A Review on Analyzing Grokking Dynamics in Swarm-Based Metaheuristic Optimization for Behavioural Disorder Classification

Author: Dr. Preeti kamra[1], Prof. (Dr.) Shashi Kant Gupta [2]

[1] PostDoctoral Researcher, Computer Science and Engineering, Lincoln University College, Petaling Jaya, Selangor, Malaysia; [2]Computer Science and Engineering, Lincoln University College, Petaling Jaya, Selangor, Malaysia.

Email ID: pdf.preetikamra@lincoln.edu.my, **Email ID:** shashigupta@lincoln.edu.my

Abstract

The classification of behavioural disorders is difficult, due to smaller dataset of patients, a large feature space dimension, and high clinical variability, which may lead to low model generalization. Although feature selection and hyper parameter tuning are the most popular applications of Swarm Intelligence (SI) algorithms, the majority of current methods stop the optimization process as soon as the optimum has been reached. On the other hand, recent research into gradient-based learning has demonstrated the appearance of a phenomenon they refer to as grokking, where long optimization training can cause late generalization. This review examines whether similar post-convergence behaviour could be relevant to non-gradient, population-based optimization. A literature review was carried out through a structured search of scientific databases from 2004 to 2026. The literature thus selected was examined in terms of optimization strategy, learning behaviour, characteristics of the dataset, and clinical applicability. The review offers four main research gaps: (1) post-convergence analysis missing from swarm intelligence algorithms, (2) insufficient knowledge of grokking in decentralized optimization, (3) lack of research in the field of behavioural healthcare, and (4) no validation performed on noisy clinical datasets. From these observations, a conceptual framework for using grokking as a link between swarm-based optimization is suggested. The review serves as a basis for further research on post-convergence optimization, and could be useful for the development of more stable and reliable feature selection methods for clinical decision supporting systems.

Keywords: Grokking, Swarm Intelligence, Metaheuristic Optimization, Behavioural Disorders, Feature Selection, Delayed Generalization, Clinical Decision Support, Healthcare Artificial Intelligence, Optimization Dynamics.

1. Introduction

Machine learning clinical decision support is difficult, due to small data sets, large numbers of features, and exhibit significant clinical heterogeneity, which can lead to overfitting and lack of generalizability. Various algorithms of Swarm Intelligence (SI) in behavioural healthcare have been suggested in literature, for feature selection and hyper parameter optimization [1]. They can search complex spaces (non-convex) without information on gradients, which makes them appropriate for clinical data. But most of the existing studies consider these algorithms from the view point of convergence time and classification accuracy and optimization should stop when the algorithm is converged. This has been disputed lately with deep learning models improving their generalization ability even after reaching near-perfect training accuracy, known as grokking [2]. This generalization of the previous time step, which has been described in a few gradient-based learning architectures, has never been explored for non-gradient, population-based optimization architectures.

In this review, the possibility of grokking and its possible interaction with swarm-based metaheuristic optimization in the context of classification of behavioural disorders is explored[3]. It summarizes the most important gaps at the intersection of these fields and introduces a conceptual framework for studying optimization after convergence of non-gradient algorithms. The objective of this review is to build a solid background on future research for developing more stable and reliable feature selection techniques in clinical decision-support systems.

2. Literature Review: The Convergence of Two Domains

2.1 Swarm Metaheuristics in Behavioural Healthcare

Swarm Intelligence (SI) metaheuristics have gained popularity in behavioural healthcare because they bypass gradient computations, search non-convex parameter spaces effectively, and resist local noise. However, existing SI-based clinical models prioritize rapid convergence. Table1 highlights common used SI algorithm in mental healthcare.

Table 1. Summary of SI algorithms in mental healthcare

Algorithm	Inspiration	Clinical Application	Dataset	Optimization Target	Advantage	Limitation
PSO (Particle Swarm Optimization) [4]	Bird flocking behaviour	ADHD and ASD detection	EEG, questionnaire data	Feature selection and classifier tuning	Fast convergence and simple implementation	May converge to local optima
GWO (Grey Wolf Optimizer) [5]	Wolf leadership hierarchy	Depression classification	Clinical and behavioural data	Hyper parameter optimization	Balanced exploration and exploitation	Sensitive to parameter settings
SSA (Salp Swarm Algorithm) [6]	Salp chain movement	Mental health forecasting	Temporal health records	Feature optimization	Robust local search capability	Slower convergence in large datasets
WOA (Whale Optimization Algorithm) [7]	Bubble-net whale hunting	EEG signal optimization	EEG and biosignals	Signal feature enhancement	Strong global exploration	Can suffer from instability
ACO (Ant Colony Optimization) [8]	Ant foraging behaviour	Behavioural feature selection	High-dimensional clinical data	Combinatorial optimization	Effective for discrete feature spaces	Computationally expensive
FA (Firefly Algorithm) [9]	Firefly flashing behaviour	Stress detection	Wearable sensor data	Feature subset selection	Handles nonlinear search spaces well	Performance decreases with noise

