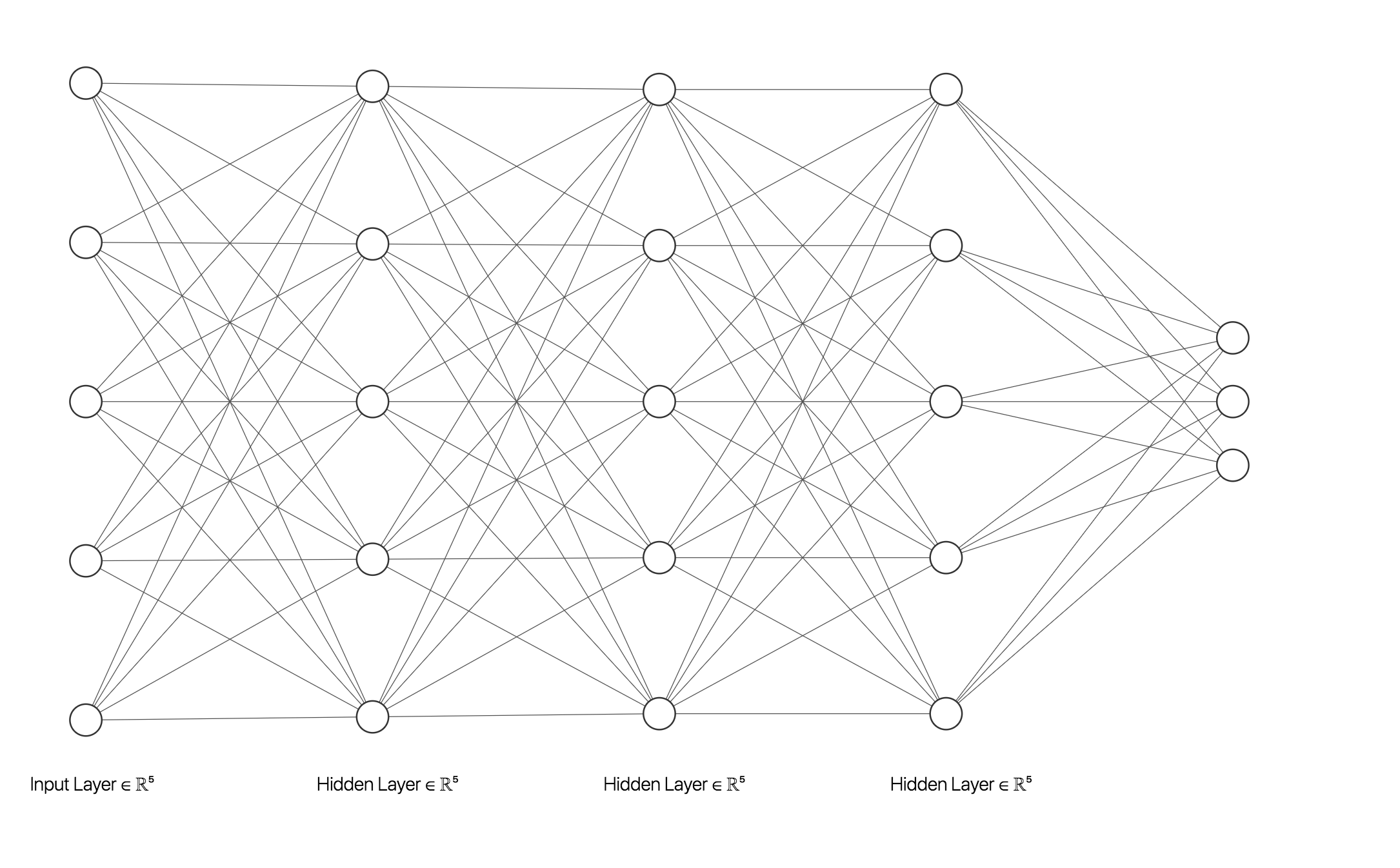
