

DECISION TREE OPTIMIZATION IN HEART FAILURE DIAGNOSTICS: A PARTICLE SWARM OPTIMIZATION APPROACH

Sumarna^{*1}, Sartini², Witriana Endah Pangesti³, Rachmat Suryadithia⁴, Verry Riyanto⁵

^{1,2}Informatics, Faculty of Information Technology, Universitas Nusa Mandiri, Indonesia

³Information Systems, Faculty of Information Technology, Universitas Nusa Mandiri, Indonesia

^{4,5}Information Systems, Faculty of Informatics Engineering, Universitas Bina Sarana Informatika, Indonesia

Email: ¹sumarna.smn@nusamandiri.ac.id, ²sartini.sar@nusamandiri.ac.id, ³witriana.weg@nusamandiri.ac.id,
⁴rachmat.rcs@bsi.ac.id, ⁵verry.vry@bsi.ac.id

(Article received: January 31, 2024; Revision: March 8, 2024; published: May 27, 2024)

Abstract

The rapid advancement of technology has made the implementation of accurate diagnostic methods for serious diseases like heart failure extremely important. Heart failure, being a leading cause of death worldwide, necessitates precise and accurate diagnostic techniques. The problem with conventional diagnostic methods is that they often fail to effectively accommodate the complexity of clinical data, leading to an increase in mortality rates due to heart failure. Previous research has employed various data analysis methods, but there are still fluctuations in the accuracy of results. The aim of this study is to enhance the accuracy of heart failure diagnosis by integrating the Decision Tree (DT) method with Particle Swarm Optimization (PSO) optimization. This research involves collecting and preprocessing heart failure data, followed by the development of a DT model. This model is then optimized using the PSO technique. The study uses a dataset from the UCI Repository, involving testing and validation processes to measure the model's effectiveness. The results show a significant improvement in accuracy and the Area Under Curve (AUC) after applying PSO. Accuracy increased from 79.92% to 85.29%, and AUC from 0.706% to 0.794%. The conclusion is that the integration of DT and PSO successfully improved the accuracy and reliability of the model in diagnosing heart failure. This innovation offers potential for further research in integrating optimization techniques in health data analysis, with the possibility of application in various clinical scenarios.

Keywords: Data Mining, Decision Tree, Heart Failure Diagnosis, Optimization, Particle Swarm Optimization.

OPTIMALISASI DECISION TREE DALAM DIAGNOSTIK GAGAL JANTUNG: SEBUAH PENDEKATAN PARTICLE SWARM OPTIMIZATION

Abstrak

Kemajuan teknologi yang cepat, penerapan metode diagnosis yang akurat untuk penyakit serius seperti gagal jantung menjadi sangat penting. Gagal jantung sebagai penyebab utama kematian di dunia, membutuhkan teknik diagnostik yang tepat dan akurat. Permasalahan yang ada metode diagnostik konvensional sering kali tidak dapat mengakomodasi kompleksitas data klinis secara efektif dan menyebabkan peningkatan angka kematian akibat gagal jantung. Penelitian sebelumnya telah menggunakan berbagai metode analisis data, namun masih terdapat fluktuasi dalam akurasi hasil. Tujuan penelitian ini adalah untuk meningkatkan akurasi diagnosis gagal jantung dengan mengintegrasikan metode *Decision Tree* (DT) dengan optimisasi *Particle Swarm Optimization* (PSO). Metode dalam penelitian ini melibatkan pengumpulan dan *preprocessing* data gagal jantung diikuti dengan pembangunan model DT. Kemudian, model ini dioptimalkan dengan teknik PSO. Penelitian ini menggunakan dataset dari *UCI Repository* dengan melibatkan proses pengujian dan validasi untuk mengukur efektivitas model. Hasil menunjukkan peningkatan signifikan dalam akurasi dan *Area Under Curve* (AUC) setelah penerapan PSO. Akurasi meningkat dari 79.92% menjadi 85.29%, dan AUC meningkat dari 0.706% menjadi 0.794%. Kesimpulan dengan dilakukannya integrasi DT dan PSO berhasil meningkatkan akurasi dan keandalan model dalam diagnosa gagal jantung. Inovasi ini menawarkan potensi untuk penelitian lebih lanjut dalam mengintegrasikan teknik optimisasi dalam analisis data kesehatan, dengan kemungkinan penerapan dalam skenario klinis yang beragam.

Kata kunci: Data Mining, Diagnosa Gagal Jantung, Decision Tree, Optimalisasi, Particle Swarm Optimization.

