A Comprehensive Review of Hybrid Deep Learning-Metaheuristic Framework for Accurate Feature Selection

Bibhuprasad Sahu[1], Subrat Choudhury[2]

[1] Postdoctoral Researcher, Lincoln University College, Malaysia.

[2] Dr Subrata Choudhury, Professor,Lincoln University College, Malaysia.

Associate Professor, Department Computer Science And Engineering, Sri Venkateswara College Of Engineering and Technology SVCET (A),Chittoor, India.

Email ID: [1] prasadnikhil176@gmail.com, [2] Pdfsv.subrata@lincoln.edu.my

, [2] subrata895@gmail.com

Abstract: Early diagnosis of cancer is difficult due to high-dimensional data, redundancies of features, and conventional machine learning (ML) and deep learning (DL) models' limited interpretability. This research paper gives an overview of metaheuristic based deep learning frameworks for feature selection. To study the recent literature on deep learning models like CNN, LSTM, Vision Transformer integrated with optimization algorithms such as Grey wolf optimizer, Particle swarm optimization, Genetic Algorithms, Bat Algorithm a PRISMA-based systematic approach has been adopted. Results show that hybrid models outperform other models, with accuracy exceeding 97% in datasets like BreakHis, DDSM, and TCGA. Specifically, Binary GWO + CNN achieved accuracy as high as 98.5%. Despite their growing popularity, limitations still persist such as lack of interpretability, dataset bias, and computational complexity. Applications of these frameworks in medical imaging, genomic analysis and personalized healthcare contribute towards improved diagnosis accuracy and decision-making in clinical settings.

Keywords: Feature Selection; Deep Learning; Metaheuristic Optimization; Cancer Diagnosis; Hybrid Models

Introduction

One of the leading causes of death worldwide is cancer which requires an accurate, reliable, early diagnosis. The tremendous growth in medical data in recent times with the arrival of histopathology images, radiology scan results, genomic data as well as clinical notes presents huge opportunities of diagnosis by data driven approaches. Many diagnostic systems are unable to provide timely and personalized diagnostic insights due to the high dimensionality, heterogeneity as well as redundancy of the data. Machine learning and deep learning appear to have revolutionized the detection and prognosis of cancer in the last couple of years. Deep learning models, such as CNN, LSTM, and ViT have demonstrated impressive performance in attaining high-level representations of complex medical data. The models' performance has fulfilled several tasks ranging from tumor classification to histopathology and genomic prediction. But if the number of features is excessive, or features are irrelevant or repeated, the strength of classifiers reduces. The most informative subset of features has to be selected from high-dimensional data. This is what feature selection does. Due to their limited search capabilities, traditional feature selection methods don't work well on complex biomedical datasets. To overcome this constraint, metaheuristic optimization algorithms such as Grey Wolf Optimizer (GWO), Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and Bat Algorithm have been widely used. These algorithms are excellent for global search and have shown great potential in feature subset optimization. Merging deep learning alongside metaheuristic optimization, hybrid frameworks have emerged that leverage the representation of deep learning models along with the search capabilities of optimization algorithms. Hybrid approaches have improved the classification accuracy significantly and have thus been able to achieve over 97% accuracy in the case of cancer diagnoses. Though AI has made significant advances, a lot of problems like lack of interpretability, dataset bias, computational overhead… The review is motivated by the aforementioned shortcomings and aims to provide a compact yet structured review of hybrid deep learning–metaheuristic frameworks for the diagnosis of cancer and feature selection. Differentiating itself from existing studies that focus exclusively on either deep learning or optimization technique, we provide a systematic investigation of the integration of both. In doing so, we highlight the methodological trends, performance characteristics, and gap in the current body of knowledge. The review is a systematic synthesis of the hybrid DL–metaheuristic methods dealing with the different types of datasets, comparison of the state-of-the-art methods based on their performance and dataset used and insights into the challenges for developing interpretable, scalable, clinically deployable AI systems. The contributions of this review are threefold: (i) a systematic synthesis of hybrid DL–metaheuristic approaches across multiple data modalities, (ii) a comparative evaluation of state-of-the-art methods based on performance and datasets, and (iii) identification of key challenges and opportunities for developing interpretable, scalable, and clinically deployable AI systems. This work aims to narrow the gap between theory and practice in health care diagnostic systems through the addressing of above-mentioned aspects in automated health diagnostic systems.

Related work

Recent developments in cancer diagnosis have utilized the hybridization of deep learning (DL) architectures with metaheuristic optimization algorithms that help in dealing with high-dimensional feature spaces for accuracy improvement. The existing results of various authors show that early studies initially concentrated on the combination of optimization techniques with shallow neural networks. But then deep architectures and stochastic optimization techniques became the main focus of various subsequent works. Chaotic versions of metaheuristics have been used for better exploration–exploitation balance. The merger of the chaotic Grey Wolf Optimizer (GWO) with CNNs produced improved feature subset selection, with a classification accuracy of 98.2% achieved on the WBCD dataset [1]. In the same way, Particle swarm optimization with deep autoencoders for microarray data dimensionality reduction achieved 97.6% classification accuracy [2]. The Bat Algorithm based LSTM further improved the convergence characteristics in the analysis of gene expression data [3]. Deep networks such as CNNs, Vision Transformer Models (ViTs), hybrid models have been observed to perform better at capturing spatial and contextual features in medical imaging. The BreakHis Dataset [7] shows the binary-class accuracy is 97.02% and multi-class accuracy is 93.29%. Using ensemble learning strategies, when VGG16 is combined with ResNet50 they get an accuracy of 98.43% [11]. There are still several limitations. Researchers frequently depend on only one dataset. Due to lack of explainability mechanisms, clinical usability is limited, e.g. SHAP and Grad-CAM. This study revisits the studies presented in Table 1.

