ASEC 2024 Conference

The 5th International Electronic Conference on Applied Sciences

04–06 December 2024 | Online

MDPI

A HYBRID SWARM OPTIMIZATION ALGORITHM FOR IMPROVING FEATURE SELECTION IN MACHINE LEARNING

Kabir Muhammad, Kabir Umar, Ruqayya Auwal, Hauwa Abubakar Musa

Computer Science Dept. of Bayero University, Kano State, Nigeria

INTRODUCTION & AIM

In recent years, the volume of data has exponentially increase which in turn decreases the quality of data required for data mining, image processing, pattern recognition and in general machine learning prediction. The increase in the volume of data leads to increase in the dimensionality of data which makes it to contain noisy, redundant and irrelevant data. When this data is fed into a machine learning algorithm, it increases the error rate of the algorithm.

To eliminate these problems, dimensionality reduction techniques such as feature selection are applied as part of the preprocessing step of machine learning. Feature selection is a technique used to minimize size of the dataset by cleaning up the noisy, irrelevant and redundant data which in turn increase the performance of machine learning prediction.

Metaheuristic algorithms, inspired by natural phenomena, have been extensively used for feature selection. Among these, Cat Swarm Optimization (CSO) and Crow Search Algorithm (CSA) have demonstrated strong individual performances in solving complex optimization problems, such as:

•**Smart Grid Optimization** (Raj Kumar, 2019)

•**Vehicle Routing with Time Windows** (Pratiwi, 2018)

However, a hybrid approach combining CSO and CSA for feature selection in machine learning has not yet been explored. This research addresses this gap by developing a novel **Hybrid CSO-CSA algorithm** to achieve optimal feature selection.

The aim of this research work is to develop a hybrid feature selection algorithm for improving performance of machine learning prediction

PROBLEM

The rapid growth of data has led to high-dimensional datasets filled with noisy, redundant, and irrelevant features. These issues negatively impact machine learning (ML) models by increasing computational complexity and reducing prediction accuracy. Feature selection is a critical preprocessing step to address these challenges, reducing dimensionality and enhancing model performance.

Metaheuristic algorithms like Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1995), Crow Search Algorithm (CSA) (Askarzadeh, 2016), and Artificial Bee Colony (ABC) (Karaboga, 2005) have been widely used for feature selection. However, these methods face limitations such as:

•**Over-selection** of redundant features.

•**Susceptibility to local optima**, especially in high-dimensional spaces.

•**Low convergence rates**, impacting computational efficiency.

Hybrid approaches, such as ECCSPSOA (Adamu et al., 2021), which combine PSO and CSA, have attempted to overcome these limitations. However, these approaches still struggle with selecting irrelevant features and lack adaptability in complex datasets.

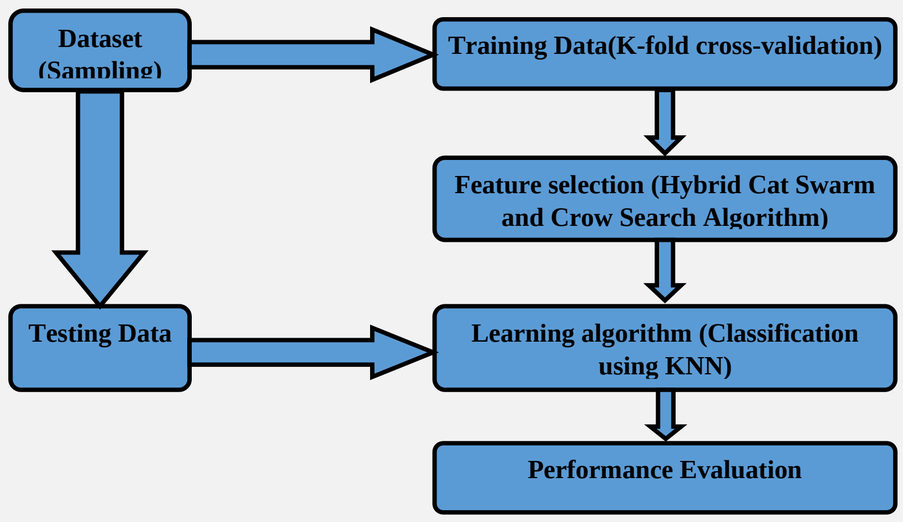
Cat Swarm Optimization (CSO), a newer algorithm inspired by the behavior of cats, outperforms PSO in local and global search strategies(Chu et al., 2006). When hybridized with CSA, this combination has shown success in solving optimization problems like smart grid optimization (Raj Kumar, 2019) and vehicle routing with time windows (Pratiwi, 2018). Despite this potential, CSO-CSA has not been explored for feature selection in ML.

This research proposes a novel hybrid CSO-CSA algorithm to improve feature selection by overcoming existing limitations, reducing redundancy, and enhancing ML prediction accuracy.

