

## ABSTRACT

HIV epidemic and high mortality rate encourage scientists and researchers in the world to find effective prevention and treatment method. P24 protein in HIV is an important part in the first step of HIV identification. Discovering epitope region in the P24 protein plays important role in the epitope-based vaccine development, it will save many people from death and opportunistic infection that caused by HIV.

One of machine learning technique that can be applied in epitope prediction is neural network. Epitope of HIV contain amino acids that have to be encoded before training using neural network. The orthogonal encoding was used to encode amino acid sequence, and sliding window method was applied to build training pattern of amino acid sequence whose different length and slightly number. This research would test the best window length of 9, 11, 13, 15, and 17. Optimal result of prediction obtained by doing twice of neural network training, called cascade neural network. The first network gave output that would be processed in the second network. The second network revised the output from the first network. Test carried out using 5-cross validation with 10 attempts.

The test results are measured with 3 indicators, sensitivity, spesificity and accuracy. Comparison of the result of first neural network and second neural network were tested with t-test statistic. While the best windows length were tested with ANOVA test. The result shows that cascade neural network could increased the sensitivity approximately 2% in each sliding window, while the accuracy and spesificity did not increase significantly. The best sensitivity score of 57,93% was obtained with 11 windows size, while the best spesificity score of 93,82% was obtained with 17 window size, and the best accuracy score of 92,41% was obtained with 9 window size.

Keyword: neural network, epitope, HIV



## INTISARI

Epidemik HIV dan tingginya angka kematian yang disebabkan oleh virus HIV mendorong para peneliti di dunia untuk menemukan obat dan metode pencegahan yang efektif. Protein P24 pada HIV merupakan bagian penting dalam langkah pertama identifikasi HIV. Menemukan lokasi epitop pada protein P24 sangat penting dalam pengembangan vaksin berbasis epitop, hal ini dapat menyelamatkan banyak orang dari kematian dan penyakit oportunistik yang disebabkan oleh HIV.

Salah satu teknik pembelajaran mesin yang dapat diaplikasikan dalam memprediksi epitop adalah jaringan syaraf. Epitop pada HIV terdiri dari asam amino yang harus dikodekan sebelum menjadi data pelatihan dalam jaringan syaraf. *Orthogonal Encoding* digunakan untuk mengkodekan sekuen asam amino dan metode *sliding window* digunakan untuk membentuk pola pelatihan asam amino yang memiliki perbedaan panjang. Penelitian ini akan menguji besar *window* yang terbaik antara 9, 11, 13, 15, dan 17. Hasil yang optimal diperoleh dengan dua kali proses pelatihan yang disebut jaringan syaraf bertingkat. Jaringan syaraf pertama memberikan output yang akan diproses pada jaringan syaraf kedua. Jaringan syaraf kedua dilatih untuk memperbaiki hasil jaringan syaraf pertama. Pengujian dilakukan dengan 5-fold *cross validation* dengan 10 kali percobaan.

Hasil pengujian diukur dengan 3 indikator yaitu sensitivitas, spesifisitas, dan akurasi. Perbandingan hasil jaringan syaraf pertama dan jaringan syaraf kedua diuji dengan uji statistik T. Sedangkan besar *window* terbaik diuji dengan uji statistik ANOV. Hasil pengujian menunjukkan bahwa jaringan syaraf bertingkat dengan dua kali pelatihan dapat meningkatkan sensitivitas sebesar 2% di setiap besar *window*, sedangkan hasil spesifisitas dan akurasi tidak naik secara signifikan. Hasil sensitivitas terbaik adalah 57,93% diperoleh dari besar *window* 11, sedangkan hasil spesifisitas terbaik adalah 93,82% diperoleh dari besar *window* 17, dan akurasi terbaik adalah 92,41% diperoleh dari besar *window* 9.

**Kata kunci** – jaringan syaraf, epitop, HIV