



# Deteksi Kanker Berdasarkan Klasifikasi Data Microarray Menggunakan Least Absolute Shrinkage and Selection Operator dan Functional Link Neural Network

Dinda Rahma Putri\*, Adiwijaya, Yuliant Sibaroni

Fakultas Informatika, Universitas Telkom, Bandung, Indonesia

Email: <sup>1</sup>dindarp@student.telkomuniversity.ac.id, <sup>2</sup>adiwijaya@telkomuniversity.ac.id, <sup>3</sup>yuliant@telkomuniversity.ac.id

Email Penulis Korespondensi: dindarp@student.telkomuniversity.ac.id

**Abstrak**—Kanker adalah penyakit berbahaya yang muncul dari perubahan sel normal menjadi sel tumor yang berkembang menjadi tumor ganas. Menurut WHO, kanker merupakan penyakit mematikan kedua di dunia. Sekitar 70% kematian akibat kanker terjadi di negara berpenghasilan rendah dan menengah seperti Indonesia. Kanker dapat dideteksi dengan mengenali pola ekspresi gen manusia. DNA Microarray adalah teknologi yang dapat mengetahui pola ekspresi gen di berbagai kondisi yang berbeda dengan cara klasifikasi data microarray. Data microarray memiliki dimensi yang sangat besar dan perlu direduksi guna mendapatkan gen yang informatif untuk mendeteksi kanker dengan optimal. Dalam penelitian ini, penulis menggunakan Least Absolute Shrinkage and Selection Operator (LASSO) sebagai metode seleksi fitur untuk mereduksi dimensi data dan Functional Link Neural Network (FLNN) sebagai metode klasifikasi dengan fungsi basis Legendre Polynomial. Dengan serangkaian proses yang telah dilakukan, didapatkan akurasi rata-rata sebesar 86.41% dan f1-score rata-rata sebesar 81.83%.

**Kata Kunci:** Deteksi Kanker, DNA Microarray, Reduksi Dimensi, Least Absolute Shrinkage and Selection Operator (LASSO), Functional Link Neural Network (FLNN).

**Abstract**—Cancer is a dangerous disease that arises from the conversion of normal cells into tumor cells that develop into malignant tumors. According to WHO, cancer is the second deadliest disease in the world. About 70% of cancer deaths occur in low and middle income countries such as Indonesia. Cancer can be detected by recognizing patterns of expression of human genes. DNA Microarray is a technology that can find patterns of gene expression in a variety of different conditions by means of microarray data classification. Microarray data has very large dimensions and needs to be reduced in order to obtain informative genes to detect cancer optimally. In this study, the authors use the Least Absolute Shrinkage and Selection Operator (LASSO) as a feature selection method to reduce data dimensions and Functional Link Neural Network (FLNN) as a classification method with Legendre Polynomial base functions. With a series of processes that have been carried out, obtained an average accuracy of 86.41% and an average f1-score of 81.83%.

**Keywords:** Cancer Detection, DNA Microarray, Dimension Reduction, Least Absolute Shrinkage and Selection Operator (LASSO), Functional Link Neural Network (FLNN).

## 1. PENDAHULUAN

Kanker adalah salah satu penyakit berbahaya yang dapat menyebabkan kematian. Kanker muncul dari perubahan sel normal menjadi sel tumor yang berkembang menjadi tumor ganas. Menurut WHO [1], kanker merupakan penyakit mematikan kedua di dunia. Sekitar 70% kematian akibat kanker terjadi di negara berpenghasilan rendah dan menengah [1]. Seperti halnya angka kejadian penyakit kanker di Indonesia yang berada di urutan 8 di Asia Tenggara [2]. Namun, peningkatan kasus kanker dapat dikurangi dengan deteksi kanker sejak dini dan menghindari faktor resiko utamanya seperti mengatur pola makan dengan baik dan menghindari konsumsi alkohol [1]. Dengan demikian, teknologi deteksi kanker sangat dibutuhkan agar upaya pencegahan maupun pengobatan kanker dapat dilakukan sesegera mungkin.

Terdapat suatu teknologi dalam bidang bioinformatika yang digunakan untuk mendeteksi kanker dalam beberapa tahun terakhir ini. DNA microarray adalah teknologi yang menggunakan data microarray untuk mengetahui pola ekspresi gen di berbagai kondisi yang berbeda [3]. Kanker dapat dideteksi dengan memahami pola ekspresi gen karena ekspresi gen dapat mendeskripsikan sel kanker yang mengalami abnormalitas yang bisa menyebabkan kanker tumbuh dan berkembang. Deteksi kanker melalui ekspresi gen jauh lebih terpercaya dan akurat dibandingkan deteksi kanker secara tradisional yang berdasarkan analisis kemunculan tumor [4]. Sehingga, teknologi ini dapat membantu para medis memahami pola ekspresi gen seseorang yang terkena kanker maupun tidak [5].

