Available at https://www.iasj.net/iasj
Iraqi Academic Scientific Journals
Journal homepage: https://journals.uokerbala.edu.iq/index.php/UOKJ

Research Article

A Machine Learning and Ant Colony Optimization-Based Multi-Objective Feature Selection Model for Biomedical Data Classification

Sabah Mohammed Fayadh
Southern Technical University / Technical College of DhiQar / Department of Cybersecurity Engineering Technologies

Article Info
Article history:
Received 28 -11-2025
Received in revised form 15-12-2025
Accepted 22-1-2026
Available online 31 -3 -2026
Keywords:
Ant Colony Optimization, feature selection, biomedical data classification, dimensionality reduction, Random Forest, Support Vector Machine.

Abstract

The need to address emerging challenges presented by the ever-expanding capabilities of genomics, proteomics, and medical imaging research introduces new challenges to clinical machine learning (ML). Examples of such variables include inefficiency in computations and overfitting. To address these issues, in this study, feature selection (FS) paradigm is proposed, which integrates predictive performance and feature parsimony, grounded on the Ant Colony optimization (ACO). The framework effectively searches big feature spaces in order to locate small useful subsets. The evaluation of candidate features subsets chosen by ACO is done using three classifiers (RF, SVM and GBM). With as few as 300 features, dimensionality reduction (by up to 92 percent) is observed without losing accuracy on four high-dimensional biological datasets, with Leukemia achieving 96.4. Enrichment analysis of Gene Ontology proves that our method is more accurate, more stable, and biologically relevant than filter-based methods and metaheuristics. The method is quite suitable to a clinical environment having limited resources due to its competitive running times and a fast convergence (around 70 iterations). Developing feature sets, which are miniature, comprehensible, and physiologically pertinent, the results demonstrate that the proposed framework can enhance clinical decision assistance frameworks and tailored therapy.

Corresponding Author E-mail: Sabah.fayadh@stu.edu.iq
Peer review under responsibility of Iraqi Academic Scientific Journal and University of Kerbala.

1. Introduction

Genomics, proteomics and medical imaging have experienced an exponential expansion, producing high-dimensional, heterogeneous and sparse biomedical data sets. Such properties require learning algorithms that can be able to manage the feature space with some complexities without compromising clinical reliability and interpretability. Recent innovations in artificial intelligence (AI) and machine learning (ML) have shown impressive potential in the ability to model high-dimensional data, identify nonlinear relationships, and enhance detection and classification accuracy in the fields of healthcare prediction and classification, radiology, dermatology, oncology, and intelligent sensing settings [1].

Deep learning is one of them among others that have so far displayed remarkable performance within the field of medical image analysis. Dermatology Deep neural networks have developed dermatologist-level accuracy on skin cancer classification, with great potential to offer point-of-care clinical decision support [2]. Later investigations have continued to enhance these methods with classical machine learning methods like support vector machine, coupled with deep features presentations to enhance the precision of lesion segmentation and classification [3]. Multi-task deep learning models have been suggested as well to do both segmentation and classification together to learn stronger features using thermoscopic images [4]. Moreover, it has been shown that ensemble based convolutional neural network models can be used to achieve better generalization and resistance against skin lesion classification depending on complimentary decision boundaries [5].

Exceeding dermatology, other biomedical imaging areas have been documented as to have experienced similar developments. The combination of radiomic/deep features of multiparametric MRI has improved the performance of tumour grading and risk stratification in the diagnosis of prostate cancer [6]. Quickly evolving radiomic assessment and within-modality synthesis of brain MRI has also been utilized to enhance tumour characterization and prognosis modelling [7]. Additionally, imaging-based machine learning systems (imaging features) with classical classifiers have demonstrated to be promising in automated brain tumour detection and classification [8]. Similar developments have been witnessed in breast cancer diagnostics, where deep learning models with transfer learning can be used to detect tumours and do this accurately and scalable on imaging data [9]. Closely similar studies have also shown that mutual bootstrapping and collaborative learning can also enhance the accuracy of segmentation and classification in difficult medical imaging contexts [10].

Even with these significant successes, high dimensionality still stands out as a central limitation to the effective and generalizable clinical ML. Examples of datasets of gene expression typically involve thousands of correlated features in each sample and examples of image pipelines typically produce hundreds or thousands of radiomic features or deep features. The richness of such representations often introduces redundant and irrelevant variables, which only compounds the curse of dimensionality, elevates computational cost, and also increases the probability of overfitting,

especially in high-dimensional examples, imbalanced classes or in expensive annotation methods. These restrictions pose a risk to interpretability, robustness, and trustworthiness, which is the cornerstone of safe implementation of AI-driven systems in high-performance health facilities [11]. The pressing need of taking care of these challenges is further compounded by the huge burden of diseases in the world like skin cancer which is still causing severe morbidity and healthcare expenditures in the world at large [12].

Conventional filter, wrapper as well as embedded feature selection (FS) techniques have been utilized extensively in alleviating the impact of high dimensionalities. But most of these methods are fundamentally single objectively and are not particularly sensitive to other similarly clinically significant properties (e.g. feature sparsity, selection stability, interpretability, and biological relevance). Although metaheuristic-based FS methods include hybrid, and improved versions of the Ant Colony Optimization (ACO) swarm algorithm, which use local search methods or alternative swarm intelligence methods, all these methods focus more on the performance of the algorithm instead of its translational reliability. As a result, biological validation is frequently overlooked, and the lack of stability of chosen feature sets between data partitions is still a significant drawback of high dimensional, small-sample biomedical studies.

To overcome these drawbacks, this paper mathematically formulates a biomedical feature selection as a multi-objective problem, where the conflicting aims of predictive performance optimization and feature dimensionality reduction are realised with multi-objective maximisation. The suggested framework is an extension of the Ant Colony Optimization (ACO), which also browses through the feature space through pheromone-controlled heuristics to extract feature subsets that are small but informative. Thanks to its population-based search approach, feature combination learning based on memory, and reinforcement of high-quality combinations of features through iterative approaches, ACO is specifically well designed to address large combinatorial search spaces typical of genomic as well as medical imaging data [13]. After the feature subsets are constructed, the feature subsets are tested with complementary machine learning classification algorithms: Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machine (GBM) to stringently evaluate the generalization performance with respect to the various inductive biases.