1. PENDAHULUAN

Di masa kini, di mana teknologi maju dan canggih dengan cepat menerapkan metode diagnosis penyakit menjadi sangat krusial. Terlebih lagi khususnya untuk penyakit serius seperti gagal jantung. Penyakit Gagal jantung, sebagai penyebab utama kematian nomor satu di dunia dapat menyebabkan komplikasi kardiovaskular serius termasuk *stroke* dan serangan jantung serta membutuhkan teknik diagnostik yang tepat dan akurat untuk menentukan pengobatan yang tepat. Di kalangan umum, diperkirakan risiko mengalami gagal jantung pada usia 40 tahun adalah satu dari lima orang yang terkena. Hal ini terfokus pada tantangan menghadapi kerumitan gagal jantung, yang seringkali susah didiagnosis dengan cepat dan akurat karena gejala dan pemeriksaan klinis yang bervariasi. Saat ini kita melihat peningkatan yang mengkhawatirkan dalam insiden terkena suatu penyakit seperti menjadi gaya hidup masyarakat saat ini. Penyakit yang menjadi *trend* saat ini seperti Penyakit Jantung, Hipertensi, dan Obesitas. Di antara semua, Penyakit Jantung berada di puncak daftar sebagai penyebab utama kematian di banyak negara. Ini termasuk penyakit pada otot jantung, katup, sistem konduksi, dan serangan jantung. *Infark miokard*, atau serangan jantung, merupakan penyakit utama di antara penyakit jantung lainnya.

Penyakit jantung tidak lagi terbatas pada kelas sosial tertentu atau orang tertentu. Terkadang sering kali terjadi pada seseorang tanpa gejala yang jelas dan menjadikannya sebagai pembunuh senyap yang mematikan. Karena sifat penyakit ini, meningkatkan kekhawatiran terhadap masyarakat serta konsekuensinya yang didapat. Oleh karena itu upaya terus dilakukan untuk mendeteksi penyakit ini lebih awal, karena sangat penting untuk menemukan peluang pencegahan dengan langkah dasar awal dapat melakukan sebuah penilaian risiko terhadap penyakit gagal jantung ini. Berbagai alat dan teknik sedang dikembangkan secara teratur untuk memenuhi kebutuhan kesehatan saat ini. Teknik data mining misalnya dapat sangat membantu dalam hal mengidentifikasi masalah dan melacak perkembangan faktor risiko penyakit jantung.

Data mining dapat menyoroti keterbatasan metode diagnosa konvensional yang menangani gagal jantung. Maka tujuan dilakukan penelitian ini dapat mengidentifikasi dan melacak perkembangan faktor risiko penyakit jantung dengan metode *Decision Tree* (DT) dan *Particle Swarm Optimization* (PSO) dalam mengatasi masalah ini dengan mengintegrasikan kelebihan algoritma DT dalam klasifikasi data klinis dan efisiensi PSO dalam menemukan kombinasi parameter terbaik untuk diagnosis yang akurat.

Permasalahan yang ada metode konvensional yang sering kali tidak dapat mengakomodasi kompleksitas data klinis pasien secara efektif dan

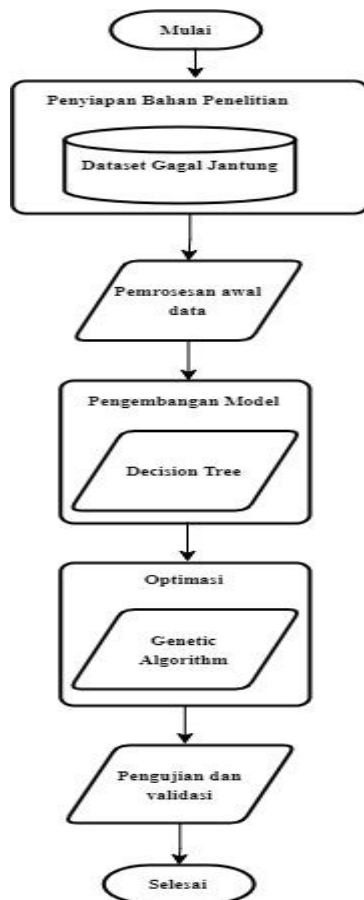
banyaknya kematian yang semakin meningkat, sehingga menjadi prioritas utama untuk semua pusat kesehatan masyarakat dengan mengurangi angka kematian penyakit gagal jantung dengan sebuah tehnik datamining menggunakan metode yang populer saat ini. Metode-metode yang sudah dilakukan masih belum memaksimalkan hasil akurasi. Dari metode yang sudah digunakan dalam hal penyakit gagal jantung sebagai berikut.

