

1. Pendahuluan

Kanker adalah penyakit mematikan yang terkenal di dunia. Berdasarkan *World Health Organization* (WHO), kanker adalah penyebab utama kematian kedua di dunia dan bertanggung jawab atas sekitar 9,6 juta kematian pada tahun 2018 [1]. Pendeteksian kanker dengan cepat dan akurat sangat dibutuhkan agar penyakit kanker dapat langsung ditangani dan diatasi dengan tepat.

Salah satu teknik terkenal untuk pendeteksian kanker adalah teknik DNA *microarray*. DNA *microarray* adalah kumpulan dari ribuan DNA mikroskopis dalam bentuk fragmen DNA yang diletakkan pada sebuah chip. Teknik DNA *microarray* memberikan kesempatan bagi para peneliti untuk menganalisis ribuan profil *gene expression* dalam waktu yang bersamaan. Dengan menganalisis profil *gene expression* melalui proses klasifikasi, dapat diketahui apakah seseorang menderita kanker atau tidak. Kendala yang dialami dalam proses pemrosesan profil *gene expression* ini adalah data DNA *microarray* mengandung banyak fitur yang tidak signifikan dan tidak relevan yang mempengaruhi proses dan hasil klasifikasi. Sedangkan dalam mendapatkan model yang akurat, diperlukan data sampel dan fitur yang informatif [18]. Oleh karena itu, fitur yang relevan harus terlebih dahulu ditentukan [2]. Salah satu cara untuk mengatasi hal tersebut adalah dengan melakukan *feature selection* terhadap data. *Feature selection* dilakukan untuk memilih fitur-fitur penting dan relevan terhadap data dan membuang fitur-fitur yang tidak berpengaruh terhadap proses klasifikasi kanker.

Penelitian ini akan menganalisis pengaruh *feature selection* pada data DNA *microarray* untuk mengetahui apakah *feature selection* dapat meningkatkan hasil akurasi dalam pengklasifikasian *gene expression*-nya. Metode *feature selection* yang akan digunakan adalah *Support Vector Machine-Recursive Feature Elimination* (SVM-RFE) dan *Chi-Square*. Sebuah metode klasifikasi juga dipilih untuk mengklasifikasikan *gene expression* DNA *microarray* yaitu *Naïve Bayes*. Pemilihan metode-metode tersebut didasarkan pada penelitian-penelitian sebelumnya yang menyebutkan bahwa metode-metode tersebut memiliki performansi yang baik dalam pengklasifikasian *gene expression* DNA *microarray*. Data yang digunakan adalah data *colon cancer*, *breast cancer*, *lung cancer*, dan *ovarian cancer* yang bersumber dari Kent-Ridge Biomedical [3]. Pengujian performa pengklasifikasian kanker dilakukan dengan cara membandingkan akurasi klasifikasi data dengan dan tanpa *feature selection* pada metode klasifikasi. Hasil akurasi antara penggunaan kedua metode *feature selection* juga akan dibandingkan, untuk mencari metode *feature selection* mana yang lebih baik jika digabungkan dengan metode klasifikasi *Naïve Bayes*. Untuk mendapatkan gambaran keseluruhan tentang perbandingan performa, penelitian ini juga mempertimbangkan *precision*, *recall*, dan *F1-score*.