

OPTIMASI RANDOM FOREST-SSO UNTUK KLASIFIKASI RISIKO DIABETES BERBASIS REKOMENDASI

Zelviah^{1*}, Darniati², Ida Mulyadi³

^{1,2,3}Universitas Muhammadiyah Makassar

¹105841105922@student.unismuh.ac.id, ²darniati@unismuh.ac.id,

³idamulyadi@unismuh.ac.id

* Corresponding Author

Info Artikel

Pengajuan : 31-07-2026
Peninjauan : 10-08-2026
Revisi : 25-08-2026
Diterima : 10-09-2026
Diterbitkan : 16-09-2026

Kata Kunci:

Diabetes Melitus, Random
Forest, Salp Swarm
Optimization, Klasifikasi
Multiclass, Rekomendasi.

Abstrak

Diabetes melitus merupakan penyakit kronis yang memerlukan klasifikasi risiko secara sistematis untuk mendukung proses skrining dan pemantauan pasien. Penelitian ini mengoptimalkan Random Forest menggunakan Salp Swarm Optimization (SSO) untuk mengklasifikasikan risiko diabetes ke dalam kategori Normal, Pra-Diabetes, dan Diabetes serta memberikan rekomendasi awal. Dataset terdiri atas 500 data hasil pemeriksaan pasien UPTD Puskesmas Larompong tahun 2025 dengan 11 variabel masukan. Tahapan penelitian mencakup preprocessing, feature engineering, pembagian data 80% untuk pelatihan dan 20% untuk pengujian, pembangunan model baseline, optimasi enam hyperparameter menggunakan empat salp selama 40 iterasi, serta evaluasi model. Random Forest baseline memperoleh accuracy 88,00% dan macro F1-score 88,90%. Optimasi SSO menghasilkan *n_estimators* 247, *max_depth* 16, *max_features* 0,80, *min_samples_split* 4, *min_samples_leaf* 1, dan *class_weight* balanced. Model Random Forest-SSO memperoleh accuracy 90,00% serta macro precision, macro recall, dan macro F1-score sebesar 90,75%. Hasil tersebut menunjukkan bahwa SSO meningkatkan kinerja Random Forest dan mengurangi kesalahan klasifikasi antara kelas Pra-Diabetes dan Diabetes. Hasil prediksi digunakan sebagai dasar rekomendasi awal tanpa menggantikan pemeriksaan atau keputusan tenaga kesehatan.

How to Cite : Zelviah, Z., Darniati, D., & Mulyadi, I. (2026). Optimasi Random Forest-SSO Untuk Klasifikasi Risiko Diabetes Berbasis Rekomendasi. *Journal of Computer Science and Information Technology (JCSIT)*, 3(4), 429-435.

DOI : 10.70248/jcsit.v3i4.4362

This is an open access article under <https://creativecommons.org/licenses/by/4.0/>
Copyright© 2026 by Author. Published by Yayasan Nuraini Ibrahim Mandiri



PENDAHULUAN

Diabetes melitus merupakan penyakit metabolik kronis yang ditandai oleh peningkatan kadar glukosa darah akibat gangguan produksi atau penggunaan insulin. Penyakit ini dapat menimbulkan komplikasi apabila tidak dikenali dan ditangani sejak dini. Oleh karena itu, klasifikasi risiko diperlukan untuk mendukung skrining dan pemantauan kondisi pasien (Noviyanti & Alamsyah, 2024), (Ismail et al., 2021). UPTD Puskesmas Larompong memiliki data hasil pemeriksaan pasien yang memuat karakteristik individu, parameter klinis, serta faktor risiko diabetes. Data tersebut dapat dikembangkan melalui pendekatan machine learning untuk membantu mengelompokkan risiko pasien ke dalam kategori Normal, Pra-Diabetes, dan Diabetes secara lebih sistematis tanpa menggantikan keputusan tenaga kesehatan (Anjani & Abiyasa, 2023), (Aryani & Wijayanto, 2023).

Random Forest merupakan algoritma ensemble yang membangun sejumlah pohon keputusan dan menentukan kelas akhir melalui majority voting. Walaupun mampu menangani banyak fitur dan hubungan nonlinier, performanya dipengaruhi oleh pemilihan hyperparameter, khususnya *n_estimators*, *max_depth*, *max_features*, *min_samples_split*, *min_samples_leaf*, dan *class_weight* (Ardilla et al., 2022).