Berawal pada tahun 2016 penelitian gagal jantung telah dilakukan dengan metode *Logistic Regression* (LR), *neural network*, *svm*, *Random Forest*, *Decision Tree* (DT) dan *Naïve Bayes* (NB) dengan menghasilkan model *decision tree* menjadi akurasi tertinggi yang ke dua setelah model *naïve bayes* dengan nilai akurasi sebesar 86,60% [1]. Pada tahun 2017 penelitian gagal jantung telah dilakukan kembali dengan metode NB dan DT dengan hasil akurasi tertinggi 98,03% yang diperoleh oleh metode DT [2]. Pada tahun 2019 penelitian gagal jantung telah dilakukan kembali dengan metode DT, NB, LR, RF dan *Support Vector Machine* (SVM) dengan hasil bahwa metode DT memiliki akurasi terendah dari metode lain dengan nilai 82,22% [3], dan peneliti lain meneliti di tahun yang sama menggunakan metode *multi-task deep and wide neural network* dengan hasil AUC 0,9393 [4]. Pada tahun 2020 penelitian gagal jantung telah dilakukan kembali dengan metode RF, *Gradient boosting*, LR, *artificial neural network*, NB, SVM, *k-nearest neighbors* dan DT dengan hasil nilai akurasi metode DT menjadi akurasi tertinggi 73,70% peringkat ke tiga dari metode lain [5], dan peneliti lain meneliti di tahun yang sama menggunakan metode *random forest* dengan hasil AUC 0,918 [6]. Pada tahun 2021 penelitian gagal jantung telah dilakukan kembali dengan metode RF [2], *adaptive boosting classifier*, LR, *stochastic gradient classifier*, RF, *gradient boosting classifier*, *extra tree classifier*, *gaussian naive bayes classifier*, SVM dan DT dengan hasil nilai akurasi yang menurun 78,89% dan menjadi dua terbawah dari pada metode lain [7]. Pada tahun 2022 penelitian gagal jantung telah dilakukan dengan metode K-Nearest Neighbors (KNN), NB, LR, SVM, RF, DT dan NN, dimana hasil akurasi menunjukkan bahwa metode LR, SVM dan NN memberikan hasil akurasi yang tinggi mencapai hingga 94,57% [8]. Pada tahun 2024 penelitian gagal jantung telah dilakukan kembali dengan metode *neural network* dan *Particle Swarm Optimization* (PSO) dengan hasil nilai akurasi sebesar 77,26% naik menjadi 85,61% setelah dilakukan penggabungan pada kedua metode tersebut [9]. Pada tahun 2024 juga dilakukan penelitian gagal jantung dengan metode RF dan PSO dengan hasil akurasi 71,91% naik menjadi 83,64% [10]. Maka dari itu, Setelah penelitian yang sudah dilakukan maka penelitian ini menggunakan metode DT dengan mengoptimalkan PSO. Tujuan dengan dilakukan metode tersebut untuk meningkatkan

akurasi pada data gagal jantung yang sebelumnya belum pernah dilakukan pada penelitian gagal jantung, sehingga dapat memperbaiki fitur yang tidak relevan.

Metode DT telah terbukti efektif sebagai alat analisis prediktif karena kemudahannya dalam interpretasinya serta kemampuannya dalam menangani data kategorikal dan numerik [11]. Namun, terdapat kelemahan pada model ini seperti risiko *overfitting* dan kurangnya efisiensi dalam pemilihan fitur serta pembentukan struktur pohon [12]. Sebagai solusi teknik optimasi digunakan PSO, yang terinspirasi dari perilaku gerombolan hewan seperti burung atau ikan mengoptimalkan model prediktif dengan menyesuaikan bobot fitur dan parameter [13]. Teknik ini efektif dalam menemukan kombinasi parameter yang terbaik, sangat berguna untuk data kompleks dan multidimensi seperti dalam kasus data medis [14]. PSO meningkatkan akurasi model prediktif, terutama dalam mendeteksi risiko gagal jantung, dengan cara terus-menerus menyesuaikan posisi partikel berdasarkan pengalaman mereka sendiri dan partikel lain, sehingga efektif dalam menyelesaikan masalah optimasi yang rumit [15].

2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Gambar 1 diagram alur tahapan penelitian menjelaskan proses penelitian untuk optimasi metode

DT dalam diagnosa gagal jantung dengan pendekatan PSO. Berikut penjelasannya dari alur tersebut.

Persiapan bahan penelitian

Pada tahap ini, peneliti mengumpulkan data yang diperlukan untuk studi tersebut. Ini meliputi dataset gagal jantung yang mungkin terdiri dari data medis pasien seperti umur, anemia, jenis kelamin, merokok atau tidak dan faktor lain yang akan digunakan untuk penelitian. Dalam hal ini, data yang dikumpulkan adalah data set gagal jantung, yang kemungkinan berisi informasi tentang pasien dan hasil penyakit jantung mereka.

Preprocessing Data

Sebelum data dapat diaplikasikan dalam pembangunan model prediktif, langkah krusial yang perlu dilakukan adalah *preprocessing*. Proses ini meliputi berbagai kegiatan seperti membersihkan data yang mencakup penghapusan atau penggantian data yang tidak lengkap, normalisasi atau standarisasi nilai, *encoding* data kategorikal, serta pengurangan dimensi data bila diperlukan. Tujuan utama dari *preprocessing* adalah untuk mengolah data menjadi format yang optimal untuk pemodelan, serta meminimalisir distorsi dalam hasil model yang dapat disebabkan oleh data yang kurang tepat. Ini juga melibatkan transformasi format data, penanganan data yang hilang, dan pembersihan data dari nilai-nilai yang tidak konsisten untuk memastikan data tersebut siap digunakan dalam model pembelajaran mesin.

