

Analyzing and exploring Graph Attention Networks and protein-based language models for predicting *Porphyromonas gingivalis* resistant efflux protein sequences

Pradeep Kumar Yadalam^{1,A}, Prabhu Manickam Natarajan^{2,B}, Naresh Shetty^{3,C}, Maria Maddalena Marrapodi^{4,D}, Hande Uzunçibuk^{5,D}, Diana Russo^{6,E}, Marco Ciccù^{7,E,F}, Giuseppe Minervini^{8,9,D,E}

¹ Department of Periodontics, Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences (SIMATS Deemed University), Chennai, India

² Department of Clinical Sciences, Center of Medical and Bio-Allied Health Sciences and Research, Ajman University, UAE

³ Department of Clinical Sciences, College of Dentistry, Ajman University, UAE

⁴ Department of Woman, Child and General and Specialist Surgery, University of Campania Luigi Vanvitelli, Naples, Italy

⁵ Department of Orthodontics, Faculty of Dentistry, Trakya University, Edirne, Turkey

⁶ Oral Surgery Unit, Multidisciplinary Department of Medical-Surgical and Dental Specialties, University of Campania Luigi Vanvitelli, Naples, Italy

⁷ Department of Biomedical and Surgical and Biomedical Sciences, Catania University, Italy

⁸ Department of Orthodontics, Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences (SIMATS Deemed University), Chennai, India

⁹ Multidisciplinary Department of Medical-Surgical and Dental Specialties, University of Campania Luigi Vanvitelli, Naples, Italy

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;

D – writing the article; E – critical revision of the article; F – final approval of the article

Dental and Medical Problems, ISSN 1644-387X (print), ISSN 2300-9020 (online)

Dent Med Probl. 2025;62(2):265–273

Address for correspondence

Giuseppe Minervini

E-mail: giuseppe.minervini@unicampania.it

Funding sources

None declared

Conflict of interest

None declared

Acknowledgements

None declared

Received on February 21, 2024

Reviewed on February 27, 2024

Accepted on March 15, 2024

Published online on April 30, 2025

Cite as

Yadalam PK, Natarajan PM, Shetty N, etc. Analyzing and exploring Graph Attention Networks and protein-based language models for predicting *Porphyromonas gingivalis* resistant efflux protein sequences. *Dent Med Probl.* 2025;62(2):265–273. doi:10.17219/dmp/186143

DOI

10.17219/dmp/186143

Copyright

Copyright by Author(s)

This is an article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC BY 3.0) (<https://creativecommons.org/licenses/by/3.0/>).

Abstract

Background. Antimicrobial resistance (AMR) must be predicted to combat antibiotic-resistant illnesses. Based on high-priority AMR genomes, it is possible to track resistance and focus treatment to stop global outbreaks. Large language models (LLMs) are essential for identifying *Porphyromonas gingivalis* multi-resistant efflux genes to prevent resistance. Antibiotic resistance is a serious problem; however, by studying specific bacterial genomes, we can predict how resistance develops and find better kinds of treatment.

Objectives. This paper explores using advanced models to predict the sequences of proteins that make *P. gingivalis* resistant to treatment. Understanding this approach could help prevent AMR more effectively.

Material and methods. This research utilized multi-drug-resistant efflux protein sequences from *P. gingivalis*, identified through UniProt ID A0A0K2J2N6_PORGN, and formatted as FASTA sequences for analysis. These sequences underwent rigorous detection and quality assurance processes to ensure their suitability for computational analysis. The study employed the DeepBio framework, which integrates LLMs with deep attention networks to process FASTA sequences.

Results. The analysis revealed that the Long Short-Term Memory (LSTM)-attention, ProtBERT and BERTGAT models achieved sensitivity scores of 0.9 across the board, with accuracy rates of 89.5%, 88.5% and 90.5%, respectively. These results highlight the effectiveness of the models in identifying *P. gingivalis* strains resistant to multiple drugs. Furthermore, the study assessed the specificity of the LSTM-attention, ProtBERT and BERTGAT models, which achieved scores of 0.89, 0.87 and 0.90, respectively. Specificity, or the genuine negative rate, measures the ability of a model to accurately identify non-resistant cases, which is crucial for minimizing false positives in AMR detection.

Conclusions. When utilized clinically, this LLM approach will help prevent AMR, which is a global problem. Understanding this approach may enable researchers to develop more effective treatment strategies that target specific resistant genes, reducing the likelihood of resistance development. Ultimately, this approach could play a pivotal role in preventing AMR on a global scale.

Keywords: periodontitis, antimicrobial resistance, large language models, *Porphyromonas gingivalis*, efflux protein

Highlights

- By examining efflux protein sequences, advanced artificial intelligence (AI) models, such as LSTM-attention, ProtBERT and BERTGAT, can accurately (up to 90.5%) and sensitively (~0.90) predict antimicrobial resistance (AMR) in *Porphyromonas gingivalis*.
- With the highest accuracy (90.5%) and specificity (0.90), BERTGAT performed better than other models. This indicates that adding graph-based attention mechanisms enhances AMR prediction by more accurately capturing biological relationships.
- The SHAP, UMAP, ROC, PR, and UpSet plots confirmed model interpretability and robustness, indicating their possible clinical use in detecting resistant strains and directing precise antibiotic tactics.
- The study emphasizes the significance of targeting efflux proteins for novel drug design to combat multidrug-resistant *P. gingivalis*, a keystone pathogen in periodontitis.
- Notwithstanding encouraging findings, the small dataset size and the absence of external validation are drawbacks that call for additional research with bigger and more varied datasets.

