

# Deteksi Kanker Berdasarkan Klasifikasi Data Microarray Menggunakan Functional Link Neural Network dengan Seleksi Fitur Genetic Algorithm

Putri Tsatsabila Ramadhani <sup>#1</sup>, Untari Novia Wisesty <sup>#2</sup>, Annisa Aditsania <sup>#3</sup>

*# Program Studi Ilmu Komputasi, Universitas Telkom  
Jalan Telekomunikasi 1, Bandung 40257, Indonesia*

<sup>1</sup> tsatsabila@gmail.com

<sup>2</sup> untarinw@telkomuniversity.ac.id

<sup>3</sup> aaditsania@telkomuniversity.ac.id

## Abstract

According to Data and Information Center, Ministry of Health Republic of Indonesia in 2015, cancer is one of the major causes of death in the world. In fact, in 2012 there were 8.2 million people death caused by cancer. The number of death is estimated to increase every year even, in the two next decades will increase 15 million to 22 million. The cancer can be known and detected early have possibility to get better handling, and microarray technology has a role in it. In recent years, microarray technology has a great impact for determining genes informative causes cancer. Microarray has capability to determine the expression of thousands genes simultaneously monitoring biological processes. By analyzing microarray data, the thousand genes expression that represent a human, would be classified as a cancer or not. In this final project, the author implements Functional Link Neural Network with Legendre Polynomial basis functions for the accurate classification of data and uses Genetic Algorithm as a feature selection to reduce the high dimensional does exist on microarray data. Thus after all the processes have been conducted, then we retrieved the highest performance against the classification of microarray data of Colon Tumors of 92.3% and classification for Leukemia of 87.5%. The differences of performance is caused by the difference of the acquired characteristics of data respectively.

**Keywords:** cancer detection, microarray, Functional Link Neural Network, Genetic Algorithm

## Abstrak

Berdasarkan Pusat Data dan Informasi Kementerian Kesehatan RI Tahun 2015, penyakit kanker merupakan salah satu penyebab kematian utama di seluruh dunia, dan di tahun 2012, sekitar 8,2 juta kematian disebabkan oleh kanker. Diperkirakan penderita kanker terus meningkat setiap tahunnya bahkan dalam dua dekade berikutnya akan meningkat dari 14 juta pada 2012 menjadi 22 juta. Kanker yang dapat diketahui dan dideteksi lebih awal memiliki probabilitas untuk mendapatkan penanganan yang lebih baik, dan teknologi microarray memiliki peran dalam hal itu. Di beberapa tahun terakhir, pemanfaatan teknologi microarray memiliki pengaruh besar dalam menentukan gen informatif yang menyebabkan kanker. Microarray mampu menentukan ekspresi ribuan gen dan secara simultan memantau proses biologis yang sedang berlangsung. Dengan melakukan analisa terhadap data micorarray, selanjutnya ekspresi dari ribuan gen yang merepresentasikan suatu jaringan pada manusia, akan diklasifikasikan sebagai jaringan kanker atau bukan. Dalam penelitian ini, penulis mengimplementasikan Functional Link Neural Network dengan fungsi basis Legendre Polynomial untuk klasifikasi data yang akurat dan menggunakan Genetic Algorithm sebagai seleksi fitur untuk mereduksi data berdimensi tinggi yang sering ditemukan pada data microarray. Dengan serangkaian proses yang telah dilakukan, maka diperoleh kinerja tertinggi terhadap klasifikasi data microarray

Colon Tumor sebesar 92.3% dan Leukemia sebesar 87.5%. Perbedaan kinerja yang diperoleh disebabkan oleh perbedaan karakteristik masing-masing data.

**Kata Kunci:** deteksi kanker, microarray, Functional Link Neural Network, Genetic Algorithm

## I. PENDAHULUAN

SETIAP tahunnya jumlah penderita penyakit kanker terus meningkat. Umumnya penderita baru mengetahui saat kanker tersebut sudah memasuki stadium lanjut. Terlambatnya penanganan terhadap penderita kanker bisa berakibat fatal bahkan dapat menyebabkan kematian. Oleh sebab itu, tidak heran penelitian mengenai kanker menjadi topik utama dalam bidang medis dan bioinformatika dan terus berkembang hingga saat ini, termasuk teknologi DNA microarray [7].

