

Implementasi Algoritma Genetika dan Jaringan Syaraf Tiruan Untuk Klasifikasi Penyakit Tumor Usus Besar

Aldi Riyadi Ramadhan¹, Danang Triantoro Murdiansyah², Annisa Aditsania³

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

¹aldiriyadi.telkomuniversity.ac.id,

²danangtri@telkomuniversity.ac.id, ³aaditsania@telkomuniversity.ac.id

Abstrak

Tumor usus besar merupakan salah satu penyakit paling umum yang dapat menyebabkan kematian, di mana setengah juta kasus tumor usus besar terjadi setiap tahun di seluruh dunia. Kasus kematian dapat dicegah dengan melakukan diagnosis yang akurat sehingga pasien bisa mendapatkan pertolongan yang tepat dan cepat. Diagnosis dilakukan dengan memperoleh informasi dari data DNA pasien yang berbentuk microarray memiliki ribuan atribut. Pada penelitian ini dipakai data ekspresi gen pasien tumor colon (tumor usus besar), yang diambil dari Kent Ridge Biomedical Data Set Repository. Dimensi data ekspresi gen berjumlah ribuan hingga jutaan atribut sehingga proses komputasi memakan waktu dan upaya yang lebih besar. Maka dari itu dibutuhkan metode untuk melakukan reduksi dimensi, namun tetap menggunakan data-data terbaik yang mampu memberikan hasil yang akurat. Pada penelitian ini dilakukan penggabungan metode algoritma genetika serta jaringan syaraf tiruan atau lebih sering disebut AG-JST Hybrid yang mampu melakukan reduksi dimensi dan tetap menggunakan data-data terbaik. Hasil terbaik yang diberikan pada pengujian dengan AG-JST adalah reduksi dimensi data DNA sebesar 47% dan akurasi 89%.

Kata kunci : DNA, Tumor usus besar, Algoritma Genetika, Jaringan Syaraf Tiruan, Hybrid.

Abstract

Colon tumors are one of the most common diseases that can cause death, where half a million cases of colon tumors occur every year around the world. Death cases can be prevented by making an accurate diagnosis so that patients can get the right and fast help. Diagnosis is done by obtaining information from patient DNA data in the form of textit microarray which has thousands of attributes. However, the dimensions of human DNA data amount to thousands to millions of attributes so the computational process takes more time and effort. Therefore we need a method to reduce dimensions, but still use the best data that can provide accurate results. In this study, a combination of genetic algorithm and artificial neural networks or more commonly called AG-JST textit Hybrid is capable of reducing dimensions and still using the best data. The best results given for testing with AG-ANN are the reduction of DNA data dimensions by 47 % and accuracy of 89

Keywords: DNA, Colon Tumor, Genetich Algorithm, Artificial Neural Networks, Hybrid.

1. Pendahuluan

Latar Belakang

Tumor usus besar merupakan salah satu penyakit paling umum yang dapat menyebabkan kematian, di mana setengah juta kasus tumor usus besar terjadi setiap tahun di seluruh dunia. Tumor dibedakan menjadi 2 jenis, yaitu tumor jinak dan tumor ganas (kanker)[6]. Diagnosis tumor dilakukan dengan memperoleh informasi dari data DNA pasien dalam bentuk data microarray yang memiliki ribuan atribut. Data yang memiliki banyak atribut (dimensi), memakan banyak waktu dan memanfaatkan upaya komputasi berlebihan yang sulit ditangani. Oleh karena itu, metode reduksi dimensi diperlukan untuk memilih atribut optimal sehingga upaya komputasi dapat dilakukan dengan lebih efisien [8].

Belakangan ini, beberapa penelitian untuk diagnosis telah banyak dilakukan, salah satunya adalah diagnosis kanker payudara dengan menggunakan metode Algoritma Genetika pada Struktur Backpropagation Neural Network yang mampu menghasilkan nilai rata-rata akurasi yang tinggi yaitu sebesar 97% [9]. Melihat akurasi dan kemampuan metode tersebut untuk mereduksi dimensi pada kasus kanker, metode tersebut kemudian dijadikan acuan pada penelitian ini. Maka, pada penelitian ini digunakan metode algoritma genetika (AG) dan jaringan syaraf tiruan (JST) untuk membantu diagnosa tumor usus besar.