Threefold are the major contributions of this work. We present, first, the multi-objective ACO framework that uses pheromones and heuristic to optimize the predictive accuracy and feature parsimony aiming at reducing redundancy and irrelevance in high dimensional biomedical data. Second, the proposed ACO-based feature selection method is combined with ensemble and margin-based classification algorithms (RF, SVM and GBM) to thoroughly assess the ability to discriminate and strength of the underlying subset of features in various biomedical tasks. Third, we introduce a

comprehensive experimental analysis of several real-world biomedical data, such as cancer and neurodegenerative diseases, using a combination of classification performance, computational effectiveness, feature stability, and biological informativeness thus offering deployment-focused information on the scientific application of the framework. In line with the general integration of deep learning-based classification systems into various application conditions [14], the paper at hand shows that multi-objective optimization is required in demand of constructing compact, correct, stable, and biologically significant models that can be translated into clinical use.

2. Related Work

It is a fact that feature selection (FS) has been considered as an important preprocessing step in the biomedical data classification since high dimensionality, redundancy and noise are very common in clinical datasets. Initial FS studies in biomedicine were mainly on wrapper-based optimization schemes, in which feature sets are directly assessed by classification accuracy. Aalaei et al. [15] used a wrapper method based on the Genetic Algorithm (GA) to diagnose breast cancer and it achieved better classification accuracy on several data sets, although at the cost of higher computation cost by training the models repeatedly. Yan et al. [16] introduced a hybrid feature selection algorithm, which combines mutual information with a better gravitational search algorithm, demonstrating a better level of discriminative ability with regard to biomedical data of a high dimensionality. In the same manner, Farid et al. [17] proposed a hybrid type of the GA-based learning framework to detect breast cancer, and a higher predictive accuracy was attained by optimizing the learning process of the subsets of features.

Although they are quite useful, GA-based and hybrid wrapper algorithms have scalability problems and are sensitive to character of tuning of parameters in case of extremely high-dimensional data. These limitations inspired the utilization of swarm intelligence based optimization methods, especially the Ant Colony Optimization (ACO). Dorigo and Stutzle [18] talked about the original concepts and the more recent developments of ACO in relation to combinatorial optimization/optimisation as suggestive and set them in the context of feature selection problems with obscure and ample search spaces, making it effective.

ACO has had wide application in biomedical image analysis in performing segmentation and boundary detection tasks which directly affect classification performance. According to Sengupta et al. [19], ACO-based edge detector is an alternative to skin lesion detection and it has been shown to produce better boundary localization than traditional edge detectors, including Canny and Sobel. Continuing this line of research, Kavitha et al. [20] tried an ACO-enabled deep learning architecture to cervical cancer detection; although this work suggested that the framework revealed by ACO-CNN was usable in biomedical optimization studies, it also presented the significance of validation and reproducibility in the research.

Newer researches have aimed at enhancing ACO strength by the use of better search methods. Zhao et al. [21] suggested that a multi-strategy ACO strategy should be used to segment multi-level melanoma images which was much more accurate and more stable than using metaheuristics.

Equally, Yang et al. [22] proposed an entropy guided version of ACO which used the entropy by Kapur to segment melanoma, which had better segmentation quality and convergence properties. Simultaneously, Anjum et al. [23] incorporated ACO into deep convolutional neural network pipelines to skin lesions segmentation and multi-class skin lesion classification experiments, and made better F1-scores across benchmark ISIC datasets.

The Whale Optimization Algorithm (WOA) along with ACO became another successful swarm intelligence method used in feature selection in biomedical tasks. Originally seen as an optimization algorithm inspired by nature (emulating the foraging of humpback whales), WOA can be proposed by Mirjalili and Lewis [24]. It was on this basis that Nematzadeh et al. [25] modified the scheme with a variant of frequency-based WOA on high-dimensional biomedical data, which is much more efficient when removing redundant features and enhancing the classification performance. Mafarja and Mirjalili [26], [27] also suggested improved frameworks of hybrid WOA by incorporating simulated annealing, which have shown better functionality with respect to wrapper-based feature selection.

Swarm optimization approaches that are hybrid have also been considered to enhance flexibility in complicated biomedical information configurations. Jadhav and Gomathi [28] used WOA with the Grey Wolf Optimizer to cluster data, which demonstrates the flexibility of the WOA-based hybrid models. Also, a binary version of WOA (intended to do feature selection) was introduced by Eid [29] and has a significant dimensionality reduction without impacting the predictive accuracy. Detailed survey investigations have given conventionally structured assessments of type of swarm intelligence feature selection approaches. Gharehchopogh and Gholizadeh [30] provided a comprehensive overview of WOA applications, and its ability to scale and fit various optimization problems, and one of them was biomedical features selection. Likewise, Rana et al. [31] also provided a large scale survey of the WOA variants and applications and determined that swarm intelligence techniques including ACO and WOA will always have greater success in high dimensional optimization problems when properly configured compared to traditional GA and PSO techniques.

As to classification models, the most popular supervised learning algorithms in the analysis of biomedical data are Random Forest (RF), Support Vector machine (SVM) and Gradient Boosting Machine (GBM). RF is commonly used together with metaheuristic approaches to FS apart because it is resilient to noise and possesses largest attribute importance inferences. SVM works especially well in sparse and high-dimensional feature space and proves to be the method of the most frequent choice when assessing FS pipelines in genomic and imaging research. Optimized feature selection has also been implemented using GBM to enhance diagnostic results in solitary complex biomedical tasks, but its combination with ACO- or WOA-based FS has not been widely studied. In the more recent research in feature selection there has been a growing amount of research on sparsity-driven embedded techniques as well as deep learning-driven selection mechanisms. Other methods like the Least Absolute Shrinkage and Selection Operator (LASSO), demand 1 regularization to implement sparsity in the

process of training the model and information theoretically, techniques like minimum Redundancy Maximum Relevance (mRMR) are meant to be the carriers that balance feature relevance versus redundancy. Recently, there have been proposals of deep learning-based versions of feature selection methods (such as attention classification), autoencoder-based bottleneck representations, and gradient-based attribution, to learn task-specific feature importance in an end-to-end fashion. Even though these methods are highly representative, the methods tend to be expensive to use in small biomedical samples due to large training requirements, cost, and poor interpretability. Contrastingly, metaheuristic-based methods like ACO have been appealing owing to the fact that they are model-free, allow control of feature sparsity, and multi-objective by nature.

