

1. Pendahuluan

Kanker merupakan salah satu penyakit yang memiliki peningkatan jumlah kasus yang tinggi di setiap tahun dan juga menjadi salah satu penyakit yang mematikan di dunia. Berdasarkan data dari *International Agency for Research on Cancer* (IARC) pada tahun 2018 terdapat 18,1 juta kasus baru untuk penyakit kanker di seluruh dunia dan diperkirakan pada tahun 2040, kasus kanker di dunia meningkat sebanyak 11,4 juta kasus [1]. Selain itu, pada tahun 2018 terdapat 9,6 juta kasus kematian yang disebabkan oleh penyakit kanker. Untuk tingkat kematian yang disebabkan oleh penyakit kanker pada tahun 2018 sebesar 57.3%[2]. Maka dari itu, diperlukannya diagnosis dini untuk menghindari peningkatan angka kematian yang disebabkan oleh penyakit kanker. Namun pendeteksi secara konvensional menggunakan media gambar seperti *CT scan* dan *PET scan* masih membutuhkan waktu yang cukup lama. Banyak faktor yang membuat pendeteksian menggunakan media gambar kurang akurat seperti kesalahan pada pengambilan gambar, kelainan pada organ yang tidak terdeteksi atau diagnosis yang salah[3]. Seiring perkembangan pembelajaran mesin, terdapat suatu teknologi *DNA microarray* yang dapat digunakan untuk mendeteksi terjangkitnya penyakit kanker. *DNA microarray* dapat mendeteksi penyakit kanker dengan waktu yang singkat menggunakan data gen dan juga dapat mengurangi faktor yang memungkinkan pendeteksian kurang akurat[4].

DNA microarray adalah suatu teknologi yang digunakan dalam mengumpulkan dan memproses ribuan ekspresi gen dalam waktu yang sama [4][5]. *Microarray* merupakan metode yang efisien dalam menentukan pola ekspresi ribuan gen. data ekspresi gen yang akan diolah berupa sebuah matriks dimana baris merepresentasikan gen dan kolom merepresentasikan sampel[3]. Langkah dalam mendapatkan data *microarray* dimulai dari sampel RNA sel kanker dan sel normal dipisahkan. Kedua RNA sampel digabungkan dengan cDNA dengan pewarna hijau. Perubahan warna terjadi pada sel kanker. Kemudian hasil tersebut di amati untuk mendapatkan gambar dari kedua sampel yang selanjutnya akan dianalisis untuk digabungkan ke dalam data *microarray*[6]. Pemrosesan ribuan ekspresi gen membuat *DNA microarray* memiliki kendala yaitu jumlah atribut yang dimiliki oleh data *microarray* sangat besar dan setiap atribut memiliki hubungan yang rumit. Kendala tersebut menyebabkan hasil pendeteksian kurang akurat dan waktu komputasi yang tinggi saat melakukan pendeteksian. Reduksi dimensi dapat menjadi solusi untuk kendala tersebut. Penggunaan reduksi dimensi dapat membantu klasifikasi dalam meningkatkan keakuratan hasil klasifikasi. Hasil yang baik bisa didapatkan jika penggabungan reduksi dimensi dan metode klasifikasi cocok.

Penelitian ini membangun sebuah sistem klasifikasi *microarray* untuk mendeteksi penyakit kanker. Penelitian ini menggunakan lima data *microarray* yaitu *Colon Tumor*, *Breast Cancer*, *Lung Cancer*, *Prostate Tumor* dan *Ovarian Cancer*. Kelima data tersebut di dapat dari <http://leo.ugr.es/elviraDBCRepository/>. Metode reduksi dimensi yang digunakan pada penelitian ini yaitu *Discrete Wavelet Transform* (DWT). Metode klasifikasi yang digunakan yaitu *Classification and Regression Tree* dan *Random Forest*. Hasil dari kedua metode klasifikasi tersebut dibandingkan untuk menentukan metode klasifikasi mana yang menghasilkan performansi paling baik jika dipasangkan dengan reduksi dimensi DWT untuk mengklasifikasikan data *microarray*.