JIKO (JURNAL INFORMATIKA DAN KOMPUTER)
Oktober 2025, Volume: **9**, No. **3** | Pages 614–625
doi: 10.26798/jiko.v9i3.2126
e-ISSN : 2477-3964 – p-ISSN : 2477-4413

ARTICLE

Peningkatan Performa Prediksi Survival Pasien Gagal Jantung Menggunakan Stacking Ensemble Learning

Improving Survival Prediction Performance of Heart Failure Patients Using Stacking Ensemble Learning

Faiza Rulla Salwa,* Mega Novita, dan Ramadhan Renaldy

Informatika, Fakultas Teknik dan Informatika, Universitas PGRI Semarang, Semarang, Indonesia

*Penulis Korespondensi: faizarullasalwa@gmail.com

(Disubmit 06-08-25; Diterima 16-09-25; Dipublikasikan online pada 20-10-25)

Abstrak

Prediksi kelangsungan hidup pasien gagal jantung merupakan hal krusial untuk mendukung pengambilan keputusan medis yang lebih dini dan tepat. Sebagian besar penelitian sebelumnya cenderung menggunakan model Machine Learning yang kompleks dengan kebutuhan komputasi tinggi, sementara pendekatan yang mengombinasikan model ringan dan efisien masih jarang dieksplorasi. Penelitian ini bertujuan untuk meningkatkan performa prediksi dengan menerapkan metode Stacking Ensemble Learning, yang menggabungkan tiga base learners yaitu Decision Tree, Naive Bayes, dan K-Nearest Neighbor, serta Support Vector Machine sebagai meta-learner. Dataset yang digunakan adalah Heart Failure Clinical Records dari UCI Machine Learning Repository yang telah dikurasi ulang oleh Aadarsh Velu, mencakup 5.000 data pasien dengan 13 atribut klinis. Tahap pra-pemrosesan meliputi standardisasi numerik dan pembagian data dengan teknik stratified sampling menggunakan rasio 80:20. Proses pelatihan dilakukan dengan 5-fold cross-validation serta tuning hyperparameter pada meta-learner menggunakan GridSearchCV untuk memperoleh kombinasi optimal parameter C dan gamma pada kernel RBF. Hasil menunjukkan bahwa model stacking mencapai accuracy 98,7% dan F1-score 0,9791, lebih unggul dibandingkan seluruh model tunggal. Temuan ini membuktikan bahwa strategi penggabungan model ringan dapat memberikan kinerja prediksi yang tinggi tanpa meningkatkan kompleksitas berlebihan, sehingga berpotensi besar untuk diterapkan dalam sistem pendukung keputusan klinis berbasis data, khususnya pada prediksi penyakit kronis.

Kata kunci: Gagal jantung; Sistem pendukung keputusan; Stacking ensemble learning; Machine learning

Abstract

Predicting the survival of patients with heart failure is crucial for enabling earlier and more precise clinical decision-making. Most prior research has utilized complex Machine Learning models with high computational demands, while the combination of lightweight and efficient models remains a less-explored approach. This study aims to enhance prediction performance by implementing a Stacking Ensemble Learning method that combines three base learners, Decision Tree, Naive Bayes, and K-Nearest Neighbor, with a Support Vector Machine as the meta-learner. We utilized the Heart Failure Clinical Records dataset from the UCI Machine Learning Repository, which was re-curated by Aadarsh Velu, comprising 5,000 patient records with 13 clinical attributes. The preprocessing stage involved numerical standardization and data splitting via stratified sampling using an 80:20 ratio. The model was trained using 5-fold cross-validation, with hyperparameter tuning on the meta-learner performed using GridSearchCV to identify the optimal combination of C and gamma for the RBF kernel. Our results demonstrate that the stacking model achieved an accuracy of 98.7% and an F1-score

How to Cite: F. R. Salwa *et al.*, "Peningkatan Performa Prediksi Survival Pasien Gagal Jantung Menggunakan Stacking Ensemble Learning", *JIKO (JURNAL INFORMATIKA DAN KOMPUTER)*, Volume: **9**, No.**3**, Pages 614–625, Oktober 2025, doi: 10.26798/jiko.v9i3.2126.

of 0.9791, significantly outperforming all individual base models. These findings demonstrate that combining lightweight models can yield high predictive performance without excessive complexity, thereby holding significant promise for implementation in data-driven clinical decision support systems, particularly for chronic disease prediction.

KeyWords: Heart failure; Decision support system; Stacking ensemble learning; Machine learning

1. Pendahuluan

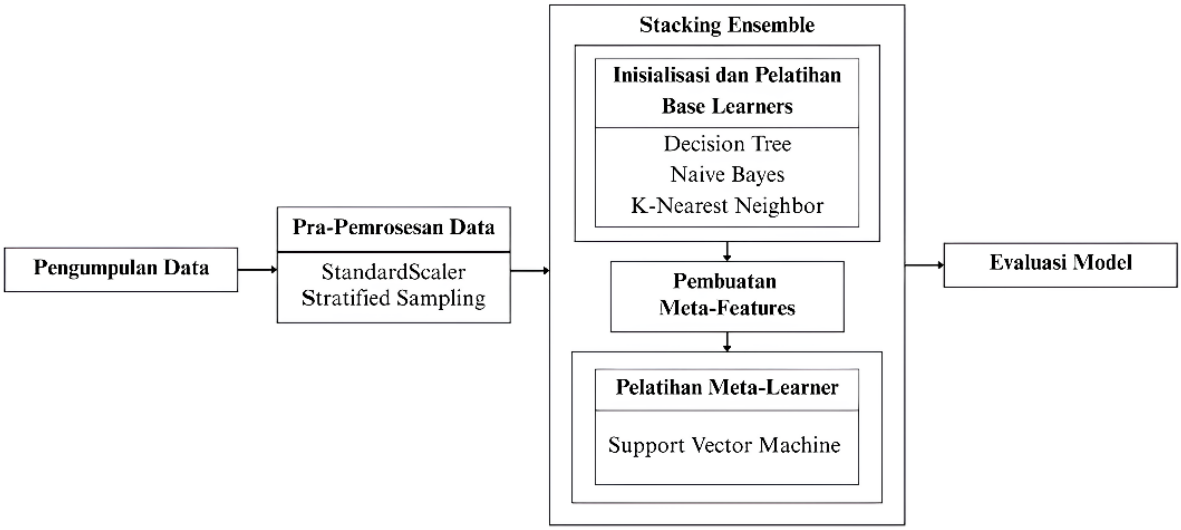
Gagal jantung merupakan salah satu penyakit kronis yang paling banyak dijumpai dalam praktik medis, ditandai oleh penurunan fungsi jantung dalam memompa darah secara efisien ke seluruh tubuh. Kondisi ini menyebabkan gejala seperti kelelahan, kesulitan bernafas, serta pembengkakan di beberapa area [1]. Di Indonesia, beban gagal jantung tergolong tinggi. Berdasarkan laporan *Global Burden of Disease* (GBD), Pada tahun 2019 Indonesia menempati peringkat kedua kasus gagal jantung tertinggi di Asia, dengan angka kejadian mencapai 900,90 per 100.000 penduduk berdasarkan standar usia. Peningkatan sebesar 7,83% dibandingkan tahun 1990 ini seiring dengan penuaan populasi dan tingginya prevalensi faktor risiko seperti hipertensi, diabetes, dan penyakit jantung koroner [2].

