

Penerapan Metode RBPNN untuk Klasifikasi Kanker Payudara

Fairudz Shahura¹, Oni Soesanto², Fatma Indriani³

^{1,2,3}Prodi Ilmu Komputer FMIPA ULM

Jl. A. Yani Km 36 Banjarbaru, Kalimantan selatan

¹Email: irashahura@gmail.com

Abstract

Breast cancer is the most commonly diagnosed cancer in women. Breast cancer cases are increasing each year. Therefore, early detection of breast cancer plays an important role in anticipating the spread of cancer. Fine-needle aspiration (FNA) biopsy is one way to detect breast cancer. FNA is a method of taking the majority of tissue with a syringe that is intended to aid in the diagnosis of various tumor diseases. The FNA samples that have been studied generate ten characteristics, namely radius, texture, perimeter, area, compactness, smoothness, concavity, concave points, symmetry, and fractal dimension. These characteristics are used to classify benign and malignant breast cancer. To classify breast cancer, Radial Basis Probabilistic Neural Network (RBPNN) required. This study aims to determine how the performance of the method of Radial Basis Probabilistic Neural Network for classifying breast cancer. The accuracy was found to be equal 93.19% for training data, and 90.35% for testing data.

Keywords: Radial Basis Probabilistic Neural Network, Classification, Breast Cancer.

Abstrak

Kanker payudara merupakan penyakit yang paling banyak menyerang kaum wanita. Penderita penyakit kanker payudara semakin meningkat pada tiap tahunnya. Oleh karena itu deteksi dini kanker payudara memegang peranan penting dalam mengantisipasi penyebaran kanker. Salah satu cara untuk mendeteksi kanker payudara adalah dengan fine-needle aspiration (FNA) biopsy. FNA merupakan suatu metode pengambilan sebagian jaringan tubuh manusia dengan jarum suntik yang bertujuan untuk membantu diagnosis berbagai penyakit tumor. Sampel FNA yang telah diteliti menghasilkan sepuluh karakteristik, yaitu radius, texture, perimeter, area, compactness, smoothness, concavity, concave points, symmetry, dan fractal dimension. Kesepuluh karakteristik tersebut digunakan untuk mengklasifikasikan kanker payudara jinak dan ganas. Untuk mengklasifikasi tingkat keganasan dari kanker payudara dapat dilakukan dengan metode Radial Basis Probabilistic Neural Network (RBPNN). Penelitian ini bertujuan untuk mengetahui bagaimana performansi metode Radial Basis Probabilistic Neural Network untuk mengklasifikasikan kanker payudara. Dari hasil penelitian didapat akurasi 93.19% untuk data training, serta 90.35% untuk data testing.

Kata kunci : Radial Basis Probabilistic Neural Network, Klasifikasi, Breast Cancer.

1. PENDAHULUAN

Kanker payudara adalah jenis kanker yang terdapat pada jaringan payudara. Kanker payudara merupakan penyakit yang paling banyak menyerang kaum wanita. Di Indonesia kanker payudara menempati urutan pertama pada pasien rawat inap di seluruh rumah sakit [1]. Banyaknya penderita penyakit kanker payudara ini semakin meningkat pada tiap tahunnya hingga mencapai 8.328 penderita pada tahun 2006. Pada tahun 2007 banyaknya penderita tidak jauh berbeda meski sedikit mengalami penurunan yaitu, 8.277 kasus [1].

Deteksi dini kanker payudara memegang peranan penting dalam mengantisipasi penyebaran kanker. Dengan deteksi dini, timbulnya sel-sel kanker dapat segera diatasi dan dicegah penyebarannya. Pemeriksaan payudara berguna untuk memastikan bahwa payudara seseorang masih normal. Bila ada kelainan seperti infeksi, tumor, atau kanker dapat ditemukan lebih awal. Kanker payudara dapat dideteksi salah satunya melalui diagnosis definitif dari massa payudara yang hanya dapat dibentuk melalui fine-needle aspiration (FNA) biopsy, core needle biopsy, atau excisional biopsy [2].