JST digunakan untuk training pada bobot dan klasifikasi sedangkan AG digunakan untuk mereduksi dimensi dan mendapatkan akurasi tertinggi dari JST.

Topik dan Batasannya

Berdasarkan latar belakang di atas, dapat dirumuskan permasalahan yang akan diangkat dalam penelitian ini adalah bagaimana implementasi metode algoritma genetika dan jaringan syaraf tiruan untuk klasifikasi dan reduksi dimensi, dan bagaimana hasil performansi dari implementasi metode algoritma genetika dan jaringan syaraf tiruan terhadap data ekspresi gen penyakit tumor usus besar.

Adapun batasan masalah pada penelitian ini adalah data set yang digunakan dalam penelitian tugas akhir ini adalah data colon tumor yang diambil dari Kent Ridge Biomedical Data Set Repository. Data yang diklasifikasikan oleh sistem terbagi kedalam 2 kelas yaitu terdiagnosa "positif " penyakit tumor usus dan yang terdiagnosa "nega-tif " penyakit tumor usus besar.

Tujuan

Tujuan dari tugas akhir ini adalah mengetahui bagaimana implementasi metode algoritma genetika dan jaringan syaraf tiruan untuk mereduksi dimensi, mengetahui bagaimana analisis hasil reduksi dimensi pada metode algoritma genetika dan jaringan syaraf tiruan, mengetahui bagaimana hasil performansi dari implementasi metode algoritma genetika dan jaringan syaraf tiruan terhadap data ekspresi gen penyakit tumor usus besar.

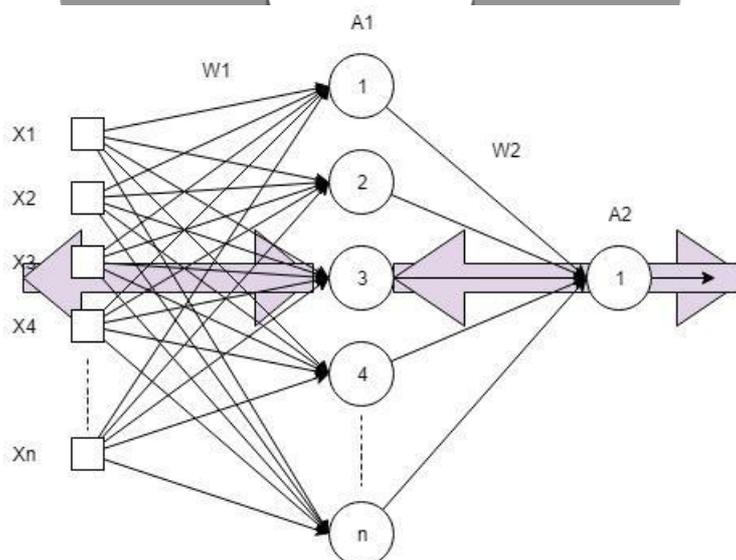
Organisasi Tulisan

Penulisan tugas akhir ini tersusun dalam beberapa bagian, yaitu sebagai berikut: bagian pertama berisi latar belakang, batasan masalah, hingga tujuan mengenai penelitian ini. Bagian kedua berisi studi terkait, yang menjelaskan hal-hal yang berkaitan dengan penelitian ini. Kemudian, bagian ketiga berisi sistem yang dibangun, akan menjelaskan rancangan sistem hingga alur dari JST-GA yang dibangun. Pada bagian keempat berisi evaluasi, mengenai hasil pengujian dan evaluasi sistem, dan bagian kelima berisi kesimpulan dari penelitian.

2. Studi Terkait

2.1 Jaringan Syaraf Tiruan

Jaringan Syaraf Tiruan (JST) merupakan salah satu algoritma pembelajaran mesin yang populer dalam penerapan ramalan rangkaian waktu dan metode yang diterima secara luas dalam melakukan prediksi [3]. JST adalah model digital otak manusia, di mana program dirancang mensimulasikan cara kerja bagaimana cara kerja otak manusia saat memproses informasi [2]. JST mengumpulkan pengetahuan mereka berdasarkan proses saat mendeteksi pola hubungan antar data dan belajar melalui pengalaman kemudian memprediksi masa depan. Salah satu con-toh pengambilan ide pada JST adalah terdapat elemen pemrosesan JST yang saling berkaitan dan berjalan secara paralel. Hal ini meniru jaringan syaraf biologis yang tersusun dari sel-sel syaraf (neuron).



