Systematic Review Article
Scientific Journal of Mathematics and Statistics
Volume 1, 2026, **pp.** 1-22, **Article ID:** PM2610303002

Feature Selection for High-Dimensional Predictive Modelling: A Systematic Review of Statistical Foundations, Stability, Causal Awareness, Performance Trends, and Practical Guidance

Madhab Chandra Jena[1, *]

[1]*Gandhi Institute for Technology (GIFT), Autonomous, Bhubaneswar, India*

Article History

Received: 15 May, 2026 Revised: 28 July, 2026 Accepted: 24 August, 2026 Published: 25 September, 2026

Abstract:

Introduction: High-dimensional predictive modelling is increasingly used in genomics, healthcare, finance, engineering, and artificial intelligence, yet its reliability is constrained by redundant variables, small-sample settings, computational burden, unstable feature subsets, and limited interpretability.

Methods: This systematic review synthesises peer-reviewed empirical studies published between 2020 and 2025 on feature selection for high-dimensional classification and regression. Following a PRISMA-guided process, 40 studies were included and examined through structured data extraction, methodological-quality appraisal, vote counting by direction of reported effect, and thematic synthesis.

Results: Conventional filter, wrapper, embedded, and hybrid methods were compared in relation to predictive performance, computational efficiency, stability, interpretability, reproducibility, and suitability for different data regimes. The evidence indicates that hybrid, ensemble, embedded, evolutionary, and learning-based approaches frequently report favourable performance, but cross-study superiority cannot be inferred because datasets, sample sizes, metrics, baselines, and validation procedures differ substantially. Filter methods remain attractive for scalability, whereas wrapper and adaptive methods provide richer interaction modelling at greater computational and tuning cost. Causal feature selection, stability reporting, external validation, and reproducible benchmarking remain insufficiently developed.

Conclusion: The review separates feature selection from feature transformation, introducing an author-proposed three-dimensional analytical framework stability awareness, causal awareness, and data-regime adaptivity reclassifying the included studies using that framework, and providing practitioner-oriented guidance for method selection. The findings support context-specific rather than universal recommendations and identify priorities for transparent, stable, causally informed, and scalable feature-selection research.

Keywords: Feature selection, high-dimensional data, predictive modelling, systematic review, stability selection, causal feature selection, machine learning, explainable artificial intelligence.

1. INTRODUCTION

The proliferation of data-driven applications in many sectors, including genomics, healthcare, finance, and remote sensing, has led to the creation of high-dimensional datasets which in many cases surpass the availability of data. High-dimensional data usually refer to data sets where the number of variables are significantly larger than the

[*]Address correspondence to this author at Gandhi Institute for Technology (GIFT), Autonomous, Bhubaneswar, India; E-mail: madhab_jena@rediffmail.com

Cite as: M.C. Jena. "Feature selection for high-dimensional predictive modelling: a systematic review of statistical foundations, stability, causal awareness, performance trends, and practical guidance," *Sci. J. Mathe. Statis.*, Vol. 1, pp. 1–22. Art No. PM2610303002, 2026.

observations hence posing an issue to predictive modelling [1, 2]. Such challenges include overfitting, multicollinearity, high computational cost, and reduced model interpretability. On one hand, high-dimensional datasets contain rich and valuable information, while on the other, the large number of features provides difficulty in extracting accurate and meaningful insights [3]. In this regard, feature selection has become a critical preprocessing method in the high-dimensional predictive modelling process and tends to establish a smaller number of useful features with the greatest significance to the predictive model, thus alleviating the above-mentioned issues.

The selection of features is integral to the enhancement in model accuracy and efficiency in computation including interpretability. The feature selection techniques can improve the efficiency and speed of the predictive models while reducing the risk of overfitting by decreasing the dimensionality of the data [1, 4]. Moreover, feature selection increases the interpretability of models as only the most important variables are stored, hence making it easier to comprehend the latent relationships among the data by the researcher and practitioner. The demand for efficient feature selection is further intensified in high-dimensional features like genomics or finance where the model usually makes use of hundreds or thousands of features [2, 4]. In contrast to dimensionality reduction algorithms, such as the Principal Component Analysis (PCA), which projects the data into a new space, feature selection algorithms preserve the original significance of the variables, which is of great importance in areas where it is important to know the role of every feature to make a decision.

Although the role of feature selection has increased, the procedure of identifying the appropriate approach is multifaceted, especially when dealing with data that has high dimensions. Various feature selection methods including filter, wrapper, embedded, and hybrid approaches are based on different statistical principles to identify relevant features and are characterized with different strengths and weaknesses [5, 6]. Filter methods evaluate feature relevance independently of a predictive model using statistical measures such as correlation or mutual information. Although these are computationally efficient and can be scaled to large dimensional datasets, they are not always good at capturing feature interactions [6]. On the contrary, wrapper methods can capture model-specific interactions but require greater computation and may be prone to overfitting [5]. Embedded methods integrate feature selection into model training and can provide a practical balance between computational efficiency and predictive performance [7]. The recent development of hybrid methods, which are based on a combination of various feature selection strategy, is a promising way of solving the issues; it utilizes the strengths of each strategy to enhance performance in a difficult high-dimensional environment [8].

There is an increasing requirement to have a systematic knowledge of these feature selection techniques, especially where high-dimensional predictive modelling is involved. This review seeks to critically review the different feature selection techniques applied in high-dimensional data analysis in relation to their statistical basis, trend of performance and practical utilization. This review entails synthesis of the recent research on feature selection by examining the studies published within the period between 2020 and 2025 to allow getting a comprehensive picture of the methods that are being used and the underlying statistical principles and their applicability to the various fields. By doing so, it will answer critical research questions, *i.e.,* which methods are most effective in various high-dimensional domains, how they are compared in terms of predictive accuracy, and obstacles and constraints while using these methods in real-life contexts.

For the purposes of this review, feature selection and feature transformation are treated as distinct operations. Feature selection identifies a subset of the original variables and therefore preserves their original meaning, measurement scale, and potential domain interpretation. By contrast, feature-transformation or dimensionality-reduction methods, including PCA, autoencoders, and representation-learning models, to construct new latent variables from combinations of the original features. A study using PCA or an autoencoder was included only when the transformation was followed by an explicit mechanism that ranked, selected, masked, weighted, or removed original input variables. Studies that produced only latent representations without identifying a subset of original features were excluded. This boundary was adopted to prevent conceptually different dimensionality-management procedures from being treated as equivalent.

This review addresses the following questions: Which feature-selection methods were evaluated in high-dimensional predictive studies published from 2020 to 2025? How do filter, wrapper, embedded, and hybrid approaches differ in their statistical foundations, strengths, and limitations? To what extent do current methods address stability, causal relevance, and adaptation to specific data regimes? What patterns emerge regarding predictive performance, computational efficiency, interpretability, and reproducibility? What evidence-based guidance can support practitioners in selecting an appropriate method?

The rationale behind this systematic review is the growing need to have precise and interpretable predictive models to be able to manage the complexity of high-dimensional data. With the ever-increasing amount and type of accessible data, particularly in fields such as genomics, healthcare and finance, there is a need to develop efficient, scalable, and interpretable feature selection procedures. Nevertheless, the large number of existing feature selection methods have their own assumptions, strengths and limitations making it difficult for researchers and practitioners to make informed decisions regarding which method to apply to a given task. Also, despite the significant advancements that have been achieved, there are still a number of issues, including the ability to scale some of the

techniques to ultra-high-dimensional data, the consistency of particular features across various data sets, and the interpretability of more complicated techniques, especially those combined with machine learning and deep learning systems.

The existing reviews of high-dimensional feature selection commonly organise methods into filter, wrapper, embedded, and hybrid categories. Although useful for describing algorithmic implementation, this conventional taxonomy provides limited guidance on why a method succeeds or fails under particular statistical conditions. Previous syntheses also tend to compare reported accuracy values without adequately addressing heterogeneity in datasets, dimensionality ratios, class distributions, validation designs, predictive models, and evaluation metrics. Consequently, their conclusions frequently remain descriptive and do not offer defensible cross-study evidence of methodological superiority.