Untuk memprediksi *survival* pasien gagal jantung, berbagai pendekatan berbasis teknologi telah dikembangkan, salah satunya melalui penerapan *Machine Learning*. Teknologi ini memungkinkan sistem belajar dari data historis untuk membuat prediksi secara otomatis [3], dan telah banyak digunakan dalam prediksi penyakit kronis [4]. Salah satu pendekatan *Machine Learning* yang berkembang pesat adalah *Ensemble Learning*, yakni model yang menggabungkan beberapa model prediktif untuk meningkatkan akurasi dan stabilitas [5]. Teknik *Stacking*, sebagai bagian dari *Ensemble Learning*, menggunakan beberapa *base learners* dan satu *meta learner* untuk menyatukan hasil prediksi [6]. Teknik ini terbukti mengurangi variansi dan mengurangi resiko *overfitting* terlebih dalam konteks data yang kompleks [7]. Beberapa studi menunjukkan keberhasilan teknik *stacking* dalam domain kesehatan, misalnya Fathima et al. mengombinasikan beberapa model *machine learning* dengan *Logistic Regression* sebagai *meta-learner* untuk meningkatkan prediksi risiko kardiovaskular [8]. Ghasemieh et al. membuktikan bahwa *Stacking Ensemble* mampu meningkatkan *accuracy* hingga 88% pada prediksi *emergency readmission* pasien jantung dengan *meta-learner* yang mempelajari *output* dari *base learners* [9], meskipun metode ini secara umum menimbulkan beban komputasi yang lebih tinggi karena pelatihan beberapa model. Sebagian besar studi terdahulu masih mengandalkan kombinasi model berat [10]. Hal ini menandakan kebutuhan penelitian yang khusus mengevaluasi strategi *Stacking Ensemble Learning* berbasis model ringan yang memperhatikan keterbatasan komputasi perangkat klinis, serta kemudahan implementasi, sehingga menghasilkan solusi prediktif yang lebih efisien untuk sistem pendukung keputusan klinis.

Sejalan dengan kebutuhan tersebut, penelitian ini berupaya meningkatkan performa dan efisiensi prediksi *survival* pasien gagal jantung melalui penerapan pendekatan *Stacking Ensemble Learning*. Model ini mengombinasikan *algoritma Decision Tree, Naive Bayes*, dan *K-Nearest Neighbor* sebagai *base learners*, serta *Support Vector Machine* sebagai *meta learner*. *Decision Tree* dipilih karena kemampuannya menangani data non-linear dan memberikan interpretasi yang jelas [11]. *Naive Bayes* dipilih karena memiliki asumsi independensi atribut yang sederhana, memerlukan estimasi parameter hanya berdasarkan distribusi atribut tunggal, dan menunjukkan kecepatan pelatihan dan inferensi yang relatif rendah [12]. *K-Nearest Neighbor* dipilih karena memprediksi label suatu sampel berdasarkan tetangga terdekat dengan metrik jarak, sehingga mampu menangkap pola lokal dalam distribusi atribut sekaligus memberikan interpretabilitas yang baik [13]. Sedangkan *Support Vector Machine* dipilih karena keunggulannya dalam klasifikasi berdimensi tinggi [14]. Berdasarkan Mahajan et al. kombinasi dari ke empat model tersebut belum diteliti secara mendalam, khususnya pada prediksi kelangsungan hidup pasien gagal jantung [10]. Dataset yang digunakan mencakup atribut klinis seperti usia, tekanan darah, kadar kreatinin serum, dan riwayat merokok [15]. Kinerja model diukur menggunakan metrik *accuracy, precision, recall*, dan *F1-score* [16]. Diharapkan pendekatan ini dapat memberikan kontribusi terhadap sistem prediksi klinis yang lebih akurat, ringan secara komputasi, dan praktis untuk mendukung keputusan pada bidang kesehatan.

2. Metode

Gambar 1 menggambarkan alur tahapan eksperimen dalam penelitian ini, yang menerapkan model *Stacking Ensemble Learning*. Proses dimulai dari pengumpulan dan pra-pemrosesan data, diikuti dengan pelatihan model *Decision Tree, Naive Bayes*, dan *K-Nearest Neighbor* sebagai *base learners*. Selanjutnya, *meta-features* yang dihasilkan digunakan untuk melatih *Support Vector Machine* sebagai *meta learner*, dan hasilnya diuji menggunakan data uji untuk dievaluasi secara menyeluruh.



Gambar 1. Prosedur Eksperimen

2.1 Pengumpulan Data

Dataset yang digunakan pada penelitian ini adalah *Heart Failure Clinical Records* dari *UCI Machine Learning Repository* [17], yang dikurasi ulang oleh Aadarsh Velu di Kaggle [15]. *Dataset* ini mencakup 5.000 data pasien dan memiliki 13 atribut klinis, seperti usia, anemia, *kreatinin fosfokinase*, diabetes, fraksi ejeksi, tekanan darah tinggi, trombosit, kreatinin serum, natrium serum, jenis kelamin, kebiasaan merokok, waktu observasi, dan status kematian [15]. *Dataset* ini telah dipastikan bebas dari *missing values* sehingga tidak memerlukan teknik imputasi. Pada Tabel 1 memberikan informasi mengenai atribut, tipe data, dan deskripsi dari *dataset*.

