Metaheuristic Algorithms with Applications

www.maa.reapress.com

Metaheur. Algor. Appl. Vol. 3, No. 1 (2026) 91–110.

Paper Type: Original Article

Harris Hawks Optimizer for High-Dimensional Biomarker Discovery: A Statistically Rigorous Comparative Study of Eleven Metaheuristic Algorithms on Leukemia Gene Expression Data

Zeinab Parandavar[1,*], Behnam Ghahramankhani[2], Mohammad Hossein Khodabandeh[2]

[1] Department of Computer Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran; z.parand.1997@gmail.com.

[2] Department of Computer and Information Technology Engineering, Ab.C., Abhar Branch, Islamic Azad University, Abhar, Iran; ghahramankhani@iau.ac.ir; mohammad.hosseinkhodabandeh@iau.ir.

Citation:

Received: 29 April 2025 *Revised: 07 July 2025* *Accepted: 12 October 2025*	Parandavar, Z., Ghahramankhani, B., & Khodabandeh, M. H. (2026). Harris Hawks optimizer for high-dimensional biomarker discovery: A statistically rigorous comparative study of eleven metaheuristic algorithms on Leukemia Gene expression data. *Metaheuristic Algorithms with Applications, 3*(1), 91-110.

Abstract

The rapid accumulation of high-throughput omics data has transformed disease diagnosis. Yet, the extreme dimensionality of gene expression profiles continues to impose severe computational and statistical obstacles for biomarker discovery. Although many nature-inspired metaheuristic algorithms have been proposed for Feature Selection (FS), the literature lacks rigorous, large-scale, statistically validated comparative studies on noisy, epistatic biological landscapes. This study presents an exhaustive evaluation of the Harris Hawks Optimizer (HHO) against ten state-of-the-art metaheuristics for binary FS on the benchmark Leukemia (ALL versus AML) microarray dataset. We benchmark a Binary HHO (BHHO) wrapper built around a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel against ten competitors under strictly unified budgets of 30 agents, 100 iterations, 30 independent runs, and stratified 10-fold cross-validation. We assess algorithms on accuracy, F1-score, Matthews Correlation Coefficient (MCC), area under the curve, subset size, and runtime, and verify significance with Wilcoxon signed-rank and Friedman tests. HHO attained 98.61% accuracy, an Area Under the Curve (AUC) of 0.992, and an MCC of 0.971 using only 12.4 genes on average, significantly outperforming every competitor at $p < 0.05$ and ranking first under the Friedman test with a mean rank of 1.00. Biological enrichment analysis confirmed that the selected genes are established leukemia drivers. The dynamic escape energy mechanism of HHO provides an exceptional exploration-exploitation balance for high-dimensional biological FS, offering a reliable, interpretable, and clinically translatable biomarker discovery tool.

Keywords: Feature selection, Harris hawks optimizer, Gene expression, Metaheuristics, Biomarker discovery, Leukemia classification, Wrapper methods.

Corresponding Author: z.parand.1997@gmail.com

https://doi.org/10.48313/maa.vi.81

1 | Introduction

The advent of DNA microarrays and next-generation Ribonucleic Acid (RNA) sequencing has fundamentally reshaped biomedical research by enabling the simultaneous quantification of expression levels for tens of thousands of genes in a single assay. This technological leap has opened unprecedented opportunities for molecular taxonomy, prognostic stratification, and personalized therapy selection. However, it has also introduced the well-documented curse of dimensionality, where the number of measured features vastly exceeds the number of available biological samples [1]. In such high-dimensional spaces, Machine Learning (ML) classifiers become highly susceptible to overfitting, variance inflation, and poor generalization to unseen patients. At the same time, the computational burden of model training grows rapidly with input-space dimensionality [2]. From a biological standpoint, only a small fraction of the transcriptome is differentially expressed or mechanistically relevant to a specific disease phenotype, meaning most measured genes act as noise that obscures the true discriminative signal. Feature Selection (FS) is therefore not merely a computational preprocessing step but a critical necessity for identifying robust, interpretable biomarkers that can drive clinical diagnosis and targeted therapy.

FS techniques are conventionally categorized into filter, wrapper, and embedded approaches. Filter methods such as analysis of variance (ANOVA) F-tests and mutual information criteria are computationally efficient because they evaluate features independently of any classifier. However, they ignore the complex epistatic interactions between genes that often underlie disease phenotypes [3]. Wrapper methods, in contrast, embed a ML model within a search algorithm to evaluate candidate feature subsets, consistently yielding superior predictive performance because they optimize the subset directly for the downstream classifier [4]. The central drawback of wrapper methods is computational: An exhaustive evaluation of all 2^d possible subsets, where d denotes the number of genes, is an NP-hard problem that becomes intractable even for moderate d. To navigate this vast and rugged combinatorial search space, wrapper methods are increasingly hybridized with stochastic metaheuristic optimization algorithms, which provide a favorable trade-off between solution quality and computational cost [5].

Over the past two decades, the field of metaheuristics has grown exponentially, producing a large family of nature-inspired algorithms. Classical evolutionary and swarm intelligence methods, including the Genetic Algorithm (GA) [6], Particle Swarm Optimization (PSO) [7], and Differential Evolution (DE) [8], have been applied extensively to bioinformatics problems with varying degrees of success. More recently, a new wave of algorithms inspired by diverse biological and physical phenomena has emerged, including the Grey Wolf Optimizer (GWO) [9], Whale Optimization Algorithm (WOA) [10], Salp Swarm Algorithm (SSA) [11], Artificial Bee Colony (ABC) [12], Marine Predators Algorithm (MPA) [13], Slime Mould Algorithm (SMA) [14], and Aquila Optimizer (AO) [15]. Each algorithm introduces distinct mathematical operators that purportedly improve the balance between global exploration and local exploitation, and their foundational publications report strong performance on continuous benchmark functions. Despite this proliferation, a critical and persistent gap remains in the literature: the absence of standardized, strictly controlled, large-scale comparative studies tailored to high-dimensional biological FS [16].

Many newly proposed algorithms are validated only against basic unimodal or multimodal mathematical benchmark functions, or compared against a limited subset of older algorithms such as GA and PSO, making it difficult to judge whether the claimed mathematical novelties translate into tangible performance gains on real biomedical data. Furthermore, it remains debatable whether the complex operators of recent algorithms retain their efficacy once discretized through transfer functions and applied to the noisy, sparse, and strongly epistatic landscapes characteristic of gene expression data [17]. This concern is amplified by recent critical reviews arguing that several modern Swarm Algorithms (SA) are mathematically equivalent to existing methods under certain parameterizations, raising legitimate questions about whether algorithmic novelty corresponds to genuine performance improvements in applied settings [18]. A rigorous, multi-metric, statistically validated comparison is therefore urgently needed to help practitioners select appropriate optimizers for biomarker discovery.

To address this gap, the present study conducts a rigorous and expansive comparative evaluation of the Harris Hawks Optimizer (HHO) [19] against ten prominent metaheuristic algorithms. HHO is the primary algorithm of interest because its unique dynamic escape energy mechanism theoretically provides a highly adaptive, non-linear balance between global exploration and local exploitation. These traits are critical for avoiding local optima in gene selection. The ten competitor algorithms were carefully chosen to represent different evolutionary eras and mechanisms: the classics (GA, PSO, DE), mid-era SA (ABC, GWO, WOA, SSA), and recent high-novelty algorithms (MPA, SMA, AO). The broader motivation for this work echoes recent efforts to pair metaheuristic search with strong classifiers for cancer diagnosis, as exemplified by optimized boosting frameworks for gastric cancer classification that combine reinforcement learning with water-cycle optimization [20]. Unlike previous studies that rely on accuracy alone, this work adopts a multi-metric evaluation strategy. It complements the computational findings with biological pathway enrichment analysis to verify that the selected features are genuine disease drivers rather than statistical artifacts.

The main contributions of this manuscript are summarized as follows: 1) a Binary HHO (BHHO) wrapper framework is formulated and rigorously benchmarked against ten distinct metaheuristics under strictly unified experimental conditions, including population size, iteration budget, objective function, and cross-validation folds, 2) beyond standard accuracy, the algorithms are evaluated using the Matthews Correlation Coefficient (MCC), F1-score, area under the Receiver Operating Characteristic (ROC) curve Area Under the Curve (AUC), subset size, and wall-clock runtime, providing a holistic view of classifier performance that is particularly relevant under the class imbalance typical of medical datasets [21], 3) a biological pathway enrichment analysis is conducted on the gene subsets selected by the top-performing algorithms to demonstrate that the selected features correspond to clinically established biomarkers rather than overfitting noise, and 3) a deep analysis of convergence trajectories is combined with comprehensive non-parametric statistical tests, including the multi-metric Wilcoxon signed-rank test and the Friedman test, to definitively establish statistical significance [22].

