

Implementasi Extreme Learning Machine dengan Seleksi Fitur Particle Swarm Optimization untuk Klasifikasi Sindrom Ovarium Polikistik

AUDYRA DEWI PUSPA MUKTI, NURISSAIDAH ULLINNUHA*,
AHMAD HANIF ASYHAR

MATEMATIKA UIN SUNAN AMPEL SURABAYA.
NURIS.ULINNUHA@UINSA.AC.ID.

Abstrak

Sindrom Ovarium Polikistik (SOPK) adalah gangguan hormonal yang sering terjadi pada wanita usia reproduktif dan menjadi salah satu penyebab utama masalah kesuburan. Sekitar 3%–15% wanita di seluruh dunia mengalami kondisi ini, yang juga dapat memicu berbagai masalah kesehatan lainnya. Penelitian ini bertujuan untuk mengembangkan metode diagnosis SOPK yang lebih efisien dan akurat dengan memanfaatkan algoritma Extreme Learning Machine (ELM) yang dikombinasikan dengan seleksi fitur menggunakan Particle Swarm Optimization (PSO). ELM dipilih karena kemampuannya dalam melakukan klasifikasi secara cepat, sedangkan PSO digunakan untuk memilih fitur-fitur yang paling relevan. Hasil seleksi fitur menghasilkan 18 fitur terpilih dari total 40 fitur. Pencarian parameter terbaik dilakukan dengan pendekatan random search dan grid search. Hasil menunjukkan bahwa random search memberikan performa terbaik, dengan akurasi 95.35%, sensitivitas 96.67%, dan spesifisitas 92.65%. Tanpa seleksi fitur, ELM hanya menghasilkan akurasi 84.20%, sensitivitas 90.10%, dan spesifisitas 70.62%. Temuan ini menunjukkan bahwa seleksi fitur menggunakan PSO mampu meningkatkan performa klasifikasi ELM secara signifikan.

Kata kunci: ELM, Klasifikasi, PSO, SOPK.

Abstract

Polycystic Ovary Syndrome (PCOS) is a hormonal disorder commonly found in women of reproductive age and is one of the leading causes of infertility. It affects approximately 3–15% of women worldwide and may also lead to various other health complications. This study aims to develop a more efficient and accurate diagnostic method for PCOS by utilizing the Extreme Learning Machine (ELM) algorithm combined with feature selection using Particle Swarm Optimization (PSO). ELM was chosen for its speed in handling classification tasks, while PSO was used to identify the most relevant features. The feature selection process resulted in 18 selected features out of a total of 40. Optimal parameter tuning was performed using both random search and grid search approaches. The results showed that random search achieved the best performance, with an accuracy of 95.35%, sensitivity of 96.67%, and specificity of 92.65%. In comparison, ELM without feature selection only reached an accuracy of 84.20%, sensitivity of 90.10%, and specificity of 70.62%. These findings indicate that applying PSO for feature selection significantly enhances the classification performance of ELM.

Keywords: Classification, ELM, PCOS, PSO.