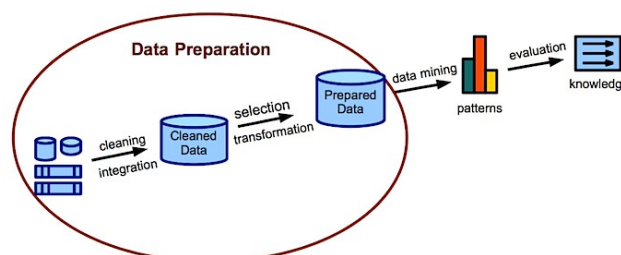
FNA adalah cara termudah dan tercepat untuk mendapatkan biopsi payudara, dan efektif bagi wanita yang memiliki kista berisi cairan. FNA menggunakan jarum yang lebih kecil dari yang digunakan untuk tes darah untuk mengeluarkan cairan, sel, dan fragmen kecil jaringan untuk diteliti di bawah mikroskop [3]. Setelah sampel FNA diambil ditetapkan sepuluh karakteristik, yaitu *radius*, *perimeter*, *area*, *compactness*, *smoothness*, *concavity*, *concave points*, *symmetry*, *fractal dimension*, dan *texture* [4]. Sepuluh karakteristik tersebut merupakan isi dari dataset *Wisconsin Diagnosis Breast Cancer* (WDBC) yang akan digunakan dalam penelitian ini untuk mengklasifikasikan kanker payudara jinak dan ganas.

Untuk mengklasifikasi tingkat keganasan dapat dilakukan dengan metode *Radial Basis Probabilistic Neural Network* (RBPNN) yang merupakan salah satu model dari Jaringan Syaraf Tiruan yang akan diterapkan pada sistem yang akan dibuat nantinya.

2. METODOLOGI PENELITIAN

2.1 Metode Pengembangan Sistem

Penelitian ini menggunakan metode Data Mining untuk pengembangan sistem.

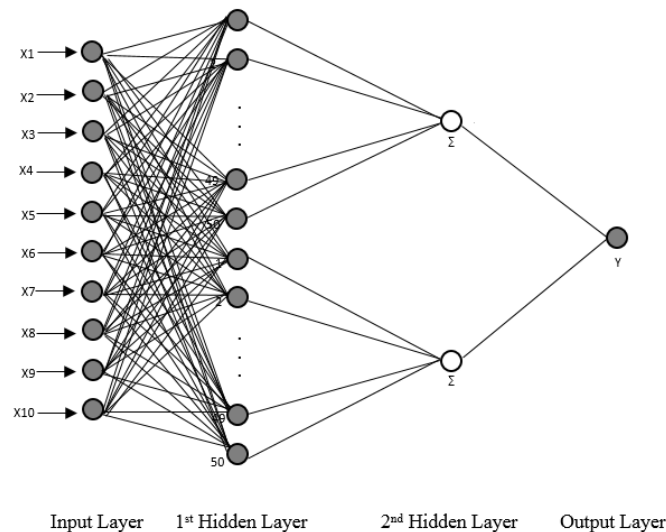


Gambar 1. Model Data Mining

Adapun tahapan-tahapan dalam model pengembangan *Data Mining* yaitu analisis kebutuhan, *data selection* dan *addition*, *preprocessing* dan *cleaning data*, *data mining (coding dan testing)*, *evaluasi*.

2.2 Analisis Kebutuhan

Model jaringan syaraf tiruan yang digunakan adalah *Radial Basis Probabilistic Neural Network (RBPNN)*. Berikut gambaran arsitektur jaringan RBPNN pada proses klasifikasi breast cancer untuk sistem yang dibuat, input layer berisi sepuluh atribut, yaitu *radius*, *texture*, *area*, *perimeter*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, dan *fractal dimension*. *Hidden layer* pertama berisi nilai *center* yang diambil secara acak, *hidden layer* kedua berisi hasil perhitungan dari persamaan Gaussian. Sedangkan *output layer* berisi hasil klasifikasi.



