

1. Pendahuluan

1.1 Latar Belakang

Kanker adalah salah satu penyakit yang dapat menyebabkan kematian manusia di berbagai negara. Kanker menjadi penyebab utama kematian manusia dalam beberapa tahun terakhir. Menurut WHO pada tahun 2018, kanker menyebabkan 9.6 juta kematian manusia di seluruh dunia (2018) [1]. Pada beberapa tahun terakhir mendeteksi kanker dilakukan dengan klasifikasi data *microarray*. Teknologi *microarray* digunakan untuk mendeteksi ekspresi ribuan gen pada saat yang bersamaan untuk menganalisis dan mendiagnosis kanker [15]. Teknologi *microarray* juga dapat membantu untuk memantau tingkat ekspresi gen dalam suatu sampel tertentu [3].

Akan tetapi data *microarray* memiliki dimensi yang tinggi karena jumlah fitur yang sangat banyak dan memiliki ketersebaran data yang rendah artinya memiliki jumlah sample data yang sedikit yang menyebabkan performansi yang rendah [4]. Untuk mengatasi permasalahan tersebut dapat dilakukan dengan reduksi dimensi data *microarray*. Metode yang digunakan untuk reduksi dimensi yaitu *Random Projection (RP)*, *Support Vector Machine (SVM)*, dan *Random Forest* sebagai metode klasifikasi. Klasifikasi menggunakan *Support Vector Machine (SVM)* karena memiliki banyak fitur matematika yang membuatnya menarik untuk menganalisis ekspresi gen, termasuk fleksibilitas dalam memilih fungsi kesamaan, ketersebaran solusi saat menangani sample data dan jumlah fitur yang besar dan mampu mengidentifikasi data *outlier* [5], klasifikasi menggunakan *Random Forest (RF)* mampu digunakan untuk mengklasifikasi fitur yang tidak lengkap, cocok untuk mengklasifikasi data yang memiliki jumlah fitur dan sample yang sangat banyak [17]. Namun masalah klasifikasi menggunakan *Random Forest (RF)* adalah membutuhkan waktu yang lebih lama untuk melatih data. Oleh karena itu, diperlukan pendekatan untuk mengurangi dimensi agar proses melatih data tidak membutuhkan waktu yang lama. Dengan menggunakan *Random Projection (RP)* sebagai metode dimensi yang memiliki keunggulan yaitu dapat memproyeksikan ke subruang berdimensi lebih rendah menggunakan matriks acak yang sampelnya memiliki panjang satuan. *Random Projection (RP)* menjadi metode komputasi yang efisien namun cukup akurat untuk mengurangi dimensi data yang tinggi [7]. *Random projection* dengan kesederhanaan kompleksitasnya tidak memberikan pengaruh yang signifikan terhadap distorsi dalam data [7].

Dalam penelitian ini, penulis mengambil referensi dari penelitian sebelumnya yang berkaitan dengan deteksi kanker yang menggunakan klasifikasi data *microarray*. Penelitian yang dilakukan pada tahun 2018 [15], metode reduksi dimensi menggunakan *Principal Component Analysis (PCA)* yang meliputi perhitungan proporsi varian untuk pemilihan vektor eigen. Sedangkan metode klasifikasi menggunakan *Support Vector Machine (SVM)* dan *Levenberg-Marquardt Algoritma Backpropagation (LMBP)*. Hasil akurasi rata-rata yang didapat dengan *LMBP* sebesar 96,07%, sedangkan dengan *SVM* sebesar 94,98%. Hasilnya metode klasifikasi menggunakan *LMBP* lebih stabil dari pada *SVM*.

Penelitian selanjutnya dilakukan pada tahun 2020 [17] dengan menggunakan algoritma *Random Forest* untuk proses klasifikasi dan *Redundancy Maximum Relevance (MRMR)* dan *Least Absolute Shrinkage and Selection Operator (LASSO)* untuk mengurangi dimensi data

microarray. Penulis mendapatkan kesimpulan bahwa LASSO lebih baik daripada MRMR karena dapat menghasilkan evaluasi 100% pada lung cancer dan ovarian cancer, 92% pada colon cancer, 93% pada prostate tumor dan 83% pada central nervous system.

Berdasarkan keunggulan algoritma Support Vector Machine (SVM) dan Random Forest (RF) yang telah dijelaskan pada penelitian yang sudah disebutkan sebelumnya, pada penelitian kali ini, penulis membangun sebuah sistem klasifikasi yang dapat mendeteksi kanker berdasarkan data *microarray* dengan menggunakan metode reduksi dimensi *Random Projection (RP)* dan klasifikasi menggunakan Support Vector Machine (SVM) dan Random Forest (RF) untuk membandingkan metode klasifikasi mana yang lebih baik.