Introduction

Antimicrobial resistance (AMR)^{1,2} is the ability of micro-organisms to resist the effects of antimicrobial drugs, such as antibiotics, antivirals and antiparasitics.^{3–5} Combating antibiotic-resistant diseases requires predicting AMR. High-priority AMR genomes can lead surveillance to track resistance and focus treatment in order to prevent global outbreaks.^{6–8}

Leveraging insights from large language models (LLMs), like ProtBERT or BERTGAT, can be employed to explore the intricate mechanisms governing the interplay between protein sequences, their structural configurations and resultant functions.^{9,10} The essence of this paradigm lies in understanding how the linear arrangement of amino acids, akin to the syntax of a sentence, dictates the three-dimensional (3D) structure of a protein, which, in turn, governs its biological functions. By adopting computational language models, traditionally used in natural language processing (NLP), we gain a valuable tool to dissect and decipher the functions of proteins.^{11–13} This approach allows researchers to unveil the nuanced relationships between amino acid sequences, the structural motifs they form and the functional roles they play in biological processes. Treating protein sequences as linguistic entities provides a powerful framework for unraveling the language of life encoded in these fundamental biological molecules.¹⁴

The attention-based Long Short-Term Memory (LSTM-attention) network is a method that analyzes big datasets and looks for patterns that point to AMR, using state-of-the-art algorithms.^{15–22} Co-AMPpred is one instance of a machine learning method for AMR prediction.^{23,24} This tool distinguishes between antimicrobial peptides (AMPs) and non-AMPs by combining physicochemical characteristics and composition-based sequences through machine learning techniques.

An important global health concern is periodontitis, an immune-inflammatory infectious disease, mostly caused by *Porphyromonas gingivalis*.^{25,26} The bacterium

exhibits a variety of omics and phylogeny information, making it a significant factor in severe periodontitis. Treatment for *P. gingivalis* is becoming more difficult due to its growing resistance to antibiotics, which highlights the need for a deeper comprehension of its resistance mechanisms. In particular, the resistance-nodulation-division (RND) family of efflux pumps is a major contributor to the AMR of *P. gingivalis*. These pumps, including proteins such as AcrA, AcrB and TolC,^{27–30} block the entry of antimicrobial drugs into the bacterial cell, contributing to multi-drug resistance (MDR).

Porphyromonas gingivalis-produced gingipains and virulence factors^{31,32} add to the complexity of the situation. Due to gingipains, *P. gingivalis* can elude the host immune system, which contributes to AMR. The integrated protein–protein interaction network (PPIN), which includes virulence regulators and efflux pump proteins, was subjected to topological and functional analysis; this analysis identified genes crucial for understanding the relationships across cellular systems in *P. gingivalis*.³¹ The bifunctional NAD(P)H-hydrate repair enzyme A0A212GBI3_PORGN is one of the most prevalent resistant efflux proteins.^{33–37} It is essential for the bifunctional enzyme that it catalyzes the dehydration of the S-form of NAD(P)HX³⁸ at the expense of ADP, which is converted to AMP, as well as the epimerization of the S- and R-forms of NAD(P)HX.

Identifying *P. gingivalis* multi-resistant efflux genes with the use of LLMs is crucial for preventing resistance. The present study aimed to analyze and explore Graph Attention Networks (GATs) and protein-based language models for predicting *P. gingivalis* resistant efflux protein sequences.

Methods

Using UniProt,³⁹ the following sequences of multi-drug resistant proteins of *P. gingivalis* were downloaded: A0A0K2J2N6_PORGN; A0A212GBI3_PORGN;

A0A2D2N4E3_PORGN; A0A0E2LNT1_PORGN; A0A829KLL9_PORGN; U2K1P7_PORGN; Q7MXT9_PORGI; A0A1R4DUJ6_PORGN; and A0A212FQN2_PORGN. The identified FASTA sequences underwent a thorough quality check to ensure that there were no biases during their entry. Additionally, the sequences were formatted according to the prescribed format based on the DeepBio tool for LLMs and deep attention networks.⁴⁰

DeepBio

Academics can construct a deep learning architecture to address any biological problem with the help of DeepBio, a one-stop web service. In addition to visualizing biological sequencing data, DeepBio compares and enhances deep learning models. It offers base functional annotation tasks, with in-depth interpretations and graphical visualizations, and conservation motif analysis to confirm site dependability, and well-trained deep learning architectures for more than 20 tasks. The sequence-based datasets were divided into the training and test sets using DeepBio. We randomly divided each dataset into 1,000 training and 200 testing sets to optimize hyperparameters and analyze performance.

BERTGAT

BERTGAT⁴¹ is a neural network model that combines the pre-trained language model Bidirectional Encoder Representations from Transformers (BERT) with GAT.^{16,42} BERT extracts text features,⁴¹ while GAT learns the sentence–word relationships.^{26,43,44} Transformer-based language models are preferred over recurrent neural networks (RNNs). Pre-trained BERT representations are fine-tuned to generate state-of-the-art models for wide-ranging text-to-structured query language (SQL) workloads with one extra output layer.

ProtBERT

The provided search results do not contain specific information about the full code architecture of ProtBERT and its detailed steps. However, based on the available information, it was possible to provide a general outline of the architecture and the steps involved in using ProtBERT for protein sequence prediction.⁴¹

ProtBERT architecture and steps for protein sequence prediction

Pre-training

ProtBERT is pre-trained on a large dataset of protein sequences, representing the entire known protein space, using a masked language modeling task combined with

a novel Gene Ontology (GO) annotation prediction task. The architecture of ProtBERT consists of local and global representations, allowing the end-to-end processing of protein sequences and GO annotations.

Fine-tuning

After pre-training, the ProtBERT model is fine-tuned on specific protein-related tasks, such as protein sequence classification or function prediction. Fine-tuning involves initializing the model from the pre-trained state, freezing some layers, training additional, fully connected layers, and then unfreezing all layers for further training.

Model evaluation

The fine-tuned ProtBERT model is evaluated on diverse benchmarks covering various protein properties to assess its performance. The ProtBERT model is built on Keras/TensorFlow and is available through the Hugging Face model hub. The code for using ProtBERT involves loading the pre-trained model, fine-tuning it on specific protein-related tasks, and utilizing it for protein sequence prediction and analysis.