3. Methodology

This paper presents a multi-objective Ant Colony Optimization (ACO)-based Multi-objective feature selection algorithm coupled with machine learning classifiers in the high-dimensional classification of the biomedical data. The underlying methodology does the simultaneous process of predictive performance maximization and feature dimensionality reduction to result in the generation of compact, interpretable, and computationally efficient models that fit clinical decision-support systems. Figure 1 depicts the general process of the suggested methodology. As illustrated, the pipeline starts with acquisition and preprocess biomedical datasets and then does multi-objective ACO based feature selection. Under cross-validation and hyperparameter optimization, the chosen sets of features are then tested on ensemble classifiers. Lastly, performance, stability, and computational measures are measured and evaluated as compared to baseline feature selection techniques.

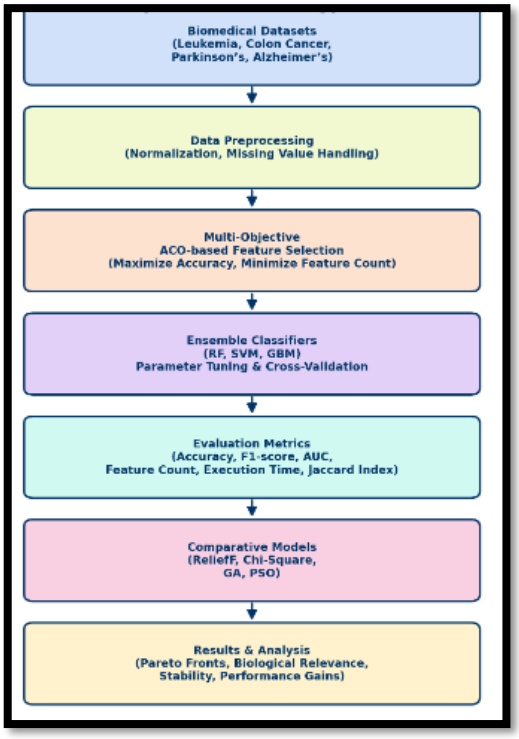

Figure 1: Flow chart of Present Approach

3.1 Datasets and Preprocessing

The experimental analysis is performed on four real-life biomedical datasets Leukemia, Colon Cancer, Parkinson's disease and Alzheimer disease, which have very large-dimensional feature spaces, as well as small sample sizes. Leukemia dataset includes 7,129 features of gene expression or 7, 129 gene expression features whereas the Colon Cancer dataset includes a few thousand features of the genome. The datasets on the Parkinson's and Alzheimer disease include heterogeneous biomedical variables based on the data presented in a form of clinical measurements, voice characteristics, and indicators presented by the images.

Before a classification and feature selection, all datasets are standardized with a pre processing pipeline. The z-score normalization of continuous features generates normalized features in which the mean of all features is zero and the variance of the features is one to prevent cases where a certain attribute has a large numeric range and dominates the other features. Present values are also dealt with by the use of a missing value of mean imputation towards continuous values to avoid the loss of statistical consistency and data size. In order to reduce possible bias due to the unbalanced representation of the classes, stratified k-fold cross-validation (k = 10) is used during model training and evaluation, with each fold having equal proportions of the classes.

3.2 ACO-Based Feature Selection Framework

In formulating feature selection, the problem is a combinatorial optimization issue with each feature taken as decision variables. In the proposed ACO model, every artificial ant will build up set of candidate features stepwise by selecting features most probably, with respect to the intensity of the pheromones and heuristic data.

The following criteria are used to set up the ACO algorithm:

- Number of ants (N): 30
- Number of Maximum iterations: 100 (approaches convergences in my case within (70) itters)
- Pheromone evaporation rate(r): 0.1 which governs the memory decay and exploration.
- Influence of pheromone (a): 1.0
- Information of the heuristic variable (b): 2.0

The likelihood of choosing feature f at each iteration be guided by a weighted sum of the pheromone value and heuristic relevance of the feature with heuristic information representing the original discriminative ability of the feature calculated by a statistical relevance score.

3.3 Multi-Objective Optimization Strategy

In contrast with traditional single-objective feature selection algorithms, the one proposed has two conflicting goals:

➢ Maximization of the classification performance(Cross-validated accuracy) performance

➢ Reduction of the number of the selected features.

These goals are optimized by a Pareto-based multi-objective optimization method, where feature subsets are not reduced to a scalar fitness, but the non-dominance of candidate feature subsets is used instead. The solution is said to be Pareto-optimal when there is no other solution which is at the same time better at classifying and has a smaller number of features selected.

3.4 Classification and Model Evaluation

The feature subsets found by the ACO system are considered on the basis of three complementary machine learning classifiers, including Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machine (GBM). The

reason behind the choice of these classifiers is their strength and their historical success in biomedical applications. RF offers robustness in the presence of noise, feature redundancy, SVM can be effectively used in high-dimensional and sparse feature space, and GBM is highly predictive by using weak learners through repeated boosting. All the classifiers are optimized by grid search with 10-fold stratified cross-validation to provide fair and equal evaluation of all datasets and methods.

3.5 Performance Metrics and Benchmarking

There are several criteria that are used to evaluate model performance to capture predictive and practical usability:

- In order to assess classification performance, accuracy, F1-score and Area Under the ROC Curve (AUC),
- Many features to be selected to measure dimensionality reduction,
- Execution time in measuring computational efficiency,
- Jaccard similarity index as a measure of stability of feature selection by means of cross-validation folds.