Teknologi DNA microarray digunakan untuk menentukan tingkat ekspresi ribuan gen yang dilakukan dalam sekali percobaan, dan secara simultan memantau proses biologi yang sedang berlangsung [6]. Teknologi ini mampu menghasilkan data yang dibutuhkan untuk proses prediksi dan klasifikasi gen yang diambil dari beberapa jaringan tertentu pada manusia untuk digolongkan ke dalam kanker atau bukan. Berdasarkan [3], yang menjadi kendala utama dalam suatu data microarray adalah besarnya dimensi, yang menyebabkan beban komputasi menjadi tidak stabil. Untuk menangani masalah itu, Genetics Algorithm (GA) digunakan sebagai seleksi fitur untuk mereduksi dimensinya. GA akan membangkitkan populasi awal secara random, kemudian fungsi fitness digunakan untuk mengevaluasi populasi tersebut, dan selanjutnya dilakukan operasi genetika seperti seleksi orang tua, *crossover*, dan mutasi hingga kondisi kriteria dicapai. Kemudian data yang dimensinya sudah direduksi akan diklasifikasi menggunakan metode Functional Link Neural Network (FLNN).

FLNN merupakan salah satu model Artificial Neural Network (ANN) yang mampu menyelesaikan tugas klasifikasi dengan keberhasilan yang cukup signifikan karena menggunakan arsitektur datar atau layar tunggal, yang mampu mengoptimasi proses, meningkatkan kemampuan diskriminasi terhadap masukan yang irrelevan, insignifikan, dan *noisy*. Sebelumnya, dalam penelitian yang dilakukan oleh Kumar et.al [3] terhadap data set Leukemia, Ovarian Cancer, dan Breast Cancer masing-masing menghasilkan akurasi sebesar 96.15%, 97.78%, 86.54% dengan menggunakan fungsi basis Legendre Polynomial. Berbeda dengan [3], dalam [8] akurasi klasifikasi yang diperoleh menggunakan metode ANN dengan data Colon Tumor dan Lymphoma masing-masing sebesar 50.92% dan 95.80%.

Dalam penelitian ini, FLNN dengan fungsi basis Legendre Polynomial diimplementasikan sebagai metode klasifikasi data micorarray, dan GA sebagai metode seleksi fitur. Selanjutnya kinerja hasil klasifikasi tersebut diukur dengan teknik validasi F1-Measure. Oleh karena itu, dengan dilakukan deteksi kanker diharapkan dapat membantu kebutuhan diagnosis, prognosis dan tindakan penanganan yang cepat dan tepat untuk penderita kanker.

## II. STUDI LITERATUR

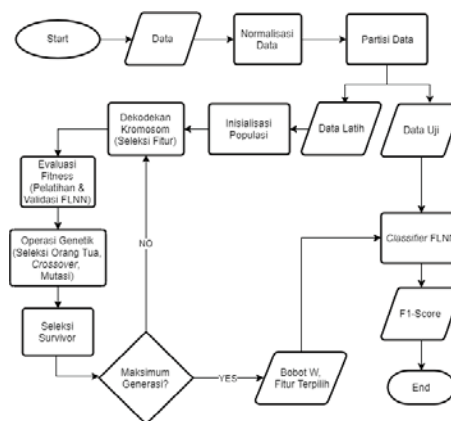
Dari berbagai penelitian yang telah dilakukan oleh beberapa para ahli dalam mengklasifikasikan data microarray, metode seleksi fitur yang digunakan untuk menangani masalah dimensionalitas dan metode klasifikasi yang digunakan tersaji dalam Tabel I.

Tabel I  
STUDI LITERATUR

Penulis	Seleksi Fitur	Classifier	Akurasi (%)
Kumar et al.(2015)	T-Statistik	Functional Link Neural Network	96.15, 97.78, 86.54
Diaz-Urriarte et al.(2006)	Random Forest	Random Forest	95
C.D.A Vanita et al.(2015)	Mutual Information	KNN, ANN, SVM Linear, SVM RBF, SVM Quadratic, SVM Polynomial	44.51, 50.92, 67.74, 60.51, 40.23, 46.83
Khare et al.(2016)	Genetic Algorithm	Bayesnet, SMO, Simple Logistic, One-R, Zero-R	75.7, 99.1, 100, 99.8, 99.1

### III. METODOLOGI PENELITIAN

Gambaran umum dari sistem klasifikasi yang dibangun dalam penelitian ini tersaji dalam diagram alur Gambar 1 di bawah ini.



Gambar 1. Diagram Alur Sistem

Data yang telah melewati tahap *preprocessing* dalam hal ini adalah normalisasi, selanjutnya akan dipartisi untuk menjadi data latih dan data uji menggunakan CVpartition dengan teknik *holdout* berdasarkan proporsi tertentu yang akan disesuaikan dengan kebutuhan. Lalu bilangan biner dibangkitkan random sebanyak fitur. Jika sebuah bit sama dengan 0 artinya fitur tersebut tidak dipilih, namun jika bit sama dengan 1 maka fitur tersebut akan dipilih untuk selanjutnya diklasifikasikan menggunakan algoritma FLNN fungsi basis Legendre Polynomial orde tiga. Kemudian, hasil akurasi dari pelatihan data akan digunakan sebagai fungsi fitness dalam proses evaluasi individu GA. Selama proses pelatihan dihasilkan model klasifikasi berupa fitur yang terpilih dan bobot jaringan yang telah diperbaharui yang akan digunakan dalam proses pengujian terhadap data uji. Selanjutnya kinerja sistem dalam mengklasifikasikan diukur dengan metode F1-Measure.