Three additional limitations remain insufficiently addressed. First, stability of the selected feature subset is rarely integrated with predictive performance, even though unstable selection can limit the reproducibility and scientific interpretation. Second, most reviews do not distinguish predictive association from causal relevance, despite the importance of this distinction in clinical, financial, and policy decisions. Third, limited guidance is available on matching feature-selection methods to specific data regimes, including $p \gg n$ settings, which strongly correlates with predictors, multimodal data, class imbalance, streaming observations, and restricted computational resources. Literature also lacks recent synthesis that combines study-quality assessment, structured evidence synthesis, statistical limitations, and practitioner-oriented method-selection guidance. This review addresses these gaps through a PRISMA-guided examination of studies published from 2020 to 2025.

2. METHODOLOGY

The study follows the PRISMA guidelines to provide a thorough study of the current developments in feature selection methods of high-dimensional predictive modelling. The review is performed according to the guidelines of systematic reviews, such as transparent identification of studies, screening, eligibility check, and evidence synthesis. To provide rigor, transparency and reproducibility of findings, a structured and replicable approach was adopted.

2.1. Search Strategy

The literature search covered peer-reviewed studies published between January 1, 2020 till December 31, 2025. The searches were last updated on July 14, 2026, but studies published after December 31, 2025 were not eligible for inclusion. Four publication platforms were searched: ScienceDirect, IEEE Xplore, SpringerLink, and Taylor & Francis Online. ScienceDirect and Scopus were treated as separate resources; therefore, the previously used term "ScienceDirect (Scopus)" was removed.

The search syntax was adapted to the indexing fields and search functions supported by each platform. The core search concepts included "feature selection", "high-dimensional data", "predictive modelling", "machine learning", "classification" and "regression". Searches were restricted to English-language, peer-reviewed journal research articles. Conference papers, review articles, editorials, book chapters, protocols, short communications, and non-peer-reviewed manuscripts were excluded.

The searches conducted across ScienceDirect, IEEE Xplore, SpringerLink, and Taylor & Francis Online identified a total of 1,245 records. Following the removal of 876 duplicate records, 369 unique records remained for title-and-abstract screening. Table **1** presents the database-specific search expressions and the 369 unique records contributing to the screened pool. Records appearing more than once were counted only once in the screened pool.

2.2. Study Selection Process

A total of 1,245 records were identified, with removal of 876 duplicate records. The remaining 369 unique records underwent title-and-abstract screening further excluding 269 records as they did not meet the predefined criteria concerning the study focus, feature-selection procedure, publication type, or empirical evaluation. Full-text reports were sought for the remaining 100 records. Six reports could not be retrieved despite searches through institutional subscriptions, publisher websites, academic repositories, and other accessible sources. Consequently, 94 full-text reports were assessed for eligibility. Fifty-four reports were excluded because they failed to meet one or more of the eligibility criteria, resulting in the inclusion of 40 empirical studies in the systematic review. The complete study-selection process is presented in Fig. (**1**) and corresponds to the search results reported in Table **1**.

2.2.1. Inclusion Criteria

Studies were included if they:

1. Investigated feature selection involving the explicit selection, ranking, weighting, masking, or removal of original input variables in a high-dimensional dataset.
2. Evaluated statistical, computational, machine-learning, optimisation-based, or hybrid feature-selection method.
3. Reported an empirical assessment using real-world, publicly available, simulated, or recognised benchmark data.
4. Applied feature selection within a classification, regression, survival-analysis, or related predictive-modelling task.
5. Published as peer-reviewed English-language journal research articles between January 1, 2020 and December 31, 2025.

2.2.2. Exclusion Criteria

Studies were excluded if they:

1. Applied only feature transformation or latent representation methods, such as PCA, autoencoders, or representation learning, without explicitly selecting, ranking, weighting, masking, or removing original variables.
2. Presented a theoretical, conceptual, or methodological proposal without empirical evaluation.
3. Did not focus on high-dimensional data or predictive modelling.
4. Were review articles, conference papers, editorials, book chapters, protocols, short communications, or non-peer-reviewed manuscripts.
5. Did not provide sufficient methodological information to determine how the original variables were selected or evaluated.

2.3. Data Extraction

A structured data-extraction protocol was applied to all included studies to improve consistency and comparability. Information was extracted concerning study characteristics, feature-selection methodology, application context, data structure, evaluation design, and reported findings. The following information was recorded for each study:

- Authors, publication year, and publication source.
- Feature-selection method and conventional methodological category.
- Statistical, computational, machine-learning, or optimisation procedures.
- Application domain.
- Dataset source and relevant data characteristics.
- Predictive learner and evaluation design, where reported.
- Performance measures, including accuracy, AUC, F1-score, MCC, RMSE, MAE, C-index, recall, or error rate.
- Direction and interpretation of the reported result relative to the study's stated comparator.

This structured extraction supported a systematic comparison of methodological categories, application domains, evaluation practices, data-regime considerations, and the direction of reported findings across the included studies (Table **2**).

This structured extraction allowed systematic comparing of cross-studies and synergy among classes of methods.

2.4. Screening and Quality Assessment

The screening process was conducted in two stages.

1. The titles and abstracts of the 369 unique records were assessed against the predefined eligibility criteria.
2. The 94 full-text reports retrieved were evaluated to determine final eligibility. Two reviewers participated in the screening and full-text assessment processes, and disagreements were resolved through item-by-item discussion and consensus.

Table 1. Database search strategy and records in the screened pool.

Platform	Database-Specific Search Expression	Publication Limits	Search-Update Date	Unique Records Contributing to Screened Pool	Included Studies
ScienceDirect	"Feature selection" AND ("high-dimensional data" OR "high dimensionality") AND ("predictive model*" OR "machine learning" OR "statistical learning") AND (classification OR regression)	2020–2025; research articles; English	July 14, 2026	130	10
IEEE Xplore	("All Metadata":"feature selection") AND (("All Metadata":"high-dimensional data") OR ("All Metadata":"high dimensionality")) AND (("All Metadata":"predictive model") OR ("All Metadata":"machine learning") OR ("All Metadata":"statistical learning")) AND (("All Metadata":"classification") OR ("All Metadata":"regression"))	2020–2025; journals; English	July 14, 2026	96	10
SpringerLink	"Feature selection" AND ("high-dimensional data" OR "high dimensionality") AND ("predictive modelling" OR "machine learning" OR "statistical learning") AND (classification OR regression)	2020–2025; research articles; English	July 14, 2026	73	10
Taylor & Francis Online	"Feature selection" AND ("high-dimensional data" OR "high dimensionality") AND ("predictive modelling" OR "machine learning") AND (classification OR regression)	2020–2025; journal articles; English	July 14, 2026	70	10
Total	-	-	-	369	40

