



Peningkatan Akurasi Klasifikasi *Backpropagation* Menggunakan *Artificial Bee Colony* dan *K-NN* Pada Penyakit Jantung

Pandito Dewa Putra, Sukemi*, Dian Palupi Rini

Fakultas Ilmu Komputer, Magister Ilmu Komputer, Universitas Sriwijaya, Palembang, Indonesia

Email: ¹panditodewaputra@gmail.com, ^{2*}sukemi@ilkom.unsri.ac.id, ³dprini@unsri.ac.id

Email Penulis Korespondensi: sukemi@ilkom.unsri.ac.id

Abstrak—Penyakit jantung menduduki peringkat sebagai penyebab utama kematian di dunia terhitung sekitar 17,3 juta kematian per tahun dengan berbagai faktor penyebabnya, seperti tekanan darah tinggi, diabetes, fluktuasi kolesterol, kelelahan dan banyak lagi lainnya yang banyak terkoleksi dalam dataset. Dataset penyakit jantung yang diterapkan yaitu cleveland heart disease dengan 14 atribut data. Metode yang diterapkan dalam penelitian ini yaitu dengan menggunakan algoritma *Backpropagation* dalam mengklasifikasi penyakit jantung, dimana akan dikombinasikan dengan algoritma *Artificial Bee Colony* dan *k-Nearest Neighbor* untuk pemilihan fitur atau pemilihan atribut sebab dengan teknik ini dapat meningkatkan akurasi model pengklasifikasi yang dihasilkan sebesar 94,23%.

Kata Kunci: Penyakit Jantung; Cleveland; *Backpropagation*; *Artificial Bee Colony*; *K-NN*

Abstract—Heart disease has ranked as the leading cause of death in the world, accounting for around 17.3 million deaths per year with some causes, as high blood pressure, diabetes, cholesterol fluctuation, fatigue, and some others which is collected on dataset. Heart disease dataset that was applied is cleveland heart disease with fourteen (14) data attribute. The method that was applied in this research was using *Backpropagation* algorithm on heart disease classifying, where will be combined *Artificial Bee Colony* and *k-Nearest Neighbor* algorithm for features or attribute choose due to this technique can increase classifier model accuracy which is produced as much as 94,23%.

Keywords: Heart Disease; Cleveland; *Backpropagation*; *Artificial Bee Colony*; *K-NN*

1. PENDAHULUAN

Penyakit jantung merupakan salah satu penyakit yang paling banyak diderita saat ini, karena berbagai faktor penyebabnya, seperti tekanan darah tinggi, diabetes, fluktuasi kolesterol, kelelahan dan banyak lagi lainnya [1]. Saat ini, dalam industri perawatan kesehatan, analisis data dapat menyelamatkan nyawa dengan meningkatkan diagnosis medis. Kemudian dengan berbagai perkembangan besar dalam rekayasa perangkat lunak, berbagai alat penambangan data tersedia bagi para peneliti, dan digunakan untuk melakukan studi dan eksperimen [2]. Penelitian ini fokus pada prediksi penyakit jantung menggunakan teknik data mining. Motivasi dari penelitian ini adalah perkiraan yang diberikan oleh WHO. Berdasarkan perkiraan WHO pada tahun 2030, hampir 23,6 juta orang akan menjadi penyebab penyakit Jantung [3]. Penyakit jantung menduduki peringkat sebagai penyebab utama kematian di dunia terhitung sekitar 17,3 juta kematian per tahun. Hal ini membawa kebutuhan yang mendesak untuk menggunakan teknik data mining di industri medis yang cukup efisien untuk menemukan pola dalam data Penyakit Jantung. Peningkatan akurasi dan kompleksitas waktu untuk mendeteksi penyakit jantung pada pasien berdasarkan beberapa parameter penting menjadi perhatian utama para ilmuwan data di seluruh dunia. Sistem prediksi penyakit jantung yang efisien memiliki kemampuan untuk menggantikan tes medis yang memakan waktu dan mahal yang harus dilalui pasien [4]. Penambangan data medis digunakan untuk mendiagnosis penyakit dengan menggunakan teknik pembelajaran mesin, sehingga praktisi medis dapat memperoleh masukan berharga tentang pasien penyakit jantung, yang memungkinkan mereka untuk memberikan perawatan yang tepat kepada pasien. Hal ini memotivasi kami untuk mengembangkan model ini untuk membantu para dokter dalam mendiagnosis pasien penyakit jantung dengan teknik data mining klasifikasi menggunakan fitur selection [5].