Gambar 2. Arsitektur Jaringan RBPNN

Sumber : Implementasi Metode Radial Basis Probabilistic Neural Network (RBPNN) untuk Klasifikasi Breast Cancer. 2016.

2.3 Data Selection dan Addition

Pada tahap ini, melakukan penentuan data dan memilih data yang akan digunakan. Pada penelitian ini dipilih data yaitu data yang telah diperoleh dari *University of Wisconsin Hospitals, Madison (UCI) Machine Learning Repository* dengan judul *Breast Cancer Wisconsin (Diagnostic)* yakni data set *Wisconsin Diagnosis Breast Cancer (WDBC)*.

2.4 Preprocessing dan Cleaning Data

Dalam tahap ini kehandalan data ditingkatkan. Termasuk pembersihan data, seperti menangani data yang tidak lengkap, menghilangkan gangguan atau outlier. Dalam penelitian ini, terdapat pembersihan beberapa data yang nantinya akan menyulitkan dalam menjadikan parameter untuk klasifikasi.

2.5 Data Mining

Pada tahap ini dilakukan pemilihan algoritma dan implementasi dari algoritma *data mining*. Tahap ini terdiri dari:

a. Coding

Tahap ini merupakan tahapan dilakukannya penerjemahan dalam bentuk pemrograman. Pada proses ini lebih diketahui mengenai metode apa yang digunakan. Proses pengklasifikasian dilakukan dengan menggunakan metode Jaringan Syaraf Tiruan dengan model *Radial Basis Probabilistic Neural Network* (RBPNN).

b. Testing

Pada tahap ini dilakukan pengujian untuk mengetahui apakah sistem yang dibuat telah sesuai dengan kebutuhan dan masih terdapat kesalahan atau tidak.

2.6 Evaluasi

Pada tahap ini dilakukan evaluasi dan penerjemahan dari pola yang diperoleh, kemudian diinterpretasikan sebagai pengetahuan yang dapat digunakan untuk dapat menjawab rumusan masalah dan tujuan dari penelitian ini. Yakni bagaimana prosedur klasifikasi *breast cancer* menggunakan *Radial Basis Probabilistic Neural Network* (RBPNN).

2.7 Radial Basis Probabilistic Neural Network (RBPNN) untuk Klasifikasi

2.7.1 Radial Basis Probabilistic Neural Network (RBPNN)

Model *Radial Basis Probabilistic Neural Network* (RBPNN) merupakan model jaringan yang menggabungkan beberapa keunggulan Jaringan Fungsi Basis Radial (RBFNN) dan Jaringan Probabilistik (PNN) [5]. *Radial Basis Function Neural Network* (RBFNN) dikenal sebagai model *neural network* yang handal dan banyak digunakan pada masalah peramalan (*forecasting*) dan klasifikasi [6]. Jaringan RBPNN yang terdiri dari empat *layer*, yaitu *layer input*, dua *hidden layer* dan *layer output*. *Hidden layer* pertama merupakan *layer* dengan proses *non linier*, yang secara umum terdiri dari *center* tersembunyi yang ditentukan dari *training input*. *Hidden layer* kedua merupakan penjumlahan dari *output layer* pertama, dan secara umum memiliki ukuran yang sama dengan *layer output*. Bobot antara *hidden layer* pertama dan *hidden layer* kedua, merupakan bobot konstan. Artinya, bobot disetting tetap sehingga tidak diperlukan *learning*. Lapisan terakhir dari RBPNN adalah lapisan *output*. Secara matematis, RBPNN dengan vektor *input* x akan menghasilkan nilai aktual untuk neuron *output* ke- i yang dinyatakan sebagai persamaan berikut [5]:

$$y_i^\alpha = \sum_{k=1}^M w_{ik} h_k(x) \quad \dots (1)$$

dimana

$$h_k(x) = \sum_{i=1}^{nk} \phi_i(x, c_{ki}) = \sum_{i=1}^{nk} \phi_i(\|x - c_{ki}\|_2), k = 1, 2, \dots, M \quad \dots (2)$$

