

## INTISARI

### **PREDIKSI STRUKTUR SEKUNDER PROTEIN MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK DAN SUPPORT VECTOR MACHINE**

Vincent Michael Sutanto

16/398531/PA/17492

Prediksi struktur sekunder protein adalah salah satu permasalahan dalam disiplin ilmu Bioinformatika. Prediksi struktur sekunder protein dilakukan untuk mengetahui fungsi dari suatu protein. Prediksi struktur sekunder protein dilakukan dengan mengklasifikasi setiap sekuens struktur primer protein ke dalam bentuk sekuens struktur sekunder protein. Permasalahan ini termasuk ke dalam kategori *Sequence Labelling* yang mana bisa diselesaikan dengan pendekatan Pembelajaran Mesin.

*Convolutional Neural Network* dan *Support Vector Machine* adalah 2 metode Pembelajaran Mesin yang sering digunakan dalam masalah klasifikasi. Pada penelitian ini *Convolutional Neural Network* dipilih karena kemampuannya untuk mengambil dan memperkaya pola hubungan dari sekuens struktur primer protein. *Support Vector Machine* juga dipilih untuk digunakan untuk memprediksi struktur sekunder protein berdasarkan data masukan berupa *feature maps* dari arsitektur *Convolutional Neural Network*.

Pada penelitian ini kombinasi arsitektur *Convolutional Neural Network* dan *Support Vector Machine* mampu menghasilkan tingkat akurasi yang lebih tinggi dibandingkan dengan arsitektur *Convolutional Neural Network* sederhana. Arsitektur kombinasi menghasilkan tingkat akurasi data uji CullPDB Q3 sebesar 80.75% dan Q8 sebesar 68.84% (Q3 meningkat 1.4% dan Q8 meningkat 0.71%), sementara itu pada data uji CB513 menghasilkan tingkat akurasi Q3 sebesar 78.51% dan Q8 sebesar 64.76% (Q3 meningkat 1.37% dan Q8 meningkat 0.67%).

Kata kunci : Prediksi struktur sekunder protein, *Convolutional Neural Network*, *Support Vector Machine*

## **ABSTRACT**

### **PROTEIN SECONDARY STRUCTURE PREDICTION USING CONVOLUTIONAL NEURAL NETWORK DAN SUPPORT VECTOR MACHINE**

Vincent Michael Sutanto  
16/398531/PA/17492

Protein secondary structure prediction is one of the problems in the Bioinformatics discipline. Protein secondary structure prediction is conducted in order to find the function of proteins. Protein secondary structure prediction is done by classifying each sequence of protein primary structure into the sequence of protein secondary structure. This problem is included in the Sequence Labeling category which can be solved with the Machine Learning approach.

Convolutional Neural Network and Support Vector Machine are 2 methods of Machine Learning that are often used in classification problems. In this study, the Convolutional Neural Network was chosen because of its ability to extract and enrich the relationship patterns of the protein primary structures sequences. Support Vector Machine was also chosen to be used to predict the protein secondary structures based on feature maps from the Convolutional Neural Network architecture.

In this study, the combination of Convolutional Neural Network and Support Vector Machine is able to produce a higher level of accuracy compared to simple Convolutional Neural Network architecture. The hybrid architecture produces accuracy as high as 80.75% (Q3) and 68.84% (Q8) for CullPDB test data (Q3 increased by 1.4% and Q8 increased by 0.71%), while in CB513 test data resulted in an accuracy of 78.51% (Q3) and 64.76% (Q8) (Q3 increased by 1.37% and Q8 increased by 0.67%).

**Keywords:** Protein Secondary Structure Prediction, Convolutional Neural Network, Support Vector Machine