

# DETEKSI KANKER BERDASARKAN KLASIFIKASI MICROARRAY DATA MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS DAN BACKPROPAGATION TERMODIFIKASI DENGAN CONJUGATE GRADIENT POWELL-BEALE

Sugeng Hadi Wirasna<sup>1</sup>, Adiwijaya<sup>2</sup>, Danang Triantoro<sup>3</sup>

Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

<sup>1</sup>sugeng.hadi.wirasna@gmail.com

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

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

## Abstract

Cancer is the main cause in death. In the each year of cancer will continue to increase due to the unhealthiness of life-style patterns. In recent decades microarray usefull an important role in the diagnosis of cancer. Microarray is a technology that can store thousands of genes taken in some human cells as well. The microarray has a very large data dimensions therefore, to improve the accuracy of cancer diagnosis is then compared with traditional techniques it is done by means of reducing dimensions by using Principal Component Analysis (PCA) and Modified Back Propagation (MBP). MBP is a modification of Standard Backpropagation (BP) which apply the method Artificial Neural Network (ANN) to the Conjugate Gradient algorithm Powell-Beale to speed up the training process. In this final project has succeeded in proving that the modified backpropagation (MBP) and data reduction using Principal Component Analisis (PCA) show results quickly bees in the process of training. Results - average of the test using a modified backpropagation and PCA is the performance of each with techniques linesearch Charalambous Goldensection by 72.38% and amounted to 79.33%. This method is also good in terms of training time, given the average time - average required for linesearch Charalambous 2:30 seconds while Goldensection require 2:50 seconds.

Keyword : *cancer, microarray, principal component analysis (PCA), modified back propagation (mbp), conjugate gradient Powell-Beale.*

## Abstrak

Kanker merupakan penyebab utama dalam kematian. Dalam setiap tahun diperkirakan kanker akan terus meningkat karena tidak sehatnya gaya pola hidup. Pada beberapa dekade terakhir *microarray* berperan penting dalam diagnosis kanker. *Microarray* merupakan teknologi yang dapat menyimpan ribuan gen yang diambil dalam beberapa sel manusia sekaligus. *Microarray* memiliki dimensi data yang sangat besar oleh karena itu, untuk meningkatkan akurasi diagnosis kanker tersebut maka dibandingkan dengan teknik tradisional hal tersebut dilakukan dengan cara pengurangan dimensi dengan menggunakan Principal Component Analysis (PCA) dan Modified Back Propagation (MBP). MBP merupakan modifikasi dari Backpropagation Standart (BP)

yang menerapkan metode Jaringan Syaraf Tiruan (JST) dengan algoritma Conjugate Gradient Powell-Beale untuk mempercepat proses pelatihan. Pada tugas akhir ini telah berhasil membuktikan bahwa Modified Backpropagation (MBP) dan reduksi data menggunakan Principal Component Analysis (PCA) menunjukkan hasil lebih cepat dalam melakukan proses pelatihan. Hasil rata – rata dari pengujian menggunakan *backpropagation* termodifikasi dan PCA adalah performansi dari masing – masing dengan teknik *linesearch Charalambous* sebesar 72.38% dan *Goldensection* sebesar 79.33%. Metode ini juga bagus dalam hal waktu pelatihan, mengingat waktu rata – rata yang diperlukan 2.30 detik untuk *linesearch Charalambous* sedangkan *Goldensection* memerlukan 2.50 detik.

**Kata Kunci** : *kanker, microarray, principal component analysis (PCA), modified back propagation (mbp), conjugate gradient Powell-Beale.*

## I. PENDAHULUAN

Penyakit kanker merupakan salah satu penyebab kematian utama di seluruh dunia. Pada tahun 2012, kanker menjadi penyebab kematian 8,2 juta orang. Berdasarkan Data GLOBOCAN, International Agency for Research on Cancer (IARC) diketahui bahwa pada tahun 2012 terdapat 14.067.894 kasus baru kanker dan 8.201.575 kematian akibat kanker di seluruh dunia [1].