1. PENDAHULUAN

Sindrom Ovarium Polistik (SOPK) merupakan gangguan hormonal yang umum terjadi pada wanita usia reproduktif dan sering menyebabkan masalah kesuburan [1]. World Health Organization (WHO) menyebutkan bahwa infertilitas telah menjadi masalah kesehatan global dengan sekitar 64% kasus disebabkan oleh faktor perempuan [2]. Prevalensi SOPK secara global diperkirakan antara 3% hingga 15%, dengan variasi antar negara, seperti 8.7% di Australia dan 15.5% di Thailand [3]. Hingga saat ini, belum tersedia data nasional resmi mengenai prevalensi SOPK di Indonesia, namun studi di Rumah Sakit Cipto Mangunkusumo mencatat 105 kasus dengan mayoritas pasien mengalami gangguan menstruasi dan hirsutisme [4]. Wanita dengan SOPK sering mengalami ketidakseimbangan hormon androgen dan estrogen yang berkaitan erat dengan resistensi insulin dan obesitas [5]. Kondisi ini tidak hanya mengganggu fungsi reproduksi, tetapi juga berdampak pada metabolisme tubuh secara menyeluruh. Meski demikian, gaya hidup sehat dapat membantu mengurangi risiko dan gejala SOPK. Diagnosis SOPK seringkali memerlukan waktu lama karena membutuhkan evaluasi klinis dan pemeriksaan lanjutan yang kompleks. Oleh karena itu, pengembangan sistem pendukung diagnosis berbasis teknologi menjadi sangat penting untuk efisiensi dan akurasi diagnosis. Beberapa penelitian telah memanfaatkan pendekatan machine learning untuk membantu proses klasifikasi dan prediksi SOPK. Kombinasi metode Support Vector Machine (SVM) dan Particle Swarm Optimization (PSO) berhasil mencapai akurasi sebesar 90,18% dalam memprediksi SOPK [6]. Selain itu, pendekatan lain menggunakan teknik seleksi fitur seperti Recursive Feature Elimination (RFE) dengan algoritma Random Forest mencapai akurasi hingga 90% [7]. Extreme Learning Machine (ELM) merupakan salah satu metode klasifikasi yang dikenal mampu melakukan proses pembelajaran dengan cepat dan akurat karena hanya memanfaatkan satu lapisan tersembunyi [8]. Penelitian sebelumnya menunjukkan bahwa ELM dapat menghasilkan akurasi yang lebih tinggi dibandingkan metode lain. Sebagai contoh, studi yang membandingkan ELM dengan Multilayer Perceptron (MLP) dalam memprediksi mahasiswa yang berpotensi drop out menunjukkan bahwa ELM dengan rasio data 80:20 mampu mencapai akurasi sebesar 95% dengan 20 unit pada hidden layer, serta nilai Mean Squared Error (MSE) sebesar 0.369. Sebaliknya, MLP dengan rasio yang sama hanya memperoleh akurasi sebesar 91% [9]. Penelitian lain yang membandingkan metode Support Vector Machine (SVM) dan ELM pada model prediksi keandalan perangkat lunak bertujuan mengevaluasi efektivitas kedua metode dalam mengklasifikasikan dan memprediksi tingkat keandalan perangkat lunak, yang merupakan aspek penting dalam proses pengembangannya. Hasil penelitian menunjukkan bahwa metode SVM menghasilkan akurasi sebesar 78.68%, sementara ELM memberikan performa yang lebih baik dengan akurasi mencapai 84.61% [10]. Dalam pengembangan sistem klasifikasi, seleksi fitur memainkan peran penting untuk mengurangi kompleksitas data dan meningkatkan performa model. PSO dikenal sebagai algoritma optimasi berbasis populasi yang efektif dalam pemilihan fitur untuk klasifikasi [11]. Beberapa penelitian telah menerapkan kombinasi ELM dan PSO dalam klasifikasi penyakit, seperti penelitian yang dilakukan oleh Rasif Nidaan Khofia Ahmadah, dkk menggunakan metode ELM-PSO untuk mendiagnosis penyakit jantung koroner yang bertujuan untuk mendukung tenaga medis dan mengurangi keterlambatan dalam penanganan pasien diperoleh nilai akurasi sebesar 86% [12]. Berdasarkan potensi tersebut, penelitian ini bertujuan untuk mengklasifikasikan apakah seseorang menderita SOPK menggunakan metode ELM yang dioptimasi dengan algoritma PSO guna meningkatkan akurasi dan efisiensi model klasifikasi. Penelitian ini diharapkan dapat mendukung pengembangan sistem klasifikasi SOPK yang lebih cepat dan akurat dengan memanfaatkan kombinasi ELM dan PSO. Sistem ini bertujuan membantu tenaga medis dalam skrining awal dan mempercepat proses diagnosis. Selain itu, hasil penelitian ini diharapkan menjadi referensi bagi penerapan machine learning dalam klasifikasi penyakit lain serta memperkaya literatur di bidang kesehatan reproduksi.

2. METODE PENELITIAN

Dalam penelitian ini digunakan Extreme Learning Machine (ELM) sebagai model klasifikasi, namun performanya sangat ditentukan oleh seberapa baik fitur yang masuk ke dalam sistem. Untuk itu, pendekatan ini dipadukan dengan Particle Swarm Optimization (PSO) agar pemilihan fitur dapat dilakukan secara otomatis dan lebih terarah. Melalui integrasi keduanya, model diharapkan memperoleh himpunan atribut yang lebih informatif sehingga hasil prediksi menjadi lebih andal. Diagram pada Gambar 1 memberikan gambaran alur kerja yang dimulai dari pengolahan data hingga proses evaluasi model.

2.1. Input Data. Penelitian ini merupakan studi kuantitatif yang menggunakan data rekam medis pasien dari dataset daring [13]. Menggunakan data dari 538 perempuan, terdiri atas 176 pasien yang terdiagnosis Sindrom Ovarium Polikistik (SOPK) dan 362 individu tanpa indikasi kondisi tersebut. Data dikonversi ke format numerik agar dapat diproses dalam analisis klasifikasi, yang bertujuan mendukung deteksi dini SOPK sebagai langkah pencegahan terhadap infertilitas. Struktur dataset terdiri atas variabel prediktor (independen) dan variabel respon (dependen), di mana variabel respon menunjukkan status diagnosis SOPK dengan kode 1 untuk positif dan 0 untuk negatif. Secara keseluruhan, dataset memuat 40 variabel, terdiri dari 32 variabel numerik seperti usia, indeks massa tubuh (IMT), kadar hormon, dan tekanan darah, serta 8 variabel kategorik yang menggambarkan riwayat klinis dan kebiasaan hidup, termasuk frekuensi olahraga, konsumsi makanan cepat saji, dan gejala yang dialami. Kelengkapan dan keberagaman jenis variabel ini memberikan landasan analisis yang komprehensif dalam memahami karakteristik umum maupun spesifik pada individu dengan SOPK.