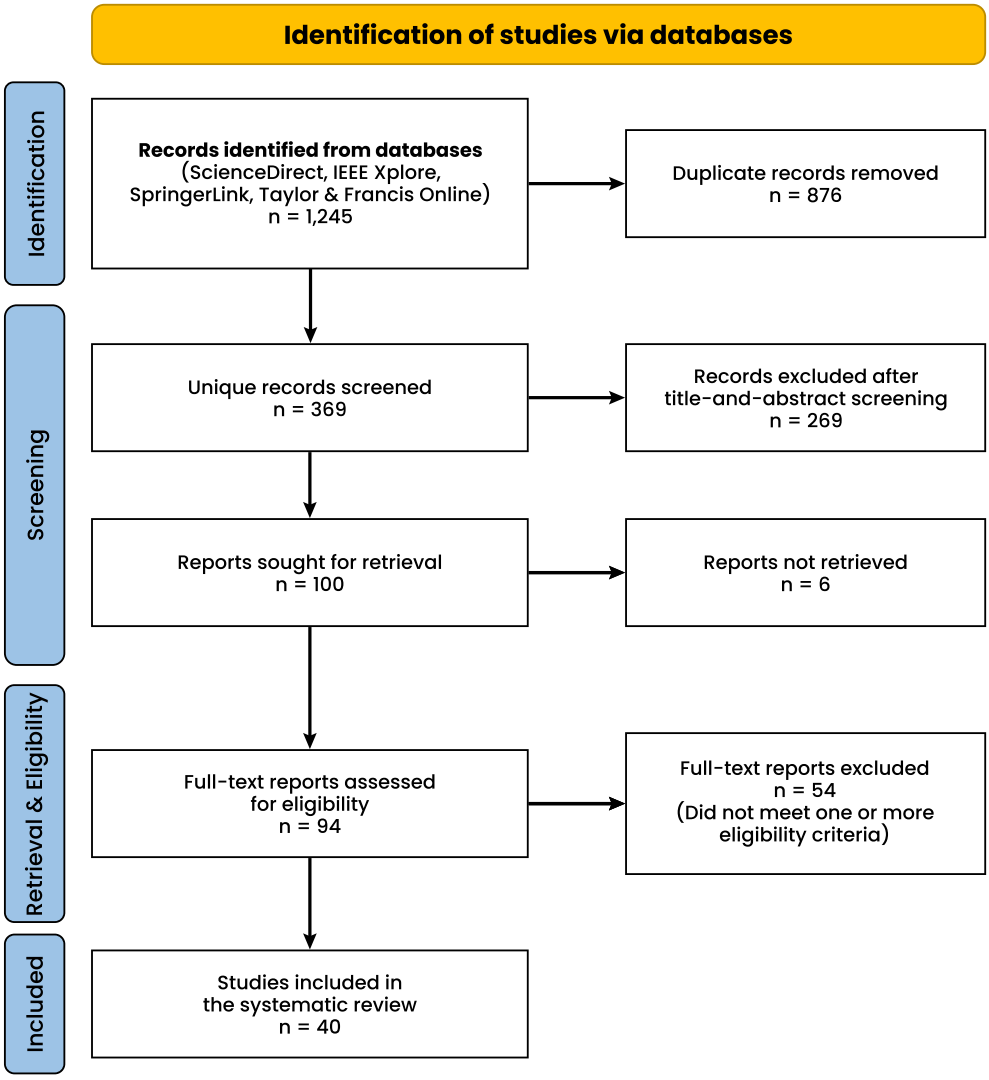

Fig. (1). PRISMA flow diagram for study identification, screening, retrieval, eligibility assessment, and inclusion.

Methodological quality was assessed using a customised checklist informed by systematic-review reporting principles and machine-learning evaluation considerations. The assessment examined the clarity of study objectives, appropriateness of the feature-selection method, adequacy of experimental validation, integrity of the feature-selection pipeline, transparency of evaluation reporting, and reproducibility of the analysis.

Inter-reviewer agreement was evaluated using the paired decisions presented in Table **3**. The resulting Cohen's kappa coefficient was $\kappa = 0.6082$. Under commonly used interpretation thresholds, this value lies at the boundary between moderate and substantial agreement and should not be described unconditionally as "strong agreement." All disagreements were resolved through discussion before the final inclusion decisions were confirmed.

The agreement matrix contains 50 paired reviewer decisions. The manuscript should clearly state whether these decisions represent a documented subsample or a particular screening stage. If the 50 decisions do not constitute the actual agreement dataset, Cohen's kappa should be recalculated using the complete retained reviewer-decision log.

2.5. Evidence-Synthesis Strategy

The included studies differed substantially in their application domains, datasets, sample sizes, dimensionality ratios, outcome prevalence, predictor structures, predictive learners, comparator methods, hyperparameter-tuning procedures, validation designs, and reported performance measures. The studies were therefore not treated as statistically exchangeable. Accuracy, AUC, F1-score, MCC, RMSE, MAE, C-index, recall, error rate, and other reported

outcomes were retained in their original forms and were not pooled or converted into a common performance scale.

The evidence synthesis combined five complementary components:

1. Structured study tabulation
2. Direction-of-effect vote counting
3. Thematic synthesis
4. Methodological quality appraisal
5. Domain-by-method evidence mapping

A favourable direction of effect indicated that an included study reported improvement relative to at least one comparator evaluated within that study. Vote counting was used only to describe the direction of published findings. It was not used to estimate a pooled effect size, calculate statistical significance across studies, or establish the superiority of one feature-selection category over another.

No cross-study means, weighted averages, pooled standard deviations, normalised performance scores, ANOVA tests, or category-level inferential comparisons were calculated. When an individual study reported results for multiple datasets, classifiers, or feature-selection configurations, those findings were described within the context of that study rather than reduced to a single aggregated value. This approach prevented studies with numerous data sets or experimental combinations from exerting disproportionate influence over the synthesis.

The conclusions of the review therefore concern recurring methodological patterns, practical trade-offs, validation quality, stability, causal awareness, and suitability for different data regimes. They do not represent numerical rankings of feature-selection categories.

2.6. Quality and Risk-of-Bias Assessment

Each study was rated from 0 to 2 across five domains:

- Validation quality
- Dataset quality
- Selection-pipeline integrity
- Reproducibility
- Reporting and risk of bias

The detailed scoring rules used for the five quality-appraisal domains are presented in Table **4**. These criteria were applied consistently across all included studies to support a structured and transparent assessment of methodological quality and risk of bias.

Using the criteria defined in Table **4**, each included study was assigned a score for validation quality, dataset quality, selection-pipeline integrity, reproducibility, and reporting or risk of bias. The resulting domain-level scores, total scores, and overall methodological-quality judgements are presented in Table **5**.

2.7. Inter-Reviewer Agreement

Inter-reviewer agreement was evaluated using Cohen's kappa. The calculated coefficient was $\kappa = 0.6082$. Under the commonly cited Landis and Koch interpretation thresholds, values from 0.41 to 0.60 are generally described as moderate agreement, whereas values from 0.61 to 0.80 are described as substantial agreement. The obtained coefficient lies close to the boundary between these categories and is therefore interpreted as moderate-to-substantial agreement rather than unequivocally strong agreement. Disagreements were reviewed individually and resolved through discussion and consensus before the final study-selection decisions were established.

3. FEATURE SELECTION METHODS FOR HIGH-DIMENSIONAL DATA

One of the key preprocessing tasks in predictive modelling, usually in a high-dimensional predictive situation where the number of variables greatly outweighs the number of observations, is feature selection. High-dimensional data is often present in areas like genomics, bioinformatics, medical diagnostics and sensor-based system where features that are redundant, irrelevant, or noisy can grossly diminish model performance. The objective of feature selection is to identify an informative subset of variables while reducing redundancy, computational burden, and the risk of overfitting.

In predictive modelling, feature selection has numerous applications. First, it increases model interpretability by lessening complexity thus facilitating better understanding of relationships between predictors and the outcomes. Second, it enhances the generalization performance through the curse of dimensionality mitigation and avoiding over-fitting, particularly when operating in small sample size and high dimensional contexts. Third, it enhances computational efficiency which is essential in large scale or real time. In the recent research, including the works of [8, 10], it is confirmed that successful feature selection may greatly contribute to the increase of classification accuracy and a decrease in the complexity of models.

The need to design robust, adaptive and scalable feature selection methods that can process noise, correlated and heterogeneous data are becoming important in the research of feature selection in contemporary times. Consequently, there exists a broad spectrum of techniques both in the statistical, machine learning, and evolutionary paradigms that have been devised and utilized in various areas of application.

3.1. Analytical Refinement of Feature Selection Taxonomy

Although the classical taxonomy of feature selection methods into filter, wrapper, embedded and hybrid categories offers a good organizational grounding, it falls short in explaining behavior in performance in modern high-dimensional learning settings. More specifically, taxonomy is primarily procedural rather than theoretical, classifying methods according to how they are conducted, as opposed to why they perform well or poorly under the particular statistical conditions.

Table 2. Characteristics of the included studies.