Kanker adalah sebutan untuk sebuah penyakit di mana saat sel-sel dalam tubuh membelah diri di luar kendali. Sel-sel abnormal ini kemudian menyerang jaringan terdekat, atau berpindah ke daerah yang jauh dengan cara masuk ke dalam pembuluh darah atau sistem limfatik. Masalah kanker di dunia semakin hari akan semakin meningkat seiring gaya hidup yang tidak sehat, hal ini juga yang banyak menyumbang angka kematian di dunia. Kematian oleh kanker bisa ditekan bila kanker tersebut dilakukan deteksi dini [2].

Deteksi dini kanker bisa menggunakan mikroarray sebagai mediana. Informasi yang terkandung di dalam rangkaian molekuler Deoxyribonucleic acid (DNA) makhluk hidup dapat diketahui melalui teknologi mikroarray. Mikroarray adalah teknologi yang mampu menyimpan ribuan ekspresi gen yang diambil dari beberapa jaringan tertentu manusia sekaligus, di dalamnya memiliki potensi yang sangat besar untuk pengetahuan baru, yang mendasari kemajuan dalam fungsional genomik dan biologi molekuler[3]. Pada pengolahan data mikroarray inilah menjadi hal yang menarik manakala didapatkan informasi penting dari data yang berukuran besar, dalam arti memiliki banyak sekali atribut bisa ratusan bahkan ribuan atribut.

Dalam penelitian ini penulis menggunakan metode Jaringan Syaraf Tiruan (JST) dengan algoritma Backpropagation (BP) yang digunakan sebagai proses learning. Modifikasi backpropagation yaitu dengan menambahkan conjugate gradient (CG), hal ini dilakukan karena untuk mempercepat pelatihan untuk Backpropagation sehingga dapat memperkecil kelemahan dari algoritma Backpropagation tersebut[6].

Kebanyakan peneliti biomedis mencari software yang tepat dan tidak hanya bisa mencapai akurasi prediksi yang tinggi tetapi juga mencakup desain yang user friendly untuk memudahkan penggunaanya. Selain itu, software tersebut sangat berguna jika source codenya tersedia. Diluar kedua hal tersebut, software harus up-to-date dengan informasi terkait untuk memastikan bahwa software tersebut kompetitif dengan perangkat lunak yang lain [4].

## II. TINJAUAN PUSTAKA

### 2.1 Mikroarray

Mikroarray Data merupakan jenis data yang dipakai dalam bioinformatika. Data ini berisi informasi gen karena itu jumlah featuranya sangat banyak misalnya banyaknya gen manusia dewasa ini diketahui jumlahnya sekitar 32 ribu.

Teknologi mikroarray data *gene expression* (DNA) dapat memonitor tingkat ekspresi ribuan gen secara bersamaan selama proses biologis dan seluruh koleksi sampel terkait. Pengetahuan yang didapat melalui analisis mikroarray data semakin penting karena mereka berguna untuk klasifikasi fenotip penyakit<sup>[3]</sup>. Pada

penelitian kali ini penulis akan menggunakan data mikroarray ekspresi gen untuk mengklasifikasikan sel kedalam golongan kanker dan bukan kanker.

Informasi yang dikumpulkan selama proses ini sangat signifikan bagi ahli biologi dan sejumlah metode analisis data yang telah digunakan untuk menganalisis data <sup>[8]</sup>.

2.2 *Klasifikasi*

Klasifikasi merupakan penyusunan bersistem dalam kelompok atau golongan menurut kaidah atau standar yang ditetapkan. Istilah klasifikasi ini untuk menunjukkan sebuah metode penyusunan data secara sistematis. Dalam penelitian ini konteks dari klasifikasi adalah golongan kanker-kanker berdasar data mikroarray yang sudah berlabel. Dalam bidang *Machine Learning*, masalah deteksi kanker dapat dilakukan dengan mengelompokkan data ke dalam kelas yang sudah ditetapkan, proses penentuan kelas data menggunakan metode seperti *Artificial Neural Network* <sup>[2]</sup>.

2.3 *Principal Component Analysis (PCA)*

*Principal Component Analysis (PCA)* merupakan teknik untuk membangun variable-variable baru yg merupakan kombinasi linear dari variable-variable asli. Jumlah maksimum dari variable - variable baru ini akan sama dengan jumlah dari variable lama, dan variable-variable baru ini tidak saling berkorelasi satu sama lain.

*Principal Component Analysis (PCA)* adalah suatu analisis yang menjelaskan struktur varian-kovarian dari suatu himpunan variabel yang melalui beberapa kombinasi linear dari variable – variabel tersebut (Johnson and Wichern, 2002).