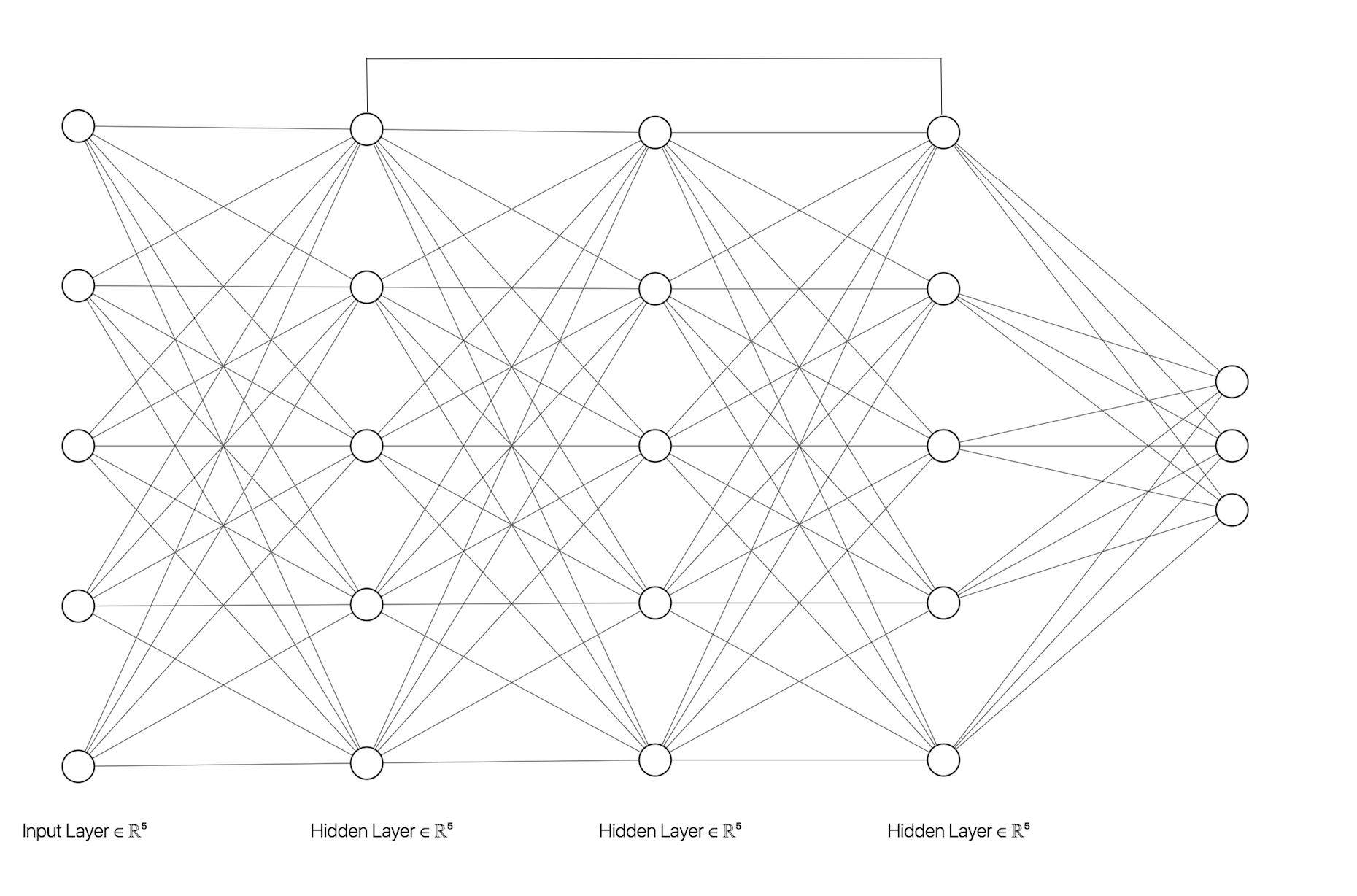
Protein Data Analysis