Tabel 1. Informasi Atribut

Atribut	Tipe Data	Deskripsi
age	Interger	Usia pasien
anaemia	Binary	“1” jika pasien mengalami anemia
creatinine_phosphokinase	Integer	Kadar CPK dalam darah
diabetes	Binary	“1” jika pasien mengidap diabetes
ejection_fraction	Integer	Fraksi ejeksi jantung (%)
high_blood_pressure	Binary	“1” jika tekanan darah tinggi
platelets	Continuous	Jumlah trombosit
serum_creatinine	Continuous	Kadar kreatinin serum
serum_sodium	Integer	Kadar natrium serum
sex	Binary	“1” untuk laki-laki, “0” untuk perempuan
smoking	Binary	“1” jika pasien merokok
time	Integer	Waktu follow-up (hari)
DEATH_EVENT	Binary	“1” jika non-survivor, “0” jika survivor

Tabel 2 menampilkan sampel data pasien yang terdiri dari 13 atribut klinis dan satu variabel target. Atribut

yang dicatat antara lain age, anaemia, creatinine_phosphokinase, diabetes, ejection_fraction, high_blood_pressure, platelets, serum_creatinine, serum_sodium, sex, smoking, dan time. Target klasifikasi yang digunakan adalah DEATH_EVENT, yang menunjukkan status kelangsungan hidup pasien selama masa observasi [15].

Tabel 2. Sampel Dataset

age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
72	1	110	0	25	0	237000	1	140	0	0	65	1
65	0	56	0	25	0	305000	5	130	1	0	207	0
45	0	582	1	38	0	319000	0.9	140	0	0	244	0
60	1	754	1	40	1	328000	1.2	126	1	0	90	0
95	1	582	0	30	0	461000	2	132	1	0	50	1

2.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk meningkatkan kualitas data dan memastikan model dapat dilatih secara optimal [18]. Proses dalam penelitian ini mencakup pemisahan fitur target "DEATH_EVENT" dan standardisasi semua atribut numerik menggunakan *StandardScaler* untuk menyamakan skala, yang penting dilakukan sebelum menerapkan algoritma pembelajaran mesin berbasis jarak atau margin seperti *K-Nearest Neighbor* atau margin seperti *Support Vector Machine* [19]. Penelitian ini tidak menerapkan teknik *oversampling SMOTE*, karena distribusi kelas dijaga melalui *stratified sampling* dengan rasio 80:20, sehingga proporsi survivor dan non-survivor tetap terwakili pada data latih dan uji [20]. Menjaga distribusi survivor dan non-survivor secara proporsional sangat krusial dalam konteks medis, karena kesalahan dalam mengklasifikasikan non-survivor sebagai kelas minoritas yang dapat berdampak besar pada reliabilitas sistem prediksi klinis.

2.2.1 StandardScaler

Dalam penelitian ini, atribut numerik dinormalisasi menggunakan *StandardScaler* untuk menyamakan skala atribut numerik dengan mengubah distribusinya menjadi memiliki *mean* nol dan standar deviasi satu. Proses standardisasi ini bertujuan mengurangi dominasi atribut dengan skala besar terhadap model, serta mempercepat konvergensi dalam pelatihan model *machine learning* [21]. Selain itu, transformasi ini membantu menjaga stabilitas numerik selama proses optimasi, terutama pada algoritma yang sensitif terhadap perbedaan skala antar atribut [22]. Rumus perhitungan *mean* (μ), standar deviasi (σ) dan hasil standarisasi (z) yang digunakan dalam proses standardisasi melalui metode *StandardScaler* ditunjukkan pada Persamaan (1), (2) dan (3).

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_1 \tag{1}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x-\mu)^2} \tag{2}$$

$$z = \frac{(x-\mu)}{\sigma} \tag{3}$$

μ (Rata-rata atau mean)	=	Nilai rata-rata seluruh data pada suatu atribut.
σ (Standar Deviasi)	=	Ukuran sebaran data dari rata-rata.
n (Jumlah Sampel)	=	Total observasi dalam *dataset*.
x (Nilai asli)	=	Nilai asli per data/observasi.
z (z-score)	=	Nilai hasil standardisasi.

2.2.2 Stratified Sampling

Teknik *Stratified Sampling* digunakan untuk membagi data dengan mempertahankan representasi proporsional dari kelas target, sehingga distribusi kelas "DEATH_EVENT" tidak berubah antara data latih dan uji. Ini sangat penting dalam konteks klasifikasi medis yang rentan terhadap ketidakseimbangan kelas, karena prediksi pada kelas minoritas sering kali lebih kritis namun sulit dikenali oleh model yang dilatih secara tidak proporsional. Selain itu, pada tahap pembagian data dilakukan pengacakan (*shuffle=True*)

sebelum stratifikasi menggunakan kode *train_test_split(X_scaled, y, test_size=0.2, stratify=y, shuffle=True, random_state=42*), sehingga distribusi tetap terjaga namun data tidak bias terhadap urutan. Dengan menggunakan metode *stratified sampling*, model menjadi lebih sensitif terhadap kelas-minoritas dan distribusi target, menghasilkan akurasi kelas yang lebih baik dan hasil yang lebih seimbang [23].

2.3 Stacking Ensemble

Penelitian ini menerapkan model *Stacking Ensemble Learning* sebagai pendekatan utama dalam klasifikasi kelangsungan hidup pasien gagal jantung. Proses diawali dengan pelatihan tiga *base learners*, yaitu *Decision Tree, Naive Bayes*, dan *K-Nearest Neighbor*, secara paralel menggunakan data pelatihan yang telah melalui proses standardisasi. Ketiga model ini dipilih karena mewakili paradigma yang berbeda, yaitu berbasis pohon keputusan, probabilitas, dan jarak antar data, sehingga diharapkan dapat menghasilkan prediksi yang saling melengkapi. Dalam implementasinya, *base learners* dikonfigurasi dengan parameter khusus, yakni *Naive Bayes* dengan asumsi distribusi *Gaussian*, yang dikonfigurasi menggunakan nilai *var_smoothing* sebesar $1e-9$ untuk mengurangi kemungkinan pembagian nol pada perhitungan probabilitas, *K-Nearest Neighbor* dengan lima tetangga terdekat ($k = 5$) dan pembobotan berbasis jarak menggunakan metrik *Euclidean* ($p = 2$), serta *Decision Tree* dengan kedalaman maksimum empat dan jumlah minimum lima sampel pada setiap pemisahan, disertai *random state 42* untuk menjamin reprodusibilitas. Hasil prediksi dari setiap *base learners* kemudian digabungkan untuk membentuk *meta-features*, yakni fitur baru yang merepresentasikan performa awal masing-masing model terhadap data pelatihan [24]. Pada tahap selanjutnya, setiap *base learner* dalam klasifikasi biner menghasilkan dua probabilitas (kelas survivor dan non-survivor), sehingga tiga model dasar membentuk total $23 = 6$ *meta-features*. Enam fitur probabilistik ini kemudian digunakan oleh *meta-learner* untuk mempelajari pola gabungan dan menghasilkan keputusan akhir yang lebih akurat.