Dengan,

- $h_k(x)$: nilai output ke- k dari lapisan *hidden* kedua
 w_{ki} : bobot synaptik antara neuron ke- k pada lapisan *hidden* kedua dan neuron ke i pada lapisan *output*
 c_{ki} : vektor *center hidden* untuk klas ke- k
 nk : jumlah vector *center hidden* untuk klas ke- k
 $\|\bullet\|_2$: norm Eucledian
 M : jumlah *neuron* pada lapisan output dan *hidden layer* kedua atau jumlah klas untuk sample *training*
 Untuk $\varphi_i(\bullet)$ adalah fungsi kernel yang umumnya fungsi kernel Gaussian, $\varphi_i(\|x - c_{ki}\|_2)$, dapat ditulis sebagai berikut:

$$\varphi_i(\|x - c_{ki}\|_2) = \exp\left(\frac{\|x - c_{ki}\|_2^2}{\sigma_i^2}\right) \quad \dots(3)$$

Dimana σ_i adalah parameter pada fungsi kernel Gaussian.

Metode penentuan inisialisasi *center* untuk RBFNN dapat juga dilakukan pada RBPNN. Salah satu metode yang digunakan untuk menentukan inisialisasi *center* RBFNN yaitu dengan menggunakan teknik klustering [6].

Berikut adalah algoritma RBPNN:

- a. Inisialisasi: Tahap inisialisasi RBPNN menggunakan input hasil klustering. Data akan dibagi dua bagian, bagian pertama digunakan sebagai data *training* sebanyak 80% dari total data dan yang lainnya untuk data *testing* sebesar 20% [7].
- b. *Training*: Proses *training* merupakan pelatihan dari data input RBPNN kedalam fungsi aktivasi Gaussian.
 - 1) Untuk setiap vektor input pelatihan dihitung jaraknya terhadap *center* dengan persamaan fungsi Gaussian berikut:

$$\phi_i(\|x - c_{ki}\|_2) = \exp\left(-\frac{\|x - c_{ki}\|_2^2}{\sigma_i^2}\right)$$

dengan σ_i adalah parameter pada fungsi kernel Gaussian diberikan sebagai berikut:

$$\sigma = \frac{d_{max}}{\sqrt{K}}$$

Dimana d_{max} adalah jarak Euclid maksimal dari kelompok pelatihan dan K adalah jumlah total kelompok pelatihan.

- 2) Hitung matriks H yaitu output pada *hidden layer* pertama dengan persamaan sebagai berikut:

$$h_k(x) = \sum_{i=1}^{nk} \phi_i(x, c_{ki}) = \sum_{i=1}^{nk} \phi_i(\|x - c_{ki}\|_2), k = 1, 2, \dots, m$$

