

Deteksi Risiko Terkena Kanker Payudara Berdasarkan Karakteristik Tumor Menggunakan Metode Fuzzy Inference System (FIS)

Fadhline Frima¹, Agung Toto Wibowo, S.T., M.T.², Mahmud Dwi Sulityo, S.T., M.T.³

^{1,2}Fakultas Informatika IT Telkom, Bandung

¹alin.fadhlinefrima@gmail.com, ²atw@ittelkom.ac.id, ³mahmuddwis@gmail.com

ABSTRAK

Kanker payudara adalah penyakit yang sering menimbulkan kematian pada wanita. Kemunculan gumpalan pada payudara merupakan suatu tanda yang harus diwaspadai, tetapi hal tersebut tidak selalu mengindikasikan kanker. Untuk itu diperlukan suatu tindakan pendeteksian dini untuk mendeteksi tingkat keganasan suatu kanker. Salah satu fungsi penanda tumor adalah *skrining* dan deteksi awal[5].

Data penanda tumor yang digunakan yaitu Database Breast Cancer Wisconsin. Sistem memiliki lima variabel input yaitu *Cell Size*, *Cell Shape*, *Bare Nuclei*, *Bland Chromatin* dan *Epithelial Size*, serta memiliki satu variabel output yaitu *Class Jenis Kanker*. Sistem ini menggunakan metode *Fuzzy Inference System* dengan model *inference* mamdani dalam deteksi dini risiko terkena kanker payudara. Sistem ini diharapkan dapat membantu ahli medis dalam deteksi dini risiko terkena kanker payudara secara otomatis.

Kata kunci: kanker payudara, penanda tumor, sistem fuzzy, FIS, inferensi mamdani

ABSTRACT

Breast cancer is a disease that often causes death in women. The appearance of a lump in the breast is a sign to look out for, but it does not always indicate cancer. It required an act of initial detection to detect the degree of malignancy of a cancer. One function of tumor markers is skrining and initial detection [5].

Tumor markers used are Wisconsin Breast Cancer Database. The system has five input variables, namely Cell Size, Cell Shape, Bare Nuclei, Bland Chromatin and epithelial-Size, and has a variable output of the Class Type of Cancer. The method of the system is Fuzzy Inference System with Mamdani inference model in the initial detection of breast cancer risk. The system is expected to help medical experts in the initial detection of breast cancer risk automatically.

Keywords: kanker payudara, penanda tumor, sistem fuzzy, FIS, inferensi mamdani

1. PENDAHULUAN

Kanker payudara adalah penyakit yang sering menimbulkan kematian pada wanita. Kemunculan gumpalan pada payudara merupakan suatu tanda yang harus diwaspadai, tetapi hal tersebut tidak selalu mengindikasikan kanker. Untuk itu diperlukan suatu tindakan pendeteksian dini untuk mendeteksi tingkat keganasan suatu kanker. Pada penelitian sebelumnya penanda tumor yang digunakan yaitu penanda tumor molekuler (*biomarker*)[1]. Penanda tumor molekuler tersebut adalah *Her2*, *Hormone Receptor*, *Age*, *Tumor Grade*, *Tumor Size* dan *Lymph Node*[1]. Penanda tumor molekuler lebih digunakan untuk pendidagnosaan kanker payudara[1].

Dalam penelitian tugas akhir kali ini, penanda tumor yang digunakan adalah penanda tumor seluler. Salah satu fungsi penanda tumor adalah *skrining* dan deteksi awal[5]. Penanda tumor yang digunakan yaitu penanda tumor *seluler*, dimana pendeteksian dilakukan pada tingkat sel[5]. Penanda tumor *seluler* yang digunakan adalah *Clump Thickness*, *Cell Size*, *Cell Shape*, *Marginal Adhesion*, *Epithelial Size*, *Bare Nuclei*, *Bland Chromatin*, *Normal Nucleoli*, *Mitoses* serta *Class*[10]. Dalam melakukan deteksi

dini melibatkan ilmu pengetahuan dan intuisi manusia. Sedangkan intuisi manusia yang satu dengan yang lainnya dapat berbeda. Oleh karena itu dibutuhkan sebuah sistem yang mampu meminimalisasi keberagaman keputusan individu dalam pengambilan keputusan. Sistem ini diharapkan dapat membantu ahli medis dalam deteksi dini risiko terkena kanker payudara secara otomatis.