2.2. Normalisasi Data. Normalisasi data adalah proses yang bertujuan untuk menyeimbangkan nilai-nilai yang memiliki jarak yang berbeda [14]. Proses normalisasi yang diterapkan dalam penelitian ini adalah Min-Max Normalization.

$$d' = \frac{d - d_{\min}}{d_{\max} - d_{\min}} \quad (1)$$

Keterangan

- d' : Hasil normalisasi
- d : Nilai data
- $d_{\min}$: Nilai minimum data
- $d_{\max}$: Nilai maksimum data

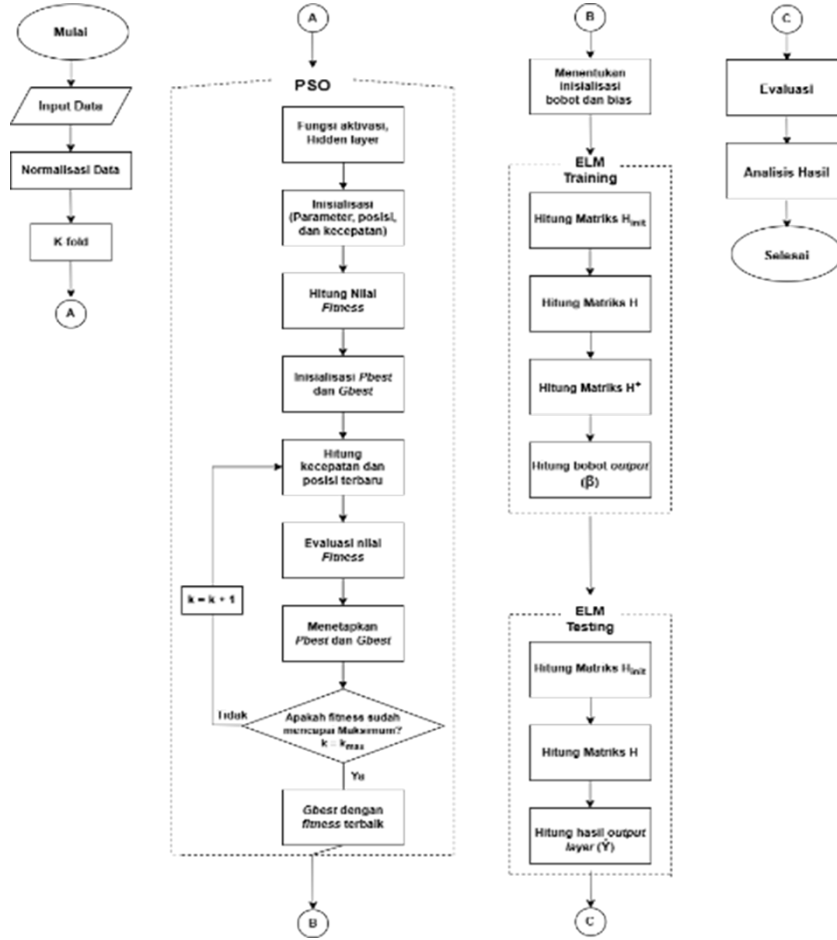
2.3. K-fold Cross-Validation. K-Fold Cross Validation merupakan teknik evaluasi statistik yang membagi data menjadi k bagian untuk melatih dan menguji model secara bergantian, sehingga menghasilkan penilaian performa yang lebih stabil dan akurat [15]. Metode ini efektif dalam mengurangi risiko overfitting serta memberikan estimasi kinerja yang lebih andal dibandingkan pembagian data konvensional. Dalam penelitian ini, digunakan skema 5-fold, di mana data dibagi secara merata untuk proses cross-validation guna meningkatkan akurasi evaluasi model.

2.4. Seleksi Fitur PSO. Particle Swarm Optimization (PSO) adalah metode optimasi yang bekerja dengan menggerakkan partikel-partikel dalam ruang pencarian untuk menemukan solusi terbaik. Setiap partikel mewakili kemungkinan solusi dan dievaluasi berdasarkan fungsi objektif. Partikel-partikel ini saling berbagi informasi untuk bergerak menuju posisi optimal [16]. Berikut merupakan proses perhitungan PSO [17]:

2.4.1. Menginisialisasi nilai awal.

- Menginisialisasi kecepatan awal dengan nilai 0.
- Menginisialisasi posisi awal yang bernilai sesuai dengan inisialisasi input bobot dan bias.
- Menghitung fitness.

$$Fitness = Accuracy$$



GAMBAR 1. Diagram Alir Penelitian

Menginisialisasi Pbest dan Gbest awal, dimana Pbest bernilai sama dengan posisi awal partikel. Dan untuk nilai Gbest merupakan nilai dari Pbest yang memiliki fitness terbesar.

2.4.2. Melakukan Update pada nilai partikel.

Update kecepatan partikel

$$v_{i,j}^{t+1} = \omega \cdot v_{i,j}^t + c_1 \cdot r_1 \cdot (pbest_{i,j}^t - x_{i,j}^t) + c_2 \cdot r_2 \cdot (gbest_{i,j}^t - x_{i,j}^t) \quad (3)$$

Keterangan

- $v_{i,j}^{t+1}$: Kecepatan partikel ke-i pada dimensi ke-j pada iterasi ke-(t+1).
- ω : Bobot inersia.
- c_1, c_2 : Koefisien percepatan kognitif dan sosial.
- r_1, r_2 : Bilangan acak antara 0 dan 1.
- $pbest_{i,j}^t$: Posisi terbaik partikel ke-i pada dimensi ke-j hingga iterasi ke-t.
- $gbest_{i,j}^t$: Posisi terbaik global pada dimensi ke-j hingga iterasi ke-t.
- $x_{i,j}^t$: Posisi partikel ke-i pada dimensi ke-j pada iterasi ke-t.

