them ideal for peptide datasets composed of diverse physicochemical descriptors like this one.

In contrast, the Decision Tree algorithm offers a more interpretable single-tree structure that sequentially splits the data based on feature thresholds to maximize information gain to understand the underlying structure of the peptide data.

Given the strong predictive performance obtained in the initial evaluation, we further assessed the generalization capacity of RF and DT models on the independent test set.

This test set contained peptide sequences that were not used during training, with the same feature representation (403 descriptors) as the training data. The test set included a balanced distribution of samples from both classes, positive and negative. The confusion matrices in Fig. 4 summarize the classification results for both models.

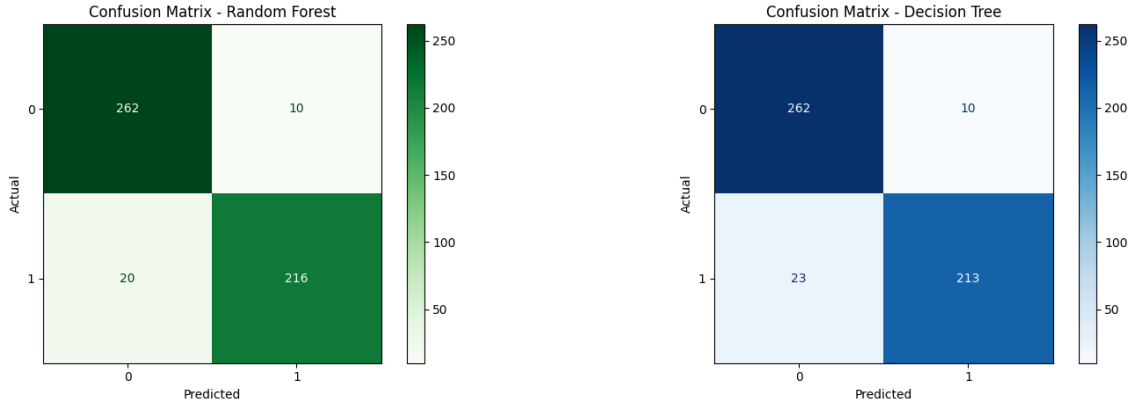

Fig. 4. Confusion matrices for the Random Forest (left) and Decision Tree (right) classifiers applied to the prediction of dual-activity AMPs.

As illustrated, the RF model correctly classified 262 of the positive samples and 216 of the negative ones, while misclassifying 10 negative samples as positive (false positives) and 20 positive samples as negative (false negatives), resulting in a total of 30 misclassifications out of 508 predictions. The DT model, on the other hand, classified 262 positive samples and 213 negative ones, with 33 misclassifications.

Both models demonstrated robustness, achieving high F1-scores on the test set, with the RF model maintaining slightly better precision and recall compared to the Decision Tree. These results confirm that both models generalize well to unseen data, with RF exhibiting slightly stronger overall performance.

Most important feautures

To better understand the factors associated with antimicrobial and anticancer activity in AMPs, we analyzed the top 20 ranked features identified by the RF model. These features capture key molecular and physicochemical properties that are most predictive of dual biological function.

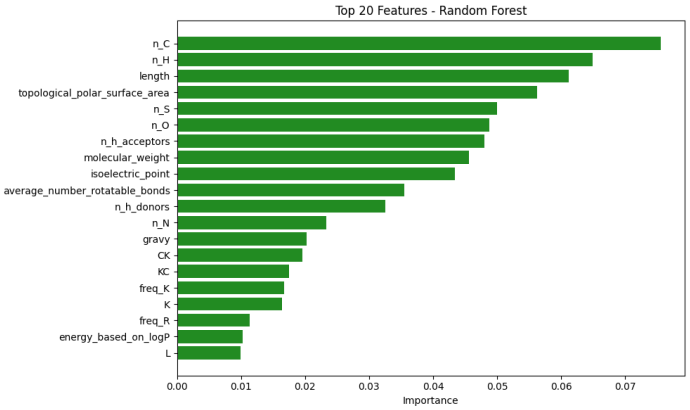

Fig. 5. Importance of the top 20 features for the Random Forest model in predicting dual-function AMPs.