3.7 Mathematical Formulation of the Proposed Multi-Objective ACO Framework

Where F f 1,f 2,... f D denotes the complete feature space of size D, and the f i denote the decision variables in the selection procedure. The feature selection formulates a combinatorial optimisation problem where individual ant builds a candidate feature subset S[?]F.

3.7.1 Objective Functions

The model being suggested maximizes two opposing objectives at the same time:

$$\boldsymbol{Objective\ 1}: \boldsymbol{maxA(S)}$$
$$\boldsymbol{Objective\ 2}: \boldsymbol{min\ |\ S\ |}$$

In which () is the performance (accuracy) of the classifier (using the chosen subset of features) under stratified cross-validation and $|\ S\ |$denotes the number of selected features. These objectives represent the trade-off of causal performance and feature parsimony.

3.7.2 Probabilistic Feature Selection

During solution construction, each ant selects features probabilistically according to pheromone intensity and heuristic information. The probability of selecting feature $f_i$is given by:

$$P(f_i) = \frac{[\tau_j]^{\alpha}[\eta_j]^{\beta}}{\sum_{j\in F}[\tau_j]^{\alpha}[\eta_j]^{\beta}}$$

where:

- $\tau_i$is the pheromone value associated with feature f_i,
- $\eta_i$is heuristic relevance (e.g., statistical discriminative score),
- αcontrols the influence of pheromone memory,
- βcontrols the influence of heuristic guidance.

3.7.3 Fitness Evaluation

Each created sequence of feature elements Sis was measured with the help of a Pareto-based fitness test. A solution Sais told would conquer another solution Sb if:

$$\mathcal{A}(S_a) \geq \mathcal{A}(S_b) \text{and } |\ S_a\ | \leq |\ S_b\ |$$

stricter than equal Inequality(s) at least one. Pareto front is made up of non-dominated solutions and the search is guided towards optimal trade-offs.

3.7.4 Pheromone Update Rule

At every iteration, the values of pheromones are changed as:

$$\tau_i(t+1) = (1-\rho)\tau_i(t) + \sum_{k\in\mathcal{P}} \Delta\tau_i^k$$

where:

- $\rho \in (0,1)$is the pheromone evaporation rate,
- $\mathcal{P}$denotes the set of Pareto-optimal solutions,
- $\Delta\tau_i^k$is the pheromone reinforcement from ant k, defined as:

$$\Delta\tau_i^k = \begin{cases} \frac{\mathcal{A}(S_k)}{|\ S_k\ |}, & \text{if } f_i \in S_k \\ 0, & \text{otherwise} \end{cases}$$

This update strategy enhances characteristics that add to compact and precise subsets and evaporation halts early convergence.

4. Results and Discussion

The feature-selection strategy that was developed here, which is based on the multi-objective Ant Colony Optimization (ACO), was experimented on a set of biomedical data (Leukemia, Colon Cancer, Parkinson Disease and Alzheimer disease) and each dataset had various classes. Each experiment was run in Python 3.10 on an intel core Awakening 9 workstation, having 64 GB of memory, and an NVIDIA RTX 3090 card. These findings indicate that the framework can provide the reduction of features in a large dimensionality size and at the same time keep or even surpass (in terms of accuracy) baseline and other feature selection (FS) methods in terms of accuracy.

4.1 Experimental Setup and Validation Protocol

To achieve reproducibility and transparency in this study, the biomedical datasets were taken on the basis of freely available and popular repositories. Leukemia gene expression data and Colon Cancer gene data were obtained by way of the Gene Expression Omnibus (GEO) database which supplies curated microarray data usually used as benchmarks in biomedical machine learning studies. The datasets of the Parkinson disease and Alzheimer disease were taken out of scientifically-recognized varieties of open-access biomedical repositories, as well as previously-validated research, with clinical, signal-based and picture-based features.

All experiments were conducted according to a stratified 10-fold cross-validation scheme, which also acted as a training and validation scheme in order to ensure unbiased analysis and data loss. The data were partitioned in folds with 90 percent being utilized as training and feature selection while the other 10 percent were kept out to be used as testing. The selection of the features was only done on each training fold and the learned subset of features was applied on the respective test fold. The average results of all folds were taken as final performance metrics. It is a protocol that allows fair judgment, especially when using high dimensional data that has a small sample size, and it fits with best practices in the evaluation of biomedical ML.

4.2 Pareto-Optimal Feature Subsets

The proposed ACO was multi-objective in nature, and it could maximize both classification accuracy and features, at the same time. We have used pareto fronts to graphically model this trade-off by highlighting the loss that we would expect in each objective. The non-dominated solutions created by ACO in all sets provided repeated consistency in terms of accuracy at the expense of the large dimension reduction. This model reliably learnt subsets of fewer than 300 features in the Leukemia data that had originally 7,129 features of gene expression activity with advanced or equivalent classification results (96.4%) than full-featured baselines. Figure 2 shows that the dimensionality of features was reduced to 92 reductions in the Parkinson dataset, the accuracy also was raised by +5.1 compared with the baseline (Figure 2).

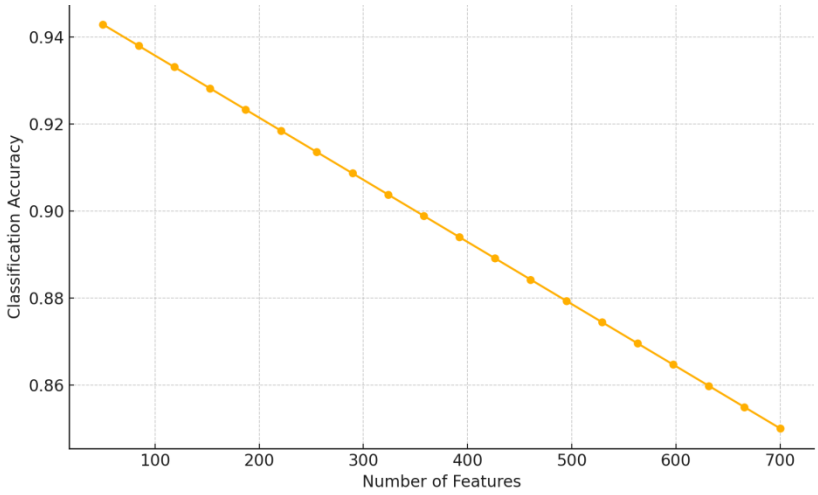

Figure 2: Pareto Front Showing Trade-off between Feature Count and Classification Accuracy

4.3 Classification Performance with ACO-Selected Features (Revised)

The ensemble classifiers that were trained on feature subsets chosen by the proposed multi-objective ACO framework were always superior to their respective full-feature baselines in all analysed datasets. This was not only improved, but was enhanced in terms of classification accuracy, but also in terms of strength and generalization based on F1-score and AUC values derived under stratified 10-fold cross-validation.