BA (Bat Algorithm) [10]	Echolocation of bats	Sleep disorder classification	Physiological signals	Neural network tuning	Adaptive frequency search	Risk of premature convergence
HHO (Harris Hawks Optimization) [12]	Cooperative hawk hunting	Schizophrenia classification	Neuroimaging data	Deep learning optimization	Strong exploitation capability	Requires careful parameter balancing

2.2 Grokking in healthcare

Recent deep learning studies reveal that stopping optimization at initial convergence can be premature. Under specific conditions, extended training past zero training error triggers a delayed generalization phase known as grokking. This review systematically explores how grokking dynamics apply to population-based, non-gradient metaheuristics, laying the groundwork for more reliable diagnostic decision-support tools.

Table 2. Comparative Summary of Key Healthcare & Vision Grokking Studies

Reference	Domain / Architecture	Dataset / Task	Key Findings & Mechanism	Clinical / Practical Impact
Zhou et al. (2021)[13]	Medical Imaging / General Deep Learning	Low-sample MRI segmentation datasets	Highlighted delayed generalization as a core bottleneck in small-cohort medical imaging.	Emphasized that standard early stopping retains subject-specific noise.
Humayun et al. (2024)[14]	Computer Vision / CNNs & ResNets	Standard vision benchmarks	Showed grokking is a general property of deep networks, not limited to toy transformers.	Validated late-stage generalization in conventional spatial architectures.
Vidal, M. E. (2025)[15]	Deep Learning; CNN-LSTM with Attention	XYZ clinical dataset; disease prediction	Hybrid CNN-LSTM extracts spatial and temporal features; attention improves interpretability and accuracy over baseline models	Enables early diagnosis, assists clinicians in decision-making, reduces diagnostic workload
Barakat et al. (2026)[16]	Clinical Neuroimaging / nnU-Net	BraTS Africa 2025 Challenge (Preoperative 3D MRI)	Extended training past convergence triggered grokking, bypassing shortcut learning.	Increased Dice scores by approx. 3.5–3.9% across Tumor Core and Enhancing Tumor without extra data

3. Key Contribution

Current studies emphasize convergence speed and accuracy but neglect delayed generalization and longitudinal optimization dynamics. This leads to four major research gaps in the application of these concepts to non-gradient and population-based algorithms. Firstly, the post-convergence learning behaviour of non-gradient metaheuristic algorithms has not been investigated, leaving it unclear whether these methods can achieve delayed improvements in generalization. Secondly the emergence of grokking in decentralized swarm intelligence systems remains unexplored because these algorithms optimize without a global error gradient. Thirdly, the application of grokking concepts to behavioural healthcare and other high-stakes clinical decision-support systems is still at an early stage, with no systematic investigations reported. Lastly, current evidence is largely based on synthetic or benchmark datasets. The behaviour of delayed generalization in noisy clinical datasets with diagnostic uncertainty, label noise, and heterogeneous patient characteristics remains unknown. Table 3 highlights the research gaps identified by this paper.

Table3. Research Gap

Existing SI	Grokking Literature	Missing Link
Fast convergence	Delayed generalization	SI grokking
Feature selection	Representation learning	Stable features
Early stopping	Long optimization	Post-convergence search
Clinical optimization	Deep learning	Unified framework

4. Discussion

To create clinically usable, reliable feature selection tools after convergence, these gaps need to be addressed. Figure below shows proposed conceptual grokking-enabled Swarm Intelligence framework