#### A. Data

Klasifikasi data dilakukan terhadap dua jenis data microarray yang berbeda yaitu data Colon Tumor dan Leukemia yang merupakan data sekunder yang diperoleh dari Repository Data Set Kent Ridge Biomedical dan sudah sering digunakan di berbagai jurnal ilmiah salah satunya dalam [3]. Setiap nilai pada data tersebut merupakan nilai representasi dari ekspresi gen, *protein profiling* dan *genomic sequence* yang diperoleh dari *chip* microarray.

#### B. Normalisasi Data

Tahap pertama yang harus dilakukan terhadap data adalah *preprocessing*. Karena tidak ditemukan *missing value* dalam data, maka *preprocessing* yang dilakukan hanya meliputi normalisasi data. Kemudian

Tabel II  
 RINCIAN DATA MICROARRAY

Data	Jumlah Fitur	Jumlah Kelas	Jumlah Basis		Ukuran
Colon Tumor	2000	2	22 positive	40 negative	0.9 MB
Leukemia	7129	2	25 AML	47 ALL	1.8 MB

untuk normalisasi data digunakan teknik Min-Max Normalization dengan persamaan berikut.

$$Normalisasi(x) = \frac{x - Min(X_i)}{Max(X_i) - Min(X_i)} \quad (1)$$

$Min(X_i)$  dan  $Max(X_i)$  merupakan nilai minimum dan nilai maksimum dari kolom fitur ( $X_i$ ) pada data. Jika ditemukan  $Min(X_i)$  sama dengan  $Max(X_i)$ , maka normalisasi terhadap  $x$  adalah 0.5 [3]. Setelah dilakukan normalisasi data, maka diperoleh nilai masukan data yang berada dalam jangkauan nilai [0,1]. Selanjutnya menetapkan sejumlah proporsi data latih dan data uji yang akan diproses pada sistem klasifikasi.

### C. Seleksi Fitur GA

Setelah melalui tahap *preprocessing* seperti pada penjelasan sebelumnya, maka dilakukan tahap seleksi fitur menggunakan GA. Seleksi fitur dilakukan untuk mereduksi data berdimensi tinggi agar menghindari masalah *curse of dimensionality* dan resiko *overfitting*. Menurut Erick Cantu-Paz dalam [2], GA biasanya memberikan hasil yang baik dan dapat menghasilkan akurasi yang lebih tinggi pada data set tertentu. Sebelum masuk ke proses GA, proporsi data latih dan data uji harus ditetapkan. Lalu merujuk pada [4] dan [1] maka dilakukan tahap-tahap GA yang telah disesuaikan dengan kebutuhan penelitian ini sebagai berikut.

1) *Representasi Individu*: Setiap individu direpresentasikan ke dalam deretan bilangan biner 0 atau 1. Kemudian dilakukan inialisasi populasi berdasarkan bilangan biner tersebut yang dibangkitkan secara acak sebanyak jumlah fitur dan ukuran populasi. Setiap individu pada populasi merepresentasikan sebuah solusi kandidat terhadap masalah seleksi fitur. Jika sebuah bit sama dengan 0 artinya fitur tersebut tidak dipilih, namun apabila bit berupa bilangan biner sama dengan 1, maka fitur tersebut dipilih.

2) *Evaluasi Fitness (Akurasi FLNN)*: Algoritma FLNN dengan fungsi basis Legendre Polynomial digunakan untuk menghasilkan akurasi yang dibutuhkan sebagai fungsi fitness dengan persamaan 2. Kelas klasifikasi yang dihasilkan akan dihitung dengan metode Confusion Matrix dengan persamaan 3.

$$fitness(x) = Accuracy(x) \quad (2)$$

$$Accuracy(x) = \frac{TP + TN}{TP + FN + FP + TN} * 100\% \quad (3)$$

dimana TP (*True Positive*) adalah proporsi positif dalam data set yang diklasifikasikan positif, FP (*False Positive*) adalah proporsi negatif dalam data set yang diklasifikasikan positif, TN (*True Negative*) adalah proporsi negatif dalam data set yang diklasifikasikan negatif, dan FN (*False Negative*) proporsi positif dalam data set yang diklasifikasikan negatif. Setelah diperoleh nilai fitness pada tiap-tiap individu, maka individu yang memiliki nilai fitness terbaik akan disalin sebagai Elitisme. Agar nilai fitness tersebut yang merupakan kandidat solusi terbaik tidak hilang selama operasi genetik yang berlangsung nantinya.