Dalam mendeteksi kanker menggunakan data microarray, terdapat masalah yang harus diatasi yaitu dimensi data yang sangat besar dimana jumlah gen lebih besar daripada jumlah sampel [6]. Solusinya yaitu dilakukan reduksi dimensi data karena dapat membantu dalam memaksimalkan data dengan mengetahui gen yang sangat berpengaruh untuk mendeteksi kanker [7]. Pada penelitian yang berjudul "Feature Selection Algorithm Based on Mutual Information and Lasso for Microarray Data" [8] telah dilakukan penelitian tentang klasifikasi data microarray menggunakan Mutual Information (MI), Least Absolute Shrinkage and Selection Operator (LASSO) dan Improved LASSO untuk metode reduksi dimensinya dan menggunakan empat metode klasifikasi yaitu Support Vector Machine (SVM), K- Nearest Neighbors KNN, C4.5 Decision Tree and Random Forest. Data yang digunakan yaitu Colon, Prostate, Lymphoma, Leukimia, dan Lung Cancer. Hasil dari penelitian tersebut diperoleh rata-rata akurasi yang menggunakan LASSO sebagai metode seleksi fiturnya adalah 93,562% dan yang



menggunakan MI diperoleh rata-rata akurasi sebesar 92,97%. LASSO telah banyak digunakan dalam berbagai penelitian untuk pemilihan fitur karena kinerjanya yang efisien [8, 9].

Berbagai penelitian mengenai klasifikasi microarray telah banyak dilakukan untuk membantu bidang medis dalam mendeteksi kanker dengan mengamati ekspresi genetik dari berbagai gen. Informasi penting dari gen dapat diekstraksi menggunakan teknik pembelajaran mesin [10]. Pada penelitian yang berjudul “Implementation of Minimum Redundancy Maximum Relevance (MRMR) and Genetic Algorithm (GA) for Microarray Data Classification with C4.5 Decision Tree.” [11] telah dilakukan klasifikasi data microarray menggunakan Minimum Redundancy Maximum Relevance (MRMR) dan Genetic Algorithm (GA) dengan metode klasifikasi C4.5 Decision Tree. Pada percobaan pertama menggunakan MRMR yang dikombinasikan dengan GA sebagai metode optimasi dan klasifikasi C4.5 menghasilkan rata-rata akurasi sebesar 79%. Pada percobaan kedua menggunakan GA sebagai metode seleksi fitur dan klasifikasi C4.5 menghasilkan rata-rata akurasi sebesar 78%. Pada penelitian yang berjudul “Implementation of Mutual Information and Bayes Theorem for Classification Microarray Data.” [12] telah dilakukan klasifikasi data microarray menggunakan Mutual Information (MI) dan Teorema Bayes. Hasil yang didapatkan yaitu F1-score tertinggi sebesar 91.06% menggunakan Bayesian Network dan 88.85% menggunakan Naive Bayes.

Pada penelitian yang berjudul “Deteksi Kanker berdasarkan Klasifikasi Data Microarray menggunakan Functional Link Neural Network dengan Seleksi Fitur Genetic Algorithm.” [13] telah dilakukan klasifikasi data microarray menggunakan Genetic Algorithm (GA) sebagai metode reduksi dimensi dan Functional Link Neural Network (FLNN) sebagai metode klasifikasinya. Pada penelitian tersebut dijelaskan bahwa FLNN merupakan salah satu model Artificial Neural Network (ANN) yang berhasil dalam melakukan klasifikasi karena memiliki arsitektur layer tunggal yang mampu mengoptimasi proses. Berbeda dengan metode ANN yang menggunakan multi layer, waktu komputasi FLNN dapat lebih cepat karena memiliki arsitektur layer tunggal. Pada klasifikasi FLNN, terdapat fungsi basis untuk mengekspansi vektor masukan. Pada pengujiannya, terdapat beberapa skenario pengujian dan pengujian terbaik dihasilkan dari skenario pertama dengan proporsi data latih berbanding data uji sebesar 70:30. Kemudian, diperoleh akurasi sebesar 92.3% dan 87.5% yang masing-masing dihasilkan dari kombinasi parameter ukuran populasi 100, dan Legendre Polynomial orde 3 dengan LR 0.6. Sedangkan pada penelitian yang berjudul “Cancer Detection based on Microarray Data Classification Using Principal Component Analysis and Functional Link Neural Network” [14] telah dilakukan analisis pengaruh PCA dan orde polinomial FLNN terhadap performansi sistem. Orde polinomial yang digunakan pada penelitian tersebut yaitu 1, 2, 3, dan 4. Hasil rata-rata akurasi yang didapatkan sebesar 76.08%.

Pada penelitian yang berjudul “Classification of Microarray Data Using Functional Link Neural Network.” [15] telah dilakukan klasifikasi DNA microarray menggunakan Functional Link Neural Network (FLNN) sebagai metode klasifikasi dan T-Test sebagai metode seleksi fiturnya. Mereka membandingkan semua fungsi basis yang ada pada metode FLNN tersebut yaitu Power Series Polynomial FLNN, Trigonometric FLNN, Chebyshev Polynomial FLNN, dan Legendre Polynomial FLNN. Mereka menyimpulkan bahwa basis Legendre Polynomial dengan akurasi 0.8557 menjadi basis yang terbaik daripada tiga basis lainnya.

