Prediction and design of antifungal peptides using artificial neural networks

Ivan Sanchis; Alvaro Siano. Bioactive Peptide Laboratory, FBCB-National University of the Littoral, Santa Fe, Argentina. CONICET, Argentina. Correspondence: asiano@fbcb.unl.edu.ar

INTRODUCTION

The rising incidence of fungal infections has prompted the exploration of novel therapeutic avenues, with antimicrobial peptides (AMPs) emerging as promising candidates for antifungal therapies. Various computational methodologies, including template-based approaches, docking simulations, alignment methods, and machine learning techniques, have been harnessed for predicting and designing antifungal peptides (AFPs).
In this study, we developed an artificial neural network (ANN) based deep-learning model to predict antifungal activity of peptides using their amino acid sequence. Leveraging a diverse dataset of experimentally validated antifungal peptides, our model predicts antifungal activity and facilitates the design of new peptides with high in silico predicted efficacy.

METHODS

Positive dataset: 1478 unique AFPs from Antifp, Uniprot and APD3 datasets. Negative dataset: 739 random sequences from Uniprot (not classified as AFPs or AMPs) + 739 randomly generated sequences. Several combinations of CNN with LSTM layers, accompanied by dropout layers and dense layers with different hyperparameters were tested as illustrated in the image below. Model performance was evaluated based on ROC (Receiver Operating Characteristic) curve area, sensitivity [SE = true positive / (true positive + false negative)], and specificity [SP = true negative / (true negative + false positive)]. Models were built using Keras framework.

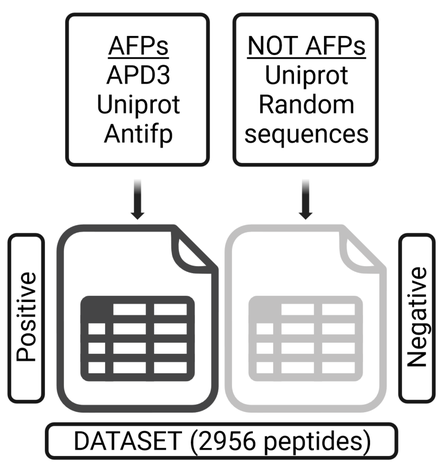

Fig. 1 Scheme of the dataset construction.

METHODS

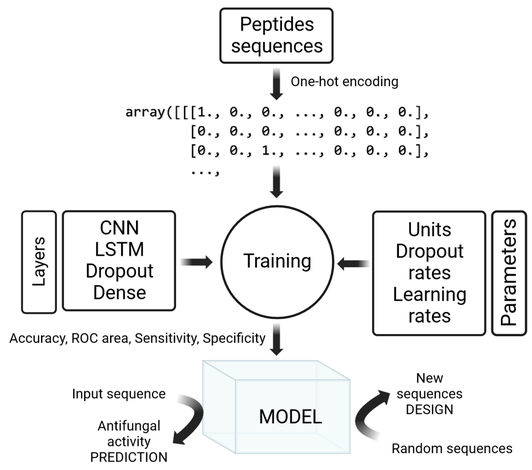

Fig. 2 Scheme of the methodology developed.

RESULTS

Table1. Best model summary

Layer (type)	Output Shape	Parameters
Conv1D	(None, 38, 64)	filters=64, kernel size=3
MaxPooling1D	(None, 19, 64)	pool size=2
LSTM	(None, 19, 50)	units=50
Dropout	(None, 19, 50)	rate=0.5
LSTM	(None, 32)	units=32
Dense	(None, 64)	units=64
Dropout	(None, 64)	rate=0.5
Dense	(None, 64)	units=64
Dropout	(None, 64)	rate=0.9
Dense	(None, 16)	units=16
Dense (output)	(None, 1)	units=11, act.='sigmoid'

Except for the output layer, all layers' activation functions were ReLu.

Table 2. Best model performance

Metric	Value
Accuracy	92.46%
Sensitivity	92.98%
Sensibility	91.95%
ROC curve area	0.9770

RESULTS

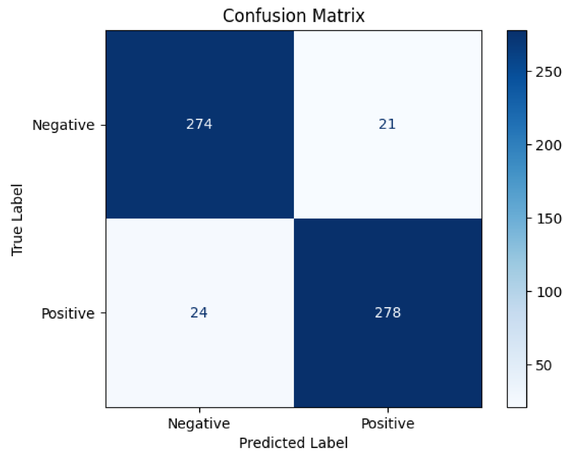

Fig. 3 Confusion matrix of the best performing AFP prediction model.

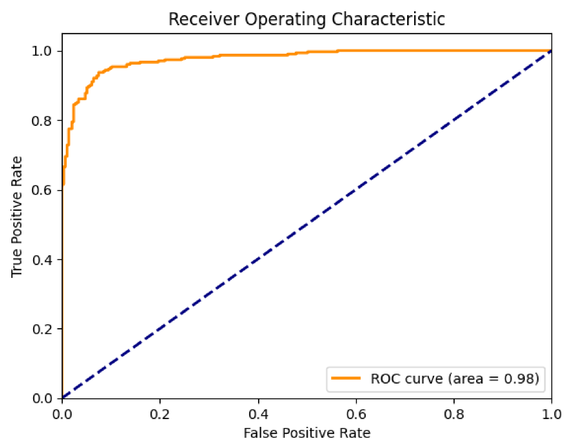

Fig. 4 ROC curve graph of the best performing AFP prediction model.

RESULTS

The best-performing model featured a single CNN layer followed by LSTM, dense, and dropout layers. While similar architectures have been explored in previous AFP studies, a notable distinction lies in the incorporation of dropout layers with distinct dropout rates in all LSTM and dense layers. The model achieved remarkable performance metrics: Accuracy = 92.46%, SE = 92.98%, SP = 91.95%, and ROC curve area = 0.98. Furthermore, we employed the model in reverse to generate sequences with high predicted antifungal activity. Random 5-40 mer sequences were iteratively generated and subjected to model prediction. The iterations stopped when 1000 sequences surpassing a prediction threshold of 0.99 were found. This calculation took 178534 iterations (approx. 3 h) on a T4 GPU. The 10 sequences with the highest predicted value are listed in Table 3.

Table 3. Top ten predicted sequences

Iteration	Sequence
20514	NPTALKKLHKAR
24119	CTRRPCIA
33892	KSCNVNCACVR
65074	TKCCGVMKAVNGPCYCW
92060	ECKCYPSCPVRHKY
93946	KRRSCSLRNPALNQCICRCHASN
96430	LKCPCSCIFQMC
78512	SPNLLVKLF
123297	FCPFKC
146486	VFSPCGTHA

CONCLUSION

Our work's next steps entail synthesizing and evaluating the identified sequences against various fungal species. Additionally, we plan to develop and make available an online application for public use, enabling the prediction and design of AFPs using our model. These outcomes underscore the potency of ANN-based approaches in predicting biological peptide activity solely from their amino acid sequences, with significant implications for the tailored design of novel AFPs.

Created with BioRender Poster Builder