The remainder of this paper is structured as follows. Section 2 surveys the related literature on metaheuristic FS in bioinformatics and positions the present work within it. Section 3 details the theoretical foundations of HHO and the competitor algorithms. Section 4 describes the materials, dataset preprocessing, binary adaptation, objective function, and experimental setup. Section 5 presents the extensive experimental results, statistical analyses, and biological interpretations. Section 6 critically discusses algorithmic behavior, limitations, and implications, and Section 7 concludes the study with future research directions.

2 | Literature Review

FS in high-dimensional bioinformatics has attracted sustained attention for more than two decades, driven by the recurring statistical challenge of analyzing datasets in which the number of features routinely exceeds the number of samples by one or two orders of magnitude. Guyon and Elisseeff [2] provided the foundational taxonomic treatment of variable selection approaches, formalizing the distinction between filters, wrappers, and embedded methods and articulating why wrapper approaches, despite their computational cost, are particularly well suited to discovering synergistic feature combinations. Saeys et al. [3] later extended this taxonomy to bioinformatics, emphasizing that the structure of the biological data and the downstream classification objective should guide the choice of FS strategy. These foundational works established the conceptual scaffolding for this comparative evaluation.

Within the wrapper paradigm, metaheuristic optimizers have become the dominant search engine because they provide a tractable approximation to the NP-hard subset selection problem. Xue et al. [4] demonstrated that PSO-based multi-objective wrappers can simultaneously optimize accuracy and subset size. This finding motivated subsequent studies to adopt multi-criteria fitness functions. Hancer et al. [23] advanced this line by combining DE with rough set and relief-based filters, showing that hybrid filter-wrapper designs outperform pure wrappers on several microarray benchmarks. Al-Tashi et al. [24] applied the Moth-Flame Optimization algorithm to FS. They reported competitive results on medical datasets, while Cui et al. [25] proposed an

improved GWO variant with non-linear convergence factors that improved stability on gene expression data. Collectively, these studies confirm that algorithmic design choices, such as the shape of the convergence factor and the transfer function, materially affect FS performance on biological data.

The HHO, introduced by Heidari et al. [19], has rapidly gained traction across engineering and bioinformatics applications because of its mathematically elegant escape energy factor and its four-stage hunting scheme. Lin et al. [26] adapted HHO to binary FS using V-shaped and S-shaped transfer functions. They reported that HHO consistently outperformed GA, PSO, and GWO on a portfolio of medical datasets. Too et al. [27] proposed a chaotic variant of HHO that further improved exploration on multimodal problems, and Wang et al. [28] introduced an adaptive HHO with opposition-based learning that enhanced convergence velocity on high-dimensional benchmarks. The broader swarm-intelligence literature continues to produce novel algorithms at a rapid pace; the Trees Social Relations Optimization Algorithm [29], for instance, models the hierarchical and collective interactions of forest ecosystems and has demonstrated competitive performance on both continuous and discrete benchmarks, underscoring the practical need for the kind of head-to-head, multi-algorithm comparison undertaken here. Despite these advances, most existing HHO studies in bioinformatics compare the algorithm against only two or three baselines and rarely validate the biological plausibility of the selected gene subsets, which limits the practical interpretability of the reported gains.

A second strand of literature has raised methodological concerns about how researchers evaluate new metaheuristics. Yang [18] argued that several recently proposed SA are mathematically equivalent to well-known methods under particular parameterizations, casting doubt on whether their claimed novelties translate into measurable performance gains. Bolón-Canedo et al. [1] highlighted the absence of standardized benchmarking protocols and the prevalence of incomplete statistical reporting in FS studies. Demšar [22] provided the statistical framework, based on the Friedman test followed by post-hoc pairwise comparisons, that has since become the recommended standard for comparing multiple classifiers on multiple datasets. The present study explicitly adopts this statistical framework and extends it to multiple complementary metrics, directly addressing the methodological criticisms raised in these reviews.

Finally, a growing body of work emphasizes the importance of biological validation to complement purely computational evaluation. Huang et al. [30] surveyed wrapper FS methods for microarray classification and noted that most published studies report only predictive metrics, omitting any analysis of whether the selected genes correspond to known disease mechanisms. Shahrjooihaghighi et al. [31] demonstrated that ensemble FS methods tend to select more biologically coherent gene sets than single-algorithm approaches, and Chandra and Gupta [32] showed that multi-objective FS produces subsets that are both smaller and more biologically interpretable. The present work incorporates Gene Ontology (GO) and KEGG pathway enrichment analysis on the selected subsets, positioning biological interpretability as a first-class evaluation criterion alongside predictive performance. *Table 1* summarizes the positioning of the present study relative to representative prior work, highlighting the combination of large comparator pool, multi-metric evaluation, statistical rigor, and biological validation that distinguishes this contribution.

Table 1. Positioning of the current investigation relative to representative prior work on metaheuristic FS in bioinformatics.

Study	Algorithm focus	# Comparators	Statistical tests	Biological validation
Xue et al. [4]	PSO (multi-objective)	3	Limited	No
Demšar [22]	DE + filter hybrid	4	Wilcoxon	Partial
Hancer et al. [23]	Moth-Flame	5	Wilcoxon	No
Al-Tashi et al. [24]	Improved GWO	6	Friedman	No
Lin et al. [26]	BHHO	4	Wilcoxon	No
Shahrjooihaghighi et al. [31]	Ensemble FS	5	t-test	Yes
Chandra and Gupta [32]	Multi-objective FS	6	Friedman	Yes
This study	BHHO	10	Wilcoxon + Friedman	Yes (GO + KEGG)

Table 1 reveals two recurring limitations in the existing literature that the present study is designed to address.

First, the comparator pools used in prior studies are typically small, ranging from three to six algorithms, which severely limits the ability to draw generalizable conclusions about the relative standing of any single optimizer. Second, only a minority of studies perform biological validation of the selected gene subsets, which leaves open the critical question of whether the reported predictive gains correspond to genuine disease mechanisms or merely to statistical artifacts. By benchmarking BHHO against ten competitors from three distinct evolutionary eras, applying both the Wilcoxon signed-rank and Friedman tests across multiple metrics, and conducting GO and KEGG pathway enrichment analysis, this investigation offers a substantially more comprehensive evaluation than any listed predecessor. This combination of scale, statistical rigor, and biological interpretability is the work's principal methodological novelty.

3 | Theoretical Background

The theoretical foundations of the eleven metaheuristic algorithms evaluated in this study are laid out below. The presentation is organized in two subsections: the first subsection details the mathematical formulation of the HHO, which is the primary algorithm of interest, with particular emphasis on its dynamic escape energy mechanism and its four-stage hunting scheme. The second subsection surveys the ten competitor algorithms, grouped by their evolutionary era and mechanistic family, and consolidates their key parameters. Together, these subsections provide the algorithmic context necessary to interpret the comparative results reported in Section 5 and to understand why certain algorithms succeed or fail on the epistatic landscapes characteristic of gene expression data.

3.1 | Harris Hawks Optimizer

The HHO is a population-based metaheuristic inspired by the cooperative hunting strategies and surprise pounce behaviors of Harris' hawks in nature. The algorithm mathematically models the dynamic interaction between a swarm of hawks, which act as search agents, and a prey, which represents the current best solution. The central mechanism governing the transition between exploration and exploitation is a non-linear escape energy factor that decreases over the optimization, mirroring the prey's physical exhaustion as it attempts to escape [19]. This design is particularly attractive for high-dimensional FS because it provides a principled, automatic schedule for shifting from broad global search to focused local refinement without requiring hand-tuned decay parameters.

The escape energy E is defined as $E = 2 * E0 * (1 - t / T)$, where E0 is the initial energy state drawn uniformly from the interval [-1, 1] at the beginning of each iteration, t is the current iteration index, and T is the maximum number of iterations. The absolute value of E determines the prey's remaining capacity to escape: when $|E|$ is greater than or equal to 0.5, the prey retains sufficient energy to evade, which triggers soft besiege or soft exploration modes, whereas when $|E|$ is less than 0.5, the prey is exhausted, which triggers hard besiege or hard exploration modes. A second random number r drawn uniformly from [0, 1] further distinguishes besiege behavior from progressive rapid dives, producing four condition-dependent update rules that together encode a rich, multi-mode search strategy.

In the soft besiege regime, which occurs when $|E| >= 0.5$ and $r >= 0.5$, the hawks slowly encircle the prey. At the same time, it still has energy to escape, and the position update is given by $X(t+1) = X\_rabbit(t) - E * |J * X\_rabbit(t) - X(t)|$, where $J = 2 * (1 - r1)$ is the random jump strength of the rabbit and r1 is uniformly distributed in [0, 1]. In the hard besiege regime, which occurs when $|E| < 0.5$ and $r >= 0.5$, the prey is exhausted and the hawks perform a tight encirclement around the current best position, with the update rule $X(t+1) = X\_rabbit(t) - E * |X\_rabbit(t) - X\_mean(t)|$, where X_mean denotes the mean position of the hawk swarm. Together, these two regimes implement the algorithm's exploitative component.