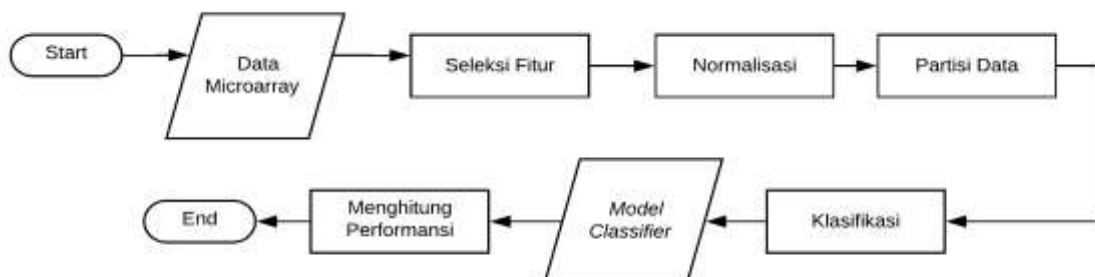
Dalam penelitian ini, FLNN dengan fungsi basis Legendre Polynomial digunakan pada klasifikasi dan LASSO pada reduksi dimensinya. Kemudian, performansi hasil klasifikasi dianalisis berdasarkan pengaruh dari metode seleksi fitur dan metode klasifikasi.

Pada bab pendahuluan ini penulis menjelaskan latar belakang, rumusan masalah, dan tujuan. Kemudian pada bab metode penelitian, penulis membahas penelitian terkait metode dan alur sistem yang digunakan. Selanjutnya pada bab hasil dan pembahasan, penulis membahas hasil dan analisis. Setelah itu, pada bab terakhir, penulis menjelaskan kesimpulan dari penelitian ini.

## 2. METODOLOGI PENELITIAN

### 2.1 Desain Sistem

Pada penelitian kali ini, skema sistem yang dibangun tersaji dalam diagram pada gambar 1.



**Gambar 1.** Desain Sistem



Proses pertama, data harus melewati tahap proses seleksi fitur menggunakan LASSO. Kemudian, data hasil dari seleksi fitur harus dinormalisasi. Selanjutnya, data yang telah dinormalisasi akan dipartisi menggunakan K-Fold menjadi dua jenis data yaitu data training dan data testing. Setelah itu, data training digunakan dalam pembuatan model klasifikasi. Terakhir, model yang dihasilkan dari proses tersebut akan digunakan untuk mengklasifikasi data testing dan menghitung performansi sistem.

## 2.2 Data Set

Klasifikasi dilakukan terhadap tiga data yaitu Colon Tumor, Breast Cancer, dan Lung Cancer yang didapat dari Kent-Ridge Bio-medical Data Set Repository [16].

**Tabel 1.** Spesifikasi Data Penelitian

Data	Jumlah Fitur	Jumlah Data	Jumlah Kelas	Kelas Positif	Kelas Negatif
Colon Tumor	2000	62	2	Positif	Negatif
Breast Cancer	24481	97	2	Relapse	Non-relapse
Lung Cancer	12533	181	2	Mesothelioma	ADCA

## 2.3 Praproses Data

Data yang digunakan memiliki nilai yang bersifat kontinu dan memiliki perbedaan jangkauan yang signifikan. Menurut penelitian yang berjudul "Evaluation of Normalization Methods for Microarray Data." [17], pada penggunaan data microarray, banyak variasi yang tidak diinginkan namun tetap teramati dalam ekspresi gen. Normalisasi merupakan proses penghilangan variasi data yang signifikan sehingga dalam penggunaan data microarray diperlukan normalisasi. Pada praproses ini dilakukan normalisasi data untuk mengubah rentang nilai data menjadi rentang 0 sampai 1. Teknik yang digunakan adalah Min-Max Normalization dalam persamaan 1 [3].

$$\text{Normalisasi}(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Dimana,  $x$  adalah nilai data yang belum dinormalisasi,  $x_{\min}$  adalah nilai terkecil dari semua data, dan  $x_{\max}$  adalah nilai terbesar dari semua data dalam atribut yang dinormalisasi.

## 2.4. LASSO

Least Absolute Shrinkage and Selection Operator (LASSO) adalah metode yang bekerja untuk regularisasi dan seleksi fitur. Pada parameter modelnya terdapat jumlah nilai absolut koefisien yang diberikan batasan agar tidak lebih dari nilai tetap. Metode ini menerapkan proses regularisasi atau penyusutan yang menyebabkan sebagian koefisien dari fitur mendekati nol atau menjadi nol. Fitur yang memiliki koefisien tidak nol akan dipilih untuk digunakan dalam klasifikasi [9].

Menurut penelitian yang berjudul "Predicting Patient Survival from Microarray Data by Accelerated Failure Time Modeling Using Partial Least Squares and LASSO." [18], LASSO menggunakan model linier regresi yang digambarkan pada persamaan 2.

$$y = \alpha + \sum_{i=1}^k \beta_i X_i \quad (2)$$

$$\sum_{i=1}^k |\beta_i| \leq t \quad (3)$$

Dimana,

$\alpha$  = parameter alpha

$\beta$  = vektor koefisien regresi

$x$  = variable bebas

$y$  = variable terikat

$t$  = parameter tuning yang mengontrol besarnya penyusutan

## 2.5 K-Fold

K-Fold adalah salah satu metode partisi data set menjadi data training dan data testing. Metode ini membagi data set sebanyak jumlah partisi K secara acak dan eksperimen dilakukan sebanyak K. Pada penelitian ini, setelah data direduksi dan dinormalisasi, akan dilakukan partisi data menjadi data training dan data testing. Nilai K yang digunakan adalah 5 yang artinya perbandingan data training dan testing sebesar 80:20.

