AMP_0: Species-Specific Prediction of Antimicrobial Peptides using Zero and Few Shot Learning

Sadaf Gull, Fayyaz Minhas*

Abstract—Evolution of drug-resistant microbial species is one of the major challenges to global health. Development of new antimicrobial treatments such as antimicrobial peptides needs to be accelerated to combat this threat. However, the discovery of novel antimicrobial peptides is hampered by low-throughput biochemical assays. Computational techniques can be used for rapid screening of promising antimicrobial peptide candidates prior to testing in the wet lab. The vast majority of existing antimicrobial peptide predictors are *non-targeted* in nature, i.e., they can predict whether a given peptide sequence is antimicrobial, but they are unable to predict whether the sequence can target a particular microbial species. In this work, we have used zero and few shot machine learning to develop a targeted antimicrobial peptide activity predictor called AMP_0. The proposed predictor called AMP_0 takes the peptide sequence and any N/C-termini modifications together with the genomic sequence of a microbial species to generate targeted predictions. Cross-validation results show that the proposed scheme is particularly effective for targeted antimicrobial prediction in comparison to existing approaches and can be used for screening potential antimicrobial peptides in a targeted manner with only a small number of training examples for novel species. AMP_0 webserver is available at http://ampzero.pythonanywhere.com.

Index Terms— Antibiotic resistance, Antimicrobial peptides, Zero/Few shot learning, Target microbial species.

1 Introduction

Antibiotics play a significant role in protecting humans from microbial infections. The discovery and use of antibiotics since the 1930s has helped in treating serious infections and saved many lives [1]. Resistance against antibiotics in microbes was detected in the 1960s and it prompted an evolutionary arms race between microbes and antibiotics [2]. Antimicrobial resistance is currently a major global health crisis. The number of deaths due to infections caused by antibiotic resistance annually is increasing and is estimated to reach up to 10 million by 2050 [3]. The World Health Organization (WHO) has generated a list of antibiotic resistant bacterial species that are a major threat to global health and require urgent development of novel therapeutics against them: *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* [4].

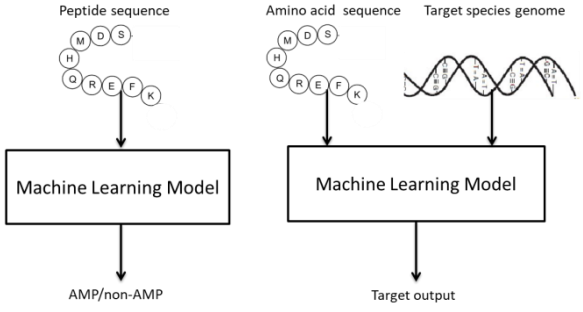
To handle the issue of antibiotic resistance, the development of novel antibiotics is necessary [1]–[5]. In comparison to the rate of development of antimicrobial resistance, the pace of discovery or development of new antibiotics is very slow: in the last 2 decades only two new classes of antibiotics were introduced for clinical use [4]. Consequently, the use of vaccines, lysins, antibodies, probiotics, bacteriophages and antimicrobial peptides (AMPs) is becoming popular in therapeutics as alternatives to antibiotics [1]. For designing new drugs, the use of AMPs is rapidly gaining attention [1], [6]–[8]. AMPs exhibit different biological activities against microbes, e.g., bacteria, viruses, fungi, etc. [1], have higher inhibition rates than antibiotics, and can potentially slow down the evolution of antibiotic resistance as well [8].

Machine learning approaches and artificial intelligence tools can potentially deal more effectively with biological data especially proteins [9]–[13]. Numerous machine learning based tools/models have been developed for predicting protein functions which uses Deep Neural Networks (DNN) and Support Vector Machines (SVM) [10], [14]–[18]. Potential AMP candidates need to be tested and evaluated experimentally before entering clinical trials. The prediction of AMPs using machine learning techniques reduces the cost of identifying the effectiveness of a peptide sequence against microbial species in the wet lab by pre-screening potential antimicrobial peptides. A number of machine learning based AMP predictors are available in the literature [19]–[24]. The primary issue with these *un-targeted* predictors is that they are unable to predict whether a given peptide sequence will be effective against a given target microbial species or not (see Fig 1). Only a small number of targeted predictors exist in the literature but they are not able to generate predictions for novel microbial species [25]–[27]. Vishnepolsky et al. developed a predictor for 6 different gram-negative bacterial strains [26]. The AMP predictor by Kleandrova et al. used 70 different gram-negative strains

- *Sadaf Gull is with the PIEAS Biomedical Informatics Lab, Department of Computer & Information Sciences, Pakistan Institute of Engineering and Applied Sciences, PO Nilore, Islamabad 45650, Pakistan. E-mail: sadafzakarkhan@gmail.com.*
- *Fayyaz Minhas is with the Department of Computer Science, University of Warwick, Coventry, UK and PIEAS Biomedical Informatics Lab, PIEAS. E-mail: fayyaz.minhas14@alumni.colostate.edu*

of bacteria in training to predict antimicrobial and cytotoxic activity of individual amino acids in a peptide sequence for different strains [25]. Although they covered a large set of bacterial species, their method can generate predictions for only specific strains of gram-negative bacterial strains. Unavailability of their predictor for public use is also a limitation [25]. The major drawback in targeted predictors is their inability of predicting peptide's antimicrobial activity for novel microbial species. The prediction of antimicrobial activity of a peptide without knowing the microbial species against which the peptide is effective is not meaningful.

In this work, we have developed a machine learning model to overcome this limitation. The proposed model takes amino acid sequence of a peptide and the genomic sequence of a target microbial species to predict the effectiveness of the peptide against that species in a targeted manner.



(a) **(b)**

Fig. 1. A general framework of machine learning predictors for (a) non targeted and (b) targeted predictions

A targeted predictor which predicts a peptide's effectiveness for species on which model was not trained with better accuracy, requires a new strategy in modeling. Design of such predictors promoted ZSL strategy to be used in modeling. ZSL is a new concept in the field of machine learning whose learning strategy learns attributes of class labels instead of the labels. The concept of Zero shot learning (ZSL) is to predict the instances of a class whose zero instances were available for training. Many variants of ZSL strategies have been proposed from which are being used in the field of machine learning [28]–[34]. Few shot learning also strengthens the concept of ZSL in which it is assumed that very few examples were available at training time for a class but the prediction of instances of that class at test time ensures better generalization. Different techniques for FSL have also been proposed and their results are far better than conventional machine learning models [35]–[39]. The use of ZSL/FSL in the field of machine learning for classification of objects [34], [38], [40] classification of videos [41], and transfer learning [42], [43] is very common. However the application of zero-shot/few-shot learning in the biological domain is still not very common. This manuscript focuses on using this strategy for prediction of antimicrobial peptides by introducing ZSL/FSL in biological domain.