2. LANDASAN TEORI

2.1 Kanker Payudara

Tumor adalah pertumbuhan jaringan tubuh dimana terjadi pertumbuhan yang pesat dalam menghasilkan jaringan baru (*proliferasi*). Tumor berupa massa padat atau berisi cairan yang ukurannya membesar. Kanker adalah sebuah keganasan sedangkan tumor dapat bersifat jinak atau ganas. Tumor terjadi ketika sel membelah dan tumbuh berlebihan dalam tubuh. Sel-sel baru diciptakan untuk menggantikan sel yang lebih tua atau untuk melakukan fungsi-fungsi baru. Sel yang rusak atau tidak diperlukan lagi mati untuk

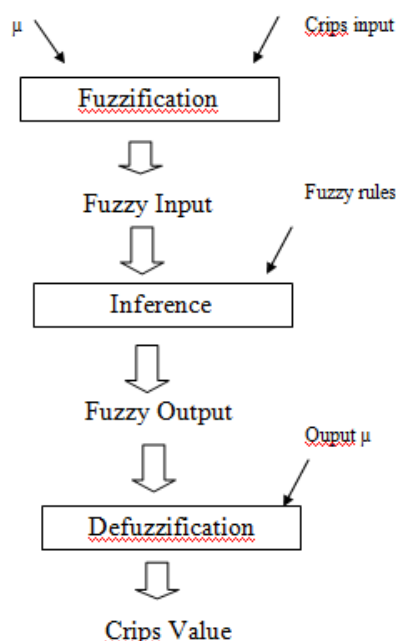
mendapatkan ruang untuk pengganti yang sehat. Jika keseimbangan pertumbuhan sel dan kematian sel terganggu, maka akan terbentuk tumor [2].

Istilah “jinak” atau “benigna” memiliki arti ringan atau tidak progresif. Dikatakan tumor jinak jika gangguan *proliferasi* sel-sel tidak menyerang jaringan yang berada di dekatnya atau tidak menyebar ke bagian lain dari tubuh. Tumor jinak biasanya dikelilingi oleh selubung fibrosa (kapsul) yang menghambat kemampuan mereka untuk berperilaku ganas [2].

Istilah “ganas” atau “maligna” yang berarti buruk dan berpotensi mematikan. Tumor ganas adalah kanker, tetapi kanker belum tentu berbentuk tumor. Sel-sel kanker dapat menyerang dan merusak jaringan dan organ di dekat tumor. Sel-sel kanker dapat melepaskan diri dari tumor ganas dan memasuki aliran darah. Kanker menyebar dari tumor asli untuk membentuk tumor baru dari bagian tubuh yang lainnya. Kebanyakan tumor ganas tidak memiliki kapsul seperti tumor jinak[2].

2.2 Fuzzy Inference System

Suatu sistem yang berbasis aturan *fuzzy* terdiri dari tiga komponen utama : *Fuzzification*, *Inference* dan *Defuzzification*[6].



Gambar 2.1 Diagram Blok Sistem Berbasis Fuzzy

2.3 Information Gain

Untuk mengevaluasi semua atribut yang ada maka diperlukan suatu alat ukur statistik yang digunakan untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan kumpulan sampel data[7]. Ukuran efektifitas ini disebut *information gain*. Untuk menghitung *information gain* terlebih dahulu kita harus memahami suatu ukuran lain yang disebut *entropy*[7]. Di dalam bidang *Information Theory*, kita sering menggunakan *entropy* sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dari suatu kumpulan sampel data. Jika kumpulan sampel data semakin heterogen, maka nilai *entropy*-nya semakin besar. Secara matematis, *entropy* dirumuskan sebagai berikut :

