Soft Computing Fusion with Applications

www.scfa.reapress.com

Soft. Comput. Fusion. Appl. Vol. 3, No. 4 (2026) 230–246.

Paper Type: Original Article

Computational Intelligence for Crop Improvement: A Critical Taxonomic Survey of Metaheuristic Optimization

Ana Lucia Montenegro-Ramírez[1,*], Devang Suri[2], Helena Voss[3], Tobias Brandt[4]

[1] International Center for Tropical Agriculture (CIAT/Alliance Bioversity-CIAT), Cali, Colombia; a.montenegro@cgiar.org.

[2] Department of Computer Science & Engineering, Indian Institute of Technology Ropar, Rupnagar, Punjab 140001, India.

[3] Wageningen University & Research, Plant Sciences Group, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

[4] IBM Research Europe, Sustainability & Agricultural AI, Säumerstrasse 4, 8803 Rüschlikon, Switzerland.

Citation:

Received: 25 February 2026 *Revised: 11 June 2026* *Accepted: 18 August 2026*	Montenegro-Ramírez, A. L., Suri, L., Voss, L., & Brandt, T. (2026). Computational intelligence for crop improvement: A critical taxonomic survey of metaheuristic optimization. *Soft Computing Fusion with Applications, 3*(4), 230-246.

Abstract

Crop improvement combines biological prediction, experimental design and management decisions under constrained resources and environmental variation. This narrative taxonomic survey organizes metaheuristic optimization by problem class, search mechanism, application role and evaluation design. It considers yield prediction, genomic selection, irrigation, pest detection, phenotyping and sensor deployment, while distinguishing optimization of a computational criterion from demonstrated agronomic benefit. A documented genetic-algorithm study of training-population selection illustrates a concrete link between genomic information, a prediction-error criterion and phenotyping decisions. A published crop-rotation framework illustrates scenario-based multi-objective search without establishing a universal algorithm advantage. The survey also audits the interpretation of a retained numerical compilation: missing instance-level records, invalid complete mean ranks, inconsistent sensitivity summaries and unreconciled literature percentages prevent it from supporting inferential conclusions. No new benchmark experiment or validated systematic corpus is asserted. Research priorities are transparent objectives and constraints, independent validation across environments and genotypes, matched computational budgets, reproducible data and uncertainty-aware decision support.

Keywords: Metaheuristic optimization, Crop improvement, Genomic selection, Precision agriculture, Multi-objective optimization, Statistical comparison.

1 | Introduction

Crop improvement involves linked decisions about genotypes, field management, measurements and predictive models. Each decision has its own feasible set, objective and validation requirements.

Corresponding Author: a.montenegro@cgiar.org

https://doi.org/10.22105/scfa.v3i4.113

Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

Computational intelligence can support those decisions when the search mechanism is matched to the formulation [1]. This survey examines that connection rather than inferring agricultural benefit from generic benchmark performance or unsupported market and population projections.

A problem-centered taxonomy helps readers distinguish an optimizer from the biological or agronomic evaluator it uses. Algorithmic terminology should be linked to operators, representations and information flow [2]. Comparative evidence must then specify the data, budgets and independent units being compared [3], [4]. The survey does not estimate the percentage of inadequately reported studies without an auditable corpus.

Agricultural decisions can involve seasonal constraints, related genotypes, spatially correlated observations and expensive field evaluations. These characteristics do not make all exact or gradient-based methods unsuitable. Linear, convex, integer and specialized formulations should be considered where applicable. Metaheuristics are a candidate search strategy when the resulting formulation justifies them, not a substitute for a valid crop response model.

The scope comprises yield and trait prediction, genomic and breeding decisions, irrigation and resource allocation, pest detection, phenotyping and sensor deployment. The literature is used narratively to connect formulation and evaluation principles. The reported 284-study distribution lacks a reproducible search export, deduplicated inclusion list and coding matrix; it therefore does not define a verified systematic-review population. A DOI alone is neither a quality assessment nor an adequate criterion for excluding a conference paper.

Section 2 explains search mechanisms and the decision context; Section 3 presents the taxonomic axes. Sections 4–8 consider application roles. Section 9 examines the evidence status of the numerical compilation and conditions for valid comparisons, Section 10 discusses validity, and Section 11 identifies research directions. The conclusion concerns methodological implications rather than a demonstrated field-wide optimizer ranking.

2 | Background

2.1 | Metaheuristic Optimization: A Brief Refresher

A metaheuristic guides candidate generation, evaluation and selection in a stated feasible search space [1]. Single-objective problems compare one criterion; multi-objective problems compare a vector of criteria under a specified preference or non-dominance rule. Search variability, biological uncertainty and uncertainty in a fitted prediction model are different quantities and should be evaluated separately.

Evolutionary, swarm and trajectory-based methods offer different mechanisms. Differential evolution builds trial vectors from population differences [5]; particle swarm optimization uses individual and social information [6]; ant-system methods construct combinatorial candidates [7]; NSGA-II provides a non-dominated evolutionary selection architecture [8]. Inspiration categories overlap and do not establish disjoint mathematical families or biological suitability.

2.2 | Optimization in Agriculture: A Historical Lens

Mathematical programming and heuristic search address different formulations of farm planning. New sensing and predictive models can change the evaluator and the available data, but do not demonstrate a universal historical transition from exact methods to metaheuristics. Method selection should follow the resulting variables, objectives, constraints and evaluation cost.

A trained prediction model can become an evaluator inside optimization, while a process-based simulation can evaluate candidate management actions. In either case, search quality is conditional on evaluator validity. Statistical comparison across independent tasks can assess computational performance [3], but it does not by itself demonstrate agronomic improvement after deployment.

2.3|The Agri-Biotechnology Lifecycle as a Decision Cascade

The agri-biotechnology lifecycle can be viewed as linked decisions: which individuals to phenotype, which parents to cross, which management schedule to use and where to place sensors. The links should be explicit because an upstream selection changes the data available to downstream models. A computationally convenient criterion is useful only when its relationship to the intended biological or operational objective is justified.

3|Taxonomy of Metaheuristics in Agri-Biotech

The taxonomy records four complementary descriptors: problem class, algorithm mechanism, application role and evaluation protocol. These descriptors are not statistically independent axes, and the schematic hierarchy is an organizing convention rather than a proof of separability.

Problem class identifies the decision, such as yield prediction, genomic selection, resource allocation or phenotyping.

Algorithm mechanism identifies the operators and state, such as population variation, velocity updates, differential mutation or combinatorial construction.

Application role identifies what search changes: features, model parameters, individuals selected for phenotyping, cross assignments, irrigation schedules or sensor locations.