Gambar 1. Jaringan Syaraf Tiruan

Arsitektur dasar pada JST terdiri dari tiga tahap yaitu input, hidden layer, dan output layer. Dalam arsitektur jaringan maju, sinyal mengalir dari input hingga output secara ketat ke arah maju [1]. Pada JST terdapat algoritma Backpropagation dimana terdapat perhitungan maju dan mundur. JST memiliki hidden layer dan output layer yang jumlahnya ditentukan sesuai kebutuhan. Berikut perhitungan maju dan mundur pada JST:

Pada hidden layer:

$$V_1 = P \cdot W_1 + B_1 \tag{1}$$

$$A_1 = \frac{1}{1 + e^{-a \cdot v_1}} \tag{2}$$

Pada output layer:

$$V_2 = W_2 \cdot A_1 + B_2 \tag{3}$$

$$A_2 = \frac{1}{1 + e^{-a \cdot v_2}} \tag{4}$$

$$E = T - A_2 \tag{5}$$

$$MSE = \frac{SE^2}{N} \tag{6}$$

Perhitungan Mundur:

$$D_2 = A_2 \cdot (1 - A_2) \cdot E \tag{7}$$

$$D_1 = A_1 \cdot (1 - A_1) \cdot (W_2 \cdot D_2) \tag{8}$$

$$dW_1 = lr \cdot D_1 \cdot P \tag{9}$$

$$dW_2 = lr \cdot D_2 \cdot P \tag{10}$$

$$dB_1 = lr \cdot D_1 \tag{11}$$

$$dB_2 = lr \cdot D_2 \tag{12}$$

$$W_1 = W_1 + dW_1 \tag{13}$$

$$W_2 = W_2 + dW_2 \tag{14}$$

$$B_1 = B_1 + dB_1 \tag{15}$$

$$B_2 = B_2 + dB_2 \tag{16}$$

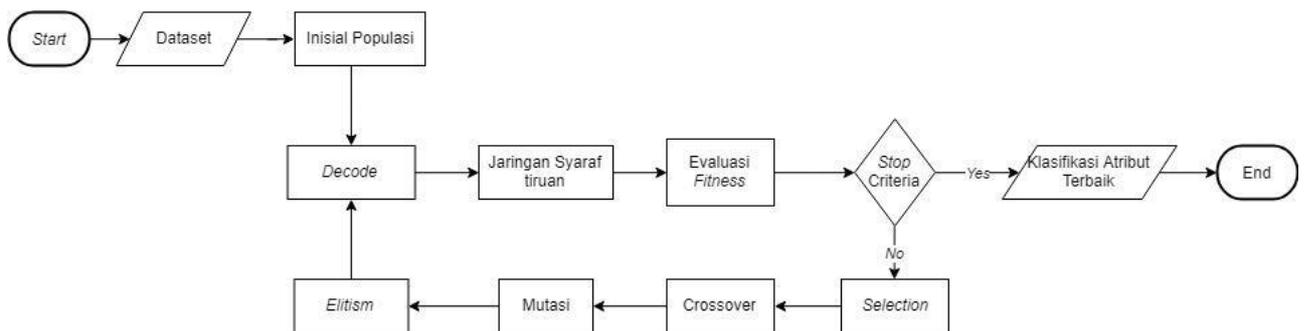
2.2 Algoritma Genetika

Algoritma Genetika (AG) telah terbukti efektif dalam penjelajahan ruang yang kompleks dengan cara adap-tif, yang dipandu oleh mekanisme evolusi biologis seleksi, crossover, dan mutasi [5]. Algoritma ini didasarkan pada kehidupan makhluk hidup yaitu perkembangan generasi dalam sebuah populasi, yang pada akhirnya akan melalui proses seleksi alam atau "siapa yang kuat, dia yang bertahan". AG memungkinkan populasi yang terdiri dari banyak individu yang berevolusi dibawah aturan seleksi yang spesifik agar mendapatkan nilai maksimal dari fitness.