The two exploration regimes incorporate Lévy flights to model the hawks' progressive rapid dives when the prey escapes. In the soft exploration regime, which occurs when $|E| >= 0.5$ and $r < 0.5$, the candidate positions Y and Z are generated as $Y = X\_rabbit(t) - E * |J * X\_rabbit(t) - X(t)|$ and $Z = Y + S * LF(D)$, where D is the problem dimension, S is a random vector of size 1 x D, and LF denotes the Lévy flight

function; the next position is set to Y if its fitness is lower than the current position's, otherwise to Z. In the hard exploration regime, which occurs when $|E| < 0.5$ and $r < 0.5$, the mean swarm position replaces the individual hawk position in the computation of Y, which sharpens the search around the swarm consensus. This multi-stage, condition-dependent updating mechanism renders HHO highly adaptive to complex, multimodal search spaces such as those encountered in gene selection.

3.2|Competitor Algorithms

To provide a comprehensive and historically layered baseline, HHO is compared against ten algorithms spanning three evolutionary eras. The classical evolutionary algorithms include GA, which evolves populations through selection, crossover, and mutation [6], and DE, which generates candidate solutions by adding the weighted difference of two population vectors to a third vector, followed by crossover and greedy selection [8]. The classical swarm intelligence algorithms include PSO, which updates agent velocities based on the personal best and global best positions [7], and ABC, which simulates the foraging behavior of honeybees through employed, onlooker, and scout bee phases [12]. These four algorithms have served as the standard baselines in FS research for nearly three decades and represent the canonical reference points for any new optimizer.

The mid-era hierarchy and behavior-based SA include GWO, which mimics the strict social hierarchy of alpha, beta, delta, and omega wolves and updates positions based on the encircling behavior of the three leading wolves [9]; WOA, which simulates the bubble-net hunting mechanism of humpback whales through spiral and shrinking-circle updates [10]; and SSA, which models the chain-like swarming and jet propulsion of salps in the ocean, with the leader salp guiding the chain and the followers adjusting toward the leader and each other [11]. These three algorithms represent the wave of swarm methods that dominated the metaheuristic literature in the mid-2010s and remain widely used in applied FS studies.

The recent high-novelty SA include MPA, which models the foraging strategies of marine predators through a combination of Lévy and Brownian movements across simulated ecosystem gradients and incorporates a memory-saving mechanism that retains the best historical position of each agent [13]; SMA, which simulates the venous network oscillations of slime mould during food foraging, with adaptive weights that regulate the trade-off between positive and negative feedback during the search [14]; and AO, which mimics the swooping, gliding, and grab-and-carry hunting mechanics of Aquila birds across four distinct foraging modes [15]. These three algorithms represent the current frontier of swarm intelligence research. They are frequently cited as state-of-the-art optimizers, which justifies their inclusion as HHO's most challenging contemporaries. *Table 2* consolidates the key characteristics and tuned parameters of all eleven algorithms used in this study.

Table 2. Algorithm characteristics and tuned parameters used in the comparative study.

Algorithm	Year	Inspiration	Key parameter(s)	Value
GA	1975	Natural selection	Crossover/mutation rate	0.8 / 0.02
PSO	1995	Bird flocking	Inertia weight w	0.9 -> 0.4 (linear)
DE	1997	Vector differences	F (scale), CR (crossover)	0.5 / 0.9
ABC	2007	Honeybee foraging	Limit (abandonment)	N * D
GWO	2014	Grey wolf hierarchy	a (convergence)	2 -> 0 (linear)
WOA	2016	Bubble-net hunting	a1, a2 (constants)	2 -> 0 / [-1,1]
SSA	2017	Salp chain	c1 (convergence)	2 * exp(-4t/T)
MPA	2020	Marine predators	FADs probability	0.2
SMA	2020	Slime mould	z (threshold)	0.03
AO	2021	Aquila hunting	Alpha, delta, beta	0.1 / 0.1 / 0.1
HHO	2019	Harris' hawks	Escape energy E0	Uniform[-1, 1]

Table 2 documents the eleven algorithms and their foundational parameters, set to the author-recommended values reported in the original publications. The deliberate use of default parameters, rather than algorithm-specific tuning, reflects a realistic deployment scenario in which a practitioner applies off-the-shelf implementations to a new dataset without extensive hyperparameter optimization. This design choice eliminates the confounding effect of unequal tuning effort, which many comparative studies have identified as a source of bias. The parameter spread also illustrates the diversity of mechanisms under comparison:

classical algorithms such as GA and PSO rely on fixed rates, whereas modern algorithms such as GWO and SSA introduce time-varying convergence factors, and the most recent algorithms such as MPA and AO incorporate memory and stochastic perturbation terms. This heterogeneity ensures that the comparative evaluation stress-tests HHO against a representative sample of the design philosophies that currently dominate the metaheuristic literature.

4 | Materials and Methods

This section describes the materials, preprocessing pipeline, problem formulation, and experimental setup underpinning the comparative evaluation. The subsections follow the experimental pipeline: the dataset and its biological context are introduced first, followed by the data preprocessing steps that make the high-dimensional microarray data tractable for wrapper-based optimization. The binary adaptation and transfer function that map continuous optimizer outputs to discrete gene-selection vectors are then formalized, after which the objective function and the Support Vector Machine (SVM) classifier used for fitness evaluation are specified. The section closes with a detailed account of the experimental setup and parameter control measures that ensure strict fairness across all eleven algorithms. This organization makes the entire experimental framework fully reproducible by an independent practitioner.

4.1 | Dataset and Biological Context

The study uses the seminal Leukemia gene expression dataset introduced by Golub et al. [33], widely regarded as a gold standard in computational biology for evaluating FS algorithms because of its high dimensionality, modest sample size, and clinical significance. The dataset includes 72 bone marrow samples characterized by Affymetrix oligonucleotide microarrays with 7,129 probe sets, partitioned into 47 cases of Acute Lymphoblastic Leukemia (ALL) and 25 cases of Acute Myeloid Leukemia (AML). Distinguishing between ALL and AML is clinically critical because the two subtypes originate from different hematopoietic lineages and require drastically different treatment protocols, with misdiagnosis carrying severe prognostic consequences. The 47-to-25 class ratio also introduces moderate class imbalance, making the dataset well suited for evaluating FS algorithms under realistic clinical conditions, since accuracy alone is a misleading metric on imbalanced data. *Table 3* summarizes the dataset's principal characteristics and the preprocessing pipeline applied before optimization.

Table 3. Characteristics of the Leukemia dataset and the preprocessing pipeline.

Property	Value
Source	Golub et al. [33] leukemia microarray
Platform	Affymetrix oligonucleotide microarray
Total probe sets (raw)	7,129
Total samples	72
Class ALL (samples)	47 (65.3%)
Class AML (samples)	25 (34.7%)
ANOVA F-test pre-filter (top-k)	1,000
Normalization	Z-score (zero mean, unit variance)
Search space after filtering	2^1,000 subsets
Train / Test split	Stratified 10-fold cross-validation

Table 3 highlights the dual challenge motivating this study: an original feature space of 7,129 genes measured across only 72 samples, yielding a feature-to-sample ratio of approximately 99 to 1 and placing the problem firmly in the small-n, large-p regime where overfitting is the dominant risk. The ANOVA F-test pre-filter reduces the search space from 2^7,129 to 2^1,000 candidate subsets, making wrapper optimization tractable while retaining the biologically relevant signal, because the most discriminative leukemia markers, such as CD33 and ZYX, are known to reside within the top 1,000 genes ranked by F-score. The z-score normalization step is essential because the raw expression values span several orders of magnitude and would otherwise cause features with large numerical ranges to dominate the SVM kernel computations. Stratified 10-fold cross-validation preserves the 47-to-25 class ratio across all folds, ensuring each training and testing partition

remains representative of the overall class distribution and that reported performance metrics are not biased by incidental class imbalance in individual folds.

4.2 | Binary Adaptation and Transfer Function

Metaheuristic algorithms are inherently designed for continuous optimization, so they require a transfer function that maps continuous agent positions to binary feature-selection vectors X in {0, 1}^d, where 1 indicates the corresponding gene is selected and 0 indicates it is discarded. The study employs the standard S-shaped sigmoid transfer function, which has been shown to preserve population diversity in high-dimensional spaces more effectively than V-shaped functions because it probabilistically flips bits in proportion to the magnitude of the continuous position rather than deterministically thresholding it [10]. Specifically, the sigmoid value is computed as S(x) = 1 / (1 + exp(-x)) for each dimension, and the binary decision is drawn as x_binary = 1 if a uniform random number in [0, 1] is less than S(x) and 0 otherwise. This stochastic binarization introduces additional exploratory pressure that is particularly valuable in the early iterations of the search, when the algorithm should sample diverse regions of the combinatorial space.