Evaluation design identifies the independent unit, data separation, computational budget, outcome measure and uncertainty or comparison procedure.

Recording these descriptors together prevents a model-tuning algorithm from being confused with a crop-management policy and prevents repeated seeds from being confused with independent environments. The taxonomy is a proposed reporting structure. Establishing a distribution of published studies within it requires a traceable coding exercise.

Table 1. Unverified literature-distribution compilation.

Problem Class (L1)	Dominant Algorithm Family (L2)	Application Stage (L3)	Evaluation Protocol (L4)	Studies
Yield & trait prediction	GA / PSO / DE	Feature selection, hyperparameter tuning	5-fold CV + Wilcoxon	77
Genomic selection	GA / NSGA-II / DE	Marker subset, cross design	Bootstrap + Friedman	54
Irrigation & resource allocation	PSO / WOA / GA / ABC	Scheduling, dosing	RMSE + paired-t	43
Pest & disease detection	PSO / GWO / WOA / FA	Hyperparameter tuning, architecture search	Accuracy + 5-fold CV	48
Phenotyping & sensor networks	GA / ACO / PSO	Sensor layout, image segmentation	F1 + Wilcoxon	31
Cross-cutting (review/survey)	Multiple	Multiple	Qualitative	31
Total				284

Note: The reported study counts are retained but are not a validated corpus or prevalence estimate.

Fig. 1 illustrates the intended descriptors, while the prose specifies their meaning. A particular algorithm can serve several application roles and more than one evaluation design. Lines in a taxonomy diagram should not be read as exclusive algorithm–application relationships.

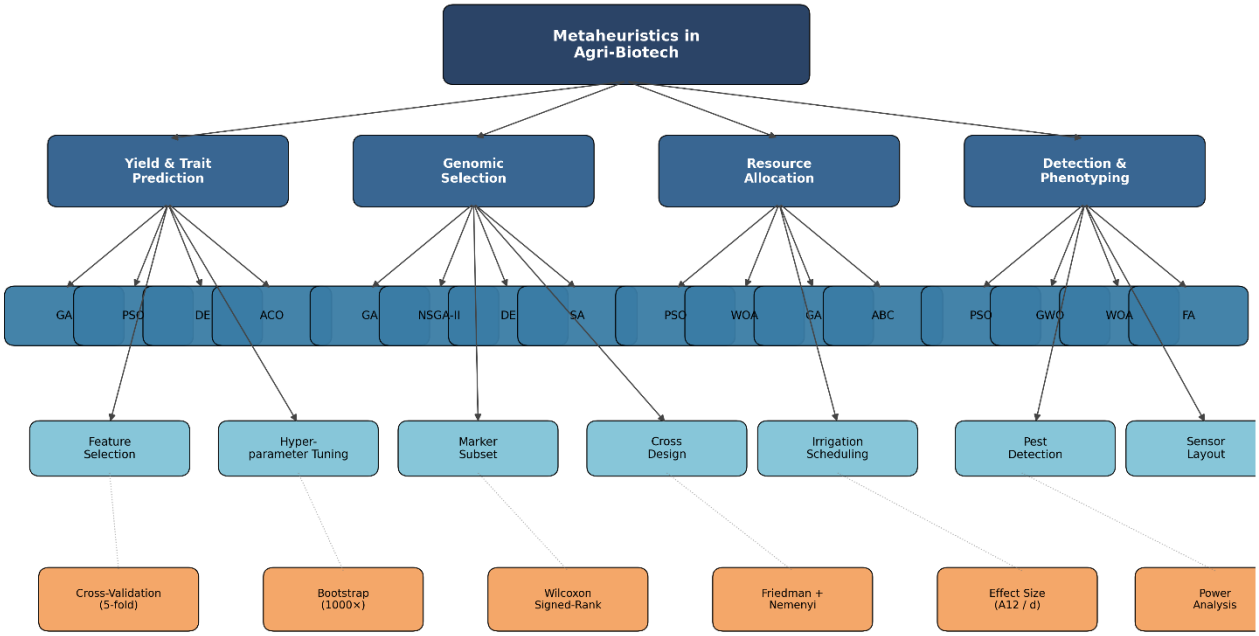


Fig. 1. Taxonomic schematic.

The algorithm–application links are illustrative; the diagram does not certify a complete survey or prescribe one evaluation test for each problem class (*Fig. 1*).

Figs. 2 and *3* retain publication and application-count graphics without a traceable corpus. They cannot establish growth, causal inflection points or domain prevalence. A statistical uncertainty band around a cumulative count requires a justified sampling or stochastic model; it is not created automatically by the fact that papers are counted.

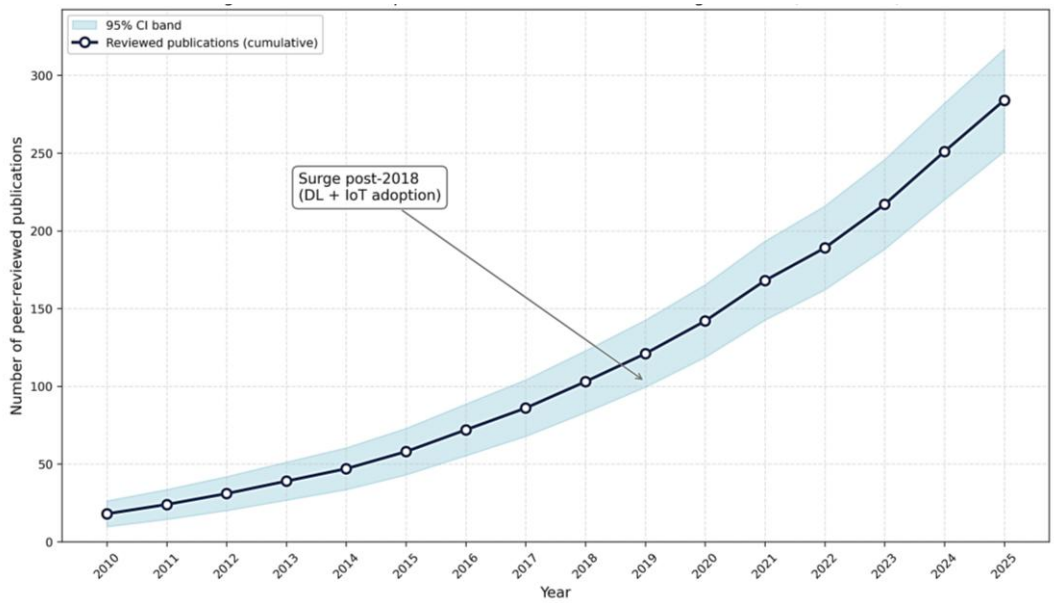


Fig. 2. Unverified cumulative-publication graphic.