Pembangunan Model

Pada tahap ini, dengan membangun model DT yang merupakan sebuah metode pembelajaran mesin yang efektif untuk klasifikasi dan regresi [16]. Model ini menggambarkan prediksi sebagai struktur pohon keputusan, dimana data dibagi menjadi berbagai cabang berdasarkan kriteria dan parameter tertentu untuk menghasilkan prediksi [17]. Model DT yang tidak hanya memudahkan interpretasi, tetapi juga sering diaplikasikan dalam konteks klasifikasi medis [18]. Model ini memetakan observasi dari item data melalui serangkaian keputusan yang mengarah pada kesimpulan nilai target [19]. Selanjutnya, proses optimisasi dilakukan untuk meningkatkan efektivitas model ini.

Optimisasi

Setelah model *Decision Tree* dibangun, langkah selanjutnya adalah mengoptimalkan model tersebut menggunakan PSO. PSO digunakan untuk meningkatkan kinerja model DT yang telah dibangun. dengan sebuah teknik optimisasi berdasarkan simulasi perilaku sosial dari gerombolan burung atau ikan [20]. PSO digunakan untuk menemukan kombinasi terbaik dari parameter model atau struktur pohon untuk memaksimalkan akurasi prediksi gagal jantung.

Pengujian dan Validasi

Seluruh proses ini diakhiri dengan pengujian dan validasi model untuk memastikan bahwa model yang dioptimalkan dapat memprediksi gagal jantung

dengan akurat dan dapat diandalkan sebelum digunakan dalam setting klinis atau penelitian lebih lanjut. Ini melibatkan penggunaan metrik seperti akurasi, *Area Under Curve* (AUC). Untuk menentukan seberapa baik model dalam memprediksi gagal jantung. Berikut rumus yang digunakan untuk akurasi pada pengujian yang dapat dilihat rumus (1) dan rumus AUC yang didapat dari hasil penggabungan antara *sensitivity* dan *specificity* pada rumus (2) dan (3).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{1}$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \tag{2}$$

$$specificity = \frac{TN}{TN+FP} \times 100\% \tag{3}$$

3. HASIL DAN PEMBAHASAN

Persiapan bahan penelitian

Data yang digunakan dalam studi ini diperoleh dari UCI *Repository*, sebuah sumber data publik, dengan fokus pada dataset tentang kegagalan jantung.

Dataset ini berisi 300 data dan mencakup 12 atribut, termasuk variabel seperti umur, anemia, diabetes, hipertensi, jenis kelamin, status merokok, dan lainnya. Studi ini juga menggunakan data historis dari tes medis yang dijalani pasien. Catatan kesembuhan pasien dari penyakit jantung juga diambil untuk menilai hubungan dengan faktor-faktor yang telah disebutkan. Data ini akan dianalisis menggunakan serangkaian metode statistik, termasuk *Decision Tree* (DT) dan metode untuk meningkatkan akurasi, yang akan dioptimalkan melalui algoritma *Particle Swarm Optimization* (PSO).

Preprocessing Data

Data digunakan untuk mengembangkan model prediktif yang melibatkan berbagai proses, termasuk pembersihan data yang berfokus pada penghapusan atau penggantian data yang tidak lengkap dan normalisasi data untuk memastikan formatnya sesuai untuk pemodelan yang dapat dilihat pada tabel 1. Langkah-langkah ini bertujuan untuk mengoptimalkan data dan mengurangi distorsi dalam hasil model yang mungkin terjadi karena data yang tidak akurat. Hasil dari proses ini dapat dilihat dalam Tabel 2, yang memperlihatkan sampel data gagal jantung yang digunakan dalam penelitian ini.

Tabel 1. Data gagal jantung sebelum preprocessing

age	anaemi a	creat i	diabete s	ejection fraction	high_bloo d pressure	platelet s	serum creatinin e	serum sodiu m	se x	smokin g	tim e	DEAT H EVENT
69	0	582	1	35	0	228000	3.5	134	1	0	30	1
90	1	60	1		0	226000	1	134		0		1
82	1		1	50	1	321000	1	145	0	0	30	
60	0	2656	1	30	0	305000	2.3		1	0		0
60	0	235	1	38	0		3	142	0	0	30	1
70	0	582	0	20	1		1.83	134	1	1	31	1