Langkah pertama pada PCA adalah mencari data  $X^*$  dengan metode *zero-mean* yaitu dengan mengurangi semua nilai  $X_{i,j}$  pada matrik  $X$  dengan nilai rata – rata matrik tersebut (2.1)

$$X^* = X - \bar{X} \tag{2.1}$$

Langkah selanjutnya adalah dengan mencari kovarian dari matrik  $X^*$ , kovarian disimbolkan dengan  $C_x$ ,

$$C_x = \frac{1}{n-1} X^{*T} \cdot X^* \tag{2.2}$$

Dimana  $C_x$  adalah matriks kovariansi  $n \times n$ , dan  $m$  yaitu jumlah pengamatan. Pada matrik  $C_x$  hitung nilai eigen  $\lambda$  dan vector eigen  $v$ , dengan persamaan :

$$|C_x - \lambda I| = 0 \tag{2.3}$$

dan

$$(C_x - \lambda I)v = 0 \tag{2.4}$$

Vektor eigen yang didapatkan merupakan komponen utama untuk membentuk variable baru. Variable baru tersebut merupakan perkalian antara vector eigen ( $v$ ) terhadap matriks  $X^*$  Sedangkan persentase variansi yang dapat dijelaskan oleh variable baru tergantung kontribusi  $P_i$  dari masing masing nilai eigen, yang dihitung dengan persamaan :

$$P_i = \frac{\lambda_i}{\sum_{j=1}^m \lambda_j} \times 100\% \tag{2.5}$$

Sedangkan penentuan jumlah variable baru yang digunakan tergantung kepada persentase kontribusi kumulatif variansi  $P_i$ , dihitung dengan persamaan :

$$\sum_{j=1}^k \lambda_j = \sum_{j=1}^m \lambda_j \times 100\% \tag{2.6}$$

Setelah mendapatkan hasil persentase kumulatif varian, Maka dari hasil itu dapat menentukan hasil reduksi data berapa PC, dengan jumlah PC tersebut sudah dapat menjelaskan sebaran data sebesar nilai yang didapat pada PC tersebut.

#### 2.4 Jaringan Syaraf Tiruan

Jaringan saraf tiruan (JST) adalah salah satu algoritma pembelajaran mesin yang meniru cara kerja jaringan saraf makhluk hidup. JST terdiri dari sekelompok unit pemroses kecil yang disebut *neuron*. JST dapat digunakan untuk memodelkan hubungan yang kompleks antara *input* dan *output* untuk menemukan pola-pola pada data.

Pada referensi [10], ada dua jenis arsitektur yang sering dipakai dalam jaringan saraf tiruan, antara lain :

a. *Single Layer Network*

Arsitektur jst ini hanya terdiri dari *input* layer dengan node sumber yang terproyeksi ke *output* layer, tetapi tidak sebaliknya. Dengan kata lain, jaringan ini adalah jaringan jenis umpan maju (*feedforward*).

b. *Multi Layer Network*

Arsitektur layar tunggal dapat dikembangkan menjadi layar jamak dengan menambahkan satu atau lebih lapisan tersembunyi (*hidden layer*).

#### 2.5 Algoritma Pelatihan Jaringan Syaraf Tiruan

Salah satu dari sekian banyak algoritma pelatihan untuk jaringan saraf tiruan yang sangat populer adalah *Backpropagation*. Pada proses *learning*, *backpropagation* proses standar memiliki dua fase [7], yaitu :

Fase 1 : Propagasi

1. Propagasi Maju : Dengan bobot-bobot yang telah ditentukan pada inisialisasi awal, hitung keluaran dari *hidden layer*. Hasil keluaran dari *hidden layer* dipakai untuk mendapatkan keluaran *output layer*. Selanjutnya, keluaran jaringan ini dibandingkan dengan nilai target, selisih antara nilai target dengan keluaran jaringan adalah *error* atau MSE [7].
2. Propagasi Mundur : MSE yang diperoleh dipakai sebagai parameter dalam pelatihan. Pelatihan akan selesai jika MSE yang diperoleh sudah dapat diterima. Error tersebut dipropagasikan balik untuk memperbaiki bobot-bobot sinaptik dari semua neuron pada *hidden layer* dan *output layer* [10].