LSTM-attention model

LSTM^{15,17} and attention mechanisms are combined in LSTM-attention, a deep learning architecture, to enhance sequence prediction task performance. The following steps are needed to put the LSTM-attention model into practice:

1. Data Preparation: The first stage is to prepare the input data for the model. This could entail activities like feature extraction, encoding and tokenization.
2. Model Architecture: An LSTM layer and an attention layer form the LSTM-attention model. After processing the input sequence, the LSTM layer creates a series of hidden states. The more pertinent states are given more weight when the attention layer computes a weighted sum of the hidden states.
3. Training: The model is trained using the appropriate loss function and optimization technique with the prepared data. The parameters of the model are adjusted during training to minimize the loss function.
4. Evaluation: After training, the performance of the model is assessed on an independent test set. This entails calculating metrics like the F1 score, recall, accuracy, and precision.
5. Prediction: The model can forecast new sequences after evaluation. The trained model receives the input sequence and the learned weights generate the output.
6. Fine-tuning: The model can be further adjusted on particular tasks or datasets to boost performance. This involves changing the hyperparameters or architecture of the model to fit a given task better (Table 1).

Table 1. Parameters of the Protein Language Model (PLM)

Cuda:	TRUE	TRUE2	TRUE3
Seed:	43	43	43
num_workers:	4	4	4
num_class:	2	2	2
Kmer:	3	3	3
heatmap_seq:			
save_figure_type:	png	png	png
Mode:	train-test	train-test	train-test
Type:	prot	prot	prot
model:	BertGAT	LSTMAttention	prot_bert
datatype:	userprovide	userprovide	userprovide
interval_log:	10	10	10
interval_valid:	1	1	1
interval_test:	1	1	1
Epoch:	50	50	50
Optimizer:	Adam	Adam	Adam
loss_func:	CE	CE	CE
batch_size:	4	8	32
LR:	1.00E-05	0.0001	0.0001
Reg:	0.0025	0.0025	0.0025
Gamma:	2	2	2
Alpha:	0.25	0.25	0.25
max_len:	35	207	52
dim_embedding:	32	32	32
minimode:	modelCompare	modelCompare	modelCompare
if_use_FL:	0	0	0
if_data_aug:	0	0	0
if_data_enh:	0	0	0
CDHit:	['1']	['1']	['1']

Results

LSTM-attention, ProtBERT and BERTGAT were used to find the hidden features and weights in the FASTA protein sequences; then, backpropagation algorithms with ADAM optimizer and 50 iterations fine-tuned the model.

LSTM-attention, ProtBERT and BERTGAT had sensitivity of 0.90, 0.90 and 0.91, respectively ($TP / (TP + FN)$; TP – true positive, FN – false negative). Specificity, or the true negative rate, is the proportion of actual negatives correctly predicted as negatives. The specificity of LSTM-attention, ProtBERT and BERTGAT was 0.89, 0.87 and 0.90, respectively ($TN / (TN + FP)$; TN – true negative, FP – false positive).

ROC curve

The receiver-operating characteristic (ROC) curve shows the trade-off between the true positive rate (sensitivity) and false positive rate (1-specificity) of the model

over the categorization thresholds. Regarding LSTM-attention, ProtBERT and BERTGAT, high true positive rates are shown by the ROC curve in the upper left corner of the plot.

PR curve

The trade-off between recall and precision for binary classifiers with different probability thresholds is depicted by the precision–recall (PR) curve. While precision is the fraction of positive predictions, recall is the percentage of accurately expected positives. This model's performance with uneven classes is made public. The area under the PR curve (AUC-PR) is a widely used metric to summarize the classifier performance. Higher AUC-PR values for LSTM-attention, ProtBERT and BERTGAT denote improved model performance.

An epoch plot is a graph showing the accuracy and loss of a machine learning model over training. It is an effective diagnostic tool for overfitting and other model issues. The number of epochs or iterations the model has been trained on is shown by the X-axis in an epoch plot. The accuracy or loss of the model is plotted on the Y-axis. The loss indicates how effectively the model predicts the proper output for a given input. Accuracy gauges whether the predictions of the model are accurate.

UpSet plot

The frequency of common items between groups can be ascertained by comparing the intersection diameters. While smaller crossings imply less overlap, larger intersections show more overlap between groups. In a vertical UpSet plot, rows represent intersections and matrix columns represent sets. Each row has filled intersection cells that show how the rows are related to each other.

Uniform Manifold Approximation and Projection (UMAP) creates a weighted graph from high-dimensional data to show clustering patterns, with the edge strength reflecting how 'close' the points are. Projecting this graph lowers its dimension. This data shows algorithm clustering. UMAP is a non-linear dimension reduction method for embedding high-dimensional data in low-dimensional space. It assumes that high-dimensional data points should be close to low-dimensional space.

SHAP values

The predictive value of each feature is quantified in a machine learning model. All possible feature combinations are considered, along with the relative contributions of each feature to the prediction when coupled with a subset of features, to compute the value. When a feature enhances the prediction, the Shapley Additive Explanations (SHAP) red value is positive. A feature with a negative SHAP blue value is less predictive.

Discussion

Antimicrobial drug-resistant periodontal bacteria^{45–47} are characterized by efflux pumps – proteins that remove antimicrobial medications from the cell, thus preventing the drugs from killing the bacteria. Bacteria can also adapt their outer membrane to block antimicrobial medications or change the target site of the drug to lessen its efficacy.^{36,48} These pathways and others cause antibiotic resistance in periodontitis patients. Whole-genome sequencing can detect AMR genes^{34,35,49,50} and mutations, assessing the resistance potential. Large genomic, phenotypic and clinical datasets can be used to train machine learning algorithms to predict resistance and discover the key AMR genes. Prolonged illness, more expensive second-line therapies and missed productivity can strain healthcare systems and national economies. Predicting AMR in globally prevailing periodontal infections, especially for the keystone pathogens like *P. gingivalis*, is important for preventing resistance from spreading across continents. Antimicrobial resistance is a growing concern in the field of periodontitis research. It refers to the ability of microorganisms, such as bacteria, to resist the effects of antimicrobial drugs.^{1,51}

Large language models have revolutionized various fields, including protein sequence prediction. In this study, models such as LSTM-attention, ProtBERT and BERTGAT demonstrated high predictive performance, with accuracy rates reaching up to 90.5% (Table 2,

Table 2. Accuracy of the LSTM-attention, ProtBERT and BERTGAT models

Model name	Accuracy	Sensitivity	Specificity	AUC
LSTM-attention	0.895	0.90	0.89	0.948
ProtBERT	0.885	0.90	0.87	0.941
BERTGAT	0.905	0.91	0.90	0.951

AUC – area under the curve (referring the receiver-operating characteristic (ROC) curve unless otherwise specified). Accuracy is defined as the overall proportion of correctly classified cases (true positives and true negatives) among all predictions.

