

# OPTIMASI KLASIFIKASI PENYAKIT GINJAL KRONIS (PGK) MENGGUNAKAN KOMBINASI METODE SELEKSI FITUR DAN ALGORITMA MACHINE LEARNING

Agus Wantoro<sup>1\*</sup>, Lukman Pura<sup>2</sup>, Ade Dwi Putra<sup>3</sup>, Adhie Thyo Priandika<sup>4</sup>, Very Hendra Saputra<sup>5</sup>, Auliya Rahman Isnain<sup>6</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknologi dan Informatika; Universitas Aisyah Pringsewu, Lampung

<sup>2</sup>Poli Penyakit Dalam; Rumah Sakit Umum Daerah Abdul Moeloek, Lampung

<sup>3</sup>Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Lampung

<sup>4,6</sup>Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Lampung

<sup>5</sup>Pendidikan Matematika, Fakultas Sastra dan Ilmu Pendidikan, Universitas Teknokrat Indonesia, Lampung

## Keywords:

Ginjal Kronis;  
Seleksi Fitur;  
Machine Learning;

## \*Correspondent Email:

aguswantoro@aisyahuni  
versity.ac.id

**Abstrak.** Penyakit ginjal kronis (PGK) merupakan salah satu penyakit yang memiliki tingkat prevalensi tinggi dan sering terlambat terdeteksi karena minimnya gejala pada tahap awal. Oleh karena itu, diperlukan metode klasifikasi yang akurat untuk membantu proses deteksi dini. Penelitian ini bertujuan untuk mengoptimalkan kinerja model klasifikasi PGK dengan mengkombinasikan metode seleksi fitur dan model algoritma Machine Learning. Metode seleksi fitur yang digunakan dalam penelitian ini meliputi Information Gain (IG), Gain Rasio (GR), dan Fast Correlation Based Filter (FCBF). Sedangkan algoritma klasifikasi yang digunakan adalah Naive Bayes, Support Vector Machine (SVM), Tree, Random Forest, dan K-Nearest Neighbor (KNN). Dataset yang digunakan adalah Chronic Kidney Disease Dataset yang diperoleh dari UCI Machine Learning Repository. Proses evaluasi dilakukan menggunakan metode k-fold cross validation dengan metrik akurasi. Hasil penelitian menunjukkan bahwa penggunaan seleksi fitur mampu meningkatkan performa model klasifikasi 5.84%. Kombinasi terbaik diperoleh pada metode Gain Ratio dengan algoritma Naive Bayes yang menghasilkan akurasi sebesar 99.9%. Selain itu, seleksi fitur juga berhasil mengurangi jumlah fitur yang digunakan sehingga meningkatkan efisiensi model. Penelitian ini membuktikan bahwa optimasi melalui kombinasi metode seleksi fitur dan algoritma Machine Learning dapat meningkatkan akurasi dan efisiensi dalam klasifikasi penyakit ginjal kronis, serta berpotensi untuk diterapkan dalam sistem pendukung keputusan di bidang medis.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

**Abstract.** Chronic kidney disease (CKD) is a disease that has a high prevalence rate and is often detected late due to the lack of symptoms in the early stages. Therefore, an accurate classification method is needed to assist the early detection process. This research aims to optimize the performance of the PGK classification model by combining feature selection methods and Machine Learning algorithm models. The feature selection methods used in this research include Information Gain (IG), Gain Ratio (GR), and Fast Correlation Based Filter (FCBF). Meanwhile, the classification algorithms used are Naive Bayes, Support Vector Machine (SVM), Tree, Random Forest, and K-Nearest Neighbor (KNN). The dataset used is the Chronic Kidney Disease Dataset obtained from the UCI Machine Learning Repository. The evaluation process was carried out using the k-fold cross validation method

*with accuracy metrics. The research results show that the use of feature selection can increase the performance of the classification model by 5.84%. The best combination was obtained from the Gain Ratio method with the Naive Bayes algorithm which produced an accuracy of 99.9%. In addition, feature selection also succeeded in reducing the number of features used, thereby increasing model efficiency. This research proves that optimization through a combination of feature selection methods and machine learning algorithms can increase accuracy and efficiency in the classification of chronic kidney disease, and has the potential to be applied in decision support systems in the medical field..*

## 1 PENDAHULUAN

Penyakit Ginjal Kronis (PGK) merupakan salah satu masalah kesehatan global yang terus meningkat prevalensinya [1]. Penyakit ini ditandai dengan penurunan fungsi ginjal secara bertahap dan seringkali tidak menunjukkan gejala pada tahap awal, sehingga banyak kasus terdeteksi pada kondisi yang sudah lanjut [2]. Oleh karena itu, diperlukan metode deteksi dini yang akurat untuk membantu tenaga medis dalam proses diagnosis.