## 2.6 FLNN

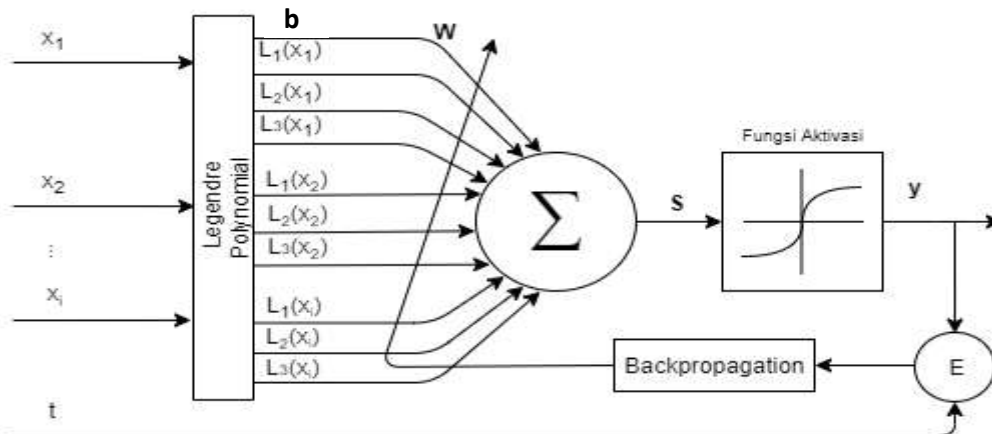
Selanjutnya, pada proses klasifikasi menggunakan metode FLNN yang merupakan algoritma jaringan syaraf tiruan dengan arsitektur layer tunggal dan hanya memiliki input layer dan output layer. Sehingga, metode ini memiliki kecepatan performansi lebih cepat daripada jaringan syaraf tiruan dengan arsitektur multi layer yang memiliki hidden layer [19]. Pada FLNN, vektor masukan diperluas ke dalam dimensi yang lebih tinggi berdasarkan fungsi basis yang ditentukan [13]. Terdapat beberapa jenis fungsi basis antara lain Power Series Polynomial,



Trigonometric, Chebyshev Polynomial, dan Legendre Polynomial. Dalam penelitian ini, basis yang digunakan adalah Legendre Polynomial yang didukung oleh penelitian yang berjudul “Classification of Microarray Data Using Functional Link Neural Network.” [15] yang telah membandingkan hasil setiap basis pada FLNN dan menyimpulkan bahwa FLNN dengan basis Legendre Polynomial menghasilkan hasil yang lebih baik bila dibandingkan dengan tiga basis lainnya.

Proses yang ada pada FLNN dalam penelitian ini yaitu memperluas vektor masukan berdasarkan fungsi basis Legendre Polynomial. Kemudian, dalam proses training, dimulai dengan inisiasi bobot secara acak hingga hasil dari nilai keluaran yang telah diaktivasi dilakukan evaluasi untuk mengetahui hasil prediksi. Proses tersebut akan terus berulang karena dalam alur kerja FLNN ini menggunakan teknik Back-Propagatioan guna meminimalisasi kesalahan melalui proses learning [19].

Adapun ilustrasi FLNN dengan Legendre Polynomial orde 3 pada gambar 2 [13].



**Gambar 2.** Skema Klasifikasi FLNN

Menurut penelitian yang berjudul “Funtional Link Neural Network Approach to Solve Structural System Identification Problems.”[20], rincian langkah-langkah klasifikasi menggunakan algoritma FLNN sebagai berikut.

1. Nilai input  $x$  diekspansi dengan fungsi basis Legendre Polynomial menggunakan persamaan 4.

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

Dimana,  $L$  adalah notasi untuk fungsi basis Legendre Polynomial dengan  $n+1$  adalah orde polynomial dan  $x$  adalah nilai masukan dari data asli. Contoh nilai masukan  $x = [x_1 x_2 x_3]^T$ , setelah diekspansi menggunakan Legendre Polynomial orde 2, maka nilai masukan menjadi  $x' = [1, L_1(x_1), L_2(x_1), L_1(x_2), L_2(x_2), L_1(x_3), L_2(x_3)]^T$  dimana  $L_0(x) = 1$  adalah nilai orde ke-0,  $L_1(x) = x$  adalah nilai orde ke-1 dan akan terus diekspansi ke orde-2 menjadi  $L_2(x) = \frac{1}{2}(3x^2 - 1)$ .