2.3 AG-JST Hybrid

Algoritma genetika dan jaringan syaraf tiruan metode Hybrid bisa di singkat AG-JST Hybrid. JST memiliki beberapa kelemahan atau kerugian seperti waktu pelatihan yang lama, bukan global optimal, dan angka parameter yang besar [3]. Oleh karena itu untuk memperbaiki beberapa kelemahan yang dimiliki JST, maka JST digabungk-an dengan algoritma lain. Algoritma yang sering digabungkan dengan JST adalah AG [3]. Pada tahun 2006, Kim [5] mengusulkan model hybrid AG-JST untuk melakukan contoh dari reduksi dimensi. Pada tahun 2012, AG digunakan untuk menemukan satu set bobot untuk koneksi ke setiap node dalam model JST dan menentukan korelasi kepadatan dalam nanofluida [4]. Sangwan [7] mengusulkan AG-JST untuk pemodelan prediktif dan optimalisasi penentuan parameter untuk meminimalkan kekasaran permukaan. Dan beberapa penerapan AG-JST yang berhasil dilakukan seperti mendeteksi intrusi jaringan dan deteksi penyakit pada penderita kanker [3].

3. Sistem yang Dibangun



Gambar 2. Gambaran sistem umum

Berdasarkan gambaran sistem umum diatas terdapat beberapa proses yang dilakukan untuk mendapatkan hasil akhir. Proses yang terjadi sebagai berikut:

1. Data set yang digunakan pada penelitian ini merupakan data ekspresi gen penyakit tumor usus besar. Data set ini diambil dari Kent Ridge Biomedical Data Set Repository.

Tabel 1. Gambaran data set yang dipakai pada penelitian ini

No.	attribute1	attribute2	attribute3	attribute2000
1	4913.799	5215.048	4288.616	89.45375
2	4985.219	4735.793	4075.123	21.63125
3	5012.02	1383.489	1269.649	48.375
4	9112.373	6824.486	5982.846	32.16625
.	
.	
62	8753.239	8978.134	7777.841	117.19625

Karena range data masing-masing atribut terlalu jauh maka dilakukan normalisasi data,

$$a = \min(\text{data}) \tag{17}$$

$$b = \max(\text{data}) \tag{18}$$

$$ra = 0:9 \tag{19}$$

$$rb = 0:1 \tag{20}$$

$$\text{Datanormalisasi} = \frac{(ra - rb) (\text{data } a)}{(b - a)} + rb \quad (21)$$

Untuk a = mengambil data paling rendah, b = mengambil data paling tinggi, ra = untuk menentukan data paling tinggi nilainya, rb = untuk menentukan data paling rendah nilainya, dan Datanormalisasi = Rumus normalisasi data.

2. Pada sistem yang dibangun, dataset dibagi menjadi 2 yaitu data training dan data testing. Pembagian data dibagi menjadi 2, pertama 70% data training dan 30% data tesing, kedua 80% data training dan 20% data testing. Sebelum dataset diproses lebih lanjut, perlu dilakukan normalisasi data karena range antar data terlalu jauh sehingga sulit untuk melakukan klasifikasi lebih lanjut.
3. Proses pertama pada sistem yaitu decode /generate chromosomes dengan menggunakan angka bit 1 atau 0 secara acak, ukuran populasi dan chromosomes yang digunakan sebanyak 10 individu dan 2000 chromoso-mes disebut populasi 1.
4. Pada tahap kedua membaca chromosomes untuk menemukan variabel mana yang akan dipilih. Setelah mendapatkan variabel masukan, masuk pada tahap JST model untuk menemukan nilai fitness pada setiap populasi.
5. Kemudian pada tahap fitness evaluatin akan mengambil nilai akurasi dari setiap chromosomes dari JST menjadi nilai fitness pada AG.
6. Selanjutnya akan masuk pada kondisi stopping criterion dimana jika jumlah iterasi sudah memenuhi maka proses AG akan selesai, jika belum memenuhi maka proses AG akan terus berlanjut.
7. Jika kondisi iterasi belum memenuhi, maka akan masuk pada tahap selection dimana akan memilih chromo-somes dengan menggunakan teknik roulette wheel disebut populasi 2.
8. Kemudian setelah melalui tahap selection masuk pada tahap crossover, crossover akan dilakukan jika nilai pc pada populasi diatas nilai yang ditentukan maka akan dilakukan crossover begitupun sebaliknya.
9. Tahap selanjutnya yaitu mutation dengan memasukkan gen baru pada populasi dimana jika nilai acak mutasi pada setiap populasi terpenuhi akan dilakukan.
10. Selanjutnya Elitism yaitu memasukan 2 individu yang memiliki nilai fitness terbaik dari populasi 1 (sebelum masuk selection) dan menukarnya dengan individu yang memiliki nilai fitness terendah dari populasi 2 (setelah mutation).
11. Kemudian kembali ke tahap dua yaitu decode dan akan terus berulang hingga kriteria sudah terpenuhi.

