

Implementasi *Minimum Redudancy Maksimum Relevance* (MRMR) dan *Genetic Algorithm* (GA) untuk Reduksi Dimensi pada Klasifikasi Data Micorarray Menggunakan *Functional Link Neural Network* (FLNN)

Bisma Pradana¹, Adiwijaya², Annisa Aditsania³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹bismapradana@students.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id,

³aaditsania@telkomuniversity.ac.id

Abstrak

Dalam bidang kesehatan DNA *microarray* banyak digunakan untuk memprediksi penyakit kanker, dimana sel kanker dapat mengalami abnormalitas dalam mengekspresikan gennya. DNA *microarray* ini memungkinkan untuk mengetahui tahapan perkembangan sel kanker dengan melihat level ekspresi gennya. Analisis data ekspresi gen dalam bentuk *microarray* dapat memudahkan ahli medis dalam mendeteksi dan mendiagnosis apakah seseorang menderita kanker atau tidak. Data *microarray* sendiri memiliki dimensi yang besar sehingga dapat mempengaruhi proses dan akurasi klasifikasinya. Oleh karena itu, untuk melakukan proses pengklasifikasian pada data *microarray*, perlu dilakukan sebuah proses yaitu reduksi dimensi. Reduksi dimensi ini bertujuan agar mengurangi *redundancy* dan meningkatkan *relevance* pada data yang dimiliki. Dalam hal ini, penulis menggunakan metode *Minimum Redudancy Maksimum Relevance* (MRMR) yang dioptimasi menggunakan metode *Genetic Algorithm* (GA). Secara singkat MRMR adalah teknik yang digunakan untuk mengidentifikasi himpunan gen yang memiliki relevansi tinggi terhadap atribut kelas akan tetapi memiliki redundansi yang rendah antar atributnya. MRMR berupaya untuk mengatasi permasalahan ini dengan menghapus subset atribut yang dianggap tidak diperlukan. Kemudian algoritma *Functional Link Neural Network* (FLNN) dengan basis *Legendre Polynomial* digunakan untuk proses pengklasifikasian pada data *microarray*. Pada pengujian MRMR GA dengan klasifikasi *Functional Link Neural Network* (FLNN) didapatkan kenaikan hasil akurasi pada data Colon Tumor sebesar 5,55% - 16,66%, pada data Lung sebesar 2,78 - 5,56% dan pada data Ovarian sebesar 2% - 2,67%.

Kata kunci: Klasifikasi, FLNN, MRMR, GA, DNA *microarray*

Abstract

In the field of health DNA microarrays are widely used to predict cancer, in which cancer cells can experience abnormalities in expressing their genes. This DNA microarray makes it possible to determine the stages of development of cancer cells by looking at the level of gene expression. Analysis of gene expression data in the form of microarrays can facilitate medical experts in detecting and diagnosing whether a person has cancer or not. Microarray data itself has large dimensions so that it can affect the process and classification accuracy. Therefore, to do the classification process in microarray data, it is necessary to do a process that is dimension reduction. This dimension reduction aims to reduce redundancy and increase the relevance of the data owned. In this case, the author uses the Minimum Reduction Maximum Relevance (MRMR) method which is optimized using the Genetic Algorithm (GA) method. In short MRMR is a technique used to identify a set of genes that has high relevance to class attributes but has low redundancy between attributes. MRMR seeks to overcome this problem by removing the subset of attributes that are considered unnecessary. Then the Functional Link Neural Network (FLNN) algorithm with Legendre Polynomial base is used for the classification process in microarray data. In the GA MRMR test with the Functional Link Neural Network (FLNN) classification we found an increase in the accuracy of the Colon Tumor data of 5.55% - 16.66%, in the Lung data of 2.78 - 5.56% and in the Ovarian data of 2% - 2.67%.

Keywords: *classification*, FLNN, MRMR, GA, DNA *microarray*

1. Pendahuluan

Latar Belakang

Kanker adalah salah satu jenis penyakit paling berbahaya di dunia [1]. Penyebab utamanya adalah akibat dari pertumbuhan tidak normal sel - sel jaringan tubuh yang berubah menjadi kanker. Di Indonesia sendiri menurut data Riskesdes tahun 2013, prevalansi tumor atau kanker adalah 1,4 per 1.000 penduduk atau sekitar 347.000 orang, dan yang paling banyak dialami oleh perempuan adalah kanker payudara dan kanker leher rahim. Dari sisi

pembiayaan, berdasarkan data dari BPJS sehingga bulan September 2017, penyakit kanker telah menghabiskan biaya 2,1 triliun rupiah [13].

Seiring dengan berkembangnya teknologi, terdapat juga cara dalam mendeteksi kanker yaitu dengan menggunakan teknologi DNA *microarray* yang mana teknologi ini memungkinkan tingkat ribuan ekspresi gen dapat dipantau secara bersamaan dalam satu percobaan [2]. Teknologi ini juga memungkinkan untuk mengetahui tahapan perkembangan dari sel kanker itu sendiri dengan melihat level ekspresi gennya. Analisis ekspresi gen ini juga sangat penting karena dengan ini memungkinkan untuk bisa lebih kompleks dalam mendeteksi kanker. Data *microarray* memiliki dimensi yang sangat besar. Maka dari itu untuk memperoleh hasil yang maksimal diperlukan algoritma yang cukup kompleks dalam mempelajari karakteristik dari ekspresi gen ini. Akan tetapi sebelum melakukan proses klasifikasi diperlukan yang namanya reduksi dimensi terlebih dahulu, karena data *microarray* ini memiliki dimensi yang sangat besar. Tujuan dari reduksi dimensi ini ialah untuk mengurangi dimensi atau mengisolasi dimensi yang dianggap tidak diperlukan dalam pengklasifikasiannya, sehingga akan meningkatkan terhadap hasil akurasi.