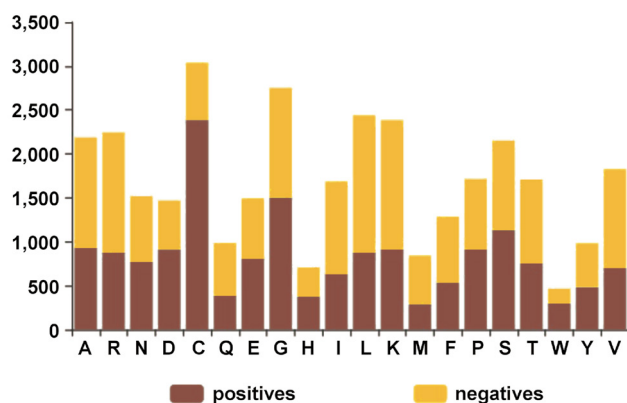


Fig. 1. Distribution of positive (antimicrobial drug-resistant) and negative (non-resistant) sequences in the training and test datasets

The letters on the X-axis represent the subsets or identifiers of the input sequences used during model training and evaluation.

Fig. 1–5). Large language models have also shown strong results in broader protein-related tasks, such as structure prediction and protein design, in previous studies.

The observed performance differences between LSTM-attention, ProtBERT and BERTGAT, with accuracy rates of 89.5%, 88.5% and 90.5%, respectively, deepen the

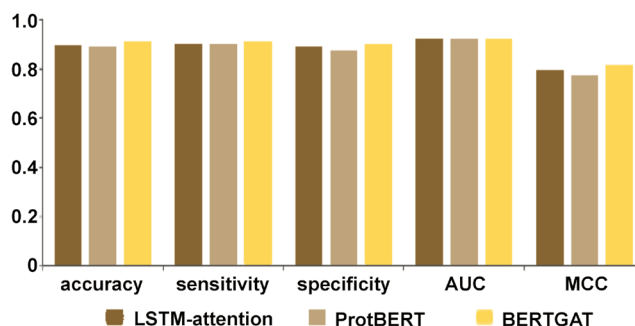


Fig. 2. Accuracy of the models

MCC – Matthews correlation coefficient.

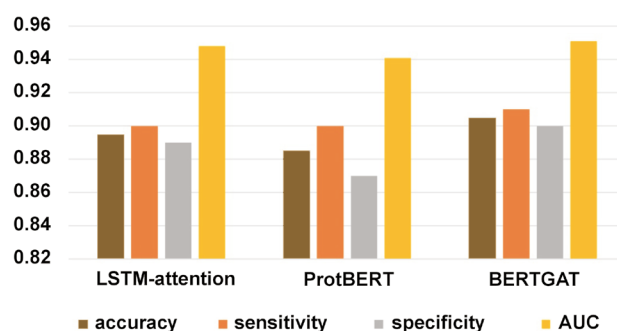


Fig. 3. Bar chart of the accuracy of the models

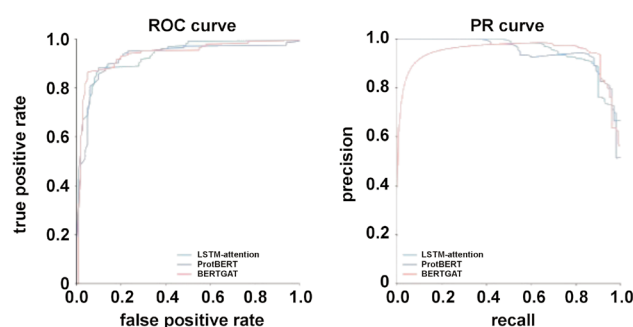


Fig. 4. Receiver-operating characteristic (ROC) and precision–recall (PR) curves of the plot

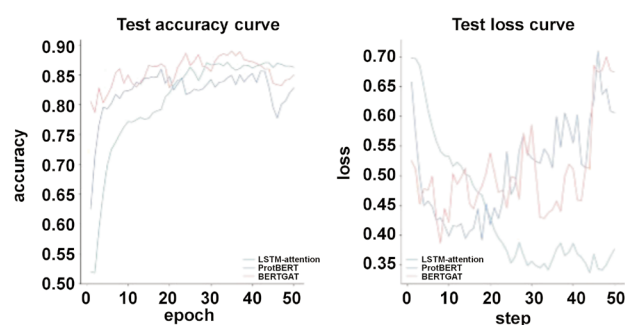


Fig. 5. Epoch plot of all iterations with algorithms

interpretation of the results in the context of model architecture implications.^{52,53} LSTM-attention utilizes long short-term memory units and attention mechanisms, while ProtBERT incorporates a transformer-based architecture, specifically designed for protein sequence data, and BERTGAT incorporates graph attention mechanisms. The higher accuracy of BERTGAT suggests that its increased model complexity and ability to capture graph structures in the data have contributed to improved performance. Data representation is another important factor to consider. The comparable accuracy of LSTM-attention and ProtBERT suggests that their respective data representations are effective for a given task. Biological relevance is a critical consideration when evaluating model performance. Protein sequence analysis is inherently tied to biology, and it is important to assess how well the

models align with biological knowledge.^{50,52,54} While all 3 models demonstrated high accuracy, it is necessary to delve deeper into the interpretation to understand if the superior accuracy of BERTGAT is biologically relevant or if other factors drive it. Overall, the observed performance differences between LSTM-attention, ProtBERT and BERTGAT highlight the impact of model complexity, data representation and biological relevance. Further analysis and interpretation are required to uncover the specific advantages of each architecture and their implications in the context of protein sequence analysis.