Meta-learner yang digunakan adalah *Support Vector Machine*, yang dilatih menggunakan himpunan *meta-features* untuk menghasilkan prediksi akhir. *Support Vector Machine* dipilih karena kemampuannya menemukan *hyperplane* optimal yang memisahkan kelas secara efektif, serta kemampuannya dalam menangani data berdimensi tinggi dan pola non-linear secara akurat [14]. Pada tahap implementasi, digunakan *kernel Radial Basis Function* (RBF) dengan opsi *probability estimation* serta *class weight balancing* untuk mengatasi distribusi kelas yang tidak seimbang. Selain itu, dilakukan proses *hyperparameter tuning* menggunakan *GridSearchCV* dengan *5-fold cross-validation* guna memperoleh kombinasi parameter terbaik, yakni nilai $C\{1, 10, 50\}$ dan $gamma\{0.01, 0.1, 1\}$, dengan akurasi sebagai metrik evaluasi utama. Seluruh proses pelatihan dilakukan menggunakan teknik *cross-validation* untuk menghindari tercampurnya data latih dan data uji, serta meningkatkan kemampuan generalisasi model terhadap data baru [24]. Dengan demikian, struktur *Stacking Ensemble* tersebut dirancang untuk memanfaatkan keunggulan dari masing-masing model, sehingga menghasilkan performa prediksi yang lebih baik dibandingkan model tunggal.

2.4 Evaluasi Model

Evaluasi performa model dilakukan menggunakan matriks *accuracy, precision, recall*, dan *F1-score* untuk menilai kinerja secara menyeluruh. *Accuracy* mengukur rasio prediksi yang benar terhadap seluruh prediksi yang dilakukan. *Precision* menunjukkan seberapa tepat model dalam mengidentifikasi kasus positif, sementara *recall* menilai kemampuannya menangkap seluruh kasus positif yang sebenarnya. *F1-score* digunakan sebagai ukuran gabungan untuk mengatasi ketidakseimbangan kelas karena merupakan rata-rata harmonik antara *precision* dan *recall* [16]. Matriks evaluasi tersebut sebagai penilaian kinerja model *Stacking Ensemble Learning* serta pembandingan dengan performa model tunggal seperti *Decision Tree, Naive Bayes, K-Nearest Neighbor* dan *Support Vector Machine*. Rumus perhitungan untuk *accuracy, precision, recall*, dan *F1-score* masing-masing ditulis pada persamaan (4), (5), (6), dan (7).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

TP (True Positive) : Jumlah kasus positif yang diprediksi benar oleh model
TN (True Negative) : Jumlah kasus negatif yang diprediksi benar oleh model
FP (False Positive) : Jumlah kasus negatif yang secara keliru diprediksi sebagai positif oleh model.
FN (False Negative) : Jumlah kasus positif yang secara keliru diprediksi sebagai negatif oleh model.

3. Hasil dan Pembahasan

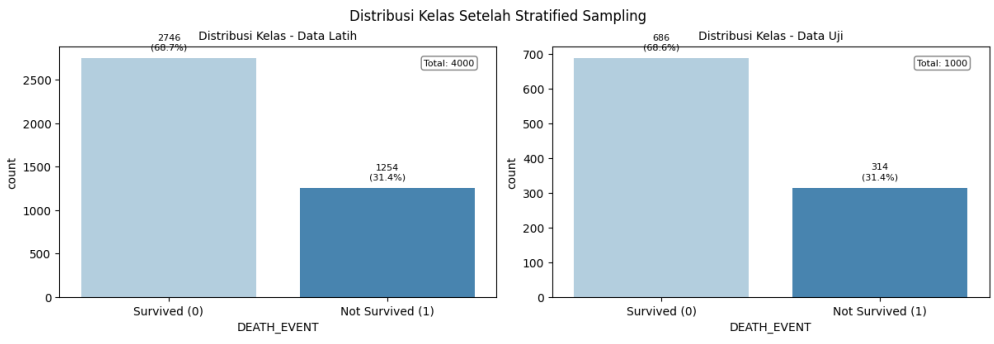
3.1 Hasil Pra-Pemrosesan Data

Agar algoritma dapat belajar secara efisien, seluruh data numerik dalam *dataset* terlebih dahulu melalui tahap standardisasi menggunakan *StandardScaler*. Proses ini bertujuan menghilangkan ketidakseimbangan skala antar atribut, yang jika dibiarkan dapat mempengaruhi kinerja model, terutama yang berbasis jarak seperti *K-Nearest Neighbor* atau margin seperti *Support Vector Machine*. *StandardScaler* mentransformasikan setiap atribut agar memiliki nilai rata-rata sekitar nol dan standar deviasi mendekati satu. Sehingga setiap atribut berkontribusi secara seimbang dalam perhitungan model, mempercepat proses pelatihan, dan mencegah dominasi atribut berskala besar. Hasil proses ini dapat dilihat pada Tabel 3 di bawah, yang memperlihatkan perubahan distribusi nilai atribut.

Tabel 3. Hasil Standarisasi Data

Atribut	Sebelum		Sesudah		
	Mean	Standar Devisiasi	Min	Median	Max
age	60.29	11.7	-1.735	-0.025	2.968
creatinine_phosphokinase	586.76	976.73	-0.577	-0.347	7.448
ejection_fraction	37.73	11.51	-2.061	0.023	3.671
platelets	265075.4	97999.76	-2.449	-0.018	5.969
serum_creatinine	1.37	1.01	-0.861	-0.267	7.954