### 3) Operasi Genetik:

- Seleksi Orang Tua

Setiap individu akan diseleksi untuk menjadi orangtua dengan metode *Roulette Wheel*, dengan menempatkan setiap kromosom individu pada *Roulette Wheel* sesuai dengan proporsi nilai fitness yang dimiliki masing-masing individu. Tentunya semakin besar nilai fitness suatu kromosom maka proporsi yang dimilikinya dalam *Roulette Wheel* akan semakin besar pula sehingga peluang individu tersebut terpilih untuk menjadi orang tua juga semakin besar.

- Crossover  
Lalu, crossover dilakukan terhadap kromosom orang tua yang telah terpilih melalui proses sebelumnya sehingga menghasilkan kromosom *offspring* (anak). Setiap kromosom *offspring* yang telah dibentuk sudah pasti mewarisi gen dari orang tuanya. Dalam penelitian ini dilakukan *single-point crossover*. Dengan probabilitas *crossover* merujuk dalam jurnal [1]  $P_c$  sebesar 0.8.
  - Mutasi  
Mutasi dilakukan dengan membangkitkan kromosom *offspring* secara acak dalam bilangan biner dengan syarat probabilitas mutasi tertentu. Jika bilangan biner yang telah dibangkitkan secara acak memenuhi kriteria kurang dari probabilitas mutasi  $P_m$ , maka gen tersebut akan diubah dengan bilangan biner sebaliknya (0 diubah menjadi 1, dan 1 diubah menjadi 0). Dengan probabilitas mutasi merujuk dalam jurnal [1]  $P_m$  sebesar 0.1.
- 4) *Seleksi Survivor: Generational Replacement* digunakan sebagai seleksi survivor yang nantinya kromosom pada suatu generasi diperbaharui atau digantikan sekaligus dengan kromosom baru hasil *crossover* dan mutasi, juga serta kromosom terbaik yang sudah disimpan dalam Elitisme.
- 5) *Kriteria Terminasi*: Maksimum generasi atau maksimum iterasi sebagai kriteria terminasi GA dalam melakukan seleksi fitur.

Adapun parameter yang dibutuhkan dalam mengimplementasikan GA sebagai seleksi fitur dengan 500 individu yang akan dievaluasi pada setiap kombinasi ditentukan seperti yang tersaji dalam Tabel III.

Tabel III  
PARAMETER GA

Parameter GA	Nilai
Jumlah Individu	500
Ukuran Populasi	100
Maksimum Generasi	5
Skema Pengkodean	<i>Binary Encoding</i>
Fungsi Fitness	akurasi FLNN
<i>Crossover</i>	<i>Single Point Crossover</i>
Peluang <i>Crossover</i>	0.8
Mutasi	<i>Flip Bit Mutation</i>
Peluang Mutasi	0.1
Mekanisme Seleksi	<i>Roulette Wheel</i>
Seleksi Survivor	<i>Generational Replacement</i>

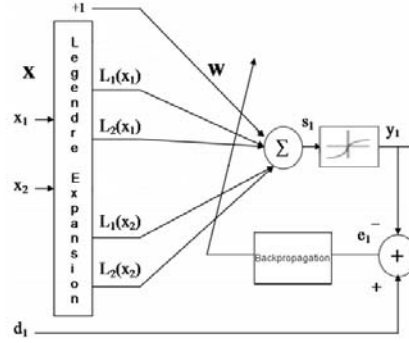
#### D. Classifier FLNN

Selanjutnya, tahapan yang menjadi inti dari penyelesaian masalah pada penelitian ini adalah melakukan klasifikasi terhadap data micorarray menggunakan FLNN sehingga diperoleh hasil klasifikasi data microarray ke dalam kelas kanker atau bukan. Jika dibandingkan dengan *neural network* yang menggunakan *hidden layer*, maka bisa dikatakan bahwa FLNN memiliki komputasi yang lebih efisien dan lebih cepat jika dibandingkan dengan Multilayer Neural Network (MNN) [5]. FLNN merupakan jaringan syaraf tiruan yang memiliki arsitektur layer tunggal, sehingga tidak memiliki *hidden layer* ataupun *hidden node*. Dalam FLNN, vektor masukan diekspansi ke dimensi yang lebih tinggi sesuai blok fungsional atau fungsi basis yang telah ditentukan. Dalam hal ini, penulis memilih Legendre Polynomial. Didukung dalam [5] Legendre Polynomial yang dikombinasikan terhadap FLNN melibatkan waktu komputasi yang lebih sedikit sehingga menawarkan proses pelatihan yang lebih cepat dibandingkan dengan MNN. Hal ini didukung dalam [3] yang menjelaskan bahwa Legendre Polynomial dianggap paling sesuai dalam permasalahan pengklasifikasian data microarray jika dibandingkan dengan fungsi basis lainnya.