4. Evaluasi

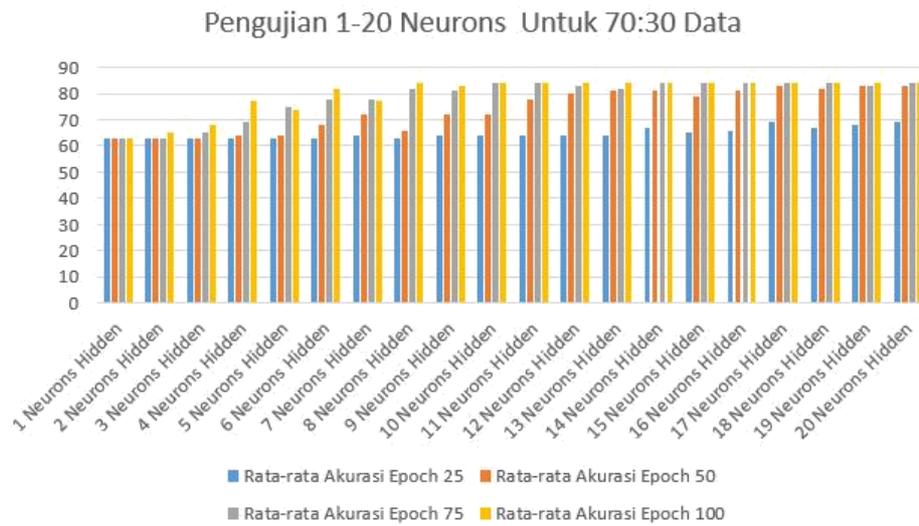
4.1 Skenario Pengujian

Pada penelitian ini terdapat 3 skenario yang dilakukan skenario pertama menentukan epoch dan neurons hidden pada metode jaringan syaraf tiruan (JST), skenario kedua menentukan PC (peluang crossover) dan PM (peluang mutasi) pada metode algoritma genetika(AG), skenario ketiga membandingkan performansi metode JST dengan metode AG-JST.

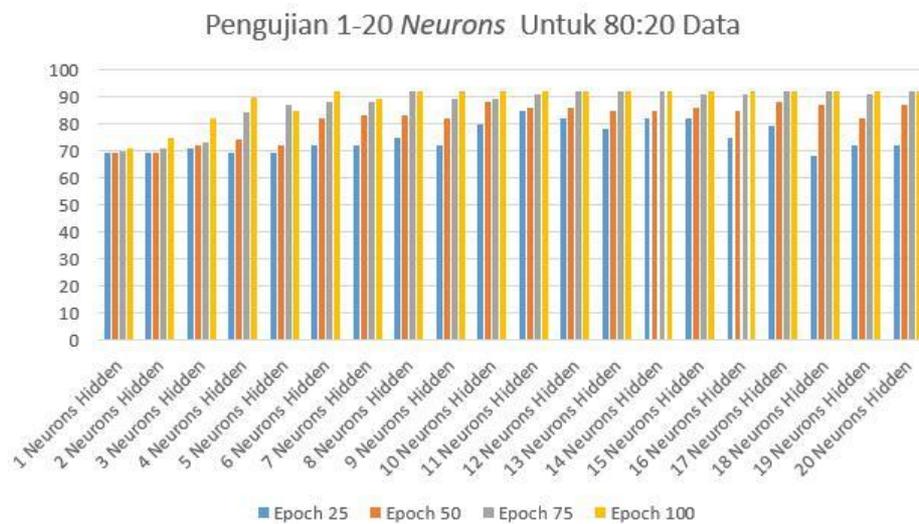
4.1.1 Menentukan Epoch dan Neuron Hidden pada JST

Pengujian ini digunakan untuk mengetahui pengaruh jumlah epoch dan neurons hidden terhadap hasil akurasi. Jumlah epoch yang digunakan ada 4 yaitu 25, 50, 75, dan 100 untuk kedua pembagian data set. Sedangkan jumlah neurons hidden yang di gunakan ada 20 dari 1 sampai 20. Karena JST setiap di running akurasinya berubah-ubah maka pada pengujian ini dilakukan 10x JST, untuk mendapatkan rata-rata akurasi.

