

Evaluasi Logistic Regression dan Neural Network pada Klasifikasi Gagal Jantung Berbasis Threshold

Leslie Anggraini^{*1}, Attar Akram Abdillah², Muhammad Qaessar Kartadilaga³,
Miranti Verdiana⁴, Eko Dwi Nugroho⁵, Aidil Afriansyah⁶, Andre Febrianto⁷, Radhinka
Bagaskara⁸

^{1,2,3,4,5,6,7,8}Program Studi Teknik Informatika, Institut Teknologi Sumatera;
Lampung Selatan, Indonesia

e-mail: ^{*1}leslie.anggraini@if.itera.ac.id, ²attar.121140013@student.itera.ac.id,

³muhammad.121140119@student.itera.ac.id, ⁴miranti.verdiana@if.itera.ac.id,

⁵eko.nugroho@if.itera.ac.id, ⁶aidil.afriansyah@if.itera.ac.id, ⁷andre.febrianto@if.itera.ac.id,

⁸radhinka.bagaskara@if.itera.ac.id

Abstrak

Kardiovaskular adalah sistem jantung dan pembuluh darah dalam tubuh manusia yang bertanggung jawab atas sirkulasi darah dalam jantung, pembuluh darah, dan darah sendiri. Gangguan pada fungsi sistem ini dapat menyebabkan penyakit kardiovaskular, seperti gagal jantung, yang menjadi salah satu penyebab utama kematian di seluruh dunia. Kematian yang disebabkan oleh gagal jantung mempengaruhi 1.5 juta pasien di seluruh dunia. Dikarenakan oleh data statistik tersebut, maka ada kebutuhan untuk dapat memprediksi dampak gagal jantung untuk membantu tingkat kelangsungan hidup pasien. Sebagai bentuk kontribusi terhadap kebutuhan tersebut, penelitian ini akan menganalisis sebuah dataset pelayanan kesehatan, yaitu dataset rekam gagal jantung dari UCI. Dataset tersebut akan digunakan untuk mengklasifikasi dan memprediksi peluang kematian dari pasien gagal jantung. Kami akan membandingkan antara dua metode klasifikasi dari machine learning, yaitu Logistic Regression (LR), dan deep learning, yaitu Shallow Neural Network (SNN). Mutual Information (MI) dipilih sebagai metode pemilihan fitur. Hasil menunjukkan bahwa SNN menghasilkan akurasi lebih tinggi dengan skor 0.75, dibandingkan LR dengan akurasi sebesar 0.63.

Kata kunci—Gagal Jantung, Klasifikasi, Logistic Regression, Shallow Neural Network

Abstract

Cardiovascular is a system in the human body that is responsible for blood circulation, encompassing the heart, blood vessels, and blood itself. Disruption of this system's function can lead to cardiovascular diseases that become one of the leading causes of death worldwide, such as heart failure. Death caused by heart failure is one of the leading causes of death globally, affecting over 1.5 million of patients. Given these concerning statistics, there's a pressing need for predicting heart failure outcomes to improve patient's survival rates. To contribute to this goal, this study will analyze a prominent healthcare datasets, the UCI Heart Failure Clinical Records dataset. Data from this dataset will be used to classify and predict mortality of heart failure patients. We will compare the two classification methods between a machine learning method, which is Logistic Regression (LR), and a deep learning method, which is Shallow Neural Network (SNN). Mutual Information (MI) is chosen as the feature selection method. The results show that SNN produces a higher accuracy for predicting the dataset, with the score of 0.75, compared to LR which only has 0.63 accuracy.

Keywords—Heart Failure, Classification, Logistic Regression, Shallow Neural Network

1. PENDAHULUAN

Kardiovaskular adalah sistem dalam tubuh manusia yang bertanggung jawab atas sirkulasi darah, mencakup jantung, pembuluh darah, dan darah itu sendiri. Sistem kardiovaskular ini sangat penting dalam tubuh manusia karena fungsinya dalam mengatur peredaran darah. Sistem kardiovaskular dapat terganggu oleh beberapa faktor eksternal, seperti tekanan darah tinggi, kadar kolesterol yang tinggi dalam darah, diabetes, dan kebiasaan merokok [1]. Gangguan pada sistem ini dapat menyebabkan penyakit

kardiovaskular seperti hipertensi, stroke, penyakit jantung koroner, dan gagal jantung.

