`

## Feature Prediction Using Meta-Heuristic Optimization Algorithms

*Review-2*

### B. Tech

**Computer Science and Engineering**

**By**

### A SAI CHARAN (20BDS0354)

### R PRADEEP KUMAR (20BDS0183)

*Under the guidance of*

*Krishna Raj N,*

*Assistant Professor,*

*SCOPE, VIT, Vellore*



**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

### VELLORE INSTITUTE OF TECHNOLOGY

**VELLORE 632 014, TAMILNADU, INDIA**

**April 2024**

**Abstract**

Feature selection is paramount in enhancing the performance and interpretability of machine learning models, particularly in supervised classification and regression tasks. This paper presents an investigation into the integration of convolutional neural network (CNN) models such as ResNet, VGG, Inception, and GoogLeNet with binary versions of meta-heuristic algorithms for feature selection in supervised learning tasks, followed by their application to image datasets. The primary focus lies in efficiently identifying discriminative feature subsets for supervised classification and regression, leveraging the strengths of CNN architectures and meta-heuristic techniques including genetic algorithms, particle swarm optimization, simulated annealing, and ant colony optimization. In addition, clustering methods such as the isolation forest algorithm are incorporated to refine feature subsets extracted by CNN models and meta-heuristic algorithms. Through a comprehensive review of existing literature and empirical analysis, we assess the efficacy of these integrated approaches on benchmark datasets for supervised classification and regression tasks. We also investigate their performance on real-world image datasets, providing insights into the challenges and opportunities of image-based feature selection and classification. By demonstrating the effectiveness of these methodologies in supervised learning contexts, particularly in addressing feature selection for classification and regression, this research contributes to advancing the understanding of feature selection techniques and provides valuable guidance for practitioners and researchers in machine learning.

**Introduction**

Computational intelligence and optimization encompass the topic of meta-heuristic optimization techniques. These algorithms are made to effectively search solution spaces for the best or almost best answers to challenging situations. Meta-heuristic algorithms are derived from abstract ideas such as natural events, rather than typical optimization techniques that depend on mathematical equations and derivatives. Feature selection is the process of removing the redundant or the other components which doesn’t contribute much to the dataset while modelling. Existing methodologies like feature selection using Gini index, wrapper and embedded method. In the contemporary world the size of data is usually getting bigger. For processing such large amounts of data heavy compute and storage power is required, which can’t be compensated by all users with existing methods. Existing wrapper models such as RandomForest and Unsupervised methods such as clustering all will face the curse of dimensionality. The Clustering method’s may require the optimized inputs like cluster size, which would again use Elbow-Curve method to find a better value for the number of clusters. The project domain involving meta-heuristic optimization algorithms encompasses various fields, including Engineering Design: Meta-heuristic algorithms are widely used in engineering design problems such as structural optimization, aerodynamic design, and mechanical component design. These algorithms help in finding the best configuration or parameters that satisfy multiple constraints and objectives. Operations Research: Meta-heuristic optimization techniques are extensively applied in operations research problems such as scheduling, logistics, and resource allocation. These algorithms are adept at solving complex combinatorial optimization problems where traditional methods often fail due to the exponential growth of search space. Machine Learning and Data Mining: Meta-heuristic algorithms play a significant role in training complex machine learning models and optimizing their hyperparameters. They are also employed in data mining tasks such as clustering, feature selection, and association rule mining. Finance and Economics Meta-heuristic optimization techniques find applications in financial modeling, portfolio optimization, risk management, and algorithmic trading. They assist in making informed decisions under uncertainty and optimizing investment strategies.