METHOD

This study presents a hybrid feature selection algorithm combining **Cat Swarm Optimization (CSO)** and **Crow Search Algorithm (CSA)** to enhance machine learning (ML) performance by reducing redundant and irrelevant features. The methodology consists of the following steps:

- Datasets: Twelve standard datasets from the UCI repository were used. Missing values were handled using mean imputation, and data was normalized with Min-Max normalization. Each dataset was divided into training and testing sets using k-fold cross-validation to ensure robust evaluation.
- Hybrid Algorithm: The algorithm alternates between CSO (focused on local search efficiency) and CSA (optimized for global exploration) to dynamically update solutions and identify the most relevant features.
- Evaluation: Features were tested using the K-Nearest Neighbors (KNN) classifier with k-fold cross-validation to ensure robust accuracy measurements.
- Metrics: Accuracy, precision, recall, F1-score, and optimization consistency were used to validate performance.
- Implementation: The algorithm is implemented in Python, leveraging tools such as NumPy, Pandas, and scikit-learn in a Google Colab environment.



RESULTS & DISCUSSION

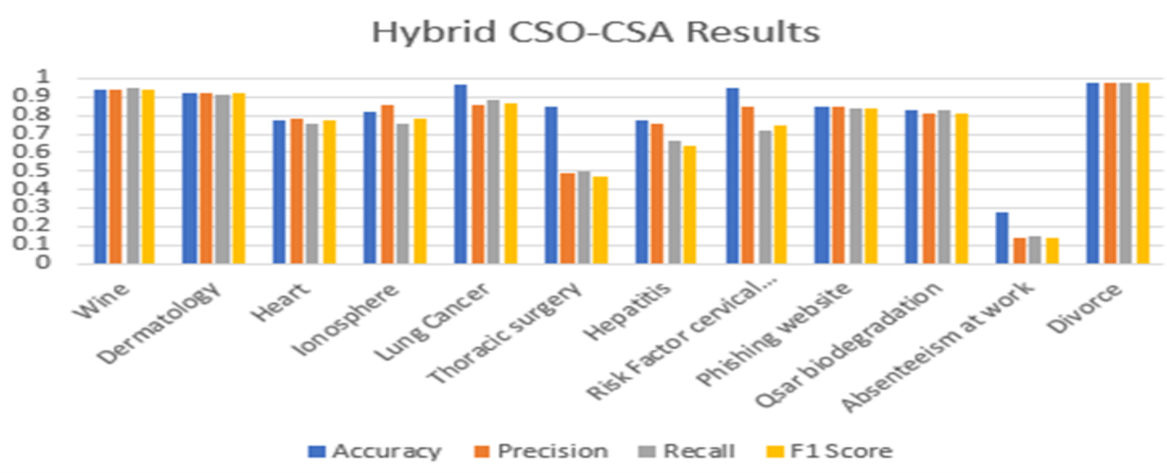


Table 1: Result Presentation

- The proposed CSO-CSA hybrid algorithm achieved an average accuracy of 87%, outperforming the baseline algorithm (ECCSPSOA) with an accuracy of 83%.
- The hybrid approach demonstrated improved feature selection by reducing redundancy and irrelevant data while maintaining model consistency.
- Performance Metrics:
  - Mean Fitness: Higher compared to standalone methods.
  - Best Fitness: Significantly better convergence rates.
  - Standard Deviation: Lower variability, indicating robust optimization.

CONCLUSION

The proposed CSO-CSA hybrid algorithm effectively enhances feature selection for machine learning by reducing noisy, redundant, and irrelevant features, resulting in improved prediction accuracy. The algorithm outperforms traditional methods, achieving an average accuracy of 87%, compared to 83% in previous approaches. By leveraging the strengths of both Cat Swarm Optimization (CSO) and Crow Search Algorithm (CSA), the hybrid approach ensures better convergence, consistency, and optimization efficiency.

FUTURE WORK / REFERENCES

Future research will focus on testing the CSO-CSA hybrid algorithm with other classifiers like SVM and Random Forest to assess its robustness. Additionally, applying the algorithm to real-time datasets in domains such as healthcare and finance can help evaluate its practical applicability. Finally, exploring parallel computing techniques could further enhance the algorithm's efficiency for large-scale problems.

- Raj Kumar. (2019). Hybrid Cat Swarm and Crow Search Algorithm to Solve the Combined Economic Emission Dispatch Model for Smart Grid. Journal of Computational Mechanics, Power System and Control, 2(3), 10–18. https://doi.org/10.46253/jcmps.v2i3.a2
- Pratiwi, A. B. (2018). A hybrid cat swarm optimization - Crow search algorithm for vehicle routing problem with time windows. Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017, 2018-Janua, 364–368. https://doi.org/10.1109/ICITISEE.2017.8285529
- Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. 1942–1948.
- Askarzadeh, A. (2016). A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm. Computers and Structures, 169, 1–12. https://doi.org/10.1016/j.compstruc.2016.03.001
- Karaboga, D. (2005). AN IDEA BASED ON HONEY BEE SWARM FOR NUMERICAL OPTIMIZATION. Journal of Physics A: Mathematical and Theoretical, 44(8), 1–29. https://doi.org/10.1088/1751-8113/44/8/085201
- Adamu, A., Abdullahi, M., Junaidu, S. B., & Hassan, I. H. (2021). An hybrid particle swarm optimization with crow search algorithm for feature selection. Machine Learning with Applications, 6(April), 100108. https://doi.org/10.1016/j.mlwa.2021.100108