ELM Performance Using Pre -Trained Deep Neural Network Features- A Comprehensive Literature Review and Analysis

Nilesh Rathod[1], *Shashi Kant Gupta*[32]
[1] Research Scholar Lincoln University College; Adjunct Professor, Faculty of Information Technology, Victorian Institute of Technology, Melbourne VIC 3000, Australia; [2]
drntrathod@lincoln.edu.my, raj2008enator@gmail.com,

Abstract: Extreme Learning Machines (ELMs) represent a significant departure from traditional gradient-based training paradigms for single-hidden layer feedforward networks, offering substantial advantages in terms of learning speed and generalization capability. Despite these benefits, the inherent stochasticity of hidden layer initialization often leads to suboptimal stability and performance inconsistency across complex datasets. This report provides an exhaustive literature review and thematic analysis of feature-driven optimization techniques for ELMs. It systematically explores the fundamental challenges of random initialization, the efficacy of meta-heuristic optimization algorithms—specifically Genetic Algorithms, Particle Swarm Optimization, and Grey Wolf Optimizer—and the transformative role of feature engineering and selection. Furthermore, the analysis evaluates domain-specific applications ranging from biomedical diagnostics to environmental forecasting. By synthesizing contemporary research, this report identifies critical gaps in cross-domain evaluation, the integration of optimization within deep ELM architectures, and the computational overhead of deep feature extraction on lightweight models. The findings provide a roadmap for the development of high-performance, resource-efficient intelligent systems suitable for real-time industrial and clinical applications

Keywords: Extreme Learning Machines (ELMs); Particle Swarm Optimization (PSO); Grey Wolf Optimization (GWO); Deep Neural Network (DNN).

Introduction

The emergence of the Extreme Learning Machine (ELM) has provided a robust alternative to traditional backpropagation-based training for single-hidden layer feedforward neural networks (SLFNs). Proposed as a faster and more efficient learning paradigm, ELM bypasses the iterative gradient descent process by randomly assigning input weights and hidden layer biases, which remain fixed throughout the training cycle [1]. This structural configuration allows for the analytical determination of output weights through a single-step calculation using the Moore-Penrose generalized inverse, transforming the learning problem into a linear least-squares task [2]. The primary motivation behind ELM development was to overcome the slow convergence and susceptibility to local minima inherent in classical feedforward networks [4].
The theoretical foundation of ELM is anchored in the Universal Approximation Theorem, which posits that an SLFN with randomly generated hidden nodes can approximate any continuous function given a sufficient number of neurons [6]. This paradigm has been extended to various learning tasks, including

classification, regression, and unsupervised representation learning.3 Evidence indicates that ELM often achieves an order-of-magnitude faster training process compared to deep neural networks (DNNs) while maintaining competitive generalization performance.6 These characteristics make ELM highly desirable for big data analytics, real-time diagnostic systems, and deployments on resource-constrained hardware such as Field Programmable Gate Arrays (FPGAs) [3].

However, the "randomness nature" that grants ELM its speed also introduces significant performance sensitivity. The model's reliability is heavily dependent on the quality of the initial hidden layer parameters and the choice of activation functions.1 In many instances, the random initialization results in redundant nodes or insufficient feature mapping, leading to suboptimal generalization on high-dimensional or noisy datasets [2]. To address these deficiencies, a recent research shift has focused on "feature-driven optimization," a framework that integrates meta-heuristic search algorithms and advanced feature engineering to refine both the internal model structure and the input data space [1].

This paper presents a detailed analysis of optimization strategies designed to enhance ELM performance. It begins with an evaluation of fundamental stability challenges and proceeds to categorize optimization techniques into meta-heuristic parameter tuning and feature-driven engineering. The report then explores domain-specific applications, highlighting the real-world impact of optimized ELMs, and concludes by identifying persistent research gaps as of 2024–2025

THEMATIC ANALYSIS OF META-HEURISTIC OPTIMIZATION TECHNIQUES

Meta-heuristic optimization algorithms serve as the primary mechanism for replacing random initialization with an informed search process. These algorithms treat the selection of input weights and biases as a global optimization problem, aiming to minimize the validation error or structural risk.[1]

A. Genetic Algorithms (GA) and Evolutionary Optimization

Genetic Algorithms utilize selection, crossover, and mutation to evolve a population of candidate ELM parameters.

