SVM-Based Diagnosis of Pneumonia with MGSPLO Wrapper Feature Selection on Blood Routine and Inflammatory Markers

Abstract

In computationally expensive optimization problems (EOPs), surrogate-assisted evolutionary algorithms (SAEAs) are a mainstream approach for reducing costly function evaluations. Most existing SAEAs manually design the search strategy and select criteria based solely on model uncertainty, which limits both dynamic solution space exploration and model accuracy. To address these issues, we introduce a bandit-driven adaptive search mechanism and an explainable uncertainty criterion into SAEAs. Instead of explicitly defining the optimizer for exploration and exploitation, the proposed bandit-driven adaptive search autonomously adjusts its search mode based on environmental feedback using reinforcement learning, providing a simple yet effective solution for exploring the search space. Additionally, by incorporating entropy of feature differences into the expected improvement (EI) criterion, the explainable uncertainty criterion selects more informative samples, enhancing surrogate model accuracy. Numerical experiments demonstrate the superiority of the proposed BEXEA over seven state-of-the-art SAEAs on two complex benchmark sets and a real-world oil reservoir production optimization problem. This work presents an efficient SAEA framework for expensive optimization. The source code for BEXEA is available at `https://github.com/jonawon/BEXEA.git`.

Keywords: Expensive optimization, Surrogate-assisted evolutionary algorithm, Bandit-driven adaptive search, Explainable uncertainty criteria

1. Introduction

Pneumonia is a common and sometimes life-threatening lung infection that causes major illness and death around the world.[1, 2] Community-acquired pneumonia (CAP) ranges from mild disease that can be treated as an outpatient to severe illness with respiratory failure and sepsis requiring intensive care.[3, 5] In current practice, clinicians combine symptoms and signs with chest imaging (usually chest X-ray, and sometimes CT) and basic laboratory tests to make a diagnosis.[3, 4] However, chest radiography can be unavailable in some settings, may be inconclusive, and can even be negative early in the disease course; CT is more sensitive but not always feasible.[7, 8, 9] These limitations create diagnostic uncertainty, delay treatment, and may encourage unnecessary antibiotic use. Better, data-driven tools that use routine clinical information could help diagnose pneumonia earlier and more reliably.

Blood biomarkers reflecting the host response to infection are widely used in suspected pneumonia. These include complete blood count (CBC) indices and acute-phase reactants such as C-reactive protein (CRP) and procalcitonin (PCT).[10, 11, 12] Elevated leukocyte counts with neutrophil predominance and higher CRP levels are typical in bacterial CAP and are linked to worse outcomes.[13, 14, 15] Ratios derived from the CBC — the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and monocyte-to-lymphocyte ratio (MLR) — summarize inflammation and have shown promise for risk stratification and prognosis in CAP.[16, 17, 18, 19] At the same time, recent work has explored machine-learning (ML) models that combine multiple laboratory and clinical variables to predict pneumonia or its complications, often improving discrimination beyond any single marker or traditional scores.[20, 21, 22, 23, 24] Many ML studies still emphasize imaging-derived features from chest X-rays or CT scans,[25, 26] or they use a small, fixed set of lab variables, which may miss useful information present in the full panel of hematologic and inflammatory indices.

From a data-analysis perspective, even a "modest" blood panel becomes high-dimensional once we include derived indices, ratios, and simple interaction terms. Training ML models directly on such expanded feature sets can cause overfitting, high variance, and lower interpretability, especially in clinical datasets with limited sample sizes.[27, 28, 29] Feature selection (FS) is therefore important to find compact, informative subsets of markers that maintain (or improve) performance and remain clinically interpretable.[27, 28] Broadly, FS methods are grouped as filters (fast, model-agnostic), embedded methods (e.g., LASSO or tree-based models), and wrappers (which evaluate candidate subsets using a downstream classifier).[29, 30, 31, 32] Wrappers often perform well because they account for feature interactions, but they can be computationally expensive.[32]

The main challenge for wrapper FS is the combinatorial search: with D candidate features there are $2^D - 1$ non-empty subsets to consider.[27] Metaheuristic optimization has become popular for wrapper FS because it is flexible, derivative-free, and designed to escape local optima in rugged, high-dimensional landscapes.[33, 34] Representative algorithms include Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), Harris Hawks Optimization (HHO), and the Slime Mould Algorithm (SMA).[35, 36, 37, 38, 39] These methods have delivered strong empirical results in biomedical tasks, including clinical diagnosis and prognosis from heterogeneous data.[40, 41] Yet standard metaheuristics can still suffer from premature convergence, unstable exploration–exploitation balance, and sensitivity to hyperparameters, problems that are amplified when each fitness evaluation retrains a classifier in a wrapper loop.[42, 43, 44]

