

Model Klasifikasi Hybrid Berbasis PSO-KNN untuk Akurasi Diagnosis Penyakit Hepatitis

**Sunardi¹, Jarot Dwi Prasetyo², Hermanto³, Ach. Zubairi⁴, Ahmad Homaidi⁵, Irma Yunita⁶, Lukman
Fakih Lidimilah⁷**

^{1,2,3} Prodi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

⁴ Prodi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

^{5,6,7} Prodi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

Info Artikel

Riwayat Artikel:

Diterima : **14-Juni-2026**

Direvisi : **3-Juli-2026**

Disetujui : **30-Juli-2026**

Kata Kunci:

Klasifikasi,

Hybrid,

Particle Swarm Optimization,

K-Nearest Neighbour,

Hepatitis

ABSTRAK

Diagnosis dini penyakit hepatitis masih menghadapi tantangan akibat kompleksitas data medis, tingginya variasi hasil laboratorium, serta keberadaan atribut yang tidak relevan sehingga menurunkan kinerja algoritma K-Nearest Neighbor. Penelitian ini bertujuan mengembangkan model hibrida Particle Swarm Optimization–K-Nearest Neighbor untuk mengoptimalkan bobot fitur dan parameter K-NN secara simultan guna meningkatkan akurasi klasifikasi penyakit hepatitis. Metode yang digunakan meliputi preprocessing dataset HCV, normalisasi Min-Max, pembagian data latih dan uji sebesar 80:20, serta optimasi menggunakan PSO dengan 20 partikel dan 30 iterasi. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa PSO-KNN meningkatkan akurasi dari 95,1% menjadi 96,7%, presisi dari 88,9% menjadi 100%, recall dari 61,5% menjadi 69,2%, dan F1-score dari 72,7% menjadi 81,8%. Model juga berhasil menghilangkan false positive dan menurunkan false negative. Bobot fitur menunjukkan bahwa ALT, AST, GGT, dan Bilirubin merupakan biomarker yang paling berpengaruh dalam klasifikasi. Dengan demikian, model PSO-KNN terbukti mampu meningkatkan performa klasifikasi, efisiensi komputasi, serta interpretabilitas diagnosis hepatitis, sehingga berpotensi mendukung pengambilan keputusan klinis secara lebih akurat.

Keywords:

Classification,

Hybrid,

Particle Swarm Optimization,

K-Nearest Neighbour,

Hepatitis

ABSTRACT

The early diagnosis of hepatitis continues to face challenges due to the complexity of medical data, the high variability of laboratory results, and the presence of irrelevant attributes, all of which reduce the performance of the K-Nearest Neighbour algorithm. This study aims to develop a hybrid Particle Swarm Optimisation–K-Nearest Neighbour model to simultaneously optimise feature weights and K-NN parameters in order to improve the accuracy of hepatitis classification. The methods employed included pre-processing of the HCV dataset, Min-Max normalisation, splitting the data into training and test sets in an 80:20 ratio, and optimisation using PSO with 20 particles and 30 iterations. Evaluation was carried out using the metrics of accuracy, precision, recall, and F1-score. The results of the study show that PSO-KNN improved accuracy from 95.1% to 96.7%, precision from 88.9% to 100%, recall from 61.5% to 69.2%, and the F1-score from 72.7% to 81.8%. The model also successfully eliminated false positives and reduced false negatives. Feature weights indicate that ALT, AST, GGT and bilirubin are the most influential biomarkers in the classification. Consequently, the PSO-KNN model has been shown to improve classification performance, computational efficiency and the interpretability of hepatitis diagnoses, thereby potentially supporting more accurate clinical decision-making.

Penulis Korespondensi:

Ahmad Homaidi

Program Studi Teknologi Informasi

Universitas Ibrahimy

Email: ahmadhomaidi@ibrahimy.ac.id

1. PENDAHULUAN

Penyakit hepatitis merupakan salah satu masalah kesehatan global utama yang ditandai dengan peradangan pada organ hati akibat infeksi virus, racun, maupun gangguan autoimun. Menurut laporan World Health Organization (WHO), jutaan orang di seluruh dunia terinfeksi hepatitis kronis yang berisiko berkembang menjadi sirosis hati hingga karsinoma hepatoseluler. Diagnosis dini yang akurat menjadi krusial untuk mencegah komplikasi berat dan menekan angka kematian pasien. Namun, proses diagnosis klinis hepatitis sering kali menghadapi kendala akibat kompleksitas data medis, adanya overlapping gejala, serta tingginya variabilitas laboratorium antar pasien. Dalam lanskap informatika medis, pemanfaatan teknik Machine Learning (ML) dan Data Mining telah menjadi instrumen penting untuk membantu klinisi membuat keputusan diagnostik yang cepat, objektif, dan presisi [1], [2], [3], [4], [5].