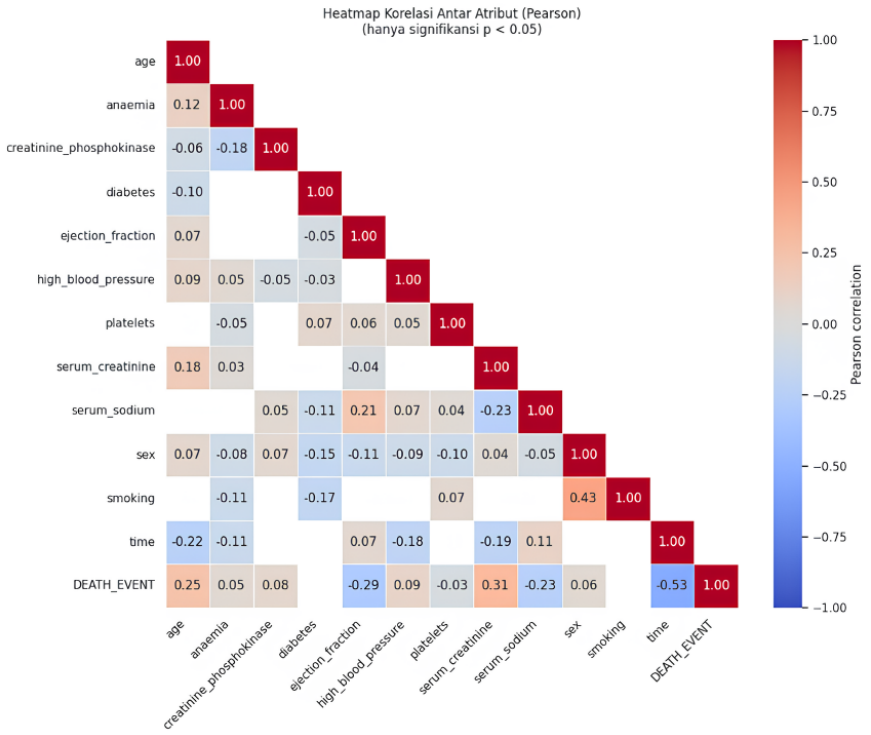
Setelah proses standardisasi dilakukan, pembagian *dataset* dilakukan menggunakan teknik Stratified Sampling dengan rasio 80:20. Dari total 5.000 data pasien, sebanyak 4.000 digunakan sebagai data latih dan 1.000 sebagai data uji. Teknik *Stratified Sampling* dipilih karena mampu menjaga proporsi kelas target "DEATH_EVENT" secara seimbang di kedua subset data. Seperti ditampilkan pada Gambar **??**, data latih terdiri atas 2.746 survivor dan 1.254 non-survivor, sementara data uji mencakup 686 survivor dan 314 non-survivor. Proporsi yang tetap konsisten ini penting untuk menghindari bias model terhadap kelas mayoritas dan memastikan representasi yang adil dalam proses pelatihan dan evaluasi. Pengaturan `random_state=42` digunakan untuk menjamin konsistensi hasil pembagian data.



Gambar 2. Visualisasi Pembagian Data dengan Stratifikasi

Heatmap pada Gambar 3 menggambarkan matriks korelasi antar atribut dengan metode Pearson, di mana hanya korelasi signifikan ($p < 0,05$) yang ditampilkan untuk menyaring hubungan yang relevan. Visualisasi ini berfungsi sebagai bagian dari eksplorasi data untuk memahami pola linier antar variabel serta

potensi redundansi antar atribut. Meskipun sebagian besar atribut menunjukkan korelasi lemah hingga sedang, hubungan yang lebih menonjol terlihat antara *time* dengan DEATH_EVENT (negatif) dan *age* dengan DEATH_EVENT (positif). Temuan ini memberikan wawasan awal mengenai faktor-faktor klinis yang potensial berkontribusi terhadap kelangsungan hidup pasien gagal jantung, misalnya usia yang lebih tua dan *time* yang lebih singkat berkaitan dengan peningkatan risiko kematian. Informasi ini penting sebagai langkah validasi awal dalam pra-pemrosesan data, sebelum atribut dimanfaatkan lebih lanjut pada tahap pemodelan prediktif.

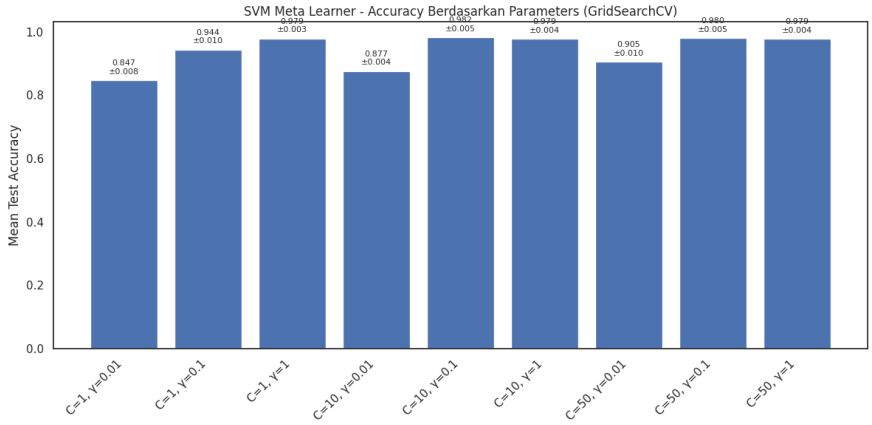


Gambar 3. Heatmap Korelasi Atribut

3.2 Hasil *Stacking Ensemble*

Model *stacking* pada penelitian ini diawali dengan proses *hyperparameter tuning* terhadap model *Support Vector Machine* yang digunakan sebagai *meta-learner*. *Tuning* dilakukan menggunakan pendekatan *GridSearchCV* dengan validasi silang lima lipatan (*5-fold cross-validation*) untuk mengoptimalkan dua parameter utama, yaitu C yang mengatur regularisasi, di mana nilai kecil menghasilkan margin yang lebih lebar dengan toleransi *error* lebih tinggi, sedangkan nilai besar mempersempit margin dan meningkatkan risiko *overfitting* dan *gamma* sebagai pengatur kepekaan *kernel RBF*, nilai besar dapat meningkatkan kompleksitas dan sensitivitas terhadap *noise*, sedangkan nilai kecil menghasilkan batas keputusan yang lebih halus. Rentang nilai yang diuji adalah1, 10, 50 untuk C dan 0.01, 0.1, 1 untuk *gamma*, sehingga total terdapat 9 kombinasi parameter yang dievaluasi. Visualisasi hasil tuning disajikan pada gambar 4 dalam bentuk *bar chart*, yang menunjukkan bahwa kombinasi optimal diperoleh ketika C = 10 dan *gamma* = 0.1. Kombinasi ini menghasilkan nilai akurasi tertinggi, yaitu 0.982, selama proses *cross-validation*. Selain itu, penambahan *class_weight='balanced'* memungkinkan model untuk menangani ketidakseimbangan kelas, yang sering menjadi isu pada data medis. Dalam konteks klinis, kesalahan prediksi memiliki implikasi serius, seperti *false positive* dapat menunda intervensi medis yang seharusnya segera dilakukan, sedangkan *false negative* dapat mengurangi kualitas perawatan. Dengan *kernel RBF*, *Support Vector Machine* mampu mengakomodasi pola non-linear, sedangkan pemilihan parameter yang tepat membantu menjaga keseimbangan antara bias dan variansi, yang sangat penting dalam struktur *meta-learner*.

