

# PREDIKSI PASIEN DENGAN PENYAKIT KARDIOVASKULAR MENGUNAKAN RANDOM FOREST

Mochammad Anshori<sup>1</sup>, Nindynar Rikatsih<sup>2</sup>, M. Syauqi Haris<sup>3</sup>

<sup>1, 2, 3</sup>Fakultas Sains, Teknologi dan Kesehatan, ITSK RS.DR. Soepraoen Kesdam V/BRW, Malang  
<sup>1</sup>moanshori@itsk-soepraoen.ac.id, <sup>2</sup>nindynar@itsk-soepraoen.ac.id,  
<sup>3</sup>haris@itsk-soepraoen.ac.id

Diterima pada 26 Desember 2022; disetujui pada 30 Januari 2023; dan diterbitkan pada 3 Maret 2023.

---

## Abstrak

*Cardiovascular disease* adalah salah satu penyakit mematikan di dunia. Hal ini dibuktikan dengan data yang dikeluarkan oleh WHO yang menunjukkan kasus kematian sekitar 18 juta. Penyakit ini mengakibatkan berhentinya detak jantung yang menjadi sumber utama kehidupan dari tubuh manusia. Penyakit ini disebabkan oleh berbagai macam hal termasuk gaya hidup yang kurang sehat. Contohnya adalah mengonsumsi rokok dan alkohol. Selain itu disebabkan juga oleh faktor lain yaitu masalah kesehatan seperti darah tinggi, kolesterol, diabetes, depresi ataupun kecemasan. Penyakit *cardiovascular* cenderung susah untuk disembuhkan, oleh karena itu diperlukan prediksi yang tepat dan akurat dalam mendiagnosis pasien. Salah satu metode dalam melakukan prediksi dengan menggunakan teknik pembelajaran mesin. Didalam pembelajaran mesin terdapat berbagai macam metode yang bisa digunakan, salah satunya adalah dengan metode berbasis pohon keputusan yaitu *random forest*. Sebelum *random forest* diimplementasikan untuk membuat model, data dilakukan pra-proses dengan normalisasi dan menerapkan *cross-validation* dengan *k-fold* = 10. Hasil prediksi dengan *random forest* pada penelitian ini memberikan akurasi sebesar 98%. Akurasi ini lebih tinggi jika dibandingkan dengan penelitian sebelumnya dengan dataset yang sama yaitu 96.75% menggunakan *ensemble method* dan 91.61% dengan *logistic regression*. Dengan dasar tersebut membuktikan bahwa *random forest* mampu digunakan untuk melakukan prediksi penyakit kardiovaskular.

**Kata Kunci:** penyakit kardiovaskular, *tree model*, *random forest*, pembelajaran mesin.

---

## Abstract

*Cardiovascular disease* is one of the deadliest diseases in the world. This is evidenced by data released by WHO which shows around 18 million deaths. This disease causes the cessation of the heartbeat which is the main source of life for the human body. This disease is caused by various things including an unhealthy lifestyle. Examples are consuming cigarettes and alcohol. In addition, it is also caused by other factors, namely health problems such as high blood pressure, cholesterol, diabetes, depression, or anxiety. The cardiovascular disease tends to be difficult to cure, therefore a precise and accurate prediction is needed in diagnosing patients. One method of making predictions is using machine learning techniques. In machine learning, there are various methods that can be used, one of which is the decision tree-based method, namely random forest. Before the random forest is implemented to create a model, the data is pre-processed by normalizing and applying cross-validation with *k-fold* = 10. The prediction results with the random forest in this study provide an accuracy of 98%. This accuracy is higher when compared to previous studies with the same dataset, namely 96.75% using the ensemble method and 91.61% with logistic regression. On this basis, it proves that the random forest can be used to predict cardiovascular disease.