$$Entropy(S) = \sum_i^c -p_i \log_2 p_i$$

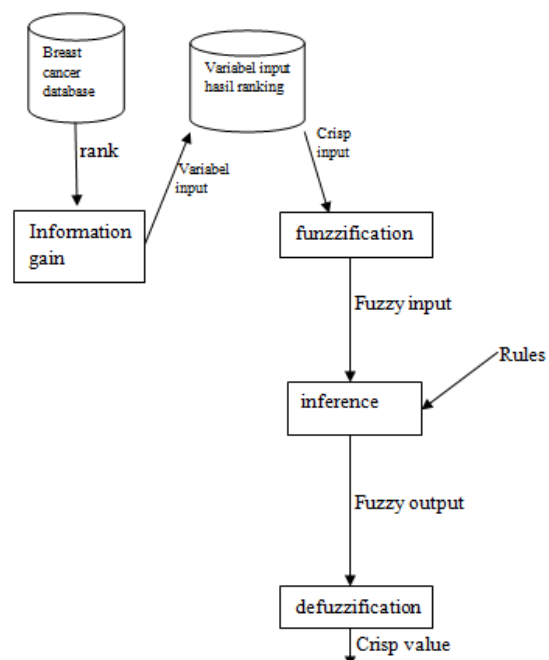
di mana c adalah jumlah nilai yang ada pada atribut target (jumlah kelas klasifikasi). Sedangkan p_i menyatakan jumlah sampel untuk kelas i .

Setelah mendapatkan nilai *entropy* untuk suatu kumpulan data, maka kita dapat mengukur efektivitas suatu atribut dalam mengklasifikasi data. Ukuran efektifitas ini disebut *information gain*.

3. MODEL DAN PERANCANGAN SISTEM

3.1 Gambaran Umum Sistem

Sistem yang dibangun pada tugas akhir ini adalah sistem Deteksi Risiko Terkena Kanker Payudara Berdasarkan Karakteristik Tumor Menggunakan Metode Fuzzy Inference System (FIS) Tools yang digunakan dalam pembuatan sistem yaitu Matlab R2009b dan Weka 3.7.



Gambar 3.1 Alur Sistem

Data penanda tumor (nilai *crisp*) akan diinputkan ke dalam sistem. Pada tahap *fuzzifier*, nilai *crisp* penanda tumor akan di konversi ke bentuk *fuzzy input* yaitu berupa nilai *membership function*. Pada tahap inference dilakukan penalaran menggunakan *fuzzy input* dan *fuzzy rules*, sehingga didapatkan *fuzzy output*. Pada *defuzzifier* akan diubah *fuzzy output* menjadi nilai *crisp* berdasarkan fungsi keanggotaan yang telah di tentukan.

3.2 Wisconsin Breast Cancer Database

Dalam perancangan sistem ini, penulis menggunakan Wisconsin Breast Cancer Database yang berasal dari University of Wisconsin Hospitals, Madison. Data ini merupakan hasil penelitian dari Dr. William H. Wolberg. Total keseluruhan data yang dimiliki yaitu 699 data dengan 10 atribut, terdiri dari[10] :

1. Clump Thickness [1 – 10]
Pada Clump Thickness, sel jinak cenderung dikelompokkan dalam monolayer sedangkan sel-sel kanker sering dikelompokkan dalam multilayer.
2. Uniformity of Cell Size[1 – 10]
Sel-sel kanker cenderung memiliki ukuran dan bentuk yang bervariasi. Oleh karena itu parameter ini sangat berpengaruh dalam penentuan sel kanker atau tidak.
3. Uniformity of Cell Shape[1 – 10]
Sel-sel kanker cenderung memiliki ukuran dan bentuk yang bervariasi. Oleh karena itu parameter ini sangat berpengaruh dalam penentuan sel kanker atau tidak.
4. Marginal Adhesion[1 – 10]
Sel-sel normal cenderung untuk terkumpul secara bersama-sama. Sementara sel-sel kanker cenderung terpisah-pisah dan ini menjadi tanda dari keganasan.
5. Single Epithelial Cell Size [1 – 10]
Parameter ini terkait dengan uniformity cell size dan cell shape. Sel-sel epitel yang membesar secara signifikan berkemungkinan menjadi tanda keganasan sel.
6. Bare Nuclei[1 – 10]
Parameter ini digunakan untuk inti yang tidak dikelilingi oleh sitoplasma (sisa sel). Biasanya terlihat pada tumor jinak.
7. Bland Chromatin [1 – 10]
Parameter ini menunjukkan sebagai sebuah “tekstur” inti yang terlihat pada sel jinak.

Dalam sel-sel kanker kromatin cenderung memiliki “tekstur” yang lebih kasar.