Legendre Polynomial dinotasikan dengan  $L_n(X)$ ,  $n$  merupakan orde polinomial, dan  $X$  adalah nilai masukan dari data asli. Untuk membangkitkan nilai masukan ke orde yang lebih tinggi dengan Legendre Polynomial digunakan persamaan rekursif seperti berikut.

$$L_{n+1}(x) = \frac{1}{n+1} [(2n+1)xL_n - nL_{n-1}(x)] \quad (4)$$

Misalkan  $X = [x_1 x_2]^T$ . Untuk orde ke-nol dan orde pertama Legendre Polynomial diberikan  $L_0(x) = 1$  dan  $L_1(x) = x$ , maka akan diekspansi ke orde dua menjadi  $L_2(x) = \frac{1}{2}(3x^2 - 1)$  sehingga vektor masukan ke *neural network* bukanlah data yang sebenarnya melainkan nilai data yang sudah diekspansi menjadi  $X^e = [1, L_1(x_1), L_2(x_1), L_1(x_2), L_2(x_2)]^T$ . Adapun ilustrasinya seperti Gambar 2.



Gambar 2. Legendre Polynomial Orde 2 [5]

Kemudian nilai keluaran jaringan diaktivasi dengan fungsi sigmoid dengan persamaan 6 dan selanjutnya diklasifikasikan ke dalam kelas kanker atau bukan dengan nilai *threshold* 0.5. Selanjutnya, untuk mengevaluasi hasil klasifikasi digunakan fungsi *square error* yang dinotasikan dengan  $E$  dalam persamaan 7, dimana  $y_i$  merupakan kelas prediksi dan  $d_i$  merupakan target kelas dari masing-masing masukan  $i$  dan sejumlah  $n$  pola pelatihan.

$$s = wx_i - \Theta \quad (5)$$

$$f(s) = \frac{1}{1 + e^{-s}} \quad (6)$$

$$E = \frac{1}{2} \sum_{i=1}^n [d_i - y_i]^2 \quad (7)$$

Selain itu, algoritma pembelajaran yang digunakan pada FLNN adalah *backpropagation*. *Backpropagation* akan melakukan dua tahap perhitungan, yaitu perhitungan maju untuk menghitung galat antara kelas prediksi dengan target kelas, dan perhitungan mundur untuk mempropagasikan mundur galat tersebut untuk memperbaharui bobot  $w$  dengan persamaan:

$$w_i(new) = w_i(old) + \Delta w_i \quad (8)$$

$$\Delta w_i = \eta(d_i - y_i)L(x_i) \quad (9)$$

dimana  $\eta$  merupakan *learning rate* yang digunakan selama proses pelatihan data terhadap model klasifikasi. Kemudian akurasi klasifikasi berdasarkan *confusion matrix* dilakukan dengan persamaan 2 dan 3 yang selanjutnya digunakan sebagai nilai fitness pada proses evaluasi individu GA.

#### E. F1-Measure

Model klasifikasi yang dihasilkan FLNN dengan fungsi basis Legendre Polynomial terhadap data latih akan divalidasi dengan data uji berdasarkan bobot  $W$  yang telah diperbaharui dan nilai fitness selama proses pelatihan data. Kemudian, kinerja klasifikasi dari sistem yang telah dibangun akan diukur dengan memperhitungkan beberapa kriteria kinerja seperti *precision*, *recall*. Oleh sebab itu penulis menggunakan F1-Measure untuk mengukur beberapa kriteria kinerja tersebut untuk mendukung kesimpulan dan penyelesaian masalah dalam penelitian ini. Adapun persamaan yang dibutuhkan seperti berikut.

$$F_1 = \frac{2.P.R}{P + R} \quad (10)$$

P adalah *precision* yang merupakan nilai kecocokan atau kesesuaian antara informasi yang diminta dengan hasil yang diberikan oleh sistem yang dapat diperoleh dari persamaan:

$$P = \frac{TP}{TP + FP} \quad (11)$$

Sedangkan R adalah *recall* yang merupakan nilai keberhasilan sistem dalam menemukan kembali sebuah informasi yang dapat diperoleh dari persamaan:

$$R = \frac{TP}{TP + FN} \quad (12)$$

TP (*True Positive*) adalah proporsi positif dalam data set yang diklasifikasikan positif, FP (*False Positive*) adalah proporsi negatif dalam data set yang diklasifikasikan positif, TN (*True Negative*) adalah proporsi negatif dalam data set yang diklasifikasikan negatif, dan FN (*False Negative*) proporsi positif dalam data set yang diklasifikasikan negatif. Untuk TP, FP, TN dan FN diperoleh dengan memanfaatkan *Confusion Matrix*.

#### IV. HASIL DAN PEMBAHASAN

##### A. Skenario Pengujian

Pada penelitian ini dilakukan beberapa skenario pengujian dengan kombinasi parameter yang berbeda-beda seperti pada Tabel IV.

