

ABSTRAK

Kanker payudara merupakan salah satu jenis kanker yang paling umum menyerang jaringan kelenjar susu dan menjadi penyebab utama kematian pada wanita di seluruh dunia. Penentuan stadium kanker sangat penting dalam memilih strategi pengobatan yang tepat. Penelitian ini bertujuan untuk mengklasifikasikan stadium kanker payudara menggunakan algoritma Random Forest serta mengidentifikasi fitur klinis yang paling berpengaruh dalam proses klasifikasi. Dataset yang digunakan berasal dari *SEER Breast Cancer Data – Labeled* yang dipublikasikan oleh National Cancer Institute (NCI) dan tersedia di platform Kaggle. Dataset ini terdiri dari 4,024 data pasien wanita dengan informasi klinis lengkap, seperti ukuran kanker, status kelenjar getah bening, dan tingkat diferensiasi sel kanker. Setelah dilakukan tahap pra-pemrosesan dan pelatihan model, algoritma *Random Forest* berhasil mencapai performa klasifikasi yang tinggi dengan tingkat akurasi sebesar 77,01% pada data uji sebanyak 805 entri. Model mampu mengklasifikasikan keempat stadium kanker dengan performa sangat baik pada stadium 0 dan 1, serta hasil yang memadai pada stadium 2 dan 3. Penelitian ini menunjukkan bahwa algoritma *Random Forest* efektif digunakan sebagai alat bantu diagnosis berbasis data untuk klasifikasi otomatis stadium kanker payudara.

Kata Kunci: Kanker Payudara, Random Forest, Stadium Kanker, Pembelajaran Mesin, Pentingnya Fitur, Klasifikasi, Dataset SEER.