Pemilihan fitur adalah istilah umum untuk proses otomatis yang hanya memilih fitur yang paling penting dan relevan dari suatu objek. Fitur yang berlebihan atau tidak relevan akan dibuang atau diabaikan sebelum memiliki kesempatan untuk menurunkan performa sistem klasifikasi. Penggunaan pemilihan fitur dengan demikian menghasilkan akurasi yang lebih tinggi dan mengurangi biaya komputasi pemodelan secara keseluruhan [6]. Pemilihan fitur atau pemilihan atribut telah menjadi fokus penting dalam banyak aplikasi penelitian, sebab dengan teknik ini dapat meningkatkan akurasi model pengklasifikasi yang dihasilkan. Selain itu, penting untuk menemukan subset fitur prediktif yang relevan. Misalnya, seorang dokter mungkin mengambil keputusan tentang kekritisannya suatu penyakit berdasarkan klasifikasi yang dilakukan dengan menggunakan fitur yang dipilih. Kemudian akurasi saat proses klasifikasi terhadap suatu penyakit ditingkatkan dengan mengoptimalkan pemilihan fitur [7].

Beberapa metode populer yang telah digunakan oleh beberapa peneliti dalam mengklasifikasi penyakit jantung (*cleveland heart disease*) adalah algoritma *CART* dengan akurasi yang dicapai 83% [8], algoritma *naïve bayes* dengan akurasi mencapai 83% dan *support vector machine* yang mencapai akurasi 84% [9], algoritma *K-NN* (IBK) mencapai akurasi 79% [10], kemudian algoritma *random forest* mencapai akurasi 86% dan *random tree* yang mencapai 79% [11]. Kemudian terdapat juga beberapa metode kombinasi dalam mengklasifikasi penyakit jantung (*cleveland heart disease*) yaitu algoritma kombinasi *CSF-PSO-MLP* mencapai akurasi 90% [12], algoritma kombinasi *GA-SVM-SMO-RBF* mencapai akurasi 88% [13]. Metode *backpropagation* sendiri mencapai



akurasi 85% [14]. Kemudian beberapa kombinasi algoritma *backpropagation* yaitu *DE-Backpropagation* mencapai akurasi 80%, algoritma *PSO-Backpropagation* mencapai akurasi 83%, algoritma *DEGI-Backpropagation* mencapai 87% [15]. Algoritma metaheuristik yang diterapkan pada penelitian ini yaitu *artificial bee colony*, dimana *ABC-SVM (Cleveland dataset)* mencapai akurasi 86.76 % [16], kemudian *ABC-kNN (autism dataset)* mencapai akurasi 85% [17]. Penelitian ini akan berfokus pada pemilihan fitur penting sambil membuang fitur yang berlebihan dengan algoritma *artificial bee colony* dan algoritma *K-Nearest Neighbor* untuk mendapatkan nilai *fitness* dan atribut-atribut terbaik agar terjadi peningkatan akurasi optimal pada algoritma *backpropagation* dalam mengklasifikasi penyakit jantung (*cleveland heart disease*).

2. METODOLOGI PENELITIAN

Metodologi yang diterapkan dalam penelitian ini yaitu dengan menerapkan algoritma *Backpropagation* yang akan dikombinasikan dengan algoritma *Artificial Bee Colony* dan *k-NN*.

2.1 Artificial Bee Colony (ABC)

Algoritma *ABC* dapat dikategorikan menjadi tiga kelompok: lebah yang dipekerjakan, lebah pengamat dan lebah pengintai. Pada algoritma *ABC*, sumber makanan merupakan solusi dalam masalah pengoptimalan. Kualitas suatu sumber pangan dapat diukur dengan fungsi kesesuaian yang berkaitan dengan lokasi sumber pangan tersebut. Proses pencarian sumber makanan oleh lebah dapat digunakan untuk menemukan solusi yang optimal di ruang solusi [18].

Lebah pengamat buatan memilih sumber makanan tergantung pada nilai probabilitas yang terkait dengan sumber makanan tersebut :

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \quad (1)$$

Untuk menghasilkan posisi makanan kandidat dari yang lama dalam ingatan, *ABC* menggunakan :

$$v_{ij} = z_{ij} + \phi_{ij}(z_{ij} - z_{kj}) \quad (2)$$

Pada *ABC*, ini disimulasikan dengan menghasilkan posisi secara acak dan menggantinya dengan yang ditinggalkan. Dalam *ABC*, asalkan suatu posisi tidak dapat ditingkatkan lebih lanjut melalui jumlah siklus yang telah ditentukan, maka sumber pangan tersebut diasumsikan telah ditinggalkan. Nilai jumlah siklus yang telah ditentukan merupakan parameter kontrol penting dari algoritma *ABC*, yang disebut "batas" untuk pengabaian. Kemudian lebah pengintai menemukan sumber makanan baru :

$$z_i^j = z_{min}^j + rand(0,1)(z_{max}^j - z_{min}^j) \quad (3)$$

Setelah masing-masing kandidat sumber posisi v_{ij} diproduksi dan kemudian dievaluasi oleh lebah tiruan, kinerjanya dibandingkan dengan yang lama. Jika sumber makanan baru memiliki nektar yang sama atau lebih baik dari sumber lama, itu diganti dengan yang lama di memori. Jika tidak, yang lama disimpan dalam memori. Dengan kata lain, mekanisme seleksi greedy digunakan sebagai operasi seleksi antara yang lama dan yang kandidat [18].