**Literature Survey**

|  |  |  |  |
| --- | --- | --- | --- |
| Title | Author | Description Of Invention | Advantages |
| Innovative  Feature Selection Method Based on Hybrid Sine Cosine and Dipper Throated Optimization Algorithms  (2023) | ABDELAZIZ A. ABDELHAMID , EL-SAYED M. EL-KENAWY 3 , (Senior Member, IEEE), ABDELHAMEED IBRAHIM 4 , (Member, IEEE), MARWA METWALLY EID5 , (Member, IEEE), DOAA SAMI KHAFAGA 6 , AMEL ALI ALHUSSAN 6 , SEYEDALI MIRJALILI7,8, (Senior Member, IEEE), NIMA KHODADADI9 , (Member, IEEE), WEI HONG LIM10, (Senior Member, IEEE), AND MAHMOUD Y. SHAMS11 | Feature selection is crucial in pattern recognition and data mining for identifying relevant features and improving classification algorithms' performance. This problem can be seen as an optimization issue, tackled by meta-heuristic techniques.  This paper proposes a novel hybrid binary meta-heuristic algorithm named **bSCWDTO** for feature selection, combining the **Dipper Throated Optimization (DTO)** and **Sine Cosine (SC)** algorithms. | bSCWDTO Algorithm: This algorithm leverages the exploration strengths of the SC algorithm and the exploitation capabilities of the DTO algorithm. It incorporates:  Sine and cosine search operators: Guide exploration and enhance global search  Local search: Improves exploitation around promising solutions  Levy flight: Enhances exploration diversity  Selection mechanism: Balances exploration and exploitation |
| Centroid mutation-based Search and Rescue optimization algorithm for feature selection and classification(2023) | Essam H. Houssein ∗ , Eman Saber, Abdelmgeid A. Ali, Yaser M. Wazery | Several meta-heuristic algorithms have been explored for this task, including Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Harmony Search (HS).  The **Centroid Mutation-based Search and Rescue (CMSR) optimization algorithm** emerges as a novel approach for feature selection and classification. It draws inspiration from search and rescue operations, mimicking the collaborative efforts to locate missing individuals.  Fitness: Evaluated based on a combination of classification accuracy and feature relevance.  Search Operators:  Migration: Selected individuals move towards promising areas guided by high-performing solutions.  Local Search: Refines solutions within promising regions using a centroid-based mutation operator.  High-performing individuals guide others towards better solutions.  Selection: Based on fitness and diversity to maintain balance between exploration and exploitation | * Effective exploration and exploitation: Leverages diverse operators for both broad search and focused improvement. * Enhanced classification accuracy: Achieves competitive performance compared to state-of-the-art algorithms on benchmark datasets. * Adaptability: Flexible for various classification tasks and datasets. * Collaboration and rescue mechanisms: Promote population diversity and prevent premature convergence. |
| BMPA-TVSinV: A Binary Marine Predators Algorithm using time-varying sine and V-shaped transfer functions for wrapper-based feature selection.(2023) | Zahra Beheshti | BMPA-TVSinV addresses these challenges by proposing a novel Binary Marine Predators Algorithm (BMPA) with time-varying sine and V-shaped transfer functions. This method builds upon the original MPA algorithm, known for its exploration capabilities, but struggles with exploitation and binary conversion.  Methodology:  Core principles of MPA make  Predators (solutions) move in the search space based on prey location (high-performing solutions).  Lévy flight enhances exploration.  Social interaction fosters collaboration.  Innovations:  Time-varying sine function: Dynamically controls exploration and exploitation throughout iterations.  V-shaped transfer function: Efficiently converts continuous search space values to binary feature selections. | BMPA-TVSinV achieves statistically significant accuracy improvement compared to GA, PSO, and other state-of-the-art feature selection methods on several benchmark datasets, including high-dimensional ones.  The dynamic exploration-exploitation balance driven by the time-varying functions and V-shaped transfer function is credited for this improved performance. |
| Multiclass classification of leukemia cancer data using Fuzzy Support Vector Machine (FSVM) with feature selection using Principal Component Analysis (PCA)(2023) | [Fauzi, Rustam, and Wibowo (2021)](https://www-sciencedirect-com.egateway.vit.ac.in/science/article/pii/S0957417421015463?via%3Dihub" \l "b20) | PCA method coupled with the Fuzzy Support Vectors Machines (FSVM) | Accuracy = 96.92% (obtained by using 60 features). |
| Hierarchical Harris hawks optimizer for feature selection | Lemin Peng, Zhennao Cai, Alisghar Heidari ,Lejun Zhang,Huiling Chen | The Hierarchical Harris Hawks Optimizer (EHHO) addresses these limitations by introducing a hierarchical structure to the traditional HHO algorithm. This approach aims to improve its effectiveness in feature selection tasks.  Methodology:  Hierarchical structure:  Multiple levels, each representing a subset of features.  Information exchange between levels allows exploitation of promising features across different scales.  Modified operators:  Exploration and exploitation phases adapted for the hierarchical structure, including modified levy flight and escape mechanisms.  Feature subset evaluation integrated into the optimization process.  Selection:  Selection within and between levels ensures diversity and promotes convergence towards improved feature subsets. | Accuracy:  EHHO exhibits superior accuracy compared to traditional HHO, GA, PSO, and other feature selection methods, particularly on high-dimensional and complex datasets.  The hierarchical structure facilitates efficient information exchange and exploitation of relevant features across different scales, contributing to this improved performance. |

