Wrapper Feature Selection for Pneumonia Diagnosis from Routine Blood Markers using MGSPLO

1 Introduction

Pneumonia is a common and serious infection worldwide. It causes substantial illness and death and places a heavy burden on health systems [1]. Community-acquired pneumonia (CAP) ranges from mild outpatient disease to severe illness with respiratory failure and sepsis that requires intensive care [2, 3].

In routine practice, diagnosis combines symptoms and signs with chest imaging and basic laboratory tests [3, 4]. Chest radiographs may be non-diagnostic early in the course of disease [5]. In such cases a CT scan can reveal pneumonia even when the concurrent radiograph is negative [6, 7]. These gaps motivate data-driven tools that work well with information available at the bedside.

Host-response blood biomarkers are widely used in suspected pneumonia. Typical examples are complete blood count (CBC) indices and acute-phase reactants such as C-reactive protein (CRP) and procalcitonin (PCT) [8, 9]. Higher leukocyte counts with neutrophil predominance and elevated CRP have been linked to disease severity and outcomes [9, 10]. PCT-guided strategies can also support antibiotic stewardship in acute respiratory infections, including CAP [11, 12].

Ratios derived from the CBC summarize inflammatory imbalance. The neutrophil-to-lymphocyte ratio (NLR) has shown prognostic value in hospitalized CAP [13, 14]. Combining NLR with clinical scores can further improve risk stratification in older adults [15, 16]. Related ratios such as the platelet-to-lymphocyte ratio (PLR) and monocyte-to-lymphocyte ratio (MLR) also correlate with severity indices and adverse outcomes [17, 18].

At the same time, machine-learning (ML) models that integrate multiple clinical and laboratory variables show strong discrimination for diagnosis and prognosis in CAP [19, 20]. Some recent studies report gains for outcomes such as mortality and length of stay using only routinely collected features [21, 22]. By contrast, a large ML literature still emphasizes imaging-derived features from chest radiography or CT [23, 24].

From a data perspective, even a modest blood panel becomes high dimensional once derived ratios and simple interactions are included. Training directly on such expanded sets increases overfitting risk and reduces interpretability, especially with limited sample sizes. Feature selection (FS) aims to find a compact, informative subset that preserves or improves performance and remains clinically interpretable [25, 26]. FS methods are commonly grouped as filters, embedded methods, and wrappers [27, 28]. Wrappers evaluate candidate subsets using a downstream classifier and naturally capture interactions, though at higher computational cost [29]. The search space is combinatorial: with D candidates there are $2^D - 1$ nonempty subsets [30].

Because exhaustive wrapper search is infeasible, metaheuristic optimization is popular in FS. These algorithms are derivative-free and designed to escape local optima in high-dimensional landscapes [31]. Classic examples include Particle Swarm Optimization (PSO) and Differential Evolution (DE) [32, 33]. Newer nature-inspired methods—such as Grey Wolf Optimizer (GWO) and Harris Hawks Optimization (HHO)—are also widely used [34, 35]. The Slime Mould Algorithm is another recent approach [36]. Despite strong empirical performance, metaheuristics

can exhibit premature convergence, an unstable exploration–exploitation balance, and hyperparameter sensitivity [37]. No-free-lunch results further underline that careful design and tuning are problem dependent [38].

Many enhanced or hybrid strategies have been proposed to improve convergence and solution quality. Examples include comprehensive-learning/topology variants of PSO and opposition-based learning [39, 40]. Local search and differential-evolution hybrids can also help maintain diversity and refine elites [41**?**]. For binary FS, transfer functions provide a principled way to convert continuous updates into probabilistic inclusion decisions [42].

Our approach. We propose a **Multi-Guide Slime-Predator Learning Optimizer (MGSPLO)** for wrapper FS over routine blood and inflammatory markers. MGSPLO is designed to (i) use *multi-anchor guidance* so each candidate learns from several strong subsets, (ii) preserve population diversity via adaptive exploration inspired by predator–prey dynamics and slime-mould-like oscillations, and (iii) apply a tailored *binary mapping* for fine-grained, probability-driven feature inclusion. In our pipeline, MGSPLO wraps a Support Vector Machine (SVM) and uses cross-validated performance as the fitness signal [43, 44].

Study setting and goals. We analyze adults evaluated for pneumonia, with age, comorbidities, CBC (including differentials and red cell/platelet indices), and CRP measured at presentation. From these we derive composite inflammation ratios (e.g., NLR, PLR) to build a feature pool that reflects clinical practice and biomarker evidence. Our goals are threefold. First, we quantify the diagnostic value of routine blood and inflammatory markers—alone and in combination—for pneumonia detection. Second, we test whether MGSPLO-driven wrapper FS can identify small, stable, interpretable subsets that match or improve SVM performance versus using all markers or standard FS baselines. Third, we relate the selected markers to current biomarker literature.

Contributions.