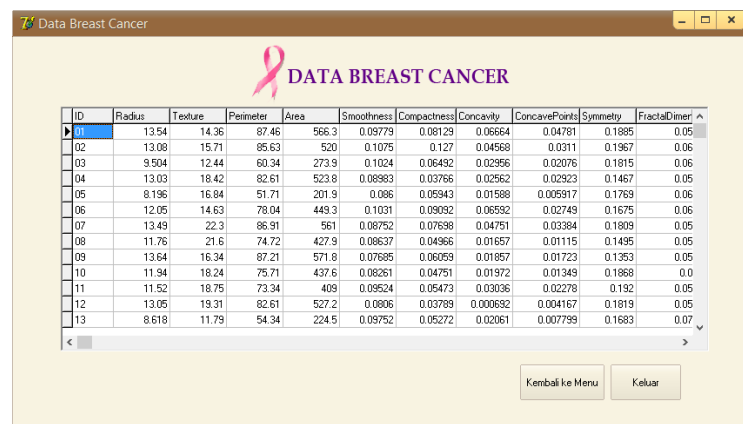
- 3) Selanjutnya dilakukan dekomposisi orthogonal matriks H secara rekursif
 - 4) Hitung output: $Y = HW$
- c. *Testing* : Pada proses *testing*, vektor input untuk *testing* dimasukkan proses (1) dan (2) pada *training* untuk mendapatkan matriks H . Selanjutnya matriks H tersebut digunakan untuk mencari output Y dengan menggunakan bobot W yang telah diperoleh pada proses *training* [6].

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Adapun tampilan dari sistem yang dibuat ini yaitu sebagai berikut:

- a. Form Data Breast Cancer




ID	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
01	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781	0.1885	0.05
02	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311	0.1967	0.06
03	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956	0.02076	0.1815	0.06
04	13.03	18.42	82.61	523.8	0.08983	0.03766	0.02962	0.02923	0.1457	0.05
05	8.196	16.84	51.71	201.9	0.086	0.05943	0.01588	0.005917	0.1769	0.06
06	12.05	14.63	78.04	449.3	0.1031	0.09032	0.06532	0.02749	0.1675	0.06
07	13.49	22.3	86.91	561	0.08752	0.07688	0.04751	0.03384	0.1809	0.05
08	11.76	21.6	74.72	427.9	0.08537	0.04966	0.01657	0.01115	0.1495	0.05
09	13.64	16.34	67.21	571.8	0.07885	0.06059	0.01857	0.01723	0.1353	0.05
10	11.94	18.24	75.71	437.6	0.08261	0.04751	0.01972	0.01349	0.1868	0.0
11	11.52	18.75	73.34	409	0.08524	0.05473	0.03036	0.02278	0.192	0.05
12	13.05	19.31	82.61	527.2	0.0806	0.03789	0.00682	0.004167	0.1819	0.05
13	8.618	11.79	54.34	224.5	0.09752	0.05272	0.02061	0.007798	0.1683	0.07

Gambar 3. Form Data Breast Cancer

Sumber : Implementasi Metode *Radial Basis Probabilistic Neural Network* (RBPNN) untuk Klasifikasi *Breast Cancer*. 2016.