Penyakit gagal jantung menjadi salah satu penyebab utama kematian di seluruh dunia. Gagal jantung adalah suatu kondisi di mana organ jantung tidak dapat mengedarkan darah dengan baik yang disebabkan oleh kelainan pada struktur fisik dan/atau fungsinya. Pasokan darah yang tidak mencukupi ke organ-organ dalam tubuh manusia akan menyebabkan fungsi organ-organ tersebut tidak berjalan dengan baik dan pada akhirnya menyebabkan kematian pada pasien gagal jantung. Menurut penelitian terdahulu, terdapat lebih dari 1,5 juta pasien gagal jantung dari berbagai jenis, yang memperkirakan angka kelangsungan hidup sebesar 87% setelah 1 tahun, 73% setelah 2 tahun, 57% setelah 5 tahun, dan hanya 35% setelah 10 tahun [2, 3].

Melihat angka statistik yang mengkhawatirkan ini, terdapat kebutuhan mendesak untuk memahami dan memprediksi hasil dari gagal jantung dengan lebih baik guna meningkatkan perawatan pasien dan angka kelangsungan hidup. Sebagai bentuk berkontribusi terhadap tujuan tersebut, penelitian ini akan menganalisis *dataset* layanan kesehatan terkemuka, yaitu *dataset UCI Heart Failure Clinical Records*, untuk mengembangkan model prediktif terhadap hasil gagal jantung. *Dataset UCI Heart Failure Clinical Records* terdiri atas catatan medis yang mencakup usia, fraksi ejeksi, kadar kreatinin serum, dan indikator medis relevan lainnya.

Penelitian terkini dalam bidang prediksi penyakit jantung menitikberatkan pada optimalisasi metode seleksi fitur guna meningkatkan kinerja model *machine learning*. Studi oleh Pratama dan Kusnawi membandingkan berbagai metode seleksi fitur pada klasifikasi penyakit gagal jantung menggunakan *dataset* dari Kaggle [4]. Hasilnya menunjukkan bahwa penggunaan *Pearson Correlation* dengan algoritma *K-Nearest Neighbor* (KNN) memberikan akurasi tertinggi sebesar 90,58%. Selain itu, penelitian oleh Azizah et al. menggunakan kombinasi seleksi fitur *Information Gain* dan *Gain Ratio* pada algoritma *Naive Bayes*, yang menghasilkan akurasi tertinggi mencapai 91,80% menggunakan data dari *UCI Machine Learning Repository* [5]. Studi-studi ini secara keseluruhan menunjukkan betapa pentingnya metode seleksi fitur yang canggih dalam meningkatkan model prediksi penyakit jantung.

Dalam penelitian ini, *dataset UCI Heart Failure Clinical Record* akan digunakan untuk memprediksi kematian pada pasien gagal jantung yang akan diklasifikasikan ke dalam dua kategori menggunakan dua metode: metode *machine learning* dengan *Logistic Regression* (LR), dan metode *deep learning* yaitu *Shallow Neural Network* (SNN). LR adalah metode klasifikasi yang bekerja dengan menerapkan fungsi sigmoid sebagai fungsi biaya pada variabel dependen kategorikal, sehingga menghasilkan nilai probabilitas antara ‘ya’ atau ‘tidak’ [6]. Sedangkan SNN adalah jenis jaringan saraf tiruan (*Artificial Neural Network*) yang terdiri dari hanya satu lapisan tersembunyi (*hidden layer*) antara lapisan *input* dan lapisan *output* [7].

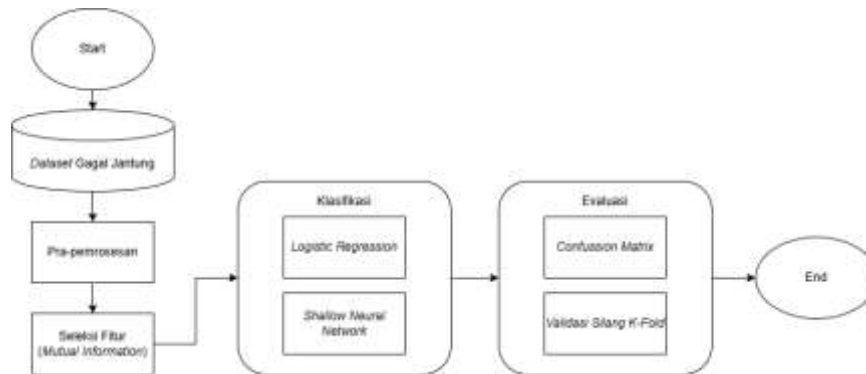
Untuk memastikan bahwa model menghasilkan akurasi terbaik, diperlukan suatu metode untuk seleksi fitur. *Mutual Information* (MI) merupakan salah satu algoritma *feature selection* yang digunakan untuk mengukur sejauh mana suatu fitur memberikan informasi terhadap variabel target [8, 9]. MI membantu mengoptimalkan kinerja model dengan memilih fitur-fitur yang paling relevan, sekaligus mengurangi redundansi antar fitur. Pemilihan fitur yang minimal namun informatif melalui MI dapat memberikan dampak positif terhadap berbagai aspek, seperti efisiensi memori, waktu komputasi, performa model, dan interpretabilitas hasil.