Update posisi partikel

$$x_{i,j}^{t+1} = Sig(v_{i,j}^{t+1}) = \frac{1}{1 + \exp(-v_{i,j}^{t+1})} \quad (4)$$

Dimana,

$$x_{i,j}^{t+1} = \begin{cases} 1, & \text{if } \text{sigmoid} > \text{rand}(0, 1) \\ 0, & \text{Otherwise} \end{cases} \quad (5)$$

Keterangan

- $x_{i,j}^{t+1}$: Posisi baru partikel ke-i pada dimensi ke-j.
 $\text{Sig}(x_{i,j}^{t+1})$: Nilai probabilitas aktivasi update posisi ke-i pada dimensi ke-j.
 $v_{i,j}^{t+1}$: Nilai update kecepatan.

Update nilai $Pbest$ dan $Gbest$

$Pbest$ diperbarui dengan membandingkan nilai *fitness* sebelum dan sesudah posisi partikel diubah, nilai tertinggi di antara keduanya dipilih. Sedangkan $Gbest$ adalah $Pbest$ terbaik dari seluruh partikel berdasarkan nilai *fitness* tertinggi.

2.5. Klasifikasi ELM. *Extreme Learning Machine* (ELM) merupakan jenis jaringan saraf tiruan *feedforward* dengan satu lapisan tersembunyi (SLFN) yang dirancang untuk mengenali pola dalam data secara efisien [18]. Dalam ELM, bobot input dan bias pada lapisan tersembunyi diinisialisasi secara acak dan tidak diperbarui selama pelatihan, sementara bobot keluaran dihitung secara analitis melalui operasi invers sederhana [19]. Hal ini membuat proses pelatihan lebih cepat karena tidak memerlukan backpropagation. Arsitektur ELM terdiri dari tiga lapisan utama: lapisan input yang menerima data, lapisan tersembunyi yang menerapkan fungsi aktivasi pada kombinasi bobot dan bias, serta lapisan keluaran yang menghasilkan prediksi akhir berdasarkan bobot keluaran yang telah dihitung [14].

Berikut merupakan tahapan-tahapan yang dilakukan dalam proses Training dan Testing data [14]:

2.5.1. Proses training.

Menginisialisasi untuk input weight dan bias yang diambil secara random.

Menghitung matriks output hidden layer dengan persamaan berikut

$$H_{\text{init}} = X_{\text{train}} \cdot W^T + b \quad (6)$$

Keterangan

- H_{init} : Hasil output dari hidden layer.
 X_{train} : Matriks data input dari data training.
 W^T : Transpose dari matriks bobot input.
 b : Vektor bias.

Perhitungan fungsi aktivasi dari output hidden layer.

$$H = \frac{1}{1 + \exp(-H_{\text{init}})} \quad (7)$$

Keterangan

- H : Hasil fungsi aktivasi.
 H_{init} : Nilai output dari hidden layer.

Moore-Penrose Inverse

$$H^+ = (H^T \cdot H)^{-1} \cdot H^T \quad (8)$$

Keterangan

- H^+ : Hasil dari Moore-Penrose inverse.
 $(H^T \cdot H)^{-1}$: Invers dari perkalian matriks H^T dan H .
 H^T : Transpose dari matriks H (hasil fungsi aktivasi).

Menghitung nilai matriks output weight menggunakan persamaan berikut

$$\beta = H^+ \cdot Y \quad (9)$$

Keterangan

- β : Matriks output weight.
 H^+ : Moore–Penrose inverse dari matriks H .
 Y : Vektor target yang telah dinormalisasi.

2.5.2. Proses testing.

Nilai weight dan bias menggunakan nilai yang sama dengan proses training. Perhitungan matriks output hidden layer menggunakan persamaan berikut.

$$H_{\text{init}} = X_{\text{testing}} \cdot W^T + b \quad (10)$$

Keterangan

- H_{init} : Hasil output dari hidden layer.
 X_{testing} : Matriks data input dari data testing.
 W^T : Transpose dari matriks bobot input.
 b : Vektor bias.

Perhitungan nilai fungsi aktivasi dari H_{init} dilakukan dengan cara yang sama seperti tercantum dalam persamaan (2.5) pada tahap pelatihan.

Perhitungan Prediksi (Output Layer)

$$\hat{y} = H \cdot \beta \quad (11)$$

Keterangan

- $\hat{y}$: Hasil dari output layer.
 H : Fungsi aktivasi data testing.
 β : Matriks bobot output hasil training.

2.6. Evaluasi Hasil (*Confusion Matrix*). Untuk memastikan akurasi metode klasifikasi dan mengurangi kesalahan, evaluasi dilakukan dengan menggunakan confusion matrix, yang merupakan alat penting dalam menilai kinerja model klasifikasi [12]. Pada Tabel 1 merupakan matriks yang menyajikan hasil prediksi dalam bentuk tabel yang menggambarkan jumlah klasifikasi benar dan salah, sehingga memudahkan dalam mengevaluasi tingkat ketepatan model [20].

TABEL 1. Confusion Matrix

	Kelas Prediksi SOPK	Kelas Prediksi Normal
Kelas Aktual SOPK	TP	FN
Kelas Aktual Normal	FP	TN

Keterangan:

- **TP (True Positive)** : Data aktual untuk kategori SOPK yang diklasifikasikan sebagai SOPK, menunjukkan bahwa pasien tersebut memang terdiagnosis mengidap penyakit SOPK.
- **FP (False Positive)** : Data aktual kategori normal tetapi diklasifikasikan sebagai SOPK, menunjukkan bahwa pasien sebenarnya sehat namun salah diklasifikasikan sebagai penderita SOPK.
- **FN (False Negative)** : Data aktual kategori SOPK tetapi diklasifikasikan sebagai normal, menunjukkan bahwa pasien tersebut sebenarnya menderita SOPK namun sistem salah mengklasifikasi sebagai sehat.
- **TN (True Negative)** : Data aktual kategori normal yang diklasifikasikan sebagai normal, menunjukkan bahwa pasien sehat teridentifikasi dengan tepat.