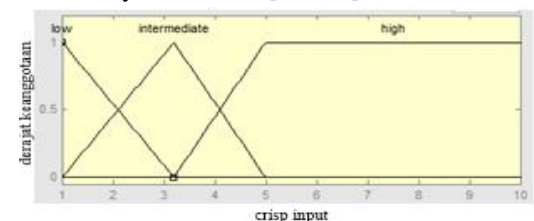
8. Normal Nucleoli[1 – 10]
Nukleolus merupakan struktur terkecil yang terlihat pada inti. Dalam sel normal nukleolus biasanya sangat kecil. Dalam sel-sel kanker nukleolus menjadi lebih menonjol.
9. Mitoses[1 – 10]
Merupakan proses pembelahan sel. Yang nantinya akan menandakan tergolong sel jinak atau ganas.
10. Class[2 untuk benign, 4 untuk malignant]
Merupakan kelas pengelompokan sel.

3.3 Fuzzification

Proses pengubahan inputan-inputan yang nilai kebenarannya bersifat pasti (*crisp input*) kedalam bentuk *fuzzy input*, yang berupa nilai linguistic yang semantiknya ditentukan berdasarkan fungsi keanggotaan tertentu. Atribut-atribut pada data Wisconsin Breast Cancer Database digunakan sebagai inputan. Atribut ini berupa Clump Thickness, Cell Size, Cell Shape, Marginal Adhesion, Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, serta Mitoses yang akan diubah kedalam bentuk *fuzzy input*[6].

A. Input Variabel

1. Uniformity of Cell Size[1 – 10]

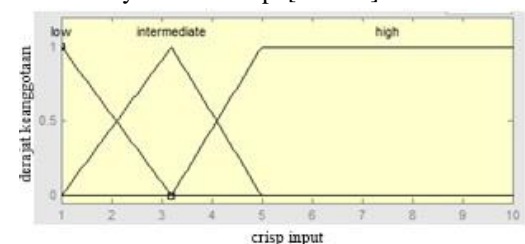


Gambar 3.2 Membership function CellSize

Tabel 3.1 Fuzzy set CellSize

Variabel Input	Range	Fuzzy set
CellSize	$1 \leq x \leq 3,2$	Low
	$1 \leq x \leq 5$	Intermediate
	$3,2 \leq x \leq 10$	High

2. Uniformity of Cell Shape[1 – 10]

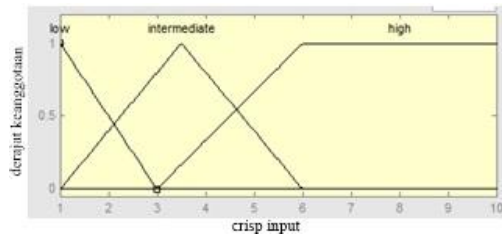


Gambar 3.3 *Membership function CellShape*

Table 3.2 Fuzzy set CellShape

Variabel Input	Range	Fuzzy set
CellShape	$1 \leq x \leq 3,2$	Low
	$1 \leq x \leq 5$	Intermediate
	$3,2 \leq x \leq 10$	High

3. Bare Nuclei [1 – 10]

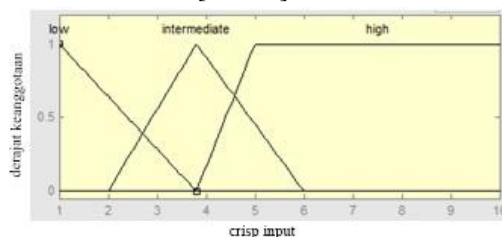


Gambar 3.4 *Membership function BareNuclei*

Tabel 3.3 Fuzzy set BareNuclei

Variabel Input	Range	Fuzzy set
BareNuclei	$1 \leq x \leq 3$	Low
	$1 \leq x \leq 6$	Intermediate
	$3 \leq x \leq 10$	High

4. Bland Chromatin [1 – 10]

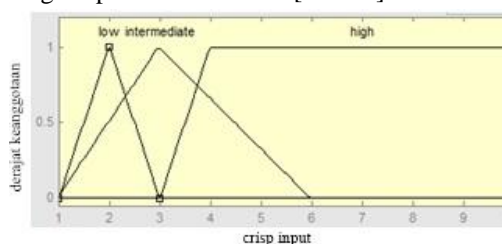


Gambar 3.5 *Membership function BlandChromatin*

Tabel 3.4 Fuzzy set BlandChromatin

Variabel Input	Range	Fuzzy set
BlandChromatin	$1 \leq x \leq 3,8$	Low
	$2 \leq x \leq 6$	Intermediate
	$3,8 \leq x \leq 10$	High