Setelah parameter optimal dari model *Support Vector Machine* diperoleh melalui *GridSearchCV*, model *Stacking Ensemble* dibentuk dengan tiga *base learners* yang mewakili pendekatan klasifikasi berbeda, yakni *Naive Bayes* (probabilistik), *K-Nearest Neighbors* (berbasis jarak), dan *Decision Tree* (berbasis aturan). Ketiga *base learners* tersebut dikonfigurasi dengan parameter khusus, yakni *Naive Bayes* dengan asumsi distribusi



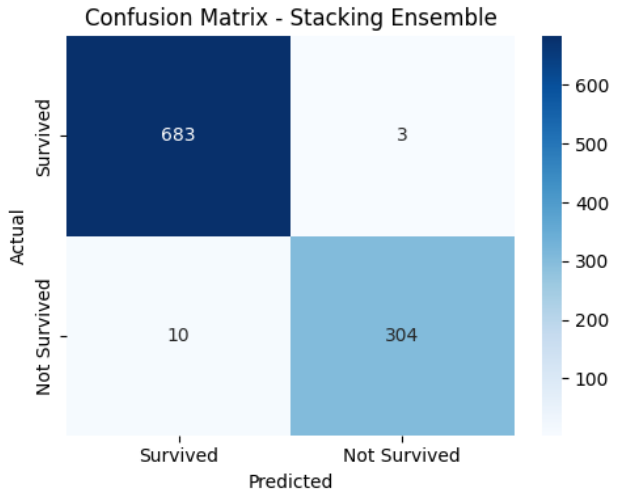
Gambar 4. Visualisasi Hasil Hyperparameter Tuning Meta Learner

Gaussian, yang dikonfigurasi menggunakan nilai *var_smoothing* sebesar 1e-9 untuk mengurangi kemungkinan pembagian nol pada perhitungan probabilitas, *K-Nearest Neighbo*r dengan lima tetangga terdekat (k=5) dan pembobotan berbasis jarak menggunakan metrik *Euclidean* (p=2), serta *Decision Tree* dengan *max_depth=4* dan *min_samples_split=5*, disertai *random state=42* untuk menjamin reprodusibilitas. Konfigurasi terbaik model *Support Vector Machine* kemudian digunakan sebagai *meta-learner* untuk menggabungkan *output* dari ketiga model tersebut. Penggunaan *passthrough=False* dalam model *stacking* berarti bahwa *meta-learner* hanya menerima prediksi dari *base learners* sebagai *input*, tanpa menyertakan atribut asli dari data. Pendekatan ini dipilih untuk menjaga agar model tetap sederhana dan menghindari kemungkinan *overfitting* akibat informasi berlebih dari fitur mentah. Meskipun *meta-learner* tidak memiliki akses langsung ke fitur awal, metode ini tetap efektif karena fokus pada pola prediktif yang telah dipelajari oleh *base learners.*

3.3 Hasil Evaluasi Model

Pada tahap evaluasi model, performa model *Stacking Ensemble Learning* diukur menggunakan metrik evaluasi seperti *accuracy, precision, recall*, dan *F1-score.* Selain itu, visualisasi *confusion matrix* juga digunakan untuk memberikan pemahaman lebih lanjut mengenai distribusi prediksi model terhadap data aktual. Dapat dilihat pada Gambar 5, model *Stacking Ensemble Learning* mampu mengklasifikasikan dengan benar 683 pasien *survivor (True Negative)* dan 304 pasien *non-survivor (True Positive).* Hanya ditemukan 3 kesalahan prediksi *False Positive*, yang berarti pasien diperkirakan tidak bertahan hidup padahal sebenarnya bertahan, serta 10 kasus *False Negative*, yaitu pasien yang diprediksi bertahan hidup tetapi pada kenyataannya tidak. Kedua jenis kesalahan ini memiliki implikasi klinis yang berbeda. *False Positive* dapat menyebabkan intervensi medis yang tidak perlu, sedangkan *False Negative* lebih berisiko karena dapat mengurangi kualitas perawatan intensif bagi pasien yang sebenarnya membutuhkan. Dengan jumlah kesalahan yang rendah, model ini memperlihatkan kemampuan prediksi yang kuat dan konsisten, sehingga layak dipertimbangkan sebagai alat bantu dalam prediksi klinis. Tabel 4 memberikan gambaran lengkap performa dari lima model yang diuji, yaitu *Decision Tree, Naive Bayes, K-Nearest Neighbors, Support Vector Machine*, dan *Stacking Ensemble.* Hasil menunjukkan bahwa *Stacking Ensemble* mencapai performa tertinggi di semua metrik utama, dengan nilai *accuracy* sebesar 0.987, *precision* 0.9902, *recall* 0.9682, dan *F1-score* 0.9791. Dibandingkan model tunggal, *Stacking Ensemble* tidak hanya unggul secara angka tetapi juga menunjukkan keseimbangan yang lebih baik antara *precision* dan *recall.* Hal ini penting terutama dalam konteks prediksi medis, di mana keseimbangan antara kesalahan *false positive* dan *false negative* memiliki peran penting terhadap keputusan klinis.