2. Selanjutnya, untuk menghitung jumlah linier  $S$  menggunakan persamaan 5 dengan menjumlahkan semua hasil kali bobot dengan nilai fungsi basis yang telah didapatkan dan ditambah dengan nilai bias.

$$S = \sum_{i=1}^N w_i L_i(x) + b \quad (5)$$

Dimana  $L(x)$  adalah nilai yang telah diekspansi dengan Legendre Polynomial,  $w$  = bobot yang diinisiasi secara acak,  $b$  adalah bias, dan  $N$  adalah jumlah data berdasarkan hasil ekspansi dengan fungsi basis.

3. Setelah itu, pada fungsi aktivasi menggunakan fungsi sigmoid  $f(S)$  dalam persamaan 6.

$$y_i = f(S_i) = \frac{1 - \exp(-S)}{1 + \exp(S)} \quad (6)$$

Dimana jumlah linier  $S_i$  sebagai nilai inputan ke- $i$  dalam fungsi tersebut dan  $y_i$  sebagai hasil dari fungsi sigmoid yang merupakan nilai prediksi.

4. Kemudian, dilakukan proses perhitungan *error* menggunakan fungsi *square error* dalam persamaan 7 guna mengevaluasi hasil klasifikasi. Perhitungan *error* ini akan digunakan untuk memperbaharui bobot dan meminimalisasi *error*.

$$E = \frac{1}{2} [t_i - y_i]^2 = \frac{1}{2} e_i^2 \quad (7)$$

Dimana,  $y_i$  adalah kelas prediksi dan  $t_i$  sebagai kelas target.



- Setelah proses pada langkah 4, bobot  $w$  akan diperbaharui menggunakan persamaan 8 dan 9 berdasarkan hasil evaluasi yang telah dilakukan menggunakan persamaan 7.

$$w_i(\text{baru}) = w_i(\text{lama}) + \Delta w_i \tag{8}$$

$$\Delta w_i = [-\eta \frac{\partial E}{\partial w_i}] \tag{9}$$

$$= [-\eta (t_i - y_i)(1 - y_i^2) L_i(x)]$$

Dimana,  $\Delta w_i$  adalah perubahan bobot,  $\eta$  adalah parameter *learning*,  $y_i$  adalah kelas prediksi dan  $t_i$  sebagai kelas target,  $L_i(x)$  adalah nilai Legendre Polynomial data  $x$ .

Tahap dua (2) hingga lima (5) akan terus berulang hingga mendapatkan nilai parameter bobot yang terbaik dengan melihat hasil evaluasi perhitungan error. Jika nilai error kurang dari nilai error terkecil yang ditentukan maka pembaharuan bobot akan berhenti. Setelah proses learning, tahap (1) sampai (3) akan dilakukan terhadap data testing. Nilai aktivasi pada tahap (3) akan diklasifikasi berdasarkan perbandingan dengan nilai threshold untuk dimasukkan ke dalam kelas kanker atau tidak kanker.

### 2.7 Skenario Pengujian

Pada penelitian ini dilakukan pengujian menggunakan data Colon Tumor, Breast Cancer, dan Lung Cancer dengan skenario yang sama. Skenario pengujian ini adalah mengklasifikasi semua data menggunakan FLNN dengan didahului reduksi dimensi menggunakan LASSO. Hasil yang akan dievaluasi adalah pengaruh dari parameter alpha pada LASSO terhadap performansi sistem dan pengaruh dari parameter orde pada FLNN terhadap performansi sistem. Parameter alpha yang digunakan adalah 0.1, 0.01, dan 0.001, sedangkan parameter orde yang digunakan adalah 1, 2, 3, dan 4.

### 2.8 Performansi

Pada penelitian ini, evaluasi dilakukan menggunakan table *confusion matrix* pada tabel 2.

**Tabel 2.** Confusion Matrix

Actual / Predicted	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Komponen yang digunakan dalam table confusion matrix yaitu True Positive, False Positive, True Negative, dan False Negative. Dimana, True Positive adalah total banyaknya hasil yang berhasil diprediksi kanker. False Positive adalah total banyaknya hasil yang tidak berhasil diprediksi tidak kanker. False Negative adalah total banyaknya hasil yang tidak berhasil diprediksi kanker. True Negative adalah jumlah banyaknya hasil yang berhasil diprediksi tidak kanker. Akurasi memperhitungkan hasil dari prediksi benar kanker dan prediksi benar tidak kanker. Sedangkan F1-Score memperhitungkan prediksi benar kanker terhadap semua kelas kanker.

Perhitungan akurasi dan f1-score dihitung dengan persamaan 10 dan 13 sebagai berikut [14].

$$\text{Akurasi} = \frac{(TN+TP)}{(TN+FN+TP+FP)} \tag{10}$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{11}$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \tag{12}$$

$$\text{F1-Score} = 2 * \frac{(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \tag{13}$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengaruh Alpha Terhadap Performansi

**Tabel 3.** Data Setelah Seleksi Fitur

Data	Jumlah Fitur	Alpha	Jumlah Fitur Setelah Seleksi Fitur
Colon Tumor	2000	0.1	133
		0.01	360
		0.001	949



		0.1	234
Lung Cancer	12533	0.01	503
		0.001	1315
		0.1	38
Breast Cancer	24481	0.01	89
		0.001	297

Pada tahap seleksi fitur, metode yang digunakan adalah Least Absolute Shrinkage and Selection Operator (LASSO). Metode tersebut memiliki parameter alpha untuk mendapatkan koefisien dari setiap fitur data. Terdapat beberapa parameter alpha yang digunakan yaitu 0.1, 0.01, dan 0.001. Dimensi data berkurang setelah data dilakukan seleksi fitur.