To improve search quality, many enhanced or hybrid strategies have been proposed, such as multi-population/cooperative learning, dynamic neighborhood structures, opposition-based learning (OBL), and problem-specific local search.[45, 46, 47, 48, 49] However, even these designs may over-rely on a single best solution or lose population diversity over iterations; mapping continuous updates to binary feature decisions can also be coarse. In practice, this can lead to unstable feature subsets and variable performance across resampling folds, making it hard to extract a small, interpretable, and reliable set of markers for real-world deployment. When binary decisions are needed, refined transfer functions (e.g., S-shaped and V-shaped) can make that mapping more granular.[50]

Our approach. We propose a *Multi-Guide Slime-Predator Learning Optimizer* (MGSPLO) for wrapper-based feature selection on blood routine and inflammatory markers. MGSPLO is designed to: (i) use *multi-anchor guidance*, letting each solution learn from multiple elite subsets rather than a single incumbent best; (ii) maintain population diversity via adaptive exploration operators inspired by predator–prey dynamics and slime-mould-like oscillatory behavior; and (iii) apply a binary mapping scheme that turns continuous trajectories into probability-driven feature inclusion decisions. In our diagnostic pipeline, MGSPLO wraps a support vector machine (SVM), using cross-validated performance as the fitness signal to drive the search toward compact, high-performing subsets.

We study a cohort of patients evaluated for pneumonia, with age, comorbidities, blood routine indices (WBC and differential counts; red cell and platelet parameters) and CRP measured at presentation. From these we derive composite markers (e.g., NLR, PLR), creating a feature pool that reflects both routine clinical practice and emerging biomarker evidence. Our goals are to: (1) quantify the diagnostic value of blood routine and inflammatory markers—alone and in combination—for pneumonia detection; (2) test whether MGSPLO-driven wrapper FS can identify small, stable, interpretable subsets that match or improve SVM performance compared with using all markers or standard FS baselines; and (3) relate the selected markers to current biomarker literature.

Contributions.

- We introduce MGSPLO, a multi-guide, diversity-preserving metaheuristic for wrapper-based feature selection that reduces premature convergence and single-anchor bias in high-dimensional clinical feature spaces.
- We derive a binary MGSPLO with a refined transfer mechanism supporting probability-driven inclusion of blood routine and inflammatory markers.
- We build an end-to-end SVM pneumonia pipeline that integrates MGSPLO with cross-validated evaluation and compare against standard FS techniques and alternative wrapper metaheuristics.

- On a real-world cohort, we show that MGSPLO–SVM achieves competitive or superior diagnostic accuracy while selecting compact, interpretable panels of routine laboratory markers, which is helpful where imaging resources are limited.

The rest of the paper is organized as follows: Section 2 reviews related work on pneumonia biomarkers, ML-based diagnosis, and feature selection. Section 3 describes MGSPLO, its binary version, and the SVM framework. Section 4 details the dataset, features, and experiments. Section 5 reports results and ablations. Section 6 discusses limitations, clinical implications, and future work.

References

World Health Organization. *Pneumonia*. Available at: https://www.who.int/health-topics/pneumonia. Accessed 2025.

GBD collaborators. Trends in the global burden of lower respiratory infections: the knowns and unknowns. *Lancet Infect Dis*. 2022. Available at: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(22)00445-5/fulltext.

Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. *Am J Respir Crit Care Med*. 2019;200(7):e45–e67. (ATS/IDSA guideline). https://www.atsjournals.org/doi/10.1164/rccm.201908-1581ST.

National Institute for Health and Care Excellence (NICE). *Pneumonia: diagnosis and management* (NG250). Updated 2025. https://www.nice.org.uk/guidance/ng250.

Musher DM, Thorner AR. Community-Acquired Pneumonia. *JAMA*. 2014;311(20):2197–2206. (See also JAMA 2020 updates). https://jamanetwork.com/journals/jama/fullarticle/2760882.

NICE Quality Standard QS110. Chest X-ray and antibiotic treatment within 4 hours. Updated 2025. https://www.nice.org.uk/guidance/qs110.

