

BAB 5 SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini berhasil merancang dan mengimplementasikan algoritma klasifikasi stadium kanker payudara berdasarkan data ekspresi gen (*gene expression*) dan ekspresi miRNA. Proses seleksi fitur dilakukan melalui kombinasi metode *Logistic Regression (L1)* dan *Recursive Feature Elimination (RFE)*, yang secara konsisten mampu mengidentifikasi fitur-fitur paling relevan dalam membedakan stadium awal dan stadium lanjut. Hasil analisis menunjukkan bahwa pendekatan seleksi fitur berbasis statistik ini memberikan kinerja klasifikasi yang lebih optimal dibandingkan metode filtrasi *Differentially Expressed Genes (DEG)* menggunakan *limma*. Dengan demikian, penggunaan *Logistic Regression (L1)* dan *RFE* terbukti lebih efektif dalam menyeleksi biomarker yang memiliki kontribusi signifikan terhadap prediksi stadium kanker payudara, dibandingkan pendekatan DEG konvensional.

Hasil evaluasi menunjukkan bahwa data ekspresi gen menghasilkan performa model yang lebih unggul dibandingkan dengan data miRNA maupun gabungan keduanya, baik dari segi akurasi maupun nilai *F1-score*. Salah satu hasil terbaik diperoleh pada data gen dengan 41 fitur terpilih, yang menghasilkan akurasi sebesar 92,6%, *precision* sebesar 92,3%, *recall* sebesar 89,7%, dan *F1-score* sebesar 90,8%, dengan waktu komputasi selama 27,53 detik. Hasil tersebut diperoleh menggunakan model *Voting Classifier* yang mengombinasikan tiga algoritma yaitu SVM, Random Forest, dan Logistic Regression. Penggunaan *Voting Classifier* terbukti mampu meningkatkan stabilitas dan konsistensi hasil prediksi secara keseluruhan. Temuan ini juga menunjukkan bahwa jumlah fitur yang relatif sedikit (kurang dari 50) tetap mampu menghasilkan performa klasifikasi yang tinggi, yang berarti pendekatan ini memiliki potensi untuk diterapkan dalam konteks klinis dengan efisiensi biaya yang lebih baik.

Selain itu, temuan penelitian ini menunjukkan bahwa kombinasi beberapa fitur ekspresi gen secara simultan mampu menghasilkan performa klasifikasi yang lebih unggul dibandingkan penggunaan fitur tunggal secara individual. Sebagai contoh, fitur-fitur seperti *PFN3*, *ALI39390.1*, dan beberapa fitur lainnya memberikan kontribusi yang lebih signifikan ketika digunakan secara terpadu

dibandingkan secara terpisah. Hal ini sejalan dengan karakteristik biologis kanker payudara yang kompleks dan dipengaruhi oleh interaksi *multifaktorial* antar gen. Fitur-fitur yang secara individu mungkin memiliki kekuatan prediktif terbatas, dapat saling melengkapi dan memperkuat sinyal biologis yang relevan ketika digunakan dalam kombinasi, sehingga meningkatkan akurasi model secara keseluruhan. Oleh karena itu, strategi pemanfaatan kombinasi biomarker dinilai lebih efektif dan memiliki potensi aplikatif yang lebih menjanjikan dalam prediksi stadium kanker payudara.

5.2 Saran

Penelitian ini membuka peluang untuk pengembangan lebih lanjut dalam bidang klasifikasi stadium kanker payudara berbasis data ekspresi gen dan miRNA. Salah satu arah pengembangan yang dapat dipertimbangkan adalah penerapan metode seleksi fitur yang lebih canggih, seperti *autoencoder* atau pendekatan pembelajaran representasi lainnya, untuk memperoleh fitur yang lebih informatif dan berkontribusi langsung terhadap peningkatan performa model. Integrasi data klinis tambahan, seperti status hormonal, usia pasien, dan riwayat keluarga, juga dinilai berpotensi memberikan nilai tambah dalam meningkatkan konteks analisis, sehingga menghasilkan model klasifikasi yang tidak hanya akurat secara statistik tetapi juga relevan secara biologis dan klinis.

Langkah ini penting dalam mengukur kemampuan model dalam menghadapi variasi data di dunia nyata. Selain itu, *pipeline* klasifikasi yang telah dibangun memiliki potensi untuk dikembangkan lebih lanjut ke dalam sistem pendukung keputusan klinis. Implementasi ini dapat membantu tenaga medis dalam proses diagnosis dan penentuan strategi pengobatan secara lebih tepat dan efisien, berbasis data molekuler pasien. Mengingat pentingnya efisiensi dalam konteks klinis, pengoptimalan terhadap jumlah biomarker yang digunakan menjadi langkah strategis. Menemukan kombinasi biomarker dengan jumlah *seminimal* mungkin, namun tetap memberikan performa klasifikasi yang tinggi, akan sangat bermanfaat dalam menekan biaya uji laboratorium dan meningkatkan efisiensi penerapan dalam skala layanan kesehatan yang lebih luas.