Indikator evaluasi klasifikasi:

2.6.1. Akurasi. Mewakili tingkat keberhasilan keseluruhan klasifikasi.

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

2.6.2. *Sensitivitas*. Menunjukkan kemampuan sistem mengidentifikasi pasien SOPK secara akurat.

$$\text{Sensitivitas} = \frac{TP}{TP + FN} \times 100\%$$

2.6.3. *Spesifisitas*. Mewakili kemampuan sistem mengidentifikasi pasien sehat (tidak SOPK) secara tepat.

$$\text{Spesifisitas} = \frac{TN}{TN + FP} \times 100\%$$

2.7. **Hyperparameter Tuning**. Meningkatkan akurasi model prediksi dapat dilakukan melalui berbagai cara, salah satunya dengan menyetel hyperparameter secara optimal. Proses ini sangat penting dalam pengembangan deep learning karena dapat berdampak besar pada kinerja model. Peneliti terus mengeksplorasi bagaimana pengaruh pengaturan parameter ini terhadap hasil akhir, agar proses penyesuaian menjadi lebih efektif. Beberapa strategi yang umum digunakan antara lain melatih model pada sebagian kecil data untuk efisiensi, serta memanfaatkan pendekatan berbasis pengalaman sebelumnya yang menggabungkan metode optimasi dan peningkatan model alternatif [21].

2.7.1. *Random Search*. Random Search adalah metode populer untuk melakukan optimasi hyperparameter dengan memilih kombinasi nilai secara acak dari rentang yang telah ditentukan. Metode ini bertujuan menemukan nilai terbaik dari parameter yang berada dalam batas bawah dan atas yang telah ditetapkan (iterasi) [20].

2.7.2. *Grid Search*. Grid Search merupakan salah satu metode tuning yang paling sederhana untuk diterapkan. Namun, metode ini menjadi kurang efisien saat jumlah hyperparameter yang dicoba semakin banyak. Prinsip kerja Grid Search ialah setiap kombinasi nilai parameter disusun dalam bentuk grid antar dua parameter berbeda. Titik optimal biasanya berada pada perpotongan kombinasi nilai dari kedua parameter tersebut. Grid Search akan menguji seluruh kombinasi parameter yang sudah ditentukan sebelumnya dalam ruang pencarian, lalu memilih konfigurasi terbaik berdasarkan hasil evaluasi performa model [20].

Melakukan uji coba hyperparameter tuning menggunakan algoritma Random Search dan Grid Search. Hasil parameter terbaik akan dipilih untuk dilatih menggunakan metode PSO-ELM. Hyperparameter tuning yang akan diuji dalam penelitian ini terdapat pada Tabel 2.

TABEL 2. Hyperparameter Tuning

No	Parameter	Deskripsi	Value	Default
1	ω (Inertia Weight)	Mengontrol kontribusi kecepatan sebelumnya terhadap kecepatan saat ini. Nilai ω mempengaruhi kemampuan eksplorasi dan eksploitasi partikel.	0.4, 0.6, 0.8, 1.0	0.8
2	n_{hidden} (Jumlah Neuron Tersembunyi)	Jumlah neuron dalam hidden layer neural network. Semakin besar nilainya, semakin tinggi kompleksitas model.	5, 10, 15, 20	10
3	c_1 (Cognitive Coefficient)	Mengontrol pembelajaran partikel berdasarkan pengalaman pribadi (self-learning).	0.5, 1.0, 1.5, 2.0, 2.5	2.0
4	c_2 (Social Coefficient)	Mengontrol pembelajaran partikel berdasarkan pengalaman terbaik seluruh populasi (social learning).	0.5, 1.0, 1.5, 2.0, 2.5	2.0

3. HASIL DAN PEMBAHASAN

3.1. **Normalisasi Data**. Normalisasi data adalah proses untuk menyamakan skala nilai-nilai dalam suatu dataset agar berada dalam rentang yang seragam. Tujuannya adalah untuk menghindari dominasi fitur tertentu yang memiliki nilai jauh lebih besar dibandingkan fitur lainnya. Dalam penelitian ini, beberapa variabel memiliki rentang nilai yang sangat berbeda,

sehingga dilakukan proses normalisasi ke dalam rentang antara 0 dan 1. Dengan cara ini, setiap fitur memiliki kontribusi yang seimbang dalam proses analisis dan pemodelan. Tabel 3 memperlihatkan sebagian hasil dari normalisasi data SOPK.

TABEL 3. Hasil Normalisasi Data

No	Age (yrs)	Weight (kg)	Height (cm)	Endometrium (mm)
1	0.286	0.177	0.263	0.472
2	0.571	0.442	0.513	0.206
3	0.464	0.491	0.605	0.556
4	0.607	0.442	0.158	0.417
⋮	⋮	⋮	⋮	⋮
538	0.107	0.662	0.605	0.383

3.2. Pembagian Data (*K-fold Cross Validation*). Dataset yang digunakan terdiri dari 538 data, dengan 432 data untuk pelatihan dan 106 untuk pengujian. Metode 5-fold cross-validation membagi data menjadi lima bagian seimbang. Setiap putaran menggunakan empat bagian untuk melatih model dan satu bagian untuk menguji. Dengan cara ini, seluruh data akan bergiliran menjadi data uji, sehingga hasil evaluasi model menjadi lebih menyeluruh dan adil.