2 METHODS

2.1 Data collection and preprocessing

For constructing the dataset used for training and evaluation of our machine learning models, we have used DBAASP version 2 [44]. DBAASP has been widely used in recent studies in this field [25]–[27], [45], [46]. It contains a total of 12, 984 peptide sequences and their experimentally verified minimum inhibitory concentrations (MICs) against various target microbial species. In order to construct our dataset from DBAASP, we have used peptides with length greater than 5 amino acids whose experimentally validated MICs are available in micro molar (μM) or microgram per milliliter ($\mu g/mL$). We also ensured that the genomes of the target species are available in NCBI [47] and that each peptide in our dataset has at least one target species for which its MIC was $\leq 25\ \mu g/mL$ [26]. The details of different filtration stages to extract the dataset of our interest are given in Table-1. DBAASP reports the effectiveness of a peptide sequence against multiple strains of a microbial species. We have taken the minimum MIC of a peptide across different strains of a species as its MIC against that species. All MIC values have been converted to $\mu g/mL$ [25]. Our final dataset comprises of 5,710 peptides that are effective against a total of 336 different microbial species. The details of individual peptides and their MICs against their target species is given in supplementary material.

As an additional preprocessing step, we have scaled the MIC scores using a sigmoidal curve such that MIC scores $\leq 25\ \mu g/mL$ are mapped onto +1 and those $\geq 100\ \mu g/mL$ are mapped to -1 (see Fig. 2). For this purpose, we have utilized a sigmoid rescaling function which maps raw MIC scores y as follows:

$$y' = s\left(-\frac{y-55}{10}\right), \text{ with } \ s(z) = 2\left(\frac{e^z}{1+e^z}\right) - 1.$$

This rescaling ensures that subsequent processing and machine learning models are not affected by large variations in MICs across different target species and peptides which can vary from a few $\mu g/mL$ to more than 2000 $\mu g/mL$. If the MIC of a peptide is not known for a species, its rescaled score is set at 0.0.

TABLE 1.

FILTERING CRITERIA APPLIED TO DBAASP DATABASE TO OBTAIN REQUIRED DATASET

Filtering criteria	**Number of peptides**
DBAASP monomer peptides	12,984
Sequences with length >5	12,517
Sequences with microbial targets (excluding cancers)	9,890
Sequences with MIC in (μM) or (μg/mL)	8,045
Sequences with target species genomes available in NCBI [47]	8,025
Sequences with at least one target species with MIC ≤ 25 μg/mL	5,710

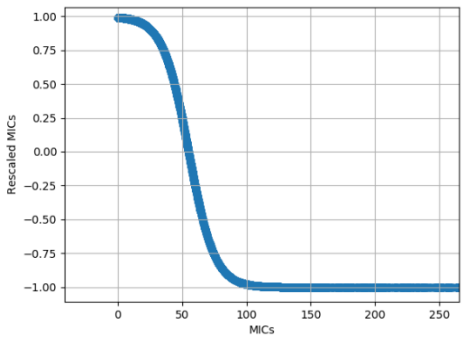


Fig. 2. MICs converted to continuous labels between -1 to +1 using bipolar sigmoid function

2.2 Feature extraction

To predict antimicrobial activity of a peptide against given species through machine learning, we need features of peptide and genomic sequence of target microbial species as discussed below (see Fig. 3).

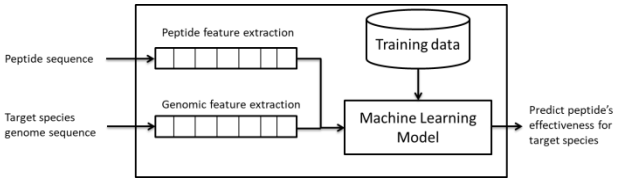


Fig. 3. Proposed model framework using features of peptide and genomic sequences

Amino Acid Sequence features

In order to obtain peptide-level features, we have used one-hot encoding of the peptide sequence that results in a 40-dimensional feature vector (frequency count of 20 L-amino acids and 20 D-amino acids). The feature representation models the type of amino acid (L and D) in the peptide sequence separately as peptide bioactivity is dependent upon the type of amino acids [48]–[51]. The resulting feature vectors for a given peptide is normalized to unit norm. We have also analyzed 2-mer composition which results in a $40^2 = 1600$-dimensional feature vector [52].

DBAASP [44] also provides information about N-terminus and C-terminus modifications of peptides which can play a significant role in their antimicrobial activity. Modification at N-terminus and C-terminus of peptides can change their biological activity [53]. We have used one-hot encoding to capture information about C- and N-terminus modifications in our feature representation. The sequence features are concatenated with C and N termini features. Details about the different types of C and N termini modifications are given in supplementary information.

Genomic features

In order to perform targeted prediction of antimicrobial activity of a peptide sequence against a particular species through machine learning, we need to extract species-level features as well. The literature reports the use of mono, di, tri and tetra-nucleotide composition of genomic sequences for comparison or clustering of genomes [54]–[61]. As a consequence, we have extracted features from complete genomes of species downloaded from NCBI [47]. For feature extraction the counts of 1-mer, 2-mer, 3-mer and 4-mer are calculated from a given genome sequence and normalized to unit norm resulting in a 340-dimensional feature representation.

2.3 Prediction Models

To predict whether a given peptide sequence will be effective against a target microbial species or not, we have proposed a zero-shot machine learning model. We compare the proposed model to a conventional machine learning model as a baseline as discussed below. In order to aid the reader in understanding our modeling approach for baseline and zero-shot predictors, we denote a peptide sequence by its d-dimensional feature vector x_i, $i = 1, \ldots, 5710$ whereas a particular microbial species is represented by an a-dimensional attribute vector s_j, for $j = 1, \ldots, 336$ based on its genomic sequence. We denote the rescaled MIC of a peptide x_i against species s_j by the target variable y_{ij}. The prediction problem can then be expressed as finding a mathematical function $f(x_i, s_j; \Theta)$ parameterized by learnable parameters Θ that can predict the effectiveness of a sequence x_i for microbial species s_j.

