Department of Computer Science
University of Science and Technology of Oran Mohamed Boudiaf (USTO-MB)
Oran, Algeria
Email: amine.bellatreche@univ-usto.dz

November 24, 2025

Editor-in-Chief
Journal Name
Journal Address

Dear Editor,

We are pleased to submit our manuscript titled **"Dholes-Inspired Optimization for Simultaneous Feature Selection and Hyperparameter Tuning: Cross-Domain Validation from Medical Diagnostics to Computer Vision"** for consideration for publication in [Journal Name].

*Research Significance

This work makes four major contributions to the metaheuristic optimization and machine learning communities:

*1. Discovery of Algorithm-Dependent Optimization Overfitting

We identify and characterize a previously under-documented phenomenon: *optimization overfitting*—where hyperparameters optimized on a single train/test split achieve perfect accuracy (100%) on that partition but generalize poorly across different data splits (94.72% average). Critically, we demonstrate that this phenomenon is **algorithm-dependent**:

- **Random Forest:** Requires expensive cross-validation within the optimization loop (7.9 hours) to achieve robust performance (96.26%, rank #3)
- **XGBoost:** Built-in regularization eliminates need for CV-based optimization entirely—single-split optimization achieves best results (96.34%, rank #1) in only 54 seconds (526× faster)

This finding has immediate implications for the hyperparameter optimization community, suggesting that algorithm selection should precede methodology selection.

*2. Rigorous Statistical Validation with 30-Run Protocol

Unlike many metaheuristic studies that report single-run results, we conducted **30 independent experimental runs** with different random seeds (42-71) for breast cancer classification. Each run employed stratified train/test splits, enabling:

- Wilcoxon signed-rank tests for paired statistical comparison (p-values: 0.0084**, ¡0.001***)
- Confidence intervals and variance analysis (standard deviations: 1.23-1.43%)
- Reproducibility assurance through fixed seeds
- Robust performance estimates across diverse data partitions

Our CV-optimized Random Forest significantly outperformed defaults using the same 6 features (p=0.0084), confirming that proper optimization methodology successfully avoids overfitting.

*3. Cross-Domain Generalizability Validation (68× Dimensional Scale-Up)

We validate DIO's effectiveness across two fundamentally different domains:

**Medical Diagnostics (Breast Cancer):**

- 30-D tabular features, binary classification, 569 samples
- **Best Result:** DIO-XGBoost achieved 96.34% accuracy with 43% feature reduction (17/30) in 54 seconds—rank #1 overall
- **Maximum Interpretability:** DIO-RF-CV achieved 96.26% accuracy with 80% reduction (6/30 features)—ideal for resource-constrained clinical settings

**Computer Vision (CIFAR-10):**

- 2048-D ResNet50 deep features, 10-class classification, 60,000 images
- **Proof-of-Concept:** DIO-XGBoost achieved 83.6% accuracy (+2.8% over 80.8% baseline) with 58.35% feature reduction (2048→853) on 2,000-sample subset
- 2.4× inference speedup enables edge deployment on IoT devices

The consistent improvement patterns across 68× dimensional scale-up (30-D → 2048-D) and different problem characteristics (binary → 10-class) demonstrate genuine framework robustness, not dataset-specific tuning.

*4. Deployment-Ready Configurations for Clinical Practice

We provide three validated models representing the Pareto frontier, each optimized for different clinical contexts:

- **Maximum Accuracy:** DIO-XGBoost (96.34%, 17 features, 54-sec optimization)—high-stakes diagnosis
- **Maximum Interpretability:** DIO-RF-CV (96.26%, 6 features, 7.9-hour optimization)—rural clinics, point-of-care testing
- **Rapid Prototyping:** DIO-RF Single-Split (94.72%, 8 features, 1-min optimization)—research, non-critical screening

The 80% feature reduction (30→6) translates to 5× faster inference, 80% reduction in laboratory costs, and substantially improved clinical interpretability.