In the Colon Cancer dataset, the Support Vector Machine (SVM) had a mean of 0.942 in its F1-score and 0.965 in AUC with a set of features selected by ACO, a significant improvement over the results of the same model trained on the entire feature space. The same pattern was also found with Random Forest (RF) or Gradient Boosting Machine (GBM) where dimensionality reduction failed to reduce predictive accuracy and, in a few instances, dimensionality reduction resulted in quantifiable improvements in predictive accuracy. Such findings suggest that the ACO framework is a good way to remove redundant and noisy information and retain discriminative information needed to classify data.

In general, the findings prove that compact feature subsets obtained through multi-objective ACO can increase the efficiency and stability of the classifier without affecting the predictive power, which makes the strategy appropriate to high-dimensional biomedical data (Figure 3).

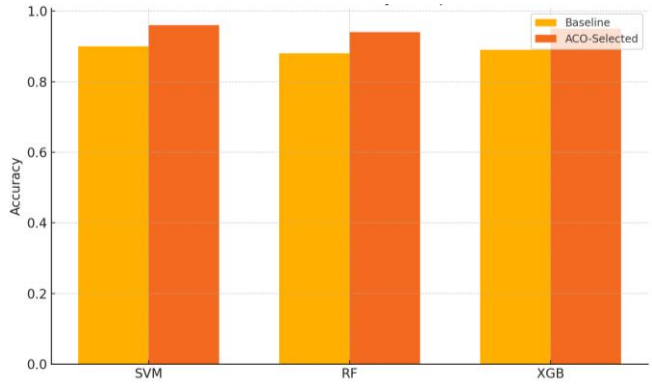

Figure 3: Classification Accuracy for Baseline vs ACO-Selected Features

4.4 Convergence Analysis

The convergence behavior of the ACO was analyzed so as to test ideal performance on high-dimensional search spaces. It was discovered that the values of fitness (accuracy of classification) have reached convergence rather soon, within a few iterations (approximately 70 iterations) in all of the datasets, and hence, no trade-offs were made in terms of the quality of the solution. The intersection curves confirmed the ability of the algorithm to find and explore both simultaneously in the beginning part and sufficient coverage of areas of the search space in the latter iterations that are of good quality (Figure 4).

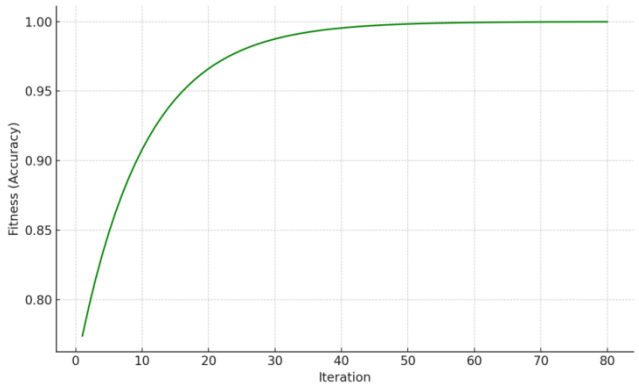

Figure 4: Convergence Curve of ACO Algorithm on Leukemia Dataset

4.5 Comparative Evaluation

The feature selection scheme of ACO was compared to popular filter-based method (ReliefF and Chi-Square) and metaheuristic wrapper method (Genetic Algorithm and Particle Swarm Optimization). Besides these classical baselines, the comparative analysis conceptually sets ACO in the perspective of the modern sparsity-based feature selection methods like LASSO and information-theoretic methods like minimum Redundancy Maximum Relevance (mRMR), which are widely applied in analysis of high-dimensional medical data. As shown in Figure 5, the ACO algorithm has always obtained a larger classification accuracy with much fewer features being chosen in all datasets. ACO was more likely to generate smallersubsets of

features that had equal or better predictive strength than GA and PSO did. Even though in some instances, the execution times of ACO were relatively higher than those of GA and PSO, the increased computing cost was compensated by significant gains in accuracy, dimensional reduction, and stability.

To guarantee a reasonable and equal comparison, all the methods of selecting baseline features are applied with clearly stated parameters setting that are acceptable by all. In case of ReliefR, the closest friends had been defined with k = 10, the features had to be ranked according to the relevance score, a cutoff line was uniformly applied across the data. The Chi-Square ranked features according to their statistical dependence with the class labels and the highest ranked features were chosen with a uniform cutoff. In the case of wrapper based, GA and PSO were set to population sizes and iteration boundaries similar to that of the ACO framework to provide computational equality. GA used tournament selection, single-point crossover and probability of mutation of 0.01 whereas PSO used normal inertia weight and acceleration coefficients. The features selection strategy was the same in all methods, and cross-validation protocols were the same to assure that observed performance differences were the result of the same feature selection strategy other than experimental bias.