Self WH, et al. The clinical utility of chest radiography for identifying pneumonia. *AJR Am J Roentgenol*. 2020;214(6):1208–1212. https://www.ajronline.org/doi/pdf/10.2214/AJR.19.21521.

Okada F, et al. Community-Acquired Pneumonia with Negative Chest Radiography Findings: Clinical and Radiological Characteristics. *Respiration*. 2018;97(6):508–516. https://karger.com/res/article/97/6/508/290847/Community-Acquired-Pneumonia-with-Negative-Chest.

Claessens YE, et al. Early chest CT-scan to assist diagnosis and guide treatment decision for suspected CAP in the ED. 2015 randomized trial. https://research.pasteur.fr/en/publication/early-chest-ct-scan-to-assist-diagnosis-and-guide-treatment-decision-for-suspected-community-acqu

Minnaard MC, et al. The added value of CRP in diagnosing pneumonia in primary care: IPD meta-analysis. *CMAJ*. 2017;189(2):E56–E63. https://www.cmaj.ca/content/189/2/E56.

Schuetz P, et al. Procalcitonin to guide initiation and duration of antibiotics in acute respiratory infections: IPD meta-analysis. *Clin Infect Dis*. 2012;55(5):651–662. https://academic.oup.com/cid/article/55/5/651/350305.

Schuetz P, et al. Effect of procalcitonin-guided antibiotic treatment on mortality in acute respiratory infections. *Lancet Infect Dis*. 2018;18(1):95–107. https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(17)30592-3/fulltext.

Krüger S, et al. Procalcitonin predicts severity and outcome in CAP; comparison with CRP and leukocyte count. *Eur Respir J*. 2008;31(2):349–355. https://publications.ersnet.org/content/erj/31/2/349.full.pdf.

Chalmers JD, et al. CRP, severity of pneumonia and mortality in elderly patients. *Age Ageing*. 2009;38(6):693–697. https://academic.oup.com/ageing/article-abstract/38/6/693/40913.

Menéndez R, et al. Utility of C-reactive protein in assessing severity and outcomes in CAP. *J Infect*. 2008;56(5):355–362. https://www.sciencedirect.com/science/article/pii/S1198743X14617985.

Lee JH, et al. Prognostic value of serial neutrophil-to-lymphocyte ratio in hospitalized CAP. *PLoS One*. 2021;16(4):e0250067. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0250067.

Enersen CCE, et al. NLR and PLR and association with mortality in CAP: derivation–validation cohort. *Infection*. 2023;51:1339–1347. https://link.springer.com/article/10.1007/s15010-023-01992-2.

Colosi IA, et al. The Neutrophil/Lymphocyte Ratio and Outcomes in Hospitalized CAP. *Biomedicines*. 2024;12(2):260. https://www.mdpi.com/2227-9059/12/2/260.

Simadibrata DM, et al. NLR, MLR, PLR and RDW to predict outcome and differentiate etiologies in pneumonia. *Sci Rep*. 2022;12:16109. https://www.nature.com/articles/s41598-022-20385-3.pdf.

Ye Y, et al. Machine learning–assisted prediction of pneumonia based on symptoms and routine data. *Front Public Health*. 2022;10:938801. https://www.frontiersin.org/articles/10.3389/fpubh.2022.938801/full.

Kang SJ, et al. Machine Learning for CAP diagnosis using only clinical and laboratory data. *Comput Biol Med*. 2024;169:108100. https://www.sciencedirect.com/science/article/pii/S2213716524004090.

Pan J, et al. Mortality prediction in severe CAP in ICU using ML. *Sci Rep*. 2025; https://www.nature.com/articles/s41598-025-85951-x.pdf.

Liu X, et al. ML model for mortality prediction in severe pneumonia using accessible clinical data. *BMJ Open Respir Res*. 2025;12:e001983. https://bmjopenrespres.bmj.com/content/12/1/e001983.

Delaney BC, et al. Applying ML to EHR to predict CAP after RTI consultations in primary care. *J Biomed Inform*. 2022;128:104028. https://www.sciencedirect.com/science/article/pii/S0895435622000154.

Rajpurkar P, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-rays with Deep Learning. arXiv:1711.05225 (2017). https://arxiv.org/abs/1711.05225.

Khan A, et al. Deep learning for pneumonia detection in CXR: a review of 2012–2023. *J Imaging*. 2024;10(8):176. https://www.mdpi.com/2313-433X/10/8/176.

Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003;3:1157–1182. https://jmlr.org/papers/volume3/guyon03a/guyon03a.pdf.

Chandrashekar G, Sahin F. A survey on feature selection methods. *Computers & Electrical Engineering*. 2014;40(1):16–28. https://www.sciencedirect.com/science/article/pii/S0045790613003066.

Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics*. 2007;23(19):2507–2517. https://academic.oup.com/bioinformatics/article/23/19/2507/185254.

Tibshirani R. Regression shrinkage and selection via the LASSO. *J R Stat Soc Ser B*. 1996;58(1):267–288.

Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5–32.

Kohavi R, John GH. Wrappers for feature subset selection. *Artificial Intelligence*. 1997;97(1–2):273–324. https://www.sciencedirect.com/science/article/pii/S000437029700043X.

Xue B, Zhang M, Browne WN. A survey on evolutionary computation approaches to feature selection. *Pattern Recognit.* 2016;53:121–143.

Sayed GI, Hassanien AE, et al. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing.* 2022;509:99–126. https://www.sciencedirect.com/science/article/pii/S092523122200474X.

Holland JH. *Adaptation in Natural and Artificial Systems.* University of Michigan Press; 1975.

Kennedy J, Eberhart R. Particle swarm optimization. In: *Proc. IEEE Int. Conf. Neural Networks.* 1995:1942–1948.

Mirjalili S, et al. Grey Wolf Optimizer. *Advances in Engineering Software.* 2014;69:46–61.

Heidari AA, et al. Harris Hawks Optimization. *Future Gener Comput Syst.* 2019;97:849–872.

Li S, et al. Slime mould algorithm. *Future Gener Comput Syst.* 2020;111:300–323.

Wazirali R, et al. A systematic review on metaheuristic optimization techniques for disease diagnosis. *Arch Comput Methods Eng.* 2023;30:1–34. https://link.springer.com/article/10.1007/s11831-022-09853-1.

Alrefai N, Ibrahim O. PSO-based feature selection with ensemble learning for cancer microarray classification. *Neural Comput Appl.* 2022;34:13513–13528. https://link.springer.com/article/10.1007/s00521-022-07147-y.

Milano M, et al. A better balance in metaheuristic algorithms: Does it exist? *Appl Soft Comput.* 2020;93:106381. https://www.sciencedirect.com/science/article/pii/S2210650219304080.

Ibrahimov M, et al. Why tuning the control parameters of metaheuristic algorithms is so important. *MENDEL.* 2014;20(1):7–14. https://mendel-journal.org/index.php/mendel/article/download/120/141/.

Wolpert DH, Macready WG. No Free Lunch Theorems for Optimization. *IEEE Trans Evol Comput.* 1997;1(1):67–82. https://www.cs.ubc.ca/~hutter/earg/papers07/00585893.pdf.

van den Bergh F, Engelbrecht AP. A cooperative approach to particle swarm optimization. *IEEE Trans Evol Comput.* 2004;8(3):225–239. https://phoenixwilliams.github.io/PersonalWebsite/LargeScaleEA/A_Cooperative_approach_to_particle_swarm_optimization.pdf.

Liang JJ, Qin AK, Suganthan PN, Baskar S. Comprehensive Learning PSO. *IEEE Trans Evol Comput.* 2006;10(3):281–295. https://tiezhongyu2005.github.io/resources/popularization/CLPSO_2006.pdf.

Zhang Y, et al. A dynamic neighborhood learning based PSO (DNLPSO). *Inf Sci.* 2012;190:19–34. https://www.sciencedirect.com/science/article/pii/S0020025512002927.

Tizhoosh HR. Opposition-Based Learning: A new scheme for machine intelligence. CIMCA 2005. https://www.researchgate.net/publication/4242497_Opposition-Based_Learning_A_New_Scheme_for_Machine_Intelligence.

Rahnamayan S, Tizhoosh HR, Salama MMA. Opposition-Based Differential Evolution. *IEEE Trans Evol Comput.* 2008;12(1):64–79. https://www.researchgate.net/publication/3419015_Opposition-based_differential_evolution_IEEE_Trans_Evol_Comput.

Mirjalili S, Lewis A. S-shaped vs V-shaped transfer functions for binary PSO. *Swarm Evol Comput.* 2013;9:1–14. https://www.sciencedirect.com/science/article/pii/S2210650212000648.