Tabel IV  
KOMBINASI PENGUJIAN

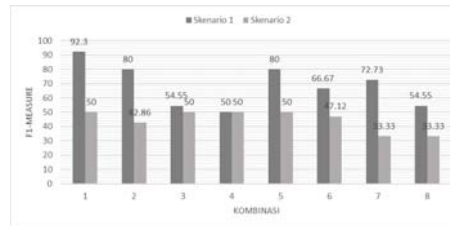
Kombinasi	Ukuran Populasi	Orde Legendre Pynomial	Learning Rate
1	100	3	0.6
2	50	3	0.6
3	100	3	0.01
4	50	3	0.01
5	100	2	0.6
6	50	2	0.6
7	100	2	0.01
8	50	2	0.01

Seperti yang tersaji dalam Tabel IV merupakan kombinasi pengujian yang akan dilakukan terhadap sistem klasifikasi yang telah dibangun. Selanjutnya data yang sudah melalui tahap *preprocessing* akan dipartisi menggunakan teknik CVPartition HoldOut dan diklasifikasikan dengan dua skenario pengujian seperti berikut:

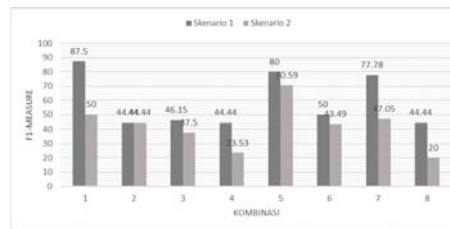
- 1) Skenario 1  
Perbandingan proporsi data latih dan data uji yang digunakan pada skenario pertama ini adalah 70:30.
- 2) Skenario 2  
Kemudian pada skenario kedua, digunakan pengujian sistem klasifikasi dengan perbandingan data latih dan data uji sebanyak 80:20.

##### B. Analisis Hasil Skenario Data

Berdasarkan pengujian yang telah dilakukan terhadap dua skenario dengan kombinasi parameter yang telah disebutkan dalam Tabel IV, maka selanjutnya dilakukan analisis terhadap skenario yang menghasilkan kinerja terbaik dalam mengklasifikasikan data Colon Tumor dan Leukemia. Gambar 3 dan Gambar 4 merupakan grafik kinerja klasifikasi yang diperoleh dari masing-masing data.



Gambar 3. Grafik Kinerja Klasifikasi Colon Tumor

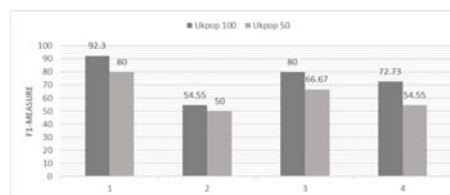


Gambar 4. Grafik Kinerja Klasifikasi Leukemia

Jika grafik yang tertera dalam Gambar 3 dan Gambar 4 diperhatikan dengan seksama, hasil klasifikasi dari kedua skenario dengan masing-masing kombinasi menghasilkan kinerja yang berbeda-beda. Namun, dari kedua skenario, Skenario 1 merupakan skenario terbaik untuk klasifikasi data Colon Tumor dan Leukemia. Hal ini dapat dilihat dari kinerja yang menurun pada Skenario 2 bahkan hingga 42.3%. Begitu juga yang terjadi dalam klasifikasi Leukemia yang mengalami penurunan kinerja sebesar 37.5% dari skenario satu ke skenario dua. Penurunan kinerja yang terjadi pada Skenario 2 membuktikan bahwa Skenario 1 merupakan skenario terbaik. Hal ini terjadi akibat pemilihan data latih dan data uji yang kurang tepat dan juga jumlah data uji yang jauh lebih sedikit pada skenario 2, sehingga menghasilkan kinerja klasifikasi yang cenderung lebih rendah.

### C. Pengaruh Ukuran Populasi (Ukpop)

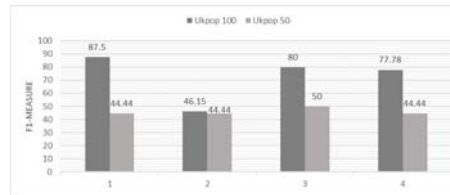
Berikut ini adalah grafik kinerja klasifikasi pada data Colon Tumor berdasarkan ukuran populasi yang berbeda-beda pada skenario 1.



Gambar 5. Grafik Kinerja Klasifikasi Colon Tumor berdasarkan Ukuran Populasi

Ukuran populasi merupakan salah satu parameter yang penting dalam penelitian ini, sehingga ukuran populasi yang telah ditentukan akan dilihat pengaruhnya terhadap kinerja klasifikasi. Dari serangkaian pengujian yang telah dilakukan, ukuran populasi ternyata memberikan pengaruh signifikan terhadap kinerja klasifikasi karena ukuran populasi memiliki dampak terhadap fitur terpilih. Ukuran populasi memiliki indikasi terhadap cakupan awal penelusuran fitur dan dampaknya terhadap evaluasi fitness dalam memperoleh fitur terpilih. Pada Gambar 5 dapat diamati bahwa dengan ukuran populasi 50, kinerja klasifikasi yang dihasilkan lebih rendah jika dibandingkan dengan kinerja klasifikasi yang dihasilkan dari ukuran populasi 100. Hal yang sama juga dapat dilihat kinerja klasifikasi Leukemia dalam Gambar 6. Meskipun kinerja klasifikasi yang dihasilkan berbeda-beda pada setiap skenario, namun dapat diperhatikan bahwa



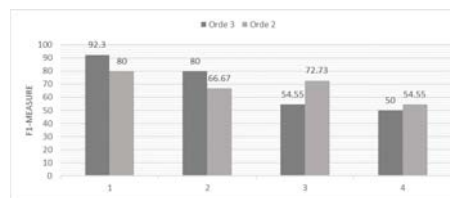


Gambar 6. Grafik Kinerja Klasifikasi Leukemia berdasarkan Ukuran Populasi

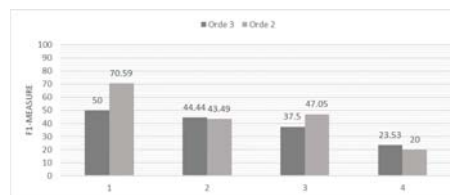
dengan ukuran populasi 50 tidak lebih baik jika dibandingkan kinerja klasifikasi yang dihasilkan dengan ukuran populasi 100. Pengujian data Colon Tumor dan Leukemia pada skenario 1 & 2 menghasilkan kinerja terbaik pada ukuran populasi 100. Oleh karena itu, dapat disimpulkan bahwa parameter GA dalam menyeleksi fitur yang paling sesuai digunakan untuk pengujian data Colon Tumor dan Leukemia adalah ukuran populasi 100.

*D. Pengaruh Orde Legendre Polynomial*

Dalam FLNN, parameter orde polynomial menjadi hal yang patut diperhatikan dalam menerapkan fungsi basis, dalam hal ini adalah Legendre Polynomial. Pengujian sistem yang dilakukan menggunakan orde polinomial yang berbeda untuk menganalisis seberapa besar pengaruh orde tersebut terhadap kinerja klasifikasi yang diperoleh. Orde polinomial yang digunakan dalam Legendre Polynomial adalah orde 3 dan orde 2. Berikut ini merupakan grafik kinerja klasifikasi yang dipengaruhi orde Legendre Polynomial pada data Colon Tumor dan Leukemia pada skenario 1.



Gambar 7. Grafik Kinerja Klasifikasi Colon Tumor berdasarkan Orde Polynomial

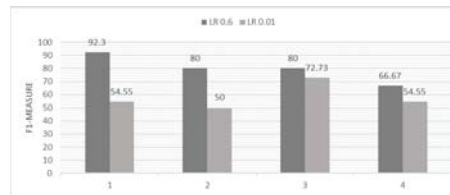


Gambar 8. Grafik Kinerja Klasifikasi Leukemia berdasarkan Orde Polynomial

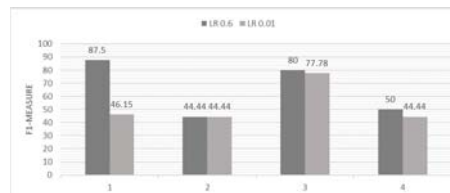
Pada Gambar 7 dan Gambar 8 merupakan hasil klasifikasi yang diperoleh dari pengujian dari masing-masing data. Grafik dalam kedua gambar tersebut menunjukkan bahwa semakin besar orde yang digunakan tidak diikuti dengan peningkatan kinerja dalam mengklasifikasikan data microarray seperti Colon Tumor dan Leukemia. Namun, kinerja yang dihasilkan berbeda-beda pada setiap orde polynomial akan memiliki pengaruh dan meningkat jika diiringi dengan LR sebesar 0.6. Selain itu, semakin besar orde polynomial tidak menjadi jaminan akan semakin besar pula kinerja yang dihasilkan. Hanya saja peningkatan orde polynomial akan meningkatkan *input space* sehingga membutuhkan proses pelatihan yang lebih lama dan meningkatkan waktu yang dibutuhkan untuk menjalankan program (*running time*), namun kinerja klasifikasi yang dihasilkan belum tentu jauh lebih baik.

### E. Pengaruh Learning Rate (LR)

Selain orde polynomial, terdapat parameter LR yang juga sangat berpengaruh, karena LR memiliki peran selama proses pelatihan. Pada penelitian ini, nilai parameter LR yang digunakan untuk pengujian yaitu 0.6 dan 0.01. Nilai yang digunakan untuk masukan parameter ini berbeda-beda sengaja dilakukan untuk mengetahui seberapa besar pengaruh nilai parameter LR tersebut terhadap kinerja yang akan dihasilkan dalam mengklasifikasikan data, baik itu Colon Tumor maupun Leukemia.



Gambar 9. Grafik Kinerja Klasifikasi Colon Tumor berdasarkan LR



Gambar 10. Grafik Kinerja Klasifikasi Leukemia berdasarkan LR

Berdasarkan berbagai pengujian data Colon Tumor yang telah dilakukan, dalam Gambar 9 menunjukkan hampir semua kombinasi parameter LR 0.6 pada skenario 1 menghasilkan kinerja yang lebih tinggi jika dibandingkan dengan nilai LR 0.01. Berbeda halnya dengan klasifikasi Leukemia, dalam Gambar 10 secara keseluruhan menunjukkan kinerja yang dihasilkan dengan LR 0.6 lebih baik. Adapun perbedaan kinerja klasifikasi yang diperoleh dapat terjadi akibat karakteristik yang dimiliki data Colon Tumor dan Leukemia berbeda. Penentuan nilai pada parameter LR menjadi salah satu aspek yang patut dipertimbangkan dalam FLNN, karena sangat memengaruhi kinerja jaringan syaraf dalam mencapai hasil yang diharapkan. Penentuan nilai LR akan berdampak pada kinerja pembelajaran *backpropagation*. Dimana, LR menjadi parameter yang digunakan dalam proses memperbaharui bobot untuk setiap masukan. Jika LR terlalu kecil maka algoritma akan membutuhkan waktu yang lama untuk mencapai konvergen, dan sebaliknya jika LR terlalu besar maka algoritma menjadi divergen. Dari kedua kombinasi nilai LR yang telah digunakan, nilai LR 0.6 dianggap paling sesuai dengan pola data microarray yaitu Colon Tumor dan Leukemia.

## V. KESIMPULAN

Dari serangkaian pengujian yang telah dilakukan terhadap sistem klasifikasi, maka diperoleh kinerja klasifikasi yang beragam. Namun jika dilihat berdasarkan skenario pengujian, kinerja terbaik diperoleh dari skenario 1 yang proporsi data latih berbanding data uji sebesar 70:30. Kemudian berdasarkan kombinasi parameter seperti ukuran populasi 100 dan LR 0.6 dianggap paling sesuai karena menghasilkan kinerja klasifikasi yang lebih tinggi. Namun, peningkatan orde pada Legendre Polynomial tidak menjamin akan meningkatkan kinerja klasifikasi yang dihasilkan. Hanya saja peningkatan kinerja klasifikasi akan cenderung meningkat jika diiringi dengan LR sebesar 0.6. Selain itu peningkatan orde polynomial akan meningkatkan ruang masukan sehingga membutuhkan proses pelatihan lebih lama. Kinerja klasifikasi yang dihasilkan terhadap Colon Tumor mencapai nilai tertinggi sebesar 92.3% dan pada klasifikasi Colon

Tumor sebesar 87.5% yang masing-masing dihasilkan dari kombinasi parameter ukuran populasi 100, Legendre Polynomial orde 3 dengan LR 0.6.

#### DAFTAR PUSTAKA

- [1] O Babatunde, Leisa Armstrong, Jinsong Leng, and Dean Diepeveen. A genetic algorithm-based feature selection. *International Journal of Electronics Communication and Computer Engineering*, 5(4):889–905, 2014.
- [2] Erick Cantu-Paz. Feature subset selection, class separability, and genetic algorithms. In *Genetic and Evolutionary Computation—GECCO 2004*, pages 959–970. Springer, 2004.
- [3] Mukesh Kumar, Sandeep Singh, and Santanu Kumar Rath. Classification of microarray data using functional link neural network. *Procedia Computer Science*, 57:727–737, 2015.
- [4] Il-Seok Oh, Jin-Seon Lee, and Byung-Ro Moon. Hybrid genetic algorithms for feature selection. *IEEE Transactions on pattern analysis and machine intelligence*, 26(11):1424–1437, 2004.
- [5] Deepti Moyi Sahoo and S Chakraverty. Functional link neural network approach to solve structural system identification problems. *Neural Computing and Applications*, pages 1–12, 2017.
- [6] Tan Ching Siang, Ting Wai Soon, Shahreen Kasim, Mohd Saberi Mohamad, Chan Weng Howe, Safaai Deris, Zalmiyah Zakaria, Zuraini Ali Shah, and Zuwairie Ibrahim. A review of cancer classification software for gene expression data. *International Journal of Bio-Science and Bio-Technology*, 7(4):89–108, 2015.
- [7] Rabindra Kumar Singh and M Sivabalakrishnan. Feature selection of gene expression data for cancer classification: a review. *Procedia Computer Science*, 50:52–57, 2015.
- [8] C Devi Arockia Vanitha, D Devaraj, and M Venkatesulu. Gene expression data classification using support vector machine and mutual information-based gene selection. *Procedia Computer Science*, 47:13–21, 2015.