Berdasarkan kebutuhan untuk mengetahui metode klasifikasi prediksi untuk *dataset* penyakit gagal jantung, maka penelitian ini akan berfokus untuk melihat performa antara metode LR dan SNN dalam mengklasifikasi. Kedua metode tersebut akan dibantu dalam proses pemilihan fiturnya menggunakan metode MI. Performa hasil klasifikasi akan diukur menggunakan *confusion matrix* untuk mengukur akurasi, presisi, *recall*, dan *F1-score* dari kedua metode.

2. METODE PENELITIAN

2.1 Diagram Flowchart

Penelitian ini melibatkan serangkaian tahapan, yang meliputi pra-pemrosesan data, seleksi fitur, klasifikasi, dan evaluasi. Hal ini dilakukan dengan tujuan mengevaluasi kedua metode klasifikasi yang digunakan pada penelitian ini. Selengkapnya, alur kerja penelitian ini disajikan dalam bentuk diagram alir yang ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

2.2 Dataset

Dataset UCI Heart Failure Clinical Records terdiri dari catatan medis dari 299 pasien gagal jantung. *Dataset* ini dikembangkan oleh *University of California*, yang mencakup 12 fitur klinis berupa informasi demografis, pengukuran fisiologis, dan hasil kondisi pasien. Fitur-fitur ini akan menjadi indikator penting dalam analisis gagal jantung dan peningkatan model prediksi. Penjelasan rinci mengenai fitur-fitur dalam *dataset* UCI ditampilkan pada Tabel 1 berikut.

Tabel 1. Seleksi Fitur menggunakan *Mutual Information* (MI)

ID	Nama Fitur
U-F1	<i>age</i>
U-F2	<i>anaemia</i>
U-F3	<i>creatinine phosphokinase</i>
U-F4	<i>diabetes</i>
U-F5	<i>ejection fraction</i>
U-F6	<i>high blood pressure</i>
U-F7	<i>platelets</i>
U-F8	<i>serum creatinine</i>
U-F9	<i>serum sodium</i>
U-F10	<i>sex</i>
U-F11	<i>smoking</i>
Kelas Target	<i>death event</i>
-	<i>time</i>

2.3 Pra-pemrosesan Data

Preprocessing data sangat penting untuk menyiapkan data mentah dari *dataset*, guna memastikan kualitas dan integritas data sebelum dilakukan analisis model. Metode normalisasi akan diterapkan pada *dataset* UCI untuk mempersiapkan data. Normalisasi adalah teknik pra-pemrosesan data yang mengubah nilai fitur ke dalam rentang atau skala standar, sehingga dapat meningkatkan kinerja algoritma klasifikasi [10]. Metode yang digunakan dalam penelitian ini adalah normalisasi Min-Max, yang melakukan penskalaan linier pada setiap fitur ke dalam interval terbatas, seperti 0 hingga 1. Teknik ini bertujuan untuk mencegah fitur dengan rentang nilai besar memiliki pengaruh yang

tidak proporsional terhadap model, sekaligus menjaga hubungan relatif antara nilai dalam setiap fitur [11]. Penerapan metode normalisasi Min-Max telah banyak digunakan dalam penelitian kardiovaskular, mengingat parameter medis umumnya memiliki skala pengukuran yang beragam. Pendekatan standarisasi ini sangat penting terutama bagi algoritma yang sensitif terhadap skala, karena dapat mencegah dominasi fitur dengan nilai numerik besar tanpa mengurangi keterbacaan atau interpretabilitas data yang telah ditransformasi.

2.4 Seleksi Fitur dengan Mutual Information (MI)

Seleksi fitur adalah proses sistematis untuk mengidentifikasi dan mengekstraksi fitur-fitur yang paling informatif dari suatu dataset, sekaligus mengeliminasi fitur yang bersifat redundan atau tidak relevan, dengan tujuan mengoptimalkan kinerja model dan mengurangi dimensi data [12]. Reduksi dimensi ini meningkatkan efisiensi model dan mengurangi risiko overfitting dengan mengidentifikasi subset optimal dari fitur-fitur yang bersifat diskriminatif. Proses ini tidak hanya meningkatkan kinerja komputasi, tetapi juga memperkuat ketahanan (*robustness*) performa model melalui penghapusan fitur yang tidak diperlukan. Penelitian ini akan melakukan penerapan metode *filter*, yaitu *Mutual Information* (MI) untuk melakukan seleksi fitur pada dataset. *Mutual Information* (MI) merupakan salah satu metode dalam *feature selection* yang digunakan untuk mengevaluasi sejauh mana suatu fitur memberikan informasi terhadap target kelas [8, 9]. Metode ini mengukur ketergantungan statistik antara fitur dan label, sehingga fitur yang memiliki nilai MI tinggi dianggap memiliki kontribusi besar dalam meningkatkan akurasi model klasifikasi.