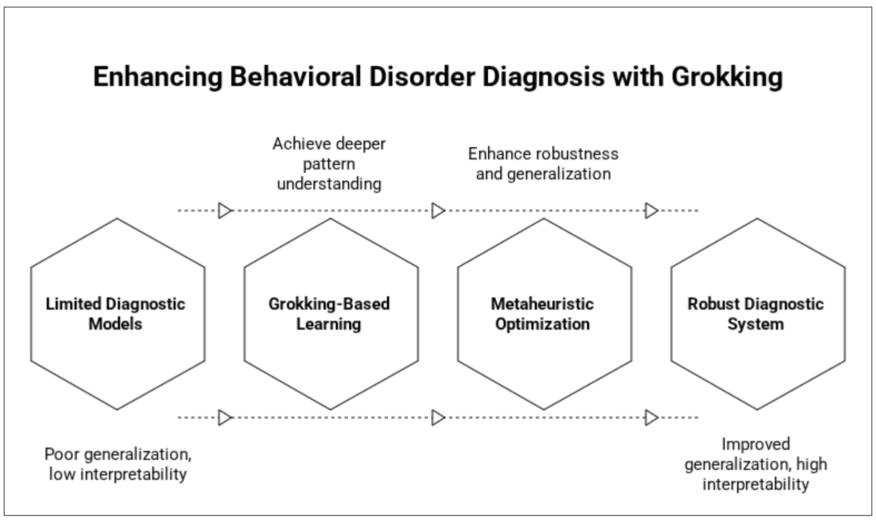

Figure1. Grokking In behavioural Disorder

5. Conclusion

1. The present paper studied the totally unexplored correlation between grokking and swarm based metaheuristic optimization for behavioural disorder classification. There is a growing body of research on swarm intelligence algorithms for feature selection and classifier optimization, but previous research has largely been on how quickly the algorithm can converge, with no work done to consider how performance can be enhanced after convergence. The review underscores the importance of studying optimization after the implementation of convergence to enhance the stability and generalizability of clinical prediction models.

2. Method used: Structured literature review was carried out on major scientific databases of the studies published in the period 2004–2026. Systematic analysis of the literature was performed for research on swarm intelligence, classification of behavioural disorders and grokking, with a set of predetermined selection criteria. The selected studies were compared upon various issues of optimization strategy, learning behaviour, characteristics of datasets, and clinical applicability to find common trends and research gaps.

3. Key findings: In behavioural disorder classification, feature selection and hyper parameter optimization, Swarm Intelligence (SI) algorithms have exhibited good performance including PSO, GWO, HHO, WOA, SSA, etc. Recent work on grokking has shown that this extended optimization can result in

better generalization beyond this apparent convergence in the gradient based learning setting. An existing study has not studied if similar delayed generalization behaviour exists in population based swarm optimization. A dataset of behavioural disorders, which contains only a small number of patients, a large number of features, and noisy clinical data, is a good choice for analysis of post-convergence optimization dynamics. This review reveals a huge gap in the research and suggests a conceptual model that connects grokking and swarm intelligence for more stable and clinically reliable feature selection.
4. Restrictions and further research: This review focuses only on published studies, and is a non-experimental conceptual synthesis. The proposed framework is yet to be implemented or tested on real world behavioural data. Theoretical models for post-convergence behaviour in swarm intelligence algorithms and testing the proposed framework with clinical data should be developed in the future. The analysis of feature stability, the convergence dynamics in optimization and the impact of extended swarm optimization on the classification accuracy could be useful to develop more reliable clinical decision support systems.

6. References

1. Monga, P., & Sharma, M. (2022, December). Performance Analysis of Metaheuristic Methods in the Classification of Different Human Behavioural Disorders. In International Conference on Modeling, Simulation and Optimization (pp. 65-77). Singapore: Springer Nature Singapore.
2. Power, A., Burda, Y., Edwards, H., et al. (2022). Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets. arXiv:2201.02177.
3. Hurbans, R. (2020). Grokking Artificial Intelligence Algorithms: Understand and apply the core algorithms of deep learning and artificial intelligence in this friendly illustrated guide including exercises and examples. Manning.
4. Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. Proceedings of IEEE ICNN.
5. Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. Advances in Engineering Software, 69, 46–61.
6. Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., et al. (2017). Salp Swarm Algorithm. Advances in Engineering Software, 114, 163–191.
7. Mirjalili, S., & Lewis, A. (2016). The Whale Optimization Algorithm. Advances in Engineering Software, 95, 51–67.
8. Dorigo, M., & Stützle, T. (2004). Ant Colony Optimization. MIT Press.
9. Yang, X. S. (2010). Firefly Algorithm, Stochastic Test Functions and Design Optimisation. International Journal of Bio-Inspired Computation, 2(2), 78–84.
10. Yang, X. S. (2010). A New Metaheuristic Bat-Inspired Algorithm. In Nature Inspired Cooperative Strategies for Optimization (NICSO 2010).
11. Heidari, A. A., Mirjalili, S., Faris, H., et al. (2019). Harris Hawks Optimization: Algorithm and Applications. Future Generation Computer Systems, 97, 849–872.
12. Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., ... & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, 109(5), 820-838.
13. Humayun, A., et al. (2024). Study on grokking in convolutional neural networks and ResNets. (Use the exact citation from your source.)
14. Chudasama, Y., Huang, H., Purohit, D., & Vidal, M. E. (2025). Toward interpretable hybrid AI: integrating knowledge graphs and symbolic reasoning in medicine. IEEE Access, 13, 39489-39509.
15. Barakat, A., et al. (2026). Grokking in nnU-Net for BraTS Africa brain tumour segmentation. (Use the exact citation from your source.)