2.2 K-NN

Dalam algoritma *KNN*, sampel pendidikan dijelaskan dengan fitur n-dimensi numerik. Setiap sampel ditunjukkan oleh titik dalam fitur dimensi-n numerik. Jadi, semua sampel pendidikan disimpan dalam pola pendidikan n-dimensi [19]. Ketika ada sampel yang tidak diketahui yang perlu diklasifikasikan, algoritma ini akan mencari pola eksperimental untuk sampel K yang paling dekat dengan sampel yang tidak diketahui. Pendekatan ini (kedekatan) ditentukan oleh jarak Euclidean. Jika dua titik tersebut menyertakan $X_1 = (x_{11}, x_{12}, x_{13}, \dots, x_{1n})$ dan $X_2 = (x_{21}, x_{22}, x_{23}, \dots, x_{2n})$, maka jarak Euclidean persamaan tersebut adalah sebagai berikut :

$$dist(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (4)$$

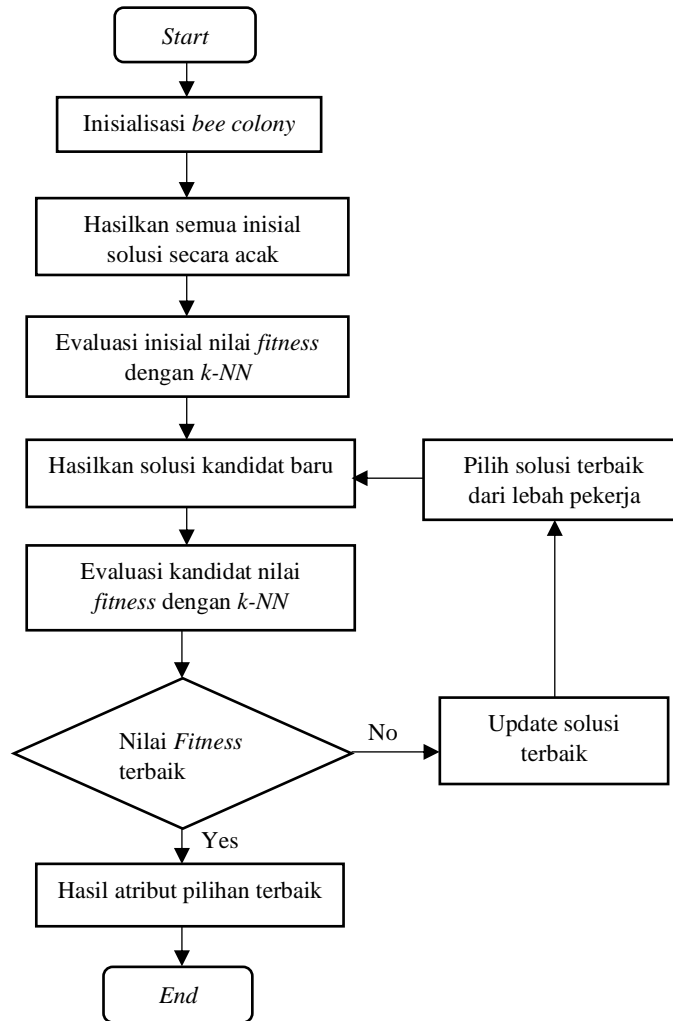
Misalnya, jarak *Euclidean* digunakan untuk data numerik dan jarak *Hamming* untuk data kategorikal. Setelah semua jarak dihitung, mereka diurutkan. Kategori data pengujian ditentukan berdasarkan kategori mayoritas dari *k-neighbours* terdekat dengan item data pengujian. Langkah ini akan diulang sampai semua data pengujian diklasifikasikan [17].

2.3 Algoritma ABC - kNN

Metode yang diusulkan menerapkan algoritma *ABC* dengan *kNN*. *kNN* digunakan untuk evaluasi *fitness* untuk memperkirakan nilai *fitness* dari sumber makanan *ABC*. Setelah lebah yang dipekerjakan dan lebah pengamat menghasilkan kandidat sumber makanan baru, yang merupakan himpunan bagian dari fitur terpilih, *kNN* dilakukan untuk mengevaluasi ketepatan klasifikasi (fungsi objektif) dari kandidat sumber makanan baru. Akurasi digunakan sebagai kriteria untuk memilih subset fitur yang optimal [17]. Dalam penelitian ini menetapkan jumlah *nearest*



neighbors dan k menjadi satu sehingga setiap item data masukan akan diklasifikasikan ke kelas data tetangga terdekat. Metode *leave-one-out cross validation* (LOOCV) digunakan untuk mengevaluasi akurasi klasifikasi k -NN. LOOCV akan menggunakan satu item data sampel dari dataset asli sebagai data validasi; sisa item akan digunakan sebagai data pelatihan. Proses pelatihan ini diulangi sampai semua item data dalam dataset telah digunakan sebagai item data validasi. Kami menerapkan metode ABC-kNN untuk pemilihan fitur dan masalah klasifikasi yang dijelaskan di atas. Tujuannya adalah untuk menemukan subset fitur yang optimal dari kumpulan fitur asli sementara subset yang dihasilkan memenuhi akurasi klasifikasi tertinggi. Pemilihan fitur dan proses klasifikasi ABC-kNN diilustrasikan pada Gambar 1.