2.5 Klasifikasi

Klasifikasi merupakan teknik fundamental dalam *supervised learning* pada *machine learning*, yang bertujuan untuk memprediksi label kategorik berdasarkan fitur *input* [13]. Proses klasifikasi melibatkan pelatihan model menggunakan *dataset* berlabel, memungkinkan algoritma untuk mempelajari pola dan hubungan yang mendasari data. Tujuan utama dari klasifikasi adalah membangun model prediktif yang mampu secara akurat memberikan label pada data baru yang sebelumnya belum pernah dilihat. Penelitian ini membandingkan dua metode klasifikasi yaitu *Logistic Regression* (LR) dan *Shallow Neural Network* (SNN).

2.5.1 Logistic Regression (LR)

Logistic Regression (LR) menggunakan fungsi aktivasi *sigmoid* untuk mengubah kombinasi linier dari fitur-fitur menjadi skor probabilitas, sehingga memungkinkan prediksi biner dengan menetapkan ambang batas antara dua keluaran kategorikal seperti “ya” atau “tidak” [6]. LR seringkali dipilih karena sifatnya yang mudah diinterpretasikan, efisien secara komputasi, serta telah terbukti efektif dalam tugas klasifikasi biner, di mana LR memodelkan probabilitas suatu keluaran melalui pendekatan linier terhadap hubungan antar fitur.

2.5.2 Shallow Neural Network (SNN)

Shallow Neural Network (SNN) merupakan arsitektur jaringan syaraf tiruan yang hanya terdiri dari satu lapisan tersembunyi (*hidden layer*) yang terletak di antara lapisan input dan output [7]. SNN memanfaatkan fungsi aktivasi non-linear, seperti ReLU atau sigmoid, untuk mentransformasi fitur-fitur masukan ke dalam representasi non-linier yang memungkinkan pemodelan hubungan kompleks dalam data. Menurut penelitian oleh Agliari et al. di tahun 2021, meskipun struktur SNN lebih sederhana dibanding jaringan dalam (*deep network*), SNN tetap mampu menangkap konsep-konsep dasar atau archetype dari data yang tidak terstruktur, asalkan jumlah dan kualitas data pelatihan mencukupi [14]. Studi tersebut juga menunjukkan bahwa terdapat ambang batas ukuran dataset di mana jaringan dapat secara efektif belajar untuk menggeneralisasi, baik sebagai pengklasifikasi maupun model generatif.

2.6 Evaluasi

Pengujian akan menggunakan matrik evaluasi dan validasi silang *k-fold* untuk mengetahui performa model secara menyeluruh.

2.6.1 Matrik Evaluasi

Performa model akan dievaluasi dengan standar yang terdiri dari akurasi, presisi, *recall*, dan *f1-score*. Metrik-metrik ini dihitung berdasarkan *confusion matrix*, yang membandingkan hasil prediksi dengan klasifikasi aktual pada data uji [14]. Definisi masing-masing matrik evaluasi adalah sebagai berikut:

1. Akurasi mengukur proporsi prediksi yang benar, baik *true positive* maupun *true negative*, dari seluruh prediksi yang dilakukan, sehingga memberikan gambaran umum terhadap kinerja model.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Presisi menghitung proporsi true positive dari seluruh prediksi positif yang ditemukan dalam dataset, menunjukkan kemampuan model dalam menghindari *false positive*.

$$Presisi = \frac{TP}{TP + FP}$$

3. Recall mengukur proporsi kasus positif aktual yang berhasil diidentifikasi oleh model, mencerminkan kemampuan model dalam mendeteksi semua instansi positif.

$$Recall = \frac{TP}{TP + FN}$$

4. *F1-score* merupakan rata-rata harmonik dari nilai presisi dan *recall*. Metrik ini memberikan ukuran seimbang terhadap kinerja model, dan sangat berguna pada *dataset* yang tidak seimbang.

$$F1 - score = \frac{2 \times Presisi \times Recall}{Presisi + Recall}$$

Nilai yang lebih tinggi pada semua metrik menunjukkan kinerja model yang lebih baik, dengan nilai 1.0 merepresentasikan klasifikasi yang sempurna.