Tabel 2. Preprocessing Data Gagal Jantung

age	anaemi a	creat i	diabete s	ejection fraction	high_bloo d pressure	platelets	serum creatinin e	serum sodiu m	se x	smokin g	tim e	DEAT H EVENT
69	0	582	1	35	0	228000	3.5	134	1	0	30	1
90	1	60	1	50	0	226000	1	134	1	0	30	1
82	1	855	1	50	1	321000	1	145	0	0	30	1
60	0	2656	1	30	0	305000	2.3	137	1	0	30	0
60	0	235	1	38	0	329000	3	142	0	0	30	1
70	0	582	0	20	1	263358.0 3	1.83	134	1	1	31	1
50	0	124	1	30	1	153000	1.2	136	0	1	32	1
70	0	571	1	45	1	185000	1.2	139	1	1	33	1
...
50	0	582	1	38	0	310000	1.9	135	1	1	35	1

Pada tabel 1 merupakan hasil dari *preprocessing* data, sehingga data tersebut siap digunakan dalam pengujian dengan metode DT.

Pembangunan Model

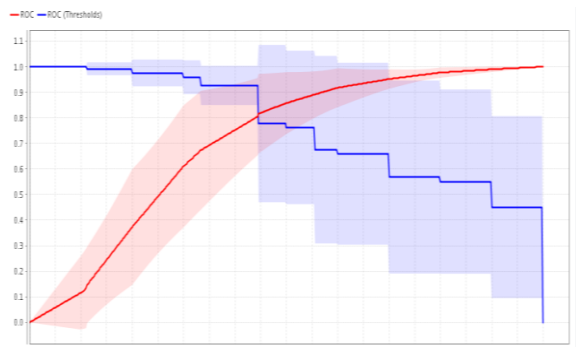
Informasi dari tabel 1 akan digunakan untuk melakukan proses pengujian dan validasi data menggunakan model *Decision Tree* (DT) dan perangkat lunak *Rapid Miner*. Proses ini akan memanfaatkan metode validasi silang sebanyak 10 kali untuk menjamin akurasi dan reliabilitas hasil.

Hasil pengujian akan ditampilkan dalam bentuk *confusion matrix*, yang rinciannya tersedia di tabel 3. Pendekatan ini bertujuan untuk memberikan pemahaman lebih dalam mengenai efisiensi dan efektivitas model DT dalam penelitian ini.

Table 3. Confussion Matrix DT

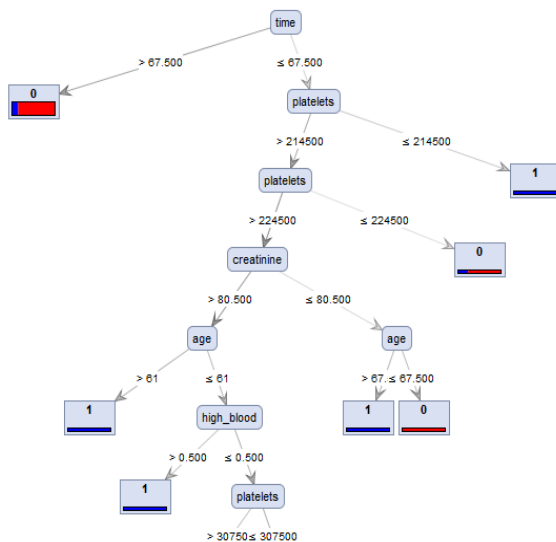
	true 1	true 0
pred 1	50	14
pred 0	46	189

Ini menunjukkan jumlah kasus di mana model secara benar memprediksi gagal jantung. Dalam hal ini, ada 50 kasus di mana model dengan benar mengidentifikasi gagal jantung. Tabel menunjukkan 14 kasus menunjukkan jumlah kasus di mana model keliru memprediksi gagal jantung (yaitu, memprediksi gagal jantung ketika sebenarnya tidak terjadi). Ada 46 kasus dimana gagal jantung ada tetapi model tidak mengidentifikasinya. Ada 189 kasus di mana tidak ada gagal jantung dan model juga memprediksi hal yang sama. Dari *confussion matrix* DT maka akan didapatkan kurva gagal jantung dengan model DT yang dapat dilihat pada gambar 2.



Gambar 2. Kurva Gagal Jantung Model DT

Gambar 2 menunjukkan bahwa penggunaan metode *decision tree* pada dataset kegagalan jantung menghasilkan kurva yang menandakan model yang relatif baik, ditandai dengan nilai AUC sebesar 0,706%. Sehingga dataset gagal jantung akan memperoleh sebuah pohon keputusan yang terdapat pada gambar 3.



Gambar 3. Kurva Gagal Jantung Model DT

Pada gambar 3 pohon keputusan ini dibagi menjadi beberapa jalur berdasarkan nilai-nilai tersebut. Misalnya, jika waktu kurang dari atau sama dengan 67.500 itu mengarah ke tidak gagal jantung, jika lebih dari itu selanjutnya membagi berdasarkan jumlah *platelets*. Ini berlanjut hingga mencapai titik

di mana sebuah keputusan dapat dibuat. Sehingga dari pohon keputusan tersebut dapat digunakan oleh profesional kesehatan untuk membantu dalam pengambilan keputusan berbasis bukti. Pohon keputusan seperti ini sering digunakan dalam pengolahan data besar untuk mengidentifikasi pola yang mungkin tidak terlihat melalui analisis konvensional.