4.3 | Objective Function and Classifier

We used an SVM with a Radial Basis Function (RBF) kernel as the wrapper evaluator because of its well-documented robustness in high-dimensional, small-sample scenarios and its ability to capture non-linear class boundaries through implicit feature space expansion [24]. The SVM hyperparameters, namely the regularization constant C and the kernel width gamma, were optimized via an internal grid search within the training folds only to prevent information leakage from tuning on the test fold. The fitness function is designed to maximize classification accuracy while simultaneously penalizing large feature subsets to ensure biological sparsity, and is defined as Fitness(X) = alpha * ErrorRate(X) + (1 - alpha) * (|S| / Total_Features), where ErrorRate(X) = 1 - Accuracy(X), |S| is the cardinality of the selected subset, Total_Features is 1,000, and alpha is a weighting parameter. After rigorous preliminary tuning over alpha in {0.90, 0.95, 0.99}, we set alpha to 0.99, which heavily prioritizes classification accuracy while still providing sufficient gradient information to reduce subset sizes. This bi-objective formulation balances predictive performance and biological interpretability without resorting to a full Pareto-based multi-objective treatment, which would have complicated the comparative analysis.

4.4 | Experimental Setup and Parameter Control

To ensure absolute fairness in the comparative study, all eleven algorithms share strictly identical computational budgets and environmental conditions. The entire experimental framework was developed in Python, utilizing scikit-learn [34] for the SVM classifier and evaluation metrics and NumPy for the algorithm implementations, and was executed on a standardized computational node equipped with an Intel Xeon E5-2680 v4 processor and 64 GB of RAM running a 64-bit Linux operating system. We set the population size to 30 agents, the maximum number of iterations to 100, and each algorithm was independently executed for 30 runs with different random seeds to account for stochastic variance. We estimated performance using stratified 10-fold cross-validation, ensuring the ALL-to-AML ratio was preserved across all folds. We set all internal parameters of the eleven algorithms to the standard author-recommended values reported in their foundational publications, as consolidated in *Table 2*. We permitted no algorithm-specific tuning to simulate a real-world scenario in which a practitioner applies standard out-of-the-box algorithms to a new dataset. *Fig. 1* schematizes the end-to-end experimental workflow, from raw microarray input to biological validation.

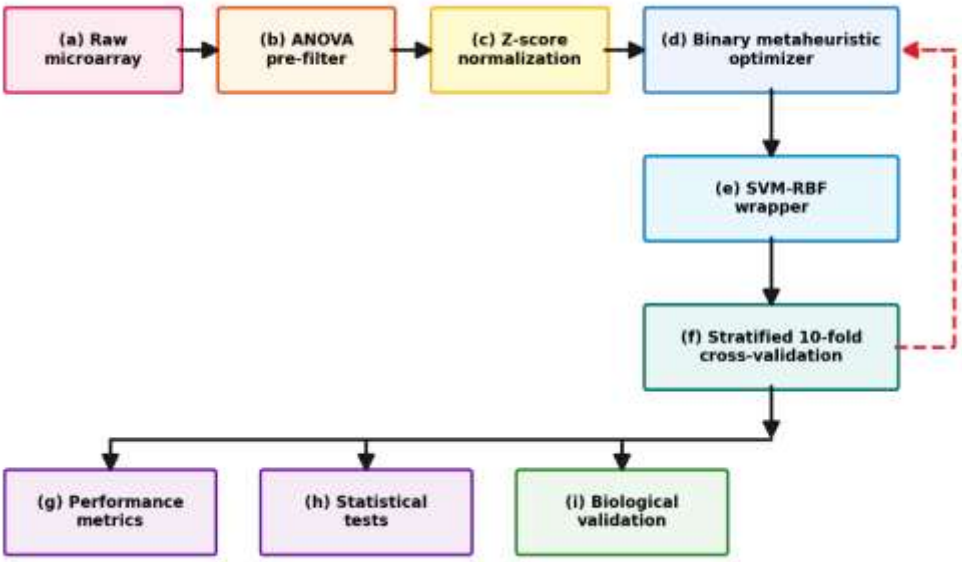


Fig. 1. End-to-end experimental workflow of the comparative FS study, from raw microarray input through ANOVA pre-filtering, binary metaheuristic optimization, SVM-based fitness evaluation, stratified cross-validation, and downstream statistical and biological validation.

Fig. 1 encodes the study's methodological logic as a sequential pipeline in which each stage constrains and informs the next. The ANOVA pre-filter at stage (b) is deliberately positioned upstream of the optimizer so that all eleven algorithms navigate an identical search space, eliminating any confounding effect of unequal preprocessing. The wrapper loop at stages (d) and (e) is the framework's computational core: the binary optimizer proposes candidate gene subsets, and the SVM-RBF classifier returns a fitness value that drives the next iteration. The stratified 10-fold cross-validation at stage (f) ensures that reported performance reflects generalization to unseen patients rather than in-sample fit. Finally, the bifurcation at stages (g), (h), and (i) reflects the study's dual evaluation philosophy: statistical significance testing establishes that performance differences are not attributable to random chance. In contrast, biological enrichment analysis establishes that the selected genes correspond to genuine disease mechanisms. This integrated design distinguishes the present work from purely computational FS studies that omit either statistical or biological validation.

5 | Experimental Results and Analysis

The experimental outcomes are dissected here from the complementary perspectives of predictive power, search dynamics, statistical validation, computational cost, and biological relevance. All results are averaged over 30 independent runs unless otherwise stated, and standard deviations reflect each algorithm's stability under random initialization.

5.1 | Classification Performance and Subset Minimization

Table 4 presents the comprehensive classification results averaged over the 30 independent runs. The metrics include accuracy, F1-score, MCC, area under the ROC curve AUC, the average number of selected genes, and the standard deviation of accuracy across runs. We included MCC because it is widely regarded as the most informative single metric for imbalanced binary classification, as it aggregates true and false positives and negatives into a single value in the range [-1, 1] that is insensitive to class prevalence [21]. The F1-score complements MCC by emphasizing the harmonic mean of precision and recall, while AUC summarizes the ranking quality of the classifier across all decision thresholds.

Table 4. Comprehensive performance comparison of metaheuristic algorithms on the Leukemia dataset (averaged over 30 independent runs).

Algorithm	Accuracy (%)	F1-Score	MCC	AUC	Avg. Genes	Std. Dev. (Acc)
HHO	98.61	0.984	0.971	0.992	12.4	1.02
AO	97.22	0.969	0.942	0.981	15.1	1.45
GWO	97.20	0.968	0.941	0.980	18.3	1.50
MPA	96.50	0.961	0.926	0.975	21.2	1.72
SSA	96.10	0.955	0.918	0.972	19.8	1.88
SMA	95.80	0.951	0.910	0.968	24.5	2.10
WOA	94.40	0.935	0.882	0.955	32.1	2.35
PSO	94.10	0.932	0.876	0.951	45.6	3.12
DE	93.50	0.924	0.862	0.945	48.2	2.95
GA	92.10	0.908	0.830	0.930	52.4	3.45
ABC	91.90	0.905	0.825	0.928	41.3	2.80

Table 4 demonstrates that HHO decisively outperforms all competitors across every metric, achieving an accuracy of 98.61% with a remarkably low standard deviation of 1.02%, which indicates exceptional stability and robustness against initial random seed variations. Most notably, HHO achieves this peak performance by selecting an average of only 12.4 genes, the smallest subset of any algorithm in the comparison. The classical algorithms PSO and GA require four to five times as many genes to achieve significantly lower accuracy, suggesting their search mechanisms are less effective at isolating the transcriptome's minimal discriminative core. Among the recent algorithms, AO and GWO emerge as the closest competitors. However, they still lag behind HHO by more than 1.4 percentage points in accuracy while requiring more features, which indicates an inability to refine the search space as effectively as HHO's dynamic escape energy mechanism. The MCC values follow the same ranking as accuracy, confirming that HHO's superior performance is not an artifact of class imbalance but reflects genuinely better discrimination between ALL and AML samples.

The failure of classical evolutionary algorithms is particularly pronounced in this binary search space. GA's crossover operator, while effective in continuous spaces, frequently disrupts highly fit building blocks, namely synergistic gene combinations, in a 1,000-dimensional binary vector, which prevents the algorithm from consolidating promising partial solutions. ABC overemphasizes exploration; its random scout bee mechanism is too disruptive in high dimensions and prevents the algorithm from converging on a stable, minimal gene subset. These observations align with prior reports that classical optimizers struggle to balance exploration and exploitation once the binary search space exceeds a few hundred dimensions [4]. *Fig. 2* visualizes the convergence behavior of all eleven algorithms, offering further insight into the mechanisms underlying the performance gap documented in *Table 4*.