2.6.2 Validasi Silang K-Fold

Validasi silang *k-fold* adalah metode evaluasi model yang membagi *dataset* menjadi *k* bagian (*fold*) yang sama besar [15]. Proses dilakukan sebanyak *k* kali, di mana pada setiap iterasi, satu *fold* digunakan sebagai data uji, sementara *k-1 fold* sisanya digunakan untuk pelatihan. Setelah seluruh iterasi selesai, hasil evaluasi dari tiap *fold* dirata-ratakan untuk memberikan gambaran menyeluruh terhadap kinerja model. Metode ini membantu mengurangi bias akibat pembagian data dan memastikan bahwa setiap bagian data digunakan baik untuk pelatihan maupun pengujian.

3. HASIL DAN PEMBAHASAN

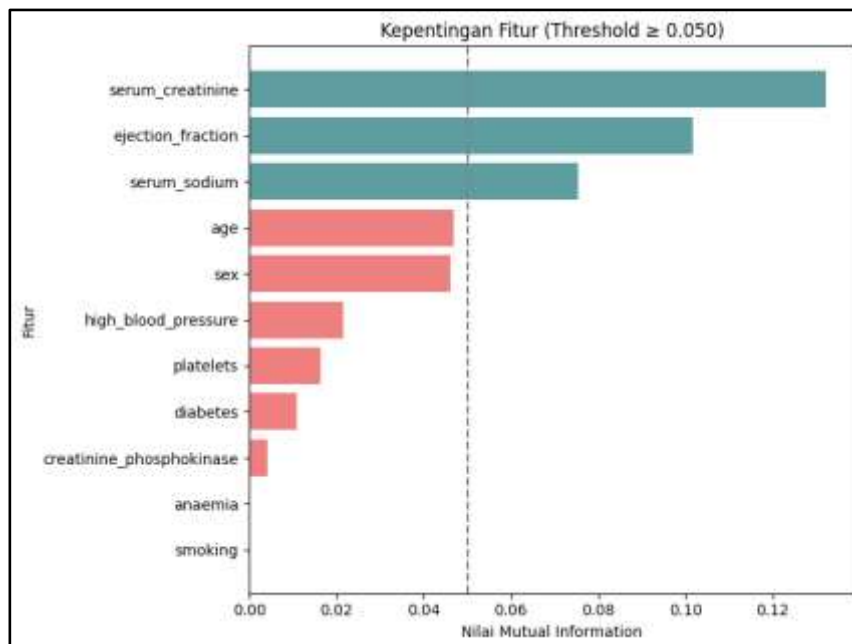
3.1 Seleksi Fitur

Seleksi fitur dilakukan untuk mengidentifikasi fitur yang paling relevan dalam memprediksi kematian pasien gagal jantung. Metode *Mutual Information* (MI) digunakan untuk mengukur relevansi atau kepentingan setiap fitur terhadap kelas target, dimana semakin tinggi nilai MI suatu fitur, maka semakin besar kontribusi fitur tersebut dalam prediksi. Meskipun keseluruhan 12 fitur yang terdapat pada dataset *UCI Heart Failure*, namun fitur *time* tidak digunakan. Fitur *time* dikecualikan karena bersifat deskriptif, yang lebih merepresentasikan durasi rawat inap dan bukan penyebab terjadinya kematian. Penelitian ini menerapkan seleksi fitur berbasis ambang batas (*threshold*) nilai MI sebesar $\geq 0,05$, tanpa membatasi jumlah fitur terpilih. Hanya fitur-fitur yang memenuhi *threshold* yang dipilih sebagai *subset* fitur untuk pemodelan klasifikasi, sehingga pendekatan ini memungkinkan model hanya dilatih menggunakan fitur-fitur yang sangat signifikan. Hasil seleksi fitur ini

dapat dilihat pada Tabel 1.

Tabel 2. Seleksi Fitur menggunakan *Mutual Information* (MI)

No.	Fitur	Nilai <i>Mutual Information</i> (MI)	Memenuhi Threshold ($\geq 0,05$)
1	<i>serum creatinine</i>	0.1323	Iya
2	<i>ejection fraction</i>	0.1018	Iya
3	<i>serum sodium</i>	0.0754	Iya
4	<i>age</i>	0.0469	Tidak
5	<i>sex</i>	0.0462	Tidak
6	<i>high blood pressure</i>	0.0216	Tidak
7	<i>platelets</i>	0.0164	Tidak
8	<i>diabetes</i>	0.0109	Tidak
9	<i>creatinine phosphokinase</i>	0.0042	Tidak
10	<i>anaemia</i>	0	Tidak
11	<i>smoking</i>	0	Tidak



Gambar 2. Seleksi Fitur Berdasarkan Nilai *Mutual Information* (MI)

Berdasarkan Tabel 1, terdapat tiga fitur dengan nilai *Mutual Information* (MI) yang memenuhi *threshold* $\geq 0,05$, yaitu *serum creatinine*, *ejection fraction*, dan *serum sodium*. Ketiga fitur ini menunjukkan fungsi organ vital (jantung dan ginjal) serta keseimbangan elektrolit, yang berperan penting dalam menentukan kondisi klinis pasien gagal jantung, sehingga sangat relevan dalam memprediksi risiko kematian. Ketiga fitur tersebut merupakan *subset* dari hasil seleksi fitur. Berikut visualisasi perbandingan nilai MI antar fitur ditampilkan pada Gambar 2.