Baseline models

We have chosen Radial Basis Function SVM [62], XGBoost [63] , Neural network [64] and k-nearest neighbor [65] as baseline models due to their widespread use and ease of modeling. For this purpose, in order to predict the effectiveness of a given peptide sequence against a microbial species, we construct a joint feature representation $\boldsymbol{\phi}_{ij} = \begin{bmatrix} \boldsymbol{x_i} \\ \boldsymbol{s_j} \end{bmatrix}$ by concatenating peptide and species level features with the associated training label y_{ij} set to +1 (antimicrobial) if the MIC of peptide $\boldsymbol{x_i}$ for species $\boldsymbol{s_j}$ is $\leq 25\,\mu g/mL$ and -1 (non-antimicrobial) if the MIC is $\geq 100\,\mu g/mL$. A conventional machine learning model like SVM, XGBoost or neural network can then be trained over such a data set.

Zero and Fewshot learning

In this work, we propose to model the problem of targeted antimicrobial activity prediction through zero shot learning (ZSL) [34]. Widely used in object classification and computer vision, ZSL allows a classification model to generate predictions for novel classes which were not available at training time [30]–[32]. This is achieved by learning the definition of a class through an attribute vector representation instead of predicting class labels directly as in conventional classification. Many variants of ZSL have been proposed in the literature [28]–[34]. While ZSL assumes that no examples of a novel class presented during testing are available for training, the related case of few-shot learning aims at building a machine learning model such that only a few training examples are available for the target class [35]–[39]. Few Shot Learning (FSL) techniques perform significantly better than conventional classification methods when the number of training examples is very small [35]–[37].

The problem of targeted antimicrobial activity prediction is

ideally suited to zero and few shot learning: in typical machine learning guided design of wet lab experiments for screening potential peptides that are effective against a target microbial species, no or very few peptides with known labels are available for training. Furthermore, in order to predict how effective a peptide is against a novel microbial species for which no or very few training examples are available, we can model the target microbial species as a class represented by an attribute vector based on its genomic sequence. In this work, we have used the ZSL scheme given by Romera-Paredes and Torr [34]. For predicting the MIC of a peptide sequence for a target species, the discriminant function used by the ZSL model of Romera-Paredes and Torr [34] can be written as $f(x_i, s_j; \Theta) = x_i^T \Theta s_j$ with the learnable weight matrix $\Theta \in \mathbb{R}^{d \times a}$. If the number of peptides and species (classes) available during training are m and z, respectively and the rescaled MIC scores for each of the peptide against each microbe is represented by the $m \times z$ matrix $Y \in [-1,1]^{m \times z}$, the learning problem for ZSL can be formulated as the following optimization problem:

$$\Theta^* = \underset{\Theta \in \mathbb{R}^{d \times a}}{\operatorname{argmin}} \|X^T \Theta S - Y\|_{Fro}^2 + (\gamma \|\Theta S\|_F^2 + \lambda \|X^T \Theta\|_F^2 + \gamma\lambda \|\Theta\|_F^2)$$

Here, $X \in \mathbb{R}^{d \times m}$ and $S \in \mathbb{R}^{a \times z}$ represent matrices of all peptide features (m examples each with a d-dimensional feature vector) and attributes of microbial species (z classes each with a attributes), respectively. The first term represents the loss function with the aim of minimizing the error between predicted and target MICs. The second term $(\gamma \|\Theta S\|_F^2 + \lambda \|X^T \Theta\|_F^2 + \gamma\lambda \|\Theta\|_F^2)$ is the regularization factor that ensures smoothness of the prediction function $f(x, s; \Theta)$ and sparsity of the weight matrix Θ through penalization of the Frobenius norm $\|\cdot\|_F^2$ of respective matrices. γ and λ are regularization hyper-parameters. In addition to better performance over benchmark datasets, another reason for choosing this ZSL implementation is the existence of a computationally efficient closed-form solution of its underlying optimization problem which can be written as follows:

$$\Theta^* = (XX^T + \gamma I)^{-1} X Y S^T (SS^T + \lambda I)^{-1}$$

Once the optimal weight matrix Θ^* has been obtained, the predictions for a peptide (represented by the feature vector x) for species (represented by the attribute vector s) can be generated by the decision function $f(x, s; \Theta^*) = x^T \Theta^* s$. Note that this decision function can be used for generating predictions both for novel peptides and novel species provided their attribute representation s is available. The most likely target species for a given peptide can be identified by simply ranking the resulting decision function scores across a given list of potential target species.

This formulation can be kernelized for non-linear kernels as well by applying the Representer theorem to the underlying optimization problem [34]. For this purpose, an $m \times m$ sized kernel matrix K with $K_{ij} = k(x_i, x_j)$ is computed over the training data using a kernel function such as the radial basis function (RBF) $k(a, b) = \exp(-\kappa \|a - b\|^2)$ with the hyperparameter $\kappa > 0$. The closed form solution of the kernelized ZSL optimization problem requires calculation of an $m \times a$ sized instance-attribute association matrix A from training data as follows (see [34] for details):

$$A = (K^T K + \gamma I)^{-1} K Y S (S^T S + \lambda I)^{-1}$$

For inference or prediction of effectiveness of a peptide represented by a feature vector x against a microbial species represented by its attribute vector s, an m-dimensional vector of kernel scores $k(x) = [k(x, x_1) \quad k(x, x_2) \quad \cdots \quad k(x, x_m)]^T$ of the test example with each training example is computed and used in the kernelized prediction function $f(x, s; A) = k(x)^T A s$.

It is important to note that this framework extends seamlessly to FSL by simply adding further training instances for a target class. The hyperparameters of the model $(\gamma, \lambda, \kappa)$ are tuned through cross-validation.

The best performance of the model was found using $\gamma = 2.0$, $\lambda = 0.0001$, and the hyperparameter κ of RBF kernel is set to 2.0.

2.4 Performance Evaluation

We consider two practical use-cases of our system: 1) Target Species Ranking (TSR): given a set of microbial species for which labeled peptide sequences are available for training, predict the microbe that is most-likely to be targeted by a novel peptide sequence and, 2) Peptide Activity Prediction for Novel Species (PAP): predict whether a peptide is effective against a given species or not such that no or very few peptide examples for that species are available during training (i.e., Zero Shot or Few Shot Learning) (see Fig. 4). It is important to note that both these scenarios reflect practical use cases for biologists who are interested in machine-learning guided discovery for targeted antimicrobial peptides.