Seiring dengan perkembangan teknologi, penerapan Machine Learning (ML) dalam bidang kesehatan telah menjadi solusi yang menjanjikan dalam membantu klasifikasi penyakit, termasuk penyakit ginjal kronis [3]. Berbagai algoritma klasifikasi seperti Support Vector Machine (SVM), Decision Tree, dan Random Forest telah banyak digunakan untuk meningkatkan akurasi diagnosis berbasis data medis [4].

Namun demikian, salah satu tantangan utama dalam pengolahan data medis adalah tingginya jumlah fitur yang tidak semuanya relevan terhadap proses klasifikasi [5]. Fitur yang tidak relevan (*redundan*) dapat menurunkan performa model, meningkatkan kompleksitas komputasi, serta berpotensi menyebabkan *overfitting* [6]. Oleh karena itu, diperlukan teknik seleksi fitur untuk memilih atribut yang paling berpengaruh terhadap hasil klasifikasi.

Metode seleksi fitur sendiri terdiri dari beberapa pendekatan, seperti Information Gain (IG), Gain Rasio (GR), dan Fast Correlation Based Filter (FCBF). Masing-masing metode memiliki kelebihan dan kekurangan dalam hal akurasi, efisiensi, dan kompleksitas. Selain itu, kombinasi antara metode seleksi fitur dengan

algoritma klasifikasi tertentu juga dapat menghasilkan performa yang berbeda-beda

Penelitian terkait klasifikasi penyakit ginjal kronis telah banyak dilakukan dengan memanfaatkan berbagai model ML. Beberapa studi menunjukkan bahwa algoritma seperti Random Forest dan SVM mampu memberikan akurasi yang tinggi dalam klasifikasi PGK [7].

Selain itu, penerapan metode seleksi fitur juga terbukti dapat meningkatkan performa model [8]. Metode seperti IG, GR, dan FCBF yang termasuk dalam kategori *filter methods* sering digunakan karena memiliki waktu komputasi yang relatif cepat [9]. Di sisi lain, metode seperti *Recursive Feature Elimination* (RFE) sebagai bagian dari *wrapper methods* mampu memberikan hasil yang lebih optimal, meskipun dengan biaya komputasi yang lebih tinggi [10]. Penelitian terbaru juga mulai mengkombinasikan metode seleksi fitur dengan algoritma klasifikasi untuk mendapatkan performa terbaik. Namun, hasil yang diperoleh masih bervariasi tergantung pada jenis dataset. Hal ini menunjukkan bahwa belum ada kombinasi metode yang secara universal paling optimal untuk semua kasus.

Dengan demikian, diperlukan kajian lebih lanjut untuk membandingkan dan mengoptimalkan kombinasi metode seleksi fitur dan algoritma klasifikasi khususnya pada kasus penyakit ginjal kronis. Deteksi dini penyakit ginjal kronis sangat penting untuk mencegah perkembangan penyakit ke tahap yang lebih parah. Kesalahan dalam klasifikasi dapat berdampak serius terhadap penanganan pasien [11]. Oleh karena itu, diperlukan model klasifikasi yang tidak hanya akurat tetapi juga efisien.

Urgensi penelitian ini terletak pada (a) Kebutuhan akan model klasifikasi yang lebih optimal dalam mendeteksi penyakit ginjal

kronis. (b) Pentingnya pemilihan fitur yang relevan untuk meningkatkan performa model. (c) Belum adanya kesepakatan mengenai kombinasi metode seleksi fitur dan algoritma terbaik. (d) Potensi penerapan hasil penelitian dalam sistem pendukung keputusan di bidang medis. Dengan adanya penelitian ini, diharapkan dapat memberikan kontribusi dalam pengembangan metode klasifikasi yang lebih efektif dan dapat diimplementasikan secara nyata

Tujuan dari penelitian ini adalah sebagai berikut (a) Menganalisis pengaruh penggunaan metode seleksi fitur terhadap performa klasifikasi PGK. (b) Membandingkan beberapa metode seleksi fitur dalam meningkatkan akurasi model klasifikasi. (c) Mengoptimalkan kombinasi metode seleksi fitur dan algoritma ML untuk memperoleh hasil terbaik. (d) Menghasilkan model klasifikasi yang akurat dan efisien untuk deteksi penyakit ginjal kronis.

