

BAB V

SIMPULAN SARAN

5.1. Simpulan

Penelitian ini memanfaatkan data DNA methylation kanker prostat dari XenaBrowser untuk membangun model klasifikasi biner menggunakan MLP. Setelah pemrosesan awal, hanya CpG site bertipe “cg” yang digunakan karena dominasinya lebih dari 90% serta kelengkapan datanya, sementara tipe lain dieksklusi akibat banyaknya nilai hilang. Proses seleksi fitur dilakukan melalui dua tahap: analisis DMP dengan $p\text{-value} < 0.05$, dan RFE menggunakan model klasifikasi SVM. Fitur yang terseleksi digunakan untuk melatih model MLP dalam berbagai konfigurasi, dan performanya dievaluasi menggunakan metrik akurasi, F1-score, dan AUC. Model terbaik dari 150 skenario berhasil mencapai akurasi pengujian 98% dan nilai AUC 1.00 dengan *loss* 15% pada kelompok ras kulit putih. Namun, pengujian terhadap kelompok ras minoritas menunjukkan hasil yang sangat kontras. Akurasi model turun drastis hingga 0% untuk sampel Asia dan sekitar 37% untuk sampel kulit hitam, bahkan saat jumlah fitur tetap tinggi pada 90. Nilai F1-score dan AUC yang rendah pada kelompok ini memperkuat adanya kegagalan model dalam melakukan generalisasi ras.

Hasil ini menegaskan adanya potensi disparitas rasial dalam performa model MLP, yang berasal dari rancangan awal penelitian yang berfokus pada ras mayoritas. Seleksi fitur dilakukan berdasarkan pola metilasi pada sampel kulit putih, dan proses pelatihan model juga didominasi oleh data dari kelompok tersebut. Akibatnya, model belajar dari pola biologis yang lebih representatif untuk ras mayoritas, namun tidak mampu mengenali biomarker metilasi dari kelompok ras minoritas. Hal ini menandakan pentingnya pendekatan berbasis ras dalam pengembangan model prediktif berbasis epigenomik, agar hasil yang diperoleh adil dan dapat diandalkan untuk populasi yang beragam. Perbedaan biologis ini dapat mencerminkan faktor genetik, epigenetik, lingkungan, dan sosial yang saling berhubungan yang memengaruhi pola metilasi setiap ras. Dampak dari

ketidakseimbangan ini tidak hanya terbatas pada kinerja model MLP, tetapi juga mencapai area yang lebih luas dalam dunia praktik medis dan kesehatan masyarakat. Jika model prediktif dibangun berdasarkan data yang tidak inklusif secara rasial, risiko bias dalam diagnosis, prognosis dan pengambilan keputusan klinis menjadi lebih tinggi, terutama untuk ras minoritas. Oleh karena itu, hasil penelitian ini menekankan pentingnya aspek representasi populasi pada data metilasi DNA, serta perlunya metode yang mempertimbangkan keragaman biologis dan demografis dalam setiap tahapan pembangunan model kecerdasan buatan untuk pengaplikasian di biomedis.

5.2. Saran

Penelitian di masa depan disarankan untuk memprioritaskan peningkatan keragaman dan representasi ras dalam dataset awal untuk memastikan rancangan model yang lebih adil dan dapat digeneralisasi. Teknik-teknik seperti pemilihan fitur bertingkat, adaptasi domain, atau pelatihan yang sadar akan ras dapat membantu mengungkap biomarker spesifik subkelompok dan meningkatkan keadilan model. Selain itu, integrasi atribut fenotipe eksternal (misalnya, PSA, skor Gleason) dan data multi-omik dapat memberikan pandangan yang lebih terpadu tentang perkembangan kanker prostat di berbagai latar belakang ras, yang berpotensi mengurangi bias dan meningkatkan relevansi klinis.