Ref.	Authors (Year)	Feature Selection Method	Statistical Analysis Used	Application Domain	Data Source	Evaluation Metrics	Performance Outcomes
[1]	Reddy and Mishra (2024)	Fuzzy memetic algorithm	Fuzzy logic, GA	Pattern recognition	UCI datasets	Accuracy, convergence	Stable performance
[2]	Yu *et al.* (2025)	Evolutionary multitask FS	Multi-objective optimization	Bioinformatics	Gene expression	Accuracy, sparsity	Good trade-off balance
[3]	Venâncio and Batista (2025)	Cooperative co-evolution	Evolutionary computation	High-dimensional datasets	Public benchmarks	Accuracy, convergence	Effective dimensionality reduction
[4]	Wang *et al.* (2025)	Dual-objective evolutionary FS	Multi-objective optimization	Gene expression	Accuracy, sparsity	Balanced trade-offs	Parameter tuning
[5]	Shaer and Shami (2024)	RL wrapper FS	Reinforcement learning	Industrial datasets	Accuracy	Improved stability	Training overhead
[6]	Rojas-Velazquez *et al.* (2025)	MCC-based recursive FS	Correlation analysis	Genomics	Gene expression datasets	MCC, accuracy	Strong low-sample performance
[7]	Bhattacharjee *et al.* (2022)	highMLR (ML-based FS)	Survival modeling	Cancer genomics	TCGA datasets	C-index	Handles censoring well
[8]	Salhi *et al.* (2025)	Hybrid AI-driven FS	ML ensemble	High-dimensional data	Accuracy	High performance	Risk of overfitting
[9]	Baba *et al.* (2025)	Correlation-based FS + SVM	Correlation analysis, SVM classification	General high-dimensional data	Benchmark datasets	Accuracy, F1-score	Improved accuracy *vs* baseline FS
[10]	Wei *et al.* (2025)	RL-enhanced PSO	Reinforcement learning, PSO	Generic high-dimensional data	Benchmark datasets	Accuracy, convergence rate	Faster convergence, higher accuracy
[11]	Cao *et al.* (2025)	Contrast-based FS	Contrast statistics	ML datasets	UCI datasets	Accuracy, stability	Robust to noise
[12]	Rossi *et al.* (2025)	Variational explainable neural networks	Variational inference	Biomedical imaging	Biomedical datasets	AUC, accuracy	Strong interpretability
[13]	Wu *et al.* (2024)	ML-based FS	Supervised learning	Clinical outcome prediction	Clinical cohort data	AUC, sensitivity	High predictive accuracy
[14]	Belenguer-Llorens *et al.* (2025)	Bayesian multimodal FS	Bayesian inference	Biomedical imaging	Multimodal medical data	Accuracy, AUC	Robust multimodal fusion
[15]	Zubair and Kim (2022)	Group FS *via* dimensionality reduction	Statistical filtering	Genomics	Microarray data	Accuracy	Strong group discrimination
[16]	Al-Helali *et al.* (2024)	Genetic programming-based feature selection	Genetic programming and symbolic regression	High-dimensional symbolic regression	Symbolic-regression benchmark datasets	RMSE, MAE	Strong predictive performance, but with high computational runtime
[17]	Du *et al.* (2025)	Deep RL-based FS	Reinforcement learning	Medical diagnostics	Clinical data	AUC, recall	Robust to imbalance
[18]	Chen *et al.* (2025)	Fuzzy autoencoder FS	Deep learning	Biomedical datasets	Accuracy	Noise-resistant	Training complexity

[19]	Chen *et al.* (2020)	Hybrid forest-based FS	Ensemble learning	High-dimensional data	Accuracy	Improved generalization	Higher runtime
[20]	Chen *et al.* (2021)	Cost-sensitive FS	Cost-based learning	Healthcare	Accuracy, cost-efficiency	Budget-aware	Sensitive to cost weights
[21]	Shekhawat *et al.* (2021)	Binary Salp Swarm	Metaheuristic optimization	Benchmark datasets	Accuracy	Fast convergence	Local minima risk
[22]	Zhong *et al.* (2023)	Ensemble FS + CV	Ensemble learning	Biological datasets	AUC, accuracy	Stable performance	Computationally intensive
[23]	Islam *et al.* (2025)	Two-stage causal FS	Causal inference	Healthcare	Bias reduction	Improved causal estimates	Model complexity
[24]	Liu *et al.* (2025)	MOEA-based FS	Multi-objective optimization	ADMET modeling	Accuracy, sparsity	High precision	Expensive computation
[25]	Sun and Barbu (2025)	Stochastic annealing FS	Stochastic optimization	Streaming data	Accuracy	Adaptability	Sensitive to hyperparameters
[26]	Sharma *et al.* (2025)	Hybrid DL + FS	Deep learning	Software reliability	Accuracy	High prediction rate	Data-hungry
[27]	Yu *et al.* (2023)	GP-based discriminant analysis	Bayesian modeling	Functional data	Accuracy	Robust classification	Computation cost
[28]	Rayan *et al.* (2024)	Modified MI-based FS	Mutual information	Medical data	Accuracy	Improved feature relevance	Noise sensitivity
[29]	Demirarslan and Suner (2024)	OCtS scoring method	Statistical scoring	Biomedical datasets	Accuracy	Outlier robustness	Parameter tuning
[30]	Shi *et al.* (2024)	LASSO-based FS	Regularization	Clinical data	AUC	Sparse model	Linear assumptions
[31]	Chen *et al.* (2024)	Tree-guided FS	Structured regularization	EHR data	Accuracy	Interpretability	Dependency on tree quality
[32]	Huang *et al.* (2025)	Q-learning DE FS	Reinforcement learning	Medical datasets	Accuracy	Robust optimization	Parameter sensitivity
[33]	Teke *et al.* (2025)	Ensemble FS	Ensemble learning	Medical diagnostics	Accuracy	Robust classification	Data imbalance
[34]	Ghaffarzadeh-Esfahani *et al.* (2025)	LLM *vs* ML comparison	Transformer-based learning	Clinical records	Accuracy, F1	LLM underperformed ML	Resource intensive
[35]	Ayad *et al.* (2025)	Ontology-based FS	Semantic modeling	Regression datasets	RMSE	Knowledge-driven	Ontology dependence
[36]	Mallidi and Ramisetty (2025)	Bowerbird metaheuristic FS	Metaheuristic optimization	Benchmark datasets	Accuracy	Efficient search	Local optima
[37]	Braik *et al.* (2025)	Chameleon swarm FS	Swarm intelligence	High-dimensional data	Accuracy, F1	Stable results	Hyperparameter sensitivity
[38]	Asha and Johnpeter (2025)	Ensemble FS framework	Ensemble learning	Classification tasks	Accuracy	Improved stability	Overfitting risk
[39]	Garcia-Torres (2025)	Scatter search FS	Metaheuristic search	Multivariate data	Accuracy	Strong exploration	Computational load
[40]	Wang *et al.* (2025)	Shapley-based FS	Cooperative game theory	Symbolic regression	Error rate	High interpretability	Computational cost

Table 3. Inter-reviewer agreement matrix.

-	Reviewer 2: Include	Reviewer 2: Exclude	Total
Reviewer 1: Include	40	3	43
Reviewer 1: Exclude	2	5	7
Total	42	8	50

Table 4. Quality-appraisal criteria.

Domain	Score 0	Score 1	Score 2
Validation	Training-set evaluation or seriously inadequate validation	Basic holdout or incompletely reported cross-validation	Nested, repeated, external, temporal, or appropriately leakage-controlled validation
Dataset quality	Poorly described or unsuitable data	Appropriate but limited or incompletely described data	Well-described, relevant, representative, or independently sourced data
Pipeline integrity	Selection before splitting or probable leakage	Timing unclear or incompletely protected	Feature selection conducted inside training folds with protected test data
Reproducibility	Code, data, parameters, and seeds unavailable	Partial implementation or parameter details	Code/package or sufficiently complete data, preprocessing, parameters, and seeds
Reporting and bias	Selective or overstated reporting	Some concerns or uncertainty omitted	Balanced outcomes, limitations, uncertainty, and comparator reporting

The failure of the classical taxonomy to describe cross-category convergence is a critical weakness of the taxonomy. As an example, the current embedded techniques using regularization (*e.g.* LASSO) implicitly adopt filtering by shrinking coefficients, and the hybrid techniques often impose stability constraints or causal priors and operate without depending on a wrapper-based optimization approach. Through this, there is blurring of strict categorical boundaries in practice.

To overcome these shortcomings, this review suggests an improved analytic taxonomy using statistical behavior instead of algorithm structure. In particular, the methods of feature selection can be reclassified in the three orthogonal dimensions:

(i) Methods that are aware of stability, and take explicit control over selection variability to perturbation of the data (*e.g.* stability selection, ensemble-based filters)

(ii) Causal-conscious approaches that seek to detect features whose relevance is structural and not merely predictive (*e.g.* causal feature selection, treatment-effect-oriented selection)

(iii) Data-regime-adaptive methods, which change their selection mechanism depending on the sample size and number of dimensions compared to noise ($p \gg n$), structure of noise, and how much the features are correlated.

Within this framework, the traditional filter, wrapper, embedded, and hybrid methods are not mutually exclusive classes but implementation-level realizations of more profound statistical goals. This approach allows making a more principled comparison between studies and will not result in the repetition of taxonomy.

3.2. Categories of Feature Selection Methods

Methods of feature selection are generally described as either filter, wrapper, embedded or hybrid, depending on the way that feature relevance is measured and incorporated into the learning process. However, as machine learning evolves and datasets become increasingly complex, these categories need to be refined and extended to account for the growing diversity in real-world applications. We propose the following extended taxonomy:

3.2.1. Filter Methods

Filter methods consider features without any predictive model based on the statistical measures of relevance. The typical measures can be correlation coefficients, mutual information, chi-square and entropy-based measures. While they are computationally efficient, they assume feature independence and do not account for feature interactions. These methods can be extended to include stability-aware filters, where feature importance is evaluated by stability over multiple runs of different data splits or noise perturbations.

