Preface

Artificial intelligence (AI) has made significant strides in protein and peptide sequence modeling, revolutionizing bioinformatics and molecular biology. AI in peptide sequence modeling relies on large datasets, advanced algorithms, and computational power to analyze vast amounts of data. These technologies continuously evolve, leading to better predictions and deeper insights into the structure and function of peptides, with implications for fields like alternative medicine. Various deep learning models help predict peptide functions from amino acid sequences by efficiently identifying similarities, differences, and conserved regions of functionally similar peptides. This aids in classifying peptides into families, which helps in inferring the characteristic functions of newly discovered or poorly characterized peptides. Moreover, AI-driven generative models can generate novel peptide sequences optimized for specific characteristics or functions, aiding accelerated drug discovery, engineering, and design.

Researchers have proposed multiple frameworks for classifying and subsequently identifying peptides as therapeutic or non-therapeutic. But little or no thought has been invested in tackling the computational challenges of building efficient and reliable frameworks. Some of the major challenges are: (i) decreasing the time to train a model based on deep learning algorithms like biLSTM, which are sequential and hence consume a lot of resources to learn patterns in a dataset; (ii) since the end-users of these models are biologists, it is imperative to make them available as web or mobile applications so that the potential therapeutic peptides can be screened before synthe-

sis, experimental validation, and production; (iii) since the applications based on deep learning need to be deployed on resource-constrained devices, reducing the size of the model without compromising its performance is the need of the hour; (iv) the models should be prevented from going into obsolescence post their deployment (as an application) because data is constantly increasing; and (v) the models must be explainable, and this very feature should be used to understand the facilitating/ inhibiting properties of the peptides that grant them their characteristic function. This understanding may lead to the design and development of novel and optimal therapeutic peptide sequences.

In this thesis, all the aforementioned challenges have been tackled by proposing robust novel frameworks for classifying, identifying, optimizing, and discovering therapeutic peptides, including antimicrobial peptides (AMPs), neurological peptides (NPs), and blood-brain barrier penetrating peptides (B3P2s).