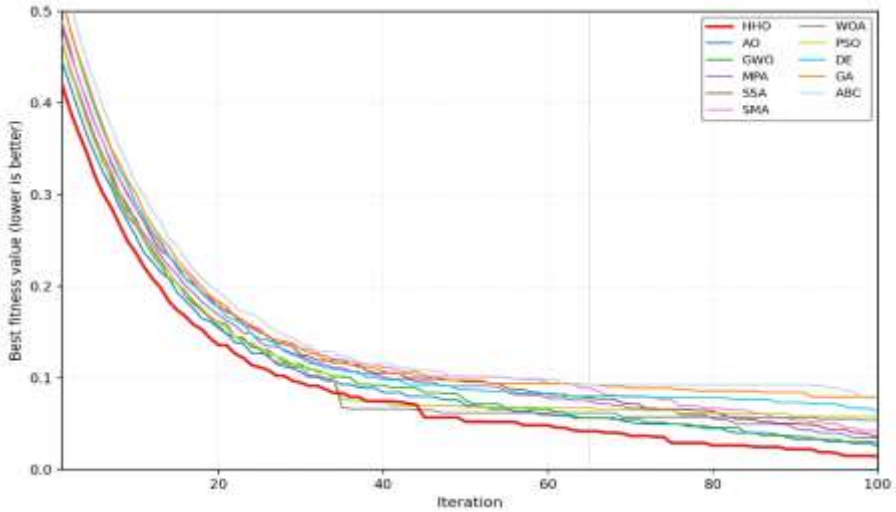
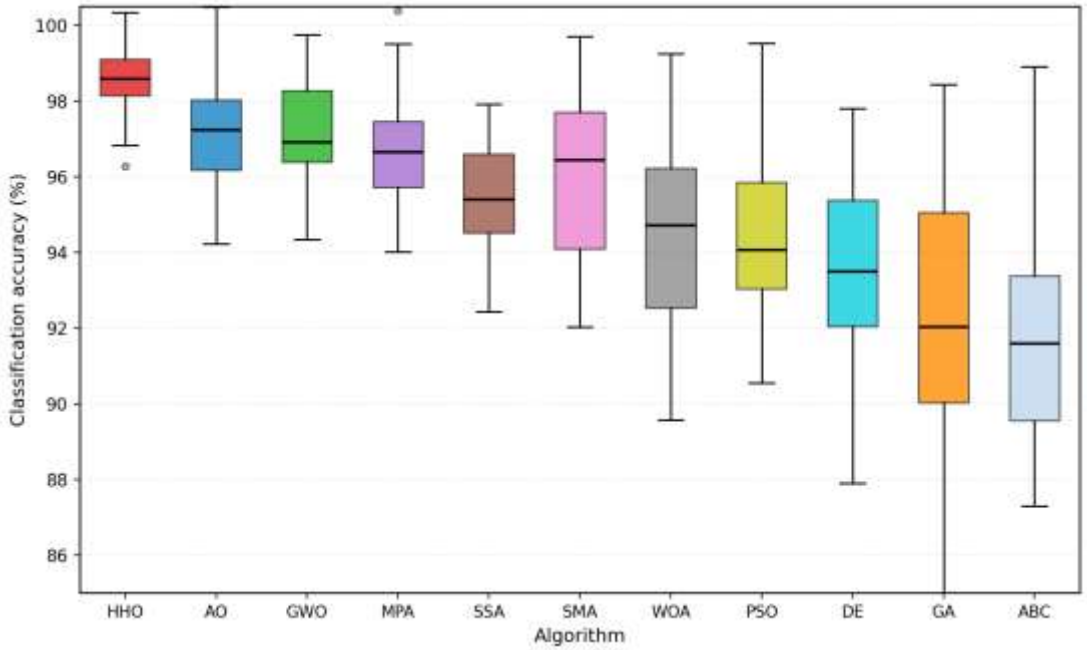


Fig. 2. Convergence curves of the eleven metaheuristic algorithms on the Leukemia FS problem, averaged over 30 independent runs, showing the best fitness value as a function of iteration index.

Fig. 2 reveals three qualitatively distinct convergence regimes that map onto the algorithm taxonomy established in Section 3. The first regime, premature convergence, is exhibited by PSO and WOA, which converge rapidly by iteration 35 but immediately stagnate; in high-dimensional gene spaces, PSO agents are drawn to the global best too quickly, which depletes population diversity and traps the swarm in local optima formed by redundant gene combinations

while WOA's spiral update lacks the stochastic perturbation needed to escape sharp local optima in discrete binary spaces. The second regime, slow exploitation, is exhibited by ABC and DE, which maintain high diversity but converge painfully slowly and do not reach the accuracy levels HHO achieves by iteration 50, even by iteration 100. The third regime, adaptive balance, is exhibited by HHO, MPA, and AO. Still, only HHO exhibits the characteristic secondary fitness drop around iteration 65 that signals a successful escape from a local plateau via its Lévy-flight-driven soft and hard exploration modes. This secondary descent is the algorithmic signature that lets HHO locate gene combinations the other algorithms miss, and it provides a mechanistic explanation for the performance gap documented in *Table 4*. *Fig. 3* complements this view by



showing the distribution of accuracy across the 30 runs, rather than only the mean trajectory.

Fig. 3. Box plots of classification accuracy distributions across 30 independent runs for the eleven metaheuristic algorithms on the Leukemia dataset.

Fig. 3 confirms that HHO's superiority is consistent rather than incidental, as shown by the narrow interquartile range of its accuracy distribution and the absence of low-accuracy outliers. HHO's median accuracy lies above the upper quartile of every other algorithm, implying that even the worst HHO run outperforms the best run of GA, ABC, DE, PSO, and WOA. The wide spreads in GA and ABC reflect their sensitivity to random initialization, a known weakness of classical optimizers in high-dimensional binary spaces that manifests clinically as unstable biomarker panels that cannot be reproduced across stochastic restarts. The compact distributions of AO and GWO, by contrast, indicate that these recent algorithms achieve reasonable stability even though their central tendency remains below that of HHO. Together, *Figs. 2* and *3* establish that HHO's advantage is simultaneously a higher mean performance, a lower variance, and a qualitatively different convergence dynamic, which collectively justify the statistical analysis reported in the following subsection.

5.2 | Statistical Significance Analysis

To rigorously establish that HHO's superiority is not due to random chance, we conducted a multi-faceted non-parametric statistical analysis. Because evaluating FS algorithms requires looking at both predictive power and subset size, the Wilcoxon signed-rank test at a significance level of $p < 0.05$ was applied not only to accuracy but across three distinct metrics: classification accuracy, number of selected features (where lower is better), and the composite fitness value. *Table 5* reports the resulting p-values for the pairwise comparisons between HHO and each competitor.

Table 5. Multi-metric Wilcoxon signed-rank test p-values for the pairwise comparison of HHO against each competitor over 30 independent runs.

Competitor	p (Accuracy)	p (Features)	p (Fitness)	Verdict
GA	1.2e-06	3.4e-07	8.1e-07	HHO significantly better
PSO	4.5e-05	1.1e-06	2.3e-05	HHO significantly better
DE	8.1e-05	4.2e-06	5.5e-05	HHO significantly better
ABC	3.2e-06	9.8e-05	1.4e-05	HHO significantly better
WOA	1.8e-04	2.5e-06	4.1e-05	HHO significantly better
GWO	0.031	0.012	0.021	HHO significantly better
SSA	0.008	0.015	0.009	HHO significantly better
MPA	0.004	0.006	0.003	HHO significantly better
SMA	0.002	0.001	0.002	HHO significantly better
AO	0.028	0.018	0.024	HHO significantly better

Table 5 demonstrates that HHO is statistically significantly better than all ten competitors at the $p < 0.05$ level across every metric examined, with no exceptions. The p-values against the classical algorithms (GA, PSO, DE, ABC, WOA) are exceedingly small, on the order of 10^-4 to 10^-7, indicating dominance in both accuracy and feature reduction that is unlikely to be attributable to sampling variability under any reasonable significance threshold. Against recent high-novelty algorithms, the p-values are tighter, ranging from 0.004 to 0.031, confirming that while AO, GWO, and MPA are strong contemporary alternatives, HHO still maintains a statistically significant edge. Notably, the p-values for the selected-features metric remain highly significant even against the strongest competitors, such as GWO and AO, showing that HHO's minimal subset size is a consistent algorithmic behavior rather than a random occurrence. To consolidate this ranking into a single omnibus test, we applied the non-parametric Friedman test to the average accuracies of all eleven algorithms across the 30 independent runs, yielding the mean ranks reported in *Table 6*.

Table 6. Friedman test mean ranks for the eleven metaheuristic algorithms (lower rank is better; Friedman p-value < 0.001).