## 2 TINJAUAN PUSTAKA

### 2.1 Penyakit Ginjal Kronis (PGK)

PGK merupakan kondisi medis yang ditandai dengan penurunan fungsi ginjal secara bertahap dalam jangka waktu yang lama [12]. Fungsi utama ginjal adalah menyaring limbah dan kelebihan cairan dari darah, yang kemudian dikeluarkan melalui urin. Pada penderita PGK, fungsi ini terganggu sehingga dapat menyebabkan penumpukan zat berbahaya dalam tubuh [1].

PGK umumnya disebabkan oleh penyakit lain seperti diabetes melitus dan hipertensi [2]. Gejala pada tahap awal seringkali tidak terlihat, sehingga diagnosis dini menjadi sangat penting. Oleh karena itu, penggunaan teknologi seperti machine learning dalam membantu proses klasifikasi penyakit menjadi solusi yang potensial

### 2.2 Data Mining dan Machine Learning (ML)

Data mining adalah proses penggalian informasi penting dari sekumpulan data dalam jumlah besar [13]. Salah satu teknik dalam data mining adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas tertentu berdasarkan karakteristik yang dimiliki.

Machine Learning (ML) merupakan cabang dari kecerdasan buatan yang memungkinkan sistem untuk belajar dari data dan meningkatkan kinerjanya tanpa diprogram secara eksplisit [5]. Dalam konteks penelitian ini, ML digunakan untuk membangun model klasifikasi penyakit ginjal kronis berdasarkan data pasien

### 2.3 Klasifikasi

Teknik dalam ML yang digunakan untuk memprediksi label atau kelas dari suatu data. Model klasifikasi dibangun menggunakan data latih (*training data*) dan kemudian diuji menggunakan data uji (*testing data*) [14].

Beberapa algoritma klasifikasi yang umum digunakan antara lain Support Vector Machine (SVM): Algoritma yang bekerja dengan mencari *hyperplane* terbaik untuk memisahkan data antar kelas. Decision Tree: Metode berbasis pohon keputusan yang mudah diinterpretasikan. Random Forest: Kombinasi dari banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi *overfitting*. K-Nearest Neighbor (KNN): Metode yang mengklasifikasikan data berdasarkan kedekatan jarak dengan data lain. Pemilihan algoritma yang tepat sangat mempengaruhi hasil klasifikasi.

### 2.4 Seleksi Fitur (SF)

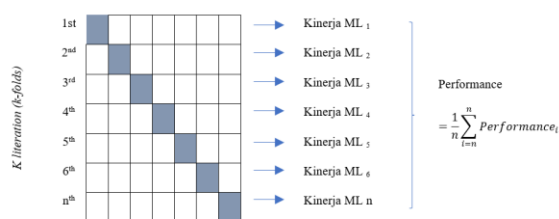
Proses memilih subset fitur yang paling relevan dari dataset untuk meningkatkan performa model ML [15]. Tujuan utama seleksi fitur adalah: (a) Mengurangi dimensi data, (b) Meningkatkan akurasi model, (c) Mengurangi waktu komputasi, (d) Menghindari *overfitting* [9]

### 2.5 Optimasi Model Klasifikasi

Optimasi model bertujuan untuk meningkatkan performa model klasifikasi dengan cara mengatur parameter (*hyperparameter tuning*) dan memilih fitur yang paling relevan. Beberapa teknik optimasi yang umum digunakan (a) Cross Validation, (b) Kombinasi seleksi fitur dan algoritma klasifikasi. Dalam penelitian ini, optimasi dilakukan dengan mengkombinasikan metode seleksi fitur dengan beberapa model algoritma klasifikasi untuk mendapatkan performa terbaik

### 2.6 Cross-validations (k-fold)

Klasifikasi merupakan salah satu bentuk teknik penambangan data yang sedang populer saat ini [16]. Strategi ini menggunakan berbagai metode untuk menilai data yang tersedia guna menghasilkan prediksi diabetes [17]. Model klasifikasi akan divalidasi menggunakan *k-fold cross-validation*. Metode *cross-validation* umumnya digunakan untuk training set [18]. Gambar 1 menampilkan prosedur *cross-validation*



Gambar 1. Prosedur k-fold Validation

### 2.7 Confusion Matrix

Evaluasi model klasifikasi dilakukan untuk mengukur seberapa baik model dalam melakukan prediksi. Studi ini mengkaji bagaimana Confusion Matrix dapat digunakan untuk mengukur akurasi dan tingkat kesalahan [19]. Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam ML. Tabel ini membandingkan hasil prediksi model dengan nilai sebenarnya (aktual) dari data, sehingga dapat memberikan gambaran mendalam tentang kelebihan dan kekurangan mode [20].