Gambar 1. Flowchart Algoritma ABC-kNN

2.4 Backpropagation (BP)

Algoritma *Backpropagation* adalah metode pembelajaran yang diawasi untuk melatih jaringan saraf tiruan multilayer [20] dengan tahapan seperti pada Gambar 2. Pertimbangan arsitektur jaringan neural yang terdiri dari beberapa lapisan tersembunyi. Jika kita menganggap neuron milik lapisan tersembunyi atau keluaran, aktivasi unit-unit ini, dilambangkan dengan y_i , dapat ditulis sebagai:

$$y_i = g \left(\sum_{j=1}^L w_{ij} \cdot s_j \right) = g(h), \tag{5}$$

Dimana w_{ij} adalah bobot sinaptik antara neuron i pada lapisan saat ini dan neuron pada lapisan sebelumnya dengan aktivasi s_j . Pada persamaan sebelumnya, h adalah potensial sinaptik dari suatu neuron, yang mencakup fungsi aktivasi tipe sigmoid. Tujuan dari algoritma pembelajaran yang diawasi BP adalah untuk meminimalkan perbedaan antara keluaran yang diberikan (target) untuk satu set data masukan dan keluaran dari jaringan, diukur dengan fungsi *error* (E) yang bergantung pada nilai bobot sinaptik, dan itu dapat didefinisikan sebagai:

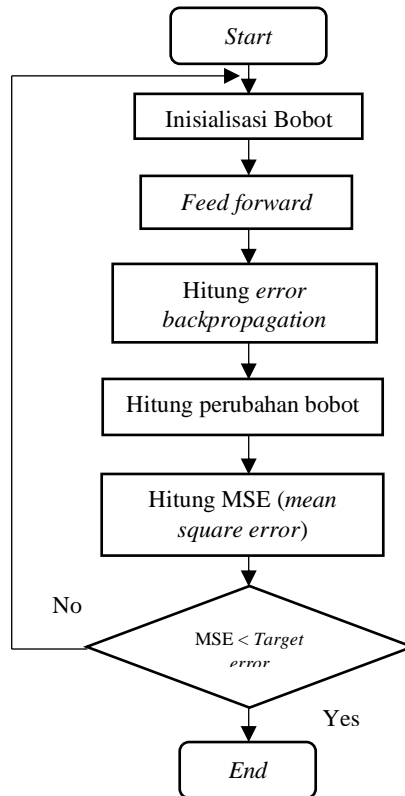
$$E = \frac{1}{2} \sum_{k=1}^P \sum_{i=1}^M (z_i(k) - y_i(k))^2, \tag{6}$$



Dimana jumlah pertama ada pada pola p dari kumpulan data dan jumlah kedua ada pada neuron keluaran M . $z_i(k)$ adalah nilai target untuk keluaran neuron i untuk pola k , dan $y_i(k)$ adalah keluaran respons yang sesuai dari jaringan. Bobot sinaptik antara dua lapisan terakhir neuron diperbarui sebagai:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta [z_i(k) - y_i(k)]g'_i(h_i)s_j(k), \tag{7}$$

Dimana η adalah kecepatan pembelajaran, g' adalah turunan dari fungsi sigmoid dan h adalah potensial sinaptik yang telah ditentukan sebelumnya, sedangkan bobot lainnya dimodifikasi sesuai dengan persamaan serupa dengan pengenalan sekumpulan nilai yang disebut "delta" (δ), yang menyebarkan kesalahan dari lapisan terakhir ke lapisan dalam.



Gambar 2. Flowchart Backpropagation

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Penelitian ini akan menggunakan *cleveland heart disease dataset* yang bersumber dari *uci machine learning repository*. Kemudian akan digunakan 14 atribut dengan 303 record data didalamnya.

Tabel 1. Cleveland Heart Disease

Kode Atribut	Atribut	Keterangan
A1	Age	Umur pasien
A2	Sex	1 = male; 0 = female
A3	Cp (chest pain)	Value 1: typical angina; Value 2: atypical angina; Value 3: non-anginal pain; Value 4: asymptomatic
A4	trestbps	tekanan darah
A5	chol	kolesterol
A6	fbs	Gula darah berdasarkan nilai 1 = true; 0 = false
A7	restcg	Elektrokardiografi berdasarkan 0 = normal; 1 = memiliki kelainan gelombang ST-T; 2 = menunjukkan hipertrofi ventrikel kiri