Tabel 3 merupakan hasil dari penggunaan LASSO untuk seleksi fitur yang menyebabkan dimensi data tereduksi sebelum masuk ke dalam proses klasifikasi. Jumlah fitur yang terpilih cukup beragam di setiap jenis data yang digunakan. Selain itu, dapat terlihat bahwa LASSO mampu mereduksi dimensi dengan cukup baik karena jumlah fitur yang tereduksi cukup banyak. Hal ini akan membantu waktu eksekusi pada proses klasifikasi menggunakan FLNN.

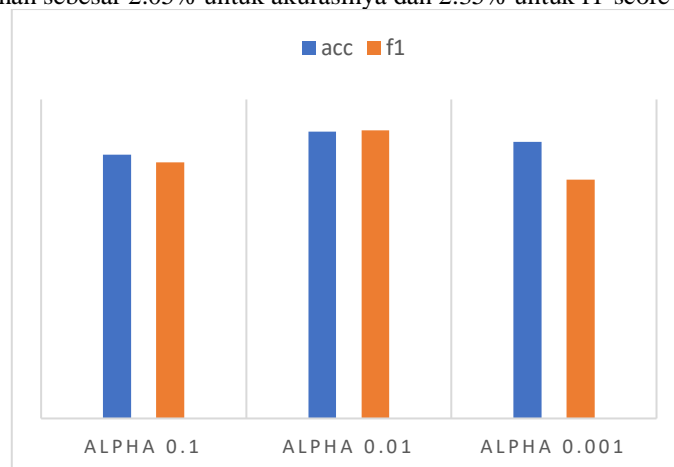
**Tabel 4.** Performansi tiap alpha terhadap semua orde polynomial

Alpha	Colon Tumor		Breast Cancer		Lung Cancer		Rata-rata	
	acc	f1	acc	f1	acc	f1	acc	f1
0.1	87	89.85	61.3	51.8	99.7	99.4	82.66667	80.35
0.01	86.665	89.47	83.7	82.7	99.4	98.8	89.92167	90.32333
0.001	86.765	89.6875	81.65	80.35	91.5	54.4	86.63833	74.8125

Tabel 4 menampilkan performansi percobaan yang menggunakan berbagai nilai alpha untuk setiap data microarray yang digunakan. Nilai parameter alpha akan berpengaruh pada jumlah fitur yang terpilih. Semakin kecil nilai alpha maka akan semakin besar jumlah fitur yang terpilih. Jumlah dari fitur terpilih akan terlihat pengaruhnya terhadap performansi sistem yang telah dibangun.

Berdasarkan hasil yang tertera pada table 4, hasil rata-rata tertinggi didapatkan dari percobaan yang menggunakan data Lung Cancer dengan parameter alpha sebesar 0.1 yang menghasilkan nilai akurasi sebesar 99.7 % dan nilai f1-score nya sebesar 99.4 %. Pada percobaan menggunakan data Colon Tumor menghasilkan nilai rata-rata f1-score yang tidak jauh berbeda di setiap nilai parameter alpha. Selain itu, pada percobaan menggunakan data Lung Cancer menghasilkan nilai rata-rata f1-score yang semakin menurun seiring dengan menurunnya nilai parameter alpha dan mengalami penurunan yang cukup besar dari nilai rata-rata f1-score sebesar 98.8% dengan parameter alpha 0.01 menurun sebesar 44.4% pada percobaan dengan parameter alpha 0.001. Hal ini menunjukkan bahwa kemampuan sistem dengan parameter alpha 0.001 pada data Lung Cancer sangat baik dalam mendeteksi kanker namun kurang baik dalam mendeteksi tidak kankernya.

Pada percobaan yang menggunakan data Breast Cancer menghasilkan pengaruh yang berbeda dari hasil kedua data lainnya. Hasil rata-rata percobaan yang menggunakan parameter alpha 0.1 terlihat rendah. Namun, pada percobaan dengan parameter alpha sebesar 0.01 mengalami kenaikan sebesar 22.4% untuk akurasinya dan 30.9% untuk f1-score nya. Berbeda dengan hasil rata-rata yang menggunakan parameter alpha 0.001 yang mengalami sedikit penurunan sebesar 2.05% untuk akurasinya dan 2.35% untuk f1-score nya.