As can be seen, sequence length appears as one of the most influential features in RF. This is consistent with established biological characteristics of AMPs, short sequences ranging from 10 to 50 amino acids. Such compact structures are believed to favor membrane interaction and penetration mechanisms that are central to both antimicrobial and anticancer activities.

In addition to length, the model emphasized the critical role of positive charge. Features related to cationicity, such as the frequency of lysine (freq_K) and arginine (freq_R) residues, as well as the isoelectric point, were among the most informative. This reflects the importance of positive charge in facilitating electrostatic interactions with negatively charged bacterial and cancer cell membranes, thereby promoting selective targeting and membrane disruption. The presence of specific dipeptide motifs, KC and CK, further suggests the role of residue position in regulating this activity. At the top of the feature ranking, the number of carbon and hydrogen atoms also emerged as highly influential, capturing aspects of overall hydrophobicity.

Together, these features reflect the properties that contribute to the dual bioactivity of AMPs, more precisely, positive charge, hydrophobicity, and structural compactness.

Overall, our results demonstrate that short length and positive charge are key determinants in predicting dual-function AMPs.

These results support the biochemical understanding of AMP function and confirm the effectiveness of ML models in revealing the physicochemical foundations of peptide activity.

ROC Curves and AUC Values

The Receiver Operating Characteristic (ROC) curve is a widely used tool to evaluate the capacity of prediction in classification models [45]. It demonstrates the balance between the True Positive Rate and the False Positive Rate, making it particularly valuable in ML for evaluating a model's ability to distinguish between classes [45].

The ROC curves presented in Figure 6 illustrate the performance of the ML models considered on the task of identifying dual-activity AMPs. The Area Under the Curve (AUC) serves as a single scalar value that summarizes the classification ability of each model.

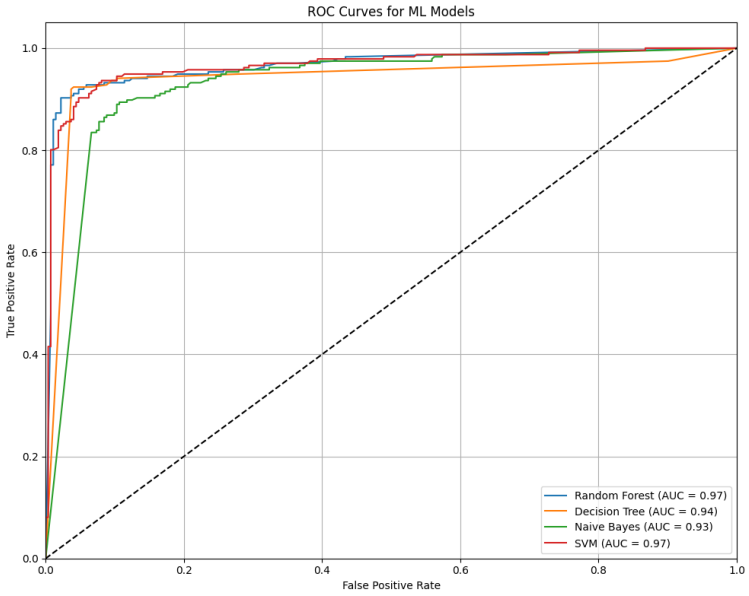

Fig. 6. ROC curves comparing the performance of four ML models for predicting dual-activity antimicrobial peptides.

Both RF and SVM achieved the highest AUC values (0.97), indicating great performance. Their curves reflect high sensitivity (true positive rate) and specificity (true negative rate).

Decision Tree also showed strong performance with an AUC of 0.94, suggesting a solid balance between true and false positive rates.

Naive Bayes, with an AUC of 0.93, also demonstrated a good classification ability. However, its slightly lower AUC suggests that it is less robust in handling more complex decision boundaries, suggesting it offers effective discrimination but with slightly less precision compared to the other models.

4.1 Conclusion

In this study, we implemented a bioinformatics workflow to classify antimicrobial peptides with dual antimicrobial and anticancer activities. A data set was constructed by integrating validated 2362 positive sequences and 2713 negative sequences, followed by the extraction of 403 informative descriptors.