Fase 2 : Perbaharui Bobot

Setelah *neuron-neuron* mendapatkan nilai yang sesuai dengan kontribusinya pada error keluaran, bobot bobot jaringan diperbaiki agar *error* dapat diperkecil [7].

Bias (b) merupakan sebuah unit masukan pada JST yang berpengaruh untuk menaikkan atau menurunkan masukan pada fungsi aktivasi. Persamaan bias sebagai berikut:

$$v = W * P + b \tag{2.7}$$

Keterangan :

- v = merupakan variabel masukan pada fungsi aktivasi.
- W = Bobot
- P = matriks masukan
- b = bias

#### 2.6 Algoritma Conjugate Gradient Powell-Beale

Algoritma conjugate gradient (CG) merupakan algoritma pencarian yang arahnya berdasarkan pada arah conjugasi. Secara umum algoritma ini lebih cepat *convergen* daripada metode penurunan tercepat. CG menggunakan vektor tidak nol yang orthogonal dan bebas linier[6]. Dua vektor  $d_i$  dan  $d_j$  dikatakan orthogonal jika perkalian dalamnya bernilai nol [12]:

$$d_i \cdot d_j = 0 \tag{2.8}$$

Peran CG dalam algoritma backpropagation yaitu untuk memperkecil error yang terdapat pada bobot-bobot yang terhubung diantara neuron sehingga dapat dituliskan fungsi objektif untuk meminimalkan error, yaitu:

$$E = \frac{1}{2N} \sum_{n=1}^N \sum_{i=1}^n (t_i - y_i)^2 \tag{2.9}$$

Conjugate gradient merupakan metode untuk meminimasi fungsi turunan dengan menghitung pendekatan  $w_{k+1}$  secara iteratif berdasarkan :

$$w_{k+1} = w_k + \alpha_k d_k \tag{2.10}$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{2.11}$$

Keterangan :  $\alpha$  dan  $\beta$  = parameter momentum (untuk menghindari konvergensi lokal).

Sehingga algoritma pelatihannya menjadi sebagai berikut [12]:

1. Definisikan masalah, misalkan matriks masukan (P) dan matrik target (T).
2. Inisialisasi, menentukan arsitektur jaringan, nilai batas MSE dan batas jumlah epoch sebagai kondisi berhenti, *line search* yang digunakan, membangkitkan bobot-bobot (W) dan bias (b) dengan bilangan acak kecil.
3. Jika kondisi penghentian belum terpenuhi, lakukan langkah 3

• **Propagasi maju**

4. Hitung keluaran dari *hidden* layer  $z_j (j=1,2,3,\dots,p)$

$$z_{net_j} = v_{j0} + \sum_{i=1}^n x_i v_{ji} \tag{2.12}$$

$$z_j = f(z_{net_j}) = \frac{1}{1 + \exp^{-z_{net_j}}}$$

5. Lanjutkan dengan menghitung keluaran dari *output* layer  $y_k (k=1,2,\dots,m)$ ,

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} \tag{2.14}$$

$$y_k = \frac{1}{1 + \exp^{-y_{net_k}}} \tag{2.15}$$

• **Propagasi mundur**

6. Hitung faktor kesalahan di unit *output* berdasarkan perbedaan (error) nilai aktual dan nilai prediksi (*output* dari unit *output*)

$$\delta_k = (t_k - y_k) f'(y_{net_k}) = (t_k - y_k) y_k (1 - y_k) \tag{2.16}$$

7. Hitung faktor kesalahan di unit *hidden layer* berdasarkan faktor sebelumnya

$$\delta_{net_j} = \sum_{k=1}^m \delta_k W_{kj} \tag{2.17}$$

$$\delta_j = \delta_{net_j} f'(z_{net_j}) \delta_j = \delta_{net_j} z_j (1 - z_j) \tag{2.18}$$

8. Hitung gradien di unit *output* dari fungsi objektif yang sudah ditentukan

$$g_{k+1} = \frac{1}{N} \sum_{n=1}^N \delta_{nk} y_{nk} \tag{2.19}$$

9. Hitung gradien di unit *hidden layer*

$$g_{j+1} = \frac{1}{N} \sum_{n=1}^N \delta_{nj} z_{nj} \tag{2.20}$$

10. Hitung parameter  $\beta$  untuk semua *neuron* di unit *hidden layer* dan unit *output*. Parameter  $\beta$  untuk conjugate gradient Powell-Beale dapat dihitung dengan persamaan:

$$\beta_{i+1} = \frac{g_i^T (g_i - \beta_i g_i)}{g_i^T g_i} \tag{2.21}$$



di mana :