Previous state-of-the-art models, like ProteinBERT,^{55–57} a universal deep learning model for protein sequences, leveraging the transformer architecture,^{58–60} are commonly used in NLP tasks. In addition to language models, various machine learning methods and algorithms are used in protein sequence prediction, such as graph neural networks and deep learning-based algorithms like BERTGAT and LSTM-attention.^{15–17} ProtBERT is a transformer-based language model trained on a large corpus of protein sequences to learn representations that capture important structural and functional information.²⁴ This study compared LLMs vs. GAT-based algorithms^{57,61,62} in predicting AMR sequencing, and the performance of the model was shown using the SHAP,^{63–65} UMAP and UpSet plot analysis (Fig. 6–8), similar to previous studies for performance.

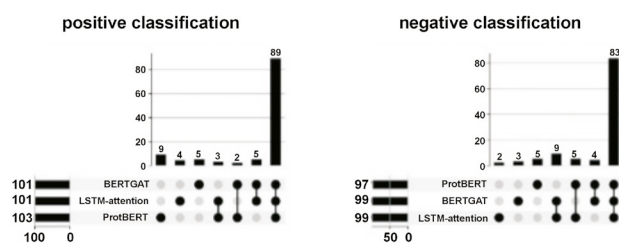


Fig. 6. UpSet intersection diagram of the models with positive and negative classification

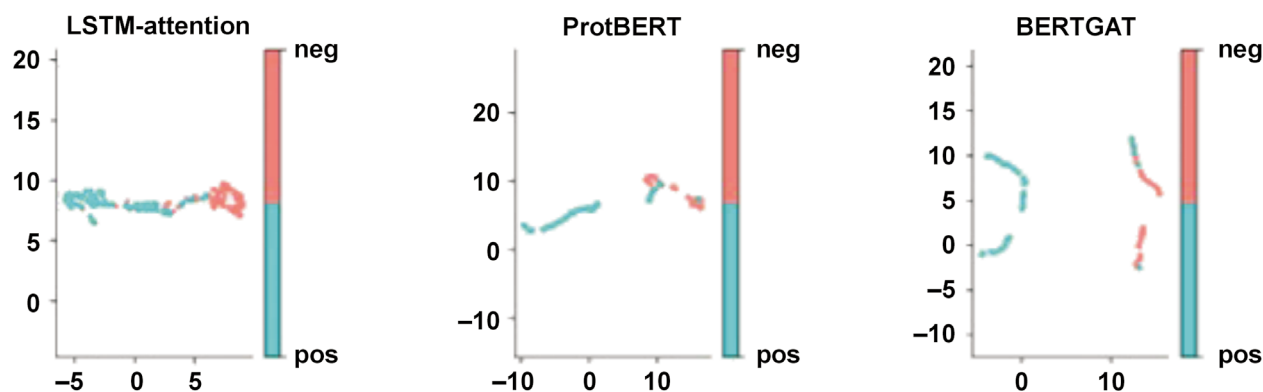


Fig. 7. Uniform Manifold Approximation and Projection (UMAP) plot of the models
neg – negative; pos – positive.

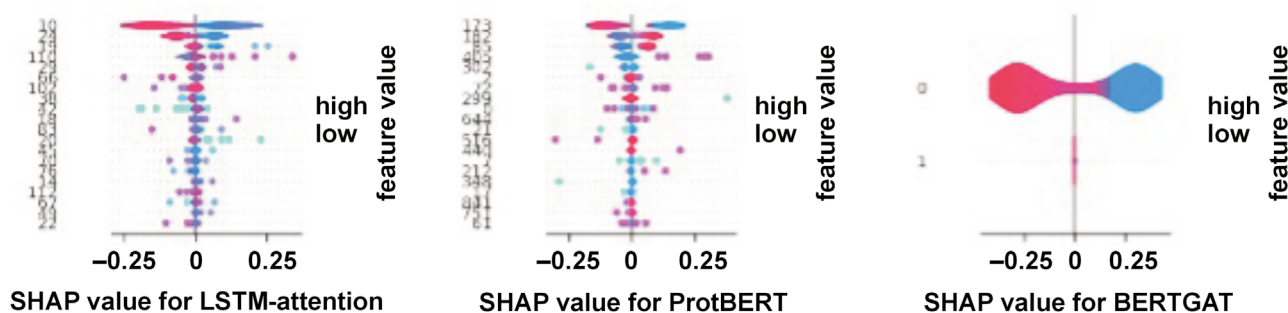


Fig. 8. Shapley Additive Explanations (SHAP) performance of the models

Targeting *P. gingivalis* efflux proteins is important for novel antibiotic drug design. These prediction models could point to resistance mutation sequences and prevent the development of AMR in periodontitis patients.^{66,67}

This study compared the performance of LLMs and GAT-based algorithms in predicting AMR sequencing. The model's performance was evaluated using the SHAP, UMAP and UpSet plot analysis, previously employed to assess the performance of similar prediction models. The study also highlighted the significance of targeting *P. gingivalis* efflux proteins to design novel antibiotic drugs. However, it is important to acknowledge that the current study has limitations. One major limitation is the small sample size and the lack of the external validation of the independent datasets used in the study.⁵² Future research should address this limitation by including larger sample sizes to ensure the reliability and generalizability of the prediction model. Furthermore, further investigations are needed to validate the model's performance in diverse datasets and to explore its applicability for other oral microbes.

Conclusions

Preventing the spread of antimicrobial resistance (AMR) is a primary global concern, and large language models (LLMs), when applied clinically, may help prevent this phenomenon.

Ethics approval and consent to participate

Not applicable.

Data availability

The datasets supporting the findings of the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Use of AI and AI-assisted technologies

Not applicable.