|  |  |  |
| --- | --- | --- |
| Title | Author | Inference |
| Feature selection using Ant Colony optimization for microarray data classification | Sanjay Prajapati  Himansu Das  Mahendra Kumar Gourisaria | Microarray datasets have high dimensions and feature selection is important for high-dimensional data. The paper proposes Ant Colony Optimization (ACO) combined with Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF) algorithms for feature selection.  The ACO algorithm is used to select the next feature based on a probability value, and the selected features are added to the ant's solution.  The solutions obtained by the ants are represented in binary form.  The accuracy and fitness levels of different algorithmic models and datasets are compared to evaluate the performance of the proposed approach.  The paper results that both Random forest and logistic regression perform well when used with ACO feature selection.  Limitation - The paper does not provide a detailed analysis of the computational complexity or runtime of the proposed ACO feature selection method. |
| Adaptive fuzzy neighbourhood based multilabel feature selection with ant colony optimization | Lin Sun, Yusheng Chen, Weiping ding, jiucheng xu, yuanyuan ma | The proposed adaptive fuzzy neighbourhood-based multilabel feature subset selection approach with ant colony optimization (ACO) achieved excellent results in terms of five metrics (AP, CV, RL, HL, and OE) on six multilabel datasets. It outperformed several comparative algorithms, including PMU, MUCO, MCLS, MFSR, MFSFS, and MFSMR, in terms of classification efficiency .  The algorithm demonstrated optimal performance in terms of AP, CV, RL, and HL indicators on all datasets, while slightly lower performance was observed in the OE index on the Computer and Yeast datasets due to the presence of negative feature values and large differences between feature values .  Comparative analysis with state-of-the-art algorithms, such as CDR, MLFSNRS, RDPM, NRFSFN, and MLACO, further confirmed the effectiveness of the proposed algorithm in achieving optimal feature subsets for multilabel classification.  The paper also acknowledges some limitations, such as the inability to fully weaken the influence of redundant features and the need for further exploration of uncertainty measures to improve multilabel classification |
| Ant-ehfs: Ant colony optimization equipped with an ensemble of heuristics through fuzzy logic for feature selection | N.Z. Joodaki , M.B. Dowlatshahi , and M. Joodaki | The paper utilizes Ant Colony Optimization (ACO) as the main optimization algorithm for feature selection .  An ensemble of heuristics is employed, which combines multiple feature selection methods to improve the efficiency and performance of feature selection .  Fuzzy logic is used to address uncertainty in the ranks of feature selection methods and calculate the trustworthiness rate for each feature  Three feature selection methods (Maximal Information Coefficient, t-test, and RF S) are individually operated to rank the features  The Euclidean Distance between each pair of features is computed as a heuristic, and a weighted graph is constructed based on the combination of the two heuristics  ACO is applied to the complete graph, where each ant considers the reliability rate and Euclidean Distance of the destination node for moving between nodes.  The proposed method is compared with five ensemble feature selection methods and eight primary feature selection methods on high-dimensional datasets to evaluate its performance |
| The Integration of Genetic and Ant Colony Algorithm in a Hybrid Approach | Apostolos Tsagaris Panagiotis Kyratsis Gabriel Mansour | The paper proposes an improved genetic algorithm that combines ant colony algorithm for path optimization .  The hybrid algorithm utilizes the global search features of the ant colony optimizer and the stepwise search features of the genetic algorithm to improve the optimal parameters and accelerate the finding of the optimal solution .  The experimental results show that the proposed algorithm can automatically obtain better parameters, especially in its initial values, and achieve better solution accuracy, robustness, and efficiency compared to a simple genetic algorithm .  The hybrid algorithm was tested on a TSP problem and has applications in spatial mechanics systems such as CNC machining, robotic systems, and Coordinate Measuring Machines (CMM)  The proposed methodology starts with the ants' algorithm, which feeds the initial conditions to the genetic algorithm, aiming to improve the application of the hybrid model. |
| Improved ACO Rank-Based Algorithm for Use in Selecting Features for Classification Models | Roberto Alexandre Delamora, Bruno Nazario Coelho and Jodelson Aguilar Sabino | The paper proposes an improvement in the general construction of the Ant Colony Optimization (ACO) algorithm, specifically the Rank-based version by Bullnheimer et al., for feature selection in classification models.  The proposed approach aims to increase the overall efficiency of the ACO algorithm by making improvements and adjustments to the subset evaluation process in the original Rank-based version.  The modified version, called ACOFS rank, randomly places ants on the remaining features and uses the statistical correlation metric, specifically Spearman Correlation, to build the model's matrix.  The proposed algorithm is evaluated on several real-life datasets from the UCI machine-learning repository and compared with the WFACOFS method by Ghosh et al., showing better performance in most cases.  The limitations of the ACO algorithm itself, such as its sensitivity to parameter settings or its convergence properties, are not discussed in the paper. |

**Requirement Analysis**