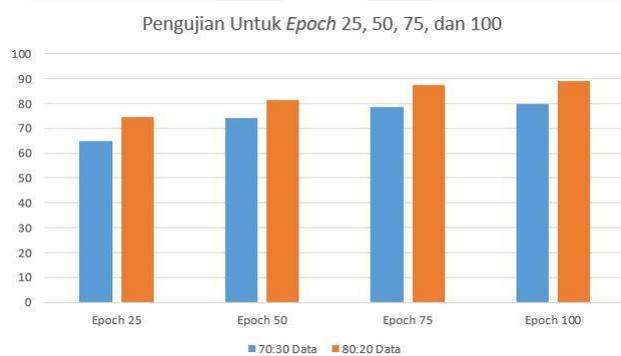
Dilihat dari kedua tabel diatas, semakin banyak neurons hidden rata-rata akurasi juga semakin tinggi. Dilan-jutkan pengujian untuk menentukan epoch.



Gambar 3. Pengujian Neurons Hidden pada rata-rata akurasi untuk 70:30 data



Gambar 4. Pengujian Neurons Hidden pada rata-rata akurasi untuk 80:20 data



Gambar 5. Pengujian Epoch pada rata-rata akurasi

Berdasarkan ketiga tabel di atas terlihat bahwa semakin banyak epoch yang dilakukan maka hasil akurasi akan semakin tinggi. Begitu juga dengan neurons hidden, semakin banyak neurons hidden akurasi juga akan semakin tinggi dan konsisten.

4.1.2 Menentukan PC (peluang crossover) dan PM (peluang mutasi)

Pengujian ini digunakan untuk mengetahui pengaruh PC dan PM terhadap hasil akurasi dari JST. Jumlah PC yang digunakan ada 3 yaitu 0.5, 0.6, dan 0.7. Sama dengan PC, PM yang digunakan ada 3 juga yaitu 0.1, 0.05, dan 0.015.

Tabel 2. Hasil Pengujian PM dan PC

epoch(100) neurons hidden (20)	70:30 Data	80:20 Data
PC (0,5) & PM (0,1)	84	92
PC (0,5) & PM (0,05)	84	92
PC (0,5) & PM (0,015)	84	92
PC (0,6) & PM (0,1)	84	92
PC (0,6) & PM (0,05)	84	92
PC (0,6) & PM (0,015)	84	92
PC (0,7) & PM (0,1)	84	92
PC (0,7) & PM (0,05)	84	92
PC (0,7) & PM (0,015)	84	92

Dilihat dari tabel diatas bahwa PM dan PC tidak mempengaruhi akurasi dari AG-JST maka dilakukan pengujian lebih lanjut lagi dengan menggunakan 25 epoch untuk melihat apakah PC dan PM mempengaruhi akurasi dari AG-JST.

Tabel 3. Hasil Pengujian PM dan PC dengan 25 epoch

epoch(25) neurons hidden (20)	70:30 Data	80:20 Data
PC (0,5) & PM (0,1)	84%	92%
PC (0,5) & PM (0,05)	79%	92%
PC (0,5) & PM (0,015)	84%	92%
PC (0,6) & PM (0,1)	84%	92%
PC (0,6) & PM (0,05)	89%	92%
PC (0,6) & PM (0,015)	84%	92%
PC (0,7) & PM (0,1)	84%	100%
PC (0,7) & PM (0,05)	79%	100%
PC (0,7) & PM (0,015)	84%	92%

Berdasarkan nilai akurasi dari tabel 5 Parameter epoch = 25, neuronshidden = 20, PC = 0; 6 dan PM = 0; 05 adalah parameter terbaik untuk 70:30 data menghasilkan akurasi 89%. Sedangkan parameter epoch = 25, neuronshidden = 20, PC = 0; 7 dan PM = 0; 1 adalah parameter terbaik untuk 80:20 data menghasilkan akurasi 100%.