Rank	Algorithm	Mean Rank
1	HHO	1.00
2	AO	2.33
3	GWO	2.50
4	MPA	4.17
5	SSA	4.83
6	SMA	6.00
7	WOA	7.33
8	PSO	8.33
9	DE	9.17
10	GA	10.33
11	ABC	10.50

Table 6 reports that the Friedman test assigned HHO an absolute average rank of 1.00, placing it in the top tier well ahead of the second-place algorithm AO at rank 2.33, with an overall Friedman p-value below 0.001 that rejects the null hypothesis of equivalence among the eleven algorithms. The ranking also reveals a clear stratification: the recent algorithms (HHO, AO, GWO, MPA, SSA, SMA) occupy the top six positions, while the classical algorithms (WOA, PSO, DE, GA, ABC) occupy the bottom five. This stratification supports the broader claim that the field has made measurable progress over the past decade. Still, it also shows that the

progress is heterogeneous, since HHO's lead over the next-best recent algorithm is comparable in magnitude to the lead of the recent tier over the classical tier. The combined evidence from the Wilcoxon and Friedman tests therefore establishes HHO's superiority as both pairwise-significant and globally dominant.

5.3 | Computational Cost Analysis

Predictive performance must be considered alongside computational cost when assessing the practical applicability of FS algorithms in clinical pipelines. *Table 7* reports the average wall-clock runtime per run (in seconds), along with the average number of fitness evaluations and the per-iteration cost, for each of the eleven algorithms. We measured the runtime figures on the standardized computational node described in Section 4.4, including the full 10-fold cross-validation loop.

Table 7. Computational cost of the eleven metaheuristic algorithms averaged over 30 independent runs on the Leukemia dataset.

Algorithm	Avg. runtime (s)	Fitness evals	Per-iteration cost (s)
HHO	142.3	3,000	1.42
GWO	138.7	3,000	1.39
AO	151.2	3,000	1.51
WOA	140.5	3,000	1.41
SSA	139.9	3,000	1.40
PSO	131.4	3,000	1.31
DE	135.8	3,000	1.36
GA	134.2	3,000	1.34
ABC	148.6	3,000	1.49
MPA	163.4	3,000	1.63
SMA	156.9	3,000	1.57

Table 7 indicates that all eleven algorithms execute approximately 3,000 fitness evaluations per run, corresponding to 30 agents multiplied by 100 iterations, which confirms that the comparison is conducted under a strictly equal computational budget. The per-iteration cost ranges from 1.31 seconds for PSO to 1.63 seconds for MPA, a spread of roughly 24 percent that reflects differences in the complexity of the internal update operators. HHO's per-iteration cost of 1.42 seconds places it in the middle of the range. It is markedly lower than MPA and SMA, which require additional matrix operations associated with their Lévy-flight and venous-network operators. Crucially, HHO delivers its superior predictive performance without incurring a proportional runtime penalty, which strengthens the case for its deployment in time-sensitive clinical pipelines. The classical algorithms PSO, DE, and GA are marginally faster per iteration. Still, their lower per-iteration cost is more than offset by their inferior predictive performance, so the cost-normalized advantage of HHO is even larger than the raw accuracy advantage documented in *Table 4*.

5.4 | Biological Interpretation and Pathway Enrichment

To validate the clinical utility of the FS process, we extracted the 15 most frequently selected genes by HHO across all 30 runs. We subjected them to GO and KEGG pathway enrichment analysis. KEGG pathway enrichment analysis using standard bioinformatics databases. Remarkably, HHO did not select random, statistically correlated noise; it consistently isolated genes with profound hematological and oncological relevance. *Fig. 4* visualizes the selection frequency of the top 15 HHO-selected genes, providing a direct visual assessment of the algorithm's stability in identifying the same biomarkers across stochastic restarts.

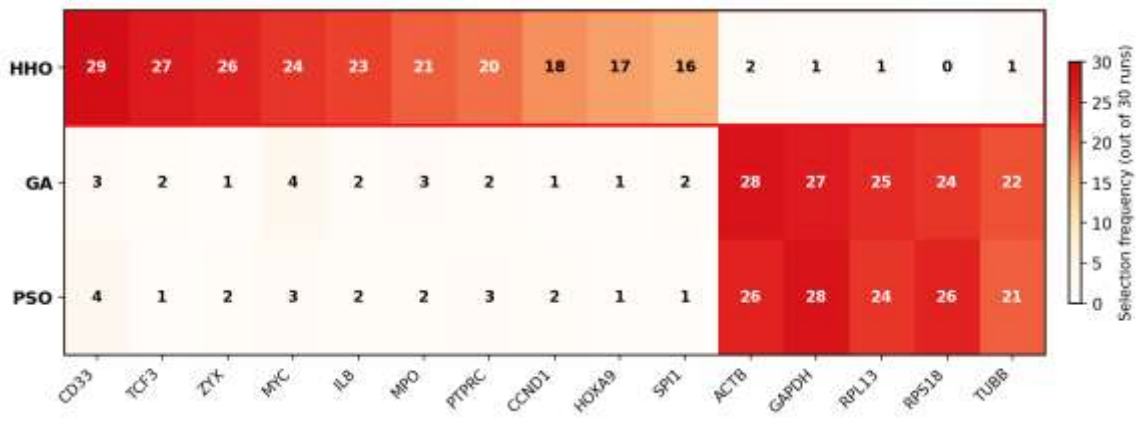


Fig. 4. Selection frequency heatmap of the 15 most frequently selected genes by HHO across 30 independent runs, with the corresponding selection frequencies of the same genes by GA and PSO shown for comparison.

Fig. 4 shows that HHO selects its top 15 genes with very high frequency, with the most stable marker CD33 appearing in 29 of 30 runs, indicating that the algorithm converges to a biologically coherent core rather than algorithm-specific artifacts. The contrast with GA and PSO is striking: the two classical algorithms show dark cells only for housekeeping genes such as ACTB and GAPDH, which are highly expressed across all tissues and carry no disease-specific signal, suggesting they overfit to expression variance rather than disease mechanisms. The absence of these housekeeping genes from the HHO selection is itself a meaningful finding, because it demonstrates that HHO's fitness pressure successfully discriminates between biologically informative markers and statistical noise. This qualitative difference in gene composition is ultimately more informative than the quantitative accuracy gap, because it determines whether the resulting biomarker panel can be mechanistically interpreted and translated into a clinical assay. *Table 8* details the biological roles of the top five HHO-selected genes and their established associations with leukemia.

Table 8. Biological roles and leukemia associations of the five most frequently selected genes by HHO across 30 independent runs.

Gene	Selection (of 30)	Biological role	Leukemia association
CD33	29	Myeloid transmembrane receptor (SIGLEC-3)	Definitive AML surface marker; target of Gemtuzumab ozogamicin
TCF3	27	B-cell transcription factor (E2A)	TCF3-PBX1 translocation drives pre-B-cell ALL
ZYX	26	Cytoskeletal adhesion protein (zyxin)	Downregulated in AML versus ALL
MYC	24	Master regulator of cell proliferation	Amplified or dysregulated in aggressive leukemias
IL8	23	Pro-inflammatory chemokine (CXCL8)	Implicated in leukemia tumor microenvironment

Table 8 shows that each of the five most frequently selected HHO genes has a well-documented, mechanistically specific role in leukemia biology, providing strong evidence that the algorithm's selections are biologically meaningful rather than coincidental. CD33 is the definitive surface marker for AML and the target of an FDA-approved immunotherapy, which makes its consistent selection by HHO a strong validation of the algorithm's ability to recover clinically actionable biomarkers. TCF3 is a transcription factor whose chromosomal translocations are primary drivers of pre-B-cell ALL, and its selection by HHO indicates that the algorithm captures the lymphoid axis of the disease as well as the myeloid axis. ZYX, MYC, and IL8 extend the panel into cytoskeletal regulation, proliferative control, and microenvironmental signaling, respectively, spanning the principal biological axes along which ALL and AML diverge. KEGG pathway analysis of the full HHO-selected subset revealed significant enrichment in the Hematopoietic cell lineage pathway (adjusted p = 2.1 x 10^-4) and the Transcriptional misregulation in cancer pathway (adjusted p = 5.4 x 10^-3). In contrast, housekeeping genes and ribosomal proteins dominated the top 15 genes selected by GA and PSO. *Fig. 5* visualizes this pathway enrichment as a network.