GridSearchCV combined with cross-validation was used to train and optimize several classifiers. Among the models evaluated, Random Forest (F1 = 0.9366) achieved a strong performance, with an AUC value of 0.97, confirming robustness in handling this biologically classification task. The SVM model, despite achieving a lower F1-score (F1 = 0.9142), it still demonstrated strong results with an AUC value of 0.97, making it a promising alternative.

Overall, this study underscores the potential of ML models, with Random Forest being the most effective for the goal of our study, predicting dual-function antimicrobial peptides.

These results reinforce the idea that AMPs are good candidates to overcome AMR and the emerging resistance in conventional cancer therapies, while also demonstrating the important role of machine learning in this process.

4.2 Future Work

This framework aligns with previous researchs demonstrating the efficacy of computational approaches in predicting peptides activity.

The results suggest that this proposed pipeline, particularly the Random Forest model, can serve as an *in silico* strategy for the identification of multifunctional AMPs and in the discovery of therapeutic candidates with dual bioactivity profiles.

As a next step, it would be valuable to integrate deep learning-based embeddings to capture higher-order biological information. Using deep learning techniques, we can improve the prediction ability, including, accuracy and the capacity to generalize and identify promising therapeutic candidates.

Code and Data Availability

The pipelines implemented in this project were entirely developed in Python, and are available at https://github.com/barbarafreitas22/Project

References

1. World Health Organization. (2023). Antimicrobial resistance. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance
2. European Centre for Disease Prevention and Control. (2020). Surveillance of antimicrobial resistance in Europe 2019. https://www.ecdc.europa.eu/en/publications-data/surveillance-antimicrobial-resistance-europe-2019
3. Panjla, A., Joshi, S., Singh, G., Bamford, S. E., Mechler, A., & Verma, S. (2024). Applying Machine Learning for Antibiotic Development and Prediction of Microbial Resistance. Chemistry–An Asian Journal, 19(18), e202400102.
4. El-Tanani, M. et al. "Bridging the gap: From petri dish to patient - Advancements in translational drug discovery." Heliyon, vol. 11,1, e41317, 17 Dec. 2024.
5. Boucher, H. W. et al. (2009). Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clinical Infectious Diseases, 48(1), 1–12.
6. Uddin, T. M., Chakraborty, A. J., Khusro, A., Zidan, B. R. M., Mitra, S., Emran, T. B., & Koirala, N. (2021). Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. Journal of Infection and Public Health, 14(12), 1750–1766.
7. Sun, J., Deng, Z., & Yan, A. (2014). Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. Biochemical and Biophysical Research Communications, 453(2), 254–267.
8. Kyriakidis, I., Vasileiou, E., Pana, Z. D., & Tragiannidis, A. (2021). Acinetobacter baumannii Antibiotic Resistance Mechanisms. Pathogens, 10(3), 373.
9. Lin, Q. et al. "Prevention of ESKAPE pathogen biofilm formation by antimicrobial peptides WLBU2 and LL37." International Journal of Antimicrobial Agents, vol. 52,5, 667–672 (2018).
10. Theuretzbacher, U. et al. (2020). The global preclinical antibacterial pipeline. Nature Reviews Microbiology, 18(5), 275–285.
11. Mookherjee, N. et al. (2020). Antimicrobial host defence peptides: functions and clinical potential. Nature Reviews Drug Discovery, 19(5), 311–332.
12. Fjell, C. D. et al. (2012). Designing antimicrobial peptides: form follows function. Nature Reviews Drug Discovery, 11(1), 37–51.
13. Ageitos, J. M., Sánchez-Pérez, A., Calo-Mata, P., & Villa, T. G. (2017). Antimicrobial peptides (AMPs): ancient compounds that represent novel weapons in the fight against bacteria. Biochemical Pharmacology, 133, 117–138.
14. Tornesello, A. L. et al. "Antimicrobial Peptides as Anticancer Agents: Functional Properties and Biological Activities." Molecules, vol. 25,12, 2850, 19 Jun. 2020.
15. Hennig, D. et al. "Novel insights into appropriate encapsulation methods for bioactive compounds into polymers: a study with peptides and HDAC inhibitors." Macromolecular Bioscience, vol. 14,1, 69–80 (2014).
16. Fu, Y. et al. "Peptide Modified Albumin-Paclitaxel Nanoparticles for Improving Chemotherapy and Preventing Metastasis." Macromolecular Bioscience, vol. 22,3, e2100404 (2022).
17. Kordi, M., Borzouyi, Z., Chitsaz, S., hadi Asmaei, M., Salami, R., & Tabarzad, M. (2023). Antimicrobial peptides with anticancer activity: Today status, trends and their computational design. Archives of Biochemistry and Biophysics, 733, 109484.
18. Boohaker, R. J., Lee, W., Vishnubhotla, M., Perez, L. M., & Khaled, R. (2012). The use of therapeutic peptides to target and to kill cancer cells. Current Medicinal Chemistry, 19(22), 3794–3804.
19. Mollinedo, F., & Gajate, C. (2020). Lipid rafts as signaling hubs in cancer cell survival/death and invasion: implications in tumor progression and therapy. Journal of Lipid Research, 61(5), 611–635.
20. Lim, K. J., Sung, B. H., Shin, J. R., Lee, Y. W., Kim, D. J., Yang, K. S., & Kim, S. C. (2013). A cancer specific cell-penetrating peptide, BR2, for the efficient delivery of an scFv into cancer cells. PloS One, 8(6), e66084.