Its count trajectory and confidence band are excluded from the validated synthesis until records and an uncertainty model are established (*Fig. 2*).

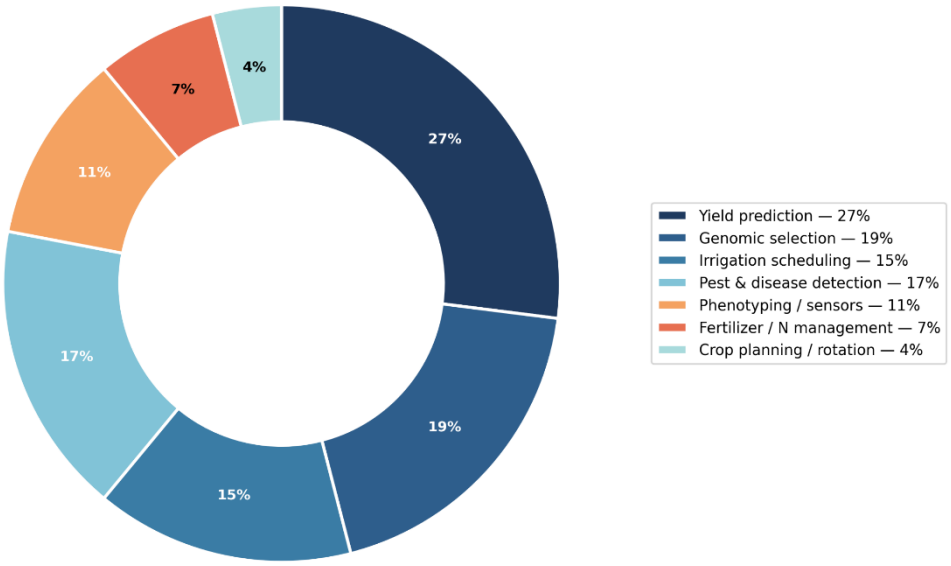


Fig. 3. Unverified application-distribution graphic.

Category definitions and record-level coding are unresolved; its percentages do not establish field prevalence (*Fig. 3*).

4 | Crop Yield Prediction

Yield prediction estimates an outcome from weather, soil, genotype and management information. Feature selection and model tuning are optimization layers within that prediction task. An evaluation intended to predict future seasons or unseen regions should hold out those units and avoid using their outcomes in preprocessing, selection or tuning.

4.1 | Feature Selection via Metaheuristics

Feature selection changes the inputs available to a predictor, while learned representations change how the inputs are processed. Khaki and Wang provide a documented deep-neural-network yield-prediction example [9]. It is a predictive-model source, not evidence for the unsourced GA reduction and metaheuristic prevalence percentages previously attributed to it. The effect of a selector should be evaluated against appropriate full-feature and simpler-selection baselines under the same outer validation split.

4.2 | Hyperparameter Optimization of Predictive Models

Hyperparameter optimization searches over model configurations using a training-only evaluation design. Reusing the final test set to compare candidate settings biases the final performance estimate. Automatic algorithm configuration tools such as irace illustrate a documented tuning procedure [10]; bandit-based operator selection and Bayesian optimization are distinct mechanisms and should not be conflated.

4.3 | Evaluation Across Locations and Seasons

A maize-yield study intended for transfer across regions must identify data sources, field or region units, prediction horizons and information available at prediction time. A credible comparison should include meaningful baseline models and uncertainty across the independent environments. No specific NSGA-II field experiment, regional ranking or percentage reduction is asserted without the primary measurements.

4.4 | Integration with Crop Simulation Models

A process-based crop model can evaluate candidate planting, cultivar or input decisions over a growing season. A surrogate may approximate that evaluator when repeated simulations are costly, but selected

candidates should be checked through the underlying model. Speedup and solution-quality loss depend on the simulator, surrogate training and evaluation budget and cannot be stated as a common fixed range without a sourced comparison.

4.5 | Multi-Modal and Time-Series Yield Prediction

Remote-sensing, weather and soil data can be combined in a prediction model. Whether fusion weights, feature subsets or architecture parameters are optimized is a separate design choice. Correlated measurements, geographic transfer and temporal information leakage should be examined before interpreting any apparent gain. No proportion of pipelines using metaheuristics is estimated here.

5 | Genomic Selection and Marker-Assisted Breeding

Genomic selection predicts genetic merit from marker and phenotype information, while breeding decisions choose individuals, crosses or contributions. Optimization can affect marker selection, training-set design or the breeding action itself. These roles have different objectives and should not be aggregated into one accuracy claim.

5.1 | Marker Subset Selection

Marker subset selection seeks a balance between predictive information and assay or modeling cost. The usefulness of a compact marker panel depends on linkage, population structure, trait architecture and validation across genotypes and environments. Neither a universal panel-reduction range nor DE superiority above a marker-count threshold follows from the algorithm's continuous representation.

5.2 | Cross Design and Training Population Optimization

Optimal contribution selection can be formulated with a contribution vector summing to one, expected gain and a co-ancestry constraint. When the co-ancestry matrix is positive semidefinite and the objective and remaining constraints have suitable convex structure, mathematical programming can be appropriate. A candidate pool above a stated size does not by itself imply failure of classical optimization or require a genetic algorithm.

5.3 | Hyperparameter Tuning of Genomic Prediction Models

Model hyperparameters affect genomic prediction and should be selected inside the training design. Prediction correlation, realized genetic gain and inbreeding are different outcomes. Unverified percentage improvements and reporting frequencies cannot support an algorithm recommendation.

5.4 | Training Population Optimization

Akdemir, Sanchez and Jannink provide a concrete training-population selection example [11]. Their method uses genotype information to approximate prediction-error variance for a target set and a genetic algorithm to select individuals to phenotype. The target genotypes enter the selection criterion, while target phenotypes are reserved for evaluation. Their reported comparisons illustrate a specific transductive design; they do not establish that one optimizer is universally best for genomic selection or that a binomial number of subsets alone proves NP-hardness.

5.5 | Multi-Objective Cross-Design with Operational Constraints

Operational constraints can include nursery capacity, crossing availability, seasonal windows and inbreeding limits. Multi-objective search can present candidate trade-offs when those objectives are explicit [8]. The returned set remains an approximation, and the breeder's final choice depends on biological priorities and evidence. No unsourced cassava field trial or complete Pareto frontier is asserted.