5. Single Epithelial Cell Size [1 – 10]



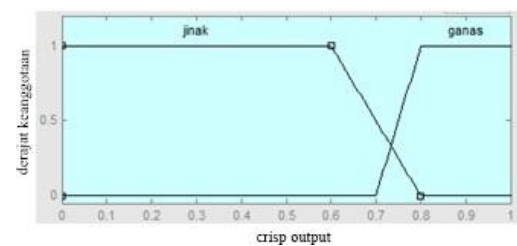
Gambar 3.6 *Membership function EpithelialSize*

Tab 3.5 Fuzzy set EpithelialSize

Variabel Input	Range	Fuzzy set
EpithelialSize	$1 \leq x \leq 3$	Low
	$1 \leq x \leq 6$	Intermediate
	$3 \leq x \leq 10$	High

B. Output Variabel

1. Jenis Tumor



Gambar 3.7 *Membership function Jenis Tumor*

Tabel 3.6 Fuzzy set Jenis Kanker

Variabel Input	Range	Fuzzy set
Jenis Tumor	$1 \leq x \leq 0,8$	Jinak
	$0,7 \leq x \leq 10$	Ganas

3.4 Inference

Proses dalam melakukan penalaran dalam deteksi resiko kanker menggunakan *fuzzy input* dan *fuzzy rules* yang telah ditentukan sehingga menghasilkan *fuzzy output*. Aturan fuzzy yang digunakan dalam pendeteksian tumor dapat diperoleh dengan cara melatih sistem dengan *training set* yang representatif dari kumpulan data-data numeric Wisconsin Breast Cancer Database dengan menggunakan *fuzzy grid table*.

3.5 Defuzzification

Proses pengubahan fuzzy output menjadi crisp value (nilai tegas) tumor jinak dan tumor ganas berdasarkan fungsi keanggotaan yang telah ditentukan.

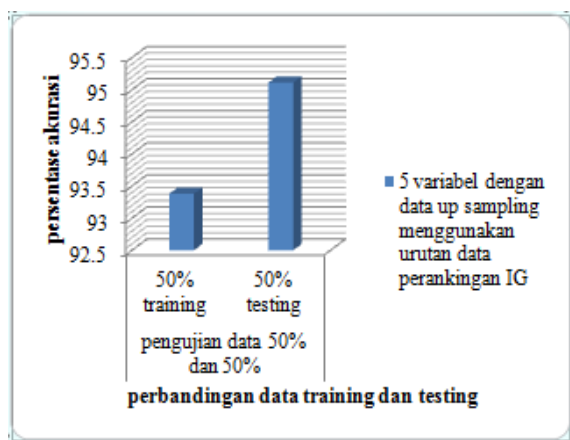
4. ANALISIS DAN PENGUJIAN SISTEM

4.1 Analisis Pengaruh Information Gain Terhadap Urutan Variabel

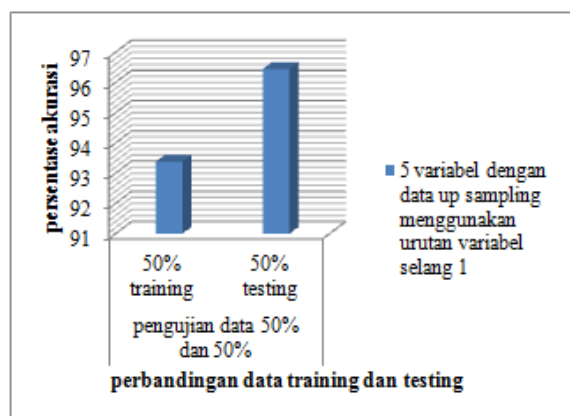
Berdasarkan grafik hasil pengujian sistem untuk semua skenario urutan variabel, tingkat akurasi yang dihasilkan oleh masing-masing skenario memiliki perbandingan akurasi yang hampir sama antara data training dan testing. Baik untuk data unbalance, data balance up sampling serta data balance down sampling, setiap skenario

menghasilkan tingkat akurasi yang hampir sama. Ini menunjukkan bahwa walaupun urutan variabel diurut secara acak dengan tetap menggunakan hasil perancangan sistem yang telah dirancang sebelumnya, akan tetap menghasilkan tingkat akurasi yang baik. Hasil perankingan variabel menggunakan *Information gain* dapat memberikan perankingan variabel terbaik yang memiliki efektifitas yang tinggi dalam melakukan pengklasifikasian data.