Kode Atribut	Atribut	Keterangan
A8	<i>thalach</i>	denyut jantung
A9	<i>exang</i>	induksi angina 1 = yes; 0 = no
A10	<i>oldpeak</i>	depresi ST
A11	<i>slope</i>	kemiringan segmen ST puncak berdasarkan nilai 1 = <i>upsloping</i> ; 2 = flat; 3 = <i>downsloping</i>
A12	<i>ca</i>	Jumlah pembuluh darah utama 0-3 diwarnai dengan <i>fluoroskopi</i>
A3	<i>thai</i>	memiliki nilai 3 = normal; 6 = cacat tetap; 7 = cacat yang dapat dibalik
A14 (class)	<i>num</i>	memiliki nilai 0 = Tidak Mengalami Penyakit Jantung; 1 = Mengalami Penyakit Jantung

Dataset yang akan digunakan yaitu *cleveland heart disease*, dimana *database* nya berisi 76 atribut, tetapi semua publikasi merujuk pada penggunaan 14 di antaranya, yang 13 terkait dengan faktor penyakit dan 1 kasus terkait dengan ada tidaknya penyakit jantung pada orang yang menjalani 13 tes sebelumnya [21] seperti pada tabel 1. Kemudian atribut *num* terdapat nilai 0 mewakili tidak adanya penyakit jantung dan semua nilai dari 1 hingga 4 mewakili pasien dengan penyakit jantung, dimana penskalaan mengacu pada tingkat keparahan penyakit (4 menjadi yang tertinggi) [22]. Jadi, attribute *num* dirancang memiliki dua tipe keluaran: 0 menunjukkan bahwa penyakit jantung tidak ada dan 1 menunjukkan bahwa penyakit jantung ada [23].

3.2 Data Preprocessing

Dataset cleveland heart disease mengandung 303 *record* data namun terdapat 6 baris data yang memiliki *missing value* sehingga data *missing* tersebut akan dibuang dan hanya akan menggunakan 297 baris data yang lengkap. Data training dan Data testing akan dipecah secara otomatis dengan *k-fold cross validation* yang akan diberikan nilai bobot sebanyak ($k=5$) sehingga rasio untuk data training-testing 80% : 20% [24].

3.3 Hasil Pengujian

Hasil penelitian akan menguji algoritma *Backpropagation* tanpa kombinasi pada *cleveland heart disease*. Kemudian pengujian selanjutnya yaitu dengan menerapkan algoritma *Backpropagation* yang akan dikombinasikan dengan algoritma *ABC-kNN* sebagai seleksi fitur pada *cleveland heart disease*. Kemudian hasil akhirnya akan di lihat perbandingan 2 pengujian tersebut.

Pengujian Pertama yaitu dengan menggunakan algoritma *Backpropagation* saja yang akan dibandingkan dengan algoritma kombinasi *Backpropagation* dengan *ABC-kNN*, kemudian atribut dari *cleveland heart disease dataset* digunakan sebanyak 14 atribut dengan 297 *record* data yang digunakan setelah dilakukan proses *preprocessing*. Dengan menerapkan *k-fold validation* yaitu menentukan bobot nilai ($k=5$) yang akan memecah data menjadi 80:20 untuk data *training* dan *testing* secara acak dan terotomatisasi.

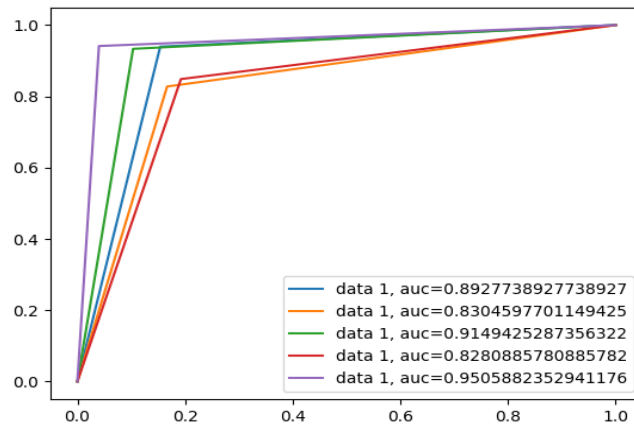
Tabel 2. *Backpropagation (Training)*

No	Class	Precision	Recall	F1-Score	Support	Accuracy
1	0	83	83	83	30	83
	1	83	83	83	29	
2	0	93	90	91	29	92
	1	90	93	92	30	
3	0	81	81	81	26	83
	1	85	85	85	33	
4	0	92	96	94	25	95
	1	97	94	96	34	

Tabel 3. *Backpropagation (Testing)*

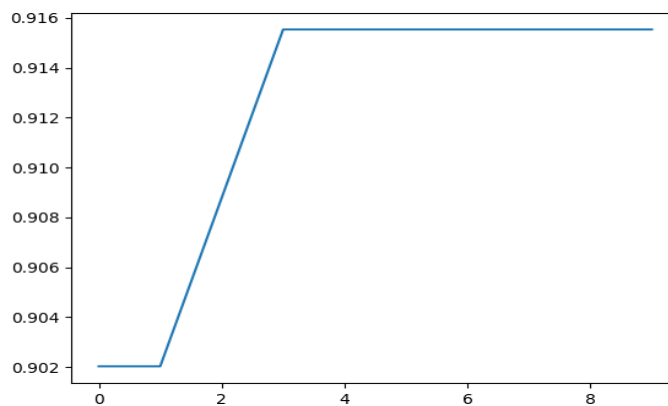
No	Class	Precision	Recall	F1-Score	Support	Accuracy
1	0	92	85	88	26	90
	1	89	94	91	33	