21. Xie, J., Bi, Y., Zhang, H., Dong, S., Teng, L., Lee, R. J., & Yang, Z. (2020). Cell-penetrating peptides in diagnosis and treatment of human diseases: from preclinical research to clinical application. Frontiers in Pharmacology, 11, 697.
22. Regberg, J., Srimanee, A., & Langel, Ü. (2012). Applications of cell-penetrating peptides for tumor targeting and future cancer therapies. Pharmaceuticals, 5(9), 991–1007.
23. Page, M. J. et al. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." BMJ, vol. 372, n71, 29 Mar. 2021.
24. Tang, Y., Yang, C., Zhao, J., Heng, H., Peng, M., Sun, L., ... & Chen, S. (2025). LTX-315 is a novel broad-spectrum antimicrobial peptide against clinical multidrug-resistant bacteria. Journal of Advanced Research.
25. Alkhulaifi, M. M., Alosaimi, M. M., Khan, M. S., Tabrez, S., Shaik, G. M., Alokail, M. S., ... & Husain, F. M. (2024). Assessment of broad-spectrum antimicrobial, antibiofilm, and anticancer potential of lactoferrin extracted from camel milk. Applied Biochemistry and Biotechnology, 196(3), 1464–1480.
26. Vernen, F. et al. "Characterization of Tachyplesin Peptides and Their Cyclized Analogues to Improve Antimicrobial and Anticancer Properties." International Journal of Molecular Sciences, vol. 20,17, 4184, 26 Aug. 2019.
27. Salnikov, E. et al. Cathelicidin-BF: A Potent Antimicrobial Peptide Leveraging Charge and Phospholipid Recruitment against Multidrug-Resistant Clinical Bacterial Isolates. Journal of the American Chemical Society, 147(13), 11199–11215 (2025).
28. Su, B. C., Hung, G. Y., Tu, Y. C., Yeh, W. C., Lin, M. C., & Chen, J. Y. (2021). Marine Antimicrobial Peptide TP4 Exerts Anticancer Effects on Human Synovial Sarcoma Cells via Calcium Overload, Reactive Oxygen Species Production and Mitochondrial Hyperpolarization. Marine Drugs, 19(2), 93.
29. Panahi Chegini, P., Nikokar, I., Tabarzad, M., Faezi, S., & Mahboubi, A. (2019). Effect of Amino Acid Substitutions on Biological Activity of Antimicrobial Peptide: Design, Recombinant Production, and Biological Activity. Iranian Journal of Pharmaceutical Research, 18(Suppl1), 157–168.
30. Piktel, E., Wnorowska, U., Gorbacz-Konończuk, J., Sienkiewicz, J., Głuszek, K., Okła, S., & Bucki, R. (2024). From antimicrobial to anticancer: unraveling the potential of pleurocidin and pleurocidin-derived peptides in the treatment of cancers. Frontiers in Pharmacology, 15, 1340029.
31. Neshani, A., Zare, H., Akbari Eidgahi, M. R., Khaledi, A., & Ghazvini, K. (2019). Epinecidin-1, a highly potent marine antimicrobial peptide with anticancer and immunomodulatory activities. BMC Pharmacology and Toxicology, 20(1), 33.
32. Kim, I. W., Choi, R. Y., Lee, J. H., Seo, M., Lee, H. J., Kim, M. A., Kim, S. H., Kim, I., & Hwang, J. S. (2021). Anticancer Activity of Periplanetasin-5, an Antimicrobial Peptide from the Cockroach *Periplaneta americana*. Journal of Microbiology and Biotechnology, 31(10), 1343–1349. `https://doi.org/10.4014/jmb.2104.04040`
33. Patil, S. M., & Kunda, N. K. (2022). Anticancer activity of D-LAK-120A, an antimicrobial peptide, in non-small cell lung cancer (NSCLC). Biochimie, 201, 7–17.
34. Yang, Y., Chen, H. Y., Hao, H., & Wang, K. J. (2022). The Anticancer Activity Conferred by the Mud Crab Antimicrobial Peptide Scyreprocin through Apoptosis and Membrane Disruption. International Journal of Molecular Sciences, 23(10), 5500.
35. Wanyan, Yangke et al. "2-Deoxy-d-glucose Promotes Buforin IIb-Induced Cytotoxicity in Prostate Cancer DU145 Cells and Xenograft Tumors." Molecules (Basel, Switzerland) vol. 25,23 5778. 7 Dec. 2020, doi:10.3390/molecules25235778
36. Han, Z., Feng, D., Wang, W., Wang, Y., Cheng, M., Yang, H., & Liu, Y. (2023). Influence of Fatty Acid Modification on the Anticancer Activity of the Antimicrobial Peptide Figainin 1. ACS Omega, 8(44), 41876–41884.
37. Timmons, P. B., & Hewage, C. M. (2021). Conformation and membrane interaction studies of the potent antimicrobial and anticancer peptide palustrin-Ca. Scientific Reports, 11(1), 22468.
38. Shen, H., Zhang, N., Kong, X., Wang, N., Hu, H. G., Cong, W., & Liu, C. (2023). Benzyl stapled modification and anticancer activity of antimicrobial peptide A4K14-Citropin 1.1. Bioorganic & Medicinal Chemistry Letters, 96, 129499.