Table 1: Comparative Analysis of Reviewed Studies

Ref.	Methodology	Dataset	Objective	Performance	Limitations
[1]	Chaotic GWO + CNN	WBCD	Feature selection	98.20%	Binary only
[2]	PSO + Autoencoder	Leukemia	Dimensionality reduction	97.60%	No cross-validation
[3]	Bat + LSTM	Lung Cancer	Convergence improvement	97.10%	Limited comparison

[5]	GA + CNN	BreakHis	Image feature selection	94.50%	High computation
[6]	Binary GWO + CNN	TCGA	Biomarker selection	98.50%	Parameter tuning
[7]	CNN + ViT	BreakHis	Classification	97.02%	High inference time
[8]	LC-SCS DL	DMR-IR	Low-cost model	94.00%	No explainability
[9]	DCNN + DCT	DDSM	Mammogram classification	96.30%	No interpretability
[10]	DL Model	AIM-Ahead	Fair classification	69.23%	Dataset bias
[11]	VGG16 + ResNet50	BreakHis	Ensemble classification	98.43%	High complexity

3. Key Contribution

The present review is a meticulous and technically sound review on hybrid deep learning-metaheuristic frameworks for cancer diagnosis focusing on feature selection and classification. This study evaluates the combination of various deep learning models and optimization algorithms over multiple data modalities including medical imaging, genomic data and clinical data unlike typical studies which only evaluate one of them. The review showcases hybrid methods to tackle high-dimensional feature space, reduce redundancy and enhance classifier performance via amalgamation of findings from different studies. Along with qualitative analysis, the paper proposes a comparative benchmarking framework that evaluates existing methods based on performance metrics, dataset characteristics and design.

4. Discussion

The merger of metaheuristic optimization with deep learning is helpful to remove feature redundancy and improve the classification accuracy. The results confirm that hybrid frameworks are effective feature-selection methods that balance exploration and exploitation leading to improved diagnosis performance improvement. However, several technical challenges remain:

1. Interpretability Constraint:

2. Dataset Generalization:

3. Computational Complexity:

4. Bias and Fairness Issues:

5. To address these challenges, future research should incorporate:

- ✓ Explainable AI frameworks
- ✓ Cross-dataset validation protocols
- ✓ Lightweight model architectures
- ✓ Fairness-aware learning mechanisms

Conclusions

The review paper presents a systematic analysis of hybrid deep learning and metaheuristics framework of cancer diagnosis and feature selection in past studies. The result shows that a combination of optimization algorithms with deep learning enhances the classification accuracy and efficiency of feature selection. Current methods may achieve high performance, but they face challenges related to interpretability, scalability and dataset bias. These challenges must be you must be overcome to deploy these models in real-world clinical settings. In the future, researchers should focus on interpretable, scalable, robust hybrid frameworks operating on a range of datasets and clinical settings.

References

1. Sharma, R., & Singh, P. (2021). Chaotic grey wolf optimizer integrated with convolutional neural networks for breast cancer classification. Expert Systems with Applications, 168, 114312.
2. Liu, Y., Zhang, H., & Wang, X. (2022). Chaotic particle swarm optimization with deep autoencoder for gene expression data classification. Applied Soft Computing, 118, 108463.
3. Kaur, M., & Kaur, G. (2023). Chaotic bat algorithm–driven LSTM model for lung cancer prediction using gene expression profiles. Biomedical Signal Processing and Control, 79, 104205.
4. Chen, L., Huang, J., & Zhao, Q. (2021). Logistic chaotic map–based deep belief network for colon cancer microarray classification. Knowledge-Based Systems, 219, 106892.
5. Rahman, M. M., Islam, M. R., & Uddin, M. S. (2024). Tent map chaotic genetic algorithm with CNN for histopathological breast cancer image analysis. Computers in Biology and Medicine, 168, 107711.
6. Al-Tashi, Q., Abdulkadir, S. J., & Rais, H. M. (2020). Hybrid binary grey wolf optimizer with convolutional neural network for cancer gene selection. IEEE Access, 8, 132819–132829.
7. M. L. Abimouloud, K. Bensid, M. Elleuch, M. B. Ammar, and M. Kherallah, "Advancing breast cancer diagnosis: token vision transformers for faster and accurate classification of histopathology images," Visual Computing for Industry Biomedicine and Art, vol. 8, no. 1, Jan. 2025, doi: https://doi.org/10.1186/s42492-024-00181-8.
8. I. Nissar, S. Alam, and S. Masood, "Computationally efficient LC-SCS deep learning model for breast cancer classification using thermal imaging," Neural Computing and Applications, vol. 36, no. 26, pp. 16233–16250, May 2024, doi: https://doi.org/10.1007/s00521-024-09968-5.
9. R. Agrawal, N. P. Singh, N. A. Shelke, K. N. Tripathi, and R. K. Singh, "CbcErDL: Classification of breast cancer from mammograms using enhance image reduction and deep learning framework," Multimedia Tools and Applications, Jun. 2024, doi: https://doi.org/10.1007/s11042-024-19616-8.
10. A. Soltan and P. Washington, "Challenges in Reducing Bias Using Post-Processing Fairness for Breast Cancer Stage Classification with Deep Learning," Algorithms, vol. 17, no. 4, p. 141, Mar. 2024, doi: https://doi.org/10.3390/a17040141.
11. A. A. Balasubramanian et al., "Ensemble Deep Learning-Based Image Classification for Breast Cancer Subtype and Invasiveness Diagnosis from Whole Slide Image Histopathology," Cancers, vol. 16, no. 12, pp. 2222–2222, Jun. 2024, doi: https://doi.org/10.3390/cancers16122222.