Hasil dari algoritma *Backpropagation* terhadap *cleveland heart disease* yaitu 83%, 92%, 83%, 95%, dan 90%, dimana akurasi rata-rata yang didapat dengan 5-fold validasi yaitu mencapai akurasi 88,47%. Selanjutnya untuk hasil AUC yang didapat dari algoritma *Backpropagation* yaitu dengan nilai AUC tertinggi mencapai 95% dan nilai AUC terendah mencapai 82% seperti pada gambar 3.



Gambar 3. Hasil AUC Algoritma *Backpropagation*

Proses seleksi atribut dilakukan terhadap *cleveland heart disease* dengan bertujuan agar membuang atribut-atribut yang dapat membuat akurasi menjadi rendah. Proses seleksi atribut ini akan menggunakan algoritma *ABC-kNN*, dimana algoritma *ABC* untuk pemilihan atribut secara random dan algoritma *k-NN* untuk mencari nilai *fitness* dari algoritma *ABC*. Kemudian terpilihlah beberapa atribut dengan jumlah yang lebih sedikit dari *cleveland heart disease dataset* agar terjadi peningkatan akurasi terhadap algoritma *Backpropagation* pada saat proses klasifikasi.



Gambar 4. Hasil Seleksi fitur Algoritma *ABC-kNN*

Cleveland heart disease memiliki 13 atribut dan 1 atribut *class*. Hasil yang didapat dari seleksi fitur dengan menggunakan algoritma *ABC* dan *k-NN* yaitu terpilihlah atribut yang akan digunakan [0 1 1 0 0 1 1 0 0 0 1 1] dimana biner angka 0 adalah atribut yang dibuang dan biner angka 1 adalah atribut yang digunakan. Atribut yang terpilih yaitu atribut A2, A3, A6, A7, A12 dan A13. Hasil nilai *fitness* terbaik dari seleksi fitur menggunakan algoritma *ABC-kNN* yaitu mencapai 91% seperti pada gambar 4. Kemudian selanjutnya melakukan pengujian untuk kombinasi algoritma *Backpropagation + ABC-kNN*.

Tabel 4. *Backpropagation + ABC-kNN (Training)*

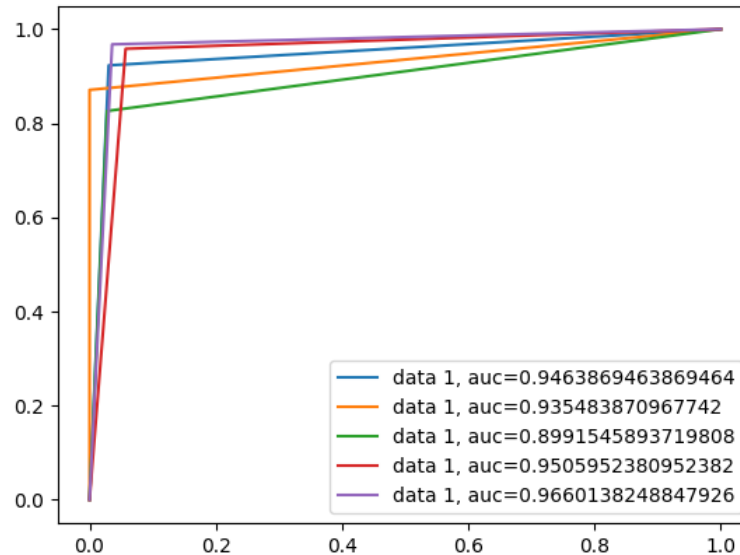
No	Class	Precision	Recall	F1-Score	Support	Accuracy
1	0	94	97	96	33	95
	1	96	92	94	26	
2	0	88	100	93	28	93
	1	100	87	93	31	
3	0	90	97	93	36	92
	1	95	83	88	23	
4	0	96	96	96	28	97
	1	97	97	97	31	

Tabel 5. *Backpropagation + ABC-kNN (Testing)*

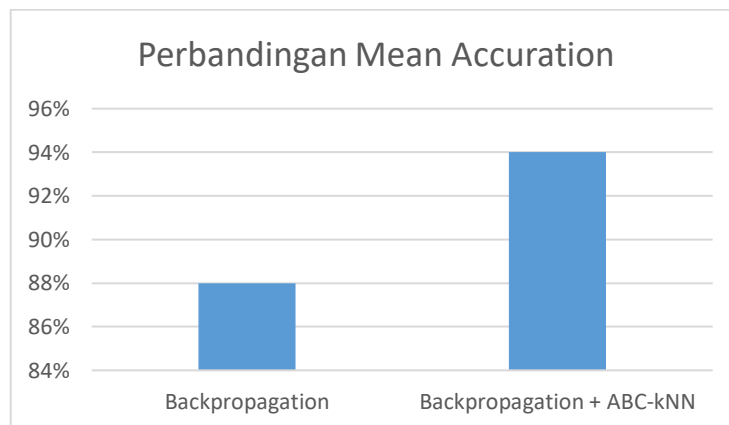
No	Class	Precision	Recall	F1-Score	Support	Accuracy
1	0	97	94	96	35	95
	1	92	96	94	24	