- b. Form Perbandingan Data



Data Training %

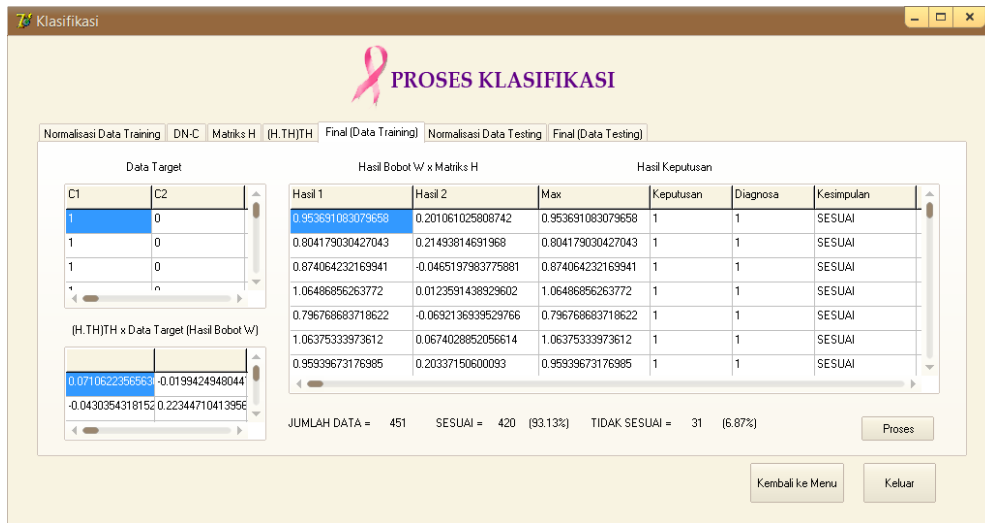
Data Testing %

Lanjut Klasifikasi Keluar

Gambar 4. Form Perbandingan Data

Sumber : Implementasi Metode *Radial Basis Probabilistic Neural Network* (RBPNN) untuk Klasifikasi *Breast Cancer*. 2016.

c. Form Proses Klasifikasi



Gambar 5. Proses Klasifikasi

Sumber : Implementasi Metode *Radial Basis Probabilistic Neural Network* (RBPNN) untuk Klasifikasi *Breast Cancer*. 2016.

3.2 Uji Coba dan Evaluasi

Tahap uji coba yaitu menggunakan analisis black box yang dilakukan dengan cara memberikan input yang valid dan non valid ke sistem dengan beberapa kondisi tertentu untuk mengetahui apakah sistem dapat tetap berjalan dan apakah output yang dihasilkan sudah sesuai dengan harapan atau tidak. Adapun uji coba yang dilakukan sebagai berikut:

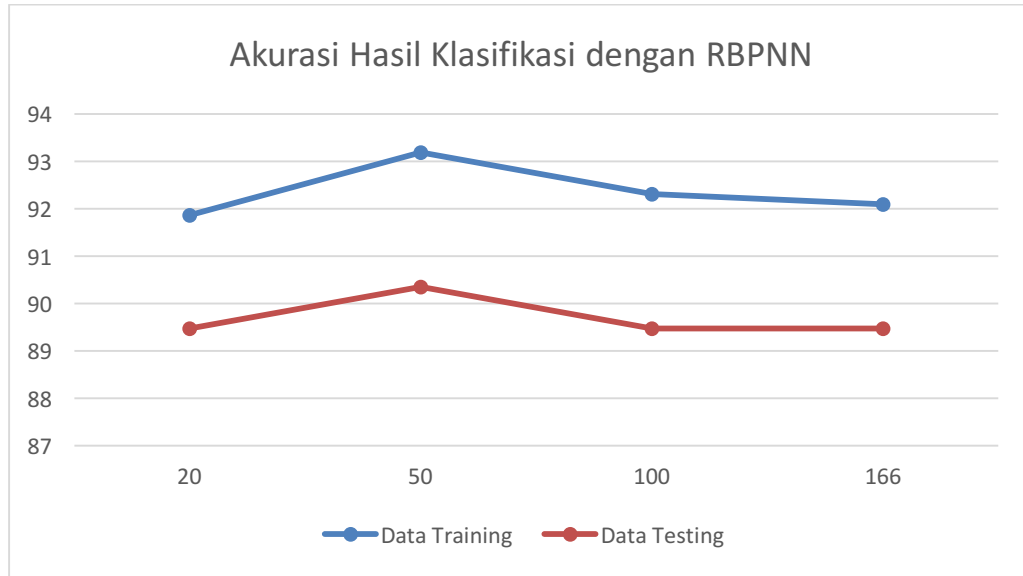
- a. Uji coba hasil perhitungan sistem dan excel
- b. Uji coba output

Pada tahap uji coba output digunakan data *breast cancer* dimana tujuannya untuk memklasifikasikan kanker payudara. Hasil dari klasifikasi tersebut dapat dilihat pada tabel dibawah ini.

Tabel 1. Hasil Klasifikasi

Jumlah center per cluster	Akurasi (%)	
	Data Training	Data Testing
20	91.87	89.47
50	93.19	90.35
100	92.31	89.47
166	92.09	89.47

Dari hasil klasifikasi diatas untuk lebih jelasnya dapat dilihat pada grafik dibawah ini.