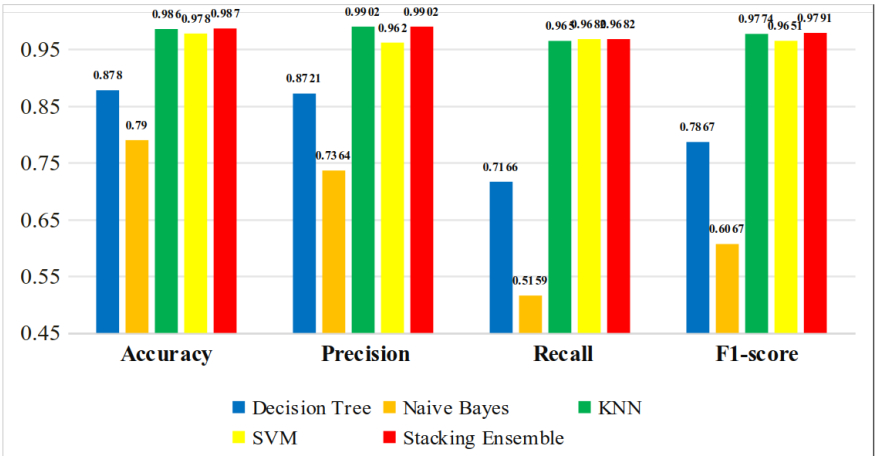
Visualisasi hasil evaluasi yang ditampilkan pada Gambar 6 memberikan representasi yang jelas dari keunggulan model *Stacking Ensemble.* Tampak bahwa stacking secara konsisten unggul di setiap metrik yang diuji, membuktikan bahwa strategi kombinasi model, baik yang berbasis aturan, probabilistik, maupun jarak, sehingga dapat meningkatkan ketahanan model terhadap *noise* serta variasi data. Konsistensi tersebut memperkuat argumen bahwa pendekatan *Stacking Ensemble* dapat menjadi solusi efektif dalam pengembangan sistem prediksi medis berbasis *machine learning* yang andal dan presisi tinggi.



Gambar 5. Confusion Matrix Stacking Ensemble

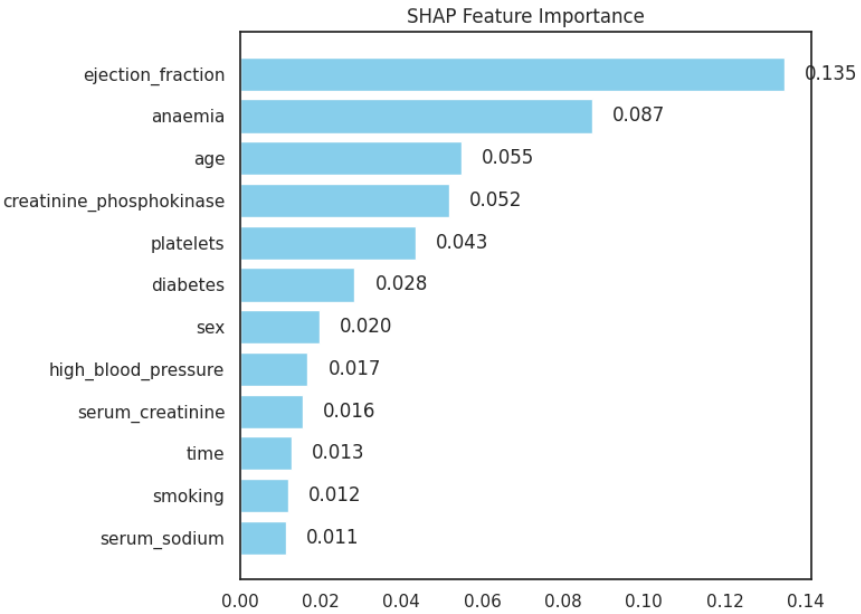
Tabel 4. Hasil Evaluasi Model

Model	*Accuracy*	*Precision*	*Recall*	*F1-score*
Decision Tree	0.878	0.8721	0.7166	0.7867
Naive Bayes	0.79	0.7364	0.5159	0.6067
KNN	0.986	0.9902	0.965	0.9774
SVM	0.978	0.962	0.9682	0.9651
Stacking Ensemble	0.987	0.9902	0.9682	0.9791



Gambar 6. Grafik Perbandingan Hasil Evaluasi Model

Gambar 7 menyajikan *bar chart* yang menunjukkan rata-rata dampak tiap atribut terhadap prediksi model berdasarkan nilai SHAP, di mana nilai yang lebih besar menunjukkan kontribusi atribut yang lebih kuat terhadap keputusan model. Dari grafik terlihat bahwa *ejection_fraction* memiliki pengaruh terbesar (≈0,135), diikuti oleh *anaemia* (≈0,087), *age* (≈0,055), dan *creatinine_phosphokinase* (≈0,052), sedangkan fitur seperti *serum_sodium* dan *smoking* memberikan pengaruh relatif kecil (≈0,011–0,012). Interpretasi ini menunjukkan bahwa variabel yang berkaitan dengan fungsi jantung, status *hematologis*, dan indikator fungsi organ paling menentukan prediksi DEATH_EVENT pada model kami, sehingga layak menjadi prioritas dalam pengamatan klinis dan validasi lebih lanjut. Perlu dicatat bahwa *SHAP* menggambarkan kontribusi prediktif dan dapat dipengaruhi oleh korelasi antar atribut; oleh karena itu hasil ini sebaiknya digunakan bersama analisis statistika dan konsultasi klinis untuk menentukan langkah intervensi atau reduksi atribut.



Gambar 7. Bar Chart dampak atribut dengan SHAP

Keberhasilan model *Stacking Ensemble* dalam penelitian ini berasal dari kemampuannya menyatukan hasil prediksi *base learners* sehingga kelemahan individu dapat tertutupi. *Naive Bayes* memberikan generalisasi yang cepat namun mengasumsikan *independensi* atribut, suatu asumsi yang jarang terpenuhi pada data klinis. *K-Nearest Neighbor* sensitif terhadap *noise* dan skala atribut, sedangkan *Decision Tree berisiko overfitting* bila pohon terlalu dalam. *Support Vector Machine* mempelajari pola gabungan tersebut dan menghasilkan keputusan akhir yang lebih stabil dan akurat, sebagaimana tercermin pada kenaikan nilai *accuracy* dan *F1-score*. Secara klinis, peningkatan stabilitas prediksi ini meningkatkan keandalan stratifikasi risiko, misalnya untuk memprioritaskan pasien dengan fraksi ejeksi rendah atau kadar kreatinin tinggi agar mendapat monitoring intensif, serta mengurangi kemungkinan *false negatives* yang paling berbahaya bagi keselamatan pasien. Dengan demikian, *Stacking Ensemble* meningkatkan performa, tetapi juga menambah nilai praktis dalam pengambilan keputusan klinis, dan efisiensi sumber daya, meskipun demikian, validasi eksternal dan kolaborasi dengan tenaga medis tetap diperlukan sebelum penerapan luas di praktik klinis.