1. Parameter Evolution: The Genetic Ensemble of ELM (GE-ELM) produces a set of ELM candidates and uses GA to identify ideal parameters, decrementing the negative effects of non-ideal random states. GE-ELM has demonstrated superior robustness on benchmark classification and regression tasks.
2. Pruning and Architecture Optimization: The GAP-ELM method uses GA to prune redundant neurons in the hidden layers, resulting in a more compact network without sacrificing accuracy.
3. Multi-Objective GA: In COVID-19 detection and Parkinson’s diagnosis, GA has been used to optimize both the hidden node count and the activation function parameters simultaneously, achieving accuracies as high as 96.81% in expert diagnosis systems

B. Particle Swarm Optimization (PSO) and Variants

PSO is favored in the ELM community for its rapid convergence and simple implementation in continuous search spaces.

1. Global Minima Search: PSO hybrids with ELM are utilized to search for global minima, overcoming the limitations of standard adaptive growth models.1 In power transformer fault diagnosis, PSO-optimized KELM (Kernel ELM) was found to be more stable and had faster learning speeds than traditional SVMs.

2. Feature Selection and Memory Effects: Fractional-Order Darwinian PSO (FODPSO) has been integrated with ELM to address high-dimensional redundancy. The introduction of "memory effects" in particle movement allows for a wider exploration of the search space, preventing premature convergence toward local optima.
3. Time-Series Forecasting: Switching delayed PSO-optimized ELMs are effective for short-term load forecasting, where they capture temporal fluctuations more reliably than standalone radial basis function networks.

C. Grey Wolf Optimizer (GWO) and Specialized Algorithms

GWO mimics the hunting behavior and leadership hierarchy of grey wolves, offering a balanced approach between exploration (searching for new regions) and exploitation (refining existing solutions).

1. Medical Diagnostic Precision: GWO-ELM models have shown significant efficacy in detecting Diabetic Retinopathy. By iteratively adjusting hidden weights to match the positions of the "alpha," "beta," and "delta" wolves in the solution space, the model avoids local minima and provides a highly reproducible diagnostic tool.
2. Deep Hybrid Kernel Optimization: As of 2024–2025, modern variants like the Improved Snake Optimization (ISO) algorithm are used to tune Deep Hybrid Kernel ELMs (DHKELM). ISO utilizes Latin hypercube sampling and Cauchy mutation strategies to enhance the expressive ability of the model, achieving R^2 values above 0.99 in environmental disaster prediction.
3. Sparrow Search and Zebra Optimization: These nature-inspired algorithms have been recently applied to Deep ELM (DELM) architectures. For example, the Improved Zebra Optimization Algorithm (IZOA) has been used to optimize color change predictions in materials engineering, reducing Mean Absolute Error (MAE) by over 56% compared to standard models.

Table 1. Comparative Table of Meta-heuristic ELM Optimization

Dataset	Algorithm	Hidden Nodes	Training Time	Accuracy/Error	Reference
Parkinson	GA-WK-ELM	22	0.21 μs	96.81% (Acc)	1
Diabetes	KELM-PSO	8	732.05 s	83.85% (Acc)	1
PV Power	ISSA-ELM	NA	Fast	$R^2 > 0.99$	16
Bearing Fault	SSA-DELM	75	Fast	98.67% (Acc)	17
Image Seg.	GE-ELM	19	243.43 s	95.87% (Acc)	1
Wood Color	IZOA-DHKELM	NA	Moderate	67.4% (RMSE Red.)	15

COMPARATIVE SUMMARY AND RESEARCH GAPS

As the field of Extreme Learning Machines matures, researchers have identified several fundamental trade-offs and persistent deficiencies in existing literature that guide contemporary studies.

A. Cross-Domain Evaluation Gap

A critical gap identified in the research is the "lack of testing across multiple data types" and domains.1 Most ELM optimization frameworks are validated on specific, isolated benchmark datasets (e.g., MNIST for images, ADNI for medical). However, these models often struggle with "domain shift" or "calibration drift" when applied to real-world scenarios where the data distribution changes over time.