$\beta_{k+1}$  = nilai parameter  $\beta$  pada iterasi saat ini  
 $g_{(k+1)}$  = gradien pada iterasi saat ini  
 $g_k$  = gradient pada iterasi sebelumnya

11. Hitung *direction* untuk semua *neuron* di unit *hidden layer* dan unit output.

$$d_{t+1} = -g_{t+1} + \beta_t d_t \quad (2.22)$$

dimana :  $d_{t+1}$  = *direction* pada iterasi saat ini  
 $g_{t+1}$  = gradient pada iterasi saat ini  
 $\beta_t$  = nilai parameter  $\beta$  pada iterasi sebelumnya  
 $d_t$  = *direction* pada iterasi sebelumnya

Untuk *direction* awal:

$$d_t = -g_t \quad (2.23)$$

12. Hitung parameter  $\alpha$  untuk semua *neuron* di unit *hidden layer* dan unit *output* dengan teknik *line search*.

• **Perubahan bobot**

13. *Update* bobot dilakukan dengan cara sebagai berikut:

$$w_{t+1} = w_t + \alpha_{t+1} d_{t+1} \quad (2.24)$$

dimana :  $w_{t+1}$  = bobot yang akan diupdate  
 $w_t$  = bobot sebelumnya  
 $\alpha_{t+1}$  = nilai alfa saat ini  
 $d_{t+1}$  = *direction* pada iterasi saat ini

Parameter  $\alpha$  disini dapat dicari dengan menggunakan teknik *line search* sehingga dapat meminimumkan kinerja selama arah pencarian, karena parameter  $\alpha$  merepresentasikan besar langkah yang diambil untuk setiap *direction*. Di sini penulis akan mencoba untuk menerapkan beberapa teknik *line search*, antara lain adalah :

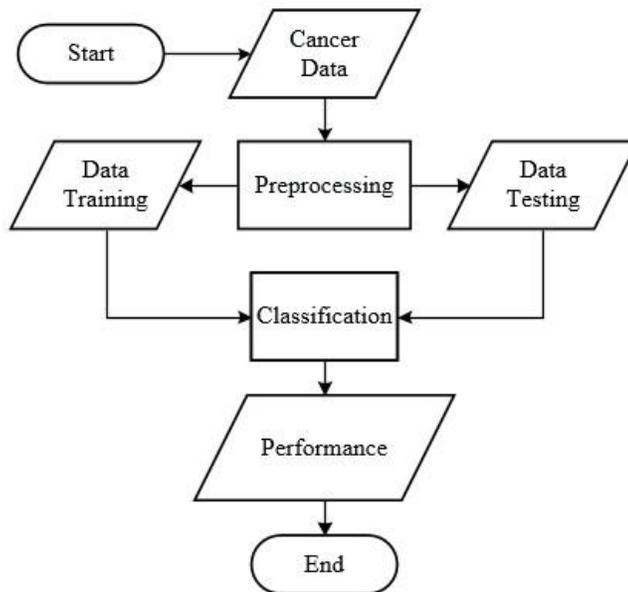
1. *Golden section search*, yakni pencarian linier sederhana tanpa harus menghitung gradient dari garis tersebut. Misalkan fungsi kinerja untuk suatu iterasi adalah  $fx$ . Proses pencarian nilai minimum dimulai dari menempatkan suatu interval (*delta*), antara  $a$  ( $= 0$ ) dan  $b$  ( $= \text{delta}$ ). Kemudian dicari fungsi kinerja untuk  $a$  ( $fa$ ) dan  $b$  ( $fb$ ). Apabila  $fb < fa$ , maka pelebaran interval dilakukan lagi sebesar  $\text{delta} = 2 * \text{delta}$ . Iterasi tersebut dilakukan hingga  $b \geq b_{\text{max}}$ , atau  $fb \geq fa$ . Tetapi, jika  $fb \geq fa$ , berarti nilai minimum terdapat pada interval  $[a, b]$  tersebut. Letakkan 2 (dua) nilai pada interval  $[a, b]$ , misalkan  $c$  dan  $d$ . Hitung fungsi kinerja  $c$  sebagai  $fc$ , dan  $d$  sebagai  $fd$ . Kedua nilai ini akan menentukan daerah mana yang dapat dibuang, dan suatu titik yang terletak di dalam interval baru. Iterasikan proses tersebut hingga lebar interval kurang dari suatu toleransi minimum.
2. *Charalambous' Search*, yakni metode pencarian yang merupakan kombinasi antara interpolasi kubik dan suatu tipe *sectioning*. Metode ini dipergunakan sebagai default dari algoritma pelatihan dengan *conjugate gradient*.