In order to evaluate the performance of baseline and proposed machine learning models for TSR, we have used 5-fold and 10-fold cross validation [66]. In 5 fold the dataset of 5,710 peptides is divided into 5 non-overlapping folds. A given model is trained on labeled examples of all peptides in

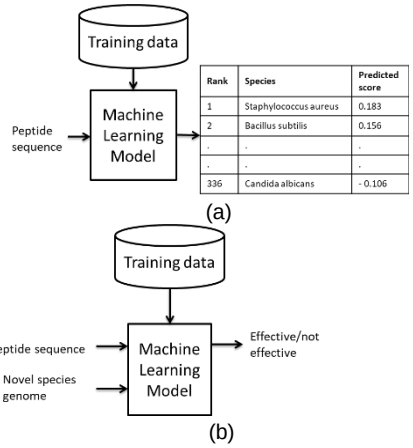


Fig. 4. (a) TSR requires a novel peptide sequence and predicts the microbe that is most likely to be targeted by that peptide (out of 336 given species); (b) PAP takes inputs of a peptide sequence and a novel species genome to predict whether a peptide is effective against a given species or not.

4 folds and tested on the remaining peptides. This process is repeated 5 times, once for each fold. For each test peptide in a fold, model scores for all 336 species are sorted in descending order. The rank of the highest scoring microbe that is a known target of the given test peptide (positive example) is used as a peptide-specific performance metric. This simple biologist-centric performance metric called Rank of First Positive Prediction (RFPP) is based on the premise that an ideal machine learning model should assign high score to a known target species of a given peptide sequence and, consequently, rank target species at lower ranks in the sorted list in comparison to non-target species [67]. As a result, for an ideal machine learning model, the RFPP for all test peptides should be 1.0. As discussed in the results section, we report the percentile-wise RFPP scores for all test peptides for different machine learning models together with a random predictor as experimental control. The RFPP score at a certain percentile p, henceforth denoted by RFPP(p) is defined as follows: RFPP(p) = q, if p% test peptides have at least one known target microbial species among their top q predictions (out of 336). Thus, for an ideal classifier RFPP(100) = 1, i.e., for every peptide, the top scoring species is a real target species of the given test peptide. RFPP is a biologist-centric metric as it tells us directly how often top-ranking predictions of a peptide can be expected to correspond to true target species and it can be directly used in experiment design. We have performed 10-fold cross validation but we did not find any significant change in results the the RFPP values (see supplementary material for results).

For PAP, i.e., predicting a peptide's effectiveness for a novel species, our proposed modeling approach takes peptide and genomic sequences as input and the score generated by the decision function of a machine learning model is used for classification of peptide sequences for individual species. In order to quantify predictive accuracy, a selected set of 17 test species from DBAASP with a small but sufficient number (75-180) of known positive and negative peptide examples is used (details given in Table-3). For ZSL, the model is trained on all examples from other species and its predictive performance is evaluated for individual species in Table-3 using area under the receiver operating characteristic curve (AUC-ROC) as a performance metric [68]. For few shots learning (FSL), a few positive and negative examples of a test species (1, 2, 4, 8 and half of all available examples for that species) are randomly sampled for training together with all examples from all other species and the model is evaluated on the remaining examples of the test species. This process is repeated 20 times with different species-level training and test examples to get average AUC-ROC scores and their standard deviation.

3 Results

In this section, we discuss the results for the two learning tasks below.

3.1 Target Species Ranking (TSR)

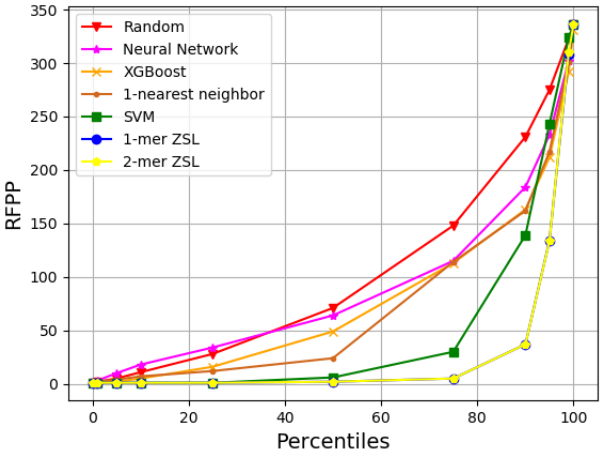


Fig. 5. Percentile-wise Rank of First Positive Prediction (RFPP) scores for various predictive models.

Fig 5 shows the percentile-wise RFPP scores for all classifiers. As discussed in section 2.4, the ideal RFPP score for all peptides is 1.0. For the random classifier that generates a random score for a given example, the median RFPP is 75, i.e., for 50% test peptides in cross-validation, a true target species is within the top 75 (out of 336) predictions. In contrast, for XGBoost and SVM baseline models, the median RFPPs are 50 and 10, respectively. However, the proposed model performs much better than these baseline models: the RFPP for the proposed model at the 75th percentile is 1.0, i.e., for up to 75% peptides, the top prediction by the model is correct. This clearly shows the effectiveness of the proposed prediction scheme for identifying the correct target species of a peptide.

TABLE 2.

RFPP PREDICTION SCORES GENERATED BY VARIOUS BASELINE AND PROPOSED MODEL

	SVM		XGBOOST		1-nearest neighbor	Neural network	ZSL	
Percentiles	1-mer	2-mer	1-mer	2-mer	1-mer	1-mer	1-mer	2-mer
0	1	1	1	1	1	1	1	1
1	1	1	1	1	2	3	1	1
5	3	3	1	1	4	10	1	1
10	6	6	1	1	7	18	1	1
25	23	15	2	1	12	34	1	1
50	65	50	9	6	24	64	2	2
75	129	112	40	30	114	115	5	3
90	176	165	161	139	162	184	37	23
95	218	213	248	243	217	233	134	113
99	277	284	328	324	301	302	308	298
100	333	329	336	336	336	335	336	336

We have analyzed the performance of both 1-mer and 2-mer as features in our proposed model as well as for baseline models. In the TSR case, both representations are used separately whose results are given in Table-2. The 2-mer representation worked well for baseline models SVM and XGBOOST while for nearest neighbor and neural network 1-mer worked well. For the proposed ZSL model, the 2-mer representation gives marginally better results as shown in Table-2. In the case of PAP the baseline model trained using XGBOOST worked well with 1-mer representation with 2-mer ZSL giving marginally better results. The proposed ZSL approach is also significantly better than k-nearest neighbor. For various values of k, RFPP scores are given in the supplementary material.

To ensure that homologous peptides are not in both training and test folds, we have used CD-Hit [4] to cluster the 5,710 peptides into 329 clusters with a threshold value of 40% identity. The cross validation strategy ensures that peptides belonging to a cluster are selected in the same fold. The results of cluster-wise analysis for various machine learning models is given below which shows that the performance of the proposed ZSL model is still significantly better than other approaches. These results have been added to the supplementary material.