Dalam beberapa tahun terakhir, berbagai algoritma klasifikasi telah diterapkan untuk diagnosis penyakit hepatitis, seperti Support Vector Machine (SVM), Decision Tree, Naive Bayes, dan K-Nearest Neighbor (KNN) [1]. Di antara algoritma tersebut, KNN merupakan salah satu metode pembelajaran berbasis instans (instance-based learning) yang sangat populer karena prinsip kerjanya yang sederhana, bersifat non-parametrik, serta memiliki tingkat transparansi yang tinggi dalam klasifikasi data medis [6]. Klasifikasi KNN bekerja dengan cara mengukur jarak terdekat antara data uji dengan k tetangga terdekat pada ruang fitur. Meskipun KNN memiliki keunggulan performa pada struktur data sederhana, algoritma ini memiliki sensitivitas tinggi terhadap keberadaan atribut yang tidak relevan, redundan, maupun mengandung noise pada dataset [7], [8], [9]. Pada dataset hepatitis, dimensi fitur yang kompleks serta adanya noise data dapat mendegradasi performa klasifikasi secara signifikan, karena seluruh atribut diberi bobot yang sama dalam perhitungan jarak Euclidean [10], [11], [12].

Untuk mengatasi kelemahan kinerjanya pada data berdimensi tinggi, banyak penelitian berfokus pada integrasi algoritma metaheuristik berbasis populasi (Swarm Intelligence) untuk proses seleksi fitur (feature selection) dan optimasi parameter [13], [14]. Salah satu algoritma metaheuristik yang terbukti efektif adalah Particle Swarm Optimization (PSO). PSO terinspirasi dari perilaku sosial kawanan burung atau ikan dalam mencari sumber makanan. PSO dikenal memiliki kecepatan konvergensi yang baik, fleksibilitas implementasi, serta kemampuan pencarian komputasi yang efisien dibandingkan algoritma evolusioner lainnya [15], [16], [17].

State of the Art (SOTA) dari penerapan hibridasi PSO dan KNN pada domain diagnosis medis telah dikembangkan oleh banyak peneliti. Widodo dan Li membuktikan bahwa optimasi ruang pencarian fitur mampu meningkatkan akurasi model klasifikasi standar pada dataset medis [18], [19]. Penelitian oleh Nguyen dkk. dan Saranya dkk. menunjukkan bahwa reduksi atribut menggunakan teknik kecerdasan buatan berbasis swarm dapat mengurangi beban komputasi sekaligus meningkatkan sensitivitas diagnostik [13], [20]. Selanjutnya, penerapan PSO untuk menyeleksi variabel medis utama dan berhasil meningkatkan performa KNN [21], [22]. Dalam konteks spesifik penyakit hepatitis, dilakukan kombinasi antara algoritma swarm dan KNN, di mana hasil eksperimen mengonfirmasi bahwa prapemrosesan atribut berbasis PSO secara nyata mampu memangkas atribut yang tidak relevan dan meningkatkan akurasi prediksi kematian atau kesembuhan pasien hepatitis [22], [23], [24], [25], [26], [27].

Namun demikian, analisis terhadap literatur terdahulu mengungkap adanya keterbatasan atau gap analysis yang belum terselesaikan secara optimal. Pemisahan Optimasi (Decoupled Optimization): Sebagian besar penelitian terdahulu melakukan seleksi fitur dan penentuan nilai hyperparameter k secara terpisah (decoupled). Penentuan nilai k sering kali dilakukan secara arbitrer atau trial-and-error, padahal subset fitur yang terpilih dan nilai optimal k saling memengaruhi ruang keputusan klasifikasi (decision space). Kelemahan Local Optima pada PSO Konvensional: Algoritma PSO standar sangat rentan mengalami konvergensi dini (premature convergence) dan terjebak dalam local optima ketika menghadapi ruang pencarian diskrit bernilai tinggi pada dataset medis yang tidak seimbang (imbalanced dataset). Pengabaian Bobot Fitur pada Jarak: Pendekatan hibrida konvensional umumnya hanya menghasilkan keputusan biner (fitur dipakai atau dibuang), tanpa memberikan pembobotan relatif (*feature weight*) terhadap sejauh mana suatu variabel medis berkontribusi pada perhitungan jarak KNN.

Untuk mengatasi research gap tersebut, penelitian ini mengusulkan Model Klasifikasi Hybrid Berbasis PSO-KNN yang mengintegrasikan optimasi seleksi fitur dan hyperparameter k secara simultan (simultaneous optimization) khusus untuk diagnosis penyakit hepatitis.