Gambar 6. Grafik Akurasi Hasil Klasifikasi dengan RBPNN

3.3 Pembahasan

Jaringan saraf tiruan yang digunakan pada penelitian kali ini adalah *Radial Basis Probabilistic Neural Network* (RBPNN). Berikut uraian proses perhitungan dengan menggunakan metode RBPNN [8]:

a. Pengelompokan Data

Data *Wisconsin Diagnosis Breast Cancer* (WDBC) yang diperoleh memiliki beberapa variabel, pada penelitian menggunakan 11 variabel yaitu variabel *radius, texture, area, perimeter, smoothness, compactness, concavity, concave points, symmetry, fractal dimension*, serta *diagnosis*. Adapun tujuan dalam pengelompokan data adalah untuk mempermudah user dalam proses normalisasi dan proses klasifikasi.

b. Normalisasi Data

Normalisasi digunakan untuk menyamakan skala atribut data ke dalam sebuah range yang spesifik misalnya dari 0 sampai 1 agar mempermudah dalam melakukan proses perhitungan nantinya. Data normalisasi pada penelitian ini menggunakan 10 atribut dari dataset *breast cancer* yaitu *radius, texture, perimeter, area, smoothness, compactness, concave points, symmetry*, dan *fractal dimension*. Rumus normalisasi yang digunakan sebagai berikut:

$$X_n = \frac{X_0}{X_{\max}} \quad \dots(4)$$

Keterangan:

X_n = Nilai data normal

X_0 = Nilai data aktual

X_{\max} =Nilai max data aktual keseluruhan

c. Penentuan Nilai Center

Tahap selanjutnya yaitu penentuan nilai center, dari data *training breast cancer* dipilih 50 data untuk masing-masing cluster yang diambil secara acak. Selanjutnya dari nilai cluster yang didapatkan akan di masukan ke dalam rumus:

$$\varphi_i(\|x - c_{ki}\|_2) = \exp\left(\frac{\|x - c_{ki}\|_2^2}{\sigma_i^2}\right)$$

Keterangan:

x = nilai normalisasi

c = nilai center

σ = parameter pada fungsi kernel Gaussian

d. Membentuk Matriks Gaussian

Berdasarkan nilai yang sudah dihitung dengan menggunakan fungsi aktivasi gaussian kemudian tiap-tiap nilai center akan dijumlahkan kemudian akan disusun membentuk Matriks Gaussian.

e. Menentukan Bobot

Menghitung bobot baru (W) dengan mengalikan pseudoinverse dari matriks H, dengan vektor target (d) dari data *training*. Rumusnya:

$$W = (H^T H)^{-1} H^T d$$

1. Menentukan hasil transpose dari Matriks Gaussian (H^T).

2. Mengalikan matriks Gaussian dan matriks transpose.

3. Mengalikan hasil perkalian matriks sebelumnya dengan transpose matriks gaussian

4. Mengalikan hasil dan target, yang kemudian menghasilkan bobot (W).

f. Menentukan Bobot (W)

Selanjutnya nilai bobot akan digunakan untuk menghitung nilai output pada jaringan, dengan menggunakan rumus:

$$Y = \sum_{i=1} w h_k(x)$$

Keterangan:

W = Nilai bobot

$h_k(x)$ = Nilai output ke-k pada *hidden layer* kedua

g. Akurasi

Setelah diketahui nilai Y diatas, akan dihitung nilai untuk mencari keakuratan dari hasil klasifikasi tersebut dengan menggunakan rumus:

$$\text{Akurasi} = \frac{\text{jumlah hasil keputusan benar}}{\text{jumlah seluruh data}} \times 100\%$$

Berikut tabel hasil klasifikasi, hasil diagnosa, dan akurasi untuk klasifikasi.