A number of reviewed studies employed filter-based approaches, for example, [6] employed correlation-based ranking and Matthews Correlation Coefficient (MCC) to rank informative features in high-dimensional gene expression data, and found that it was stable in small-sample settings. Equally, [28] used a variation of mutual information criteria to improve the COVID-19 prediction accuracy using clinical data. Although filter methods are scalable and simple, they do not in many cases resolve feature dependencies which may restrict predictive effectiveness in complicated datasets.

Table 5. Study-level quality assessment.

Ref.	Study	Validation	Dataset	Pipeline	Reproducibility	Bias Reporting	Total	Judgement
[1]	Reddy and Mishra, (2024)	2	1	1	1	1	6	Some concerns
[2]	Yu *et al.,* (2025)	2	1	1	1	1	6	Some concerns
[3]	Venâncio and Batista, (2025)	2	1	1	1	1	6	Some concerns
[4]	Wang *et al.,* (2025)	2	2	1	1	1	7	Some concerns
[5]	Shaer and Shami, (2024)	2	1	1	1	1	6	Some concerns
[6]	Rojas-Velazquez *et al.,* (2025)	2	2	1	1	1	7	Some concerns
[7]	Bhattacharjee *et al.,* (2022)	2	2	1	2	1	8	Low risk
[8]	Salhi *et al.,* (2025)	1	1	0	0	0	2	High risk
[9]	Baba *et al.,* (2025)	2	1	1	0	1	5	Some concerns
[10]	Wei *et al.,* (2025)	2	1	1	1	1	6	Some concerns
[11]	Cao *et al.,* (2025)	2	1	1	1	1	6	Some concerns
[12]	Rossi *et al.,* (2025)	2	2	1	1	1	7	Some concerns
[13]	Wu *et al.,* (2024)	1	2	1	0	1	5	Some concerns
[14]	Belenguer-Llorens *et al.,* (2025)	2	2	1	2	1	8	Low risk
[15]	Zubair and Kim, (2022)	2	2	1	1	1	7	Some concerns
[16]	Al-Helali *et al.,* (2024)	2	1	1	1	1	6	Some concerns
[17]	Du *et al.,* (2025)	2	2	1	1	1	7	Some concerns
[18]	Chen *et al.,* (2025)	2	2	1	1	1	7	Some concerns
[19]	Chen *et al.,* (2020)	2	1	1	1	1	6	Some concerns
[20]	Chen *et al.,* (2021)	1	2	1	0	1	5	Some concerns
[21]	Shekhawat *et al.,* (2021)	2	1	1	1	1	6	Some concerns
[22]	Zhong *et al.,* (2023)	2	2	2	1	2	9	Low risk
[23]	Islam *et al.,* (2025)	2	2	2	1	2	9	Low risk
[24]	Liu *et al.,* (2025)	2	1	1	1	1	6	Some concerns
[25]	Sun and Barbu, (2025)	2	1	1	1	1	6	Some concerns
[26]	Sharma *et al.,* (2025)	1	1	1	0	1	4	High risk
[27]	Yu *et al.,* (2023)	2	2	1	1	2	8	Low risk
[28]	Rayan *et al.,* (2024)	1	1	1	0	1	4	High risk
[29]	Demirarslan and Suner, (2024)	2	2	1	1	2	8	Low risk
[30]	Shi *et al.,* (2024)	1	2	1	0	1	5	Some concerns
[31]	Chen *et al.,* (2024)	2	2	2	1	2	9	Low risk
[32]	Huang *et al.,* (2025)	2	1	1	1	1	6	Some concerns
[33]	Teke *et al.,* (2025)	1	1	1	0	0	3	High risk
[34]	Ghaffarzadeh-Esfahani *et al.,* (2025)	2	2	2	1	2	9	Low risk
[35]	Ayad *et al.,* (2025)	2	2	1	1	1	7	Some concerns
[36]	Mallidi and Ramisetty, (2025)	1	1	1	0	1	4	High risk
[37]	Braik *et al.,* (2025)	2	1	1	1	1	6	Some concerns
[38]	Asha and Johnpeter, (2025)	1	1	1	0	1	4	High risk
[39]	Garcia-Torres, (2025)	2	1	1	1	1	6	Some concerns
[40]	Wang *et al.,* (2025)	2	1	1	1	1	6	Some concerns

3.2.2. Wrapper Methods

Wrapper methods also consider sets of features by direct maximization of the performance of a predictive model. These methods usually entail the use of iterative methods of searching, including genetic algorithms, particle swarm optimization or recursive feature elimination. The wrapper methods can be more accurate than the filter methods due to the ability of the former to consider the interactions between features and since they also have a higher computational cost.

A number of the studies in the examined corpus use wrapper-based approaches. As an example, [6] used classification accuracy as the feature selection evaluation metric based on a recursive ensemble-based feature selection. Likewise, [21] also introduced a Binary Salp Swarm Algorithm (BSSA), which combines swarm intelligence and wrapper evaluation, and demonstrated a good performance on various benchmark datasets. But wrapper methods are computationally costly, and are likely to be overfit, especially when used on small or noisy data.

3.2.3. Embedded Methods

Embedded methods involve the model training process that is directly embedded with feature selection. Intrinsic techniques like LASSO, Elastic Net, and tree-based methods are techniques that do variable selection by penalizing the complexity of the model or by estimating the importance of features during training.

Many studies embrace embedded strategies because of their efficiency and predictive ability. As an illustration, [7] used regularized regression in their highMLR model to choose informative genomic features to use in survival analysis. In a similar manner, the embedded mechanism that [18] utilized in a multigranularity fuzzy autoencoder was to detect discriminative features in biomedical data. Such techniques have the advantage of being less computationally expensive and having better generalization, but in many cases are model-specific and less applicable to different problem domains.

3.2.4. Hybrid Methods

Hybrid feature selection algorithms have a combination of two or more algorithms- generally filter and wrapper- to use their two-way benefits. These approaches are meant to provide a balance between computational efficiency and predictive accuracy and strength.

The hybrid strategies are a big percentage of recent studies. An example of this is the case of [3], who combined cooperative co-evolution and feature selection to efficiently search high-dimensional search spaces. On the same note, [8] suggested a hybrid AI system that incorporates statistical filtering and machine learning selection in enhancing the accuracy of classification across various datasets. Hybrid techniques have been especially useful in dealing with a large number of dimensions and removing redundancy without losing model interpretability.

3.3. Proposed Analytical Framework

The review proposes a complementary conceptual framework consisting of three dimensions:

- **Stability awareness:** whether a method evaluates or controls variation in selected features across resamples, perturbations, repeated fits, or parameter settings.
- **Causal awareness:** whether relevance is defined through causal structure, treatment effects, adjustment requirements, or intervention-related reasoning rather than predictive association alone.
- **Data-regime adaptivity:** whether the method explicitly addresses a particular statistical condition, including $p \gg n$, predictor correlation, small samples, class imbalance, censoring, multimodal data, streaming observations, structured predictors, or unequal costs.

Table **6** applies the proposed three-dimensional analytical framework to the 40 included studies. It identifies each study's conventional methodological category and indicates whether the method demonstrates stability awareness, causal awareness, or adaptation to a particular data regime, together with the principal basis for its classification.

3.4. Statistical Foundations of Feature Selection

The techniques of feature selection are based on the statistical learning theory, specifically, the notions of the bias-variance trade-off, regularization and information theory. Normalization L1 (LASSO) and L2 (Ridge) methods are regularization which promote sparsity and discourage overfitting, by sending a penalty to huge coefficients. Some of the reviewed papers, such as the ones by [4, 30], used LASSO-based frameworks to improve the generalization performance in medical prediction tasks.

Mutual information and entropy are some of the information-theoretic measures used extensively to estimate the dependency between features and target variables. Most of the filter-based and hybrid methods are based on these measures, including the modified mutual information method suggested by [28]. Also, Bayesian techniques, such as Bayesian networks and probabilistic graphical models, can be used to quantify uncertainty as well as estimate feature relevance, which is shown by [14].

More recently the use of ensemble methods as well as metaheuristic methods, including genetic algorithms, particle swarm optimization and swarm intelligence, have been increasingly prevalent because of their capability to effectively search large search spaces. [4, 37] emphasizes the knowledge on such techniques to enhance strong performance and stability in high-dimensional feature selections.