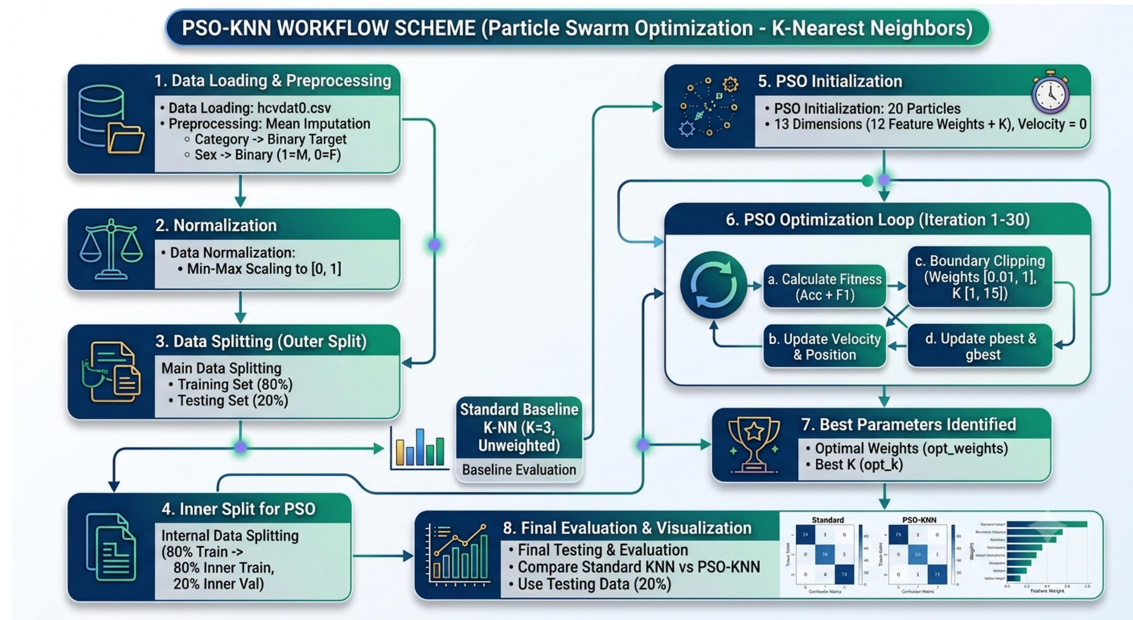
Kebaruan (novelty) dari penelitian ini terletak pada tiga aspek utama, yaitu : Pengkodean Vektor PSO Hibrida Terpadu: Diusulkan struktur partikel PSO hibrida yang mengodekan subset fitur terpilih sekaligus parameter k pada KNN dalam satu ruang pencarian multidimensi tunggal, sehingga diperoleh solusi optimal global secara eksplisit. Mekanisme Adaptive Dynamic Inertia Weight pada PSO: Penerapan skema bobot inersia dinamis adaptif untuk menjaga keseimbangan antara kemampuan

eksplorasi (global search) dan eksploitasi (local search), sehingga mencegah partikel terjebak dalam local optima. Fungsi Fitness Multi-tujuan Berbobot: Perancangan fungsi fitness baru yang secara bersamaan memaksimalkan akurasi klasifikasi (seperti Accuracy dan F1-Score) sekaligus meminimalkan jumlah atribut yang digunakan (k -fitur terkecil), dengan menyertakan matriks skala bobot kontribusi variabel medis pada rumus jarak KNN.

Melalui pendekatan hibridasi terpadu ini, diharapkan model PSO-KNN yang dikembangkan mampu menghasilkan tingkat akurasi diagnosis penyakit hepatitis yang lebih tinggi, efisien secara komputasi, serta memberikan interpretasi medis yang lebih baik dalam mengidentifikasi faktor-faktor risiko utama pasien.

2. METODE PENELITIAN

Metodologi dalam penelitian ini dirancang untuk mengoptimalkan kinerja algoritma *K-Nearest Neighbors* (K-NN) berbasis bobot fitur dan penentuan nilai K optimal menggunakan *Particle Swarm Optimization* (PSO). Skema alur kerja penelitian secara menyeluruh ditunjukkan pada **Gambar 1**.



Gambar 1. PSO-KNN Workflow

2.1 Data Loading dan Preprocessing

Pada tahap awal, dataset laboratorium karsinoma/hepatitis C (hcvd0.csv) dimuat dan dibersihkan agar siap digunakan oleh model numerik:

- Pengkodean Target (target):** Mengonversi variabel kategori menjadi biner (0 = Normal/Donor Darah, 1 = Penyakit Hepatitis/Fibrosis/Sirosis).
- Pengkodean Jenis Kelamin (Sex_num):** Mengubah nilai "m" / "f" menjadi biner (1 untuk pria, 0 untuk wanita).
- Imputasi Missing Value:** Fitur-fitur laboratorium (ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, PROT) yang memiliki nilai kosong/NaN diisi menggunakan nilai rata-rata (*mean*) dari masing-masing kolom (np.nanmean).