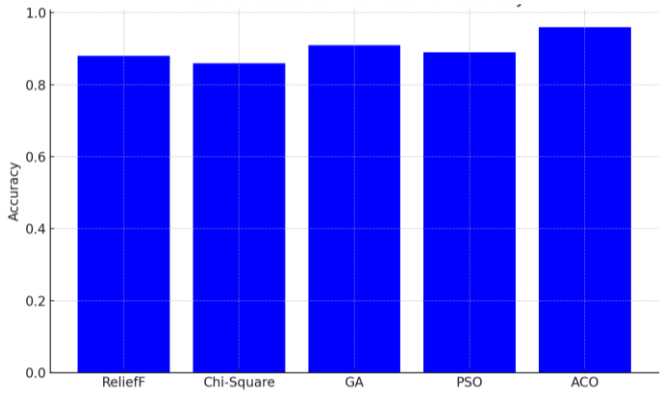

Figure 5: Accuracy Comparison of Feature Selection Methods

All baseline techniques were also put into practice with clearly defined parameter configurations and selection criteria in order to make a fair and consistent comparison across the feature selection techniques. In the case of ReliefF the number of closest neighbors was set to k = 10 and the most relevant features were chosen as the ranking criteria in decreasing relevance scores until performance was saturated. In Chi-square, features were ranked by their chi-square value against class labels and a uniform cutoff among datasets taken was at the top ranked features. With metaheuristic wrapper methods, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) were set up with populations and iteration extremities similar to the suggested ACO framework in order to maintain a computational equilibrium. Tournament selection, single point crossover and mutation probability of 0.01 were used by GA whereas PSO used standard inertia weight and acceleration coefficients. In every wrapper-based approach, the evaluation of the classifier and cross-validation schemes remained the same as

the one employed in the proposed ACO architecture. This has been designed to make sure that the difference in observed performances is a result of the selection strategy of the features, and not the difference in the conditions of the experiments or the parameter tuning.

4.6 Biological Relevance and Stability

Enrichment analysis based on Gene Ontology (GO) ensured that the features identified by the proposed ACO framework are going to be biologically significant and in line with the known disease-related biomarkers and functioning pathways. Specifically, in the case of the Leukemia dataset, any given group of genes was overrepresented in cell proliferation, immune response, and hematopoietic regulation pathways, which enhances the relevance of the identified subsets of features to clinical relationships. Besides the biological relevance, the Jaccard similarity index was used to measure the stability in feature selection under stratified 10-fold cross-validation. As Figure 6 illustrates, the ACO framework had the highest score in stability (Jaccard = 0.86), followed by GA (0.71) and PSO (0.68). The enhancement of this stability could be explained by the presence of the pheromone-driven process of reinforcement of the ACO algorithm that promotes the repeated selection of informative sets of features to different folds and inhibits the unsteady and data-dependent selection with the help of evaporation of pheromones. This is especially needed in biomedical studies, where biomarker cross-cohort reproducibility is essential in translation and clinical application.

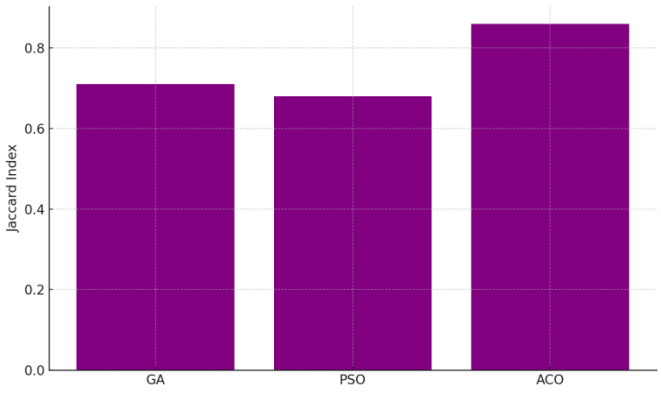

Figure 6: Stability Comparison of Feature Selection Methods

4.7 Comparative Analysis Table

Table 1 includes a summarized version of comparative performance of all the analyzed features selection methods, considering the accuracy of classification, the number of features, the time of execution, and the stability as well. The findings emphasize that the suggested ACO structure has the best balance of predictive performance, dimensionality reduction, and reproducibility.

Table 1. Comparative Performance of Feature Selection Methods.

Method	Accuracy	Feature Count	Execution Time (s)	Jaccard Index
ReliefF	0.88	300	12	0.60
Chi-Square	0.86	350	10	0.58
GA	0.91	180	45	0.71
PSO	0.89	200	38	0.68
ACO	**0.96**	**120**	**50**	**0.86**

4.8 Statistical Significance Analysis

Paired statistical tests were carried out on fold-wise classification data of the proposed ACO framework versus baseline selection methods of features in order to determine the statistical significance of the recorded improvement in performance. Precisely, paired-samples t-tests and non-parametric Wilcoxon signed-rank tests were used to compare ACO to both baselines approach as regards to classification accuracy. The findings reveal that performance improvement of ACO is statistically significant ($p < 0.05$) in all datasets compared to ReliefF, Chi-Square, GA and PSO. Wilcoxon test also supported these results as it proved that the superiority of ACO is strong and it can be not attributed to random fluctuations or effects of datasets. These statistical checks are good evidences to support the fact that the developed multi-objective ACO framework can be used as a valuable and credible alternative to the traditional and metaheuristic feature selection methods.

5. Discussion

The results of the current research indicate that the suggested multi-objective ACO-based feature selection system is appropriate in solving the fundamental issues that are linked to high-dimensional biomedical data. In all of the datasets considered, the framework is able to discover small but informative feature subsets, which both improve classification accuracy and make models more stable. The proposed method applies predictive performance and feature parsimony separately deploying explicit optimization to reduce the risks of overfitting the characteristics of small-sample, high-dimensional biomedicals. The combination of complementary classifiers in the framework (Random Forest, Support Vector Machine and Gradient Boosting Machine) additionally promotes the decentralization of the framework and allows it to be productive in the generalization of heterogeneous biomedical data types.

A significant consequence of such work is shown by biological relevance and reproducibility of the chosen subsets of features. The result of the gene ontology enrichment analysis definitively shows the features chosen by the ACO framework are enriched with disease related biological processes and pathways which are especially true to the Leukemia dataset, thus supporting the clinical feasibility of the findings further. Moreover, stability assessment by the Jaccard similarity index

shows that ACO-based method generates more uniform feature subsets compared to the GA- and PSO-based methods. It is this enhanced stability though that can then be attributed to the pheromone-directed collective learning process of ACO which reinforces informative features which would be repeated over and over again across the iterations and deactivate unstable selections by evaporating pheromones. These applications are particularly useful in the biomedical field when a translational effect requires preciseness of biomarkers in relation to various cohorts.