3.3. Seleksi Fitur PSO. Particle Swarm Optimization (PSO) digunakan sebagai metode seleksi fitur untuk menemukan kombinasi fitur yang paling mendukung performa klasifikasi. Dari 40 fitur awal dalam dataset berisi 538 sampel, PSO berhasil menyaring 18 fitur yang paling relevan. Fitur-fitur tersebut dapat dilihat pada Tabel 4 berikut.

TABEL 4. Fitur Terpilih dari PSO

No	Indeks Fitur	Nama Fitur
1	2	Height (cm)
2	3	BMI
3	5	Pulse rate (bpm)
⋮	⋮	⋮
18	38	Avg. F size (R) (mm)

3.4. Evaluasi Klasifikasi ELM tanpa Seleksi Fitur. Pada tahap ini, klasifikasi dilakukan menggunakan algoritma ELM tanpa penerapan seleksi fitur melalui PSO. Tujuannya adalah untuk mengevaluasi performa dasar model saat seluruh fitur digunakan apa adanya, tanpa proses penyaringan. Pendekatan ini memberi gambaran sejauh mana kemampuan ELM dalam mengolah data mentah tanpa dukungan teknik seleksi fitur.

TABEL 5. Hasil Klasifikasi ELM

Akurasi	Sensitivitas	Spesifisitas	Waktu
84.20%	90.10%	70.62%	8 detik

Dapat dilihat pada Tabel 5 yang merupakan tabel klasifikasi ELM, didapatkan bahwa nilai akurasi diperoleh sebesar 84.20%, dengan nilai sensitivitas 90.10% dan nilai spesifisitas 70.62%.

3.5. Evaluasi Klasifikasi PSO-ELM dengan *Random Search*. Bagian ini memaparkan hasil evaluasi model yang dioptimalkan menggunakan pendekatan Random Search. Metode ini memilih kombinasi hyperparameter secara acak dari nilai yang telah ditentukan sebelumnya, sehingga lebih efisien dalam menjelajahi ruang parameter tanpa harus menguji semua kemungkinan. Untuk mengukur performa model, digunakan confusion matrix pada setiap fold selama validasi silang. Pengujian dilakukan dalam beberapa iterasi, yaitu 10, 50, 100, 500, dan 1000, guna mengamati konsistensi hasil secara menyeluruh. Tabel 6 merupakan perbandingan akurasi terbaik dari beberapa percobaan iterasi terhadap model menggunakan Random Search:

TABEL 6. Klasifikasi PSO-ELM dengan Random Search

Iterasi	ω	n_{hidden}	c_1	c_2	Akurasi	Waktu
10	0.8	10	0.5	0.5	94.60%	2 menit 11 detik
50	0.4	20	1.0	2.0	94.79%	8 menit 28 detik
100	0.8	10	0.5	0.5	95.35%	16 menit 32 detik
500	0.5	20	2.0	2.0	86.98%	41 menit 34 detik
1000	0.8	15	0.5	1.5	85.13%	42 menit 10 detik

Dapat dilihat pada Tabel 6 yang merupakan tabel perbandingan hasil dari random search dari beberapa iterasi, didapatkan bahwa nilai akurasi terbaik diperoleh dari iterasi 100 yakni sebesar 95.35%, dengan nilai sensitivitas 96.67% dan nilai spesifisitas 92.65%.

3.6. Evaluasi Klasifikasi PSO-ELM dengan Grid Search. Bagian ini menyajikan hasil evaluasi model yang telah dioptimalkan melalui grid search. Metode ini melakukan pencarian hyperparameter secara sistematis dengan menguji seluruh kombinasi nilai yang telah ditentukan sebelumnya. Dengan pendekatan menyeluruh ini, kemungkinan untuk memperoleh konfigurasi parameter terbaik menjadi lebih tinggi karena seluruh ruang pencarian dijelajahi secara komprehensif.

TABEL 7. Klasifikasi PSO-ELM dengan Grid Search

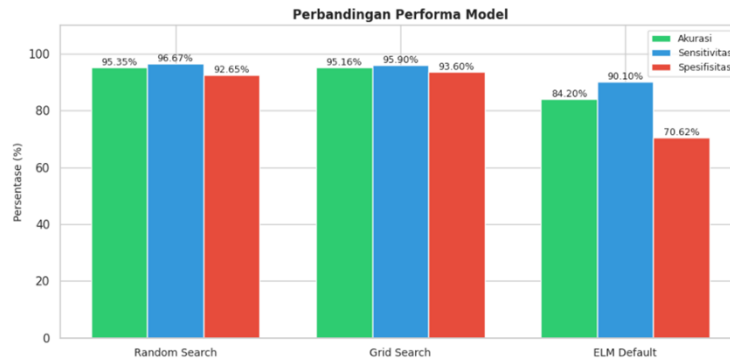
ω	n_{hidden}	c_1	c_2	Akurasi	Waktu
0.8	10	0.5	0.5	95.16%	1 menit 59 detik

Dapat dilihat pada Tabel 7 yang merupakan tabel grid search yang didapatkan bahwa nilai akurasi terbaik diperoleh sebesar 95.16%, dengan nilai sensitivitas 95.90% dan nilai spesifisitas 93.60%.

