

INTISARI

PREDIKSI STRUKTUR SEKUNDER PROTEIN MENGGUNAKAN MODEL DEEP LEARNING DENGAN KONVOLUSI DAN BIDIRECTIONAL GATED RECURRENT UNIT

M. Rizky Luthfianto
12/336268/PA/15084

Prediksi struktur sekunder protein adalah permasalahan penting di dalam bioinformatika karena struktur tiga dimensi protein menentukan fungsi dari sebuah protein. Sedangkan, protein itu sendiri adalah unit fungsional utama dalam organisme hidup dan terlibat dalam banyak proses biologis dalam sel.

Dalam penelitian ini diimplementasikan sebuah model Deep Learning untuk kasus prediksi struktur sekunder protein yang menggabungkan antara Convolutional Neural Network (CNN) dan Recurrent Neural Network (RNN). Pada model ini terdapat tiga lapisan konvolusi dengan panjang filter berbeda yang disusun secara paralel untuk mengekstrak fitur lokal. Karena adanya *long-range dependencies* dalam sekuen asam amino yang dibuktikan oleh Zhou & Troyanskaya (2014), maka digunakan Gated Recurrent Unit yang disusun secara bidireksional untuk menangkap fitur global.

Model yang dihasilkan penelitian ini mampu meraih akurasi Q8 sebesar 66.9% pada CB513 dan 69.7% pada *test set* CullPDB yang umum dijadikan sebagai *benchmark*.

Kata kunci: Prediksi Struktur Sekunder Protein, Deep Learning, Convolutional Neural Network, Bidirectional Gated Recurrent Unit, Sequence Labelling

ABSTRACT

PROTEIN SECONDARY STRUCTURE PREDICTION USING DEEP LEARNING WITH CONVOLUTIONS AND BIDIRECTIONAL GATED RECURRENT UNIT

M. Rizky Luthfianto
12/336268/PA/15084

Protein secondary structure prediction is an important problem in bioinformatics because the three-dimensional structure of a protein determines the function of the protein itself. Meanwhile, protein is the main functional units in living organisms and are involved in many biological processes in cells.

In this research, we implemented a Deep Learning model for protein secondary structure prediction that combines Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). In this model, there are three layers with different filter lengths which are arranged in parallel to extract local features. Because long-range dependencies are proven to be exists in amino acid *sequences* as discovered by Zhou & Troyanskaya (2014), we used Bidirectional Gated Recurrent Units to capture the global context features.

The resulting model is able to achieve both Q8 accuracy of 66.9% on CB513 and 69.7% on CB513, which both datasets are commonly used as benchmark.

Keywords: Protein secondary structure prediction, Deep Learning, Convolutional Neural Network, Bidirectional Gated Recurrent Unit, Sequence Labelling