

1 PENDAHULUAN

1.1 Latar Belakang

Kanker adalah sebuah penyakit yang disebabkan oleh pembelahan secara berlebihan dan tak terkendali dari sel-sel dalam tubuh. Sebanyak 8,8 juta orang meninggal akibat kanker di tahun 2015.[1] Akan tetapi kematian akibat kanker bisa dihindari sebanyak 30-50%, dengan cara mengidentifikasi jenis kanker yang menyerang pasien sedari dini. Namun, merupakan tantangan yang sulit bagi para peneliti untuk mendiagnosa kanker berdasarkan struktur morfologisnya karena perbedaan morfologis yang sangat tipis untuk jenis kanker yang berbeda.

Kesulitan di atas mendorong penggunaan algoritma penambangan data, secara spesifik penggunaan ekspresi gen untuk menentukan jenis sel kanker. Tingkat ekspresi gen dapat menunjukkan keaktifan sebuah gen di dalam sel tubuh berdasarkan jumlah mRNA di dalam sel tubuh tersebut. Tingkat ekspresi gen dalam tubuh manusia dapat diukur melalui eksperimen *microarray* DNA. Teknologi *microarray* DNA memiliki kemampuan untuk menentukan level dari beribu-ribu gen secara bersamaan di dalam sebuah eksperimen [2].

Begitu banyak pendekatan yang sudah diimplementasikan dalam melakukan klasifikasi ekspresi gen *microarray*, salah satu yang sering digunakan selama 10 tahun terakhir adalah *Support Vector Machine* (SVM). SVM dipercaya merupakan salah satu algoritma *supervised learning* yang terbaik dalam hal performansi. Furey et. al. membuktikan bahwa SVM memberikan akurasi performansi sebesar 100% dalam mengklasifikasi data kanker usus besar [8].

Karakteristik dari data *microarray* adalah jumlah gen (p) yang sangat besar, dapat mencapai puluhan ribuan (dimensi yang besar), sedangkan jumlah sampelnya (n) yang sedikit. Maka dari itu diperlukan seleksi fitur atau ekstraksi fitur pada data *microarray* untuk menghindari *overfitting* pada *classifier*. Salah satu seleksi fitur yang mampu meningkatkan kinerja *classifier* adalah *Minimum Redundancy Maximum Relevance* (MRMR), yang sudah dibuktikan oleh Ding et al. [14] dapat memberikan akurasi performansi sebesar 100% untuk data Leukemia. Sedangkan ekstraksi fitur yang sering digunakan adalah *Principal Component Analysis* (PCA), yang dapat memberikan akurasi performansi rata-rata sebesar 80% [9][17].

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka perumusan masalah dalam pembuatan tugas akhir ini adalah sebagai berikut:

1. Bagaimana membangun sistem untuk mengklasifikasi kanker berdasarkan data *microarray*?
2. Bagaimana menganalisis peranan reduksi dimensi data *microarray* yang besar pada proses klasifikasi?
3. Bagaimana menganalisis performansi dari hasil reduksi dimensi dan *classifier* yang dibangun?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam menyusun tugas akhir ini adalah data yang digunakan berupa ekspresi gen *microarray* dari jenis-jenis kanker yang berbeda, yaitu: kanker usus besar, kanker paru-paru, kanker ovarium, dan leukemia. Data didapatkan dari repositori Kent Ridge Biomedical Data Repository [15].

1.4 Tujuan

Berdasarkan perumusan masalah di atas, maka tujuan pembuatan tugas akhir ini adalah sebagai berikut:

1. Membangun model *classifier* menggunakan metode *Support Vector Machine* (SVM) untuk seleksi kanker berdasarkan klasifikasi data *microarray*.
2. Mengimplementasi dan menganalisis metode seleksi fitur *Minimum Redundancy Maximum Relevance* (MRMR) dan metode ekstraksi fitur *Principal Component Analysis* (PCA) untuk mereduksi dimensi data *microarray* pada sistem klasifikasi yang telah dibangun.
3. Mengukur performansi dan menganalisis kinerja sistem klasifikasi yang telah dibangun menggunakan F1-score dan waktu komputasi.

1.5 Sistematika Penulisan

Sistematika dalam penulisan Tugas Akhir ini akan dijelaskan dalam setiap bab dengan pembagian sebagai berikut:

1. BAB 1 PENDAHULUAN

Bab ini berisikan latar belakang pengambilan judul “Analisis Pengaruh Metode Reduksi Dimensi *Minimum Redundancy Maximum Relevance* pada Klasifikasi Kanker Berdasarkan Data *Microarray* Menggunakan *Classifier Support Vector Machine*”, perumusan masalah yang akan dikaji, tujuan pembuatan Tugas Akhir, batasan masalah dan sistematika penulisan tugas akhir yang dijelaskan secara garis besar untuk masing-masing bab.

2. BAB 2 KAJIAN PUSTAKA

Bagian ini menjelaskan tinjauan pustaka yang berkaitan dengan metode yang akan digunakan serta teori-teori dan referensi penelitian-penelitian yang mendukung pembuatan tugas akhir ini. Teori tersebut antara lain penjelasan tentang riset terkait klasifikasi data *microarray*, teknologi *microarray*, metode reduksi dimensi *Minimum Redundancy Maximum Relevance* (MRMR), metode reduksi dimensi *Principal Component Analysis* (PCA), dan *classifier Support Vector Machine* (SVM).

3. BAB 3 PERANCANGAN SISTEM

Rancangan sistem yang telah dibangun dijelaskan pada bab ini. Desain sistem ini memuat arsitektur sistem, *flowchart*, fungsionalitas sistem dan desain skenario untuk menguji performansi sistem.

4. BAB 4 PENGUJIAN DAN ANALISIS

Sistem yang telah dibangun selanjutnya diuji dan dianalisis hasilnya. Hasil pengujian dan analisis akan disajikan secara komprehensif pada bab ini.

5. BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari pengujian dan analisis yang diperoleh dan saran yang diperlukan untuk penelitian selanjutnya.