Salp Swarm Optimization merupakan metode optimasi berbasis swarm

intelligence yang dapat digunakan untuk mencari kombinasi hyperparameter terbaik. Penelitian terdahulu memperoleh akurasi 79% pada Random Forest untuk klasifikasi diabetes (Nainggolan & Sinaga, 2023), akurasi 79% pada Random Forest dengan Grid Search dan SMOTE (Aji Septa et al., 2024), serta akurasi 85% pada model ESSO-BiLSTM untuk pemantauan kesehatan (Dhand et al., 2022). SSO juga telah digunakan untuk mengoptimalkan Random Forest pada prediksi risiko penyakit lain (Bao et al., 2025). Namun, penerapannya untuk mengoptimalkan enam hyperparameter Random Forest pada klasifikasi risiko diabetes tiga kelas masih terbatas. Pengembangan dan penerapan SSO pada optimasi model machine learning telah dilaporkan pada berbagai penelitian sebelumnya (Abdulsaed et al., 2023), (Yang et al., 2024).

Berbagai metode metaheuristik seperti Particle Swarm Optimization (PSO) dan Genetic Algorithm (GA) telah digunakan dalam optimasi model machine learning. Namun, PSO memiliki ketergantungan terhadap parameter kecepatan partikel dan bobot inersia, sedangkan Genetic Algorithm membutuhkan proses seleksi, crossover, dan mutasi yang dapat meningkatkan kompleksitas pencarian. Salp Swarm Optimization (SSO) memiliki mekanisme eksplorasi dan eksploitasi melalui struktur leader dan follower dengan parameter kontrol yang relatif sederhana sehingga sesuai diterapkan untuk optimasi hyperparameter Random Forest pada ruang pencarian yang kompleks.

Penelitian ini menerapkan Random Forest-SSO pada 500 data hasil pemeriksaan pasien UPTD Puskesmas Larompong. Kontribusi penelitian terletak pada optimasi enam hyperparameter, perbandingan performa model sebelum dan setelah optimasi, analisis kontribusi fitur, serta pemanfaatan hasil klasifikasi sebagai dasar rekomendasi awal.

METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif berbasis machine learning. Data yang digunakan merupakan 500 data sekunder hasil pemeriksaan pasien UPTD Puskesmas Larompong tahun 2025. Dataset memiliki 11 variabel masukan dan satu variabel target berupa Status Diabetes. Identitas pribadi pasien tidak digunakan dalam pemodelan.

Tabel 1. Variabel yang digunakan dalam penelitian

No.	Variabel	No.	Variabel
1	Usia	7	Riwayat keluarga
2	Jenis kelamin	8	Aktivitas fisik
3	IMT	9	Pola makan
4	Lingkar perut	10	Status merokok
5	TDS	11	Gula Darah Sewaktu
6	TDD	12	Status Diabetes

Pemeriksaan awal mencakup missing value, data duplikat, konsistensi kolom, dan distribusi kelas. Variabel kategorikal dikonversi menjadi numerik, sedangkan target dikodekan menjadi Normal = 0, Pra-Diabetes = 1, dan Diabetes = 2. Feature engineering menghasilkan Indikator_GDS, IMT_Usia, Lingkar_Perut_Usia, TDS_TDD_Ratio, TDS_TDD_Diff, dan IMT_Lingkar sehingga diperoleh 16 fitur prediktor. Nilai GDS asli tidak digunakan langsung, tetapi ditransformasikan menjadi Indikator_GDS dengan nilai 1 untuk $GDS \geq 140$ mg/dL dan 0 untuk $GDS < 140$ mg/dL. Proses transformasi dilakukan menggunakan teknik encoding agar data kategorikal dapat diproses oleh algoritma Random Forest. Setiap kategori diberikan representasi numerik tanpa menghilangkan informasi karakteristik data.

Pemilihan empat salp dan 40 iterasi dilakukan berdasarkan pertimbangan

ukuran dataset penelitian serta hasil pengamatan konvergensi fitness. Jumlah populasi yang lebih besar dapat meningkatkan beban komputasi, sedangkan dataset penelitian memiliki jumlah 500 data sehingga empat salp dianggap cukup untuk melakukan eksplorasi ruang hyperparameter. Berdasarkan hasil eksperimen, nilai fitness terbaik telah tercapai sebelum iterasi ke-5 dan relatif stabil hingga iterasi ke-40 sehingga menunjukkan proses optimasi telah mengalami konvergensi.