Table 6. Reclassification of the 40 studies.

Ref.	Study	Conventional Category	Stability-Aware	Causal-Aware	Data-Regime Adaptive	Main Basis
[1]	Reddy and Mishra (2024)	Hybrid/metaheuristic	✓	—	—	Stable fuzzy-memetic search
[2]	Yu *et al.* (2025)	Hybrid/evolutionary	—	—	✓	Multitask knowledge transfer
[3]	Venâncio and Batista (2025)	Wrapper/evolutionary	—	—	✓	Decomposition of large search spaces
[4]	Wang *et al.* (2025)	Wrapper/multi-objective	—	—	✓	Accuracy-sparsity optimisation in gene data
[5]	Shaer and Shami (2024)	RL wrapper	✓	—	✓	Adaptive industrial-data selection
[6]	Rojas-Velazquez *et al.* (2025)	Filter/ensemble	✓	—	✓	Low-sample genomic stability
[7]	Bhattacharjee *et al.* (2022)	Embedded	—	—	✓	Censored survival data
[8]	Salhi *et al.* (2025)	Hybrid	—	—	✓	Multiple datasets and classifiers
[9]	Baba *et al.* (2025)	Robust filter	✓	—	✓	Outlier-resistant screening
[10]	Wei *et al.* (2025)	RL-enhanced wrapper	✓	—	✓	Adaptive search scoring
[11]	Cao *et al.* (2025)	Filter	✓	—	—	Noise-robust contrast statistics
[12]	Rossi *et al.* (2025)	Embedded deep model	—	—	✓	Nonlinear biomedical data
[13]	Wu *et al.* (2024)	Model-based	—	—	✓	High-dimensional clinical cohort
[14]	Belenguer-Llorens *et al.* (2025)	Bayesian embedded	✓	—	✓	Multimodal small-sample data
[15]	Zubair and Kim (2022)	Group filter	✓	—	✓	Correlated feature groups
[16]	Al-Helali *et al.* (2024)	GP wrapper	—	—	✓	Symbolic-regression feature space
[17]	Du *et al.* (2025)	Deep RL wrapper	✓	—	✓	Imbalanced follow-up data
[18]	Chen *et al.* (2025)	Embedded autoencoder	✓	—	✓	Noise-resistant biomedical selection
[19]	Chen *et al.* (2020)	Hybrid forest	✓	—	—	Ensemble generalisation
[20]	Chen *et al.* (2021)	Embedded/cost-sensitive	—	—	✓	Unequal feature and prediction costs
[21]	Shekhawat *et al.* (2021)	Metaheuristic wrapper	—	—	✓	Binary high-dimensional search
[22]	Zhong *et al.* (2023)	Ensemble	✓	—	✓	Nested cross-validation
[23]	Islam *et al.* (2025)	Hybrid causal	✓	✓	✓	Treatment-effect estimation
[24]	Liu *et al.* (2025)	Multi-objective wrapper	—	—	✓	Sparse ADMET data
[25]	Sun and Barbu (2025)	Stochastic selection	✓	—	✓	Streaming-data adaptation
[26]	Sharma *et al.* (2025)	Hybrid deep model	—	—	✓	Software-failure prediction
[27]	Yu *et al.* (2023)	Bayesian embedded	✓	—	✓	Functional-data nonstationarity

[28]	Rayan *et al.* (2024)	Information filter	—	—	✓	Clinical mutual-information selection
[29]	Demirarslan and Suner (2024)	Statistical filter	✓	—	✓	Outlier and correlation sensitivity
[30]	Shi *et al.* (2024)	Embedded LASSO	—	—	✓	Sparse clinical predictors
[31]	Chen *et al.* (2024)	Structured embedded	—	—	✓	Tree-guided EHR structure
[32]	Huang *et al.* (2025)	RL/evolutionary wrapper	✓	—	✓	High-dimensional medical optimisation
[33]	Teke *et al.* (2025)	Ensemble	✓	—	✓	Diagnostic and imbalance context
[34]	Ghaffarzadeh-Esfahani *et al.* (2025)	Comparative model study	—	—	✓	High-dimensional multicentre clinical data
[35]	Ayad *et al.* (2025)	Knowledge-guided hybrid	—	—	✓	Ontology-dependent regression
[36]	Mallidi and Ramisetty (2025)	Metaheuristic wrapper	—	—	✓	Combinatorial high-dimensional search
[37]	Braik *et al.* (2025)	Swarm wrapper	✓	—	✓	Heterogeneous search adaptation
[38]	Asha and Johnpeter (2025)	Ensemble	✓	—	—	Ensemble stability
[39]	Garcia-Torres (2025)	Scatter-search wrapper	—	—	✓	Search-space reduction
[40]	Wang *et al.* (2025)	Embedded/game-theoretic	—	—	✓	Interaction-sensitive symbolic regression

3.5. Bias–Variance Trade-off and Statistical Limitations of Feature Selection

The bias-variance tradeoff is a major concern when it comes to defining the usefulness of feature selection methods in high-dimensional environments but is commonly addressed only in broad strokes. In practice, feature selection algorithms are biased because they reduce dimensionality and aim to decrease the variance of estimators caused by redundant or noisy predictors. These two effects are extremely sensitive to both regime of data and model assumptions. Regularization-based embedded regression techniques, like LASSO, minimize variance by imposing sparsity with an $\ell 1$ penalty, but this advantage depends on linearity and fails when there is strong collinearity among the predictors. In $p \gg n$ regimes, LASSO can randomly choose one feature of a set of correlated features, resulting in unstable and non-identifiable solutions.

Moreover, ℓ_1 regularization is not statistically consistent in the presence of irrepresentable condition, which restricts its asymptotic accuracy in correlated high-dimensional data. The use of information-theoretic filter methods, especially those based on mutual information, inject bias due to finite-sample estimation. Mutual information estimators are also sensitive to discretization strategy, bandwidth of the kernel and the sample size, as they tend to have an upward bias in small sample regimes. The curse of dimensionality also progressively worsens the estimator consistency as dimensionality grows and in turn makes relevance scores unreliable without bias correction or ensemble estimation techniques. These failure modes underline why strategies that perform well empirically can deteriorate drastically when they are not in their optimal operating regions. Therefore, comparisons of performance that fail to correct for estimator bias, variance inflation and identifiability constraints, are likely to overestimate generalizability.

3.6. Emerging Trends and Research Directions

Recent studies point to an apparent movement towards the hybrid and learning-based feature selection models where the statistical rigor is combined with adaptive learning functions. Autoencoders and attention mechanisms are also deep learning-based methods that are being used to automatically discover the levels of feature importance in data. Research, like deep architectures, includes [12, 18], is used to illustrate the role of deep architecture in modeling the complex non-linear relationship of high-dimensional spaces.

The incorporation of stability and robustness analysis is another new development, especially in biomedical and clinical applications where reproducibility is paramount. Stability selection, ensemble learning, and repeated cross-validation are some of the techniques that are being applied to reduce variance and enhance reliability.

Also, explainable AI (XAI) and interpretability-conscious feature selection are becoming increasingly popular,

fulfilling the increasing need for transparency in high-stakes areas of healthcare and finance. The balance between predictivity performance and explainability is highlighted in recent publications, in the form of hybrid frameworks, which incorporate explainable models in deep learning structures.

4. PERFORMANCE EVALUATION AND COMPARATIVE ANALYSIS

The study provides a critical analysis of feature selection methods of high-dimension predictive modeling, based on the empirical evidence of the 40 studies assessed. The performance metrics, computational effectiveness, interpretability, and domain-specific effectiveness are aspects of the analysis and provide trends, strong points, and weaknesses of the various feature selection paradigms through synthesizing findings over heterogeneous datasets and different approaches.

4.1. Predictive Outcomes

The included studies reported multiple predictive-performance measures, including classification accuracy, AUC, F1-score, MCC, recall, RMSE, MAE, C-index, error rate, sparsity, and convergence rate. These measures represent different statistical properties and cannot be treated as interchangeable. Accordingly, AUC was not pooled with accuracy, F1-score, MCC, RMSE, MAE, C-index, or other outcomes.