**Key Words:** cardiovascular disease, tree model, random forest, machine learning.

---

## 1. Pendahuluan

Penyakit kardiovaskular menjadi ancaman bagi penduduk dunia karena terbukti menjadi salah satu penyebab kematian yang terbilang cukup tinggi [1]. Berdasarkan hasil survei dari WHO (*World Health Organization*), penyakit kardiovaskular (CVD) menjadi

penyebab kematian primer dengan data sekitar 18 juta kasus banyaknya [2]. Hampir sebanyak 31% kematian didunia disebabkan oleh penyakit ini [3]. Penyakit ini menyumbang satu dari tiga kasus kematian. Bahaya dari penyakit kardiovaskular adalah berhenti bekerjanya jantung. Jika jantung berhenti berdetak akan

menghentikan cara kerja organ tubuh yang lainnya [4]. Di Indonesia sendiri penyakit kardiovaskular diprediksi akan terus meningkat pada tahun 2030 hingga mencapai 23.3 juta kasus kematian [1]. Oleh karena itu, penyakit kardiovaskular dapat dikatakan penyakit yang mematikan karena memiliki tingkat kematian yang tinggi.

Penyakit kardiovaskular adalah istilah yang dipakai untuk menggambarkan kondisi yang mempengaruhi jantung atau pembuluh darah [3]. Dapat dikatakan penyakit kardiovaskular menyebabkan serangan jantung. Penyebab dari timbulnya penyakit kardiovaskular sangat beragam dan tidak pasti. Salah satu penyebabnya adalah karena gaya hidup yang tidak sehat yang menyebabkan kecemasan dan depresi [3]. Dengan kecemasan dapat mempengaruhi aliran darah yang beredar didalam tubuh. Apabila tidak normal seperti mendadak detak jantung berakselerasi kemudian terjadi perlambatan, hal inilah yang dapat menyebabkan serangan jantung. Selain itu karena gaya hidup yang tidak sehat seperti mengkonsumsi alkohol yang berlebih dan rokok. Selain itu bisa juga disebabkan oleh riwayat penyakit yang telah lama diderita oleh pasien. Contohnya adalah diabetes, hipertensi, tingginya tingkat kolesterol dalam darah dan denyut nadi yang tidak normal [5]. Bervariasinya hal yang menyebabkan penyakit kardiovaskular menjadikan penyakit ini susah untuk dideteksi.

Penyakit kardiovaskular dapat menyerang penderitanya dan tidak mudah untuk disembuhkan sehingga ketepatan dalam mendiagnosis pasien menjadi pekerjaan yang penting di bidang medis [6]. Disisi lain susahnya dalam mengidentifikasi penyakit jantung karena banyaknya faktor yang berkontribusi sehingga yang menjadi tantangan adalah bagaimana melakukan prediksi secara tepat dengan akurasi tinggi. Hal ini bertujuan untuk memberikan perawatan dan pencegahan yang tepat. Penelitian ini menyarankan prediksi menggunakan teknik pembelajaran mesin.

Kegunaan pembelajaran mesin mengalami kenaikan pada beberapa bidang. Termasuk pada bidang industri, pertanian, transportasi, keuangan, militer, dan lain sebagainya [7]. Seiring perkembangannya, saat ini pembelajaran mesin dapat digunakan dan diterapkan pada bidang kesehatan. Pada domain medis, pembelajaran mesin dapat mengetahui pola tersembunyi dari dataset [4]. Pembelajaran mesin dapat digunakan untuk melakukan prediksi. Analisa prediksi yang dilakukan dengan target memprediksi data yang masuk dengan membangun model statistik [8]. Model statistik ini dibentuk dari data yang diobservasi dan digunakan untuk menghasilkan aturan-aturan untuk model prediksinya. Salah satu model didalam pembelajaran mesin yang terkenal adalah sistem pakar berbasis aturan (*rule-based expert system*) [9]. Beberapa metode diantaranya adalah pohon keputusan dan *random forest*

[10] yang bentuk model pembelajaran mesinnya berbentuk pohon (*tree model*).

Penelitian ini didesain untuk mengetahui performa pembelajaran mesin berdasarkan model pohon keputusan, yaitu *random forest* yang diterapkan pada data penyakit kardiovaskular. Tujuan dari penelitian ini untuk menunjukkan bahwa dapat membuat model klasifikasi menggunakan *random forest*. Untuk membuktikan *random forest* dapat melakukan prediksi dengan baik, hasil evaluasi akan dibandingkan dengan metode pembelajaran mesin yang lain.