### III. METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum

Dalam pengertian secara umum bahwa klasifikasi ialah suatu kegiatan yang mengelompokkan benda yang memiliki beberapa ciri yang sama dan memisahkan benda yang tidak sama. Adapun tujuan klasifikasi antara lain adalah untuk mempermudah mengenali, membandingkan, dan mempelajari. Membandingkan berarti mencari persamaan dan perbedaan sifat atau ciri pada suatu benda. Pada

umumnya, kanker dirujuk berdasarkan tempat terjadinya. Misalnya kanker pada sel basal kulit dirujuk sebagai karsinoma sel basal.



Gambar 3.1 Diagram Alir Klasifikasi secara umum

Gambar di 3.1 atas adalah diagram alir klasifikasi kanker atau bukan kanker berdasar data mikroarray. Dimana proses pertama yang harus dilakukan adalah menyiapkan data mikroarray, dikarenakan data memiliki dimensi yang sangat besar, maka akan dilakukan preprosesing. Setelah data dipreprosesing, maka data siap digunakan untuk dilakukan klasifikasi, namun sebelum masuk pada klasifikasi, data harus dibagi menjadi dua bagian yaitu *data training* dan *data testing*. Pada jaringan syaraf tiruan memerlukan data training sebagai proses belajar algoritma untuk bisa menentukan apakah data lain yang akan dimasukkan merupakan kanker atau bukan kanker. Setelah jaringan syaraf tiruan melakukan proses pembelajaran, maka data testing bisa diproses sebagai inputan. Data *testing* akan diproses oleh jaringan syaraf tiruan untuk menghasilkan label, hasil label ini berupa kelas kanker dan bukan kanker. Dalam tahap akhir sistem akan mencatat performansi proses pembelajaran (*learning*) dan percobaan (*testing*).

### 3.2 Data

Data yang digunakan adalah semua data kanker maka setiap data memiliki dimensi yang berbeda (*record* dan atribut), berikut adalah spesifikasi data kanker.

Tabel 3.1 Spesifikasi Data

Data	Jumlah Kelas	Jumlah Record	Jumlah Atribut
Breast Cancer	2	97 ( 46 Relapse vs 51 non-relapse )	24482
Central Nervous	2	60 ( 21 cancer vs 39 normal )	7130
Colon Tumor	2	62 ( 22 cancer vs 40 normal )	2001
Lung Cancer	2	181 (31 Mesothelioma vs 150 ADCA)	12534
MLL Leukimia	3	72 (24 ALL vs 20 MLL vs 28 AML)	12583
Ovarian Cancer	2	253 (162 cancer vs 91 Normal)	15155
Prostate Cancer	2	136 (77 tumor vs 59 normal)	12601

Pada proses pengujian, terdapat data *training* dan data *testing* dengan porposi pembagian sebesar 80% untuk data *training* dan 20% untuk data *testing*, masing – masing diambil dari jumlah *record* data asli. Pembagian data mengacu pada *default* data yang didapat dari *Kent-Ridge Biomedical Data Repository*.

#### IV. HASIL DAN DISKUSI

##### 4.1 Analisis Perbandingan Backpropagation Standar dengan Backpropagation Termodifikasi

Proses pengujian menggunakan beberapa skenario klasifikasi diantaranya yaitu PCA dan *Modified Backpropagation* (MBP), MPB (Tanpa PCA), dan BP (*Backpropagation Standart*) untuk membandingkan hasilnya dengan hasil sistem yang diusulkan. Setiap skema klasifikasi dilakukan percobaan masing- masing sebanyak 5 kali percobaan untuk membangun model agar mendapatkan hasil arsitektur terbaik kemudian dibandingkan dengan hasil akurasi.