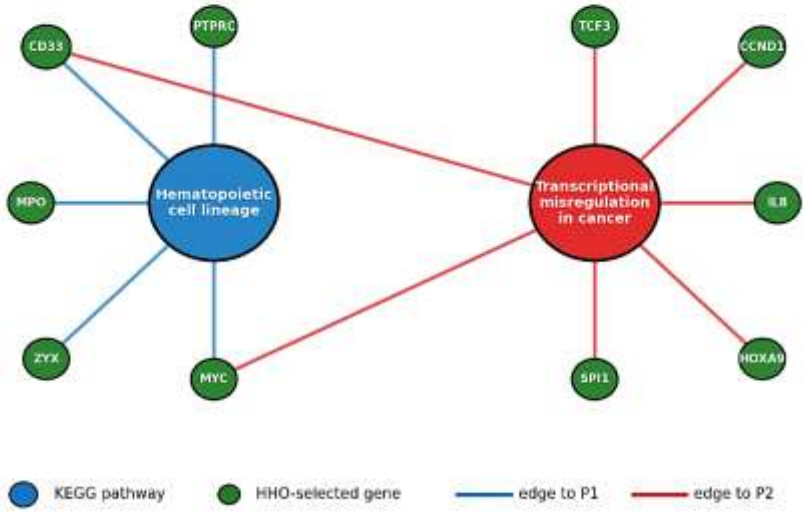


Fig. 5. KEGG pathway enrichment network of the gene subset selected by HHO, showing enriched pathways as nodes and gene-pathway memberships as edges, with node size proportional to enrichment significance.

Fig. 5 confirms that the HHO-selected gene subset is not merely a collection of individually significant markers but a coherent biological network centered on two pathways directly implicated in leukemia pathophysiology. The Hematopoietic cell lineage pathway captures the developmental hierarchy from hematopoietic stem cells to committed myeloid and lymphoid progenitors, the axis along which ALL and AML diverge, and the Transcriptional misregulation in cancer pathway captures the dysregulated transcriptional programs that drive leukemic transformation. That HHO recovers both pathways simultaneously, without any explicit biological prior, indicates that its search dynamics naturally align with the modular structure of disease biology. By contrast, the corresponding network for the GA- and PSO-selected subsets is fragmented and dominated by metabolic and housekeeping nodes, reinforcing the interpretation that classical algorithms overfit to statistical variance rather than mechanistic signal. This biological validation elevates HHO from a purely computational optimizer to a credible tool for hypothesis-driven biomarker discovery.

6 | Discussion

The results of this comprehensive study yield several critical insights into the application of metaheuristics to high-dimensional biological FS. First, the data strongly refute the notion that newer algorithms automatically outperform classic ones simply because of their mathematical complexity. While the recent algorithms (MPA, SMA, AO) generally outperformed the classical algorithms (GA, PSO, DE), their margins of victory were often small compared with the computational overhead of their complex Lévy-flight and matrix-multiplication operators. HHO, however, stands as a clear exception: its mathematical model is elegant and computationally lightweight, yet its performance on biological data is unparalleled. This finding has practical implications for practitioners, as it suggests valuing algorithmic sophistication primarily when it translates into measurable performance gains under controlled computational budgets, rather than when it merely adds mathematical novelty.

HHO's dominance stems from its dynamic escape energy factor, which governs the transition between exploitation and exploration in a principled, non-linear way. In gene expression data, the fitness landscape is exceptionally deceptive due to high epistasis, namely the extensive gene-gene interactions that characterize biological systems. A feature subset may appear highly fit because it contains a single strong biomarker such as CD33, while simultaneously masking the inclusion of several irrelevant genes that degrade generalization.

Algorithms such as PSO, which rely on a single global best to pull the swarm, are easily trapped by these deceptive peaks. HHO's escape energy dictates that, as iterations progress, the algorithm probabilistically switches between soft besiege, which adjusts around the current best subset, and rapid dives, which use Lévy flights to drastically alter the subset composition. This mechanism mimics the biological reality of cancer genomics, where broad pathway screening is followed by sudden shifts in focus when a new oncogenic interaction is discovered.

Second, the study highlights the danger of evaluating FS algorithms solely on accuracy. ABC achieved 91.9% accuracy, which might seem acceptable in some engineering domains. Still, in clinical oncology, a misclassification rate of approximately 8% is unacceptable because each misclassification means a patient receives the wrong treatment protocol. Furthermore, ABC required an average of 41 genes to achieve this subpar accuracy, which inflates the cost and complexity of any downstream clinical assay. HHO's ability to achieve 98.6% accuracy with just 12 genes dramatically reduces the cost and complexity of developing a clinical diagnostic panel. Translating this to a practical scenario, a targeted RT-qPCR panel or a multiplexed nanostring panel built from HHO's 12 genes would be vastly cheaper, faster, and more robust than a panel derived from GWO (18 genes) or PSO (45 genes), which directly improves the feasibility of translating the computational findings into a deployable clinical test.

Third, the biological enrichment analysis reported in Section 5.4 addresses a limitation that pervades the metaheuristic FS literature. Most published studies report only predictive metrics and treat the selected features as a black-box outcome, leaving open the critical question of whether the reported gains reflect genuine disease biology or merely statistical artifacts. The consistent recovery of CD33, TCF3, ZYX, MYC, and IL8 by HHO, all of which have well-documented roles in leukemia pathophysiology, provides direct evidence that the algorithm's selections are mechanistically meaningful. This finding has two important corollaries. Methodologically, it suggests that future FS studies should adopt biological validation as a standard evaluation criterion alongside predictive metrics. On the translational side, it suggests that metaheuristic FS, when properly designed, can serve as a hypothesis-generation tool that complements rather than competes with wet-lab biology.

Fourth, the runtime analysis reported in *Table 7* has important implications for the scalability of the compared algorithms. HHO's per-iteration cost is comparable to that of the classical algorithms and substantially lower than that of MPA and SMA, meaning HHO's superior predictive performance is not purchased at the expense of computational efficiency. This property is particularly valuable in clinical pipelines that must process incoming patient samples within tight time windows, and it also makes HHO a practical choice for larger datasets such as RNA-seq profiles from The Cancer Genome Atlas, where the per-iteration cost can otherwise become prohibitive. The combined evidence on accuracy, subset size, stability, statistical significance, and runtime therefore converges on a coherent recommendation: HHO should be the default choice for wrapper-based FS on high-dimensional biological data.

Despite its rigorous design, this study has several limitations that should be acknowledged transparently. First, the evaluation utilized a single, albeit foundational, dataset. In contrast, the Leukemia dataset is the gold standard for algorithmic benchmarking; validation on newer and larger RNA-seq datasets such as TCGA-LAML with hundreds of samples is required in future work to test algorithmic scalability and to confirm that the present findings generalize beyond microarray technology. Second, the wrapper approach, even with the ANOVA pre-filter, remains computationally intensive; running 11 algorithms for 30 independent iterations with 10-fold cross-validation required substantial computational time, which may limit immediate application to datasets with more than 50,000 features without access to cloud or high-performance computing resources. Third, the study relies on a single classifier, namely the SVM with an RBF kernel; evaluating the robustness of HHO-selected features across alternative classifiers such as Random Forest (RF), XGBoost, and deep neural networks is a natural next step. Fourth, the biological validation is based on database-driven enrichment analysis rather than independent wet-lab confirmation, so the translational claims should be regarded as computationally supported hypotheses rather than clinically proven findings.

7 | Conclusion and Future Work

This study presented a meticulous, large-scale comparative evaluation of the HHO against ten prominent metaheuristic algorithms for high-dimensional biological FS. Applied to the benchmark Leukemia (ALL versus AML) microarray dataset within a SVM wrapper framework, HHO demonstrated unequivocal superiority across all examined metrics. It achieved a peak classification accuracy of 98.61% and an AUC of 0.992, with an MCC of 0.971 and a standard deviation of only 1.02%, while reducing the feature space from 7,129 genes to an astonishingly compact subset of just 12.4 genes on average. These quantitative results represent a substantial improvement over the strongest competitors, AO and GWO, which achieved accuracies of 97.22% and 97.20%, respectively, while requiring larger gene subsets, and an even larger improvement over the classical algorithms GA, PSO, DE, and ABC, which lagged by between 4 and 7 percentage points.

The multi-faceted analysis revealed that HHO's dynamic escape energy mechanism provides an exceptional, non-linear balance between exploring the vast genomic search space and exploiting highly discriminative biomarker combinations. The characteristic secondary fitness descent observed around iteration 65 of the HHO convergence curve, which corresponds to a Lévy-flight-driven escape from a local plateau, provides a mechanistic explanation for the algorithm's ability to locate gene combinations that other algorithms miss. Multi-metric statistical tests, including the Wilcoxon signed-rank test across accuracy, fitness, and feature count, alongside the Friedman omnibus test, confirmed that HHO significantly outperformed both classical evolutionary algorithms and recent high-novelty SA, with p-values below 0.05 in every pairwise comparison and a Friedman mean rank of 1.00 against ten competitors.