## 2. Literature Review

### 2.1 Penelitian Sebelumnya

Terdapat penelitian yang telah dilakukan oleh peneliti sebelumnya dalam melakukan prediksi penyakit kardiovaskular. Peneliti tersebut melakukan penelitian dengan data yang sama yaitu *cardiovascular disease dataset* yang didapat dari Mendeley dataset [11]. Hasil klasifikasi dari penelitian tersebut akan digunakan sebagai pembanding pada penelitian ini. Jika dilihat akurasi yang dihasilkan dari tiap metode berada diantara 93.15% hingga 96.75%. Diharapkan pada penelitian ini, performa classifier dapat memberikan akurasi lebih baik dari penelitian terdahulu.

Peneliti terdahulu menggunakan beberapa metode, yaitu *decision tree*, *random forest*, *naïve bayes*, *log regression*, dan *support vector model* sebagai teknik dasar dalam klasifikasi. Selain itu ada tiga buah metode lainnya yang dinamakan dengan *ensemble learning*, yaitu *gradient boosting*, *XGBoost* dan metode yang diusulkan oleh peneliti sendiri. Dengan metode yang diusulkan oleh peneliti terdahulu menghasilkan akurasi paling optimal, yaitu akurasi sebesar 96.75% dari metode yang diusulkan [12] dan 91.61% didapatkan dari *logistic regression* [6]. Sedangkan metode usulan lainnya memberikan hasil akurasi yang lebih rendah.

Pembeda penelitian terdahulu dengan peneliti yang akan lakukan adalah pembagian dataset. Peneliti terdahulu melakukan pemisahan dataset dengan porsi 60% sebagai data latih dan 40% sebagai data uji. Pada penelitian ini, peneliti akan mengimplementasikan *cross-validation* sebagai pembagi data dalam membuat model prediksi.

Metode yang diusulkan adalah *random forest* karena metode tersebut termasuk dalam salah satu metode sederhana yang mampu mengatasi banyak permasalahan klasifikasi. Selain itu *random forest* pada penelitian terdahulu dapat memberikan akurasi yang terbilang baik jika dibandingkan dengan metode sederhana lainnya seperti *decision tree*, *naïve bayes*, *log regression* dan *support vector model*. Metode yang sederhana tidak selalu memberikan model yang buruk, tetapi model tersebut dapat dilakukan perbaikan tergantung dari tahap praproses dataset.

**Tabel 1. Penelitian sebelumnya**

Judul	Model	Akurasi (%)
A Reliable Machine Intelligence Model for Accurate Identification of Cardiovascular Diseases Using Ensemble Techniques[12]	Decision Tree	95
	Random Forest	95.12
	Naïve Bayes	94.25
	Log Regression	95.25
	SVM	93.15
	Gradient Boosting	95.15
	XGBoost	96.12
	<b>Proposed Ensemble Method</b>	<b>96.75</b>
Prediction of Cardiovascular Disease Using Machine Learning Algorithms[6]	SVM	88.26
	Random Forest	89.53
	Naïve Bayes	90.95
	Gradient Boosting	90.70
	<b>Log Regression</b>	<b>91.61</b>

**2.2 Evaluasi**

Evaluasi hasil klasifikasi yang digunakan pada penelitian ini berdasarkan pada akurasi, TPR (*True Positive Rate*) dan FPR (*False Positive Rate*) untuk menghasilkan ROC (*Receiver Operating Characteristic*). Akurasi digunakan untuk mengetahui ukuran seberapa bagus *random forest* dalam melakukan klasifikasi. Evaluasi menggunakan akurasi ini adalah pengukuran utama dari semua metode pembelajaran mesin [13]. Perhitungan dari akurasi didapatkan dari banyaknya jumlah prediksi yang benar dibagi dengan banyaknya total data [14], [15]. Formula dari akurasi dapat dilihat pada Persamaan (1).

$$Akurasi = \frac{total\ prediksi\ benar}{banyaknya\ data} * 100\% \quad (1)$$

TPR digunakan untuk mengetahui persentase dari data yang teridentifikasi dengan benar. Sedangkan, FPR digunakan untuk mengukur persentase dari data yang diidentifikasi dengan salah. TPR pada literatur lain