Pada tugas akhir ini metode yang digunakan dalam mereduksi dimensi adalah *Minimum Redundancy Maksimum Relevance* (MRMR) yang kemudian dioptimasi menggunakan *Genetic Algorithm* (GA). Secara istilah, MRMR memberikan informasi relevansi dan redundansi dari suatu sub himpunan. Penulis menggunakan reduksi dimensi ini karena dari penelitian yang sudah dilakukan menggunakan metode MRMR yang dioptimasi menggunakan *Genetic Algorithm* (GA) terbukti menghasilkan tingkat akurasi yang tinggi [3].

Setelah proses reduksi dimensi dilakukan kemudian dilakukan proses klasifikasi terhadap data *microarray*. Metode klasifikasi yang digunakan ialah *Functional Link Neural Network* (FLNN) dengan basis *Legendre Polynomial*, yang mana dalam tahapan ini akan diperoleh hasil dari klasifikasi berupa kelas – kelas yang termasuk kedalam kanker atau bukan kanker. *Functional Link Neural Network* (FLNN) dengan basis *Legendre Polynomial* digunakan karena pada penelitian yang telah dilakukan sebelumnya oleh Putri Tsatsabila Ramadhani dapat menghasilkan nilai akurasi yang cukup baik dibandingkan dengan metode klasifikasi lainnya [4].

Topik dan Batasannya

Pada tugas akhir ini penulis mengangkat permasalahan terkait bagaimana cara membangun sistem klasifikasi kanker berdasarkan data *microarray*, selanjutnya bagaimana menganalisis peranan reduksi dimensi pada klasifikasi data *microarray*, dan bagaimana cara menghitung nilai akurasi dari hasil reduksi dimensi pada sistem klasifikasi berdasarkan hasil pengujian. Selain itu, terdapat beberapa hal yang dijadikan sebagai batasan masalah pada tugas akhir ini yaitu data yang digunakan berupa ekspresi gen *microarray* dari jenis – jenis kanker yang berbeda sebagai berikut: kanker Ovarian, kanker Lung, dan kanker Colon. Data didapatkan dari Kent-Ridge Bio-medical Data Set Repository [14].

Tujuan

Tujuan dari penulisan tugas akhir ini antara lain dapat membangun model klasifikasi menggunakan metode *Functional Link Neural Network* (FLNN) dalam mengklasifikasikan data *microarray*, kemudian dapat menganalisis dan mengimplementasikan peranan dari metode reduksi dimensi *Minimum Redundancy Maksimum Relevance* (MRMR) dan *Genetic Algorithm* (GA) dalam mereduksi dimensi pada data *microarray* terhadap sistem klasifikasi yang dibangun, dan mengetahui cara menghitung nilai akurasi dari hasil MRMR GA pada sistem yang dibangun berdasarkan hasil pengujian.

2. Studi Terkait

Dalam beberapa tahun terakhir klasifikasi data *microarray* merupakan salah satu topik yang sering dijadikan objek penelitian. Salah satu masalah dalam mengklasifikasikan data *microarray* ialah karena data *microarray* memiliki dimensi yang sangat besar. Maka dari itu untuk memperoleh hasil yang maksimal diperlukan algoritma yang cukup kompleks dalam mempelajari karakteristik dari ekspresi gen ini.

Pada tahun 2015, Mukesh Khumar [5] melakukan sebuah penelitian tentang klasifikasi data *microarray* menggunakan *Functional Link Neural Network*. Data *microarray* yang digunakan dalam penelitian ini adalah data *leukimia cancer*, *ovarian cancer*, dan *breast cancer*. Dalam penelitian ini dihasilkan nilai akurasi sebesar 96,15% *leukimia cancer*, 97,78% *ovarian cancer*, dan 86,54 untuk *breast cancer*.

Pada tahun 2017, Putri Tsabila Ramdhani [4] melakukan penelitian tentang deteksi kanker berdasarkan klasifikasi data *microarray* menggunakan *Functional Link Neural Network* dengan seleksi fitur *genetic algorithm*. Di dalam penelitian ini, dihasilkan hasil akurasi yang cukup tinggi dalam mengklasifikasikan data *microarray*. Jenis data *microarray* yang digunakan dalam penelitian ini adalah colon tumor dan leukimia dengan tingkat akurasi mencapai 92,3% dan 87,5%.

Firda A. Ma'ruf [6] pada tahun 2018 melakukan penelitian tentang analisis pengaruh reduksi dimensi *Minimum Redudancy Maximum Relevance* (MRMR) pada klasifikasi kanker berdasarkan data *microarray* menggunakan *clasiffier Support Vector Machine* (SVM), dalam penelitian ini pengaruh reduksi dimensi terbukti lebih baik dibandingkan dengan klasifikasi data *microarray* tanpa reduksi dimensi. Hasilnya, nilai akurasi yang dihasilkan dengan menggunakan reduksi dimensi mencapai 0,81667 sedangkan nilai akurasi dengan tidak menggunakan reduksi dimensi mencapai 0,72.