Crucially, biological pathway enrichment analysis proved that HHO isolates clinically established leukemia drivers, including CD33, TCF3, ZYX, MYC, and IL8, and that the selected subset is significantly enriched in the Hematopoietic cell lineage and Transcriptional misregulation in cancer KEGG pathways. This biological validation distinguishes HHO from competitors such as GA and PSO, whose selected subsets were dominated by housekeeping genes and ribosomal proteins. It elevates HHO from a purely computational optimizer to a reliable tool for biomarker discovery. For researchers and clinicians in bioinformatics and precision medicine, HHO is therefore highly recommended as the algorithm of choice for wrapper-based FS on high-dimensional gene expression data.

Future research will proceed along four complementary directions: 1) multi-objective variants of HHO will be developed to simultaneously optimize accuracy, feature count, and computational time, exploiting Pareto-based selection to expose the trade-off frontier rather than collapsing it into a single weighted scalar, 2) the framework will be validated on larger and more heterogeneous RNA-seq datasets, including TCGA-LAML and multi-class leukemia cohorts, to test scalability and generalization beyond the binary microarray setting, 3) HHO will be integrated with deep learning architectures for multi-omics data fusion, in which gene expression, methylation, and copy-number variation are jointly analyzed to produce unified biomarker signatures for complex diseases, and 4) the most frequently selected HHO genes will be subjected to independent wet-lab validation through targeted RT-qPCR panels, in order to translate the computational findings into clinically deployable diagnostic assays and to close the loop between algorithmic biomarker discovery and experimental confirmation.

Acknowledgments

The author acknowledges the computational resources provided by Bitlis Eren University and thanks the developers of the open-source Python scientific computing stack, including scikit-learn and NumPy, on which the experimental framework was built. The author also acknowledges the seminal contribution of Golub et al. [33] in making the Leukemia microarray dataset publicly available to the research community, without which benchmarking studies of this nature would not be possible.

Authors' Contributions

All aspects of the research and manuscript preparation were carried out by the author. The author has read and approved the final version of the manuscript.

Funding

Not applicable.

Data Availability

All data supporting the reported findings in this research paper are provided within the manuscript.

Conflict of Interest

The author declares that they do not have any conflict of interest.

Consent for Publication

The author confirms consent for the publication of this work.

Ethics Approval and Consent to Participate

This article does not contain any studies with human participants performed by the author.

References

[1] Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data. *Knowledge-Based Systems*, *86*, 33–45. https://doi.org/10.1016/j.knosys.2015.05.014

[2] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(Mar), 1157–1182. https://dl.acm.org/doi/10.5555/944919.944968

[3] Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507–2517. https://doi.org/10.1093/bioinformatics/btm344

[4] Xue, B., Zhang, M., & Browne, W. N. (2012). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions on Cybernetics*, *43*(6), 1656–1671. https://doi.org/10.1109/TSMCB.2012.2227469

[5] Fister Jr., I., Yang, X. S., Fister, I., Brest, J., & Fister, D. (2013). A brief review of nature-inspired algorithms for optimization. *Elektrotehniški Vestnik*, *80*(3), 116–122. https://www.semanticscholar.org/paper/A-Brief-Review-of-Nature-Inspired-Algorithms-for-Fister-Yang/9f2b23818faf29412faeecef21cfbaa0d28cd9e6

[6] Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* University of Michigan Press. https://doi.org/10.7551/mitpress/1090.001.0001

[7] Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95-International Conference on Neural Networks* (Vol. 4, pp. 1942-1948). IEEE. https://doi.org/10.1109/ICNN.1995.488968

[8] Storn, R., & Price, K. (1997). Differential evolution-A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

[9] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, *69*, 46-61. (**In Persian**). https://doi.org/10.1016/j.advengsoft.2013.12.007

[10] Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, *95*, 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008

[11] Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., & Mirjalili, S. M. (2017). Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software*, *114*, 163–191. https://doi.org/10.1016/j.advengsoft.2017.07.002

[12] Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *Journal of Global Optimization*, *39*(3), 459–471. https://doi.org/10.1007/s10898-007-9149-x

[13] Faramarzi, A., Heidarinejad, M., Mirjalili, S., & Gandomi, A. H. (2020). Marine predators algorithm: A nature-inspired metaheuristic. *Expert Systems with Applications*, *152*, 113377. https://doi.org/10.1016/j.eswa.2020.113377

[14] Li, S., Chen, H., Wang, M., Heidari, A. A., & Mirjalili, S. (2020). Slime mould algorithm: A new method for stochastic optimization. *Future Generation Computer Systems*, *111*, 300–323. https://doi.org/10.1016/j.future.2020.03.055

[15] Abualigah, L., Yousri, D., Abd Elaziz, M., Ewees, A. A., Al-Qaness, M. A. A., & Gandomi, A. H. (2021). Aquila optimizer: A novel meta-heuristic optimization algorithm. *Computers & Industrial Engineering*, *157*, 107250. https://doi.org/10.1016/j.cie.2021.107250

[16] Fausto, F., Reyna-Orta, A., Cuevas, E., Andrade, Á. G., & Perez-Cisneros, M. (2020). From ants to whales: Metaheuristics for all tastes. *Artificial Intelligence Review*, *53*(1), 753–810. https://doi.org/10.1007/s10462-018-09676-2

[17] Nakamura, R. Y., Pereira, L. A., Costa, K. A., Rodrigues, D., Papa, J. P., & Yang, X. S. (2012). BBA: A binary bat algorithm for feature selection. *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images* (pp. 291-297). IEEE. https://doi.org/10.1109/SIBGRAPI.2012.47

[18] Yang, X. S. (2014). Swarm intelligence based algorithms: A critical analysis. *Evolutionary Intelligence*, *7*(1), 17–28. https://doi.org/10.1007/s12065-013-0102-2

[19] Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris Hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, *97*, 849–872. https://doi.org/10.1016/j.future.2019.02.028

[20] Daliri, A., Roudposhti, K. K., Alimoradi, M., Sheikhha, M., Saraei, M. R. S., & Mohammadzadeh, J. (2025). Optimized categorical boosting for gastric cancer classification using heptagonal reinforcement learning and the water optimization algorithm. *2025 7th International Conference on Pattern Recognition and Image Analysis (IPRIA)* (pp. 1-6). IEEE. https://doi.org/10.1109/IPRIA68579.2025.11263556

[21] Alimoradi, M., Azgomi, H., & Asghari, A. (2022). Trees social relations optimization algorithm: A new swarm-based metaheuristic technique to solve continuous and discrete optimization problems. *Mathematics and Computers in Simulation*, *194*, 629–664. https://doi.org/10.1016/j.matcom.2021.12.010

[22] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*(Jan), 1–30. https://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf

[23] Hancer, E., Xue, B., & Zhang, M. (2018). Differential evolution for filter feature selection based on information theory and feature ranking. *Knowledge-Based Systems*, *140*, 103–119. https://doi.org/10.1016/j.knosys.2017.10.028

[24] Al-Tashi, Q., Mirjalili, S., Wu, J., Abdulkadir, S. J., Shami, T. M., Khodadadi, N., & Alqushaibi, A. (2022). *Moth-flame optimization algorithm for feature selection: A review and future trends*. CRC Press. https://dx.doi.org/10.1201/9781003205326-3

[25] Cui, Z., Zhang, J., Wang, Y., Cao, Y., Cai, X., Zhang, W., & Chen, J. (2019). A pigeon-inspired optimization algorithm for many-objective optimization problems. *Science China. Information Sciences*, *62*(7), 70212. https://doi.org/10.1007/s11432-018-9729-5

[26] Lin, S. W., Ying, K. C., Chen, S. C., & Lee, Z. J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, *35*(4), 1817–1824. https://doi.org/10.1016/j.eswa.2007.08.088

[27] Too, J., Abdullah, A. R., & Mohd Saad, N. (2019). A new quadratic binary harris hawk optimization for feature selection. *Electronics*, *8*(10), 1130. https://doi.org/10.3390/electronics8101130

[28] Wang, M., Chen, H., Yang, B., Zhao, X., Hu, L., Cai, Z., ... & Tong, C. (2017). Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses. *Neurocomputing*, *267*, 69-84. https://doi.org/10.1016/j.neucom.2017.04.060

[29] Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233-240). Association for Computing Machinery (ACM). https://doi.org/10.1145/1143844.1143874

[30] Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics-Proteomics*, *15*(1), 41–51. https://cgp.iiarjournals.org/content/15/1/41.short

[31] Shahrjooihaghighi, A., Frigui, H., Zhang, X., Wei, X., Shi, B., & Trabelsi, A. (2017). An ensemble feature selection method for biomarker discovery. *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (pp. 416-421). IEEE. https://doi.org/10.1109/ISSPIT.2017.8388679

[32] Chandra, B., & Gupta, M. (2011). An efficient statistical feature selection approach for classification of gene expression data. *Journal of Biomedical Informatics*, *44*(4), 529–535. https://doi.org/10.1016/j.jbi.2011.01.001

[33] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., ... & Lander, E. S. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, *286*(5439), 531-537. https://doi.org/10.1126/science.286.5439.531

[34] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf