

1. Pendahuluan

Jumlah angka kematian penderita kanker mencapai 9,6 juta dari 18,1 juta penderita kanker di seluruh dunia pada tahun 2018 berdasarkan data pada The International Agency for Research on Cancer (IARC). Berdasarkan jumlah tersebut, kanker menjadi penyebab kematian kedua terbanyak di dunia [2]. Perkembangan teknologi yang bertujuan untuk menekan angka kematian yang disebabkan oleh kanker juga semakin maju. Teknologi DNA *microarray* merupakan salah satu contoh perkembangan teknologi yang berupaya untuk mendeteksi kanker dengan memanfaatkan analisis pada ekspresi gen.

Beberapa tahun terakhir, DNA *microarray* merupakan salah satu teknologi yang digunakan pada penelitian bidang bioinformatika. *Microarray* adalah data yang memiliki dimensi yang tinggi untuk setiap sampelnya. Data *microarray* digunakan sebagai representasi ekspresi gen dari untaian DNA manusia. Teknologi *microarray* sudah digunakan untuk diagnosa medis dan analisis gen, terutama untuk menganalisis perubahan pola ekspresi gen dalam keadaan tertentu [3]. Dengan beberapa pendekatan, teknologi ini akan sangat membantu dalam menentukan apakah seseorang menderita kanker atau tidak dari segi akurasi dan efisiensi waktu.

Gen merupakan bagian dari DNA yang berisi informasi yang dibutuhkan untuk membentuk semua jenis protein dalam tubuh [14]. *Microarray* menyimpan seluruh informasi gen berdasarkan sekuens dalam asam nukleat seorang manusia lalu menyimpannya menjadi satu sampel. Setiap sampel data DNA *microarray* akan menyimpan ribuan bahkan belasan ribu dimensi. DNA *microarray* sangat berguna untuk membantu mengidentifikasi perubahan atau anomali gen yang ada pada tubuh manusia.

Jumlah gen yang sangat banyak menjadikan dimensi dari data *microarray* juga akan sangat tinggi dan sangat kompleks. Sehingga akan ada ketidakseimbangan antara jumlah dimensi data dengan jumlah sampel data. Untuk menganalisis ekspresi gen diperlukan adanya identifikasi dari informasi setiap gen [14]. Hasil identifikasi dari informasi setiap gen tersebut akan digunakan sebagai parameter untuk memprediksi kelas dari sampel tersebut. Maka dari itu, dibutuhkan solusi paling optimal dalam menentukan skema klasifikasi dari ekspresi gen *microarray*, sehingga dapat menangani kompleksitas data dan mendapatkan hasil dengan akurasi tinggi dalam waktu yang relatif singkat [3].

Untuk menghasilkan nilai akurasi yang tinggi dan waktu yang singkat, dimensi pada data *microarray* haruslah direduksi. Teknik reduksi dimensi ini digunakan untuk menentukan mana saja fitur yang berpengaruh kepada hasil klasifikasi. Ada beberapa metode yang sudah digunakan untuk reduksi dimensi dalam penelitian deteksi kanker sebelumnya. Salah satu metode yang dapat digunakan untuk reduksi dimensi adalah Autoencoder. Autoencoder adalah metode yang masih tergolong baru dibandingkan dengan metode lain yang lebih populer, seperti Principal Component Analysis (PCA) ataupun Linear Discriminant Analysis (LDA). Autoencoder merupakan salah satu metode yang menggunakan jaringan syaraf tiruan untuk mereduksi dimensi dengan meminimalkan nilai *loss* rekonstruksi [12]. Jaringan tersebut akan melatih data masukan dengan proses memecah data, mengekstraksi fitur penting pada data, lalu merekonstruksi data sedemikian rupa menjadi data dengan nilai *loss* yang paling minimal.