*Methodological Rigor

Our study distinguishes itself through:

- **Complete Python Implementation:** DIO reimplemented from MATLAB specification and validated on 14 benchmark functions (convergence: $7.6 \times 10^{-26}$ on F1)
- **Nested Optimization Framework:** Simultaneous hyperparameter tuning and feature selection (vs. sequential approaches)
- **Transparent Reporting:** Full disclosure of both successes (96.34% accuracy) and limitations (optimization overfitting in RF single-split)
- **Open-Source Commitment:** Code, datasets, and reproduction instructions will be made publicly available upon acceptance

*Why This Journal?

[Journal Name]'s focus on [specific focus areas—e.g., "metaheuristic optimization, machine learning applications, and computational intelligence"] aligns perfectly with our work's emphasis on rigorous validation of nature-inspired algorithms across real-world domains. Our discovery of algorithm-dependent optimization overfitting and cross-domain validation framework directly addresses the journal's mission to advance both theoretical understanding and practical deployment of intelligent optimization systems.

The manuscript's 60 pages include:

- 24 figures (14 schemas, code snippets, statistical visualizations)
- 8 tables (performance summaries, statistical tests, comparisons)
- Comprehensive appendices (pseudocode, selected features, hyperparameters)
- 9 references to seminal works (Breiman 2001, Chen & Guestrin 2016, Dehghani et al. 2023)

*Competing Interests

We declare no competing interests. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

*Author Contributions

- **Mohamed Amine Bellatreche:** Conceptualization, Methodology, Software Implementation (DIO algorithm, nested optimization framework), Formal Analysis (statistical validation, 30-run protocol), Investigation (breast cancer and CIFAR-10 experiments), Writing—Original Draft, Visualization (all figures and tables)
- **Ghizlane Cherif:** Methodology (fitness function design, CV-based optimization strategy), Validation (cross-domain experiments), Resources (dataset curation, ResNet50 feature extraction), Writing—Review & Editing, Supervision

*Data Availability

All datasets used in this study are publicly available:

- **Breast Cancer Wisconsin (Diagnostic):** UCI Machine Learning Repository (`https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)`)
- **CIFAR-10:** University of Toronto (`https://www.cs.toronto.edu/~kriz/cifar.html`)
- **ResNet50 Pre-trained Weights:** TensorFlow/Keras (ImageNet)

Python code for DIO implementation, nested optimization framework, and all experiments will be made available on GitHub upon acceptance: `https://github.com/amine-dubs/dio-optimization`

*Suggested Reviewers

We suggest the following experts whose research intersects with our work:

1. **Dr. [Name 1]**, [University/Institution]
   Email: [email]
   Expertise: Metaheuristic optimization, nature-inspired algorithms
2. **Dr. [Name 2]**, [University/Institution]
   Email: [email]
   Expertise: Feature selection, hyperparameter optimization
3. **Dr. [Name 3]**, [University/Institution]
   Email: [email]

Expertise: Medical machine learning, breast cancer classification

4. **Dr. [Name 4]**, [University/Institution]
   Email: [email]
   Expertise: Transfer learning, computer vision optimization

*Note: We have no professional or personal relationships with any of the suggested reviewers that could be perceived as conflicts of interest.*

*Conclusion

This manuscript presents a rigorously validated, deployment-ready optimization framework with clear practical impact: 96.34% medical diagnostic accuracy using 43% fewer features, achieved in 54 seconds. The discovery that algorithm choice determines whether expensive CV-based optimization is necessary (Random Forest: yes, XGBoost: no) challenges conventional wisdom and provides actionable guidance for practitioners.

We believe this work will be of significant interest to [Journal Name]'s readership, particularly researchers working on metaheuristic optimization, medical AI, and cross-domain machine learning applications. We look forward to your editorial evaluation and welcome constructive feedback from peer reviewers.

Thank you for considering our manuscript for publication.

Sincerely,

Mohamed Amine Bellatreche
Ghizlane Cherif