3.7. Perbandingan Evaluasi Klasifikasi. Penelitian ini membandingkan dua pendekatan tuning hyperparameter, yaitu grid search dan random search, dalam mengoptimalkan performa model ELM. Berbeda dengan model ELM yang tidak melalui proses tuning maupun seleksi fitur, kedua metode pencarian parameter ini secara signifikan meningkatkan kinerja model.

TABEL 8. Perbandingan Evaluasi Klasifikasi

	Grid Search	Random Search (100 iterasi)	ELM Default
Akurasi	95.16%	95.35%	84.20%
Sensitivitas	95.90%	96.67%	90.10%
Spesifisitas	93.60%	92.65%	70.62%
Parameter Terbaik			
ω	0.8	0.8	-
n_{hidden}	10	10	20
c_1	0.5	0.5	-
c_2	0.5	0.5	-
Waktu	1m 59d	16m 32d	8d



GAMBAR 2. Perbandingan Evaluasi Klasifikasi

Dapat dilihat pada Tabel 8 dan Gambar 2 bahwa model ELM hanya menggunakan parameter bawaan tanpa proses pencarian nilai terbaik, sehingga meskipun pelatihannya lebih cepat, performanya cenderung kurang optimal. Hal ini tercermin dari akurasinya yang hanya mencapai 84.20%, sensitivitas 90.10%, dan spesifisitas yang cukup rendah, yaitu 70.62%. Nilai spesifisitas yang rendah menunjukkan tingginya kemungkinan terjadinya false positive, yang cukup krusial dalam konteks medis. Dari hasil evaluasi, random search menunjukkan performa terbaik dengan akurasi 95.35%, sensitivitas 96.67%, dan spesifisitas 92.65%. Meski grid search memiliki spesifisitas sedikit lebih tinggi (93.60%), secara keseluruhan random search memberikan hasil yang lebih seimbang dan konsisten. Oleh karena itu, model yang dihasilkan dari tuning menggunakan random search dipilih sebagai model terbaik dalam penelitian ini.

TABEL 9. Perbandingan Metode pada Penelitian Terdahulu

No	Metode	Akurasi
1	Support Vector Machine (SVM) dan Genetic Algorithm (GA) [22]	94.26%
2	Random Forest dan RapidMiner [7]	85.00%
3	Support Vector Machine (SVM) dan RapidMiner [7]	93.12%
4	K-Nearest Neighbor (KNN) dan RapidMiner [7]	91.13%
5	K-Nearest Neighbors (KNN) dengan EDA dan Recursive Feature Elimination (RFECV) [23]	93.00%
6	Extreme Learning Machine (ELM) dan Particle Swarm Optimization (PSO)	95.35%

Berdasarkan Tabel 9, penerapan metode Extreme Learning Machine (ELM) yang dikombinasikan dengan Particle Swarm Optimization (PSO) menunjukkan hasil terbaik dibandingkan pendekatan lain pada dataset [13]. PSO membantu memilih fitur paling relevan, sementara ELM menawarkan kecepatan dan akurasi tinggi dalam klasifikasi. Hasil ini menegaskan bahwa pemilihan algoritma yang tepat sangat berpengaruh terhadap performa prediksi SOPK.

4. SIMPULAN

Penelitian ini membuktikan bahwa kombinasi metode Particle Swarm Optimization (PSO) untuk seleksi fitur dan algoritma Extreme Learning Machine (ELM) sebagai model klasifikasi mampu meningkatkan akurasi analisis data medis pada pasien Sindrom Ovarium Polistik (SOPK). Dari 40 atribut awal, PSO berhasil menyeleksi 18 fitur paling relevan yang mencakup data fisiologis, hormonal, gejala klinis, dan hasil USG. Pada tahap klasifikasi, model PSO-ELM yang dioptimalkan menggunakan random search mencatat performa terbaik dengan akurasi 95.35%, sensitivitas 96.67%, dan spesifisitas 92.65%, mengungguli metode grid search yang sedikit lebih rendah. Sebaliknya, penggunaan ELM tanpa seleksi fitur hanya mencapai akurasi 84.20% dengan spesifisitas yang jauh lebih rendah, yakni 70.62%. Hasil ini menunjukkan bahwa pendekatan random search lebih efisien dalam menemukan parameter optimal dan bahwa kombinasi PSO-ELM berpotensi besar sebagai sistem pendukung keputusan berbasis kecerdasan buatan dalam diagnosis SOPK.