Penggunaan metrik evaluasi yang tepat sangat penting untuk memastikan kualitas model yang dihasilkan. Variabel confusion matrix ditampilkan dalam Table 1.

Tabel 1. Confusion Matrix

Class	Prediksi Positif	Prediksi Negatif
Aktual Positif	Jumlah True Positif (TP)	Jumlah False Negatif (FN)
Aktual Negatif	Jumlah False Positif (FP)	Jumlah True Negatif (TN)

Akurasi merupakan metode untuk mengevaluasi kinerja algoritma ML [20]. Precision merupakan tingkat ketepatan prediksi positif, sedangkan recall merupakan Kemampuan model dalam mendeteksi kelas positif. Variabel tersebut dapat diperoleh dari

Confusion Matrix pada Tabel 3 dan dihitung menggunakan persamaan (1)

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

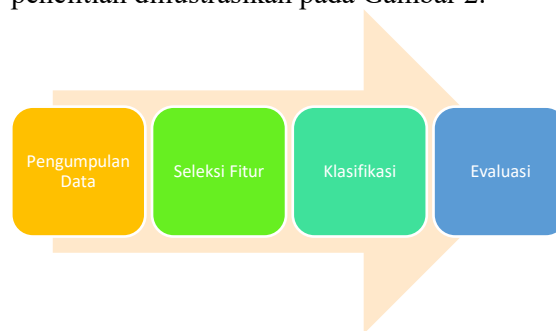
## 3 METODE PENELITIAN

### 3.1 Jenis Penelitian

Jenis penelitian yang digunakan adalah penelitian kuantitatif dengan pendekatan eksperimen. Penelitian ini berfokus pada pengujian dan perbandingan performa beberapa metode seleksi fitur yang dikombinasikan dengan algoritma ML dalam klasifikasi penyakit ginjal kronis

### 3.2 Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis yang dimulai dari pengumpulan data, dilanjutkan dengan penerapan metode seleksi fitur, lalu klasifikasi menggunakan model ML, dan terakhir evaluasi model. Tahapan penelitian diilustrasikan pada Gambar 2.



Gambar 2. Tahapan penelitian

### 3.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari dataset publik, yaitu Chronic Kidney Disease Dataset dari UCI Machine Learning Repository. Dataset yang digunakan dalam penelitian ini terdiri dari 400 data pasien dengan sejumlah atribut medis. Fitur pada dataset ditampilkan dalam Tabel 2.

Tabel 2. Fitur Dataset PKG

No	Kode	Fitur
f1	age	blood pressure
f2	bp	specific gravity
f3	sg	albumin

No	Kode	Fitur
f4	al	sugar
f5	su	red blood cells
f6	rbc	pus cell
f7	pc	pus cell clumps
f8	pcc	bacteria
f9	ba	blood glucose random
f10	bgr	blood urea
f11	bu	serum creatinine
f12	sc	sodium
f13	sod	potassium
f14	pot	hemoglobin
f15	hemo	packed cell volume
f16	pcv	white blood cell count
f17	wc	red blood cell count
f18	rc	hypertension
f19	htn	diabetes mellitus
f20	dm	coronary artery disease
f21	cad	appetite
f22	appet	pedal edema
f23	pe	anemia
f24	ane	class

Dataset ini memiliki dua kelas utama, yaitu CKD (Chronic Kidney Disease) dan Non-CKD (Not-Chronic Kidney Disease)

### 3.4 Implementasi Seleksi Fitur

Melakukan seleksi fitur menggunakan beberapa metode seperti Information Gain (IG), Gain Ratio (GR), dan Fast Correlation Based Filter (FCBF). Hal ini dilakukan untuk mendapatkan temuan berupa metode terbaik dalam kasus ini. Fitur yang akan digunakan untuk pengujian model yang memiliki bobot ( $w > 0.3$ ). Hasil pengujian yang kami lakukan, bobot tersebut merupakan bobot terbaik terhadap kinerja model. Hasil seleksi fitur menggunakan beberapa metode SF ditampilkan pada Tabel 3.