**Gambar 3.** Rata rata dari seluruh data

Pada gambar 3 memperlihatkan rata-rata hasil dari seluruh data dengan nilai parameter alpha yang berbeda-beda. Berdasarkan gambar tersebut, hasil tertinggi yang didapatkan adalah percobaan dengan nilai parameter alpha 0.01 yaitu menghasilkan akurasi sebesar 89.92 % dan nilai f1-score sebesar 90.32%. Selain itu, hasil rata-rata dari



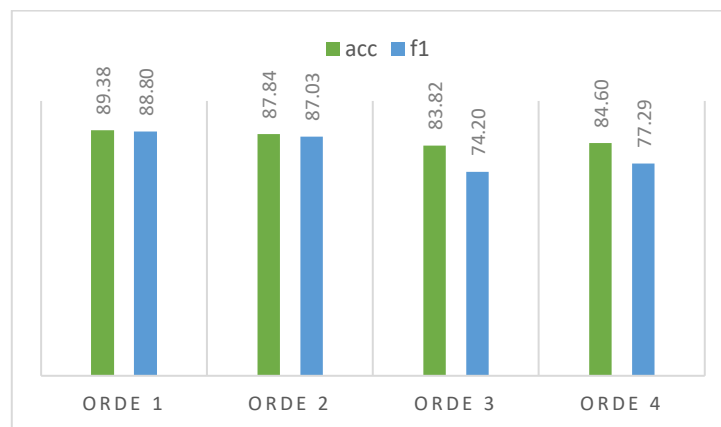
ketiga parameter alpha yang digunakan cukup konsisten karena memiliki selisih yang tidak terlalu besar satu sama lain. Sehingga, dari semua percobaan menggunakan beberapa variasi nilai parameter alpha didapatkan bahwa performansi tertinggi dihasilkan dari percobaan dengan menggunakan nilai parameter alpha sebesar 0.01. Hal tersebut menunjukkan bahwa pada penelitian ini, nilai alpha sebaiknya tidak terlalu kecil maupun tidak terlalu besar.

**3.2 Pengaruh Orde Polynomial Terhadap Performansi**

Pada klasifikasi FLNN, orde polinomial adalah sebuah parameter yang digunakan dalam proses ekspansi fitur data menggunakan fungsi basis. Pada penelitian ini, fungsi basis yang digunakan adalah legendre polinomial. Percobaan dilakukan menggunakan beberapa orde polinomial yang berbeda untuk menganalisis pengaruh orde polinomial yang digunakan terhadap performansi sistem. Parameter orde polinomial yang digunakan pada percobaan ini adalah 1, 2, 3, dan 4. Pada table 5 merupakan hasil percobaan yang didapatkan menggunakan tiga data yang berbeda.

**Tabel 5.** Performansi orde terhadap semua alpha

Orde	Colon Tumor		Breast Cancer		Lung Cancer		Rata-rata	
	Acc	f1	acc	f1	acc	f1	acc	f1
1	89.26667	91.4	79.26667	75.8	99.6	99.2	89.37778	88.8
2	86.64	89.34333	77.26667	72.53333	99.6	99.2	87.83556	87.02555
3	85.06667	88.46667	72.46667	68.26667	93.93333	65.86667	83.82222	74.2
4	86.26667	89.46667	73.2	69.86667	94.33333	72.53333	84.6	77.28889



**Gambar 4.** Grafik hasil dari keseluruhan data berdasarkan orde polinomial

Gambar 4 menampilkan hasil rata-rata dari keseluruhan data yang digunakan. Hasil tertinggi didapatkan dari percobaan yang menggunakan orde 1. Sedangkan, hasil terkecil didapatkan dari percobaan yang menggunakan orde 3. Selain itu, selisih dari hasil masing-masing orde tidak terlalu jauh. Sehingga, dapat disimpulkan bahwa peningkatan orde polinomial tidak menjamin meningkatnya performansi dan orde 1 adalah parameter orde polinomial yang menghasilkan performansi tertinggi di semua data yang digunakan pada penelitian ini.

Adapun waktu eksekusi selama melakukan percobaan yang dipaparkan pada table 6.

**Tabel 6.** Waktu Eksekusi

DATA	ORDE 1	ORDE 2	ORDE 3	ORDE 4
Colon Tumor	161.2304	181.9529	249.8385	261.3453
Breast Cancer	250.6202	311.4721	342.542	380.4112
Lung Cancer	392.1474	451.4993	481.4565	862.8089

Cara kerja FLNN adalah memperluas vektor masukan berdasarkan orde polinomialnya. Semakin tinggi orde polynomial maka akan semakin bertambah besar dimensi fitur yang akan diklasifikasi. Sehingga, terlihat pada tabel 6 bahwa waktu eksekusi akan semakin bertambah jika ordenya semakin besar. Rata-rata lama waktu eksekusi sistem adalah 360.6140 detik atau 6.010173 menit.

**4. KESIMPULAN**

Berdasarkan hasil dan pembahasan yang diperoleh, dapat disimpulkan bahwa parameter dari FLNN dan LASSO dapat mempengaruhi hasil performansi. Penggunaan parameter alpha pada seleksi fitur LASSO membantu dimensi data menjadi lebih kecil. Sehingga, dapat membantu dalam kecepatan waktu komputasinya. Performansi tertinggi yang didapatkan dari pengujian dengan perbandingan parameter alpha pada seleksi fitur adalah percobaan dengan



parameter alpha 0.01 yang menghasilkan akurasi sebesar 89.92% dan f1-score sebesar 90.32%. Sehingga, semakin besar atau semakin kecil parameter alpha tidak menjamin adanya peningkatan performansi karena dapat juga disebabkan oleh perbedaan jenis data. Pada klasifikasi FLNN, didapatkan performansi tertinggi dari percobaan dengan orde 1 yang menghasilkan akurasi sebesar 89.38% dan f1-score sebesar 88.80%. Sehingga, dapat disimpulkan bahwa peningkatan orde polinomial tidak menjamin adanya peningkatan performansi. Namun, hasil performansi yang dihasilkan cukup tinggi dan perbedaan parameter orde tidak membuat perbedaan performansi sistem yang signifikan. Selain itu, orde polinomial mempengaruhi waktu komputasi dimana semakin besar orde maka semakin lama waktu komputasinya.