ORCID iDs

Pradeep Kumar Yadalam  <https://orcid.org/0000-0002-6653-4123>
 Prabhu Manickam Natarajan  <https://orcid.org/0000-0002-4780-0465>
 Naresh Shetty  <https://orcid.org/0000-0002-4596-5215>
 Maria Maddalena Marrapodi  <https://orcid.org/0000-0002-9494-6942>
 Hande Uzuncibuk  <https://orcid.org/0000-0001-9265-1772>
 Diana Russo  <https://orcid.org/0000-0002-0915-4642>
 Marco Ciciù  <https://orcid.org/0000-0003-2311-9728>
 Giuseppe Minervini  <https://orcid.org/0000-0002-8309-1272>

References

- Inoue T, Nakayama M, Taguchi Y, et al. Characterization of the tripartite drug efflux pumps of *Porphyromonas gingivalis* aTCC 33277. *New Microbiol.* 2015;38(1):101–108. PMID:25742153.
- Grover V, Kapoor A, Malhotra R, Kaur G. *Porphyromonas gingivalis* antigenic determinants – potential targets for the vaccine development against periodontitis. *Infect Disord Drug Targets.* 2014;14(1):1–13. doi:10.2174/1871526514666140827100930
- Van Camp PJ, Haslam DB, Porollo A. Prediction of antimicrobial resistance in Gram-negative bacteria from whole-genome sequencing data. *Front Microbiol.* 2020;11:1013. doi:10.3389/fmicb.2020.01013
- Peter S, Bosio M, Gross C, et al. Tracking of antibiotic resistance transfer and rapid plasmid evolution in a hospital setting by nanopore sequencing. *mSphere.* 2020;5(4):e00525-20. doi:10.1128/mSphere.00525-20
- Suzuki M, Hashimoto Y, Hirabayashi A, et al. Genomic epidemiological analysis of antimicrobial-resistant bacteria with nanopore sequencing. *Methods Mol Biol.* 2023;2632:227–246. doi:10.1007/978-1-0716-2996-3_16
- Kuang X, Wang F, Hernandez KM, Zhang Z, Grossman RL. Accurate and rapid prediction of tuberculosis drug resistance from genome sequence data using traditional machine learning algorithms and CNN. *Sci Rep.* 2022;12(1):2427. doi:10.1038/s41598-022-06449-4
- Ren Y, Chakraborty T, Doijad S, et al. Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. *Bioinformatics.* 2022;38(2):325–334. doi:10.1093/bioinformatics/btab681
- Rahbe E, Watier L, Guillemot D, Glaser P, Opatowski L. Determinants of worldwide antibiotic resistance dynamics across drug-bacterium pairs: A multivariable spatial-temporal analysis using ATLAS. *Lancet Planet Health.* 2023;7(7):e547–e557. doi:10.1016/S2542-5196(23)00127-4
- Yang F, Wang W, Wang F, et al. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat Mach Intell.* 2022;4:852–866. doi:10.1038/s42256-022-00534-z
- Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: Pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics.* 2021;37(15):2112–2120. doi:10.1093/bioinformatics/btab083
- Sorin V, Barash Y, Konen E, Klang E. Large language models for oncological applications. *J Cancer Res Clin Oncol.* 2023;149(11):9505–9508. doi:10.1007/s00432-023-04824-w
- Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large language models in medical education: Opportunities, challenges, and future directions. *JMIR Med Educ.* 2023;9:e48291. doi:10.2196/48291
- Mallio CA, Sertorio AC, Bernetti C, Zobel BB. Large language models for structured reporting in radiology: Performance of GPT-4, ChatGPT-3.5, Perplexity and Bing. *Radiol Med.* 2023;128(7):808–812. doi:10.1007/s11547-023-01651-4
- Yadalam PK, Trivedi SS, Krishnamurthi I, et al. Machine learning predicts patient tangible outcomes after dental implant surgery. *IEEE Access.* 2022;10:131481–131488. doi:10.1109/ACCESS.2022.3228793.
- Zhang K, Yu J, Liu J, et al. LGEANet: LSTM-global temporal convolution-external attention network for respiratory motion prediction. *Med Phys.* 2023;50(4):1975–1989. doi:10.1002/mp.16237
- Zheng K, Zhang XL, Wang L, You ZH, Zhan ZH, Li HY. Line graph attention networks for predicting disease-associated Piwi-interacting RNAs. *Brief Bioinform.* 2022;23(6):bbac393. doi:10.1093/bib/bbac393
- Ming Y, Qian H, Guangyuan L. CNN-LSTM facial expression recognition method fused with two-layer attention mechanism. *Comput Intell Neurosci.* 2022;2022:7450637. doi:10.1155/2022/7450637
- Gupta A, Kumar S, Gopi A, Sharma M, Patil S, Piplani A. Assessment of knowledge, practices and attitudes of dentists toward coronavirus disease while performing aerosol-generating procedures in dentistry: A cross-sectional survey from India. *Dent Med Probl.* 2023;60(3):459–466. doi:10.17219/dmp/156197
- Dovigo S, Massariol M, Gandini A, Zuffellato N. Instantaneous dental implant loading technique by fixed dentures: A retrospective cohort study. *Dent Med Probl.* 2023;60(3):375–383. doi:10.17219/dmp/154981