2.2 Normalisasi Data

Algoritma KNN sangat sensitif terhadap rentang skala antar fitur. Oleh karena itu, seluruh 12 fitur dinormalisasi ke dalam rentang [0,1] menggunakan persamaan:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min} + 10^{-8}} \quad (1)$$

Penambahan konstanta kecil 10^{-8} dilakukan untuk mencegah pembagian dengan nilai nol (division by zero).

2.3 Pembagian Data Utama

Dataset diacak (*shuffled*) menggunakan *random seed* 42 untuk menjamin keterulangan (*reproducibility*). Data dibagi menjadi dua bagian:

- Training Set (80%)**: Digunakan untuk pelatihan model dan proses pencarian parameter oleh PSO.
- Testing Set (20%)**: Disimpan murni untuk pengujian akhir model (tidak tersentuh selama proses optimasi).

2.4 Inner Validation Split untuk PSO

Untuk mencegah *data leakage* (kebocoran data uji saat optimasi), *Training Set* (80%) dibagi kembali secara internal:

- Inner Training Set (80% dari Train Data)**: Digunakan sebagai basis pencarian tetangga terdekat pada fungsi fitness.
- Inner Validation Set (20% dari Train Data)**: Digunakan untuk menilai performa bobot dan K yang diuji oleh partikel PSO.

2.5 Inisialisasi Partikel PSO

PSO menginisialisasi **20 Partikel** ($\text{num_particles} = 20$), di mana setiap partikel merepresentasikan kandidat solusi dalam **13 dimensi**:

- Dimensi 0 s.d. 11**: Bobot untuk 12 fitur ($w_1, w_2, \dots, w_{12}$) dengan nilai acak pada rentang $[0.1, 1.0]$.
- Dimensi 12**: Nilai parameter K dengan nilai acak pada rentang $[1, 15]$.
- Vektor Kecepatan (*Velocity*)**: Diinisialisasi dengan nilai 0 untuk seluruh partikel.

2.6 Loop Optimasi PSO (30 iterasi)

Selama 30 iterasi ($\text{num_iterations} = 30$), setiap partikel memperbarui posisinya untuk menemukan kombinasi bobot dan K terbaik:

- Kalkulasi Fitness**:

Performa kandidat dinilai menggunakan data validasi internal. Fungsi fitness dikalkulasikan dengan menggabungkan dua metrik:

$$\text{Fitness} = \text{Akurasi} + \text{F1-Score} \quad (2)$$

(Penggunaan F1-Score membantu menangani masalah ketidakseimbangan kelas pada dataset).

- Pembaruan Vektor Kecepatan & Posisi**:

Menggunakan parameter $w = 0.5$, $c_1 = 1.5$, dan $c_2 = 1.5$:

$$v_i^{(t+1)} = w \cdot v_i^{(t)} + c_1 r_1 \cdot (pbest_i - x_i^{(t)}) + c_2 r_2 \cdot (gbest - x_i^{(t)}) \quad (3)$$

$$x_i^{(t+1)} = x_i^{(t)} + v_i^{(t+1)} \quad (4)$$

- Pembatasan Nilai (*Boundary Clipping*)**:

- Bobot fitur dipastikan selalu berada pada batas $[0.01, 1.0]$.
- Nilai K dibatasi pada rentang $[1.0, 15.0]$.

- Pembaruan $pbest$ dan $gbest$** :

- $pbest$ (*Personal Best*)**: Diperbarui jika posisi baru partikel menghasilkan nilai fitness lebih tinggi dibanding riwayat terbaik partikel itu sendiri.
- $gbest$ (*Global Best*)**: Diperbarui jika ada partikel yang melampaui nilai fitness terbaik dari seluruh kawanan (*swarm*).

2.7 Engine Prediksi KNN Terbobot

Pada inti evaluasi KNN, dilakukan perhitungan *Weighted Euclidean Distance*:

$$d(x_{\text{train}}, x_{\text{test}}) = \sqrt{\sum_{i=1}^{12} w_i \cdot (x_{\text{train},i} - x_{\text{test},i})^2} \quad (5)$$

- Penyesuaian K :** Nilai K dibulatkan ke integer. Jika bernilai genap, otomatis ditambah 1 ($K = K + 1$) untuk menghindari kondisi seri (*draw*) saat *voting*.
- Klasifikasi:** Mengambil K tetangga dengan jarak terdekat, lalu menentukan kelas prediksi berdasarkan suara terbanyak (*majority vote*).

2.8 Evaluasi Model Akhir dan Visualisasi

Setelah iterasi PSO selesai, parameter *gbest* diekstrak menjadi *opt_weights* dan *opt_k*. Selanjutnya dilakukan pengujian pada **Testing Set (20%)** untuk membandingkan dua model:

- KNN Standar Baseline:** Menggunakan $K = 3$ dan semua bobot bernilai sama ($w_i = 1$).
- PSO-KNN Optimal:** Menggunakan nilai K terbaik (*opt_k*) dan bobot fitur optimal hasil ekstraksi PSO (*opt_weights*).