4.1.3 Membandingkan performansi metode JST dengan metodr AG-JST

Pengujian ini digunakan untuk mengetahui JST lebih optimal digabung dengan AG. Menjadi algoritma hybrid yaitu metode algoritma genetika dan jaringan syaraf tiruan (AG-JST).

Tabel 4. Hasil Pengujian epoch = 25, neuronshidden = 20, PC = 0; 6 dan PM = 0; 05 dengan 70:30 data

70:30 Data	JST	AG-JST
Akurasi Terbaik	84%	89%
Atribut ekspresi gen	2000	1006

Tabel 5. Hasil Pengujian epoch = 25, neuronshidden = 20, PC = 0; 7 dan PM = 0; 1 dengan 80:20 data

80:20 Data	JST	AG-JST
Akurasi Terbaik	92%	100%
Atribut ekspresi gen	2000	1001

Dilihat dari tabel 6 dan tabel 7 akurasi akan naik jika JST digabungkan dengan AG. AG-JST mereduksi 49,7% atribut pada pembagian 70:30 data dan mereduksi 49,9% atribut pada pembagian 80:20 data.

5. Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, maka dapat disimpulkan bahwa :

1. Metode hybrid algoritma genetika dan jaringan syaraf tiruan dapat digunakan untuk mereduksi dimensi data ekspresi gen pada kasus diagnosis tumor usus besar.
2. Berdasarkan pengujian di atas untuk pembagian 70 data training : 30 data testing dengan parameter epoch = 25, neuronshidden = 20, PC = 0; 6 dan PM = 0; 05 mendapatkan akurasi terbaik yaitu 89% dan mereduksi 49,7% atribut. Untuk pembagian 80 data training : 20 data testing menggunakan parameter epoch = 25, neuronshidden = 20, PC = 0; 7 dan PM = 0; 1 mendapatkan akurasi terbaik yaitu 100% dan mereduksi 49,9% atribut.
3. Algoritma genetika bisa membantu meningkatkan akurasi jaringan syaraf tiruan karena saat sistem JST saja tidak dapat mendapatkan hasil yang sangat maksimal. AG dapat membantu meningkatkan hasil yang lebih baik lagi karena AG dapat membantu mengoptimisasi JST dengan melakukan seleksi atribut-atribut yang penting untuk digunakan.

Daftar Pustaka

- [1] A. Abraham. Artificial neural networks. handbook of measuring system design, 2005.
- [2] S. Agatonovic-Kustrin and R. Beresford. Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 22(5):717–727, 2000.
- [3] M. Inthachot, V. Boonjing, and S. Intakosum. Artificial neural network and genetic algorithm hybrid intelligence for predicting thai stock price index trend. Computational intelligence and neuroscience, 2016, 2016.
- [4] H. Karimi and F. Yousefi. Application of artificial neural network–genetic algorithm (ann–ga) to correlation of density in nanofluids. Fluid Phase Equilibria, 336:79–83, 2012.
- [5] K.-j. Kim. Artificial neural networks with evolutionary instance selection for financial forecasting. Expert Systems with Applications, 30(3):519–526, 2006.
- [6] M. Naiyar, Y. Asim, and A. Shahid. Automated colon cancer detection using structural and morphological features. In Frontiers of Information Technology (FIT), 2015 13th International Conference on, pages 240– 245. IEEE, 2015.
- [7] K. S. Sangwan, S. Saxena, and G. Kant. Optimization of machining parameters to minimize surface roughness using integrated ann-ga approach. Procedia CIRP, 29:305–310, 2015.
- [8] U. N. Wisesty, R. S. Warastri, and S. Y. Puspitasari. Leukemia and colon tumor detection based on microarray data classification using momentum backpropagation and genetic algorithm as a feature selection method. In Journal of Physics: Conference Series, volume 971, page 012018. IOP Publishing, 2018.
- [9] A. M. Zamani, B. Amaliah, and A. Munif. Implementasi algoritma genetika pada struktur backpropagation neural network untuk klasifikasi kanker payudara. Jurnal Teknik ITS, 1(1):A222–A227, 2012.