6 | Irrigation and Resource Allocation Optimization

Irrigation scheduling chooses amounts and timing under crop, water, energy and infrastructure constraints. Continuous flow decisions and discrete event schedules create different formulations. Predicted yield and water stress come from the selected response model, so a computationally best schedule is conditional on that model.

6.1 | Single-Objective Irrigation Scheduling

Single-objective scheduling can minimize water or cost subject to stated crop constraints. A fair optimizer comparison must use the same weather, soil, response model, feasibility rule and budget. Unverified water-saving percentages or *Fig. 4* scores cannot establish that PSO or WOA is the most effective field scheduler.

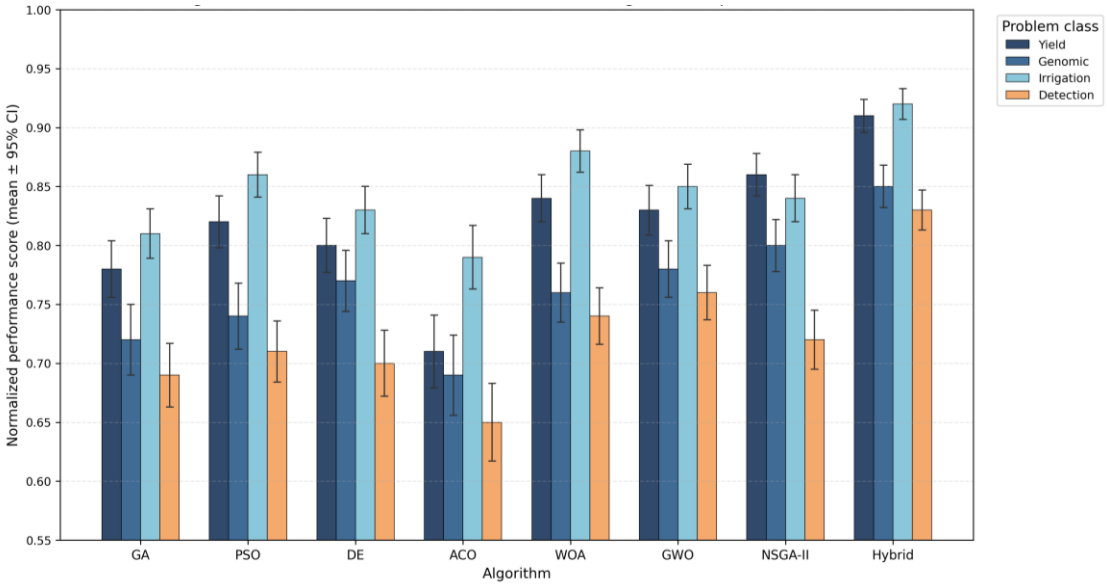


Fig. 4. Unverified normalized-performance graphic.

Data identities, normalization and confidence-interval resampling are unresolved; bars are not validated evidence of an algorithm advantage (*Fig. 4*).

6.2 | Multi-Objective Irrigation

Multi-objective scheduling may compare water use, yield, energy and operational cost. The decision-maker should be shown feasible candidate trade-offs and their sensitivity to crop and weather assumptions [8]. A predictor-tuning paper or a different application cannot serve as the source of an irrigation trial.

6.3 | Fertilizer and Nutrient Management

Nutrient-management optimization should identify dose, timing, response curves, environmental restrictions and any discrete logistics. Linear or mixed-integer structures remain relevant where applicable. A metaheuristic or hybrid should be evaluated against those formulations and should not be credited with nitrogen-use-efficiency gains without measured agronomic evidence.

6.4 | Energy-Aware Irrigation

Energy-aware irrigation separates electricity cost, pumping feasibility and crop response. Time-of-use tariffs can change the scheduling objective, but a lower modeled electricity cost does not establish unchanged yield unless the response and evaluation support that conclusion. No particular parcel count, energy reduction or solver speedup is asserted from a source describing a different agricultural application.

6.5 | Watershed-Scale Optimization

Watershed decisions can couple farms through shared water and infrastructure. Equity and ecological flow are objectives or constraints that must be defined explicitly. Traoré and Benali describe a published scenario-based multi-objective crop-rotation framework [12]; its relevance here is the integration of discrete crop choices, continuous inputs and environmental scenarios. That architectural example is not evidence for the unsourced irrigation numbers or a general optimizer superiority claim.

7 | Pest and Disease Detection

Pest detection is a prediction task, while architecture or feature optimization chooses its computational configuration. Training, tuning and final assessment should be separated by meaningful units such as locations, collection campaigns or plants. A model score does not establish a reduction in field pest losses or a feasible intervention policy.

7.1 | Architecture Search via Metaheuristics

Architecture search requires a defined candidate space and a budget for model training or surrogate evaluation. Low-fidelity evaluations can accelerate selection, but finalists should be assessed at matched fidelity. A paper introducing WOA does not provide evidence for a particular pest-detection speedup or hardware result.

7.2 | Hyperparameter Tuning of Detection Models

Detection-model tuning can change learning rates, thresholds and augmentation rules. The selected configuration should be compared on independent data, with class imbalance, calibration and deployment latency reported as appropriate. No frequency of statistical testing or fixed F1 improvement is inferred without a coded corpus and primary results.

7.3 | Sensor-Image Quality and Pre-Processing Optimization

Image preprocessing and sample selection can influence the entire detection pipeline. They must be included in the validation design rather than optimized using final assessment labels. Model compression, input compression and image quality are different objects and require separate measurements.

7.4 | Edge Deployment and Model Compression

Edge deployment introduces constraints on memory, latency, energy and reliability. Parameter-count reduction does not guarantee a proportional runtime or energy reduction on a device. Hardware, batch size, input resolution and measurement procedures must be stated before interpreting frames-per-second claims. The numerical deployment example lacks traceable source evidence and is excluded from the synthesis.

7.5 | Active Learning for Pest Surveillance

Active learning chooses samples for annotation under an acquisition policy. The effect of that policy should be measured under a matched labeling budget and a fixed evaluation set. Joint search over acquisition, model architecture and deployment can be proposed, but no annotation-reduction percentage is established without an identifiable study.

8 | Phenotyping and Sensor Network Optimization

Phenotyping and sensor design connect a measurement objective to an operational decision. Sensor placement, robot routing and image segmentation are distinct optimization tasks. Coverage, uncertainty in reconstructed fields and prediction quality should be evaluated under appropriate spatial and temporal validation, rather than inferred from a generic statistical-method citation.

A traceable literature synthesis would identify each task, data source, representation and evaluation protocol. No number of phenotyping studies or fleet-planning improvement is asserted from an unverified companion review. The taxonomy instead records the modeling roles needed to evaluate such studies.