Tabel 3. Fitur terpilih ( $w > 0.3$ )

Seleksi Fitur	Fitur Awal	Fitur Terpilih
Informatin Gain	24	8
Gain Ratio	24	5
BCBF	24	8

Setiap metode seleksi fitur menghasilkan fitur yang berbeda-beda. Metode IG menghasilkan delapan fitur yang memiliki bobot terbia yaitu hemo, pcv, sg, sc, al, rbcc, htn, dan dm. Metode GR menghasilkan lima fitur terbaik yaitu rbc,

htn, hemo, dm, dan al. Metode FCBF menghasilkan delapan fitur yaitu hemo, sg, al, sc, htn, rbc, dm, dan pc.

## 4 HASIL DAN PEMBAHASAN

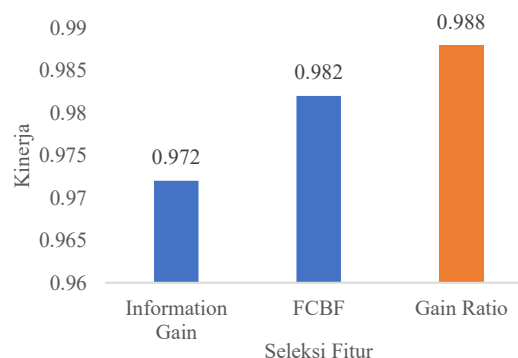
### 4.1 Evaluasi Model Klasifikasi

Evaaluasi model dilakukan untuk mendapatkan gambaran komprehensif tentang kombinasi metode seleksi fitur dan model ML. Model yang digunakan yaitu Support Vector Machine (SVM), Decision Tree, Random Forest, k-Nearest Neighbor (k-NN), dan Naïve Bayes. Model ini memiliki kemampuan yang baik dalam klasifikasi. Hasil perbandingan kinerja model ditampilkan pada Tabel 4.

Tabel 4. Kinerja model

Model	IG	GR	FCBF
k-NN	0.950	0.985	0.983
Naïve Bayes	0.955	<b>0.999</b>	0.965
Tree	0.973	0.982	0.973
Random Forest	0.990	0.990	0.998
SVM	0.993	0.988	0.995

Kami menggunakan ( $fold=5$ ) karena temaun kami nilai tersebut memiliki kinerja terbaik pada kasus ini. Tabel menunjukkan bahwa setiap model ML memiliki kinerja yang berbeda-beda. Kombinasi model Naive Bayes + Gain Ratio (GR) memiliki kinerja terbaik. Hal ini dikarenakan metode GR menghasilkan pohon keputusan yang lebih efisien dan memiliki tingkat akurasi yang lebih tinggi dibandingkan Information Gain biasa, terutama pada dataset dengan karakteristik atribut yang beragam. Nilai rata-rata kinerja metode seleksi fitur ditampilkan pada Gambar 3.



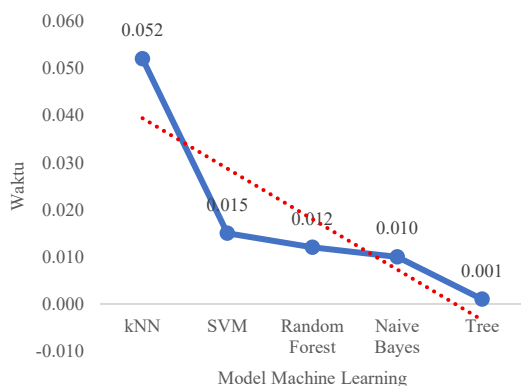
Gambar 3. Nilai rata-rata kinerja seleksi fitur

Selanjutnya kami melakukan analisis perbandingan kinerja terhadap model ML yang ditampilkan pada Gambar 4.