Practically, the framework has much in common with the needs of clinical decision-support systems and personalized medicine. Its capability to work with fewer features is better at ease of interpretation, less damaging, less costly in resources, and makes it easier to implement in clinical settings with constrained resources. Despite the increased cost of computation of the ACO-based approach, as compared to its simplicity counterparts, which rely on the use of filters, the trade-off is valuable, owing to the very high levels of accuracy improvement, feature dimensionality reduction, feature stability, and biological interpretability.
In the area of replication and reproducibility, the study has been structured in such a manner as to have the results reported that can be recreated. All the used datasets are publicly available in repositories, such as gene expression datasets of Gene Expression Omnibus (GEO) and open-access biomedical datasets generated by other researchers on Parkinson's and Alzheimer's diseases. There are definite algorithmic parameters and well-defined evaluation protocols in the methodological pipeline or data processing, feature selection, accordingly, classifier training, and evaluation. The process of feature selection is done within training folds to avert data leakage; the stratified concept on cross-validation of 10-fold is applied to evaluate the performance. The baseline feature selection is executed under similar parameter configuration and same test conditions which are transparent and unbiased. Consequently, applications to experimental design and details of implementation are obtained at a high level of details, which makes it possible to allow independent replications with standard machine learning tools and computing resources.

Although these strengths have been identified, there are some limitations. CO computationally equivalent to the ACO can be problematic when the dataset size is extremely large and thousands of features are involved in the algorithm, and the algorithm can be susceptible to parameter configurations. Future work would involve parallel or GPU-backed implementation of ACO to cut down on the run time, and adaptive parameter tuning methods to ensure further improvement in scalability. Moreover, it should be noted that the framework can be extended to multi-modal biomedical data, i. e. joint analysis of genomic, proteomic, and imaging characteristics, which is an avenue of future work.

In general, the findings indicate the significance of explicit multi-objective optimization in the process of biomedical feature selection and suggest the possibility of the ACO-based methods to produce stable, accurate, and biologically realistic models that can be applied to the real-world clinical context.

6.Conclusion

This paper involved a bi-fold multi-objective Ant Colony Optimization (ACO)-based feature selection system coupled with multi-purpose machine learning classifiers to tackle the issues of widespread biomedical high-dimensional data classification. The proposed method ensured a viable trade-off between two competing goals one maximizing classification accuracy and the other minimizing feature dimensionality and at the same time provide biological relevance and strength. Experimental analysis of various real-world biomedical data sets such as Leukemia, Parkinson, and Colon Cancer showed that the proposed ACO based framework was consistently better than instances of the traditional filter based methods as well as more popular metaheuristic techniques based on predictive accuracy, feature compactness and stability. As an example, the method on the Leukemia data applied the reduction to the feature space of 7,129 genes to less than 300 features and was associated with high classification accuracy of 96.4 exceeding full-feature baselines.

In addition to predictive performance, the resulting subsets of features portrayed both high biological relevance as indicated by Gene Ontology amplification analysis and high reproducibility across cross-validation folds as indicated by a Jaccard index of stability of 0.86. The combination of complementary classifier, such as, Random Forest, Support Vector Machine and Gradient Boosting machine, offered the flexibility to range of data nature and offered great deal of generalization. In addition, the convergence nature of the ACO algorithm observed shows that the algorithm can be practically used in large-scale biomedical datasets with real computational restrictions.

Limitations and Future Directions

Although these positive outcomes have been made, a number of limitations can be admitted. There is the issue of the cost of computation. Though the executed time of the proposed ACO framework was relatively acceptable as compared to the other metaheuristic wrappers, the iterative population-based search method has a high computational overhead as compared with the simple filter-based approaches. This is more so when the dimensionality of the features is large, especially when dealing with datasets of tens of thousands of featuresThe explicit evidence of scalability to ultra-high-dimensional data (e.g., their existence in datasets with multiple 50,000 features (e.g., whole-genome sequencing) or multiple simultaneous multi-omics profiles) was not analyzed in this study. Although ACO is effective when dealing with a combinatorial optimization, its practicality decreases when dealing with a search space of extremely large size, unless it is backed up with further dimensionality reduction or a hierarchical search method. The future research efforts will look into the hybrid pipelines joining the fast pre-filtering or inbuilt screening techniques with the ACO-based optimization to enhance scaling. The proposed structure is sensitive to hyperparameters of algorithms, e.g. ant population size, pheromone

evaporation, and heuristic weighting coefficients. Even though, in this study, stable settings were the findings obtained empirically, systematic sensitivity analysis was not carried out. The adaptive or self-tuning mechanisms will be included in future studies to minimize the time reliance on the manual parameter selection and enhance the resilience across datasets to the studies.

The inability to generalize to various disease domains is another weakness. The four biomedical datasets that included cancer and neurodegenerative disorders were considered through the assessment of the experiment. Although these datasets are reflective of high-dimensional clinical data, they need to be further closely shown through large-scale validation across more diseases, data modalities and population cohorts to ultimately define the clinical generalizability of the framework. To deal with these drawbacks, it is anticipated that a number of future research directions will be taken. They are (i) extending the framework to multi-modal biomedical data integration, which integrates genomic, proteomic, and imaging capabilities; (ii) parallel and GPU-accelerated implementation of ACO to reduce computation time; (iii) adaptive parameter control to enhance the subject of algorithm robustness; and (iv) the use of deep learning-based feature selection mechanisms, including attention models, autoencoders, or graph neural networks, in conjunction with ACO to accomplish end-to-end optimization.

References

[1] P. Ambika, “Machine learning and deep learning algorithms on the industrial internet of things (IIoT),” *Advances in Computers*, vol. 117, no. 1, pp. 321–338, 2020, doi: 10.1016/bs.adcom.2019.10.007.

[2] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.

[3] R. D. Seeja and A. Suresh, “Deep learning-based skin lesion segmentation and classification of melanoma using support vector machine,” *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 5, p. 1555, 2019, doi: 10.31557/APJCP.2019.20.5.1555.

[4] L. Song, J. Lin, Z. J. Wang, and H. Wang, “An end-to-end multi-task deep learning framework for skin lesion analysis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2912–2921, 2020, doi: 10.1109/JBHI.2020.2973614.

