ZEMING LIN is a third-year student studying computer science. As a firm believer in using computers to automate as many tasks as possible, he is fascinated by deep learning, particularly its hands-off nature. Deep learning is a set of algorithms in machine learning, a discipline that examines the construction and study of algorithms that can learn and adapt from data. The science is based on representation learning, which converts raw data input into a representation that facilitates analysis. A particularly unique feature of deep learning is that it is fast and scalable, unlike most machine learning algorithms. In an increasingly data-driven world, deep learning holds a lot of potential for processing massive volumes of data, making it applicable to various fields, such as content serving and computational biology.

Lin works in Professor Yanjun (Jane) Qi’s lab, where the research focus is on machine learning and bioinformatics. His research explores the applications of deep learning to bioinformatics. With a series of 28 tasks, he seeks to predict proteins’ functional properties using deep neural networks, which are statistical learning algorithms inspired by their biological counterpart. Due to their adaptive nature, deep neural networks are commonly used to estimate functions that can depend on many inputs. In Lin’s research, deep neural networks are used to build a model that can predict an output, like a protein local property, from a given input, like a portion of a protein sequence.

Each of the tasks in his research project has a number of labels, which are the local properties of the protein, predetermined from prior studies. The tasks include predicting proteins’ solvent accessibility, transmembrane topology, two secondary structure alphabets and identifying structural elements like DNA-binding residues, protein-binding residues, signal peptides and coiled-coil regions. Their project aims to build a model using known successful neural network techniques to create a survey of how well these techniques run on a protein sequence data set. They also hope to train their model on all tasks at the same time. This “multitask” approach allows the model to use the commonality among the tasks to achieve better results.

Using deep learning techniques, Lin and Professor Qi hope to create a comprehensive study and to release their dataset to pioneer more research into the applications of deep learning to computational biology. Deep learning has been applied heavily to other fields such as image search and facial detection, but not to computational biology, although there seems to be much potential in that field. They look to submit their