- We introduce **MGSPLO**, a multi-guide, diversity-preserving metaheuristic tailored to wrapper FS in high-dimensional clinical feature spaces.
- We provide a **binary** MGSPLO variant with a refined transfer mechanism enabling granular, probability-driven inclusion of blood routine and inflammatory markers.
- We build an end-to-end **SVM** CAP diagnosis pipeline that integrates MGSPLO with cross-validated evaluation and compare against standard FS techniques and alternative wrapper metaheuristics.
- On a real-world cohort, **MGSPLO+SVM** achieves competitive or better accuracy while selecting compact subsets of routine markers, supporting use where imaging resources are limited.

References

[1] GBD 2019 Lower Respiratory Infections Collaborators. Age–sex differences in the global burden of lower respiratory infections and risk factors, 1990–2019: results from the global burden of disease study 2019. *Lancet Infectious Diseases*, 22(11):1626–1647, 2022. doi: 10.1016/S1473-3099(22)00510-2.

[2] David M. Musher and Adam R. Thorner. Community-acquired pneumonia. *JAMA*, 311 (20):2197–2208, 2014. doi: 10.1001/jama.2014.758.

[3] Antoni Torres, Catia Cillóniz, Michael S. Niederman, Rebeca Menéndez, James D. Chalmers, Richard G. Wunderink, and Tom van der Poll. Community-acquired pneumonia. *The Lancet*, 398(10303):906–919, 2021. doi: 10.1016/S0140-6736(21)00904-0.

[4] Tomás Franquet. Imaging of community-acquired pneumonia. *Journal of Thoracic Imaging*, 33(5):W1–W12, 2018. doi: 10.1097/RTI.0000000000000374.

[5] Fumitaka Okada, Akira Ono, Yasunari Ando, et al. Community-acquired pneumonia with negative chest radiography findings: Clinical and radiological characteristics. *Respiration*, 97(6):508–516, 2018. doi: 10.1159/000489488.

[6] Wesley H. Self, Robert A. Balk, Carlos G. Grijalva, et al. Community-acquired pneumonia visualized on ct scans but not on chest radiographs: Pathogens, severity, and clinical outcomes. *Chest*, 153(3):601–610, 2018. doi: 10.1016/j.chest.2017.07.035.

[7] Yann-Erick Claessens, Marie-Pierre Debray, Florence Tubach, et al. Early chest computed tomography scan to assist diagnosis and guide treatment decision for suspected community-acquired pneumonia. *American Journal of Respiratory and Critical Care Medicine*, 192(8): 974–982, 2015. doi: 10.1164/rccm.201501-0017OC.

[8] Benedetta Casadei, Riccardo Biondi, et al. Diagnostic and prognostic roles of procalcitonin and other tools in community-acquired pneumonia and lower respiratory tract infections. *Diagnostics*, 13(11):1869, 2023. doi: 10.3390/diagnostics13111869.

[9] Raquel Menéndez, Rosana Martínez, Santiago Reyes, et al. C-reactive protein is an independent predictor of severity in community-acquired pneumonia. *American Journal of Medicine*, 121(3):219–226, 2008. doi: 10.1016/j.amjmed.2007.10.033.

[10] Otavio T. Ranzani et al. C-reactive protein as a predictor of survival and length of hospital stay in community-acquired pneumonia. *Journal of Personalized Medicine*, 12(10):1710, 2022. doi: 10.3390/jpm12101710.

[11] Philipp Schuetz, Matthias Briel, Mirjam Christ-Crain, et al. Procalcitonin to guide initiation and duration of antibiotic treatment in acute respiratory infections: An individual patient data meta-analysis. *Clinical Infectious Diseases*, 55(5):651–662, 2012. doi: 10.1093/cid/cis464.

[12] Philipp Schuetz, Yannick Wirz, Reto Sager, et al. Procalcitonin to initiate or discontinue antibiotics in acute respiratory tract infections. *Cochrane Database of Systematic Reviews*, (10):CD007498, 2017. doi: 10.1002/14651858.CD007498.pub3.

[13] Ji-Hyun Lee et al. Prognostic value of serial neutrophil-to-lymphocyte ratio in hospitalized community-acquired pneumonia. *PLOS ONE*, 16(4):e0250067, 2021. doi: 10.1371/journal.pone.0250067.

[14] Ioana A. Colosi et al. The neutrophil/lymphocyte ratio and outcomes in hospitalized community-acquired pneumonia. *Biomedicines*, 12(2):260, 2024. doi: 10.3390/biomedicines12020260.

[15] De-Yun Feng, Xiang-Li Zou, Yu-Qin Zhou, et al. Combined neutrophil-to-lymphocyte ratio and curb-65 score as an accurate predictor of mortality for community-acquired pneumonia in the elderly. *International Journal of General Medicine*, 14:1133–1139, 2021. doi: 10.2147/IJGM.S300776.

[16] Lin Huang, Bo Weng, Min Wang, et al. The improved prediction value of neutrophil-to-lymphocyte ratio to pneumonia severity scores for mortality in older people with community-acquired pneumonia. *BMC Geriatrics*, 25:485, 2025. doi: 10.1186/s12877-025-06121-2.