Pada proses pengujian, proporsi data yang digunakan sebesar 80% untuk data *Training* dan 20% untuk data *Testing*. pembagian data mengacu pada *default* data yang didapat dari *Kent Ridge Biomedical Data Repository*.

Tabel 4.0 Hasil Pengujian Pengaruh Hidden Neuron

Skema Klasifikasi	Neuron Hidden Layer	Akurasi (%)				Rata - rata
		Ovarium	Central Nervouse	Colon Tumor	Prostate	
MBP (Charalambous)	2	92.16	41.67	58.33	73.53	66.42
	4	98.04	83.33	58.33	76.47	79.04
	6	62.75	50.00	66.67	47.06	56.62
MBP (Golden Section)	2	92.16	41.67	41.67	73.52	62.25
	4	98.04	83.33	83.33	88.23	88.23
	6	60.78	75.00	58.33	29.41	55.88

Berdasarkan Tabel 4.0 didapatkan jumlah *neuron* terbaik pada *hidden layer* adalah sebanyak 4 dengan akurasi rata – rata sebesar 79.04% untuk teknik *linesearch Charalambous* dan 88.23% untuk teknik *linesearch Goldensection*. Akurasi tetinggi dianggap sebagai arsitektur terbaik yang selanjutnya digunakan kembali untuk diuji pada setiap data Kanker.

Table 4.1 Hasil testing

Data	Sistem	Akurasi (%)	Training Time (s)
Breast Cancer	PCA + MBP (Charalambous)	57.89	2.32
	PCA + MBP (Golden Section)	63.16	2.34
	PCA + BP	63.16	2.66
	MBP (Charalambous)	84.21	9.03
	MBP (Golden Section)	73.68	5.92
	BP	63.16	14.83
Central Nervous	PCA + MBP (Charalambous)	83.33	2.33
	PCA + MBP (Golden Section)	83.33	2.37
	PCA + BP	83.33	2.76
	MBP (Charalambous)	91.68	2.57
	MBP (Golden Section)	75.00	2.49
	BP	91.68	8.70
Colon Tumor	PCA + MBP (Charalambous)	58.33	2.15
	PCA + MBP (Golden Section)	66.66	2.45
	PCA + BP	58.33	3.50

	MBP (Charalambous)	75.00	2.51
	MBP (Golden Section)	75.00	2.50
	BP	75.00	6.14
Lung Cancer	PCA + MBP (Charalambous)	85.91	2.25
	PCA + MBP (Golden Section)	90.60	2.43
	PCA + BP	89.93	2.49
	MBP (Charalambous)	87.25	3.36
	MBP (Golden Section)	75.84	2.95
	BP	83.89	5.89
MLL Leukimia	PCA + MBP (Charalambous)	46.67	2.42
	PCA + MBP (Golden Section)	46.67	2.48
	PCA + BP	60.00	4.39
	MBP (Charalambous)	86.67	4.36
	MBP (Golden Section)	86.67	3.42
	BP	86.67	4.69
Ovarian Cancer	PCA + MBP (Charalambous)	98.04	2.35
	PCA + MBP (Golden Section)	100.00	2.42
	PCA + BP	98.04	3.08
	MBP (Charalambous)	100.00	8.75
	MBP (Golden Section)	98.04	6.73
	BP	100.00	62.11
Prostate Cancer	PCA + MBP (Charalambous)	76.47	2.28
	PCA + MBP (Golden Section)	88.24	2.88
	PCA + BP	82.35	9.92
	MBP (Charalambous)	97.06	5.00
	MBP (Golden Section)	97.06	4.85
	BP	76.47	10.18