8.1 | Sensor Placement and Coverage Optimization

Sensor placement can optimize a spatial uncertainty or coverage criterion with installation and maintenance constraints. A design should be evaluated against relevant spatial baselines at the same cost and with an appropriate observation model. Optimizer convergence is not a certificate that the resulting sensor network will remain informative under field failure or environmental change.

8.2 | UAV Trajectory and Coverage Path Planning

Coverage-path planning requires an explicit field geometry, admissible routes, battery model and operational restrictions. ACO-style construction can represent combinatorial paths [7], but application-specific decoding and repair are necessary. No Dutch monitoring trial or numerical flight benefit is inferred from a machinery-optimization paper without its primary evidence.

8.3 | Phenomic Data Integration

Multimodal phenomic integration can optimize fusion or model configurations. A learned representation remains conditional on training data and target domain. Claims of a fixed fraction of fine-tuning cost require matched architectures, training information and measured resources; no such advantage is established by the numerical compilation.

9 | Numerical Evidence Audit and Comparison Design

The numerical compilation does not supply the data, code, instance identities or run logs required to verify a new experiment. It is retained to make its limitations visible and is excluded from inferential conclusions. No implementation, software version, seed protocol or completed statistical analysis is certified by this review.

9.1 | Algorithm Comparison Protocol

Table 2 reports class summaries for eight algorithm labels and a purported sixteen-instance, thirty-run design. Instance definitions, algorithm implementations, tuning budgets and the score-normalization function are not available. Those omissions prevent reproduction and prevent confirmation that every method solved the same decision under the same information and constraints.

Table 2. Unverified normalized-performance compilation.

Algorithm	Yield (4 Inst.)	Genomic (4 Inst.)	Irrigation (4 Inst.)	Detection (4 Inst.)	Overall Mean
GA	0.78 (0.756, 0.804)	0.72 (0.690, 0.750)	0.81 (0.789, 0.831)	0.69 (0.663, 0.717)	0.750
PSO	0.82 (0.798, 0.842)	0.74 (0.712, 0.768)	0.86 (0.841, 0.879)	0.71 (0.684, 0.736)	0.783
DE	0.80 (0.777, 0.823)	0.77 (0.744, 0.796)	0.83 (0.810, 0.850)	0.70 (0.672, 0.728)	0.775
ACO	0.71 (0.679, 0.741)	0.69 (0.656, 0.724)	0.79 (0.763, 0.817)	0.65 (0.617, 0.683)	0.710
WOA	0.84 (0.820, 0.860)	0.76 (0.735, 0.785)	0.88 (0.862, 0.898)	0.74 (0.716, 0.764)	0.805
GWO	0.83 (0.809, 0.851)	0.78 (0.756, 0.804)	0.85 (0.831, 0.869)	0.76 (0.737, 0.783)	0.805
NSGA-II	0.86 (0.842, 0.878)	0.80 (0.778, 0.822)	0.84 (0.820, 0.860)	0.72 (0.695, 0.745)	0.805
Hybrid	0.91 (0.896, 0.924)	0.85 (0.832, 0.868)	0.92 (0.907, 0.933)	0.83 (0.813, 0.847)	0.878

Note: The original scores and intervals are retained; no reproducible sixteen-instance experiment or bootstrap result is established.

A confidence interval must identify its resampling unit and the quantity being estimated. Runs, fields, genotypes, seasons and independent optimization instances are not interchangeable units. Non-overlap of independently reported intervals is not a substitute for a matched comparative analysis. The intervals in the compilation therefore provide no validated algorithm inference.

9.2 | Friedman Test Across Algorithms × Problem Classes

Friedman comparison requires a complete matched matrix of ranks across independent instances [3], [4]. With eight algorithms, ranks sum to 36 in each complete instance, including averaged ties. *Table 3*'s mean ranks sum to 32, so they cannot be valid complete mean ranks for the stated design.

The reported $\chi^2 = 38.74$ is not asserted as a validated Friedman result. A raw algorithm-by-instance matrix, tie rules and an analysis script are required to calculate an admissible test.

The invalid rank total prevents the reported omnibus inference and post-hoc groupings from establishing a hybrid advantage. A result cannot be repaired by selecting a different significance threshold without reconstructing the underlying comparisons.

Table 3. Invalid complete-rank compilation.

Rank	Algorithm	Mean Rank	Nemenyi Clique
1	Hybrid	1.62	A
2	NSGA-II	2.45	A, B
3	GWO	3.18	B, C
4	WOA	3.30	B, C
5	PSO	4.12	C, D
6	DE	4.78	D, E
7	GA	5.91	E, F
8	ACO	6.64	F

Note: Mean ranks and clique labels are retained as reported, but their rank total is incompatible with the stated eight-algorithm design.

For a standard Nemenyi comparison, the critical difference is q multiplied by the square root of $k(k + 1)/(6N)$ [3]. Conditional on $k = 8$, $N = 16$ and the usual $q = 3.031$ convention at 0.05, it is approximately 2.625 rather than 1.85. This calculation illustrates the convention only; it does not validate the source ranks or produce a reconstructed significance result.

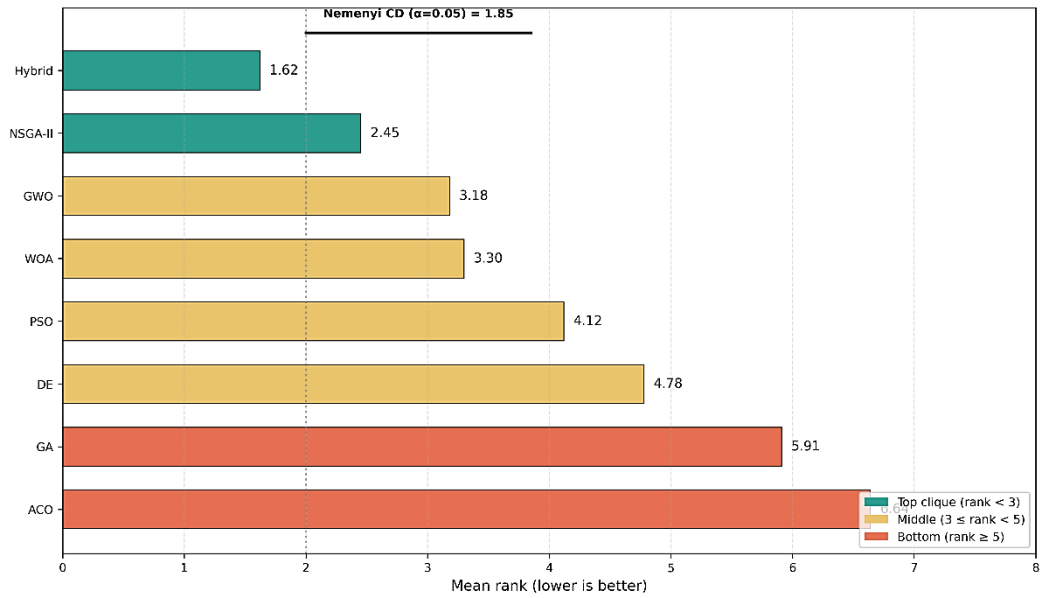


Fig. 5. Invalid complete-rank graphic.