20. Sanjuán-Navarro PS, Agudelo-Suárez AA, Mora-Cárdenas AL, et al. Frequency of symptoms and the associated factors of eating disorders in a group of dental students in Medellín, Colombia. *Dent Med Probl.* 2023;60(3):401–411. doi:10.17219/dmp/149900
21. Soundarajan S, Rajasekar A. Antibacterial and anti-inflammatory effects of a novel herb-mediated nanocomposite mouthwash in plaque-induced gingivitis: A randomized controlled trial. *Dent Med Probl.* 2023;60(3):445–451. doi:10.17219/dmp/150728
22. Paradowska-Stolarz AM, Wieckiewicz M, Mikulewicz M, et al. Comparison of the tensile modulus of three 3D-printable materials used in dentistry. *Dent Med Probl.* 2023;60(3):505–511. doi:10.17219/dmp/166070
23. Huang G, Luo W, Zhang G, et al. Enhancer-LSTMAtt: A Bi-LSTM and attention-based deep learning method for enhancer recognition. *Biomolecules.* 2022;12(7):995. doi:10.3390/biom12070995
24. Kwon HB, Choi SH, Lee D, et al. Attention-based LSTM for non-contact sleep stage classification using IR-UWB radar. *IEEE J Biomed Health Inform.* 2021;25(10):3844–3853. doi:10.1109/JBHI.2021.3072644
25. Li XY, Wang C, Xiang XR, Chen FC, Yang CM, Wu J. *Porphyromonas gingivalis* lipopolysaccharide increases lipid accumulation by affecting CD36 and ATP-binding cassette transporter A1 in macrophages. *Oncol Rep.* 2013;30(3):1329–1336. doi:10.3892/or.2013.2600
26. Eick S, Mathey A, Vollroth K, et al. Persistence of *Porphyromonas gingivalis* is a negative predictor in patients with moderate to severe periodontitis after nonsurgical periodontal therapy. *Clin Oral Investig.* 2017;21(2):665–674. doi:10.1007/s00784-016-1933-x
27. Bostanci N, Belibasakis GN. *Porphyromonas gingivalis*: An invasive and evasive opportunistic oral pathogen. *FEMS Microbiol Lett.* 2012;333(1):1–9. doi:10.1111/j.1574-6968.2012.02579.x
28. Yang Y, He X, Xia S, Liu F, Luo L. *Porphyromonas gingivalis* facilitated the foam cell formation via lysosomal integral membrane protein 2 (LIMP2). *J Periodontol Res.* 2021;56(2):265–274. doi:10.1111/jre.12812
29. Nuñez-Belmar J, Morales-Olavarria M, Vicencio E, Vernal R, Cárdenas JP, Cortez C. Contribution of -omics technologies in the study of *Porphyromonas gingivalis* during periodontitis pathogenesis: A minireview. *Int J Mol Sci.* 2022;24(1):620. doi:10.3390/ijms24010620
30. Benahmed AG, Mujawdiya PK, Noor S, Gasmi A. *Porphyromonas gingivalis* in the development of periodontitis: Impact on dysbiosis and inflammation. *Arch Razi Inst.* 2022;77(5):1539–1551. doi:10.22092/ARI.2021.356596.1875
31. Zenobia C, Hajishengallis G. *Porphyromonas gingivalis* virulence factors involved in subversion of leukocytes and microbial dysbiosis. *Virulence.* 2015;6(3):236–243. doi:10.1080/21505594.2014.999567
32. Ruan Q, Guan P, Qi W, et al. *Porphyromonas gingivalis* regulates atherosclerosis through an immune pathway. *Front Immunol.* 2023;14:1103592. doi:10.3389/fimmu.2023.1103592
33. Mysak J, Podzimek S, Sommerova P, et al. *Porphyromonas gingivalis*: Major periodontopathic pathogen overview. *J Immunol Res.* 2014;2014:476068. doi:10.1155/2014/476068
34. Latini Abreu MG, Kawamoto D, Alves Mayer MP, et al. Frequency of *Porphyromonas gingivalis* fimA in smokers and nonsmokers after periodontal therapy. *J Appl Oral Sci.* 2019;27:e20180205. doi:10.1590/1678-7757-2018-0205
35. Morandini AC, Ramos-Junior ES, Potempa J, et al. *Porphyromonas gingivalis* fimbriae dampen P2X7-dependent interleukin-1 β secretion. *J Innate Immun.* 2014;6(6):831–845. doi:10.1159/000363338
36. Liu M, Shao J, Zhao Y, Ma B, Ge S. *Porphyromonas gingivalis* evades immune clearance by regulating lysosome efflux. *J Dent Res.* 2023;102(5):555–564. doi:10.1177/00220345221146097
37. Puig-Silla M, Dasí-Fernández F, Montiel-Company JM, Almerich-Silla JM. Prevalence of fimA genotypes of *Porphyromonas gingivalis* and other periodontal bacteria in a Spanish population with chronic periodontitis. *Med Oral Patol Oral Cir Bucal.* 2012;17(6):e1047–e1053. doi:10.4317/medoral.17009
38. Yadalam PK, Anegundi RV, Hebayan A. Prediction of druggable allosteric sites of undruggable multidrug resistance efflux pump *P. gingivalis* proteins. *Biomed Eng Comput Biol.* 2023;14:11795972231202394. doi:10.1177/11795972231202394
39. Uniprot Consortium. UniProt: The Universal Protein Knowledge-base in 2023. *Nucleic Acids Res.* 2023;51(D1):D523–D531. doi:10.1093/nar/gkac1052
40. Wang R, Jiang Y, Jin J, et al. DeepBIO: An automated and interpretable deep-learning platform for high-throughput biological sequence prediction, functional annotation and visualization analysis. *Nucleic Acids Res.* 2023;51(7):3017–3029. doi:10.1093/nar/gkad055
41. Brandes N, Ofer D, Peleg Y, Rappoport N, Linial M. ProteinBERT: A universal deep-learning model of protein sequence and function. *Bioinformatics.* 2022;38(8):2102–2110. doi:10.1093/bioinformatics/btac020
42. Liu D, Xu H, Wang J, Lu Y, Kong J, Qi M. Adaptive Attention Memory Graph Convolutional Networks for skeleton-based action recognition. *Sensors (Basel).* 2021;21(20):6761. doi:10.3390/s21206761
43. Yu R, Pan C, Fei X, Chen M, Shen D. Multi-graph attention networks with bilinear convolution for diagnosis of schizophrenia. *IEEE J Biomed Health Inform.* 2023;27(3):1443–1454. doi:10.1109/JBHI.2022.3229465
44. Zhao Y, Wang L, Wang C, et al. Multi-granularity heterogeneous graph attention networks for extractive document summarization. *Neural Netw.* 2022;155:340–347. doi:10.1016/j.neunet.2022.08.021
45. Jia L, Han N, Du J, Guo L, Luo Z, Liu Y. Pathogenesis of important virulence factors of *Porphyromonas gingivalis* via toll-like receptors. *Front Cell Infect Microbiol.* 2019;9:262. doi:10.3389/fcimb.2019.00262
46. Aabed K, Moubayed N, BinShabaib MS, ALHarthi SS. Is a single session of antimicrobial photodynamic therapy as an adjuvant to non-surgical scaling and root planing effective in reducing periodontal inflammation and subgingival presence of *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans* in patients with periodontitis? *Photodiagnosis Photodyn Ther.* 2022;38:102847. doi:10.1016/j.pdpdt.2022.102847
47. Pan C, Liu J, Wang H, Song J, Tan L, Zhao H. *Porphyromonas gingivalis* can invade periodontal ligament stem cells. *BMC Microbiol.* 2017;17(1):38. doi:10.1186/s12866-017-0950-5
48. Wadhawan A, Reynolds MA, Makkar H, et al. Periodontal pathogens and neuropsychiatric health. *Curr Top Med Chem.* 2020;20(15):1353–1397. doi:10.2174/1568026620666200110161105
49. Griffen AL, Becker MR, Lyons SR, Moeschberger ML, Leys EJ. Prevalence of *Porphyromonas gingivalis* and periodontal health status. *J Clin Microbiol.* 1998;36(11):3239–3242. doi:10.1128/JCM.36.11.3239-3242.1998
50. Liu J, Gong X. Attention mechanism enhanced LSTM with residual architecture and its application for protein–protein interaction residue pairs prediction. *BMC Bioinformatics.* 2019;20(1):609. doi:10.1186/s12859-019-3199-1
51. Murakami N, Yoshikawa K, Tsukada K, et al. Butyric acid modulates periodontal nociception in *Porphyromonas gingivalis*-induced periodontitis. *J Oral Sci.* 2022;64(1):91–94. doi:10.2334/josnusd.21-0483
52. Lupo U, Sgarbossa D, Bitbol AF. Protein language models trained on multiple sequence alignments learn phylogenetic relationships. *Nat Commun.* 2022;13(1):6298. doi:10.1038/s41467-022-34032-y
53. Ferruz N, Schmidt S, Höcker B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat Commun.* 2022;13(1):4348. doi:10.1038/s41467-022-32007-7
54. Marks DS, Hopf TA, Sander C. Protein structure prediction from sequence variation. *Nat Biotechnol.* 2012;30(11):1072–1080. doi:10.1038/nbt.2419
55. Geffen Y, Ofra Y, Unger R. DistilProtBert: A distilled protein language model used to distinguish between real proteins and their randomly shuffled counterparts. *Bioinformatics.* 2022;38(Suppl 2):ii95–ii98. doi:10.1093/bioinformatics/btac474
56. Villegas-Morcillo A, Gomez AM, Sanchez V. An analysis of protein language model embeddings for fold prediction. *Brief Bioinform.* 2022;23(3):bbac142. doi:10.1093/bib/bbac142
57. Long Y, Wu M, Liu Y, Kwok CK, Luo J, Li X. Ensembling graph attention networks for human microb–drug association prediction. *Bioinformatics.* 2020;36(Suppl 2):i779–i786. doi:10.1093/bioinformatics/btaa891
58. Perez R, Li X, Giannakoulis S, Petersson EJ. AggBERT: Best in class prediction of hexapeptide amyloidogenesis with a semi-supervised ProtBERT model. *J Chem Inf Model.* 2023;63(18):5727–5733. doi:10.1021/acs.jcim.3c00817
59. Guntuboina C, Das A, Mollaei P, Kim S, Barati Farmani A. PeptideBERT: A language model based on transformers for peptide property prediction. *J Phys Chem Lett.* 2023;14(46):10427–10434. doi:10.1021/acs.jpcclett.3c02398

