

## BAB 5

### SIMPULAN DAN SARAN

Bab ini merangkum temuan akhir penelitian terkait efektivitas metode seleksi fitur Discriminant Fuzzy Pattern (DFP) dan performa Support Vector Machine (SVM) dalam klasifikasi stadium kanker payudara. Berdasarkan hasil evaluasi model dan batasan yang teridentifikasi, bab ini menyajikan simpulan komprehensif serta rekomendasi strategis untuk pengembangan penelitian selanjutnya.

#### 5.1 Simpulan

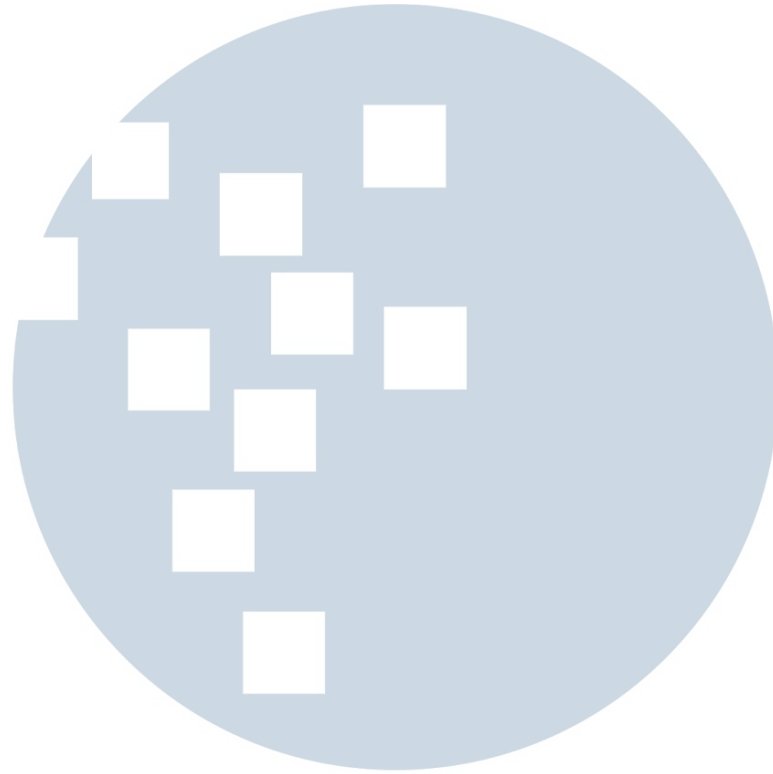
Penelitian ini berhasil menerapkan metode DFP untuk mereduksi dimensi data dari 60.661 gen menjadi 22 gen kandidat biomarker yang paling relevan. Dari hasil ini, 8 gen terbukti memiliki validasi biologis terkait mekanisme kanker, sedangkan 14 gen lainnya merupakan potensi biomarker baru yang bekerja secara sinergis. Penerapan algoritma SVM pada fitur terpilih berhasil mengklasifikasikan stadium kanker payudara (Early vs Late) dengan performa optimal pada skenario pembagian data 90:10. Model mencapai Akurasi 82%, F1-Score Makro 0.70, dan AUC 0.71, yang menunjukkan kemampuan generalisasi yang baik meskipun terdapat ketidakseimbangan data.

#### 5.2 Saran

Berdasarkan hasil analisis, kendala yang dihadapi, dan temuan selama penelitian berlangsung, berikut adalah beberapa saran yang direkomendasikan untuk pengembangan penelitian selanjutnya:

1. Mengingat kompleksitas data ekspresi genetik yang berdimensi tinggi, penelitian selanjutnya disarankan untuk mengeksplorasi teknik seleksi fitur yang lebih adaptif dan stabil dibandingkan metode tunggal. Pendekatan berbasis ansambel (*Ensemble Feature Selection*) atau metode berbasis *Deep Learning* (seperti *Autoencoders*) dapat dipertimbangkan guna menangkap interaksi non-linear antar gen yang mungkin terlewatkan oleh metode konvensional, sehingga menghasilkan subset fitur yang secara prediktif lebih kuat dan robust.
2. Peningkatan performa klasifikasi dapat dicapai melalui evaluasi komparatif menggunakan algoritma pembelajaran mesin yang lebih canggih dalam menangani *noise* pada data genomik. Pemilihan model yang tepat tidak hanya berpotensi meningkatkan akurasi prediksi, tetapi juga efisiensi komputasi dalam mengolah data skala besar dengan jumlah fitur yang lebih minimal.
3. Untuk menguji reliabilitas dan generalisasi model yang telah dibangun, penelitian mendatang sangat disarankan untuk menggunakan dataset baru yang bersumber dari repositori eksternal independen, seperti *Gene Expression Omnibus* (GEO) atau *International Cancer Genome Consortium* (ICGC). Selain memperluas cakupan populasi data, penggunaan dataset baru ini juga berfungsi sebagai sarana validasi silang untuk mengonfirmasi konsistensi

performa biomarker baru yang ditemukan dalam membedakan stadium kanker, baik secara komputasional maupun melalui verifikasi eksperimental (*Wet-Lab*).



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA