

INTISARI

MODIFIKASI PSO UNTUK SELEKSI FITUR PADA MODEL MULTILAYER ENSEMBLE UNTUK KLASIFIKASI METILASI DNA

Oleh

Abdulghoffar Lugas Aga Perwira

23/526067/PPA/06632

Identifikasi situs N4-metilsitosin (4mC) pada DNA merupakan permasalahan penting dalam studi epigenetik, namun metode eksperimental bersifat mahal dan tidak efisien. Pendekatan pembelajaran mesin menjadi alternatif yang menjanjikan, tetapi tingginya kompleksitas data biologis yang ditandai oleh dimensi intrinsik tinggi, redundansi fitur, dan overlap kelas sering menurunkan performa model serta menyebabkan konvergensi prematur pada metode optimasi konvensional.

Penelitian ini mengusulkan model klasifikasi 4mC berbasis seleksi fitur menggunakan Particle Swarm Optimization (PSO) varian, yaitu HPSO-SCAC, yang dikombinasikan dengan Multilayer Ensemble Model (MEM). Seleksi fitur bertujuan mereduksi kompleksitas data dan meningkatkan efisiensi komputasi, sedangkan MEM digunakan untuk meningkatkan akurasi klasifikasi secara bertahap. Eksperimen dilakukan pada enam dataset DNA dari spesies berbeda menggunakan fitur biologis hasil ekstraksi sekuens DNA dan protein.

Hasil eksperimen menunjukkan bahwa model baseline menghasilkan akurasi awal pada kisaran 84–88%. Penerapan seleksi fitur berbasis PSO mampu mereduksi jumlah fitur hingga 37–50% dari fitur awal, yang berdampak pada peningkatan efisiensi pelatihan model. Integrasi seleksi fitur dengan MEM meningkatkan akurasi menjadi sekitar 89–91% menggunakan PSO, dan mencapai performa terbaik menggunakan HPSO-SCAC dengan akurasi tertinggi 96 % pada dataset *E. coli* serta 88–92 persen pada dataset lainnya. HPSO-SCAC menunjukkan performa paling stabil, khususnya pada dataset dengan kompleksitas tinggi.

Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi seleksi fitur berbasis PSO adaptif dan Multilayer Ensemble Model mampu meningkatkan akurasi sekaligus efisiensi model klasifikasi 4mC, dengan performa yang bergantung pada karakteristik kompleksitas data biologis.

Kata Kunci: HPSO, PSO, Optimasi, Metilasi DNA, Ensemble Model

ABSTRACK

PARTICLE SWARM OPTIMIZATION MODIFICATION FOR FEATURE SELECTION IN MULTILAYER ENSEMBLE MODELS FOR DNA METHYLATION CLASSIFICATION

Abdulghoffar Lugas Aga Perwira

23/526067/PPA/06632

Identifikasi situs N4-methylcytosine (4mC) in DNA is an important problem in epigenetic studies; however, experimental methods are costly and inefficient. Machine learning approaches offer a promising alternative, but the high complexity of biological data, characterized by high intrinsic dimensionality, feature redundancy, and class overlap, often degrades model performance and can cause premature convergence in conventional optimization methods.

This study proposes a 4mC classification model that combines feature selection using a Particle Swarm Optimization (PSO) variant, namely HPSO-SCAC, with a Multilayer Ensemble Model (MEM). Feature selection aims to reduce data complexity and improve computational efficiency, while MEM is employed to progressively enhance classification accuracy. Experiments were conducted on six DNA datasets from different species using biologically derived features extracted from DNA and protein sequences.

The experimental results show that the baseline model achieved an initial accuracy in the range of 84–88%. PSO-based feature selection reduced the number of features by approximately 37–50% of the original set, improving training efficiency. Integrating feature selection with MEM increased accuracy to around 89–91% using PSO, and achieved the best performance with HPSO-SCAC—reaching a peak accuracy of 96% on the E. coli dataset and 88–92% on the other datasets. HPSO-SCAC demonstrated the most stable performance, particularly on datasets with high complexity.

Overall, this study demonstrates that combining adaptive PSO-based feature selection with a Multilayer Ensemble Model can improve both the accuracy and efficiency of 4mC classification, with performance depending on the complexity characteristics of the biological data.

Kata Kunci: HPSO, PSO, Optimasi, Metilasi DNA, Ensemble Model