**Tabel 2. Detail dari dataset**

Nama fitur	Deskripsi	Tipe data
patientid	ID pasien	numerik
age	Usia	numerik
gender	Jenis kelamin	biner
chestpain	Tipe sakit dada	nominal
restingBP	Tekanan darah	numerik
serumcholesterol	Serum kolesterol	numerik
fastingbloodsugar	Gula darah saat puasa	biner
restingelectro	Elektrokardiogram	nominal
maxheartrate	Detak jantung maksimum	numerik
exerciseangia	Latihan termasuk angina	biner
oldpeak	Oldpeak	numerik
slope	Slope of the peak	nominal
noofmajorvessels	Nomor dari major vessels	numerik
target	Klasifikasi	biner

disebut dengan *sensitivity*, sedangkan FPR disebut dengan *specificity*. Formula dari TPR dan FPR dapat dilihat pada Persamaan (2) dan (3).

$$TPR = \frac{TP}{TP + FN} * 100\% \quad (2)$$

$$FPR = \frac{FP}{TN + FP} * 100\% \quad (3)$$

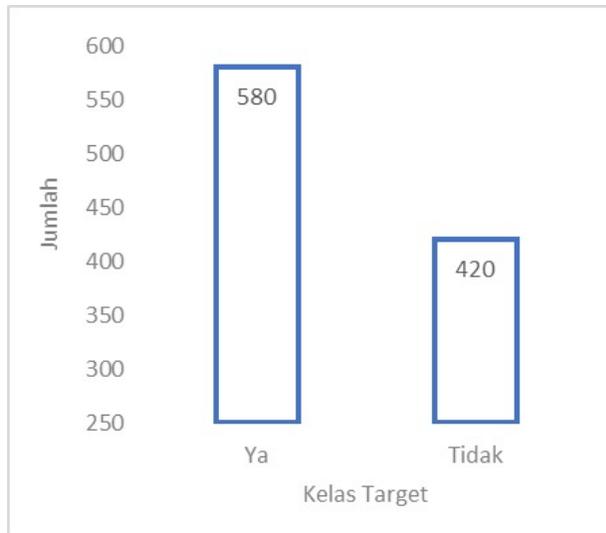
Untuk klasifikasi biner, model dibagi menjadi dua kategori, yaitu *positive* dan *negative*. Dinamakan *true positive* (TP) jika prediksi dan aktualnya benar. Jika prediksinya salah tetapi aktualnya benar, dinamakan *false negative* (FN). Jika hasil prediksinya benar dan aktualnya salah maka dinamakan *false negative* (FP). Jika prediksi dan aktualnya sama-sama salah, maka dinamakan *true negative* (TN). Pengukuran evaluasi ini dapat digunakan untuk menghitung kurva ROC yang bermanfaat untuk memodelkan *error* dari klasifikasi [16].

**3. Material dan Metode**

**3.1 Material**

Pada penelitian ini, Kami menggunakan data sekunder yang didapatkan dari arsip data Mendeley yang bernama *cardiovascular disease dataset* [11]. Dataset didapatkan dari 1000 data pasien, sehingga dataset terdiri dari 1000 instans. Banyaknya fitur pada dataset ini berjumlah 14 fitur. Detail fitur yang terdapat pada dataset ditunjukkan pada Tabel 2.

Fitur target pada dataset dalam bentuk biner dengan nilai 0 = tidak ada penyakit kardiovaskular, dan 1 =



**Gambar 1. Perbandingan Jumlah Kelas Target**

memiliki penyakit kardiovaskular. Perbandingan banyaknya data pada masing-masing kelas target dapat dilihat pada Gambar 1.

Jumlah data untuk kelas Ya sebanyak 580 instans data dan jumlah data untuk kelas Tidak sebanyak 420 data. Terlihat bahwa data dengan kelas Ya berjumlah lebih banyak daripada data dengan kelas Tidak.

**3.2 Metode**

Metode yang digunakan adalah *random forest* dengan memanfaatkan *library* dari Weka [17]. Secara garis besar metodologi yang digunakan pada penelitian ini ditunjukkan pada Gambar 2.