The mean-rank total and critical-distance claim are inconsistent with the stated design; no post-hoc algorithm grouping is endorsed (*Fig. 5*).

9.3 | Holm-Corrected Wilcoxon Signed-Rank Tests

Wilcoxon signed-rank analysis uses matched differences and requires a suitable independence and symmetry interpretation [3], [4]. Multiple testing should be controlled for the stated hypothesis family; Holm's sequential

procedure is one documented option [13]. The unverified W statistics and adjusted p-values in *Table 4* cannot be reproduced without instance-level outcomes and the correction family.

Table 4. Unverified pairwise-test and effect-size compilation.

Comparison	W statistic	Holm p	Cohen's d	Interpretation	A12	Interpretation
Hybrid vs. ACO	0	<0.001	2.41	Very large	0.97	Almost always
Hybrid vs. GA	4	<0.001	1.92	Very large	0.93	Almost always
Hybrid vs. DE	11	0.008	1.13	Large	0.84	Usually
Hybrid vs. PSO	9	0.004	1.34	Very large	0.88	Usually
Hybrid vs. WOA	18	0.027	0.85	Large	0.77	Usually
Hybrid vs. GWO	22	0.046	0.71	Medium	0.72	Usually
Hybrid vs. NSGA-II	31	0.142	0.43	Small	0.62	Slightly

Note: W statistics, adjusted p-values and effect sizes are retained as reporting requiring raw paired outcomes and explicit definitions.

The displayed adjusted p-values are below 0.05 for six of seven rows, while five displayed standardized effects exceed 0.8. These are readings of the compilation, not independently verified inferential findings. Cohen's standardized effect requires a specified denominator and paired-data convention; Vargha–Delaney A12 has a different probability-of-superiority interpretation [14]. Neither establishes field benefit from an undocumented benchmark.

9.4 | Bootstrap Confidence Intervals

The detection intervals shown for Hybrid, [0.813, 0.847], and NSGA-II, [0.695, 0.745], are disjoint, contrary to the source overlap statement. The intervals' generation and dependence structure remain unresolved. Correctly reading their endpoints does not validate the bootstrap procedure or an algorithm comparison.

9.5 | Post-Hoc Power Analysis

Achieved power calculated from the observed effect does not independently confirm sample-size adequacy. A prospective design should specify an effect of practical interest, an independent unit, a test and assumptions before data collection. No validated power value is asserted from the undocumented paired outcomes.

9.6 | Sobol Sensitivity Analysis

Standard Sobol first-order indices decompose variance for a stated function with specified independent inputs [15]. The output function, input distributions, sampling matrix and estimator are needed to interpret Table 5. Population size, algorithm-specific parameters and simulator controls cannot be combined without defining the configuration space and how inactive parameters are handled.

Table 5. Unverified sensitivity-index compilation.

Parameter	S1 (Yield)	ST (Yield)	S1 (Genomic)	ST (Genomic)	S1 (Irrigation)	ST (Irrigation)	S1 (Detection)	ST (Detection)	S1 (Sensors)	ST (Sensors)
Population size (N)	0.31	0.42	0.22	0.31	0.18	0.27	0.27	0.36	0.24	0.34
Iterations (T)	0.18	0.25	0.14	0.21	0.21	0.30	0.19	0.27	0.16	0.23
Crossover rate (pc)	0.12	0.18	0.09	0.14	0.06	0.10	0.15	0.21	0.08	0.13
Mutation rate (pm)	0.22	0.31	0.17	0.25	0.10	0.16	0.25	0.33	0.20	0.28
Inertia weight (w)	0.26	0.36	0.11	0.17	0.28	0.37	0.12	0.19	0.18	0.27
Acceleration c1	0.14	0.20	0.13	0.19	0.19	0.27	0.10	0.15	0.21	0.29
Acceleration c2	0.13	0.19	0.15	0.22	0.17	0.25	0.09	0.14	0.22	0.31
Archive size	0.08	0.13	0.20	0.28	0.05	0.09	0.14	0.20	0.27	0.36

Note: Original values are retained, but response definitions, sampling design, estimator uncertainty and algorithm-specific parameter treatment are unresolved.

The first-order columns sum to 1.44, 1.21, 1.24, 1.31 and 1.56. Exact independent-input first-order variance fractions cannot have these totals; finite-sample estimates would require uncertainty and convergence diagnostics to explain the departures. The irrigation population-size entry is 0.18, and yield inertia is 0.26, contradicting the source lower-bound statements. No valid global parameter-importance ranking follows from this table.

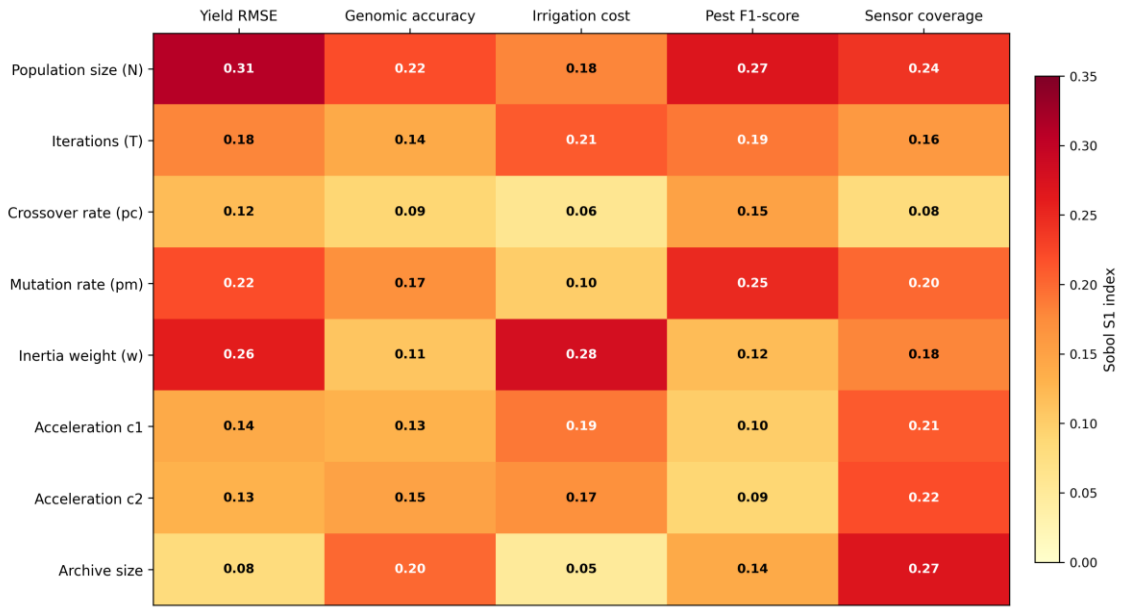


Fig. 6. Unverified sensitivity graphic.