Optimisasi

Setelah menerapkan metode *Decision Tree* (DT) pada dataset gagal jantung, langkah selanjutnya adalah mengoptimasikannya menggunakan metode *Particle Swarm Optimization* (PSO), dengan prosedur yang sama seperti yang digunakan pada metode DT. Dari kedua percobaan ini maka akan mendapatkan bobot nilai dari masing-masing atribut seperti yang terlihat pada tabel 4.

Table 4. Bobot Nilai Dataset Gagal Jantung

Attribute	Weight
age	0
anemia	0
creatinine_phosphokinase	1
diabetes	1
ejection_fraction	0
high_blood_pressure	0.583
platelets	0
serum_creatinine	0
serum_sodium	1
sex	0
smoking	1
time	0.937

Pada tabel 3 dengan bobot atribut yang bernilai 0 maka menandakan bahwa atribut tersebut tidak mempengaruhi, sedangkan atribut yang bernilai 1 merupakan atribut yang mempengaruhi untuk pengoptimalan. Dari hasil bobot atribut dengan menggunakan PSO maka akan mendapatkan *confussion matrix* yang hasilnya dapat dilihat pada tabel 5.

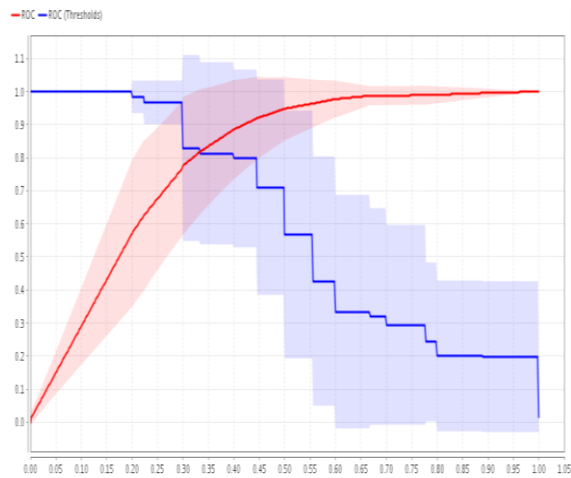
Table 5. Accuracy dan AUC Model DT+PSO

	true 1	true 0
pred 1	58	6
pred 0	38	197

Ini menunjukkan jumlah kasus di mana model secara benar memprediksi gagal jantung. Dalam hal ini, ada 58 kasus di mana model dengan benar mengidentifikasi gagal jantung. Tabel menunjukkan 6 kasus menunjukkan jumlah kasus di mana model keliru memprediksi gagal jantung (yaitu, memprediksi gagal jantung ketika sebenarnya tidak terjadi). Ada 38 kasus dimana gagal jantung ada tetapi model tidak mengidentifikasinya. Ada 197 kasus di mana tidak ada gagal jantung dan model juga memprediksi hal yang sama. Dari *confussion matrix* DT optimasi PSO maka akan didapatkan kurva gagal jantung dengan model DT optimasi PSO yang dapat dilihat pada gambar 4.

Gambar 4 menunjukkan bahwa penggunaan metode PSO pada DT pada dataset kegagalan jantung menghasilkan kurva yang menandakan model yang

relatif baik, ditandai dengan nilai AUC sebesar 0,794%.



Gambar 4. Kurva Gagal Jantung Model DT+PSO

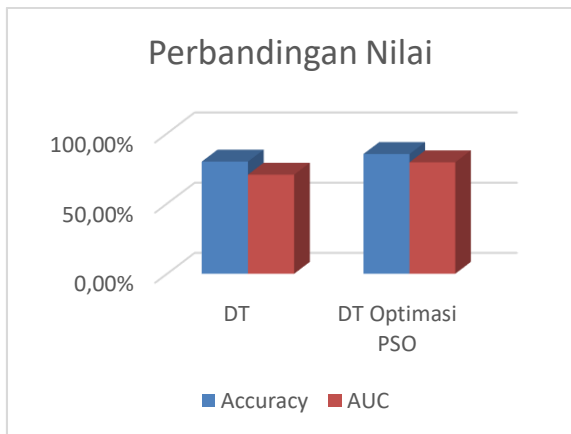
Pengujian dan Validasi

Setelah model dioptimalkan, evaluasi akan dijalankan melalui pengujian dan validasi pada hasil dari kedua metode tersebut. Terdapat kenaikan yang mencolok dalam nilai akurasi dan AUC pada data kegagalan jantung yang diteliti menggunakan metode DT bersama PSO, sebagaimana ditunjukkan di tabel 5. kinerja tiap metode dievaluasi dengan menggunakan metrik tertentu yang dirancang untuk mengukur efektivitas model, dimana akurasi dan AUC digunakan sebagai indikator perbandingan.