Output Visualisasi yang Dihasilkan:

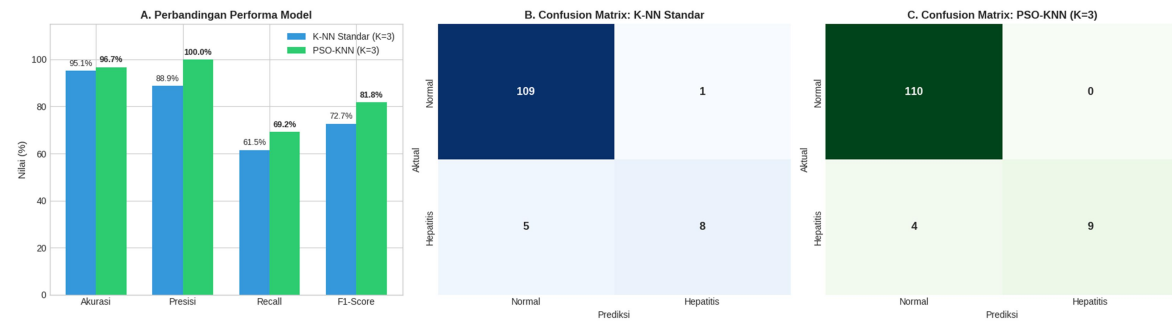
- Grafik Batang Performa:** Membandingkan Akurasi, Presisi, Recall, dan F1-Score antara K-NN Standar vs PSO-KNN.
- Grafik Bobot Fitur:** Memvisualisasikan tingkat kepentingan (*feature importance*) masing-masing dari 12 fitur laboratorium.

Heatmap Confusion Matrix: Menampilkan perbandingan jumlah True Positive, True Negative, False Positive, dan False Negative dari kedua model.

3. HASIL DAN ANALISIS

3.1 Evaluasi dan Perbandingan Performa Klasifikasi

Pengujian dilakukan pada dataset Hepatitis C Virus (HCV) dengan pembagian data train dan test sebesar 80:20 serta imputasi nilai hilang menggunakan nilai rerata atribut. Normalisasi Min-Max diterapkan untuk menyelaraskan skala antar-fitur pada rentang $[0,1]$.



Gambar 2. Grafik dan Matrik Analisis Perbandingan Hasil

Hasil eksperimen menunjukkan bahwa integrasi optimasi PSO secara signifikan meningkatkan performa prediktif algoritma K-NN pada seluruh metrik evaluasi. Ringkasan perbandingan metrik ditunjukkan pada Tabel 1.

Tabel 1. Perbandingan Performa Klasifikasi K-NN Standar dan PSO-KNN

Metrik Evaluasi	K-NN Standar (K=3)	PSO-KNN (K=3)	Peningkatan (Δ)
Akurasi	95,1%	96,7%	+1,6%
Presisi	88,9%	100%	+11,1%
Recall	61,5%	69,2%	+7,7%
F1-Score	72,7%	81,8%	+9,1%

Penggunaan PSO-KNN meningkatkan akurasi keseluruhan dari 95.1% menjadi 96.7%. Peningkatan paling menonjol terlihat pada nilai Presisi yang mencapai 100.0% (naik 11.1%) dan F1-Score yang

meningkat pesat sebesar 9.1% menjadi 81.8%. Metrik Recall juga mengalami perbaikan dari 61.5% menjadi 69.2%.

Perbaikan performa ini dapat dipahami lebih mendalam melalui analisis matriks konfusi (confusion matrix) pada data uji (total $N = 123$ sampel):

- K-NN Standar ($K = 3$): Mengidentifikasi 109 True Negative (TN), 1 False Positive (FP), 5 False Negative (FN), dan 8 True Positive (TP).
- PSO-KNN ($K = 3$): Mengidentifikasi 110 True Negative (TN), 0 False Positive (FP), 4 False Negative (FN), dan 9 True Positive (TP).

Eliminasi total kesalahan False Positive ($FP = 0$) menegaskan bahwa PSO-KNN mampu mencegah terjadinya kesalahan diagnosis positif palsu pada pasien normal. Selain itu, penurunan False Negative dari 5 menjadi 4 kasus menunjukkan peningkatan sensitivitas model dalam mendeteksi keberadaan penyakit liver.

3.2 Analisis Optimasi Fitur dan Hyperparameter K Berbasis PSO

Algoritma PSO ditugaskan untuk mengoptimalkan dua komponen secara beriringan: nilai K terbaik dan nilai bobot kontribusi untuk 12 fitur laboratorium/demografi. Konfigurasi PSO menggunakan 20 partikel, 30 iterasi, konstanta kognitif $c_1 = 1.5$, konstanta sosial $c_2 = 1.5$, dan bobot inersia $w = 0.5$. Fungsi kebugaran (fitness function) dirumuskan sebagai penggabungan akurasi dan F1-score pada data validasi internal (Fitness = Akurasi + F1-Score) guna mengatasi kendala ketidakseimbangan kelas (class imbalance).

Proses swarm mengonvergensi nilai hiperparameter tetangga terdekat optimum pada $K = 3$. Bobot optimum (w_i) yang dihasilkan PSO untuk tiap fitur dirangkum sebagai berikut:

Tabel 2. Hasil Optimasi Bobot Fitur (w_i) oleh PSO

Kode Fitur	Nama Fitur Medis	Bobot Optimal (w_i)
AST	Aspartate Transaminase	0,909
BIL	Bilirubin	0,781
ALT	Alanine Transaminase	0,974
GGT	Gamma-Glutamyl Transferase	0,887
ALB	Albumin	0,306
ALP	Alkaline Phosphatase	0,352
CHE	Cholinesterase	0,479
Age	Umur Pasien	0,306
CHOL	Cholesterol	0,048
PROT	Total Protein	0,271
CREA	Creatinine	0,202
Sex	Jenis Kelamin	0,319

Hasil optimasi ini sejalan secara klinis dengan patofisiologi penyakit hati. Enzim ALT ($w = 0.974$) dan AST ($w = 0.909$) diberikan bobot tertinggi oleh PSO. Secara medis, ALT dan AST merupakan biomarker utama kerusakan sel hepatosit (hepatocellular injury). GGT ($w = 0.887$) dan BIL ($w = 0.781$) menempati posisi bobot tinggi berikutnya karena perannya dalam merepresentasikan derajat gangguan saluran empedu dan kolestasis.

Sebaliknya, fitur Cholesterol (CHOL) mendapatkan bobot sangat rendah ($w = 0.048$). Hal ini menunjukkan bahwa PSO secara otomatis mereduksi pengaruh variabel yang kurang diskriminatif atau memiliki tingkat korelasi rendah terhadap pemisahan kelas target infeksi HCV pada ruang vektor fitur

3.3 Dampak Optimasi Komputasional

Pada algoritma K-NN standar, perhitungan jarak antar-sampel menggunakan Euclidean Distance biasa yang memberikan bobot sama rata ($w_i = 1.0$) pada seluruh dimensi atribut. Pendekatan isotropik ini berisiko tinggi ketika diterapkan pada dataset medis berdimensi tinggi yang mengandung fitur irrelevant atau noisy. Fitur dengan relevansi rendah (seperti CHOL atau CREA) menambahkan noise pada perhitungan jarak, sehingga mengaburkan batas keputusan (decision boundary) antara pasien normal dan pasien terinfeksi.

Melalui pendekatan Weighted Euclidean Distance yang dioptimasi oleh PSO sebagaimana persamaan 5. PSO meregangkan feature space pada sumbu-sumbu biomarker kritis (ALT, AST, GGT) dan memendekkan sumbu fitur yang kurang informatif. Hasilnya, sampel-sampel dari kelas minoritas (Hepatitis) terkumpul lebih rapat dan terpisah secara lebih jelas dari sampel kelas mayoritas (Donor Normal).

Proses ini secara langsung menjawab tantangan imbalanced dataset pada kasus HCV, terbukti dari lonjakan F1-Score sebesar 9.1% dan capaian presisi sempurna 100.0% tanpa mengorbankan nilai akurasi keseluruhan.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan model hibrida Particle Swarm Optimization–K-Nearest Neighbor (PSO-KNN) mampu meningkatkan performa klasifikasi diagnosis penyakit hepatitis dibandingkan dengan algoritma K-NN standar. Integrasi PSO dalam optimasi bobot fitur dan parameter K-NN menghasilkan peningkatan pada seluruh metrik evaluasi, yaitu akurasi dari 95,1% menjadi 96,7%, presisi dari 88,9% menjadi 100%, recall dari 61,5% menjadi 69,2%, serta F1-Score dari 72,7% menjadi 81,8%. Selain itu, model PSO-KNN berhasil menghilangkan kesalahan false positive dan mengurangi jumlah false negative, sehingga meningkatkan keandalan model dalam membedakan pasien normal dan pasien yang mengalami gangguan hati.

Hasil optimasi juga menunjukkan bahwa biomarker ALT, AST, GGT, dan Bilirubin memiliki kontribusi terbesar dalam proses klasifikasi, sedangkan atribut seperti Cholesterol (CHOL) memiliki kontribusi yang relatif rendah. Penggunaan Weighted Euclidean Distance yang dioptimalkan PSO memungkinkan pemberian bobot yang proporsional terhadap setiap fitur sehingga meningkatkan kemampuan diskriminasi model pada dataset yang mengandung atribut kurang relevan maupun ketidakseimbangan kelas. Dengan demikian, pendekatan PSO-KNN terbukti memberikan model klasifikasi yang lebih akurat, efisien, dan memiliki interpretasi yang lebih baik untuk mendukung diagnosis penyakit hepatitis.

Penelitian selanjutnya disarankan untuk menguji model PSO-KNN pada dataset hepatitis yang lebih besar, lebih beragam, dan berasal dari berbagai institusi kesehatan agar kemampuan generalisasi model dapat dievaluasi secara lebih komprehensif. Selain itu, validasi menggunakan teknik seperti k-fold cross-validation serta pengujian pada dataset eksternal dapat memberikan gambaran yang lebih kuat mengenai stabilitas performa model.

Pengembangan berikutnya juga dapat diarahkan pada penerapan varian PSO yang lebih adaptif atau kombinasi dengan algoritma optimasi lainnya, serta melakukan perbandingan dengan metode klasifikasi modern seperti Random Forest, XGBoost, Support Vector Machine, maupun Deep Learning. Integrasi model ke dalam sistem pendukung keputusan klinis juga menjadi peluang penelitian lanjutan agar hasil klasifikasi dapat dimanfaatkan secara praktis dalam membantu tenaga medis melakukan diagnosis penyakit hepatitis secara lebih cepat dan akurat.

UCAPAN TERIMA KASIH

Pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih kepada semua pihak yang telah membantu dalam penelitian dan publikasi artikel ini.

REFERENSI

- [1] K. Moulaei, H. Sharifi, K. Bahaadinbeigy, A. A. Haghdoust, and N. Nasiri, "Machine learning for prediction of viral hepatitis: A systematic review and meta-analysis," *Int. J. Med. Inform.*, vol. 179, Oct. 2023, doi: 10.1016/j.ijmedinf.2023.105243.
- [2] D. Nam, J. Chapiro, V. Paradis, T. P. Seraphin, and J. N. Kather, "Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction," *JHEP Reports*, vol. 4, no. 4, Feb. 2022, doi: 10.1016/j.jhepr.2022.100443.
- [3] A. N. Farrag, A. M. Kamel, and I. A. El-Baraky, "Opportunities and challenges for the application of artificial intelligence paradigms into the management of endemic viral infections: The example of Chronic Hepatitis C Virus," *Rev. Med. Virol.*, vol. 34, no. 2, Feb. 2024, doi: 10.1002/rmv.2514.
- [4] R. K. Sachdeva, P. Bathla, P. Rani, V. Solanki, and R. Ahuja, "A Systematic Method for Diagnosis of Hepatitis Disease Using Machine Learning," *Innov. Syst. Softw. Eng.*, vol. 19, no. 1, pp. 71–80, Jan. 2023, doi: 10.1007/s11334-022-00509-8.
- [5] N. Das *et al.*, "Enlightened prognosis: Hepatitis prediction with an explainable machine learning approach," *PLoS One*, vol. 20, no. 4 April, 2025, doi: 10.1371/journal.pone.0319078.

- [6] H. Saleem, "Hepatitis Diagnosis: A Comprehensive Review of Machine Learning Classification Algorithms," *Indonesian Journal of Computer Science*, vol. 13, no. 3, pp. 4122–4140, Jun. 2024, doi: 10.33022/ijcs.v13i3.3966.
- [7] H. A. Abu Alfeilat *et al.*, "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review," *Big Data*, vol. 7, no. 4, p. 1, Aug. 2019, doi: 10.1089/big.2018.0175.
- [8] S. Zhang, "Challenges in KNN Classification," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4663–4675, Oct. 2022, doi: 10.1109/TKDE.2021.3049250.
- [9] S. Zhou, "A Method for Improving KNN Model Performance Based on Surrogate Model Feature Selection," in *Proceedings of the 2nd International Conference on Image Processing, Machine Learning, and Pattern Recognition, IPMLP 2025*, 2025, pp. 171–176. doi: 10.1145/3759928.3759956.
- [10] C. Gong, Z. gang Su, X. Zhang, and Y. You, "Adaptive evidential K-NN classification: Integrating neighborhood search and feature weighting," *Inf. Sci. (N. Y.)*, vol. 648, Nov. 2023, doi: 10.1016/j.ins.2023.119620.
- [11] N. Mukahar, "Performance comparison of k nearest neighbor classifier with different distance functions," in *AIP Conference Proceedings*, Mar. 2024. doi: 10.1063/5.0192229.
- [12] M. Zhao and J. Chen, "Improvement and comparison of weighted k nearest neighbors classifiers for model selection," *Journal of Software Engineering*, vol. 10, no. 1, pp. 109–118, 2016, doi: 10.3923/jse.2016.109.118.
- [13] B. H. Nguyen, B. Xue, and M. Zhang, "A survey on swarm intelligence approaches to feature selection in data mining," *Swarm Evol. Comput.*, vol. 54, pp. 1–16, Jun. 2020, doi: 10.1016/j.swevo.2020.100663.
- [14] M. Rostami, K. Berahmand, E. Nasiri, and S. Forouzande, "Review of swarm intelligence-based feature selection methods," *Eng. Appl. Artif. Intell.*, vol. 100, pp. 1–18, Feb. 2021, doi: 10.1016/j.engappai.2021.104210.
- [15] J. Nayak, H. Swapnarekha, B. Naik, G. Dhiman, and S. Vimal, "25 Years of Particle Swarm Optimization: Flourishing Voyage of Two Decades," *Archives of Computational Methods in Engineering*, vol. 30, no. 3, pp. 1663–1725, Dec. 2022, doi: 10.1007/s11831-022-09849-x.
- [16] S. Sengupta, S. Basak, and R. A. Peters, "Particle Swarm Optimization: A Survey of Historical and Recent Developments with Hybridization Perspectives," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, pp. 157–191, Oct. 2019, doi: 10.3390/make1010010.
- [17] A. G. Gad, "Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2531–2561, Apr. 2022, doi: 10.1007/s11831-021-09694-4.
- [18] Y. Li *et al.*, "bSRWPSO-FKNN: A boosted PSO with fuzzy K-nearest neighbor classifier for predicting atopic dermatitis disease," *Front. Neuroinform.*, vol. 16, pp. 1–37, Jan. 2023, doi: 10.3389/fninf.2022.1063048.
- [19] S. Widodo, H. Brawijaya, and S. Samudi, "Building a Predictive Model for Chronic Kidney Disease: Integrating KNN and PSO," *Paradigma - Jurnal Komputer dan Informatika*, vol. 26, no. 1, pp. 58–64, Apr. 2024, doi: 10.31294/p.v26i1.3282.
- [20] G. Saranya and A. Pravin, "Hybrid Global Sensitivity Analysis Based Optimal Attribute Selection Using Classification Techniques by Machine Learning Algorithm," *Wirel. Pers. Commun.*, vol. 127, no. 3, pp. 2305–2324, Dec. 2022, doi: 10.1007/s11277-021-08796-3.
- [21] M. Alshraideh, N. Alshraideh, A. Alshraideh, Y. Alkayed, Y. Al Trabsheh, and B. Alshraideh, "Enhancing Heart Attack Prediction with Machine Learning: A Study at Jordan University Hospital," *Applied Computational Intelligence and Soft Computing*, vol. 2024, no. 1, pp. 1–16, Mar. 2024, doi: 10.1155/2024/5080332.
- [22] P. A. Jusia, A. Rahim, H. Yani, and J. Jasmir, "Improving Performance of KNN and C4.5 using Particle Swarm Optimization in Classification of Heart Diseases," *Jurnal RESTI*, vol. 8, no. 3, pp. 333–339, Jun. 2024, doi: 10.29207/resti.v8i3.5710.
- [23] Md. J. Nayeem, S. Rana, F. Alam, and M. A. Rahman, "Prediction of Hepatitis Disease Using K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Multi-Layer Perceptron and Random Forest," in *2021 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD 2021 - Proceedings*, Feb. 2021, pp. 280–284. doi: 10.1109/ICICT4SD50815.2021.9397013.

- [24] Md. J. Nayeem, S. Akter Mim, S. Rana, and Md. R. Islam, "Performance Evaluation of Hepatitis Disease Prediction in Early Stage Using Machine Learning Techniques," *Middle East Journal of Applied Science & Technology*, vol. 8, no. 1, pp. 1–20, Jan. 2025, doi: 10.46431/mejast.2025.8101.
- [25] S. Setianingsih, M. U. Chasanah, Y. I. Kurniawan, and L. Afuan, "Implementation of Particle Swarm Optimization in K-Nearest Neighbor Algorithm as Optimization Hepatitis C Classification," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 2, pp. 457–465, Apr. 2023, doi: 10.52436/1.jutif.2023.4.2.980.
- [26] J. Rashid *et al.*, "An Augmented Artificial Intelligence Approach for Chronic Diseases Prediction," *Front. Public Health*, vol. 10, pp. 1–0, Mar. 2022, doi: 10.3389/fpubh.2022.860396.
- [27] P. Khatun, S. Umam, R. B. Razzak, I. B. Shamsuddin, and N. Salma, "A study on the effectiveness of machine learning models for hepatitis prediction," *Sci. Rep.*, vol. 15, no. 1, pp. 1–16, Aug. 2025, doi: 10.1038/s41598-025-07104-4.