3.2 Peptide Activity Prediction for Novel Species (PAP)

Table-3 shows the results of various machine learning models for the Peptide Activity Prediction (PAP) task. In this task the objective is to evaluate whether a given machine learning model can correctly predict peptides that target a novel species for which none or very few training examples are available. For this purpose, we compare the performance of conventional machine learning models (SVM, XGBoost), the proposed Zero Shot Learning (ZSL) and Few Shot Learning (FSL) models in addition to existing state of the art non-targeted antimicrobial activity predictors (CAMP [22] [70] and AMAP [19]). For this use case, XGBoost with amino acid composition features performed significantly better than SVM (results not shown for brevity). However, the prediction performance of XGBoost was typically no better than a random classifier especially when the number of training examples from a given test species was Similarly very small (see Supplementary Information for complete results)., existing state of the art methods such as CAMP [22] and AMAP [19] do not give satisfactory predictive performance for the chosen species. In contrast, the proposed few shot learning model performs significantly better than other methods with an expected increase in prediction accuracy when the number of training examples of a species is increased.

we have done an additional experiment in which we have generated 2000 random peptides whose lengths are between 6 to 80 (chosen randomly) for testing such that for each test species, the number of random peptides is kept the same as the number of original negative examples for that species.

The results of PAP are given in the supplementary material which shows that from given 17 species the results of 3 species have higher false positive rates: *Aspergillus fumigatus*, *Candida glabrata* and *proteus miriblis*. For rest of the species, the results remain largely unchanged. We have added results of this analysis to the supplementary material.

In order to analyze the performance of the proposed method in terms of the genomic distance between training and test species, we have done an additional experiment. First we define and calculate the genetic distance between two species as the Euclidean distance between their respective genomic feature representations. Then, for a given species at test time, we calculate the genomic distance of its closest species which has at least T examples in training (for T=1 and T=100). In order to study the relationship between prediction accuracy and genomic distance to training species, we plot the AUC-ROC of examples for a given test species against the genomic distance to its closest training species and calculate the correlation coefficient (for T=1 and T=100). Results can be seen in the supplementary material, which shows there is negative correlation between the predictive accuracy and genomic distance, i.e., as expected, if the test species is similar to a training species, the predictions can be expected to be more accurate. However, the plot shows that the proposed model does not undergo an abrupt degradation in predictive performance when generating predictions for test species that have no similar species in training. These results have been added to the supplementary material accompanying the paper.

3.3 Feature Analysis

For analyzing the importance of different features, we have plotted the corresponding weight values of the $d \times a$ parameter matrix θ obtained after training (where d is the number of protein features and a is the number of attributes for a given species). Note that the magnitude of a particular weight parameter reflects the relative importance of its corresponding feature. Fig. 6 shows the sum of the absolute weight values for each L- (small) and D- type (captialized) amino acid in the feature representation. The large magnitudes of weights of amino acids G, g, F, f, P, p, and w correlates with literature findings about the importance of these amino acids in AMPs. Specifically, the Proline-rich peptides (P) have capability of bacterial cell penetration. Cysteine-rich peptides (C) have excessive ability of pore formation in a membrane which leads to high antimicrobial activity. Glycine (G) improves antimicrobial activity of peptides and potentially targets fungi, Gram-negative bacteria, and cancer cells. Tryptophan (W) can penetrate a microbial cell membrane and is effective against numerous antibiotic-resistant bacteria. Phenylalanine-rich (F) AMPs have higher antimicrobial activity against Gram-positive bacteria, Gram-negative bacteria and yeast without hemolytic activity [3]. Cysteine (C) is also an important amino acid in natural antimicrobial peptides of vertebrates, invertebrates and plants [2], have excessive ability of pore formation in a membrane which leads to high antimicrobial activity [3]. We have discussed the importance of these features in the revised manuscript.

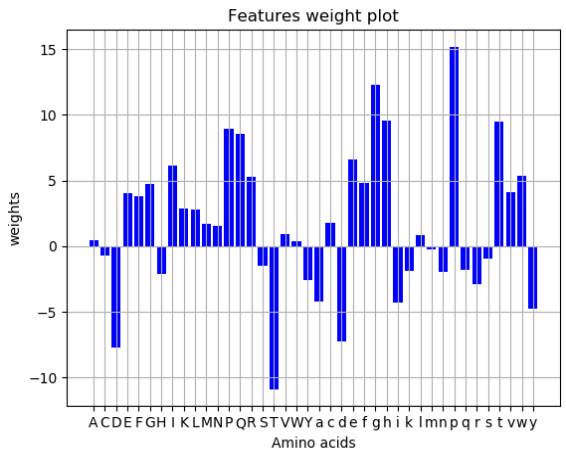


Fig. 6. Weight plot of ZSL with monomer representation of peptides

3.4 Webserver

The webserver developed for proposed model is available at the URL:http://ampzero.pythonanywhere.com. The webserver takes a peptide sequence in FASTA format along with any C-terminus and N-terminus modifications as input together with the genome of a species in order to predict the degree of effectiveness of the peptide against the given species. Additionally, the user can upload a list of known positive and negative example peptide sequences for the given species for generating few shot learning based predictions.

4 Conclusion

We have developed a targeted antimicrobial activity predictor called AMP_0 which can predict the effectiveness of a given peptide sequence against a given target species. The use of zero and few shot learning in the proposed model helps in overcoming the shortcomings of conventional machine learning techniques for this purpose. Our cross-validation analysis shows that the proposed model can perform better than existing approaches and it can be easily integrated in experimental discovery of antimicrobial peptide sequences for novel species.

TABLE 3.

Results for Peptide Activity Prediction for Novel Species. The first column indicates the type of the different test species used in this analysis. The species name together with the total number of positive (P) and negative (N) examples available for that species are given in the second column. Results for zero shot learning (ZSL) in which no examples of the given test species are included in training are shown for the proposed ZSL model. For few shot learning results for different number of training examples (1, 2, 4, 8 and Half of all available examples) of the target species are shown. In the interest of relevance and brevity results for XGBoost are shown only when half of the available examples are used for training. CAMP and AMAP are existing state of the art predictors for antimicrobial activity and the prediction results were obtained using their respective webservers. Values in bold indicate the highest prediction performance. Note that the average AUC-ROC across multiple runs is reported together with the standard deviation (in parenthesis).