Tabel 6. Perbandingan Performa Metode

Algorithm	DT	DT Optimasi PSO
Accuracy	79.92%	85.29%
AUC	0.706%	0.794%

Tabel yang membandingkan kinerja berbagai metode menunjukkan dengan jelas bahwa penelitian ini berhasil meningkatkan *performa* algoritma *Decision Tree* (DT) setelah optimalisasi dengan *Particle Swarm Optimization* (PSO). Peningkatan ini terlihat dari dua metrik evaluasi: tingkat akurasi dan Area Under Curve (AUC). Hasil perbandingan ini juga tercermin dalam grafik nilai yang ditampilkan pada Gambar 5.



Gambar 5. Grafik Perbandingan Nilai

Berdasarkan informasi pada gambar 5 jika nilai akurasi untuk DT adalah 79.92% dan setelah optimisasi dengan PSO adalah 85.29%, ini menunjukkan bahwa optimisasi PSO telah berhasil meningkatkan akurasi model DT. Peningkatan dari 79.92% menjadi 85.29% dalam akurasi adalah perubahan yang signifikan dan menunjukkan bahwa teknik optimisasi PSO dapat efektif dalam meningkatkan kinerja model prediktif. Grafik yang menunjukkan dua pasang batang yang dapat dilihat mewakili nilai sebelum dan sesudah optimisasi bahwa PSO memberikan peningkatan kinerja terhadap model DT asli.

4. DISKUSI

Penelitian ini memperlihatkan peningkatan signifikan dalam akurasi dan AUC setelah penerapan optimisasi PSO pada model DT untuk diagnosis gagal jantung. Peningkatan akurasi dari 79.92% menjadi 85.29% menandakan bahwa integrasi PSO berhasil meningkatkan kemampuan prediksi model DT dibandingkan dengan penelitian nomor 3 sebelumnya pada referensi dengan akurasi sebelumnya sebesar 82,22%. PSO memainkan peran penting dalam menentukan bobot dan parameter yang optimal, yang tercermin dari perubahan dalam confusion matrix. Ini menunjukkan efektivitas PSO dalam meningkatkan seleksi fitur dan struktur pohon keputusan pada model DT. Studi ini efektif dalam mengolah data klinis yang kompleks termasuk variabel seperti umur, anemia, dan status merokok memperlihatkan kemampuan model DT yang dioptimasi PSO dalam menangani data multidimensi.

Perbandingan dengan penelitian sebelumnya telah menggunakan beragam metode seperti LR, RF, dan NN. Namun, fluktuasi dalam akurasi menunjukkan tantangan dalam mencapai konsistensi prediksi. Model NB dan RF menunjukkan akurasi tinggi dalam beberapa studi, tetapi kedua metode tersebut mungkin tidak seefisien DT dalam interpretasi dan penggunaan data klinis. Studi ini membawa inovasi dengan menggabungkan DT dan PSO yang belum banyak dieksplorasi dalam penelitian sebelumnya. Ini menawarkan pendekatan baru dalam meningkatkan akurasi model diagnostik gagal jantung. Dengan mengkombinasikan DT yang mudah diinterpretasi dan PSO yang efisien dalam optimisasi, penelitian ini menyediakan model yang seimbang antara keakuratan dan ketergunaan praktis.

5. KESIMPULAN

Studi ini berhasil mengatasi beberapa batasan metode konvensional dalam diagnostik gagal jantung dengan mengintegrasikan DT dan PSO dengan menunjukkan peningkatan yang signifikan dalam akurasi sebesar 05,37% dan AUC 0,008%. Namun, diperlukan evaluasi lebih lanjut untuk memastikan keandalan model dalam berbagai skenario klinis. Inovasi ini membuka jalan bagi penelitian lebih lanjut

dalam penggunaan teknik optimisasi dalam analisis data kesehatan.

Penelitian masa depan dapat mengintegrasikan lebih banyak variabel klinis dan demografis untuk meningkatkan keandalan model dan membandingkan model DT yang dioptimasi PSO dengan teknik AI terbaru seperti *deep learning* dapat memberikan wawasan lebih lanjut tentang efektivitas metode ini.

DAFTAR PUSTAKA

- [1] M. Saqlain, W. Hussain, N. A. Saqib, and M. A. Khan, "Identification of Heart Failure by Using Unstructured Data of Cardiac Patients," in *Proceedings of the International Conference on Parallel Processing Workshops*, 2016, vol. 2016-Septe, pp. 426–431, doi: 10.1109/ICPPW.2016.66.
- [2] N. Priyanka and P. Ravikumar, "Usage of data mining techniques in predicting the heart diseases - Naïve Bayes & decision tree," 2017, doi: 10.1109/ICCPCT.2017.8074215.
- [3] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, pp. 261–268, 2019, doi: 10.14569/ijacsa.2019.0100637.
- [4] B. Wang *et al.*, "A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health Records," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2956859.
- [5] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–16, 2020, doi: 10.1186/s12911-020-1023-5.
- [6] D. Chicco and L. Oneto, "An Enhanced Random Forests Approach to Predict Heart Failure From Small Imbalanced Gene Expression Data," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 18, no. 6, 2021, doi: 10.1109/TCBB.2020.3041527.
- [7] A. Ishaq *et al.*, "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021, doi: 10.1109/ACCESS.2021.3064084.
- [8] R. A. M. Elghalid, A. Alwirshiffani, A. A. I. Mohamed, F. H. A. Aldeeb, and A. Andiasha, "Comparison of Some Machine Learning Algorithms for Predicting Heart Failure," 2022, doi: 10.1109/ICEMIS56295.2022.9914325.
- [9] D. Yulindari, A. Wuryanto, F. A. Sariasih, Sidik, and F. A. Sariasih, "Improving the Accuracy of Heart Failure Prediction Using the Particle Swarm Optimization Method," *Sink. J. dan Penelit. Tek. Inform.*, vol. 9, no. 1, pp. 210–220, 2024, doi: <https://doi.org/10.33395/sinkron.v9i1.13017> e-ISSN.
- [10] I. Nawawi, "OPTIMISASI PEMILIHAN FITUR UNTUK PREDIKSI GAGAL JANTUNG: FUSION RANDOM FOREST DAN PARTICLE SWARM OPTIMIZATION," *Inti*, vol. 18, no. 2, 2024.
- [11] I. M. D. Priyatama and Ridwansyah, "Klasifikasi Anak Berkebutuhan Khusus Tunagrahita Menggunakan Metode Algoritma C4.5," *Paradigma*, vol. 24, no. 1, pp. 90–95, 2022, doi: <https://doi.org/10.31294/paradigma.v24i1.1087>.
- [12] R. Ridwansyah, V. Riyanto, A. Hamid, S. Rahayu, and J. J. Purnama, "Grouping Data in Predicting Infant Mortality Using K-Means and Decision Tree," *Paradigma*, vol. 24, no. 2, pp. 168–174, 2022, doi: 10.31294/paradigma.v24i2.1399.
- [13] Ridwansyah, I. Ariyati, and S. Faizah, "PARTICLE SWARM OPTIMIZATION BERBASIS CO-EVOLUTIONER DALAM EVALUASI KINERJA ASISTEN DOSEN," *J. SAINTEKOM*, vol. 9, no. 2, pp. 165–177, 2019, doi: <https://doi.org/10.33020/saintekom.v9i2.96>.
- [14] I. Ariyati, Ridwansyah, and Suhardjono, "Implementasi Particle Swarm Optimization untuk Optimalisasi Data Mining Dalam Evaluasi Kinerja Asisten Dosen," *J. Inform. dan Komput.*, vol. 3, no. 2, pp. 70–75, 2018, [Online]. Available: <https://ejournal.akakom.ac.id/index.php/jiko/article/view/127/pdf>.
- [15] Ridwansyah and E. Purwaningsih, "Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank," *J. PILAR Nusa Mandiri*, vol. 14, no. 1, pp. 83–88, 2018.
- [16] V. Riyanto, A. Hamid, and Ridwansyah, "Prediction of Student Graduation Time Using the Best Algorithm," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, pp. 1–9, 2019, doi: <http://dx.doi.org/10.24014/ijaidm.v2i1.6424>.
- [17] J. J. Purnama, H. M. Nawawi, S. Rosyida, Ridwansyah, and Risandar, "Klasifikasi Mahasiswa Her Berbasis Algoritma Svm Dan Decision Tree," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1253–1260, 2019, doi: 10.25126/jtiik.202073080.
- [18] S. Suhardjono, A. Sudradjat, B. A. Wahid, H. Sugiarto, and H. Nurdin, "Prediction Of

Infant Mortality Using The Decission Tree And Genetic Algorithm Methods,” *Paradigma*, vol. 25, no. 1, 2023, doi: <https://doi.org/10.31294/p.v25i1.1819>.

- [19] I. Ariyati, S. Rosyida, K. Ramanda, V. Riyanto, S. Faizah, and Ridwansyah, “Optimization of the Decision Tree Algorithm Used Particle Swarm Optimization in the Selection of Digital Payments,” in *Journal of Physics: Conference Series*, 2020, vol. 1641, no. 1, doi: 10.1088/1742-6596/1641/1/012090.
- [20] Suhardjono, G. Wijaya, and A. Hamid, “PREDIKSI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN SVM BERBASIS PSO,” *Bianglala Inform.*, vol. 7, no. 2, pp. 97–101, 2019, doi: <https://doi.org/10.31294/bi.v7i2.6654.g3731>.