Tabel 2. Hasil Klasifikasi, Hasil Diagnosa dan Akurasi pada Data Training

Hasil 1	Hasil 2	Keputusan	Diagnosa	Kesimpulan
0.9536	0.201	1	1	Sesuai
0.8041	0.2149	1	1	Sesuai
0.874	-0.0465	1	1	Sesuai
1.0648	0.0123	1	1	Sesuai
...	...			
0.9593	0.2033	1	1	Sesuai
Jumlah Data=455	Sesuai =424	Akurasi=93.19%		

Tabel 3. Hasil Klasifikasi, Hasil Diagnosa dan Akurasi pada Data Testing

Hasil 1	Hasil 2	Keputusan	Diagnosa	Kesimpulan
0.9576	0.0331	1	1	Sesuai
1.1777	0.0646	1	1	Sesuai
0.8204	0.2703	1	1	Sesuai
0.9126	0.0198	1	1	Sesuai
...	...			
1.0457	0.0576	1	1	Sesuai
Jumlah Data=114	Sesuai =103	Akurasi=90.35%		

4. SIMPULAN

Kesimpulan yang diperoleh dari penelitian ini adalah:

- Radial Basis Probabilistic Neural Network* (RBPNN) dapat digunakan untuk klasifikasi kanker payudara.
- Hasil klasifikasi dengan menggunakan metode Radial Basis Probabilistic Neural Network untuk kanker payudara dari sistem yang dibuat untuk 20, 50, 100, dan 166 center per cluster dengan perbandingan data sebesar 80% untuk data training, dan 20% data testing memperoleh akurasi untuk data training berturut-turut sebesar 91.87%, 93.19%, 92.31%, dan 92.09%, dan untuk data testing berturut-turut sebesar 89.47%, 90.35%, 89.47%, dan 89.47%.
- Berdasarkan uji coba *blackbox*, setiap input yang diberikan, sistem yang dibangun sudah berjalan sesuai dengan output yang dihasilkan.

DAFTAR PUSTAKA

- [1] Makhfudhoh, Noor Uswah. 2014. **Klasifikasi Kanker Payudara dari Citra Mammografi Menggunakan Model Fuzzy Neural Network**. Jurnal Universitas Negeri Yogyakarta, edisi IV Volume III Juli - Agustus 2014.
- [2] Chester, M. 1993. **Neural Networks: A Tutorial**. Englewood Cliffs, NJ: Prentice Hall, 1993, ch.2.
- [3] Tingting, Mu. dan Nandi, A. K. 2007. **Breast Cancer Detection from FNA using SVM with Different Parameter Tuning Systems and SOM-RBF Classifier**. Journal of the Franklin Institute, Vol 344, pp. 285-311, 2007.

- [4] Al-Timemy, Ali H., Al-Naima, Fawzi M. dan Qaeab, Nebras H. 2009. **Probabilistic Neural Network for Breast Biopsy Classification**. MASAUM Journal of Computing, Volume 1 Issue 2.
- [5] Huang, DS. dan Du, JX. 2008. **A Constructive Hybrid Structure Optimization Methodology for Radial Basis Probabilistic Neural network**. IEEE Transactionon Neural network 19, 2099–2115.
- [6] Soesanto, Oni. & Irawan, Mohammad Isa. 2010. **Hybrid PCA-RBPNN Pada Klasifikasi Data Multivariat**. Prosiding Konferensi Nasional Matematika (KNM) 15, Manado.
- [7] Zhang, G. P. 1998. **Forecasting with Artificial Neural Networks : The State of the Art**. Elsevier International Journal of Forecasting, 14 , 35-62.
- [8] Shahura, Fairudz. 2016. **Implementasi Metode Radial Basis Probabilistic Neural Network (RBPNN) untuk Klasifikasi Breast Cancer**. Skripsi Ilmu Komputer, Universitas Lambung Mangkurat.
- [9] Wolberg, William H., Nick, W. & Mangasarian, Olvi L. 1992. **Breast Cancer Wisconsin (Diagnostic) Data Set**.
[https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))