Dataset dibagi menggunakan rasio 80:20 dengan random_state 42 dan stratifikasi kelas, sehingga diperoleh 400 data pelatihan dan 100 data pengujian. Model Random Forest baseline dibangun sebagai pembandingan, kemudian SSO digunakan untuk mengoptimalkan enam hyperparameter. Proses optimasi menggunakan empat salp, 40 iterasi, dan Stratified 5-Fold Cross-Validation pada data pelatihan.

Gambar 1. Alur penelitian Random Forest-SSO



Gambar 1. Alur penelitian Random Forest-SSO

Random Forest pada penelitian ini menggunakan criterion entropy. Entropy mengukur ketidakhomogenan data pada setiap node, dengan rumus:

$$Entropy(D) = -\sum_i p_i \log_2(p_i)$$

Setiap salp merepresentasikan satu kombinasi hyperparameter. Nilai fitness dihitung dari gabungan lima metrik evaluasi dengan bobot yang ditetapkan dalam laporan penelitian:

$$Fitness = 0,30A + 0,20P_m + 0,25R_m + 0,15F1_m + 0,10F1_w$$

A adalah accuracy, P_m adalah macro precision, R_m adalah macro recall, $F1_m$ adalah macro F1-score, dan $F1_w$ adalah weighted F1-score. Model akhir dievaluasi menggunakan confusion matrix, accuracy, macro precision, macro recall, macro F1-score, dan weighted F1-score. Dua rumus evaluasi utama yang digunakan adalah:

$$Accuracy = \text{jumlah prediksi benar} / \text{jumlah data uji}$$

$$F1\text{-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

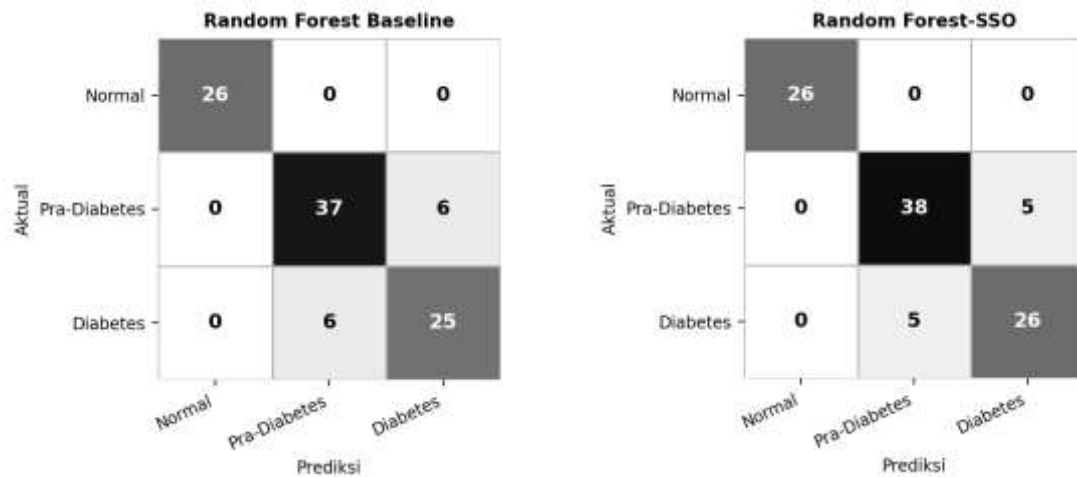
HASIL DAN PEMBAHASAN

Dataset terdiri atas 500 data dan 12 kolom tanpa missing value maupun data duplikat. Kelas Pra-Diabetes memiliki jumlah terbesar, yaitu 213 data, diikuti Diabetes sebanyak 155 data dan Normal sebanyak 132 data. Distribusi tersebut menunjukkan ketidakseimbangan kelas ringan sehingga stratifikasi digunakan pada pembagian data.

Tabel 2. Distribusi kelas risiko diabetes

Kelas	Jumlah	Persentase
Normal	132	26,40%
Pra-Diabetes	213	42,60%
Diabetes	155	31,00%
Total	500	100,00%

Random Forest baseline memperoleh accuracy 88,00%, macro precision 88,90%, macro recall 88,90%, macro F1-score 88,90%, dan weighted F1-score 88,00%. Seluruh 26 data Normal diprediksi benar. Kesalahan terjadi antara Pra-Diabetes dan Diabetes, masing-masing sebanyak enam data.