## V. KESIMPULAN

Pengambilan proporsi data mempengaruhi klasifikasi pelatihan maupun pengujian. Semakin banyak proporsi data yang diambil maka menghasilkan performansi arsitektur berupa akurasi yang baik. Hal ini terbukti pada pengujian data kanker Ovarium di mana data ini memiliki record yang paling tinggi dan hasil pengujian menghasilkan akurasi yang tinggi pula. Metode *backpropagation* termodifikasi dan reduksi menggunakan PCA dapat diimplementasikan dengan baik untuk deteksi kanker berdasarkan data mikroarray. Hasil rata – rata dari pengujian menggunakan *backpropagation* termodifikasi dan PCA adalah performansi dari masing – masing dengan teknik *linesearch Charalambous* sebesar 72.38% dan *Goldensection* sebesar 79.33%. Metode ini juga bagus dalam hal waktu pelatihan, mengingat waktu rata – rata yang diperlukan 2.30 detik untuk *linesearch Charalambous* sedangkan *Goldensection* memerlukan 2.50 detik. Metode *backpropagation* termodifikasi dan PCA memberikan hasil waktu pelatihan yang bagus dalam klasifikasi deteksi kanker, melihat perbandingan dari semua sistem. Namun dalam beberapa kasus jenis data metode MBP memberikan hasil yang lebih baik. Mengacu pada hal tersebut, bahwa MBP bisa mengurangi waktu pelatihan pada jaringan saraf tiruan.

## DAFTAR PUSTAKA

- [1] Oscar P, dkk. 2015. *Situasi Penyakit Kanker*. Jakarta: Pusat Data dan Informasi
- [2] Nurfalalah, A. Adiwijaya. and Suryani, A., 2016. *Cancer Detection Based On Microarray Data Classification Using PCA And Modified Back Propagation*. Far East Journal of Electronics and Communications, 16(2), p.269.
- [3] Vanitha, C.D.A., Devaraj, D. and Venkatesulu, M., 2015. *Gene Expression Data Classification Using Support Vector Machine and Mutual Information-based Gene Selection*. *Procedia Computer Science*, 47, pp.13-21.
- [4] Siang TC, dkk. 2015. *A Review of Cancer Classification Software for Gene Expression Data*. International Journal of Bio-Science and Bio-Technology. Vol.7, No.4 (2015), pp.89-108. <http://dx.doi.org/10.14257/ijbsbt.2015.7.4.10> (diakses tanggal 05 Februari 2016).
- [5] Kochenberg. G., Glover F., Alidaee. B. and Wang, H.. 2005. *Clustering of Mikroarray data via Clique Partitioning*. Journal of Combinatorial Optimization, 10, 77–92.
- [6] Adiwijaya, U.N., Wisesty, A., Yunanto. and , D.M, Tampubolon. 2013. *Teknik Line Search Pada Conjugate Gradient Dalam Proses Pelatihan Backpropagation Untuk Peramalan Temperatur Dan Kelembaban Udara*. Makalah Nomor: KNSI-192
- [7] Suyanto, 2008, "Soft Computing: Membangun Mesin Ber-IQ Tinggi", Bandung, Informatika.
- [8] Hussain, Syed Fawad, and Muhammad Ramazan. "Biclustering of human cancer mikroarray data using co-similarity based co-clustering." *Expert Systems with Applications* 55 (2016): 520-531.
- [9] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), 264-323.
- [10] Shewchuk, Jonathan Richard, "An Introduction to the Conjugate Gradient Method Without the Agonizing Pain", Pittsburgh, School of Computer Science Carnegie Mellon University, 1994.
- [11] Suyanto. 2007. "Artificial Intelligence Searching, Reasoning, Planing and Learning". Bandung: Informatika.
- [12] Adiwijaya, U. N. Wisesty and F. Nhita, *Study of Line Search Techniques on the Modified Back propagation for Forecasting of Weather Data in Indonesia*, FarEast J. Math. Sci. (FJMS) 86(2) (2014), 139-148.
- [13] Kusumadewi, Sri.,2004. "Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Excel Link", Yogyakarta : Graha Ilmu.
- [14] Susetyoko, R. and Purwantini, E., Teknik Reduksi Dimensi Menggunakan Komponen Utama Data Partisi Pada Pengklasifikasian Data Berdimensi Tinggi dengan Ukuran Sampel Kecil. *Dimensi, 1*, p.500.
- [15] Adiwijaya, U.N. Wisesty, Wirayuda, T.A.B., Baizal, Z.K.A. and Haryoko, U., 2013. *An Improvement Of Backpropagation Performance By Using Conjugate Gradient On Forecasting Of Air Temperature And Humidity In Indonesia*. *Far East Journal of Mathematical Sciences (FJMS)*, (Part I), pp.57-67.
- [16] Adiwijaya, 2014, "Aplikasi Matriks dan Ruang Vektor", Yogyakarta: Graha Ilmu.