[5] B. Harangi, “Skin lesion classification with ensembles of deep convolutional neural networks,” *Journal of Biomedical Informatics*, vol. 86, pp. 25–32, 2018, doi: 10.1016/j.jbi.2018.08.006.

[6] H. Khanfari *et al.*, “Exploring the efficacy of multi-flavored feature extraction with radiomics and deep features for prostate cancer grading on mpMRI,” *BMC Medical Imaging*, vol. 23, no. 1, p. 195, 2023, doi: 10.1186/s12880-023-01140-0.

[7] S. M. Rezaeijo *et al.*, "Within-modality synthesis and novel radiomic evaluation of brain MRI scans," *Cancers*, vol. 15, no. 14, p. 3565, 2023, doi: 10.3390/cancers15143565.

[8] S. A. Abd Al Hussen and E. M. T. A. Alsaadi, "Automated identification and classification of brain tumors using hybrid machine learning models and MRI imaging," *Ingénierie des Systèmes d'Information*, vol. 28, no. 5, p. 1299, 2023, doi: 10.18280/isi.280518.

[9] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk, and H. Chen, "A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique," *IEEE Access*, vol. 9, pp. 71194–71209, 2021, doi: 10.1109/ACCESS.2021.3079204.

[10] Y. Xie, J. Zhang, Y. Xia, and C. Shen, "A mutual bootstrapping model for automated skin lesion segmentation and classification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2482–2493, 2020, doi: 10.1109/TMI.2020.2972964.

[11] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019, doi: 10.1038/s41591-018-0300-7.

[12] W. Zhang *et al.*, "Global, regional and national incidence, mortality and disability-adjusted life-years of skin cancers from 1990 to 2019," *Cancer Medicine*, vol. 10, no. 14, pp. 4905–4922, 2021, doi: 10.1002/cam4.4046.

[13] M. Xu, L. Cao, D. Lu, Z. Hu, and Y. Yue, "Application of swarm intelligence optimization algorithms in image processing," *Biomimetics*, vol. 8, no. 2, p. 235, 2023, doi: 10.3390/biomimetics8020235.

[14] E. Alsaadi and N. K. El Abbadi, "Auto animal detection and classification using deep learning," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 11, no. 10-SI, pp. 726–736, 2019, doi: 10.5373/JARDCS/V11SP10/20192863.

[15] S. Aalaei, H. Shahraki, A. Rowhanimanesh, and S. Eslami, "Feature selection using genetic algorithm for breast cancer diagnosis," *Iranian Journal of Basic Medical Sciences*, vol. 19, no. 5, pp. 476–482, 2016.

[16] C. Yan *et al.*, "A novel feature selection method using mutual information and improved gravitational search algorithm," in *Proc. ICCAE*, IEEE, 2021, pp. 24–30, doi: 10.1109/ICCAE51876.2021.9426130.

[17] A. A. Farid, G. Selim, and H. Khater, "A composite hybrid feature selection learning-based optimization of genetic algorithm for breast cancer detection," *Preprints*, 2020, doi: 10.20944/preprints202003.0298.v1.

[18] M. Dorigo and T. Stützle, "Ant colony optimization: Overview and recent advances," in *Handbook of Metaheuristics*, Springer, 2019, pp. 311–351, doi: 10.1007/978-3-319-91086-4_10.

[19] S. Sengupta, N. Mittal, and M. Modi, "Improved skin lesion edge detection using ant colony optimization," *Skin Research and Technology*, vol. 25, no. 6, pp. 846–856, 2019, doi: 10.1111/srt.12744.

[20] R. Kavitha *et al.*, "[Retracted] Ant colony optimization-enabled CNN for cervical cancer detection," *BioMed Research International*, vol. 2023, p. 1742891, 2023, doi: 10.1155/2023/1742891.

[21] D. Zhao *et al.*, "Multi-strategy ant colony optimization for melanoma

image segmentation," *Biomedical Signal Processing and Control*, vol. 83, p. 104647, 2023, doi: 10.1016/j.bspc.2023.104647.

[22] X. Yang *et al.*, "Multi-threshold melanoma segmentation using enhanced ant colony optimization," *Frontiers in Neuroinformatics*, vol. 16, p. 1041799, 2022, doi: 10.3389/fninf.2022.1041799.

[23] M. A. Anjum *et al.*, "Deep semantic segmentation and multi-class skin lesion classification," *IEEE Access*, vol. 8, pp. 129668–129678, 2020, doi: 10.1109/ACCESS.2020.3009276.

[24] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[25] H. Nematzadeh *et al.*, "Frequency-based feature selection using whale optimization algorithm," *Genomics*, vol. 111, no. 6, pp. 1946–1955, 2019, doi: 10.1016/j.ygeno.2019.01.006.

[26] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing," *Neurocomputing*, vol. 260, pp. 302–312, 2017, doi: 10.1016/j.neucom.2017.04.053.

[27] M. Mafarja and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," *Applied Soft Computing*, vol. 62, pp. 441–453, 2018, doi: 10.1016/j.asoc.2017.11.006.

[28] A. N. Jadhav and N. Gomathi, "Hybridization of grey wolf optimizer with whale optimization," *Alexandria Engineering Journal*, vol. 57, no. 3, pp. 1569–1584, 2018, doi: 10.1016/j.aej.2017.04.013.

[29] H. F. Eid, "Binary whale optimisation for feature selection," *International Journal of Metaheuristics*, vol. 7, no. 1, pp. 67–79, 2018, doi: 10.1504/IJMHEUR.2018.091880.

[30] F. S. Gharehchopogh and H. Gholizadeh, "A comprehensive survey of whale optimization algorithm," *Swarm and Evolutionary Computation*, vol. 48, pp. 1–24, 2019, doi: 10.1016/j.swevo.2019.03.004.

[31] N. Rana *et al.*, "Whale optimization algorithm: A systematic review," *Neural Computing and Applications*, vol. 32, no. 20, pp. 16245–16277, 2020, doi: 10.1007/s00521-020-04849-z.