## REFERENCES

- [1] World Health Organization, "Cancer," [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/cancer>. [Accessed: 20 Januari 2020].
- [2] Kementerian Kesehatan Republik Indonesia. "Hari Kanker Sedunia 2019". [Online]. Available: <https://www.depkes.go.id/article/view/19020100003/hari-kanker-sedunia-2019.html>. [Accessed: 24 Januari 2020].
- [3] Adiwijaya, 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. *Journal of Computer Science* 14(11), 2018.
- [4] A. Bharathi & A.M. Natarajan, "Cancer Classification of Bioinformatics data using ANOVA," in *International journal of computer theory and engineering*, 2, 369, 2010.
- [5] A. Adiwijaya, "Deteksi Kanker Berdasarkan Klasifikasi Microarray Data," *J. MEDIA Inform. BUDIDARMA*, 2018, doi: 10.30865/mib.v2i4.1043.
- [6] H. Aydadenta and Adiwijaya, "A clustering approach for feature selection in microarray data classification using random forest," *J. Inf. Process. Syst.*, 2018, doi: 10.3745/JIPS.04.0087.
- [7] M. S. Pratiwi, Adiwijaya, and A. Aditsania, "Cancer Detection Based on Microarray Data Classification using Genetic Bee Colony (GBC) and Conjugate Gradient Backpropagation with Modified Polak Ribiere (MBP-CGP)," in *2018 International Conference on Computer, Control, Informatics and its Applications: Recent Challenges in Machine Learning for Computing Applications, IC3INA 2018 - Proceeding*, 2019, doi: 10.1109/IC3INA.2018.8629538
- [8] W. Zhongxin, S. Gang, Z. Jing, Zhao, J, "Feature Selection Algorithm Based on Mutual Information and Lasso for Microarray Data," in *The Open Biotechnology Journal*, 10, 278-286, 2016.
- [9] V. Fonti, "Feature Selection using LASSO," in *Research Paper in Business Analytics*. 2017.
- [10] E. M. Mashhour, E.M.F.E. Houbay, K. Wassif, A. I. Salah, "Survey on different Methods for Classifying Gene Expression using Microarray Approach," in *International Journal of Computer Applications*, 2016.
- [11] I. Mabarti, "Implementation of Minimum Redundancy Maximum Relevance (MRMR) and Genetic Algorithm (GA) for Microarray Data Classification with C4. 5 Decision Tree," in *Journal of Data Science and Its Applications*, 3(1), 38-47. 2020.
- [12] M. D. Purbolaksono, K. C. Widiastuti, M. S. Mubarak, Adiwijaya, and F. A. Ma'ruf, "Implementation of mutual information and bayes theorem for classification microarray data," in *Journal of Physics: Conference Series*, 2018, doi: 10.1088/1742-6596/971/1/012011.
- [13] P. T. RAMADHANI, "Deteksi Kanker berdasarkan Klasifikasi Data Microarray menggunakan Functional Link Neural Network dengan Seleksi Fitur Genetic Algorithm," *Indones. J. Comput.*, 2017, doi: 10.21108/indojc.2017.2.2.173.
- [14] I. Priyono, A. Adiwijaya, A. Aditsania, "Cancer Detection based on Microarray Data Classification Using Principal Component Analysis and Functional Link Neural Network," in *Journal of Data Science and Its Applications*, 3(2). 2020.
- [15] M. Kumar, S. Singh, and S. K. Rath, "Classification of Microarray Data using Functional Link Neural Network," in *Procedia Computer Science*, 2015, doi: 10.1016/j.procs.2015.07.463.
- [16] Jinyan, Li, Kent-ridge bio-medical data set repository. 2013.
- [17] Park, T., Yi, S., Kang, S. et al. "Evaluation of normalization methods for microarray data," in *BMC Bioinformatics* 4, 33, 2003, doi: <https://doi.org/10.1186/1471-2105-4-33>.
- [18] S. Datta, J. Le-Rademacher, S. Datta, "Predicting Patient Survival from Microarray Data by Accelerated Failure Time Modeling Using Partial Least Squares and LASSO," in *Biometrics* 63, 259-271. 2007. DOI: 10.1111/j.1541-0420.2006.00660.x
- [19] Y. M. Mohamad Hassim and R. Ghazali, "Functional link neural network - Artificial bee colony for time series temperature prediction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, doi: 10.1007/9783-642-39637-3-34.
- [20] D. M. Sahoo and S. Chakraverty, "Functional link neural network approach to solve structural system identification problems," in *Neural Comput. Appl.*, 2018, doi: 10.1007/s00521-017-2907-x.