Berdasarkan Gambar 2, tahap pertama mempersiapkan dataset terlebih dahulu yang diunduh dari situs Mendeley dataset. Kemudian tahap kedua dilakukan praproses dataset dengan normalisasi. Tujuan normalisasi disini untuk melakukan penskalaan data agar memiliki rentang yang sama [18]. Karena dataset memiliki rentang yang bervariasi, sehingga normalisasi diterapkan pada dataset supaya fitur menjadi normal. Normalisasi yang dilakukan adalah normalisasi *min-max* dengan formula yang ditunjukkan pada Persamaan (4). Luaran dari normalisasi ini adalah data yang memiliki rentang beragam yaitu antara 0 hingga 1.

$$x' = \frac{x - \min X}{\max X - \min X} \tag{4}$$

keterangan:

$x'$  = data normalisasi

$x$  = data yang akan dinormalisasi

$\max X$  = nilai maksimal data per kolom

$\min X$  = nilai minimal data per kolom

Tahap ketiga adalah membagi data menjadi data latih dan data uji dengan teknik *cross validation* dengan  $k = 10$ . Alasan lain menerapkan *cross validation* adalah



**Gambar 2. Metodologi penelitian**

model yang dihasilkan dapat menghindari *overfitting* dan memberikan model yang lebih general [14]. Pada kasus ini setiap prediksi berlangsung menggunakan 9 frame data sebagai data latih dan sisanya sebagai data uji. Hal ini berlangsung sebanyak 10 kali [19]. Visual representasi dari *cross validation* ditunjukkan pada Gambar 3.

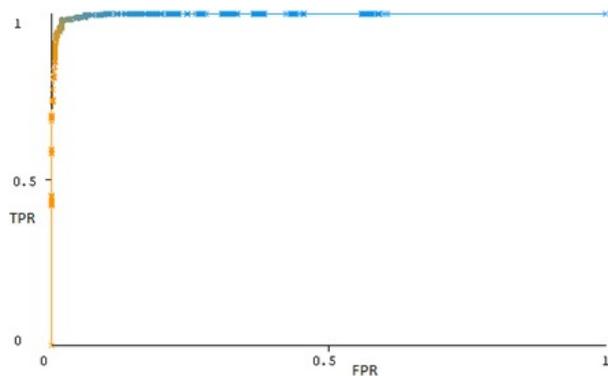
Tahap keempat adalah melakukan klasifikasi dengan *random forest* dan hasil performanya dicatat. *Random forest* adalah pembelajaran mesin yang dapat digunakan untuk permasalahan klasifikasi. Model klasifikasi pada *random forest* berdasar pada model pohon keputusan. *Random forest* terdiri dari sekelompok pohon keputusan [14][15][19][20][21][22] yang mana tiap pohon keputusan tersebut akan melakukan klasifikasi dan prediksinya berdasarkan mayoritas hasil dari masing-masing pohon keputusan [21]. Visualisasi dapat dilihat pada Gambar 6. Kelebihan dari *random forest* adalah dapat menghindari *missing values* dan *outlier* [6] serta dapat digunakan untuk data yang berukuran besar, permasalahan nonlinier dan permasalahan regresi [20]. Sedangkan kekurangan dari *random forest* adalah susah untuk direpresentasikan dalam bentuk visual [14].

Proses pembentukan *random forest* dapat dilihat pada Gambar [16]. Sedangkan urutan langkah-langkahnya yaitu:

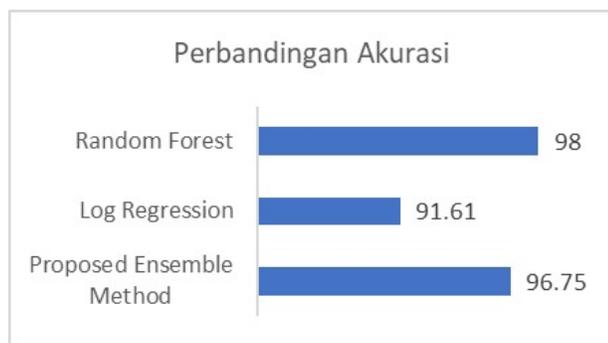
- (a) Memilih sampel  $M$  dari data *training*
- (b) Memilih  $n$  fitur ( $n < N$ ) secara acak dan fitur dengan nilai Gini yang minimum untuk memecah node dari pohon keputusan.
- (c) Ulangi tahap a dan b sebanyak  $k$

k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
k1	k2	k3	k4	k5	k6	k7	k8	k9	k10