Gambar 2. Confusion matrix sebelum dan setelah optimasi

Untuk memperkuat analisis metode optimasi, dilakukan pembahasan komparatif terhadap pendekatan optimasi konvensional yaitu Grid Search dan Random Search. Grid Search melakukan pencarian secara menyeluruh terhadap seluruh kombinasi hyperparameter yang telah ditentukan, sehingga mampu menemukan kombinasi terbaik pada ruang pencarian terbatas namun memiliki konsekuensi peningkatan waktu komputasi ketika jumlah parameter semakin besar. Random Search melakukan pengambilan sampel kombinasi parameter secara acak sehingga jumlah evaluasi dapat dikurangi, tetapi proses pencarian tidak menggunakan informasi solusi terbaik sebelumnya. Sementara itu, Salp Swarm Optimization (SSO) menggunakan mekanisme swarm intelligence melalui interaksi leader dan follower sehingga proses pencarian solusi dapat dilakukan secara adaptif. Pada penelitian ini, SSO dipilih karena mampu melakukan optimasi enam hyperparameter Random Forest dengan proses pencarian yang lebih terarah.

Sebagai tambahan evaluasi metode optimasi, Grid Search memiliki pendekatan exhaustive search yang mengevaluasi seluruh kombinasi parameter sehingga waktu komputasi dapat meningkat ketika ruang pencarian semakin besar. Random Search mengurangi jumlah evaluasi dengan memilih kombinasi secara acak, tetapi tidak memiliki mekanisme adaptasi terhadap solusi terbaik sebelumnya. SSO memiliki keunggulan melalui mekanisme swarm intelligence yang memungkinkan pencarian solusi dilakukan secara adaptif melalui interaksi leader dan follower. Perbandingan eksperimen langsung dengan Grid Search dan Random Search belum dilakukan pada penelitian ini sehingga menjadi peluang pengembangan penelitian berikutnya.

Optimasi SSO mengevaluasi 80 kandidat unik selama 40 iterasi dengan waktu komputasi 270,20 detik. Nilai fitness terbaik sebesar 0,887361 telah ditemukan sebelum iterasi ke-5 dan tetap stabil sampai iterasi ke-40. Konfigurasi hasil optimasi ditampilkan pada Tabel 3.

Tabel 3. Parameter Random Forest sebelum dan setelah optimasi

Hyperparameter	Sebelum Optimasi	Setelah Optimasi SSO
n_estimators	800	247
max_depth	12	16
max_features	0,90	0,80
min_samples_split	2	4
min_samples_leaf	1	1
class_weight	balanced_subsample	balanced
criterion	entropy	entropy
bootstrap	True	True

Hasil optimasi menunjukkan bahwa peningkatan performa tidak selalu membutuhkan jumlah pohon yang lebih banyak. SSO menurunkan n_estimators dari 800 menjadi 247, tetapi meningkatkan accuracy dari 88,00% menjadi 90,00%. Perubahan max_depth, max_features, min_samples_split, dan class_weight menghasilkan konfigurasi yang lebih sesuai dengan karakteristik data.

Model Random Forest-SSO memperoleh accuracy 90,00%, macro precision 90,75%, macro recall 90,75%, macro F1-score 90,75%, dan weighted F1-score 90,00%. Jumlah prediksi benar meningkat dari 88 menjadi 90 data. Kesalahan antara kelas Pra-Diabetes dan Diabetes berkurang dari enam menjadi lima data pada masing-masing kelas.

Tabel 4. Perbandingan performa model

Metrik	RF Baseline	RF-SSO	Peningkatan
Accuracy	88,00%	90,00%	2,00 poin
Macro Precision	88,90%	90,75%	1,85 poin
Macro Recall	88,90%	90,75%	1,85 poin
Macro F1-score	88,90%	90,75%	1,85 poin
Weighted F1-score	88,00%	90,00%	2,00 poin

Analisis feature importance menunjukkan bahwa Indikator_GDS memberikan kontribusi tertinggi sebesar 54,89%, diikuti IMT_Usia 15,03%, TDS 6,75%, TDD 4,72%, dan lingkaran perut 3,11%. Nilai tersebut menggambarkan kontribusi fitur terhadap keputusan model dan tidak menunjukkan hubungan sebab akibat secara klinis.