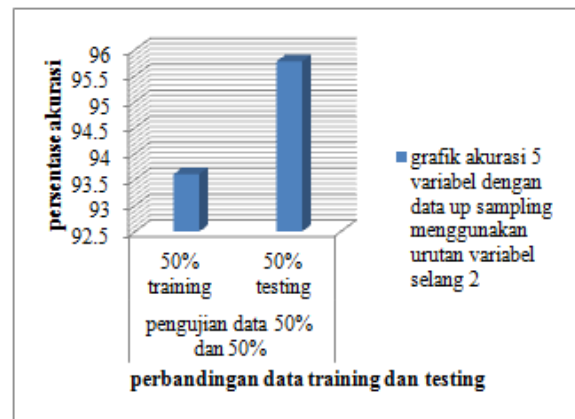
Dapat disimpulkan bahwa perhitungan *information gain* dalam meranking setiap variabel, tidak bergantung kepada variabel sebelum atau setelahnya. Sehingga apabila urutan variabel diacak tidak akan mempengaruhi tingkat akurasi dari sistem.



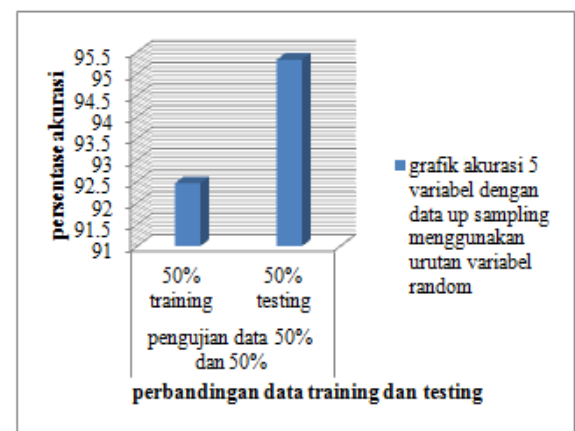
Gambar 4.1 Grafik akurasi 5 variabel dengan data up sampling menggunakan urutan variabel terurut



Gambar 4.2 Grafik akurasi variabel dengan data up sampling menggunakan urutan variabel selang 1



Gambar 4.3 Grafik akurasi 5 variabel dengan data up sampling menggunakan urutan variabel selang 2



Gambar 4.4 Grafik akurasi 5 variabel dengan data up sampling menggunakan urutan variabel random

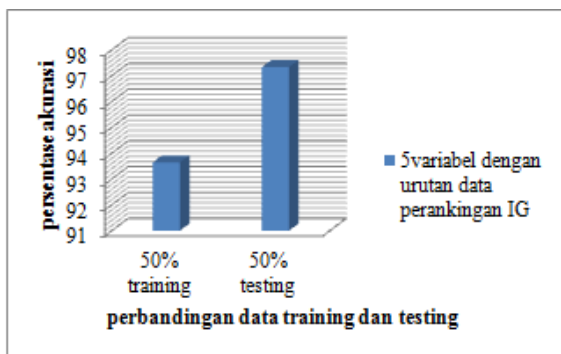
4.2 Analisis Pengujian Unbalance Data dan Balance Data

Berdasarkan grafik hasil pengujian sistem dapat dilihat unbalance data dan balance data up sampling memiliki hasil akurasi yang hampir sama antara data training dan testing. Sedangkan untuk balance data down sampling memiliki hasil akurasi yang sedikit berbeda pada data testing, 3 dari 4 skenario urutan variabel menunjukkan hasil akurasi di bawah 89%. Sedangkan pada data training memiliki akurasi yang hampir dengan unbalance data dan balance data up sampling.

Ini menunjukkan bahwa, terdapat pengaruh data unbalance dan balance terhadap perancangan sistem. Data yang pada balance data under sampling memiliki jumlah yang lebih sedikit di bandingkan data unbalance dan data balance up sampling. Data yang semulanya berjumlah 600 data dengan perbandingan 400 jinak dan 200 ganas, dibuat menjadi balance dengan mengurangi data jinak secara sehingga memiliki jumlah yang seimbang dengan data ganas. Sehingga jumlah data yang digunakan

sistem untuk proses pembelajaran semakin sedikit. Hal ini menyebabkan akurasi sistem menjadi berkurang. Sedangkan balance data dengan cara upper sampling, dengan perbandingan data yang sama, data ganas yang semulanya 200 ditambah secara random sehingga jumlahnya seimbang dengan data jinak. Data yang digunakan sistem untuk proses pembelajaran semakin banyak, sehingga akurasi sistem menjadi baik.