[17] Adam Nowiński et al. Correlation of psi and curb-65 scores with neutrophil/lymphocyte, platelet/lymphocyte, and monocyte/lymphocyte ratios in predicting in-hospital mortality for community-acquired pneumonia. *Journal of Clinical Medicine*, 14(3):728, 2024. doi: 10.3390/jcm14030728.

[18] C. C. E. Enersen et al. Neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios and mortality in community-acquired pneumonia: A derivation–validation study. *Infection*, 51: 1339–1347, 2023. doi: 10.1007/s15010-023-01992-2.

[19] Seok-Jin Kang et al. Machine learning for community-acquired pneumonia diagnosis using only clinical and laboratory data. *Computers in Biology and Medicine*, 169:108100, 2024. doi: 10.1016/j.compbiomed.2024.108100.

[20] Cheng Liu et al. Performance of machine learning algorithms for predicting adverse outcomes in community-acquired pneumonia. *Frontiers in Bioengineering and Biotechnology*, 10:903426, 2022. doi: 10.3389/fbioe.2022.903426.

[21] Mustafa Aka et al. Employing a low-code machine learning approach to predict in-hospital mortality and length of stay in community-acquired pneumonia. *Scientific Reports*, 14: 82615, 2024. doi: 10.1038/s41598-024-82615-0.

[22] Xin Liu et al. Machine learning-based model for predicting all-cause mortality in patients with severe pneumonia using accessible clinical and laboratory data. *BMJ Open Respiratory Research*, 12(1):e001983, 2025. doi: 10.1136/bmjresp-2024-001983.

[23] Daniel S. Kermany et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131, 2018. doi: 10.1016/j.cell.2018.02.010.

[24] A. Khan et al. Deep learning for pneumonia detection in chest x-rays: A review of 2012–2023. *Journal of Imaging*, 10(8):176, 2024. doi: 10.3390/jimaging10080176.

[25] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[26] G. Chandrashekar and F. Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014. doi: 10.1016/j.compeleceng.2013.11.024.

[27] Roberto Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4):537–550, 1994. doi: 10.1109/72.298224.

[28] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1):267–288, 1996.

[29] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2):273–324, 1997. doi: 10.1016/S0004-3702(97)00043-X.

[30] Avrim Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2):245–271, 1997. doi: 10.1016/S0004-3702(97)00063-5.

[31] Bing Xue, Mengjie Zhang, Will N. Browne, and Xin Yao. A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4):606–626, 2016. doi: 10.1109/TEVC.2015.2504420.

[32] James Kennedy and Russell C. Eberhart. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks*, pages 1942–1948, 1995. doi: 10.1109/ICNN.1995.488968.

[33] Rainer Storn and Kenneth Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997. doi: 10.1023/A:1008202821328.

[34] Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis. Grey wolf optimizer. *Advances in Engineering Software*, 69:46–61, 2014. doi: 10.1016/j.advengsoft.2013.12.007.

[35] Ali Asghar Heidari, Seyedali Mirjalili, Hossam Faris, Ibrahim Aljarah, Majdi Mafarja, and Hui Chen. Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, 97:849–872, 2019. doi: 10.1016/j.future.2019.02.028.

[36] Shuang Li, Hu Chen, Mengjie Wang, Ali Asghar Heidari, and Seyedali Mirjalili. Slime mould algorithm: A new method for stochastic optimization. *Future Generation Computer Systems*, 111:300–323, 2020. doi: 10.1016/j.future.2020.03.055.

[37] Ismail Boussaïd, Julien Lepagnot, and Patrick Siarry. A survey on optimization metaheuristics. *Information Sciences*, 237:82–117, 2013. doi: 10.1016/j.ins.2013.02.041.

[38] David H. Wolpert and William G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997. doi: 10.1109/4235.585893.

[39] Jing J. Liang, Anil Kumar Qin, Ponnuthurai N. Suganthan, and Subramanian Baskar. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *IEEE Transactions on Evolutionary Computation*, 10(3):281–295, 2006. doi: 10.1109/TEVC.2005.857610.

[40] Hamid R. Tizhoosh. Opposition-based learning: A new scheme for machine intelligence. In *Proc. Int. Conf. Computational Intelligence for Modelling, Control and Automation (CIMCA)*, pages 695–701, 2005. doi: 10.1109/CIMCA.2005.1631345.

[41] Saeid Rahnamayan, Hamid R. Tizhoosh, and Magdy M. A. Salama. Opposition-based differential evolution. *IEEE Transactions on Evolutionary Computation*, 12(1):64–79, 2008. doi: 10.1109/TEVC.2007.894200.

[42] Seyedali Mirjalili and Andrew Lewis. S-shaped vs v-shaped transfer functions for binary particle swarm optimization. *Swarm and Evolutionary Computation*, 9:1–14, 2013. doi: 10.1016/j.swevo.2012.09.002.

[43] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, 1995. doi: 10.1007/BF00994018.

[44] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of IJCAI*, pages 1137–1145, 1995.