Combating Antimicrobial Resistance with Artificial Intelligence

Thesis submitted in partial fulfillment
for the Award of Degree

Doctor of Philosophy

by

Ritesh Sharma

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
(BANARAS HINDU UNIVERSITY)
VARANASI - 221005

Roll No. 19071002 *Year 2023*

Chapter 7

Conclusion and Future Directions

7.1 Conclusion

The research in this thesis is focused on developing AI-assisted models and deploying them as web-based tools to help wet lab researchers in combating the alarming situation of antimicrobial resistance.

In the thesis, various models have been developed, which make use of different concepts, namely (i) Deep Learning: Different deep learning algorithms (BiLSTM, BiGRU, BiTCN, and 1DCNN) have been utilized. These deep learning algorithms can automatically extract features from the peptides, also known as DLF. (ii) Transfer Learning: The concept of transfer learning was accomplished by utilizing pretrained embeddings from [14] and from [15]. (iii) Combination of HCF and DLF : HCF can be considered while training deep learning algorithms so that they can learn the features (DLF) that are missing from HCF so that a better feature vector can be constructed by concatenating HCF and DLF. (iv) Ensemble Learning: Different ensemble strategies (Min/Max combiner, Soft voting, Hard Voting, Fuzzy non-linear, Fuzzy gompertz, Fuzzy distance) have been experimented.

While developing different models, we found that : (i) The model built by combining different models using an ensemble strategy performed better than the individual

models. This might be due to the fact the deep learning algorithms (BiLSTM, BiGRU, BiTCN, and 1DCNN) are heterogeneous. As a result, peptides misclassified by one or the other classifier might be correctly classified by the rest, resulting in improved performance. (ii) The concept of transfer learning always improves performance. The reason for this may be that transfer learning helps to learn a new task by transferring knowledge from a related task that has already been learned, which saves time and helps in better generalization [10, 11, 12]. (iii) In certain cases, the performance of the model can be improved by combining both HCF and DLF. The probable reason for this is that the peptides can have a wide range of properties. Therefore, it is difficult to have HCF that encodes all the properties of peptides. Deep learning algorithms can automatically extract features from data and, therefore, learn some complex properties that are unavailable in the form of HCF. Similarly, the importance of HCF cannot be disregarded, as it is not always possible for deep learning algorithms to learn the optimal features required for classification.

In this thesis, we developed various models that wet lab researchers can use as a pipeline to narrow down the search space while discovering new peptide-based antibiotics. For example, our proposed model XAI-INVENT can be used for identifying ABPs from the protein sequences. Then the identified ABPs can be fed to our proposed model ESKAPEE-MICpred, which provides the MIC value(s) for the identified ABPs against WHO priority ESKAPEE pathogens, which are the leading cause of nosocomial infections throughout the world. Furthermore, the potent ABPs obtained from ESKAPEE-MICpred cannot enter the clinical trial directly as they can be toxic to the host. Therefore, our proposed model EnDL-HemoLyt (which can identify low hemolytic peptides) can be used for further filtering.

7.2 Future Directions

Based on the research work presented in this thesis, the following are promising future directions that can be explored.

1. Due to the emergence of different pandemics and advancements in techniques involved in *in-vitro*, *in-vivo* testing, the antimicrobial peptide-related data is increasing rapidly. Because of this, many more peptides will be available than there are now in a few years, which can be utilized to improve the performance of the deployed tool. However, the proposed framework(s) are not designed to adapt to new data points. Therefore, the tool(s) developed as a part of it will become obsolete over time. Hence, in the future, the proposed framework(s) can be improved by considering the concept of continual learning.
2. We have developed different models and deployed them as web-based tools in different works. In the future, a hybrid tool consisting of multiple stages can be developed that takes protein sequences as input and then uses multiple tools at various stages (each stage related to some desired property of antimicrobial peptide-based drug) and directly provides peptides, which wet-lab researchers can consider for *in-vitro* testing.
3. The concept of protein-peptide docking can also be integrated to narrow the search space further in the near future.
4. Generative AI techniques like Generative Adversarial Networks, Reinforcement learning, etc., can be used in future for optimizing existing drugs and for designing entirely new molecules from scratch that can target specific diseases.
5. Advanced methodologies related to deep learning, transfer learning, and ensemble learning that have evolved through time can be considered in the future.