Gambar 3. Representasi visual *cross validation* dengan *k-fold* = 10



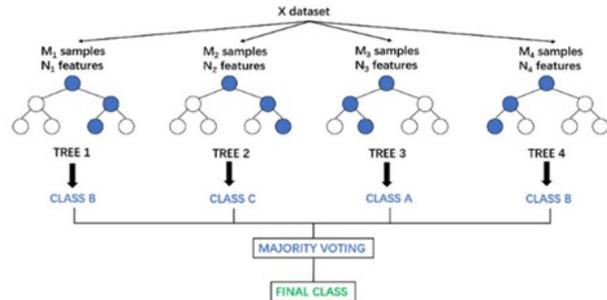
Gambar 4. Kurva ROC



Gambar 5. Perbandingan akurasi dengan penelitian sebelumnya

Tabel 3. Hasil akurasi dari *random forest*

Random Forest	Akurasi (%)	TPR	FPR
Tanpa Normalisasi	97.3	0.979	0.036
Dengan Normalisasi	98	0.983	0.024



Gambar 6. Model *random forest*

Menggabungkan pohon keputusan menjadi *random forest* dan menentukan hasil klasifikasi berdasarkan mayoritas hasil voting.

Klasifikasi dilakukan sebanyak 2 kali percobaan. Percobaan pertama dilakukan untuk dataset original tanpa dilakukan praproses data, percobaan kedua dilakukan pada dataset yang telah dilakukan praproses data menggunakan normalisasi. Hasil performa tersebut dicatat dan digunakan sebagai pembahasan. Didalam pembahasan juga berisi perbandingan hasil untuk mendapatkan hasil akhir terbaiknya. Tahap kelima adalah penarikan kesimpulan berdasarkan pembahasan yang diperoleh dari tahap sebelumnya.

Dari Tabel 3, dapat diketahui bahwa akurasi yang dihasilkan oleh *random forest* pada dataset tanpa normalisasi sebesar 97.3%. Jika dibandingkan dengan dataset yang telah dinormalisasi, akurasinya sebesar 98%. Terdapat kenaikan akurasi sebesar 0.7%. Hal ini menunjukkan bahwa normalisasi dataset mempengaruhi performa dari *classifier*. Nilai TPR dan FPR hasil dari normalisasi adalah 0.983 dan 0.024. Nilai tersebut memiliki kenaikan jika dibandingkan dengan model tanpa normalisasi, yaitu TPR=0.979 dan FPR=0.036. Selanjutnya, dilakukan analisa AUC (*Area Under Curve*) untuk mengetahui seberapa baik model dalam melakukan klasifikasi. AUC dapat menghasilkan kurva ROC. Performa ROC dapat dilihat pada Gambar 5.

Kurva ROC dibentuk berdasarkan sumbu X = FPR dan sumbu Y = TPR. Kurva ROC yang bagus adalah yang memiliki rentang nilai AUC antara 0.5 hingga 1.0. Pada Gambar 5, nilai AUC = 0.9969 yang berarti mendekati kemungkinan terbaiknya. Dapat dikatakan performa dari *classifier* telah mampu dalam melakukan prediksi penyakit kardiovaskular.

Perbandingan hasil optimal dari penelitian yang telah dilakukan dengan 2 penelitian terdahulu

ditunjukkan pada Gambar 6. Terdapat 3 buah perbandingan berdasarkan nilai akurasi dari masing-masing *classifier*, yaitu *logistic regression*, *ensemble method*, dan *random forest*. Ketiga penelitian ini menggunakan dataset yang sama sehingga dapat dilakukan perbandingan yang adil. Perbedaan mendasarnya adalah dari sisi praproses data yang mana pada penelitian sebelumnya dataset dibagi berdasarkan rasio persentase tertentu, sedangkan pada penelitian ini menggunakan *cross-validation* dengan *k-fold*= 10 sebagai pembagi datanya. Akurasi pada penelitian ini lebih baik yaitu 98%. Sedangkan pada penelitian sebelumnya menghasilkan akurasi sebesar 96.75% dengan metode *proposed ensemble method* dan 91.61% menggunakan *logistic regression*. Akurasi yang dihasilkan pada penelitian ini memiliki kenaikan lebih dari 0.2% dari penelitian terdahulunya.