39. Mishig-Ochir, T., Gombosuren, D., Jigjid, A., Tuguldur, B., Chuluunbaatar, G., Urnukhsaikhan, E., Pathak, C., & Lee, B. J. (2017). Cellular Membrane Composition Requirement by Antimicrobial and Anticancer Peptide GA-K4. Protein and Peptide Letters, 24(3), 197–205.
40. Chen, Y.-P. et al. "Pardaxin Activates Excessive Mitophagy and Mitochondria-Mediated Apoptosis in Human Ovarian Cancer by Inducing Reactive Oxygen Species." Antioxidants, vol. 10,12, 1883, 25 Nov. 2021.
41. Han, Y. et al. "In Vitro and in Vivo Anticancer Activity of Pardaxin against Proliferation and Growth of Oral Squamous Cell Carcinoma." Marine Drugs, vol. 14,1, 2, 23 Dec. 2015.
42. Ma, T., Liu, Y., Yu, B., Sun, X., Yao, H., Hao, C., Li, J., Nawaz, M., Jiang, X., Lao, X., & Zheng, H. (2025). DRAMP 4.0: an open-access data repository dedicated to the clinical translation of antimicrobial peptides. Nucleic Acids Research, 53(D1), D403–D410.
43. AxPep. (n.d.). *AXPEP: Antimicrobial Peptide Database*. Retrieved from `https://sourceforge.net/projects/axpep/support`
44. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. Retrieved from `http://jmlr.csail.mit.edu/papers/volume12/pedregosa11a/pedregosa11a.pdf`
45. Feng, K., Hong, H., Tang, K., and Wang, J. (2019). Decision making with machine learning and ROC curves. *arXiv preprint* arXiv:1905.02810.