Hasil dari algoritma *Backpropagation* + *ABC-kNN* dengan menerapkan *k-fold cross validation* dengan nilai $k=5$ yaitu 95%, 93%, 92%, 97%, dan 95%. Kemudian *Mean Accuracy* dari *5-fold cross validation* tersebut mencapai: 94,23%. Selanjutnya untuk hasil AUC yang didapat dari algoritma *Backpropagation* + *ABC-kNN* yaitu nilai AUC tertinggi mencapai 97% sedangkan AUC terendah mencapai 92%.



Gambar 5. Hasil AUC Algoritma *Backpropagation* + *ABC-kNN*



Gambar 6. Perbandingan Akurasi

Seleksi fitur yang telah diterapkan pada penelitian ini menggunakan algoritma *ABC-k-NN* telah menunjukkan peningkatan akurasi saat proses pengklasifikasi penyakit jantung dengan algoritma *Backpropagation*. Algoritma *Backpropagation* tanpa kombinasi mencapai akurasi optimal 88% sedangkan algoritma *Backpropagation* yang dikombinasikan dengan *ABC-kNN* untuk seleksi fitur berhasil mencapai akurasi yang lebih baik yaitu sebesar 94%.

4. KESIMPULAN

Kesimpulan dari penelitian ini yaitu telah terjadi peningkatan akurasi algoritma *Backpropagation* yang cukup signifikan setelah dilakukan nya proses seleksi atribut dengan menggunakan algoritma *ABC-kNN* terhadap *cleveland heart disease dataset*. Hasil nilai *fitness* tertinggi saat proses seleksi fitur *ABC-kNN* yaitu mencapai 91%. Kemudian proses seleksi fitur tersebut telah mendapatkan 6 atribut (A2, A3, A6, A7, A12 dan A13) pilihan terbaik yang selanjutnya akan digunakan untuk diklasifikasi dengan menggunakan algoritma *Backpropagation*. Hasil akurasi dari kombinasi algoritma *Backpropagation* + *ABC-kNN* yaitu mencapai *mean akurasi* 94%. Kombinasi algoritma *Backpropagation* + *ABC-kNN* telah menggungguli dari kombinasi *Backpropagation* dengan algoritma variasi lainnya, seperti *Backpropagation* + *DEGI* mencapai akurasi 87%, *Backpropagation* + *differential evolution* mencapai akurasi 80%, *Backpropagation* + *Partical Swarm Optimization* mencapai akurasi 83%, dan *Backpropagation* saja tanpa kombinasi mencapai akurasi *mean* 88%. Dengan hasil penelitian ini dapat membuktikan bahwa algoritma *Artificial Bee Colony* dan *k-NN* sangat cocok untuk dijadikan sebagai algoritma



optimasi untuk digunakan dalam melakukan seleksi fitur atribut pada *dataset* penelitian. Saran untuk penelitian selanjutnya yaitu diharapkan dapat mengkombinasikan algoritma *artificial bee colony* dan *k-NN* dengan algoritma *Artificial Neural Network* lainnya seperti algoritma *Feed Forward Propagation*, *Single Layer Perceptron*, *Multi Layer Perceptron*, *Radial Basis Function (RBF)*, dan *Convolutional Neural Network* dalam mengklasifikasi penyakit jantung.

REFERENCES

- [1] K. M. Almustafa, "Prediction of heart disease and classifiers' sensitivity analysis," *BMC Bioinformatics*, vol. 21, no. 1, pp. 1–18, 2020, doi: 10.1186/s12859-020-03626-y.
- [2] I. Tougui, A. Jilbab, and J. El Mhamdi, "Heart disease classification using data mining tools and machine learning techniques," *Health Technol. (Berl.)*, vol. 10, no. 5, pp. 1137–1144, 2020, doi: 10.1007/s12553-020-00438-1.
- [3] P. Sharma and K. Saxena, "Application of fuzzy logic and genetic algorithm in heart disease risk level prediction," *Int. J. Syst. Assur. Eng. Manag.*, vol. 8, pp. 1109–1125, 2017, doi: 10.1007/s13198-017-0578-8.
- [4] R. Chadha and S. Mayank, "Prediction of heart disease using data mining techniques," *CSI Trans. ICT*, vol. 4, no. 2–4, pp. 193–198, 2016, doi: 10.1007/s40012-016-0121-0.
- [5] G. T. Reddy, M. P. K. Reddy, K. Lakshmana, D. S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," *Evol. Intell.*, vol. 13, no. 2, pp. 185–196, 2020, doi: 10.1007/s12065-019-00327-1.
- [6] S. M. S. Shah, F. A. Shah, S. A. Hussain, and S. Batool, "Support Vector Machines-based Heart Disease Diagnosis using Feature Subset, Wrapping Selection and Extraction Methods," *Comput. Electr. Eng.*, vol. 84, p. 106628, 2020, doi: 10.1016/j.compeleceng.2020.106628.
- [7] T. Vivekanandan and N. C. Sriman Narayana Iyengar, "Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease," *Comput. Biol. Med.*, vol. 90, no. April, pp. 125–136, 2017, doi: 10.1016/j.compbiomed.2017.09.011.
- [8] V. Chaurasia, "Early Prediction of Heart Diseases Using Data Mining," *Caribb. J. Sci. Technol.*, vol. 1, pp. 208–217, 2013.
- [9] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," *Proc. - IEEE Symp. Comput. Commun.*, no. Iscc, pp. 204–207, 2017, doi: 10.1109/ISCC.2017.8024530.
- [10] N. Khateeb and M. Usman, "Efficient heart disease prediction system using K-nearest neighbor classification technique," *ACM Int. Conf. Proceeding Ser.*, pp. 21–26, 2017, doi: 10.1145/3175684.3175703.
- [11] T. R. Stella Mary and S. Sebastian, "Predicting heart ailment in patients with varying number of features using data mining techniques," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 4, pp. 2675–2681, 2019, doi: 10.11591/ijece.v9i4.pp2675-2681.
- [12] L. Verma, S. Srivastava, and P. C. Negi, "A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data," *J. Med. Syst.*, vol. 40, no. 7, 2016, doi: 10.1007/s10916-016-0536-z.
- [13] S. Iftikhar, K. Fatima, A. Rehman, A. S. Almazayad, and T. Saba, "An evolution based hybrid approach for heart diseases classification and associated risk factors identification," *Biomed. Res.*, vol. 28, no. 8, pp. 3451–3455, 2017.
- [14] S. D. Desai, S. Giraddi, P. Narayankar, N. R. Pudukalakatti, and S. Sulegaon, "Back-propagation neural network versus logistic regression in heart disease classification," *Adv. Intell. Syst. Comput.*, vol. 702, no. July, pp. 133–144, 2019, doi: 10.1007/978-981-13-0680-8_13.
- [15] N. Leema, H. K. Nehemiah, and A. Kannan, "Neural network classifier optimization using Differential Evolution with Global Information and Back Propagation algorithm for clinical datasets," *Appl. Soft Comput. J.*, vol. 49, pp. 834–844, 2016, doi: 10.1016/j.asoc.2016.08.001.
- [16] B. Subanya and R. R. Rajalaxmi, "Feature selection using artificial bee colony for cardiovascular disease classification," *2014 Int. Conf. Electron. Commun. Syst. ICECS 2014*, 2014, doi: 10.1109/ECS.2014.6892729.
- [17] T. Prasartvit, A. Banhamsakun, B. Kaewkamnerpong, and T. Achalakul, "Reducing bioinformatics data dimension with ABC-kNN," *Neurocomputing*, vol. 116, pp. 367–381, 2013, doi: 10.1016/j.neucom.2012.01.045.
- [18] D. Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm," *Appl. Soft Comput. J.*, vol. 11, no. 1, pp. 652–657, 2011, doi: 10.1016/j.asoc.2009.12.025.
- [19] F. S. Gharehchopogh, S. R. Khaze, and I. Maleki, "A new approach in bloggers classification with hybrid of K-nearest neighbor and artificial neural network algorithms," *Indian J. Sci. Technol.*, vol. 8, no. 3, pp. 237–246, 2015, doi: 10.17485/ijst/2015/v8i3/59570.
- [20] F. Ortega-Zamorano, J. M. Jerez, G. E. Juárez, and L. Franco, "FPGA Implementation of Neurocomputational Models: Comparison Between Standard Back-Propagation and C-Mantec Constructive Algorithm," *Neural Process. Lett.*, vol. 46, no. 3, pp. 899–914, 2017, doi: 10.1007/s11063-017-9655-x.
- [21] M. G. Feshki and O. S. Shijani, "Improving the heart disease diagnosis by evolutionary algorithm of PSO and Feed Forward Neural Network," *2016 Artif. Intell. Robot. IRANOPEN 2016*, pp. 48–53, 2016, doi: 10.1109/RIOS.2016.7529489.
- [22] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
- [23] T. Karayilan and Ö. Kiliç, "Prediction of Heart disease using neural network," *2nd Int. Conf. Comput. Sci. Eng. UBMK 2017*, pp. 719–723, 2017, doi: 10.1109/UBMK.2017.8093512.
- [24] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," *Proc. - 6th Int. Adv. Comput. Conf. IACC 2016*, no. Cv, pp. 78–83, 2016, doi: 10.1109/IACC.2016.25.