Because the included studies used heterogeneous datasets, predictive tasks, outcome distributions, performance metrics, comparators, and validation procedures, no performance metric was normalised or converted to a common scale. Accuracy values were retained in their originally reported form, whether expressed as percentages or proportions. AUC, F1-score, MCC, RMSE, MAE, C-index, recall, error rate, sparsity, convergence rate, and other outcomes were interpreted separately because they represent different statistical properties. Findings were assessed only relative to the comparators and experimental conditions reported within each original study. Accordingly, no pooled performance estimates, cross-study means, between-study standard deviations, category-level error bars, or inferential rankings were calculated.

Reported findings were interpreted relative to the comparators, datasets, classifiers, and validation procedures used within each original study. Salhi *et al.* [8] reported substantial variation across classifier-method combinations, including very high performance in some configurations and considerably lower accuracy in others. Such variation demonstrates why reducing a study to one category-level average could misrepresent its findings. Therefore, quantitative results were retained as study-specific evidence rather than pooled estimates of method performance.

4.2. Performance Metrics

The effectiveness of feature-selection methods was evaluated across the included studies using predictive performance, computational requirements, feature-set size, selection stability, interpretability, and reproducibility. However, these dimensions were not reported consistently. Consequently, the review provides a qualitative and study-specific synthesis rather than a pooled numerical comparison.

4.2.1. Predictive Accuracy

Predictive outcomes were retained in the form reported by each original study. A favourable result indicates that a method performed better than at least one comparator under the conditions examined in that study. It does not demonstrate superiority across other datasets, domains, classifiers, or validation procedures.

Hybrid, ensemble, embedded, evolutionary, and reinforcement-learning approaches frequently report favourable within-study findings. Nevertheless, the magnitude and direction of their performance varied according to dataset characteristics, sample size, dimensionality, class distribution, predictive learner, comparator strength, and tuning procedure. These findings suggest that adaptive and combined approaches may be beneficial in particular settings, but they do not establish consistent or universal superiority over simpler feature-selection methods.

Some individual studies reported high accuracy or AUC values. These numerical findings should be interpreted cautiously because they originated from specific experimental settings and were not directly comparable across the review corpus. High performance in a particular benchmark or clinical dataset may reflect class balance, test-set size, feature construction, model complexity, validation design, or overfitting. Several hybrid, ensemble, and reinforcement-learning studies reported favourable results within their respective experimental settings [8, 17, 33]. However, these findings were not combined numerically because the studies differed substantially in dataset characteristics, dimensionality, outcome prevalence, predictive learners, comparator methods, hyperparameter-tuning procedures, and validation designs. The reported performance values were therefore retained as study-specific evidence and were not normalised, averaged, or used to construct cross-study error bars. These results demonstrate the potential usefulness of adaptive feature-selection methods under experimental conditions but do not establish their general superiority across datasets or application domains.

It is reported by [8] for classification accuracy higher than 95% with a hybrid AI-based feature selection framework, which is 6-10% higher than baseline models. Similarly, [17] have shown that the feature selection process using deep reinforcement learning had a 6-9% improvement with absolute AUC values ranging from 0.84 to 0.92 relative to a baseline of 0.85. This overcomes the traditional feature selection methods used in medical diagnosis problems. Other studies that used ensemble learning reported high accuracies of over 98% using feature selection and ensemble classifiers as used by [33].

Simpler filter-based algorithms, *e.g.*, mutual information or correlation-based ranking, in contrast, are computationally efficient but can commonly perform poorly in predicting when the features interact in a complex manner, and are useful for large datasets. [28] found that mutual information-based selection was able to perform competitively but worse than wrapper and hybrid selections with imbalanced clinical data.

4.2.2. Computational Efficiency and Runtime Considerations

Computational requirements are important when high-dimensional datasets contain thousands or millions of candidate variables. Filter methods generally require fewer model fits and are therefore commonly suitable for large-scale screening or resource-constrained applications. Nevertheless, runtime comparisons were inconsistently reported across the included studies, and no pooled runtime analysis was undertaken [6, 29].

Wrapper, evolutionary, swarm-based, reinforcement-learning, and hybrid methods typically require repeated model training or iterative subset evaluation. These procedures may provide flexible search capabilities but usually involve greater computational cost and a larger hyperparameter space. Deep-learning-based selectors may require additional training time, memory, and specialised hardware. The evidence therefore supports a qualitative trade-off: simpler methods commonly provide greater scalability, whereas more adaptive search procedures may model complex interactions at the cost of computational burden and tuning complexity.

4.2.3. Interpretability and Model Transparency

Interpretability is particularly important in high-stakes applications such as healthcare, finance, and public decision-making. This review distinguishes intrinsic interpretability from post-hoc explainability. Intrinsic interpretability concerns whether the model or selection process can be understood directly from its structure, as in sparse linear models, rule-based methods, or transparent statistical filters. Post-hoc explainability refers to auxiliary techniques, such as Shapley values, saliency measures, or attention visualisations, that approximate feature influence after model training [28, 30].

Filter methods and sparse embedded models can often provide relatively transparent feature rankings or coefficient-based interpretations. By contrast, deep-learning, reinforcement-learning, evolutionary, and complex hybrid selection mechanisms may remain opaque even when post-hoc explanation methods are applied. Reducing the number of input variables does not automatically make the internal selection process interpretable. Claims for complex hybrid or deep methods improve interpretability and should therefore be framed as improvements in explainability or dimensional simplicity unless the decision structure itself is transparent [40].

4.3. Comparative Interpretation of Feature-Selection Approaches

The studies included were not treated as statistically exchangeable. They differed in application domain, dataset size, dimensionality, predictor dependence, outcome prevalence, predictive learner, comparator strength, hyperparameter tuning, validation procedure, and reported metric. Consequently, no meta-analysis, cross-study average, pooled standard deviation, ANOVA, category-level significance test, or confirmatory ranking of feature-selection approaches were conducted.

The comparative synthesis instead focused on recurring methodological trade-offs. Filter methods generally offered scalability, transparency, and model independence but could overlook nonlinear interactions or conditional relevance. Wrapper methods directly evaluated feature subsets using a predictive learner and could capture model-specific interactions, although they involved greater computational cost and a higher risk of overfitting. Embedded approaches integrated feature selection with model estimation and often provided a practical balance between efficiency and predictive modelling, but their performance remained dependent on model assumptions and predictor structure.

Hybrid, ensemble, evolutionary, reinforcement-learning, and other adaptive approaches combined complementary search or evaluation mechanisms. Several studies reported favourable results for these methods within their respective experimental settings. However, additional flexibility also introduced greater tuning requirements, computational burden, opacity, instability, and potential leakage risk. These methods should therefore be interpreted as potentially advantageous under suitable conditions rather than universally superior.

Terms such as “significant”, “superior”, and “outperformed” are used in this review only when referring to a formal comparison conducted within an individual included study. They are not used to imply statistical superiority across the review corpus. The strengths, limitations, and appropriate interpretation of the main methodological approaches are summarised in Table **7**.

4.3.1. Biomedical and Healthcare Applications

Hybrid, embedded, ensemble, and learning-based methods frequently reported favourable results in biomedical and healthcare studies. However, the evidence varied across clinical cohorts, imaging datasets, diagnostic tasks, sample sizes, outcome prevalence, and validation procedures. These findings support selecting methods according to dimensionality, sample size, interpretability requirements, stability, and external-validation availability rather than claiming that one methodological category performs best.

4.3.2. Bioinformatics and Genomics

Evolutionary, embedded, filter-based, ensemble, and multi-objective methods were commonly evaluated in genomic settings characterised by large numbers of

predictors and relatively small samples. Several studies reported useful trade-offs between predictive performance and feature-set sparsity. Nevertheless, the scientific value of selected genes or biomarkers depends on stability across resamples, biological plausibility, independent validation, and adequate protection against overfitting.

4.3.3. Industrial and Engineering Applications

Wrapper, hybrid, reinforcement-learning, evolutionary, and ensemble methods were evaluated in industrial and engineering datasets containing noise, heterogeneous variables, or complex interactions. Some studies reported favourable within-study findings, although increased search flexibility was generally accompanied by greater computational requirements and tuning sensitivity. Leakage-controlled, repeated, and preferably external validation is therefore important when applying these methods.

4.3.4. General Machine-Learning Benchmarks

Benchmark studies frequently reported favourable results for hybrid, swarm, ensemble, and evolutionary methods. However, benchmark outcomes are sensitive to dataset selection, classifier choice, hyperparameter tuning, comparator strength, and validation design. These findings demonstrate potential effectiveness under specific experimental conditions but do not support a universal ranking of feature-selection methods. Because the domain-specific studies were heterogeneous, no numerical domain scores, normalised comparisons, error bars, or ANOVA tests were calculated. Domain-level evidence was synthesised narratively.