3.2 Hasil dan Evaluasi Model Klasifikasi

Tahapan selanjutnya setelah seleksi fitur adalah mengembangkan model klasifikasi. Model dikembangkan dengan membandingkan dua pendekatan, yaitu *machine learning* menggunakan *Logistic Regression* (LR) dan *deep learning* menggunakan *Shallow Neural Network* (SNN). Tujuan perbandingan ini untuk mengevaluasi efektivitas kedua metode

dalam memprediksi kematian pasien gagal jantung berdasarkan *subset* fitur yang telah diseleksi oleh *Mutual Information* (MI). Penerapan kedua pendekatan ini dilakukan untuk mengetahui model mana yang memberikan performa klasifikasi terbaik, terutama pada *dataset* medis dengan jumlah data yang relatif kecil. Pemodelan menghasilkan dua kelas, yaitu kelas 0 untuk pasien bertahan hidup, dan kelas 1 terjadi kematian pasien. Berikut hasil klasifikasi dari tiga data uji yang telah dinormalisasi dalam rentang [0,1] dan dipilih secara acak ditampilkan pada Tabel 2. Model klasifikasi ini akan diuji menggunakan matrik evaluasi dan validasi silang *k-fold*.

Tabel 3. Hasil Klasifikasi Aktual dan Prediksi dari 3 Data Uji Secara Acak

No.	Fitur	Data Uji dari <i>Dataset UCI Heart Failure</i>		
		Data ke-3	Data ke-48	Data ke-59
1	<i>serum creatinine</i>	1,0	0,0225	0,0899
2	<i>ejection fraction</i>	0,3182	0,3182	0,0909
3	<i>serum sodium</i>	0,5714	0,7714	0,6571
Hasil Aktual		1	0	1
Hasil Prediksi (MI + LR)		1	0	0
Hasil Prediksi (MI + SNN)		1	0	0

3.2.1 Perbandingan Matrik Evaluasi

Berdasarkan Tabel 2, evaluasi model klasifikasi dilakukan dengan membandingkan nilai aktual dan prediksi melalui *confusion matrix*, dengan kelas 1 menunjukkan terjadi kematian dan 0 menunjukkan bertahan hidup. *Confusion matrix* terdiri dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), dimana TP dan TN merupakan prediksi yang benar, sementara FP dan FN adalah kesalahan prediksi. TP adalah kasus kematian yang berhasil diprediksi (aktual 1, prediksi 1), dan TN adalah pasien yang bertahan hidup dan diprediksi hidup (aktual 0, prediksi 0). FP terjadi saat pasien diprediksi meninggal padahal bertahan hidup (aktual 0, prediksi 1), sedangkan FN terjadi saat pasien meninggal namun diprediksi hidup (aktual 1, prediksi 0), yang berisiko tinggi karena kematian tidak terdeteksi. Berdasarkan nilai *confusion matrix* ini dihitung matrik evaluasi seperti akurasi, presisi, *recall*, dan *f1-score* untuk mengetahui performa model menyeluruh, berikut perbandingannya pada Tabel 3.

Tabel 4. Perbandingan Performa Model Klasifikasi dengan Matrik Evaluasi

Model Klasifikasi		Hasil <i>Confusion Matrix</i>				Matrik Evaluasi			
Metode Seleksi Fitur	Metode Klasifikasi	TP	TN	FP	FN	Akurasi	Presisi	<i>Recall</i>	<i>F1-score</i>
MI	LR	39	8	5	23	0,63	0,89	0,63	0,74
	SNN	37	14	7	17	0,75	0,84	0,69	0,76

Berdasarkan Tabel 3 yang membandingkan performa antara metode klasifikasi *Logistic Regression* (LR) dan *Shallow Neural Network* (SNN), terlihat bahwa SNN memiliki performa lebih baik dibandingkan LR dalam akurasi, *recall*, dan *f1-score*. SNN berhasil mencapai akurasi sebesar 0,75, *recall* sebesar 0,69, dan *f1-score* sebesar 0,76. Sementara itu, LR hanya unggul di nilai presisi yang lebih tinggi yaitu 0,89, selisih 0,05 dari SNN. Hal ini menunjukkan bahwa metode klasifikasi SNN lebih baik dalam mendeteksi kematian pada pasien gagal jantung dibandingkan dengan LR.

3.2.2 Perbandingan Validasi Silang K-Fold

Evaluasi lainnya dilakukan dengan teknik validasi silang *k-fold* sebanyak 5 lipatan (*fold*). Hasil evaluasi menunjukkan bahwa model SNN memiliki performa yang lebih stabil

dan unggul dibandingkan LR. Nilai rata-rata akurasi SNN adalah 0,76, lebih tinggi dibandingkan LR yang hanya mencapai 0,69. Hal ini menunjukkan bahwa dengan *subset* fitur yang dipilih sedikit (hanya tiga fitur saja) menggunakan metode *Mutual Information* (MI), model SNN lebih efektif dalam melakukan klasifikasi dibandingkan LR. Hasil evaluasi dengan *k-fold* dapat dilihat pada Tabel 4.

Tabel 5. Perbandingan Performa Model Klasifikasi dengan K-Fold

Model Klasifikasi		K-Fold = 5					Rata-rata Akurasi
Metode Seleksi Fitur	Metode Klasifikasi	K-1	K-2	K-3	K-4	K-5	
MI	LR	0,60	0,58	0,78	0,78	0,71	0,69
	SNN	0,68	0,70	0,83	0,82	0,75	0,76

3.3 Analisis Keseluruhan

Hasil evaluasi yang telah dilakukan menunjukkan bahwa pemilihan model klasifikasi dengan seleksi fitur yang tepat sangat berpengaruh terhadap performa sistem dalam mendeteksi kasus kritis, seperti kematian pada pasien gagal jantung. Penggunaan metode seleksi fitur *Mutual Information* (MI) terbukti efektif meskipun hanya menggunakan sedikit fitur, dan dipadukan dengan model *Shallow Neural Network* (SNN) dapat menghasilkan performa klasifikasi yang lebih baik dan stabil. Keunggulan SNN dibandingkan *Logistic Regression* (LR) salah satunya disebabkan oleh kemampuan arsitektur *neural network* dalam menangkap pola non-linear yang lebih kompleks dalam data, bahkan dengan kedalaman jaringan yang masih dangkal. Berbeda dengan LR yang bersifat linear dan cenderung kurang fleksibel dalam memodelkan hubungan antar fitur yang rumit, SNN mampu menyesuaikan bobot secara dinamis melalui proses pelatihan yang lebih adaptif. Berdasarkan hal tersebut, penggunaan model berbasis *neural network* menjadi solusi untuk melakukan klasifikasi meskipun dengan jumlah data dan fitur yang terbatas, serta perlunya evaluasi yang komprehensif seperti penggunaan *confusion matrix* dan validasi silang *k-fold* untuk memperoleh gambaran menyeluruh mengenai performa model

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan sebagai berikut.

1. Fitur yang diseleksi menggunakan *Mutual Information* (MI) pada *dataset UCI Heart Failure Clinical Records* menghasilkan tiga fitur yang memenuhi *threshold* $\geq 0,05$, yaitu *serum creatinine*, *ejection fraction*, dan *serum sodium*. Ketiga fitur tersebut menunjukkan kontribusi yang signifikan dalam memprediksi kematian pasien. Fitur-fitur ini berhubungan dengan fungsi organ vital dan keseimbangan elektrolit, yang berperan penting dalam menentukan kondisi pasien gagal jantung.
2. Metode *Shallow Neural Network* (SNN) lebih unggul dibandingkan dengan *Logistic Regression* (LR) dalam akurasi sebesar 0,75, *recall* sebesar 0,69, dan *f1-score* sebesar 0,76. Namun, LR hanya unggul dalam presisi sebesar 0,69 dengan selisih 0,05 dibandingkan dengan SNN. Hal ini menunjukkan bahwa SNN lebih efektif dalam mendeteksi kematian pasien gagal jantung meskipun hanya menggunakan tiga fitur yang terpilih, dibandingkan dengan LR.
3. Evaluasi dengan teknik validasi silang *k-fold* sebanyak 5 *fold* menunjukkan bahwa SNN memiliki performa yang lebih stabil dan lebih unggul dibandingkan LR. Nilai rata-rata akurasi untuk SNN adalah 0,76, sedangkan LR hanya menghasilkan rata-rata akurasi 0,69. Hasil ini menunjukkan bahwa SNN lebih efektif dalam mengatasi variabilitas data dan memberikan prediksi yang lebih konsisten. Hal ini mengindikasikan bahwa SNN dapat lebih diandalkan untuk melakukan klasifikasi kematian pasien gagal jantung meskipun dengan

jumlah data yang terbatas.