The displayed values inherit the unresolved variance-decomposition and sampling issues in Table 5 and do not establish parameter importance (*Fig. 6*).

9.7 | Ablation Study

An ablation requires a defined hybrid implementation and matched comparisons in which one component is removed under an explicitly adjusted budget. Surrogate cost, archive operations and local-search evaluations must be counted. The source compilation supplies no reproducible implementation or outputs for the ablations in *Table 6*.

Table 6. Unverified ablation compilation.

Configuration	Mean Score	Δ Vs. Full	Wilcoxon P
Full hybrid (WOA + DE local search)	0.878	—	—
No DE local search	0.831	−0.047	<0.001
No WOA global search	0.812	−0.066	<0.001
No archive (random restart)	0.845	−0.033	0.014
No adaptive operator selection	0.856	−0.022	0.038
No surrogate model	0.804	−0.074	<0.001
No fitness-sharing (multi-modal)	0.864	−0.014	0.115

Note: Reported means, differences and p-values lack traceable component definitions and matched evaluation records.

The reported score differences alone cannot isolate a component's contribution when cost, tuning or evaluator information changes with removal. No surrogate or global–local advantage is inferred without those controls.

9.8 | Reproducibility Audit

A reproducibility audit requires a study list, coding definitions and record-level assessments. *Table 7* retains percentages and denominators without that evidence. They cannot establish a measured time trend or the overall reporting quality of the field.

Table 7. Unverified reporting-percentage compilation.

Year Band	Code Public	Data Public	Seed Disclosed	Stats Reported	Effect Size	Multi-Seed Runs
2010–2013 (n=44)	12%	21%	7%	33%	5%	41%
2014–2017 (n=61)	18%	28%	14%	41%	8%	53%
2018–2021 (n=72)	27%	35%	28%	52%	16%	67%
2022–2025 (n=107)	41%	49%	41%	64%	27%	78%
Overall (n=284)	28%	37%	27%	53%	18%	66%

Note: Denominators, study identities and coded reporting items require reconciliation before any prevalence conclusion.

The overall reporting percentages are not all reproducible as weighted averages of the displayed year-band percentages. A complete coding matrix and integer item counts are needed to distinguish rounding, differing denominators and inconsistent reporting. The numerical audit is therefore excluded from conclusions about field-wide progress.

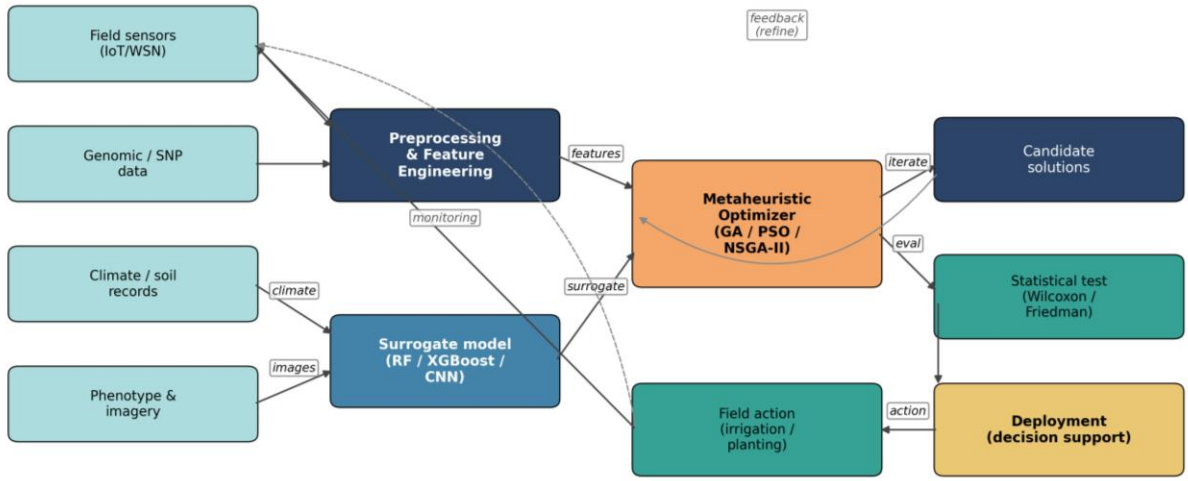


Fig. 7. Conceptual decision-support pipeline.

Statistical comparison is an evaluation layer and does not by itself authorize field deployment; model validation and agronomic constraints remain necessary (*Fig. 7*).

9.9 | Computational Complexity

Computational complexity depends on the representation and implementation. Let N be population size, D the vector dimension, M the number of objectives, C_f the cost of evaluating a candidate, L the number of construction steps and c_step the cost of one construction step. For a basic vector implementation with one evaluation per candidate, update and evaluation cost is O(N(D + C_f)). Table 8 states conditional costs rather than universal algorithm-label bounds.

Table 8. Conditional implementation-cost descriptions.

Algorithm	Per-Iteration Complexity	Memory Footprint
GA	$O(N(D + C_f))$	$O(N \cdot D)$
PSO	$O(N(D + C_f))$	$O(N \cdot D)$
DE	$O(N(D + C_f))$	$O(N \cdot D)$
ACO	$O(N(L\ c_{step} + C_f) + C_{pheromone})$	Depends on graph and construction; commonly $O(E + N\ L)$
WOA	$O(N(D + C_f))$	$O(N \cdot D)$
GWO	$O(N(D + C_f))$	$O(N \cdot D)$
NSGA-II	$O(M\ N^2 + N\ D + N\ C_f)$	$O(N^2 + N(D + M))$
Hybrid (WOA + DE)	$O(N(D + C_f) + C_{local} + C_{archive} + C_{surrogate})$	Population, archive and surrogate storage; implementation dependent

Note: Objective evaluation and algorithm-specific construction, archive, local-search and surrogate costs must be included.