DAFTAR PUSTAKA

- [1] Julia Elviethasari, Budi Santoso, Budiono, and Sulistiawati, "Knowledge of General Practitioners about Polycystic Ovarian Syndrome at the Primary Health Care in Surabaya, Indonesia," *J. Indones. Med. Assoc.*, vol. 70, no. 8, pp. 144–150, 2020, doi: 10.47830/jinma-vol.70.8-2020-228
- [2] WHO, "No Title," World Health Organization.
- [3] U. N. Alifah, "No Title," alomedika.
- [4] I. Putra, "PCOS DAN FERTILITAS TERHADAP GAYA HIDUP SEHAT," Kemenkes.
- [5] N. L. P. R. Dewi, "Pendekatan Terapi Polycystic Ovary Syndrome (PCOS)," *Cermin Dunia Kedokt.*, vol. 47, no. 11, p. 703, 2020, doi: 10.55175/cdk.v47i11.1201.
- [6] H. Jantan, U. Fatihah, M. Bahrin, and L. H. Shaufee, "Polycystic Ovary Syndrome (PCOS) Prediction System Using PSO-SVM," *J. Comput. Res. Innov.*, vol. 9, no. 1, p. 2024, 2024, doi: 10.24191/jcrinn.v9i1.414.
- [7] S. C. R. Nandipati, X. Chew, and K. K. Wah, "Polycystic Ovarian Syndrome (PCOS) Classification and Feature Selection by Machine Learning Techniques," *Appl. Math. Comput. Intell.*, vol. 9, no. 1, pp. 65–74, 2020.
- [8] J. Wang, S. Lu, S. H. Wang, and Y. D. Zhang, "A review on extreme learning machine," *Multimed. Tools Appl.*, vol. 81, no. 29, pp. 41611–41660, 2022, doi: 10.1007/s11042-021-11007-7.
- [9] M. I. S. Saad, "Perbandingan Algoritma Extreme Learning Machine dan Multilayer Perceptron Dalam Prediksi Mahasiswa Drop Out," *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 369–376, 2023, doi: 10.47065/bit.v4i3.890.
- [10] S. K. Rath, M. Sahu, S. P. Das, S. K. Bisoy, and M. Sain, "A Comparative Analysis of SVM and ELM Classification on Software Reliability Prediction Model," *Electron.*, vol. 11, no. 17, 2022, doi: 10.3390/electronics11172707..
- [11] D. Octaviana Aziza and K. I. Kurniati, "Suplementasi Vitamin D pada Wanita dengan Polycystic Ovarian Syndrome (PCOS)," *J. Ilm. Kesehat. Sandi Husada*, vol. 8, no. 2, pp. 169–177, 2019, doi: 10.35816/jiskh.v8i2.140.
- [12] R. N. K. Ahmadah, B. Rahayudi, and Y. A. Sari, "Optimasi Extreme Learning Machine Dengan Particle Swarm Optimization Untuk Klasifikasi Penyakit Jantung Koroner," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 9, pp. 3894–3900, 2021.
- [13] V. T. Shreyas Vedpathak, "PCOS Dataset," Kaggle.com.
- [14] T. D. Sukma, I. Cholissodin, and E. Santoso, "Penerapan Metode Extreme Learning Machine (ELM) dengan Optimasi Particle Swarm Optimization (PSO) untuk Memprediksi Harga Cabai Keriting di Kota Malang," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 9, pp. 3941–3949, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [15] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *J. Algoritma*, vol. 21, no. 1, pp. 239–248, 2024, doi: 10.33364/algoritma/v.21-1.1618.
- [16] M. Jain, V. Saihpal, N. Singh, and S. B. Singh, "An Overview of Variants and Advancements of PSO Algorithm," *Appl. Sci.*, vol. 12, no. 17, pp. 1–21, 2022, doi: 10.3390/app12178392
- [17] M. H. Ramdani, G. Pasek, S. Wijaya, and R. Dwiyanaputra, "Optimalisasi Pengenalan Wajah Berbasis Linear Discriminant Analysis dan K-Nearest Neighbor Menggunakan Particle Swarm Optimization (Optimization Of Face Recognition Based On Linear Discriminant Analysis and K-Nearest Neighbor Using Particle Swarm Optimiz)," vol. 4, no. 1, pp. 40–51, 2022, [Online]. Available: <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- [18] V. V. Nurdiansyah, I. Cholissodin, and P. P. Adikara, "Klasifikasi Penyakit Tuberkulosis (TB) menggunakan Metode Extreme Learning Machine (ELM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 5, pp. 1387–1393, 2020.
- [19] M. Karmila and I. Nirmala, "Prediksi Jumlah Produksi Kebutuhan Air Pada Perumda Air Minum Tirta Khatulistiwa Pontianak Menggunakan Metode Extreme Learning Machine (Elm)," *Coding J. Komput. dan Apl.*, vol. 11, no. 1, p. 137, 2023, doi: 10.26418/coding.v11i1.58052.
- [20] Shalehah, Muhammad Itqan Mazdadi, Andi Farmadi, Dwi Kartini, and Muliadi, "Implementation of Particle Swarm Optimization Feature Selection on Naïve Bayes for Thoracic Surgery Classification," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 5, no. 3, pp. 150–158, 2023, doi: 10.35882/jeemi.v5i3.305.
- [21] W. Nugraha and A. Sasongko, "Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search Hyperparameter Tuning on Classification Algorithm with Grid Search," *Sist. J. Sist. Inf.*, vol. 11, no. 2, pp. 391–401, 2022, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>.
- [22] F. Novianti and N. Ulinuha, "Data Rekam Medis Pcos Menggunakan Svm Feature Selection Using Genetic Algorithm In Pcos Medical Record Data Classification Using Svm," vol. 9, no. 1, pp. 9–19, 2024
- [23] N. T. Pitaloka and K. Kusnawi, "Pcos Disease Classification Using Feature Selection Rfcv and Eda With Knn Algorithm Method," *J. Tek. Inform.*, vol. 4, no. 4, pp. 693–701, 2023, doi: 10.52436/1.ju.tif.2023.4.4.831.