Species Type	**Species Name ↓**	**Machine Learning Models**								
		ZSL	**FSL**					**XGBoost**	**CAMP**	**AMAP**
	No. of Tr. Examples →	0	1	2	4	8	Half	Half		
Fungus	*Aspergillus fumigatus*	0.746	0.807	0.806	0.820	0.835	0.882	0.614	0.798	0.545
	(P: 44, N: 33)	(0.056)	(0.059)	(0.041)	(0.046)	(0.043)	(0.043)	(0.073)	(0.051)	(0.055)
	Candida glabrata	0.652	0.594	0.620	0.628	0.691	0.781	0.677	0.350	0.489
	(P35: , N:47)	(0.056)	(0.087)	(0.084)	(0.088)	(0.072)	(0.052)	(0.087)	(0.047)	(0.081)
	Candida parapsilosis	0.430	0.473	0.507	0.562	0.663	0.789	0.639	0.660	0.662
	(P:51 , N:33)	(0.088)	(0.094)	(0.106)	(0.093)	(0.089)	(0.055)	(0.120)	(0.069)	(0.075)
	Candida tropicalis	0.755	0.712	0.735	0.771	0.803	0.865	0.669	0.703	0.561
	(P:88 , N:16)	(0.078)	(0.072)	(0.076)	(0.078)	(0.071)	(0.042)	(0.066)	(0.076)	(0.058)
	Cryptococcus neoformans	0.504	0.497	0.487	0.518	0.628	0.628	0.541	0.576	0.581
	(P:167 , N:14)	(0.102)	(0.110)	(0.104)	(0.103)	(0.068)	(0.068)	(0.078)	(0.089)	(0.084)
	Saccharomyces cerevisiae	0.405	0.627	0.634	0.650	0.681	0.788	0.604	0.388	0.448
	(P:132 , N:36)	(0.061)	(0.064)	(0.060)	(0.064)	(0.072)	(0.052)	(0.046)	(0.053)	(0.043)
	Fusarium oxysporum	0.856	0.914	0.919	0.932	0.943	0.961	0.696	0.418	0.396
	(P:125 , N:18)	(0.045)	(0.040)	(0.043)	(0.038)	(0.033)	(0.021)	(0.094)	(0.049)	(0.033)
Gram Negative Bacteria	*Enterobacter aerogenes*	0.468	0.588	0.599	0.646	0.731	0.826	0.773	0.550	0.443
	(P:36 , N:49)	(0.061)	(0.084)	(0.063)	(0.088)	(0.078)	(0.051)	(0.069)	(0.067)	(0.069)
	Erwinia amylovora	0.478	0.385	0.450	0.543	0.714	0.892	0.907	0.877	0.385
	(P112: , N:35)	(0.059)	(0.062)	(0.068)	(0.069)	(0.069)	(0.047)	(0.032)	(0.021)	(0.046)
	Pasteurella multocida	0.722	0.745	0.807	0.876	0.924	0.957	0.914	0.528	0.295
	(P:37 , N:53)	(0.052)	(0.088)	(0.080)	(0.042)	(0.026)	(0.019)	(0.046)	(0.067)	(0.039)
	Proteus mirabilis	0.714	0.729	0.733	0.748	0.767	0.836	0.731	0.377	0.269
	(P:27 , N:105)	(0.052)	(0.045)	(0.047)	(0.054)	(0.045)	(0.048)	(0.079)	(0.046)	(0.065)
	Proteus vulgaris	0.710	0.780	0.794	0.818	0.840	0.909	0.667	0.465	0.567
	(P:84 , N:34)	(0.038)	(0.050)	(0.050)	(0.050)	(0.045)	(0.030)	(0.071)	(0.048)	(0.048)
	Serratia marcescens	0.782	0.843	0.864	0.883	0.886	0.921	0.571	0.397	0.418

	(P:48 , N:62)	(0.040)	(0.042)	(0.031)	(0.028)	(0.039)	(0.021)	(0.068)	(0.084)	(0.046)
Gram	*Listeria innocua*	0.686	0.688	0.710	0.738	0.763	0.804	0.672	0.371	0.402
Positive	(P:64 , N:36)	(0.038)	(0.067)	(0.064)	(0.056)	(0.062)	(0.057)	(0.052)	(0.048)	(0.032)
Bacteria	*Streptococcus mutans*	0.437	0.570	0.597	0.616	0.825	0.825	0.767	0.472	0.706
	(P:129 , N:11)	(0.119)	(0.118)	(0.116)	(0.136)	(0.045)	(0.045)	(0.108)	(0.099)	(0.083)
	Streptococcus pneumoniae	0.591	0.619	0.609	0.622	0.623	0.701	0.507	0.3081	0.384
	(P:86 , N:17)	(0.077)	(0.090)	(0.089)	(0.079)	(0.077)	(0.066)	(0.071)	(0.057)	(0.077)
	Streptococcus pyogenes	0.660	0.733	0.737	0.747	0.873	0.873	0.669	0.569	0.737
	(P:161 , N:09)	(0.050)	(0.044)	(0.044)	(0.053)	(0.037)	(0.037)	(0.069)	(0.156)	(0.062)

ACKNOWLEDGMENTS

Sadaf Gull is supported by a grant under indigenous 5000 Ph.D. fellowship scheme by the Higher Education Commission (HEC) of Pakistan.

REFERENCES

[1] B. Aslam *et al.*, "Antibiotic resistance: a rundown of a global crisis," *Infection and drug resistance*, vol. 11, p. 1645, 2018.

[2] C. L. Ventola, "The antibiotic resistance crisis: part 1: causes and threats," *Pharmacy and therapeutics*, vol. 40, no. 4, p. 277, 2015.

[3] J. M. Blair, "A climate for antibiotic resistance," *Nature Climate Change*, vol. 8, no. 6, p. 460, 2018.

[4] M. Lakemeyer, W. Zhao, F. A. Mandl, P. Hammann, and S. A. Sieber, "Thinking Outside the Box—Novel Antibacterials To Tackle the Resistance Crisis," *Angewandte Chemie International Edition*, vol. 57, no. 44, pp. 14440–14475, 2018.

[5] C. N. Spaulding, R. D. Klein, H. L. Schreiber, J. W. Janetka, and S. J. Hultgren, "Precision antimicrobial therapeutics: the path of least resistance?," *NPJ biofilms and microbiomes*, vol. 4, no. 1, p. 4, 2018.

[6] F. Kampshoff, M. D. Willcox, and D. Dutta, "A Pilot Study of the Synergy between Two Antimicrobial Peptides and Two Common Antibiotics," *Antibiotics*, vol. 8, no. 2, p. 60, 2019.

[7] F. Costa, C. Teixeira, P. Gomes, and M. C. L. Martins, "Clinical Application of AMPs," in *Antimicrobial Peptides*, Springer, 2019, pp. 281–298.

[8] G. Yu, D. Y. Baeder, R. R. Regoes, and J. Rolff, "Predicting drug resistance evolution: insights from antimicrobial peptides and antibiotics," *Proceedings of the Royal Society B: Biological Sciences*, vol. 285, no. 1874, p. 20172687, 2018.

[9] Z. Teng, M. Guo, Q. Dai, C. Wang, J. Li, and X. Liu, "Computational prediction of protein function based on weighted mapping of domains and GO terms," *BioMed research international*, vol. 2014, 2014.

[10] A. Sokolov, C. Funk, K. Graim, K. Verspoor, and A. Ben-Hur, "Combining heterogeneous data sources for accurate functional annotation of proteins," in *BMC bioinformatics*, 2013, vol. 14, p. S10.

[11] P. Radivojac *et al.*, "A large-scale evaluation of computational protein function prediction," *Nature methods*, vol. 10, no. 3, p. 221, 2013.

[12] A. Valencia, "Automatic annotation of protein function," *Current opinion in structural biology*, vol. 15, no. 3, pp. 267–274, 2005.

[13] B. Rost, J. Liu, R. Nair, K. O. Wrzeszczynski, and Y. Ofran, "Automatic prediction of protein function," *Cellular and Molecular Life Sciences CMLS*, vol. 60, no. 12, pp. 2637–2650, 2003.

[14] T. L. Campos, P. K. Korhonen, R. B. Gasser, and N. D. Young, "An evaluation of machine learning approaches for the prediction of essential genes in eukaryotes using protein sequence-derived features," *Computational and Structural Biotechnology Journal*, 2019.

[15] M. Kulmanov, M. A. Khan, and R. Hoehndorf, "DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier," *Bioinformatics*, vol. 34, no. 4, pp. 660–668, 2017.

[16] A. S. Rifaioglu, T. Doğan, M. J. Martin, R. Cetin-Atalay, and V. Atalay, "DEEPred: Automated Protein Function Prediction with Multi-task Feed-forward Deep Neural Networks," *Scientific reports*, vol. 9, no. 1, p. 7344, 2019.

[17] R. Fa, D. Cozzetto, C. Wan, and D. T. Jones, "Predicting human protein function with multi-task deep neural networks," *PloS one*, vol. 13, no. 6, p. e0198216, 2018.

[18] S. Hua and Z. Sun, "Support vector machine approach for protein subcellular localization prediction," *Bioinformatics*, vol. 17, no. 8, pp. 721–728, 2001.

[19] S. Gull, N. Shamim, and F. Minhas, "AMAP: Hierarchical multi-label prediction of biologically active and antimicrobial peptides," *Computers in biology and medicine*, vol. 107, pp. 172–181, 2019.

[20] P. Bhadra, J. Yan, J. Li, S. Fong, and S. W. Siu, "AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest," *Scientific reports*, vol. 8, no. 1, p. 1697, 2018.

[21] M. Torrent, V. M. Nogués, and E. Boix, "A theoretical approach to spot active regions in antimicrobial proteins," *BMC bioinformatics*, vol. 10, no. 1, p. 373, 2009.

[22] F. H. Waghu, R. S. Barai, P. Gurung, and S. Idicula-Thomas, "CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides," *Nucleic acids research*, vol. 44, no. D1, pp. D1094–D1097, 2015.

[23] W. Lin and D. Xu, "Imbalanced multi-label learning for identifying antimicrobial peptides and their functional types," *Bioinformatics*, vol. 32, no. 24, pp. 3745–3752, 2016.

[24] P. Agrawal and G. P. Raghava, "Prediction of Antimicrobial Potential of a Chemically Modified Peptide From Its Tertiary Structure," *Frontiers in Microbiology*, vol. 9, p. 2551, 2018.

[25] V. V. Kleandrova, J. M. Ruso, A. Speck-Planche, and M. N. Dias Soeiro Cordeiro, "Enabling the discovery and virtual screening of potent and safe antimicrobial peptides. simultaneous prediction of antibacterial activity and cytotoxicity," *ACS combinatorial science*, vol. 18, no. 8, pp. 490–498, 2016.

[26] B. Vishnepolsky *et al.*, "Predictive Model of Linear Antimicrobial Peptides Active against Gram-Negative Bacteria," *Journal of chemical information and modeling*, vol. 58, no. 5, pp. 1141–1151, 2018.

[27] A. Speck-Planche, V. V. Kleandrova, J. M. Ruso, and M. DS Cordeiro, "First multitarget chemo-Bioinformatic model to enable the discovery of antibacterial peptides against multiple gram-positive pathogens," *Journal of chemical information and modeling*, vol. 56, no. 3, pp. 588–598, 2016.

[28] M. Palatucci, D. Pomerleau, G. E. Hinton, and T. M. Mitchell, "Zero-shot learning with semantic output codes," in *Advances in neural information processing systems*, 2009, pp. 1410–1418.

[29] Z. Zhang and V. Saligrama, "Zero-shot learning via semantic similarity embedding," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4166–4174.

[30] R. Socher, M. Ganjoo, C. D. Manning, and A. Ng, "Zero-shot learning through cross-modal transfer," in *Advances in neural information processing systems*, 2013, pp. 935–943.

[31] M. Norouzi *et al.*, "Zero-shot learning by convex combination of semantic embeddings," *arXiv preprint arXiv:1312.5650*, 2013.

[32] Y. Fu, T. M. Hospedales, T. Xiang, and S. Gong, "Transductive multi-view zero-shot learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2332–2345, 2015.

[33] E. Kodirov, T. Xiang, and S. Gong, "Semantic autoencoder for zero-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3174–3183.

[34] B. Romera-Paredes and P. Torr, "An embarrassingly simple approach to zero-shot learning," in *International Conference on Machine Learning*, 2015, pp. 2152–2161.

[35] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, 2017, pp. 4077–4087.

[36] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1199–1208.

[37] S. Gidaris and N. Komodakis, "Dynamic few-shot visual learning without forgetting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4367–4375.

[38] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," *arXiv preprint arXiv:1711.04043*, 2017.

[39] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," 2016.

[40] X. Sun, H. Xv, J. Dong, H. Zhou, C. Chen, and Q. Li, "Few-shot Learning for Domain-specific Fine-grained Image Classification," *IEEE Transactions on Industrial Electronics*, 2020.

[41] K. Cao, J. Ji, Z. Cao, C.-Y. Chang, and J. C. Niebles, "Few-shot video classification via temporal alignment," *arXiv preprint arXiv:1906.11415*, 2019.

[42] B. Liu, X. Wang, M. Dixit, R. Kwitt, and N. Vasconcelos, "Feature space transfer for data augmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9090–9098.