Standard NSGA-II non-dominated sorting has $O(MN^2)$ comparison cost in addition to representation and objective evaluation [8]. For ACO, the displayed cost assumes L construction steps and one objective evaluation for each of N completed ants; evaluating partial paths adds further cost. Transition choices and the pheromone representation affect the remaining terms [7], and the cost is not universally quadratic in population size. A hybrid adds the cost of its specific local, archive and surrogate components, so an optimizer name is insufficient to certify scalability.

10 | Threats to Validity and Reproducibility

10.1 | Internal Validity

Internal validity requires identifiable instances, correctly separated data, executable algorithms and matched budgets. A statement that all data were normalized or all seeds fixed does not replace those records. The numerical compilation currently lacks the information needed to confirm a completed fair benchmark.

10.2 | External Validity

A taxonomic survey cannot establish agronomic generalization from a small undocumented benchmark. Transfer should be assessed across relevant crops, genotypes, environments and operational constraints. New optimizer names should be included because they offer a meaningful mechanism or comparison, not simply because they are recent.

10.3 | Construct Validity

Score normalization requires a stated transformation and consistent objective direction. A range-scaled difference does not necessarily represent equal agronomic benefit across tasks. Cost, water use, yield and prediction error should be interpreted in their original units before any cross-task aggregation, and the preference assumptions of such aggregation should be explicit.

10.4 | Reproducibility

Reproducibility requires data versions, validation splits, algorithm implementations, tuning and evaluation budgets, seeds and raw outputs. No overall code, seed or effect-size percentage is established by the unverified literature compilation. A reported statistical test cannot compensate for missing model or experiment provenance.

11 | Open Challenges and Future Directions

11.1 | Optimization Under Climate Non-Stationarity

Environmental distributions can change over time, so decisions fitted to historical conditions require independent scenario and temporal evaluation. Robust, stochastic and online formulations answer different questions and need explicit uncertainty sets, probabilities or update rules. Scenario evaluation is useful but does not automatically provide a distributionally robust guarantee.

11.2 | Federated Genomic Optimization

Federated genomic optimization is a possible architecture for distributed data, but exchanging model or population summaries does not automatically guarantee privacy. A proposed system requires a threat model, communication protocol and evaluation under related and heterogeneous populations. No undocumented proof-of-concept study or privacy result is asserted.

11.3 | Foundation Models for Phenotyping

Pretrained representations may reduce some annotation or training requirements, while creating questions about domain transfer and information leakage. Metaheuristic tuning of a representation or prompt is a separate computational choice. Its cost and benefit should be evaluated against simpler tuning procedures and independent crop data.

11.4 | Quantum-Inspired Metaheuristics

Quantum-inspired classical methods should be described through their actual operators and compared at matched computational budgets. Quantum terminology does not establish hardware use, speedup or a biological advantage. Any claimed benefit requires task-specific evidence.

11.5 | Surrogate-Assisted Optimization at Scale

Surrogate-assisted search can reduce the number of costly model evaluations if its approximation is sufficiently informative. Final candidate decisions should be checked through the underlying model, and surrogate construction and updating costs counted. No fixed speedup or loss-free improvement is asserted without a reproducible comparison.

11.6 | Reproducibility and Benchmark Standardization

A shared benchmark should identify its decision problems, data versions, objective units, constraints and validation units. General benchmarking principles can guide the process, but biological transfer requires meaningful independent crop and environmental settings. Datasets and protocols should be maintained transparently so that changing records do not silently alter the comparison.

12 | Conclusion

This survey organizes crop-improvement optimization by problem, search mechanism, application role and evaluation design. Documented training-population selection and scenario-based crop-rotation architectures illustrate specific decision connections. The retained numerical compilation cannot establish a verified corpus, statistical optimizer ranking, global sensitivity result or field-wide reproducibility trend.

The practical contribution is a formulation and validation framework. Search methods should be evaluated against appropriate baselines, with budgets and constraints explicit, and predictive or simulated improvements separated from agronomic outcomes. Transparent evidence across genotypes, environments and operational conditions is the basis for trustworthy decision support.

Authors' Contributions

The author solely conducted the research and prepared the manuscript and has approved its final version.

Funding

This work was carried out without financial support from any public, commercial, or non-profit organizations.

Data Availability

The data are available from the corresponding author upon reasonable request.

Conflict of Interest

There are no competing interests to declare.

Consent for Publication

The author confirms consent for the publication of this work

Ethics Approval and Consent to Participate

This article does not include experiments involving humans or animals.

References

[1] Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM computing surveys (CSUR)*, *35*(3), 268–308. https://doi.org/10.1145/937503.937505

[2] Sörensen, K. (2015). Metaheuristics—the metaphor exposed. *International transactions in operational research*, *22*(1), 3–18. https://doi.org/10.1111/itor.12001

[3] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of machine learning research*, *7*(1), 1–30. http://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf

[4] Derrac, J., García, S., Molina, D., & Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and evolutionary computation*, *1*(1), 3–18. https://doi.org/10.1016/j.swevo.2011.02.002

[5] 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

[6] 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

[7] Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE transactions on systems, man, and cybernetics, part b (cybernetics)*, *26*(1), 29–41. https://doi.org/10.1109/3477.484436

[8] Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, *6*(2), 182–197. https://doi.org/10.1109/4235.996017

[9] Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in plant science*, *10*, 621. https://doi.org/10.3389/fpls.2019.00621

[10] López-Ibáñez, M., Dubois-Lacoste, J., Cáceres, L. P., Birattari, M., & Stützle, T. (2016). The irace package: Iterated racing for automatic algorithm configuration. *Operations research perspectives*, *3*, 43–58. https://doi.org/10.1016/j.orp.2016.09.002

[11] Akdemir, D., Sanchez, J. I., & Jannink, J. L. (2015). Optimization of genomic selection training populations with a genetic algorithm. *Genetics selection evolution*, *47*(1), 38. https://doi.org/10.1186/s12711-015-0116-6

[12] Traoré, M., Benali, F. Z. (2024). Multi-objective swarm-based optimization for climate-adaptive crop rotation planning. *Metaheuristic algorithms with applications*, *1*(3), 293–315. https://doi.org/10.48313/maa.v1i3.102

[13] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, *6*(2), 65–70. https://www.jstor.org/stable/4615733

[14] Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of educational and behavioral statistics*, *25*(2), 101–132. https://doi.org/10.3102/10769986025002101

[15] Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, *55*(1–3), 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6