Tabel 5. Lima fitur dengan kontribusi tertinggi

Peringkat	Fitur	Kontribusi
1	Indikator_GDS	54,89%
2	IMT_Usia	15,03%
3	TDS	6,75%
4	TDD	4,72%
5	Lingkaran perut	3,11%

Hasil klasifikasi model akhir digunakan untuk memberikan rekomendasi awal sesuai kategori prediksi. Random Forest-SSO menghasilkan kelas prediksi, sedangkan rekomendasi diberikan melalui aturan yang telah ditetapkan pada sistem.

Tabel 6. Rekomendasi berdasarkan kategori prediksi

Kategori	Rekomendasi Awal
Normal	Mempertahankan pola hidup sehat, mengonsumsi makanan bergizi seimbang, beraktivitas fisik secara teratur, menjaga berat badan ideal, dan melakukan pemeriksaan berkala.
Pra-Diabetes	Memperbaiki pola makan, mengurangi konsumsi tinggi gula, meningkatkan aktivitas fisik, menjaga berat badan, memantau gula darah, dan berkonsultasi dengan tenaga kesehatan.
Diabetes	Segera melakukan konsultasi dan pemeriksaan lanjutan serta mengikuti arahan

Kategori	Rekomendasi Awal
	tenaga kesehatan terkait pola makan, aktivitas fisik, pengobatan, dan pemantauan gula darah.

Secara numerik, accuracy 90,00% lebih tinggi dibandingkan penelitian Random Forest sebelumnya yang memperoleh 79% [6], Random Forest dengan Grid Search dan SMOTE sebesar 79% [7], serta ESSO-BiLSTM sebesar 85% [8]. Perbandingan ini bersifat deskriptif karena setiap penelitian menggunakan dataset, jumlah kelas, dan skenario pengujian yang berbeda.

Penelitian ini memiliki keterbatasan pada jumlah dataset yang digunakan yaitu sebanyak 500 data yang berasal dari satu fasilitas kesehatan. Kondisi tersebut menyebabkan kemampuan generalisasi model pada populasi yang lebih luas belum dapat dipastikan. Oleh karena itu, penelitian selanjutnya disarankan melakukan validasi eksternal menggunakan dataset dari fasilitas kesehatan atau wilayah berbeda.

Sebagai pembanding pendekatan optimasi hyperparameter, Grid Search melakukan evaluasi terhadap seluruh kombinasi parameter yang telah ditentukan sehingga mampu menemukan solusi optimal pada ruang pencarian tertentu. Namun, metode ini memiliki keterbatasan berupa peningkatan waktu komputasi ketika jumlah kombinasi parameter semakin besar. Random Search mengurangi jumlah evaluasi dengan memilih kombinasi parameter secara acak, tetapi tidak menggunakan informasi dari solusi terbaik sebelumnya. Sementara itu, Salp Swarm Optimization (SSO) menggunakan mekanisme swarm intelligence melalui interaksi leader dan follower sehingga pencarian solusi dapat dilakukan secara adaptif. Karakteristik tersebut menjadi alasan pemilihan SSO sebagai metode optimasi hyperparameter Random Forest pada penelitian ini.

KESIMPULAN

Penelitian ini berhasil mengoptimalkan enam hyperparameter Random Forest menggunakan Salp Swarm Optimization untuk klasifikasi risiko diabetes tiga kelas. Model baseline memperoleh accuracy 88,00%, sedangkan Random Forest-SSO memperoleh accuracy 90,00% dengan macro F1-score 90,75%. Optimasi juga mengurangi jumlah pohon dari 800 menjadi 247 dan menurunkan kesalahan klasifikasi antara Pra-Diabetes dan Diabetes. Fitur dengan kontribusi terbesar adalah Indikator_GDS dan IMT_Usia. Hasil klasifikasi dapat digunakan sebagai dasar pemberian rekomendasi awal, tetapi tetap memerlukan pemeriksaan dan keputusan tenaga kesehatan.

Pernyataan Mengenai Penggunaan Kecerdasan Buatan (Ai)

Kecerdasan buatan (AI) digunakan sebagai alat bantu dalam penyusunan naskah ini, terutama untuk memperbaiki bahasa, tata bahasa, dan penyusunan kalimat agar lebih jelas. AI tidak digunakan dalam pengolahan data, pemodelan Random Forest-SSO, analisis hasil, maupun pengambilan keputusan penelitian. Penulis tetap bertanggung jawab atas seluruh isi, hasil, dan kesimpulan penelitian. Seluruh bagian yang dibantu AI telah diperiksa dan disesuaikan oleh penulis.