#### 4. Pembahasan dan Hasil

Penelitian pada penyakit kardiovaskular menggunakan *random forest* berbasis model pohon keputusan telah diimplementasikan. Dari dataset, Kami menghilangkan salah fitur yaitu patientID karena tidak memiliki hubungan dengan dataset. Berikutnya, klasifikasi dilakukan sebanyak 2 kali percobaan untuk mengetahui perbandingan hasilnya. Klasifikasi dengan menggunakan *random forest* dengan parameter default yang disediakan oleh Weka dan nilai *seed*= 24. Perbandingan ini dilakukan untuk mengetahui dampak perubahan akurasi dari *classifier* dalam melakukan prediksi penyakit kardiovaskular. Hasil prediksi dapat dilihat pada Tabel 3.

#### 5. Kesimpulan

Pada penelitian ini, dapat disimpulkan bahwa pembelajaran mesin berbasis pohon keputusan yaitu *random forest* dapat diimplementasikan untuk membuat model prediksi penyakit kardiovaskular. Berdasarkan hasil pengujian, dapat diketahui bahwa akurasi yang dihasilkan sebesar 98% dengan TPR=0,983 dan FPR=0.024. Nilai FPR dan TPR ini menghasilkan nilai AUC=0.9969 yang terbilang cukup tinggi karena mendekati nilai 1. Nilai-nilai performa ini didapatkan dari pengujian dengan normalisasi *min-max* sebagai teknik praproses dari dataset. *Cross-validation* juga diterapkan dengan *k-fold*= 10. Dapat diketahui juga bahwa teknik praproses mempengaruhi performa dari *random forest* karena dapat memberikan kenaikan akurasi sebesar 0.2%. Diterapkannya *cross-validation* juga memberikan kenaikan nilai akurasi dari model. Sehingga model ini memberikan akurasi yang lebih baik daripada metode pada penelitian terdahulu. Berdasarkan hal tersebut, dapat disimpulkan bahwa *random forest* berhasil dan dapat digunakan dalam melakukan prediksi penyakit kardiovaskular dan dapat memberikan akurasi

yang lebih tinggi dengan tingkat *error* yang rendah.

#### Daftar Pustaka

- [1] V. E. Septiani, A. L. Wicaksana, E. Y. A. B. Sunaryo, P. S. I. Keperawatan, F. Kedokteran, K. M. dan Keperawatan, U. G. Mada, and D. K. M. Bedah, "Gambaran tingkat risiko penyakit kardiovaskular di wilayah kerja puskesmas mlati sleman yogyakarta," *Jurnal Keperawatan Klinis dan Komunitas (Clinical and Community Nursing Journal)*, vol. 2, pp. 114–124, 7 2018. [Online]. Available: <https://jurnal.ugm.ac.id/jkkk/article/view/44259>
- [2] N. Mostafa, M. A. Azim, M. R. Kabir, and R. Ajwad, "Identifying the risk of cardiovascular diseases from the analysis of physiological attributes," *2020 IEEE Region 10 Symposium, TENSYP 2020*, pp. 1014–1017, 6 2020.
- [3] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim, and A. W. Muzaffar, "An integrated machine learning framework for effective prediction of cardiovascular diseases," *IEEE Access*, vol. 9, pp. 106 575–106 588, 2021.
- [4] S. Islam, N. Jahan, and M. E. Khatun, "Cardiovascular disease forecast using machine learning paradigms," *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, pp. 487–490, 3 2020.
- [5] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81 542–81 554, 2019.
- [6] M. Balakrishnan, A. B. A. Christopher, P. Ramprakash, and A. Logeswari, "Prediction of cardiovascular disease using machine learning," *Journal of Physics: Conference Series*, vol. 1767, p. 012013, 2 2021. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1742-6596/1767/1/012013https://iopscience.iop.org/article/10.1088/1742-6596/1767/1/012013/meta>
- [7] C. M. Cutillo, K. R. Sharma, L. Foschini, S. Kundu, M. Mackintosh, K. D. Mandl, T. Beck, E. Collier, C. Colvis, K. Gersing, V. Gordon, R. Jensen, B. Shabestari, and N. Southall, "Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency," *npj Digital Medicine*, vol. 3, 12 2020. [Online]. Available: [https://www.researchgate.net/publication/340179140\\_Machine\\_intelligence\\_in\\_healthcare\\_perspectives\\_on\\_trustworthiness\\_explainability\\_usability\\_and\\_transparency](https://www.researchgate.net/publication/340179140_Machine_intelligence_in_healthcare_perspectives_on_trustworthiness_explainability_usability_and_transparency)