Saran untuk penelitian selanjutnya dapat difokuskan pada penerapan model yang lebih kompleks, seperti meningkatkan kedalaman *neural network* pada SNN atau mencoba metode klasifikasi lain seperti *Random Forest* atau *Support Vector Machine* (SVM). Selain itu, penelitian lebih lanjut dengan data yang lebih besar dan penambahan fitur lain juga diperlukan untuk memastikan performa model yang lebih optimal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih setinggi-tingginya kepada Institut Teknologi Sumatera atas bantuan hibah penelitian (No. 1539cj/IT9.2.1/PT.01.03/2024) melalui program “Hibah Penelitian Itera 2024”.

DAFTAR PUSTAKA

- [1] American Diabetes Association Professional Practice Committee, "10. Cardiovascular Disease and Risk Management: Standards of Care in Diabetes—2024," *Diabetes Care*, vol. 47, no. 1, pp. S179-S218, 2024.
- [2] A. Groenewegen, F. H. Rutten, A. Mosterd and A. W. Hoes, "Epidemiology of heart failure," *European Society and Cardiology*, vol. 22, no. 8, p. 1342–1356, 2020.
- [3] N. R. Jones, A. K. Roalfe, I. Adoki, F. R. Hobbs and C. J. Taylor, "Survival of patients with chronic heart failure in the community: a systematic review and meta-analysis," *European Society of Cardiology*, vol. 21, no. 11, p. 1306–1325, 2019.
- [4] K. I. F. Pratama and Kusnawi, "Komparasi Algoritma Supervised Learning dan Feature Selection pada Klasifikasi Penyakit Gagal Jantung," *Indonesian Journal of Computer Science*, vol. 12, no. 6, pp. 3722-3733, 2023.
- [5] S. R. Azizah, R. Herteno, A. Farmadi, D. Kartini and I. Budiman, "KOMBINASI SELEKSI FITUR BERBASIS FILTER DAN WRAPPER MENGGUNAKAN NAIVE BAYES PADA KLASIFIKASI PENYAKIT JANTUNG," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 10, no. 6, pp. 1361-1368, 2023.
- [6] N. R. Jones, A. K. Roalfe, I. Adoki, F. R. Hobbs and C. J. Taylor, "Survival of patients with chronic heart failure in the community: a systematic review and meta-analysis," *European Society of Cardiology*, vol. 21, no. 11, p. 1306–1325, 2019.
- [7] A. Manno, E. Martelli and E. Amaldi, "AShallow Neural Network Approach for the Short-Term Forecast of Hourly Energy Consumption," *Energies*, vol. 15, no. 3, pp. 958-979, 2022.
- [8] I. P. G. H. Suputra, Linawati, I. G. Sukadarmika, N. P. Sastra, I. G. B. D. Putra and I. W. T. Wahyudi, "KLASIFIKASI JUDUL BERITA BAHASA INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE DAN SELEKSI FITUR MUTUAL INFORMATION," *Jurnal Pendidikan Teknologi dan Kejuruan*, vol. 22, no. 1, pp. 69-79, 2025.
- [9] L. G. Irham, A. Adiwijaya, and U. N. Wisesty, 'Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine', *Jurnal Media Informatika Budidarma*, vol. 3, no. 4, p. 284, 2019.
- [10] H. M. Balaha, A. O. Shaban, E. M. El-Gendy and M. M. Saafan, "A multi-variate heart disease optimization and recognition framework," *Neural Computing and Applications*, vol. 34, no. 18, pp. 15907-15944, 2022.
- [11] H. Benhar, A. Idri and J. L. F. Alemán, "Data preprocessing for heart disease classification: A systematic literature review," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105635, 2020.
- [12] S. Bashir, I. U. Khattak, A. Khan, F. H. Khan, A. Gani and M. Shiraz, "A Novel Feature Selection Method for Classification of Medical Data Using Filters, Wrappers, and

- Embedded Approaches," *Complexity*, vol. 2022, no. 1, p. 8190814, 2022.
- [13] J. Amin, M. Sharif, A. Haldorai, M. Yasmin and R. S. Nayak, "Brain tumor detection and classification using machine learning: a comprehensive survey," *Complex & Intelligent Systems*, vol. 8, p. 3161–3183, 2022.
- [14] E. Agliari, F. Alemanno, A. Barra and G. D. Marzo, "The emergence of a concept in shallow neural networks," *Neural Networks*, vol. 148, pp. 232-253, 2022.
- [15] S. M. Kasongo and Y. Sun, "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset," *Journal of Big Data*, vol. 7, no. 1, p. 105, 2020.
- [16] R. Tuntun, Kusrini and Kusnawi, "Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, pp. 2111-2119 , 2022.