Konflik Kepentingan

Penulis menyatakan tidak ada konflik kepentingan dalam penelitian ini.

Kontribusi Penulis

Konseptualisasi, Zelvia, Darniati, dan Ida Mulyadi; metodologi, Zelvia dan Darniati; pengolahan dan analisis data, Zelvia; validasi, Darniati dan Ida Mulyadi; penulisan—

penyusunan draf awal, Zelvia; penulisan—peninjauan dan penyuntingan, Darniati dan Ida Mulyadi; visualisasi, Zelvia. Seluruh penulis telah membaca dan menyetujui versi akhir naskah ini.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Makassar atas dukungan akademik dan fasilitas yang mendukung penyelesaian penelitian ini. Penulis juga menyampaikan terima kasih kepada UPTD Puskesmas Larompong atas penyediaan data sekunder yang digunakan dalam penelitian ini.

DAFTAR PUSTAKA

- C. N. Noviyanti and A. Alamsyah, "Early Detection of Diabetes Using Random Forest Algorithm," *Journal of Information System Exploration and Research*, vol. 2, no. 1, 2024, doi: 10.52465/joiser.v2i1.245.
- L. Ismail, H. Materwala, and J. Al Kaabi, "Association of risk factors with type 2 diabetes: A systematic review," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 1759-1785, 2021.
- S. Anjani and M. T. Abiyasa, *Disrupsi Digital dan Masa Depan Rekam Medis*. Selat Media, 2023.
- Y. Aryani and A. W. Wijayanto, "Klasifikasi Pengembalian Radar dari Ionosfer Menggunakan SVM, Naive Bayes dan Random Forest," *Komputika: Jurnal Sistem Komputer*, vol. 10, no. 2, pp. 111-117, 2023, doi: 10.34010/komputika.v10i2.4347.
- Y. Ardilla et al., *Data Mining dan Aplikasinya*. Penerbit Widina, 2022.
- S. P. Nainggolan and A. Sinaga, "Comparative Analysis of Accuracy of Random Forest and Gradient Boosting Classifier Algorithm for Diabetes Classification," *Sebatik*, vol. 27, no. 1, pp. 97-102, 2023, doi: 10.46984/sebatik.v27i1.2157.
- A. Aji Septa, A. Al Farizi, A. N. Khafid, D. Prasetyo, N. C. Romadhon, and F. S. Utomo, "Diabetes Detection Optimisation with Hyperparameter Tuning in Random Forest Algorithm," *Journal of Informatics and Interactive Technology*, vol. 1, no. 3, pp. 165-177, 2024, doi: 10.63547/jiite.v1i3.42.
- G. Dhand, K. Sheoran, P. Agarwal, and S. S. Biswas, "Deep enriched salp swarm optimization based bidirectional-long short term memory model for healthcare monitoring system in big data," *Informatics in Medicine Unlocked*, vol. 32, p. 101010, 2022, doi: 10.1016/j.imu.2022.101010.
- H. Bao, L. Sun, G. Cui, S. Cai, W. Zheng, H. Peng, and C. Yang, "Salp swarm-optimized machine learning models for predicting preoperative aortic rupture risk in acute type A aortic dissection patients," *Frontiers in Physiology*, vol. 16, p. 1675853, 2025, doi: 10.3389/fphys.2025.1675853.
- E. Abdulsaed, M. Alabbas, and R. Khudeyer, "Hyperparameter Optimization for Convolutional Neural Networks using the Salp Swarm Algorithm," *Informatica*, vol. 47, no. 9, 2023, doi: 10.31449/inf.v47i9.5148.
- Z. Yang, Y. Jiang, and W.-C. Yeh, "Self-learning salp swarm algorithm for global optimization and its application in multi-layer perceptron model training," *Scientific Reports*, vol. 14, 2024, doi: 10.1038/s41598-024-77440-4.
- B. H. Abed-alguni, D. Paul, and R. Hammad, "Improved Salp swarm algorithm for solving single-objective continuous optimization problems," *Applied Intelligence*, 2022, doi: 10.1007/s10489-022-03269-x.