4. Simpulan

Penelitian ini menyimpulkan bahwa penerapan model *Stacking Ensemble Learning* yang menggabungkan model *Decision Tree, Naive Bayes*, dan *K-Nearest Neighbor* sebagai *base learners*, serta *Support Vector Machine* sebagai *meta-learner*, terbukti mampu meningkatkan akurasi prediksi kelangsungan hidup pasien gagal jantung secara signifikan. Evaluasi model menunjukkan bahwa pendekatan ini tidak hanya memberikan hasil lebih akurat dibandingkan model tunggal, tetapi juga lebih stabil dan seimbang dalam menangani distribusi kelas. Keunggulan utama terletak pada kemampuannya mengompensasi kelemahan individu dari masing-masing *base learner*, yang ditunjukkan melalui *accuracy* tertinggi sebesar 98.7% dan *F1-score* 0.9791. Meskipun menjanjikan, penelitian ini terbatas karena hanya dilakukan pada satu dataset sehingga

masih terdapat risiko *overfitting* dan keterbatasan generalisasi. oleh karena itu studi lanjutan harus mencakup validasi eksternal, uji prospektif dalam alur klinis, serta eksplorasi teknik optimasi *hyperparameter*, misalnya *Bayesian optimization* dan alternatif *meta-learner* untuk memperkuat robustnes. Secara praktis, kombinasi model ringan menawarkan efisiensi komputasi yang memungkinkan *deployment on-device* atau integrasi ke dalam *Clinical Decision Support System* dan prioritisasi pasien berisiko tinggi, namun integrasi tersebut harus disertai proses verifikasi klinis, kalibrasi model, dan audit keselamatan sebelum penggunaan rutin.

Pustaka

[1] NHS (UK). (2023) Heart failure. NHS. [Online]. Available: https://www.nhs.uk/conditions/heart-failure/

[2] J. Feng, Y. Zhang, and J. Zhang, "Epidemiology and burden of heart failure in asia," *JACC Asia*, vol. 4, no. 4, pp. 249–264, 2024.

[3] A. R. Pratama, F. Wabula, H. Ilmandry, M. L. Isabela, M. Raharjo, and R. Sianipar, "Literature review the impact of machine learning in modern industries," *Nian Tana Sikk. J. Ilm. Mhs.*, vol. 3, no. 1, pp. 177–182, 2025.

[4] F. M. Delpino, Â. K. Costa, S. R. Farias, A. D. P. Chiavegatto Filho, R. A. Arcêncio, and B. P. Nunes, "Machine learning for predicting chronic diseases: a systematic review," *Public Health*, vol. 205, pp. 14–25, 2022.

[5] B. Naderalvojoud and T. Hernandez-Boussard, "Improving machine learning with ensemble learning on observational healthcare data," in *AMIA Annual Symposium Proceedings*, 2023, pp. 521–529. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10785929/

[6] M. I. H. Siddiqui *et al.*, "Accelerated and accurate cervical cancer diagnosis using a novel stacking ensemble method with explainable ai," *Informatics Med. Unlocked*, vol. 56, p. 101657, 2025.

[7] L. Nguyen Van and G. Lee, "Optimizing stacked ensemble machine learning models for accurate wildfire severity mapping," *Remote Sens.*, vol. 17, no. 5, p. 854, 2025.

[8] M. D. Fathima, S. P. Raja, K. Jayanthi, and R. Hariharan, "Optistack classifier: optimized stacking framework with ensemble feature engineering for enhanced cardiovascular risk prediction," *Inflamm. Res.*, vol. 74, no. 1, p. 88, 2025.

[9] A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef, "A novel machine learning model with stacking ensemble learner for predicting emergency readmission of heart-disease patients," *Decis. Anal. J.*, vol. 7, p. 100242, 2023.

[10] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: A review," *Healthcare*, vol. 11, no. 12, p. 1808, 2023.

[11] W. Yang, N. Ahmed, and A. L. C. Barczak, "Comparative analysis of machine learning algorithms for ckd risk prediction," *IEEE Access*, vol. 12, pp. 45 321–45 333, 2024.

[12] T. Ige, C. Kiekintveld, A. Piplai, A. Waggler, O. Kolade, and B. H. Matti, "An investigation into the performances of the current state-of-the-art naive bayes, non-bayesian and deep learning based classifier for phishing detection: A survey," *arXiv Prepr.*, vol. 2411.16751, 2024.

[13] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing k-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *J. Big Data*, vol. 11, no. 1, p. 113, 2024.

[14] R. R. Archana, S. V., A. M., V. A. P., R. M., and P. S., "Performance evaluation of cardiovascular disease prediction using supervised machine learning algorithms," in *2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC)*, 2024, pp. 1–7.

[15] A. Velu. (2023) Heart failure prediction - clinical records. Kaggle. [Online]. Available: https://www.kaggle.com/datasets/aadarshvelu/heart-failure-prediction-clinical-records

[16] S. Tabassum *et al.*, "A machine learning-based framework for heart disease diagnosis using a comprehensive patient cohort," *Comput. Mater. Contin.*, vol. 84, no. 1, pp. 1253–1278, 2025.

[17] UCI Machine Learning Repository, "Heart failure clinical records," 2020, uCI Machine Learning Repository, Irvine, CA.

[18] S. Mali and K. Veeramani, "Heart attack prediction using ensemble learning," in *2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*, 2024, pp. 1–6.

[19] Z. L. Thakker and S. H. Buch, "Effect of feature scaling pre-processing techniques on machine learning algorithms to predict particulate matter concentration for gandhinagar, gujarat, india," *Int. J. Sci. Res. Sci. Technol.*, vol. 11, no. 1, pp. 410–419, 2024.

[20] S. Saha *et al.*, "Machine learning enhanced expert system for detecting heart failure decompensation using patient reported vitals and electronic health records," *Sci. Rep.*, vol. 15, no. 1, p. 30979, 2025.

[21] H. A. Elzeheiry, S. Barakat, and A. Rezk, "Different scales of medical data classification based on machine learning techniques: A comparative study," *Appl. Sci.*, vol. 12, no. 2, p. 919, 2022.

[22] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, p. 105524, 2020.

[23] S. Shetty, P. K. Gupta, M. Belgiu, and S. K. Srivastav, "Assessing the effect of training sampling design on the performance of machine learning classifiers for land cover mapping using multi-temporal remote sensing data and google earth engine," *Remote Sens.*, vol. 13, no. 8, pp. 1–22, 2021.

[24] M. D. Teja and G. M. Rayalu, "Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance," *BMC Cardiovasc. Disord.*, vol. 25, p. 212, 2025.