Gambar 4. Perbandingan kinerja model ML

Gambar 4 menampilkan perbandingan kinerja model menggunakan semua fitur dan fitur terpilih. Metode seleksi fitur terbukti mampu meningkatkan beberapa model dengan nilai rata-rata peningkatan 5.84%. Metode Naive Bayes unggul karena kesederhanaan, kecepatan tinggi, dan efisiensinya, terutama pada dataset besar atau teks. Model ini membutuhkan data latih sedikit, menangani data biner atau multikelas dengan baik, serta mampu menangani *missing values*. Sangat efektif untuk klasifikasi teks, spam filter, dan sistem rekomendasi. Selanjutnya kami melakukan perbandingan waktu komputasi yang dibutuhkan dalam membangun model klasifikasi. Hasil perbandingan waktu ditmapilkan pada Gambar 5.



Gambar 5. Perbandingan waktu komputasi

Model Tree memiliki beberapa keunggulan signifikan dalam hal waktu efisiensi komputasi dibandingkan metode ML lainnya, terutama saat fase klasifikasi atau prediksi. Waktu yang dibutuhkan untuk memprediksi data baru sangat cepat, umumnya berbanding lurus dengan kedalaman pohon, bukan jumlah data latih.

## 5 KESIMPULAN

Penggunaan metode seleksi fitur terbukti mampu meningkatkan performa model klasifikasi dibandingkan dengan tanpa seleksi fitur. Hal ini ditunjukkan dengan peningkatan nilai akurasi 5.84% pada nilai rata-rata model algoritma yang digunakan.

Dari beberapa algoritma yang diuji, Naive Bayes menunjukkan performa paling baik dibandingkan algoritma lainnya seperti SVM, Tree, KNN, dan Random Forest.

Kombinasi terbaik dalam penelitian ini adalah penggunaan metode seleksi fitur Gain Ratio (GR) dengan algoritma Naive Bayes yang menghasilkan performa tertinggi dengan akurasi mencapai 99.9%.

Seleksi fitur tidak hanya meningkatkan akurasi, tetapi juga mampu mengurangi jumlah fitur yang digunakan sehingga model menjadi lebih efisien dan lebih cepat dalam proses komputasi

Saran untuk penelitian selanjutnya yaitu menggunakan dataset yang lebih besar dan bervariasi agar model yang dihasilkan lebih general dan robust. Menguji metode seleksi fitur lainnya, seperti metode berbasis metaheuristik (Genetic Algorithm, atau Particle Swarm Optimization) untuk mendapatkan kombinasi fitur yang lebih optimal. Mengembangkan penelitian dengan menggunakan algoritma deep learning untuk melihat potensi peningkatan performa yang lebih tinggi. Melakukan optimasi hyperparameter secara lebih mendalam untuk setiap algoritma agar diperoleh hasil yang lebih maksimal. Mengimplementasikan model yang dihasilkan ke dalam sistem nyata seperti aplikasi berbasis web atau mobile sebagai sistem pendukung keputusan medis

## UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada Fakultas Kesehatan Universitas Aisyah Pringsewu (UAP) dan RSUD Abdul Moeloek yang telah membantu memvalidasi data klinis pada penelitian ini. Kami juga berterima kasih kepada Kelompok Riset Kecerdasan Buatan UAP atas masukan berharga mereka selama fase pengembangan dan evaluasi model. Apresiasi khusus diberikan kepada para dokter yang wawasan klinisnya berkontribusi signifikan terhadap hasil pengujian sistem.

## DAFTAR PUSTAKA

- [1] J.-C. Lv and L.-X. Zhang, "Prevalence and Disease Burden of Chronic Kidney Disease BT - Renal Fibrosis: Mechanisms and Therapies," in *Prevalence and Disease Burden of Chronic Kidney Disease*, B.-C. Liu, H.-Y. Lan, and L.-L. Lv, Eds., Singapore: Springer Singapore, 2019, pp. 3–15. doi: 10.1007/978-981-13-8871-2\_1.
- [2] A. C. Webster, E. V Nagler, R. L. Morton, and P. Masson, "Chronic Kidney Disease," *Lancet*, vol. 389, no. 10075, pp. 1238–1252, Mar. 2017, doi: 10.1016/S0140-6736(16)32064-5.
- [3] V. P. Chavda, K. Patel, S. Patel, and V. Apostolopoulos, "Artificial Intelligence and Machine Learning in Healthcare Sector," in *Bioinformatics Tools for Pharmaceutical Drug Product Development*, 2023, pp. 285–314. doi: <https://doi.org/10.1002/9781119865728.ch13>.
- [4] S. Satri, "Review on Machine Learning Techniques for Medical Data Classification and Disease Diagnosis," *Regen. Eng. Transl. Med.*, vol. 9, no. 2, pp. 141–164, 2023, doi: 10.1007/s40883-022-00273-y.
- [5] A. Wantoro, A. F. Yuliana, D. Yana, A. Andini, and I. Awaliyani, "Optimizing Type 2 Diabetes Classification with Feature Selection and Class Balancing in Machine Learning," *J. Tek. Inform.*, vol. 6, no. 4, pp. 2625–2637, 2025.
- [6] A. Wantoro, Zulkifli, P. Bintoro, T. H. Andika, F. Ardhy, and A. N. Al Aziz, "Performance Evaluation of Classification Multi Algorithms on Small Dataset: A Comparative-Based Analysis," in *2025 Tenth International Conference on Informatics and Computing (ICIC)*, 2025, pp. 1–6. doi: 10.1109/ICIC68054.2025.11309491.
- [7] Z. Rustam, E. Sudarsono, and D. Sarwinda, "Random-Forest ( RF ) and Support Vector Machine ( SVM ) Implementation for Analysis of Gene Expression Data in Chronic Kidney Disease ( CKD ) Random- Forest ( RF ) and Support Vector Machine ( SVM ) Implementation for Analysis of Gene Expression Data in C," in *Materials Science and Engineering*, 2019, pp. 1–6. doi: 10.1088/1757-899X/546/5/052066.
- [8] V. Vijayarveswari *et al.*, "Development of Statistically Modelled Feature Selection Method for Microwave Breast Cancer Detection," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 50, no. 1, pp. 250–263, 2025, doi: 10.37934/araset.50.1.250263.
- [9] L. K. Singh, M. Khanna, and R. Singh, "Efficient feature selection for breast cancer classification using soft computing approach: A novel clinical decision support system," *Multimed. Tools Appl.*, vol. 83, no. 14, pp. 43223–43276, 2024, doi: 10.1007/s11042-023-17044-8.
- [10] H. Jeon and S. Oh, "Hybrid-Recursive Feature Elimination for Efficient Feature Selection," 2020. doi: 10.3390/app10093211.
- [11] M. Mizdrak, M. Kumrić, T. T. Kurir, and J. Božić, "Emerging Biomarkers for Early Detection of Chronic Kidney Disease," 2022. doi: 10.3390/jpm12040548.
- [12] M. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. W. Quinn, and M. A. Moni, "Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening," *IEEE J. Transl. Eng. Heal. Med.*, vol. 9, no. December 2020, pp. 1–11, 2021, doi: 10.1109/JTEHM.2021.3073629.
- [13] S. M. Dol and P. M. Jawandhiya, "Use of Data mining Tools in Educational Data Mining," in *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, 2022, pp. 380–387. doi: 10.1109/CCICT56684.2022.00075.
- [14] E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic Tenets of Classification Algorithms K - Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review," *J. Data Anal. Inf. Process.*, vol. 8, no. 1, pp. 341–357, 2020, doi: 10.4236/jdaip.2020.84020.
- [15] F. Miao, Y. Wu, G. Yan, and X. Si, "Dynamic multi-swarm whale optimization algorithm based on elite tuning for high-dimensional feature selection classification problems," *Appl. Soft Comput.*, vol. 169, p. 112634, 2025, doi: <https://doi.org/10.1016/j.asoc.2024.112634>.
- [16] F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, *Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques*.

- London, 2019. doi: 10.1007/979-981-13-8798-2-12.
- [17] H. Sulistiani, A. Syarif, K. Muludi, and Warsito, "Performance evaluation of feature selections on some ML approaches for diagnosing the narcissistic personality disorder," *Bull. Electr. Eng. Informatics*, vol. 13, no. 2, pp. 1383–1391, 2024, doi: 10.11591/eei.v13i2.6717.
- [18] T. Yan, S.-L. Shen, A. Zhou, and X. Chen, "Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm," *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 4, pp. 1292–1303, 2022, doi: <https://doi.org/10.1016/j.jrmge.2022.03.002>.
- [19] I. Düntsch and G. Gediga, "Confusion Matrices and Rough Set Data Analysis," *J. Phys. Conf. Ser.*, vol. 1229, no. 1, 2019, doi: 10.1088/1742-6596/1229/1/012055.
- [20] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, "Confusion-matrix-based kernel logistic regression for imbalanced data classification," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1806–1819, 2017, doi: 10.1109/TKDE.2017.2682249.