

BAB 5 SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil penelitian, model klasifikasi berbasis data ekspresi gen RNA-seq menunjukkan performa yang secara signifikan lebih unggul dibandingkan dengan model berbasis data ekspresi miRNA dalam membedakan stadium kanker prostat. Model yang dilatih menggunakan fitur-fitur genetik mampu mencapai metrik evaluasi yang lebih tinggi secara konsisten mencakup akurasi, presisi, *recall*, dan skor F1 dengan jumlah fitur yang lebih sedikit. Hal ini mengindikasikan bahwa model berbasis genetik memiliki kemampuan prediksi yang lebih stabil dan andal, serta efisiensi yang lebih tinggi dalam proses seleksi fitur.

Kombinasi metode seleksi fitur *DESeq2* dan RFE berbasis SVM menghasilkan performa terbaik dalam klasifikasi stadium kanker prostat pada pria kulit putih non-Hispanik. Model SVM yang dibangun dengan pendekatan ini mampu mencapai akurasi sebesar 96,1%, skor F1 makro 96%, dan ROC AUC sebesar 99,2%, hanya dengan menggunakan 18 fitur genetik terpilih. Jumlah fitur yang relatif sedikit ini mencerminkan efisiensi tinggi metode tersebut dibandingkan dengan pendekatan lain yang membutuhkan lebih banyak fitur untuk mencapai tingkat performa yang serupa.

Analisis lebih lanjut terhadap 18 gen yang dipilih oleh model terbaik mengidentifikasi bahwa lima di antaranya, yaitu *GLDC*, *ITLN1*, *ANKRD7*, *RECQL4*, dan *CSNK1A1L*, telah tervalidasi dalam literatur memiliki keterkaitan langsung dengan kanker prostat. Selain itu, empat gen lainnya, yaitu *SNORD83A*, *PSMD10P1*, *SOHLH2*, dan *OR1N2*, diketahui berperan dalam perkembangan berbagai jenis kanker pada organ lain. Meskipun nilai AUC individu dari gen-gen tersebut tergolong rendah, dengan maksimum hanya mencapai 0,63 dan nilai korelasi antara ekspresi gen dengan stadium kanker berada pada tingkat yang lemah hingga sedang, ketika digunakan secara kolektif melalui pendekatan *machine learning*, gen-gen ini menunjukkan performa prediktif yang sangat baik. Temuan ini menunjukkan bahwa penggunaan gen-gen secara kelompok dapat memberikan informasi yang lebih kuat dalam membedakan stadium kanker prostat, mencerminkan sifat poligenik dan kompleksitas biologis penyakit ini.

5.2 Saran

Untuk penelitian selanjutnya, disarankan menggunakan pendekatan klasifikasi *multiclass* dalam mendeteksi stadium kanker prostat. Dengan metode *multiclass*, model dapat membedakan lebih dari dua kelas stadium secara langsung, sehingga memberikan informasi yang lebih jelas dan akurat mengenai perkembangan kanker. Selain itu, penggunaan algoritma yang lebih kompleks seperti *deep learning* yang dapat meningkatkan akurasi dan kemampuan model dalam menangani data yang kompleks dan variatif, sehingga mendukung pengambilan keputusan klinis yang lebih tepat. Validasi model pada dataset yang lebih besar dan beragam dapat dilakukan untuk memvalidasi hasil yang ditemukan.

Peneliti juga dapat mempertimbangkan untuk langsung menggunakan metode RFE sejak awal sebagai pendekatan utama dalam seleksi fitur, guna menyederhanakan alur kerja dan mengurangi ketergantungan pada metode filter. Di sisi lain, jika metode filter tetap digunakan, eksplorasi ruang fitur yang lebih luas dapat dilakukan dengan memanfaatkan komputasi berkinerja tinggi, mengingat proses seleksi fitur berbasis filter bisa sangat memakan sumber daya. Terakhir, validasi eksternal pada dataset yang lebih besar dan beragam sangat dianjurkan untuk memastikan ketangguhan dan generalisasi model yang dikembangkan.