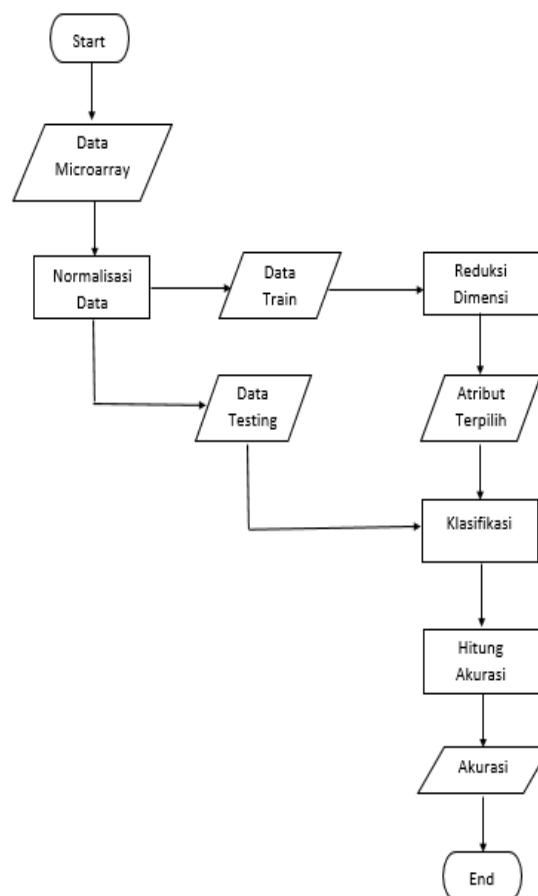
Pada tahun 2018, Adiwijaya [7] melakukan penelitian terhadap data *microarray* menggunakan metode *Principal Component Analysis* (PCA) untuk mereduksi dimensi dengan *Support Vector Machine* (SVM) dan *Levenberg-Marquardt Backpropagation* (LMBP) untuk klasifikasi. Hasil akurasi terbaik didapatkan dengan menggunakan SVM adalah sebesar 94,98% sedangkan hasil akurasi terbaik menggunakan LMBP adalah 96,07%.

Pada tahun 2018, Raden Muhammad Ananda Egantara [3] melakukan penelitian terhadap data *microarray* menggunakan *Minimum Redudancy Maksimum Relevance* (MRMR) dan *Backpropagation* termodifikasi dengan *Conjugate Gradient Polak Ribiere*. Hasil akurasi untuk fitur seleksi telah mereduksi data Lung dengan minimum akurasi sebesar 90%, Ovarian 97%, dan Colon 62%.

Pada tahun 2018, Husna Adyadenta [8] melakukan penelitian terhadap data *microarray* menggunakan seleksi fitur *Random Forest* dan empat jenis metode klasifikasi yaitu *Support Vector Machine* (SVM), *Artificial Neural Network* (ANN), *Naive Bayes*, *K-nearst Neighbor* (KNN), dan C4.5 mendapatkan hasil untuk *Random Forest* sebesar 75%, SVM 66,67%, ANN 66,67%, *Naive Bayes* 41,67%, K-NN 58,33%, dan C4.5 sebesar 58,33%.

3. Sistem yang Dibangun

Proses yang dilakukan dalam sistem ini terbagi menjadi 4 proses utama, berikut merupakan proses yang akan dilakukan dalam sistem yang dibangun yaitu terdiri dari *preprocessing* data *microarray* menggunakan metode min- max, reduksi dimensi menggunakan metode MRMR GA, klasifikasi menggunakan FLNN, dan kemudian menghitung akurasi dengan menggunakan metode *Confusion Matrix*.



Gambar 1 Rancangan Sistem Implementasi *Minimum Redudancy Maksimum Relevance* (MRMR) dan *Genetic Algorithm* (GA) untuk Reduksi Dimensi pada Klasifikasi Data *Microarray* Menggunakan *Functional Link Neural Network* (FLNN).

1. Data *Microarray*

Data yang digunakan untuk klasifikasi kanker pada tugas akhir ini adalah data *microarray* dari tiga jenis yang berbeda yaitu data kanker Colon, kanker Lung, dan kanker Ovarian. Data diperoleh dari repositori Kent-Ridge Bio-medical Data Set Repository [14]. Data *microarray* merupakan data yang memiliki dimensi sangat besar. Data *Microarray* juga merupakan perkembangan teknologi terakhir dalam biologi molekuler, yang dapat mengukur tingkat ekspresi gen dari puluhan ribu gen secara paralel sehingga dapat menghasilkan data yang sangat besar dan berharga. Teknologi *DNA microarray* ini merupakan sebuah alat yang dapat digunakan untuk mendeteksi tingkat ekspresi gen yang berbeda. Pada dasarnya, *DNA microarray* akan mengamati ekspresi genetik dari berbagai gen dengan jumlah besar secara bersamaan. Dalam analisis kanker, digunakan *DNA* normal dan *DNA* kanker yang kemudian akan di hibridisasi lalu akan dibaca menggunakan *scanner*. Berikut merupakan spesifikasi dari data yang akan digunakan dalam penulisan tugas akhir ini.

Tabel 1 Data Kanker

Data	Jumlah Kelas	Sampel	Fitur
Kanker Colon	2	62 (22 positif, 40 negatif)	2.000
Kanker Lung	2	181 (31 mesothelioma, 150 ADCA)	12.533
Kanker Ovarium	2	253 (91 normal, 162 cancer)	15.154

2. Normalisasi Data

Pada data *microarray* diperlukan *preprocessing* terhadap data input, karena data *microarray* merupakan data yang ukurannya sangat besar. Data perlu dinormalisasi ke dalam bentuk *range* 0 - 1 agar memudahkan dalam pemrosesan data, karena data yang digunakan memiliki perbedaan nilai cukup jauh antara satu dengan yang lain pada setiap atributnya. Sehingga digunakan metode normalisasi min – max dalam mengubah data, yang mana berikut adalah persamaan dari metode normalisasi min – max [6].

$$X_{ni} = \frac{X_i - \min(X)}{\max(X) - \min(X)} \quad (1)$$

dimana :

X_{ni} = data hasil normalisasi ke- i .

X_i = data aktual dengan *range* data asli ke- i .

X = data aktual dengan *range* asli.

3. Reduksi Dimensi

Reduksi dimensi yang digunakan adalah *Minimum Redudancy Maximun Relevance* (MRMR) dan *Genetic Algorithm* (GA). Secara istilah, MRMR memberikan informasi relevansi dan redundansi dari suatu sub himpunan. Pada penelitian ini MRMR diukur menggunakan *F-test Correlation Quotient* (FCQ). Kemudian akan dioptimalkan nilai FCQ menggunakan *Genetic Algorithm* (GA). Optimalisasi menggunakan *Genetic Algorithm* (GA) ini bertujuan mencari sub himpunan dengan FCQ tertinggi. Oleh karena itu FCQ pada penelitian ini dipilih sebagai fitness GA. Sehingga dapat diasumsikan individu dengan nilai *fitness* tertinggi merepresentasikan sub himpunan dengan nilai FCQ tertinggi. Pada data kontinu, untuk mencari nilai relevansi dapat menggunakan F-statistik antara gen dengan kelasnya. Nilai F-test variabel gen g_i ke atribut kelas k dilambangkan dengan h , disajikan pada persamaan berikut.

$$F(g_i, h) = [\sum_k n_k (\bar{g}_k - \bar{g}) / (k - 1)] / \sigma^2 \quad (2)$$

g merupakan rata – rata g_i di semua sampel jaringan, g_k merupakan nilai rata – rata dari g_i dalam kelas k , dan $\sigma^2 = [\sum_k(n_k - 1)\sigma_k^2 / (n - k)]$ merupakan varians gabungan (di mana n_k dan σ_k adalah ukuran varians dari kelas k). Dengan demikian, untuk relevansi himpunan set fitur S dapat dituliskan sebagai berikut.

$$\max V_f, V_f = \frac{1}{|S|} \sum_{i \in S} F(i, h) \quad (3)$$

banyak cara dalam menentukan *redundancy* minimum, apabila menggunakan koefisien korelasi pearson c . $(g_i, g_j) = c(i, j)$, maka persamaan kondisi minimum *redundancy* adalah sebagai berikut.

$$\min W_c, W_c = \frac{1}{|S|^2} \sum_{i,j} |c(i, j)| \quad (4)$$

dimana $c(i, j)$ dapat diperoleh menggunakan persamaan berikut.

$$c(i, j) = \frac{n \sum(ij) - (\sum i) (\sum j)}{\sqrt{[n \sum i^2 - (\sum i)^2][n \sum j^2 - (\sum j)^2]}} \quad (5)$$

keterangan:

n = banyaknya pasangan data i dan j .

$\sum i$ = total jumlah variabel i .

$\sum j$ = total jumlah variabel j .

$\sum i^2$ = kuadrat dari total jumlah variabel i .

$\sum j^2$ = kuadrat dari total jumlah variabel j .

$\sum(ij)$ = hasil perkalian dari total jumlah variabel i dan variabel j .

untuk data kontinyu dalam metode MRMR dapat menggunakan *F-test Correlation Quotient* (FCQ) dengan persamaan sebagai berikut.

$$\max_{i \in \Omega_S} [F(i, h) / \frac{1}{|S|} \sum_{j \in S} |c(i, j)|] \quad (6)$$

keterangan:

h = nilai uji F gen g_i pada kelas k .

\bar{g} = nilai rata – rata g_i di semua sampel jaringan.

\bar{g}_k = nilai rata – rata g_i dalam kelas k .

n_k dan σ_k = ukuran dan varians dari kelas k .

Terdapat beberapa metode dalam menyelesaikan masalah optimasi, salah satunya adalah dengan menggunakan *Genetic Algorithm* (GA). *Genetic Algorithm* (GA) sendiri merupakan cabang dari ilmu algoritma evolusi yang mana biasa digunakan dalam memecahkan permasalahan pada optimasi [9]. Berikut adalah beberapa proses atau komponen yang dilakukan pada *Genetic Algorithm* (GA).

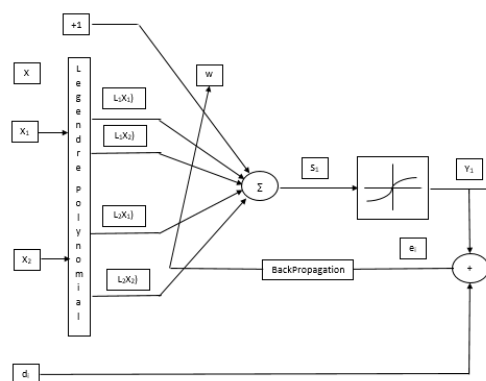
1. Inisialisasi Awal

Inisialisasi awal merupakan proses menentukan parameter – parameter pengendali jalannya *genetic algorithm* (GA), dimana berikut adalah parameter – parameter *genetic algorithm* (GA) yang digunakan dalam tugas akhir ini. (*Ukpop*) merupakan banyaknya individu yang terdapat dalam populasi, jumlah maksimum generasi (Max Gen) yaitu jumlah maksimum iterasi yang akan dijalankan pada proses *genetic algorithm* (GA), probabilitas *crossover* (*Pc*) merupakan penentuan besarnya kemungkinan individu untuk melakukan operasi pindah silang, kemudian probabilitas mutasi (*Pm*) merupakan penentuan besarnya kemungkinan individu terkena mutasi. Untuk ukuran populasi setiap individunya direpresentasikan ke dalam deretan bilangan biner 0 dan 1. Inisialisasi populasi berdasarkan bilangan biner 0 dan 1 inilah yang akan dibangkitkan secara acak sebanyak jumlah fitur dan ukuran populasinya. Setiap individu pada populasi merepresentasikan sebuah solusi kandidat pada permasalahan seleksi fitur. Apabila bitnya sama dengan 0 maka artinya fitur tersebut tidak akan terpilih sebagai calon solusi, namun sebaliknya apabila bit sama dengan 1 maka fitur tersebut akan terpilih sebagai calon solusi nantinya. Dimana untuk ukuran populasi dengan maksimum generasi pada tugas akhir ini merujuk pada jurnal [9].

2. Evaluasi *Fitness*
Dalam tugas akhir ini, *F-test Correlation Quotient* (FCQ) digunakan untuk menghasilkan nilai *fitness*, dimana individu yang memiliki nilai *fitness* lebih tinggi akan memiliki kemampuan bertahan hidup lebih tinggi dibandingkan dengan individu yang memiliki nilai *fitness* rendah.
3. *Elitisme*
Kemudian individu dengan nilai *fitness* tertinggi akan disimpan dalam *elitisme*. *Elitisme* ini bertujuan agar *fitness* tertinggi tersebut, yang mana merupakan kandidat solusi terbaik tidak hilang selama operasi genetika berlangsung.
4. Seleksi Orang Tua
Seleksi orang tua dilakukan dengan cara memilih individu berdasarkan nilai *fitness* yang telah dihasilkan. Proses seleksi orang tua dalam penelitian ini menggunakan metode *roulette-wheel*.
5. Crossover
Crossover merupakan operator dari *genetic algorithm* yang melibatkan dua orang tua untuk membentuk individu baru berdasarkan titik *crossover* yang ditentukan secara acak. Jenis *crossover* yang digunakan dalam tugas akhir ini adalah *single point crossover*, dimana untuk probabilitas *crossover* (P_c) yang digunakan dalam tugas akhir ini merujuk pada jurnal [10] yaitu P_c sebesar 0,8.
6. Mutasi
Mutasi diperlukan untuk mengembalikan informasi bit yang hilang akibat dari proses *crossover* dengan syarat probabilitas mutasi (P_m) tertentu. Dimana dalam tugas akhir ini probabilitas mutasi (P_m) merujuk pada jurnal [10] yaitu P_m sebesar 0,1.
7. Seleksi Survivor
Seleksi survivor atau pergantian populasi dilakukan apabila telah melalui proses *crossover* dan mutasi, karena seleksi survivor ini menghasilkan dampak yang signifikan terhadap performansi dan kemampuan GA dalam menemukan solusi. Dalam tugas akhir ini *general replacement* digunakan sebagai seleksi survivor, yang mana nantinya setiap individu pada suatu generasi akan diperbaharui dengan individu yang baru hasil dari *crossover* dan mutasi serta individu terbaik yang telah disimpan dalam *elitisme*.

4. Klasifikasi

Selanjutnya melakukan proses pengklasifikasian terhadap data *microarray*, yang mana bertujuan untuk mengetahui bahwa data *microarray* tersebut termasuk ke dalam kelas 0 atau 1, dimana kelas 0 mengindikasikan data terklasifikasi negatif dan kelas 1 mengindikasikan data terklasifikasi positif. Metode yang digunakan untuk proses pengklasifikasian ini adalah *Functional Link Neural Network* (FLNN) dengan menggunakan fungsi basis *Legendre Polynomial*. *Functional Link Neural Network* (FLNN) merupakan jaringan syaraf tiruan yang memiliki arsitektur *layer* tunggal, sehingga tidak memiliki *hidden layer* [4]. Dalam arsitektur FLNN node tersembunyi dikecualikan dan pola inputan diperluas untuk pola yang lebih tinggi dengan cara blok ekspansi fungsional atau fungsi basis yang telah ditentukan. Dalam kasus ini penulis menggunakan basis *Legendre Polynomial* [11]. Berikut merupakan ilustrasi arsitektur dari klasifikasi menggunakan *Functional Link Neural Network* (FLNN) dengan basis *Legendre Polynomial*.



Gambar 2 Ilustrasi Arsitektur FLNN dengan Basis Legendre Polynomial.

Legendre Polynomial sendiri dinotasikan dengan $L_x(X)$, dengan n adalah ordenya dan X adalah nilai masukan dari data asli. Untuk membangkitkan orde dari basis *Legendre Polynomial* dapat digunakan persamaan rekursif sebagai berikut [11].

$$L_{n+1}(x) = \frac{1}{n+1} [(2n+1)xL_n - nL_{n-1}(x)] \quad (7)$$

kemudian nilai *output* jaringan akan diaktivasi dengan menggunakan fungsi aktivasi *sigmoid* pada persamaan 9, selanjutnya akan diklasifikasikan berdasarkan kelas yang telah ditentukan dengan nilai *threshold* 0.5 [4]. Lalu kemudian setelah itu akan dievaluasi hasil dari klasifikasinya menggunakan fungsi *error square* yang dinotasikan dengan E pada persamaan 10.

$$s = wx_i - bias \quad (8)$$

$$f(s) = \frac{1}{1 + e^s} \quad (9)$$

$$E = \frac{1}{2} \sum_{i=1}^n [d_i - y_i]^2 \quad (10)$$

selain itu algoritma pembelajaran di dalam *Functional Link Neural Network* (FLNN) adalah *backpropagation*, yang mana di dalam *backpropagation* tersebut menggunakan dua tahap perhitungan. Pertama perhitungan maju yaitu untuk mengetahui galat antar kelasnya dan kedua perhitungan mundur untuk mempropagasikan mundur galat tersebut dalam memperbaharui bobot w nya. Berikut adalah persamaannya [4].

$$w_i(\text{baru}) = w_i(\text{lama}) + \Delta w_i \quad (11)$$

$$\Delta w_i = n(d_i - y_i)L(x_i) \quad (12)$$

dimana n merupakan *learning rate* (LR) yang digunakan selama proses pelatihan data terhadap model klasifikasi. Lalu kemudian akurasi dari klasifikasi berdasarkan *confusion matrix* dilakukan pada persamaan 13 dan persamaan 6 selanjutnya akan digunakan sebagai nilai fitness pada proses evaluasi individu GA.

5. Menghitung Akurasi

Tahap terakhir dari tugas akhir ini adalah menghitung nilai akurasi dari hasil klasifikasi yang telah dilakukan sebelumnya, dimana dengan hasil ini akan diketahui seberapa akurat hasil klasifikasi yang telah dilakukan. Metode yang digunakan untuk menghitung nilai akurasi dalam tugas akhir ini adalah *Confusion Matrix*, dimana *Confusion Matrix* merupakan sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan [12]. Perhitungan akurasi dinyatakan dalam persamaan berikut.

$$\text{Akurasi}(x) = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (13)$$

dimana pengukuran kinerja menggunakan *confusion matrix* terdapat 4 istilah sebagai representasi dari hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *True Positive* (TP) merupakan jumlah data positif yang terdeteksi benar, *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi benar, *False Positive* (FP) merupakan jumlah data negatif yang terdeteksi sebagai data positif, dan kemudian *False Negative* (FN) merupakan jumlah data positif yang terdeteksi negatif [2]. *Confusion Matrix* dapat disajikan sebagai berikut:

Tabel 2 Confusion Matrix

Kelas	Positive	Negative
Positive	(TP) True Positive	(FN) False Negative
Negative	(FP) False Positive	(TN) True Negative

4. Hasil Pengujian dan Analisis

Pada bagian ini memaparkan bagaimana hasil pengujian dari sistem yang telah dibuat dan dirancang sesuai dengan gambar (1), serta akurasi dari setiap data yang digunakan dalam penelitian tugas akhir ini.

4.1. Hasil Pengujian

Pada tugas akhir ini dilakukan beberapa skenario pengujian yaitu dengan membagi data menjadi data latih dan data uji, teknik yang digunakan dalam membagi data latih dan data uji ini adalah CVPartition HoldOut dengan proporsi seperti berikut:

1. Skenario 1
Perbandingan proporsi dengan perbandingan data latih dan data uji sebesar 70:30.
2. Skenario 2
Perbandingan proporsi dengan perbandingan data latih dan data uji sebesar 80:20.

4.2. Analisis Hasil Pengujian

Pada bagian ini akan memaparkan analisis hasil pengujian berdasarkan seluruh parameter yang digunakan pada tugas akhir ini, dimana parameter yang akan diujikan adalah sebagai berikut.

Tabel 3 Hasil Akurasi Skenario 1 Data *Microarray* pada Klasifikasi Tanpa Menggunakan Reduksi Dimensi Dan Menggunakan Reduksi Dimensi

DATA	<i>Orde Legendre Polynomial</i>	Akurasi	
		FLNN	MRMR GA + FLNN
COLON	2	77,78%	83,33%
	3	66,67%	83,33%
LUNG	2	96,30%	100%
	3	94,44%	100%
OVARIAN	2	97,33%	100%
	3	97,33%	97,33%

Tabel 3 Hasil Akurasi Skenario 2 Data *Microarray* pada Klasifikasi Tanpa Menggunakan Reduksi Dimensi Dan Menggunakan Reduksi Dimensi

DATA	<i>Orde Legendre Polynomial</i>	AKURASI	
		FLNN	MRMR GA + FLNN
COLON	2	83,33%	91,67%
	3	83,33%	91,67%
LUNG	2	97,22%	100%
	3	97,22%	100%
OVARIAN	2	98%	100%
	3	96%	96%

Tabel 3 dan 4 menyajikan hasil pengujian dengan menggunakan MRMR GA + FLNN dan tanpa menggunakan MRMR GA. Jika diperhatikan hasil klasifikasi dari skenario 1 dan skenario 2 pada masing - masing kombinasi menghasilkan kinerja klasifikasi yang berbeda dengan adanya perubahan nilai akurasi yang cenderung meningkat dibandingkan tanpa menggunakan MRMR GA. Dalam hal ini menunjukkan bahwa pengaruh dari MRMR GA dengan parameter ukuran populasi 100 dan maksimum generasi 8 kemudian orde *legendre polynomial* 2 dan 3, dapat meningkatkan kinerja klasifikasi pada setiap datasetnya. Terlihat pada skenario 1 bahwa pengaruh dari MRMR GA dapat meningkatkan nilai akurasi terhadap data Colon dengan *legendre polynomial* orde 2 sebesar 5,55% kemudian pada *legendre polynomial* orde 3 sebesar 16,66%, pada data Lung dengan *legendre polynomial* orde 2 untuk nilai akurasi meningkat sebesar 3,7% namun untuk data Lung dengan *legendre polynomial* orde 3 meningkat sebesar 5,56%, pada data Ovarian untuk *legendre polynomial* orde 2 meningkat sebesar 2,67% untuk *legendre polynomial* orde 3 nilai akurasinya tetap sama. Kemudian pada skenario 2 pun sama bahwa pengaruh dari MRMR GA dapat meningkatkan nilai akurasi terhadap data Colon sebesar 8,34% untuk *legendre polynomial*

orde 2 kemudian *legendre polynomial* orde 3 juga sama yaitu sebesar 8,34%, pada data Lung untuk *legendre polynomial* orde 2 meningkat sebesar 2,78% *legendre polynomial* orde 3 pun sama meningkatkan nilai akurasi sebesar 2,78%, dan terakhir terhadap data Ovarian yaitu untuk *legendre polynomial* orde 2 meningkatkan nilai akurasi sebesar 2% namun *legendre polynomial* orde 3 tetap sama. Sehingga dapat disimpulkan bahwa MRMR GA sangat berpengaruh peranannya terhadap nilai akurasi yang dihasilkan pada saat proses pengklasifikasian data *microarray*.

Tabel 5 Hasil Akurasi Skenario 1 Pada Data Colon, Lung, Ovarian Dengan Berdasarkan Pengaruh Dari Orde *Legendre Polynomial*

Data	Orde Legendre Polynomial	Akurasi	Running Time (second)
Colon	2	77,78%	3,21
	3	66,67%	3,46
Lung	2	96,30%	78,75
	3	94,44%	85,16
Ovarian	2	97,33%	120,92
	3	97,33%	130,93

Tabel 6 Hasil Akurasi Skenario 2 Pada Data Colon, Lung, Ovarian Dengan Berdasarkan Pengaruh Dari Orde *Legendre Polynomial*

Data	Orde Legendre Polynomial	Akurasi	Running Time (second)
Colon	2	83,33%	3,26
	3	83,33%	3,63
Lung	2	97,22%	71,89
	3	97,22%	86,55
Ovarian	2	98%	120,83
	3	96%	133,54

Tabel 5 dan Tabel 6 menyajikan hasil pengujian berdasarkan pengaruh dari orde *legendre polynomial*. Dalam FLNN pengujian berdasarkan pengaruh orde *legendre polynomial* menjadi hal yang patut diperhatikan, tujuannya agar dapat menganalisis seberapa besar pengaruh dari orde *legendre polynomial* tersebut terhadap kinerja klasifikasi yang diperoleh. Orde *legendre polynomial* yang digunakan adalah orde 2 dan orde 3. Berdasarkan hasil pengujian pada masing – masing skenario menunjukkan bahwa semakin besar orde yang digunakan, tidak menjadi jaminan akan meningkatkan kinerja klasifikasi dalam mengklasifikasikan data *microarray* seperti Colon, Lung, dan Ovarian. Kemudian peningkatan orde *legendre polynomial* juga justru malah akan meningkatkan input *space* sehingga membutuhkan waktu terhadap proses pelatihan yang lebih lama dan kinerja klasifikasi yang dihasilkan juga belum tentu jauh lebih baik.

Tabel 7 Hasil Akurasi Skenario 1 Berdasarkan Ukuran Populasi Dan Maksimum Generasi Pada Data Colon, Lung, Ovarian Menggunakan MRMR GA dan FLNN

Data	Max Generasi	Ukuran Populasi	Jumlah Fitur Terpilih	Akurasi	Rata - Rata Akurasi
Colon	5	50	950	72,22%	77,78%
	8		1016	88,89%	
	10		984	72,22%	
	5	100	1011	83,33%	87,03%
	8		993	83,33%	
	10		1003	94,44%	

Lung	5	50	6265	100,00%	98,15%
	8		6207	98,15%	
	10		6275	96,30%	
	5	100	6184	92,60%	97,53%
	8		6316	100,00%	
	10		6331	100,00%	
Ovarian	5	50	7628	97,33%	98,67%
	8		7644	98,67%	
	10		7481	100,00%	
	5	100	7625	96,00%	97,78%
	8		7581	100,00%	
	10		7657	97,33%	

Tabel 8 Hasil Akurasi Skenario 2 Berdasarkan Ukuran Populasi dan Maksimum Generasi Pada Data Colon, Lung, Ovarian Menggunakan MRMR GA dan FLNN

Data	Max Generasi	Ukuran Populasi	Jumlah Fitur Terpilih	Akurasi	Rata - Rata Akurasi
Colon	5	50	996	75,00%	80,55%
	8		980	83,33%	
	10		100	83,33%	
	5	100	1015	75,00%	80,56%
	8		1007	91,67%	
	10		969	75,00%	
Lung	5	50	6307	100,00%	99,07%
	8		6151	100,00%	
	10		6248	97,22%	
	5	100	6262	100,00%	100,00%
	8		6357	100,00%	
	10		6283	100,00%	
Ovarian	5	50	6256	98,00%	98,00%
	8		6256	98,00%	
	10		6256	98,00%	
	5	100	6256	98,00%	98,67%
	8		7618	100,00%	
	10		6256	98,00%	

Tabel 7 dan Tabel 8 menyajikan hasil pengujian dari pengaruh jumlah ukuran populasi dan maksimum generasi yang berbeda terhadap data uji. Ukuran populasi dan maksimum generasi merupakan parameter yang sangat penting dalam tugas akhir ini, sehingga ukuran populasi dan maksimum generasi yang telah ditentukan akan dapat dilihat pengaruhnya terhadap kinerja klasifikasi. Dari beberapa pengujian yang telah dilakukan, ukuran populasi memiliki dampak pada cakupan atau penelusuran awal fitur yang mana akan berpengaruh terhadap evaluasi *fitness* dalam memperoleh fitur terpilih. Terlihat pada tabel 6 dan tabel 7 dengan ukuran populasi 50 pada skenario 1 dan skenario 2 rata – rata kinerja klasifikasi yang dihasilkan cenderung lebih rendah jika dibandingkan dengan ukuran populasi 100. Sedangkan maksimum generasi berpengaruh terhadap proses pelatihan data yang mana akan berdampak pada optimalisasi nilai *fitnessnya*. Secara keseluruhan terlihat pada Tabel 6 dan Tabel 7 dengan maksimum generasi 5 dan 10, kinerja hasil klasifikasi yang dihasilkan terlihat cenderung lebih rendah jika dibandingkan dengan kinerja hasil klasifikasi pada maksimum generasi 8. Meskipun kinerja hasil klasifikasi berbeda pada skenario 1 dan skenario 2. Dapat dilihat pada skenario 1 dan skenario 2 dengan ukuran populasi 50 dengan maksimum generasi 5 dan 10 cenderung tidak lebih baik jika dibandingkan dengan ukuran populasi 100 dan maksimum generasi 8. Pengujian data Colon, Lung dan Ovarian pada skenario 1 dan skenario 2 menghasilkan kinerja hasil klasifikasi terbaik adalah dengan menggunakan ukuran populasi 100 dan maksimum generasi 8. Oleh karena itu dapat disimpulkan bahwa parameter MRMR GA dalam menyeleksi fitur yang paling sesuai digunakan

untuk pengujian data Colon, Lung, dan Ovarian adalah dengan menggunakan ukuran populasi 100 dan maksimum generasi 8.

5. Kesimpulan

Berdasarkan pengujian dan hasil analisis yang telah dilakukan pada tugas akhir ini maka diperoleh beberapa hasil dan kesimpulan sebagai berikut:

1. Dalam membangun sistem untuk klasifikasi kanker berdasarkan data *microarray*, pada tugas akhir ini membangun sistem dengan reduksi dimensi menggunakan MRMR GA dan *Functional Link Neural Network* (FLNN). Dengan cara membagi langkah menjadi 4 proses utama. Pertama melakukan *preprocessing* terhadap data *microarray* menggunakan metode min - max, selanjutnya melakukan reduksi dimensi menggunakan metode MRMR GA, kemudian melakukan klasifikasi menggunakan FLNN, dan yang terakhir melakukan perhitungan nilai akurasi dengan menggunakan *Confusion Matrix* terhadap proses pengklasifikasian yang telah dibangun.
2. Berdasarkan analisa hasil pengujian peranan reduksi dimensi pada klasifikasi data *microarray*, dapat disimpulkan bahwa MRMR GA sangat berpengaruh peranannya terhadap nilai akurasi pada saat proses pengklasifikasian data *microarray*. Dalam hal ini menunjukkan bahwa pengaruh dari MRMR GA dapat meningkatkan kinerja klasifikasi pada setiap datasetnya. Terlihat pada data Colon dapat meningkatkan nilai akurasi sebesar 5,55% - 16,66%, pada data Lung sebesar 2,78% - 5,56%, dan pada data Ovarian dapat meningkatkan nilai akurasi sebesar 2% - 2,67%.
3. Berdasarkan analisa hasil pengujian, *confusion matrix* dapat di implementasikan untuk menghitung nilai akurasi pada sistem klasifikasi yang telah dibangun menggunakan MRMR GA dan *Functional Link Neural Network* (FLNN). Jika dilihat berdasarkan skenario pengujian, kinerja klasifikasi terbaik diperoleh dari skenario 2 dengan proporsi data latih dan data uji sebesar 80:20. Namun pengaruh orde *legendre polynomial* justru meningkatkan input *space* sehingga membutuhkan waktu terhadap proses pelatihan yang lebih lama dan kinerja klasifikasi yang dihasilkan juga belum tentu jauh lebih baik. Kemudian dapat disimpulkan bahwa parameter MRMR GA dalam menyeleksi fitur yang paling sesuai digunakan untuk pengujian data Colon, Lung, dan Ovarian adalah dengan menggunakan ukuran populasi 100 dan maksimum generasi 8.

Daftar Pustaka

- [1] H. Aydadenta, "A Clustering Approach for Feature Selection in Microarray Data Classification using Random," vol. 14, no. 5, pp. 1167–1175, 2018.
- [2] S. R. Rusamsi, Adiwijaya, and A. A. Rohmawati, "Deteksi Kanker Berdasarkan Klasifikasi Microarray Data Menggunakan Wavelet Transform dan Backpropagation Termodifikasi dengan Conjugate Gradient Flechter Reeves," *E-Proceeding of Engeneering*, vol. 5, no. 1, pp. 1772–1779, 2018.
- [3] R. M. A. Egantara, "Deteksi Kanker Berdasarkan Klasifikasi Microarray Data Menggunakan Minimum Redudancy Maximum Relevance dan Backpropagation Termodifikasi dengan Conjugate Gradient Polak Ribiere," 2018.
- [4] P. T. Ramadhani, U. N. Wisesty, and A. Aditsania, "Deteksi Kanker Berdasarkan Klasifikasi Data Microarray Menggunakan Functional Link Neural Network dengan Seleksi Fitur Genetic Algorithm," vol. 2, pp. 11–22, 2017.
- [5] M. Kumar, S. Singh, and S. K. Rath, "Classification of Microarray Data using Functional Link Neural Network," *Procedia Comput. Sci.*, vol. 57, pp. 727–737, 2015.
- [6] F. A. Ma and U. N. Wisesty, "Analisis Pengaruh Metode Reduksi Dimensi Minimum Redundancy Maximum Relevance pada Klasifikasi Kanker Berdasarkan Data Microarray Menggunakan Classifier Support Vector Machine," vol. 5, no. 1, pp. 1499–1506, 2018.
- [7] U. N. Wisesty, E. Lisnawati, A. Aditsania, and D. S. Kusumo, "Dimensionality Reduction using Principal Component Analysis for Cancer Detection based on Microarray Data Classification," 2018.
- [8] H. Aydadenta and Adiwijaya, "On the Classification Techniques in Data Mining for Microarray Data Classification," *J. Phys. Conf. Ser.*, vol. 971, no. 1, pp. 0–9, 2018.
- [9] A. Andi Putri Hafisah, "Analisis Reduksi Dimensi pada Klasifikasi Microarray Menggunakan MBP Powell Beale," vol. 7, no. 1, pp. 17–24, 2018.
- [10] M. Sarmilah, Adiwijaya, and A. A. R., "Analisis Seleksi Fitur Genetic Algorithm dan Ekstraksi Fitur Wavelet pada Klasifikasi Microarray Data Menggunakan Naïve Bayes," *e-Proceeding Eng.*, vol. 5, no. 1, pp. 1715–1723, 2018.
- [11] D. M. Sahoo and S. Chakraverty, "Functional Link Neural Network Approach to Solve Structural System Identification Problems," *Neural Comput. Appl.*, vol. 30, no. 11, pp. 3327–3338, 2018.
- [12] A. Indriani, "Klasifikasi Data Forum dengan Menggunakan Metode Naive Bayes Classifier," *Semin. Nas. Apl. Teknol. Inf.*, vol. 20, no. ISSN: 1907-5022, pp. 5–10, 2014.
- [13] Riskesdes, 2013. https://kbr.id/nasional/01-2018/hari_kanker_sedunia_2018/94807.html (Accessed on October 30, 2018).
- [14] Jinyan Li, Kent-ridge bio-medical data set repository, School of Computer Engineering Nanyang Technological University, Singapore, Downloaded at April 2017 from URL: <http://leo.ugr.es/elvira/DBCRepository/>.