60. Hallee L, Rafailidis N, Gleghorn JP. cdsBERT – extending protein language models with Codon awareness [preprint]. *bioRxiv*. 2023. doi:10.1101/2023.09.15.558027
61. Wang X, Ding Z, Wang R, Lin X. DeepPro-Glu: Combination of convolutional neural network and Bi-LSTM models using ProtBERT and handcrafted features to identify lysine glutarylation sites. *Brief Bioinform*. 2023;24(2):bbac631. doi:10.1093/bib/bbac631
62. Choi J, Ko T, Choi Y, Byun H, Kim CK. Dynamic graph convolutional networks with attention mechanism for rumor detection on social media. *PLoS One*. 2021;16(8):e0256039. doi:10.1371/journal.pone.0256039
63. Wang K, Tian J, Zheng C, et al. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput Biol Med*. 2021;137:104813. doi:10.1016/j.combiomed.2021.104813
64. Fernández MV, González de Aledo AL, Delgado Moya FdP, Badía IM. SHAP model explainability in ECMO-PAL mortality prediction: A critical analysis. *Intensive Care Med*. 2023;49(12):1559. doi:10.1007/s00134-023-07252-z
65. Yi F, Yang H, Chen D, et al. XGBoost-SHAP-based interpretable diagnostic framework for Alzheimer's disease. *BMC Med Inform Decis Mak*. 2023;23(1):137. doi:10.1186/s12911-023-02238-9
66. Fan R, Zhou Y, Chen X, et al. *Porphyromonas gingivalis* outer membrane vesicles promote apoptosis via msRNA-regulated DNA methylation in periodontitis. *Microbiol Spectr*. 2023;11(1):e0328822. doi:10.1128/spectrum.03288-22
67. Yadalam PK, Arumuganainar D, Anegundi RV, et al. CRISPR-Cas-based adaptive immunity mediates phage resistance in periodontal red complex pathogens. *Microorganisms*. 2023;11(8):2060. doi:10.3390/microorganisms11082060