Dari hasil akurasi yang di dapat, hasil akurasi terbesar yaitu 96%. Ini menunjukkan bahwa terdapat 4% pasien yang mengalami kesalahan pendeteksian. Ketidakkakurata ini berdampak pada kesalahan tindakan pengobatan terhadap pasien.



Gambar 4.5 Grafik akurasi unbalance data 5variabel dengan urutan data perankingan IG

5. PENUTUP

5.1 Kesimpulan

Dari hasil pengujian dan analisa yang telah dilakukan, dapat di ambil beberapa kesimpulan sebagai berikut :

1. Sistem yang dibangun telah mampu mendeteksi risiko kanker payudara dengan melakukan klasifikasi data antara tumor jinak dan tumor ganas, dengan tingkat akurasi terbaik mencapai 96%.
2. Perhitungan *information gain* dalam proses *feature selection* bertujuan untuk memilih fitur-fitur yang penting dalam proses klasifikasi, dan urutan penggunaan fitur-fitur tersebut tidak berpengaruh terhadap akurasi sistem.
3. Penggunaan balanced data dalam proses training atau pelatihan menghasilkan sistem yang lebih baik dibandingkan penggunaan data yang tidak balanced (imbalanced data), dengan tingkat akurasi sistem yang dihasilkan sebesar 96%.
4. Penggunaan fuzzy grid table dapat menghasilkan fuzzy rules untuk sistem dan menggantikan peran *prior information* dengan hasil yang cukup memuaskan

dilihat dari tingkat akurasi yang mencapai 96%.

Bentuk *membership function* yang cukup tepat untuk setiap variable input yaitu *membership function* dengan bentuk Segitiga untuk *fuzzy set* Low dan Intermediate serta bentuk Trapesium untuk *fuzzy set* High dengan batasan-batasan tertentu. Serta bentuk *membership function* yang tepat untuk variabel output adalah trapesium dengan *fuzzy set* jinak dan ganas.

5.2 Saran

1. Bentuk *membership function* sistem fuzzy yang digunakan dapat ditambahkan jenisnya seperti penggunaan *membership function* bentuk Gaussian, dan lain sebagainya.
2. Untuk mendapatkan hasil pendefenisian *membership function* yang lebih optimal harus ditunjang dengan ilmu pengetahuan ahli yang terkait.
3. Untuk mendapatkan hasil pendefenisian rules yang lebih optimal menggunakan Multiple Fuzzy Great Table.

DAFTAR PUSTAKA

- [1] Ahmed Abou Elfetouh Saleh. Sherif Ebrahim Barakat. Ahmed Awad Ebrahim Awad.(March 2011). "A Fuzzy Decision Support System for Management of Breast Cancer". (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 2, No.3.
- [2] "Defining Cance" National Cancer Institue . Tersedia: <http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer> .
- [3] Hartini, Sri. Tersedia: http://members.tripod.com/darqcall_1/sehat05.html [25 November 2012]
- [4] Kusumadewi, Sri dan Hari Purnomo. (2010). *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu
- [5] Koesoma Aman, Adi dan Ratna Akbari Gani. Tersedia: http://ocw.usu.ac.id/course/download/1110000146electiveoncogen/elo173_slide_tumor_marker_s.pdf [25 November 2012]
- [6] Suyanto. (2008). *Soft Computing "Membangun Mesin Ber-IQ Tinggi"*. Bandung: Infomatika.
- [7] Suyanto. (2007). *Artificial Intelligence "Searching, Reasoning, Planning and Learning"*. Bandung: Informatika.
- [8] Syahrul Mubarak. (2009). *Sistem Identifikasi Plagiat Menggunakan Kombinasi Algoritma Genetika – Sistem Fuzzy*. Bandung
- [9] W.H. Wolberg. (1992). Tersedia: <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/breast-cancer->

- [wisconsin/breast-cancer-wisconsin.names](http://www.wisconsin/breast-cancer-wisconsin.names) [Februari 2012]
- [10] W.H. Wolberg. (1992). Terdesia: <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data>
- [11] <http://library.binus.ac.id/eColls/eThesis/Bab2/TSA%202012-0101%20BAB%202.pdf> [Februari 2013]
- [12] http://faculty.ucr.edu/~hanneman/linear_models/c5.html [Februari 2013]