1. Domain Shift Resilience: Recent cross-domain evaluations in video anomaly detection show that accuracy can drop by as much as 15% when moving from one surveillance environment to another.
2. Proposed Solutions: Research as of 2025 is investigating lightweight transformer-based domain adapters with Maximum Mean Discrepancy (MMD) loss to align source and target feature distributions before ELM classification.

B. Optimization Integration in Deep ELM Models

While shallow ELM models are frequently optimized using meta-heuristics, there is a distinct "lack of optimization in deep ELM models".1 Deep ELMs (DELM) and Multi-Layer ELMs (ML-ELM) typically rely on greedy unsupervised pre-training of layers. The random factor selection in these deep hierarchies remains a source of unsatisfactory reliability.30

1. Synergy Challenges: Integrating meta-heuristic algorithms into deep structures is computationally expensive, as each fitness evaluation requires a full layer-by-layer pre-training cycle.
2. The ACHOA Model: To address this, researchers have proposed the Accumulated Chimp Optimization Algorithm (ACHOA) for DELM, which provides more robust convergence and optimization stability in sports analytics and real-time COVID-19 diagnosis.

C. Computational Efficiency and Overhead

Despite the fast learning speed of the ELM classifier, the "overhead of deep feature extraction on lightweight models" presents a significant hurdle for real-time edge computing.

1. Numerical Overhead: Quantitative comparisons of ELM-DCNN hybrid models reveal a substantial training time increase when meta-heuristic optimization is added. The following table highlights the efficiency-accuracy trade-off observed in recent 2025 benchmarks.
2. Complexity Management: The transition from "high-accuracy" to "low-cost" regimes requires strategic node selection and the use of backpropagation-free frameworks like "ELM-DeepONets" for operator learning, which achieve superior accuracy while drastically reducing training costs in scientific computing.

Conclusions

The investigation into "Feature-Driven Optimization of Extreme Learning Machines" demonstrates that while the fundamental ELM architecture provides a transformative paradigm for rapid machine learning, its reliance on random initialization necessitates sophisticated optimization to achieve industrial-grade reliability. The thematic analysis of meta-heuristic algorithms—such as GA, PSO, and GWO—reveals their critical role in transforming ELM into a stable, high-performance tool across biomedical, environmental, and industrial domains. Furthermore, the integration of deep learning feature extractors and autoencoders has allowed ELM to bridge the gap between shallow classifiers and deep hierarchical models.

Despite these advances, the challenges of cross-domain generalization, the computational overhead of deep feature extraction, and the limited optimization of deep architectures remain significant hurdles. Addressing these gaps requires a concerted effort toward domain-adaptive frameworks and more efficient meta-heuristic engines that respect the real-time constraints of modern IoT and clinical systems. As researchers continue to refine the balance between high-speed training and improved predictive reliability, the optimized ELM is poised to remain a dominant technology for the next generation of scalable and intelligent analytics.

References

1. A Fast and Effective Extreme Learning Machine Algorithm Without Tuning - CMAP, accessed on January 12, 2026, http://www.cmap.polytechnique.fr/~nikolaus.hansen/proceedings/2014/WCCI/IJCNN-2014/PROGRAM/N-14064.pdf
2. Extreme Learning Machines - CORE, accessed on January 12, 2026, https://core.ac.uk/download/pdf/622196093.pdf
3. Improving Classification Performance through an Advanced ..., accessed on January 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC5435980/
4. PCA-ELM: A Robust and Pruned Extreme Learning Machine Approach Based on Principal Component Analysis | Request PDF - ResearchGate, accessed on January 12, 2026, https://www.researchgate.net/publication/257631823_PCA-ELM_A_Robust_and_Pruned_Extreme_Learning_Machine_Approach_Based_on_Principal_Component_Analysis
5. Fast Learning in Quantitative Finance with Extreme Learning Machine - arXiv, accessed on January 12, 2026, https://arxiv.org/html/2505.09551v1
6. ELM-DeepONets: Backpropagation-Free Training of Deep Operator Networks via Extreme Learning Machines - IEEE Xplore, accessed on January 12, 2026, https://ieeexplore.ieee.org/iel8/6287639/10820123/11005528.pdf
7. Learning deep representations via extreme learning machines | Request PDF, accessed on January 12, 2026, https://www.researchgate.net/publication/275590644_Learning_deep_representations_via_extreme_learning_machines
8. The Study on Initialization Aspects of the Extreme Learning Machine Parameters by Random Values - ResearchGate, accessed on January 12, 2026, https://www.researchgate.net/publication/387481623_The_Study_on_Initialization_Aspects_of_the_Extreme_Learning_Machine_Parameters_by_Random_Values
9. A Critical Analysis of the Theoretical Framework of the Extreme Learning Machine - arXiv, accessed on January 12, 2026, https://arxiv.org/html/2406.17427v1
10. Research on Transmission Line Icing Prediction for Power System Based on Improved Snake Optimization Algorithm-Optimized Deep Hybrid Kernel Extreme Learning Machine - MDPI, accessed on January 12, 2026, https://www.mdpi.com/1996-1073/18/17/4646