[43] Z. Luo, Y. Zou, J. Hoffman, and L. F. Fei-Fei, "Label efficient learning of transferable representations acrosss domains and tasks," in *Advances in Neural Information Processing Systems*, 2017, pp. 165–177.

[44] M. Pirtskhalava *et al.*, "DBAASP v. 2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides," *Nucleic acids research*, vol. 44, no. D1, pp. D1104–D1112, 2015.

[45] M. Youmans, C. Spainhour, and P. Qiu, "Long short-term memory recurrent neural networks for antibacterial peptide identification," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 498–502.

[46] T. S. Win, A. A. Malik, V. Prachayasittikul, J. E. S Wikberg, C. Nantasenamat, and W. Shoombuatong, "HemoPred: a web server for predicting the hemolytic activity of peptides," *Future medicinal chemistry*, vol. 9, no. 3, pp. 275–291, 2017.

[47] N. R. Coordinators, "Database resources of the national center for biotechnology information," *Nucleic acids research*, vol. 44, no. Database issue, p. D7, 2016.

[48] F. Cava, H. Lam, M. A. De Pedro, and M. K. Waldor, "Emerging knowledge of regulatory roles of D-amino acids in bacteria," *Cellular and Molecular Life Sciences*, vol. 68, no. 5, pp. 817–831, 2011.

[49] M. L. Mangoni *et al.*, "Effect of natural L-to D-amino acid conversion on the organization, membrane binding, and biological function of the antimicrobial peptides bombinins H," *Biochemistry*, vol. 45, no. 13, pp. 4266–4276, 2006.

[50] R. H. Baltz, "Daptomycin: mechanisms of action and resistance, and biosynthetic engineering," *Current opinion in chemical biology*, vol. 13, no. 2, pp. 144–151, 2009.

[51] Y. Kawai *et al.*, "Structural and functional differences in two cyclic bacteriocins with the same sequences produced by lactobacilli," *Appl. Environ. Microbiol.*, vol. 70, no. 5, pp. 2906–2911, 2004.

[52] C. Leslie, E. Eskin, and W. S. Noble, "The spectrum kernel: A string kernel for SVM protein classification," in *Biocomputing 2002*, World Scientific, 2001, pp. 564–575.

[53] E. Crusca Jr *et al.*, "Influence of N-terminus modifications on the biological activity, membrane interaction, and secondary structure of the antimicrobial peptide hylin-a1," *Peptide Science*, vol. 96, no. 1, pp. 41–48, 2011.

[54] S. Karlin and I. Ladunga, "Comparisons of eukaryotic genomic sequences," *Proceedings of the National Academy of Sciences*, vol. 91, no. 26, pp. 12832–12836, 1994.

[55] S. Karlin, A. M. Campbell, and J. Mrazek, "Comparative DNA analysis across diverse genomes," *Annual review of genetics*, vol. 32, no. 1, pp. 185–225, 1998.

[56] S. Kariin and C. Burge, "Dinucleotide relative abundance extremes: a genomic signature," *Trends in genetics*, vol. 11, no. 7, pp. 283–290, 1995.

[57] S. Karlin, "Global dinucleotide signatures and analysis of genomic heterogeneity," *Current opinion in microbiology*, vol. 1, no. 5, pp. 598–610, 1998.

[58] H. Nakashima, K. Nishikawa, and T. Ooi, "Di. erences in Dinucleotide Frequencies of Human, Yeast, and Escherichia coli Genes," *DNA Research*, vol. 4, no. 3, pp. 185–192, 1997.

[59] H. Nakashima, M. Ota, K. Nishikawa, and T. Ooi, "Genes from nine genomes are separated into their organisms in the dinucleotide composition space," *DNA Research*, vol. 5, no. 5, pp. 251–259, 1998.

[60] D. T. Pride, R. J. Meinersmann, T. M. Wassenaar, and M. J. Blaser, "Evolutionary implications of microbial genome tetranucleotide frequency biases," *Genome research*, vol. 13, no. 2, pp. 145–158, 2003.

[61] M. Takahashi, K. Kryukov, and N. Saitou, "Estimation of bacterial species phylogeny through oligonucleotide frequency distances," *Genomics*, vol. 93, no. 6, pp. 525–533, 2009.

[62] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[63] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[64] K. Gurney, *An introduction to neural networks*. CRC press, 1997.

[65] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[66] Z. John Lu, "The elements of statistical learning: data mining, inference, and prediction," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 173, no. 3, pp. 693–694, 2010.

[67] F. ul A. Afsar Minhas, B. J. Geiss, and A. Ben-Hur, "PAIRpred: Partner-specific prediction of interacting residues from sequence and structure," *Proteins: Structure, Function, and Bioinformatics*, vol. 82, no. 7, pp. 1142–1155, 2014.

[68] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 233–240.

[69] Y. Huang, B. Niu, Y. Gao, L. Fu, and W. Li, "CD-HIT Suite: a web server for clustering and comparing biological sequences," *Bioinformatics*, vol. 26, no. 5, pp. 680–682, 2010.

[70] M. N. Gabere and W. S. Noble, "Empirical comparison of web-based antimicrobial peptide prediction tools," *Bioinformatics*, vol. 33, no. 13, pp. 1921–1929, 2017.

[71] K. Pearson, "VII. Note on regression and inheritance in the case of two parents," *proceedings of the royal society of London*, vol. 58, no. 347–352, pp. 240–242, 1895.

[72] A. de Breij *et al.*, "The antimicrobial peptide SAAP-148 combats drug-resistant bacteria and biofilms," *Science translational medicine*, vol. 10, no. 423, p. eaan4044, 2018.

[73] J.-L. Dimarcq, P. Bulet, C. Hetru, and J. Hoffmann, "Cysteine-rich antimicrobial peptides in invertebrates," *Peptide Science*, vol. 47, no. 6, pp. 465–477, 1998.

[74] J. Wang *et al.*, "Antimicrobial peptides: Promising alternatives in the post feeding antibiotic era," *Medicinal Research Reviews*, vol. 39, no. 3, pp. 831–859, 2019.

Sadaf Gull is a PhD scholar in the Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan. She is doing her PhD under indigenous PhD fellowship by Higher Education Commission (HEC). Her area of research is "Machine Learning in Biomedical Informatics".

Fayyaz Minhas is currently with the Department of Computer Science, University of Warwick, Coventry, UK and the Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan. Dr. Minhas received his PhD degree in Bioinformatics from Colorado State University, USA on a Fulbright Scholarship. He has also been awarded the National Youth Award by the Government of Pakistan for his contributions to science and technology. His research focuses on applications of machine learning in Bioinformatics and the analysis of biomedical data