Predicting PH, temperature, and PI (isoelectric point) using the output of an NMR (nuclear magnetic resonance). The output would be helpful in quickly diagnosing patients and acquiring appropriate treatment.

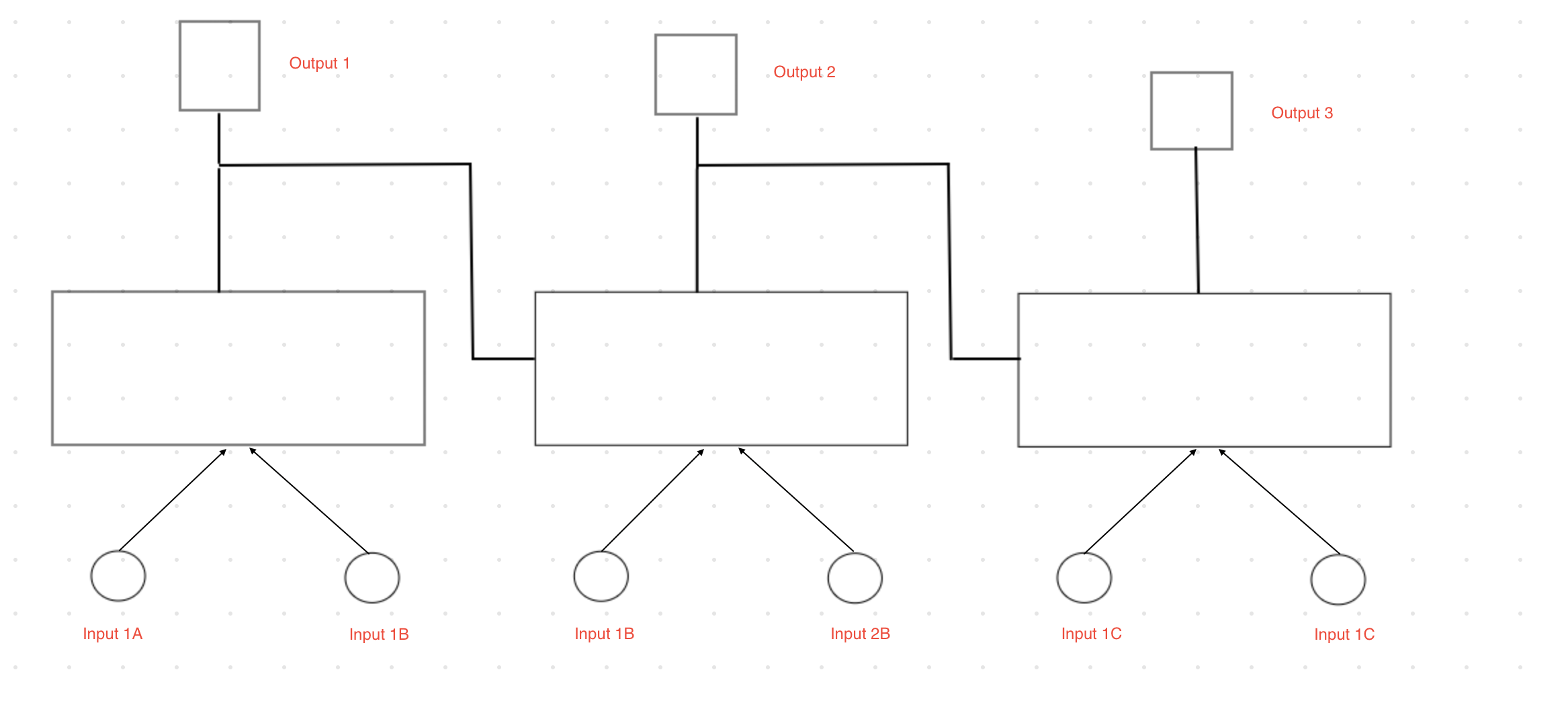
The first model trained on is the following uses a variation of the structure below which represents the traditional fully connected layers.



This structure unfortunately fails at deeper layers due to various complications which end up reducing accuracy hence the structure below, which is inspired from ResNet was used. This structure allows the models to learn faster due to the outputs influencing deeper layers and reduces the problem of cost functions increasing on long iterations.



The model used to train the current data uses a 22-layer model that can acquire a 95% accuracy at an 85% confidence. Additional work and deeper models would prove helpful for a higher accuracy.



The next cycle of development would focus on the prediction of the sequence of amino acids on various proteins given their features. The features would first pass through a fully connected layers as given above and those value would then be used fed into the RNN (recurrent neural network) to predict individual amino acids, similar to an NLP (natural language processing) model.

This would allow doctors to diagnose irregularities within patients much faster as a it brings up most likely options for a problem, saving time in a doctor’s already busy schedule. A well-trained model would also allow prediction of protein structures of new or evolving diseases. Identifying the structures would also advance medical development for new or more resistant diseases.