11. Prediction of Color Change in Heat-Treated Wood Based on Improved Zebra Algorithm Optimized Deep Hybrid Kernel Extreme Learning Machine Model (IZOA-DHKELM) - MDPI, accessed on January 12, 2026, https://www.mdpi.com/1999-4907/16/2/253
12. A Novel Method to Predict Laying Rate Based on Multiple Environment Variables, accessed on January 12, 2026, https://www.researchgate.net/publication/354113522_A_Novel_Method_to_Predict_Laying_Rate_Based_on_Multiple_Environment_Variables
13. Application of machine learning methods in photovoltaic output power prediction: A review, accessed on January 12, 2026, https://www.researchgate.net/publication/359769080_Application_of_machine_learning_methods_in_photovoltaic_output_power_prediction_A_review
14. (PDF) A Scalable Hybrid Autoencoder–Extreme Learning Machine Framework for Adaptive Intrusion Detection in High-Dimensional Networks - ResearchGate, accessed on January 12, 2026, https://www.researchgate.net/publication/391787027_A_Scalable_Hybrid_Autoencoder-Extreme_Learning_Machine_Framework_for_Adaptive_Intrusion_Detection_in_High-Dimensional_Networks
15. Machine Learning for Sensor Analytics: A Comprehensive Review and Benchmark of Boosting Algorithms in Healthcare, Environmental, and Energy Applications - PubMed Central, accessed on January 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12694449/
16. ResTN: Residual Transfer Network for Cross Domain Network Intrusion Detection, accessed on January 12, 2026, https://www.researchgate.net/publication/394592461_ResTN_Residual_Transfer_Network_for_Cross_Domain_Network_Intrusion_Detection
17. Lightweight CNN–MIL Models for Cross-Domain Video Anomaly Detection - Informatica, accessed on January 12, 2026, https://informatica.si/index.php/informatica/article/view/12037/6069
18. TGDHTL: Hyperspectral Image Classification via Transformer–Graph Convolutional Network–Diffusion with Hybrid Domain Adaptation - MDPI, accessed on January 12, 2026, https://www.mdpi.com/2072-4292/18/2/189
19. Improving women football tactics analysis by using extreme learning ..., accessed on January 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12753685/
20. Meta-heuristics and deep learning for energy applications: Review and open research challenges (2018–2023) - ResearchGate, accessed on January 12, 2026, https://www.researchgate.net/publication/380824429_Meta-heuristics_and_deep_learning_for_energy_applications_Review_and_open_research_challenges_2018-2023
21. ELM-DeepONets: Backpropagation-Free Training of Deep Operator Networks via Extreme Learning Machines - arXiv, accessed on January 12, 2026, https://arxiv.org/html/2501.09395v1
22. A Survey of Cross-domain Graph Learning: Progress and Future Directions - arXiv, accessed on January 12, 2026, https://arxiv.org/html/2503.11086v2
23. Few-shot learning and explainable AI for colon cancer histopathology - Mid Sweden University, accessed on January 12, 2026, https://miun.diva-portal.org/smash/get/diva2:2016433/FULLTEXT01.pdf