- [8] A. Callahan and N. H. Shah, "Machine learning in healthcare," *Key Advances in Clinical Informatics: Transforming Health Care through Health Information Technology*, pp. 279–291, 1 2017.
- [9] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthcare Journal*, vol. 6, pp. 94–98, 6 2019. [Online]. Available: [https://www.researchgate.net/publication/334078404\\_The\\_potential\\_for\\_artificial\\_intelligence\\_in\\_healthcare](https://www.researchgate.net/publication/334078404_The_potential_for_artificial_intelligence_in_healthcare)
- [10] N. P. Jayasri and R. Aruna, "Big data analytics in health care by data mining and classification techniques," *ICT Express*, vol. 8, pp. 250–257, 6 2022.
- [11] B. P. Doppala and D. Bhattacharyya, "Cardiovascular disease dataset," vol. 1, 2021.
- [12] B. P. Doppala, D. Bhattacharyya, M. Janarthanan, and N. Baik, "A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques," *Journal of Healthcare Engineering*, vol. 2022, 2022.
- [13] V. K. Gupta, A. Gupta, D. Kumar, and A. Sardana, "Prediction of covid-19 confirmed, death, and cured cases in india using random forest model," *Big Data Mining and Analytics*, vol. 4, pp. 116–123, 6 2021.
- [14] M. Anshori, F. Mar'i, and F. A. Bachtiar, "Comparison of machine learning methods for android malicious software classification based on system call," *Proceedings of 2019 4th International Conference on Sustainable Information Engineering and Technology, SIET 2019*, pp. 343–348, 9 2019.
- [15] S. Kabiraj, M. Raihan, N. Alvi, M. Afrin, L. Akter, S. A. Sohagi, and E. Podder, "Breast cancer risk prediction using xgboost and random forest algorithm," *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*, 7 2020.
- [16] Z. Huang and D. Chen, "A breast cancer diagnosis method based on vim feature selection and hierarchical clustering random forest algorithm," *IEEE Access*, vol. 10, pp. 3284–3293, 2022.
- [17] I. H. Witten, E. Frank, and M. A. Hall, "Data mining: Practical machine learning tools and techniques, third edition," *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*, pp. 1–629, 1 2011.
- [18] V. Gajera, Shubham, R. Gupta, and P. K. Jana, "An effective multi-objective task scheduling algorithm using min-max normalization in cloud computing," *Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016*, pp. 812–816, 4 2017.
- [19] Y. Sugianela, Q. L. Sutino, and D. Herumurti, "Eeg classification for epilepsy based on wavelet packet decomposition and random forest," *Jurnal Ilmu Komputer dan Informasi*, vol. 11, p. 27, 2 2018.
- [20] C. Zhan, Y. Zheng, H. Zhang, and Q. Wen, "Random-forest-bagging broad learning system with applications for covid-19 pandemic," *IEEE Internet of Things Journal*, vol. 8, pp. 15906–15918, 11 2021.
- [21] P. Anuradha and V. K. David, "Super learner model in prediction of heart attack based on cardiac biomarkers," *Indian Journal of Computer Science and Engineering*, vol. 12, pp. 1702–1712, 11 2021.
- [22] Y. Guo, Y. Zhou, X. Hu, and W. Cheng, "Research on recommendation of insurance products based on random forest," *Proceedings - 2019 International Conference on Machine Learning, Big Data and Business Intelligence, MLBDBI 2019*, pp. 308–311, 11 2019.