4.4. Discussion of Findings

The evidence does not identify one feature-selection category as universally superior. Instead, effectiveness depends on the alignment among the data regime, predictive model, selection objective, available computational resources, and validation design. Filter methods provide scalability and transparency but may fail to capture nonlinear or conditional relationships. Wrapper, evolutionary, swarm-based, and reinforcement-learning methods explore more complex subsets but introduce greater computational cost, tuning sensitivity, and overfitting risk. Embedded methods frequently offer a practical compromise but remain dependent on the assumptions and inductive biases of the selected predictive learner.

The most defensible cross-study conclusion concerns trade-offs rather than rankings. Increased search flexibility may improve predictive fit within a particular study while simultaneously increasing instability, runtime, opacity, and susceptibility to leakage. These concerns are especially important in small-sample, high-dimensional genomic and clinical datasets, where different samples can produce different selected subsets despite similar predictive scores.

Table 7. Strengths, limitations, and appropriate interpretation.

Category	Strengths	Limitations	Appropriate interpretation
Filter	Fast, scalable, relatively transparent, and model-agnostic	May overlook nonlinear interactions and conditional relevance	Suitable as an initial screening approach, particularly in very high-dimensional datasets
Wrapper	Captures model-specific relationships and feature interactions	Computationally expensive and potentially prone to overfitting	Useful when interaction modelling justifies the additional computational and validation burden
Embedded	Integrates selection with model estimation and can balance efficiency with predictive performance	Model-specific and sensitive to assumptions, correlation, and regularisation choices	Often provides a practical compromise when its assumptions match the data structure
Hybrid	Combines complementary selection and search mechanisms	Greater tuning burden, computational cost, opacity, instability, and leakage risk	Frequently favourable within individual studies but not universally superior
Stability-aware	Evaluates repeatability across resamples, perturbations, or repeated fits	Stability definitions and thresholds are not standardised	Particularly important for biomarker discovery and scientific interpretation
Causal-aware	Can identify variables relevant to intervention or adjustment objectives	Requires defensible causal assumptions and appropriate causal data	Necessary when the goal extends beyond prediction to intervention or causal interpretation
Data-regime adaptive	Addresses structures such as (p \gg n), imbalance, censoring, streaming, multimodality, or unequal costs	May generalise poorly outside the statistical regime for which it was designed	Should be selected according to the operating data conditions and decision objective

The proposed reclassification indicates that data-regime adaptation is widely represented and stability awareness is moderately represented, whereas explicit causal awareness remains rare. Most included studies identified predictors associated with an outcome rather than variables shown to have structural, mechanistic, or intervention-relevant effects. Predictive feature importance should therefore not be interpreted as evidence of causality.

The quality appraisal further indicates that only a minority of studies combined strong validation, protected feature-selection pipelines, reproducible implementation, and balanced reporting. Highly favourable predictive results warrant particular caution when test samples are small, feature selection is conducted before data splitting, nested validation is absent, or numerous selector-classifier combinations are evaluated without adequate control for optimisation bias.

Overall, feature-selection performance is conditioned by three interacting considerations: the data regime, model capacity, and selection objective. The data regime influences estimator reliability and selection stability; model capacity affects the bias-variance trade-off, interaction modelling, and interpretability; and the selection objective determines whether evaluation should prioritise prediction, stability, sparsity, computational cost, or causal relevance. Feature-selection methods are most defensible when these three considerations are aligned.

5. PRACTICAL APPLICATIONS FOR FEATURE SELECTION

Feature selection has become a key focus of the current generation of data-driven applications which have the challenge of high dimensionality, redundancy, and noise as major factors that affect the model performance and interpretability. In fields like bioinformatics, finance, healthcare, and medical imaging, efficient feature selection can be used to facilitate more accurate prediction and can be used with lower computational cost and better model clarity. In this section, an overview of the practical applications of feature selection methods in major areas of application has been synthesized based on the empirical evidence provided by the studies reviewed.

5.1. Applications in Bioinformatics

Bioinformatics is one of the most highly visible areas of application of feature selection because of the high dimensionality of biological data, especially in genomics, transcriptomics, and proteomics. The datasets used in this field have thousands of features (genes or biomarkers) and relatively few samples, which are highly overfitted. The role of feature selection is therefore important in discovering biologically important variables, as well as minimizing noise and redundancy.

Studies provide evidence of the usefulness of sophisticated feature selection techniques in biological and biomedical studies. Here, [7] used feature selection framework based on machine learning (highMLR) to select prognostic genes in cancer survival prediction with better predictive capabilities and interpretability. Equally, [6] used correlation-based and ensemble feature ranking methods to the high-dimensional gene expression datasets and was able to obtain strong classification results even in small sample settings.

Genomic data analysis has also been performed through transformation and hybrid solutions. It was shown by [4] and [24] that multi-objective evolutionary feature selection techniques are an effective balance between accuracy and sparsity in situations where the number of genomic features is thousands. They have found applications, especially in omics studies where it is important to determine small sets of biologically meaningful genes that can be used to perform downstream analysis and clinically make decisions.

In addition, feature selection algorithms that are based on deep learning, including the ones suggested by [18, 12], are based on autoencoders and representation learning to discover intricate nonlinear interactions in high dimensional biological data. However, as much as such approaches have high predictive power, they are not easy to interpret, which supports the significance of hybrid and explainable methods in biomedical studies.

5.2. Applications in Finance

In finance, feature selection is valuable for handling high-dimensional, noisy, and correlated data that are generated from transactional records, market data, and customer behavioral information. In the use of such applications as credit scoring, fraud detection, and risk assessment, the choice of informative features is important to boost predictive accuracy without affecting the model interpretability and regulatory compliance.

Some of the studies reviewed can show that feature selection is effective in a financial setting. As an example, [30] used LASSO-based feature selection to credit risk modelling, which is capable of identifying the main financial indicators with a high level of reliability, but minimized the complexity of the model. On the same note, hybrid feature selection procedures which incorporate filter and wrapper methods have been depicted to enhance prediction power in financial classification tasks with the concurrent computational efficiency.

Research findings like those by [8, 35] also show that hybrid and ensemble-based feature selection frameworks can be useful in dealing with heterogeneous financial data, such as transactional, demographic and behavioral features. These methods not only improve prediction but also help in regulatory transparency because they facilitate the ability to understand the choice of risk factors.

Besides, feature selection has been demonstrated to minimize overfitting and enhance generalization in continually changing market conditions in algorithmic trading and financial forecasting. The capability to predictive features is especially important in financial applications where the model decisions may carry strong economic implications.

5.3. Applications in Healthcare and Medical Imaging

Healthcare is one of the most influential application areas of feature selection, especially in medical diagnosis, prognosis and decision support systems. Clinical data in high dimension tend to harbor redundant and noisy information, including electronic health record, medical imaging, and genomics samples which may obstruct significant patterns.

Some of the studies that have been reviewed prove the usefulness of feature selection in enhancing diagnostic accuracy. As an illustration, [13] used machine learning-based feature selection to forecast the outcome of chronic pain, and they reported significant enhancement in model performance. Equally, [34] found that deep learning models compared to conventional machine learning methods in predicting COVID-19 mortality were more predictive and the judiciously chosen features enhanced predictive reliability.

In medical imaging case, feature selection is important in dimension reduction and interpretability. Such methods as convolutional feature extraction with selection mechanisms including attention or saliency-based selection have been demonstrated to enhance diagnostic accuracy and decrease computing costs. The effectiveness of the hybrid representation learning and selective feature pruning methods can be proven with references to studies that use deep feature selection techniques, similar to those provided by [18].

5.4. Emerging Applications and Future Directions

In addition to more conventional fields, feature selection is being used in more emerging applications in the Internet of Things (IoT), remote sensing and smart city analytics. These applications include high dimensional and heterogeneous data, which are often streaming data, and require scalable and adaptive feature selection methods.

Real-time feature selection can be used in IoT settings to process sensor data efficiently and reduce the use of energy and communication overheads. Experiments done using evolutionary and reinforcement learning based feature selection show promising outcomes in dynamic settings where the data distributions change with time. Equally important, feature selection is useful in remote sensing and analysis of satellite images to minimize spectral redundancy and improve the accuracy of land-cover classification.

6. CHALLENGES AND FUTURE DIRECTIONS

The procedure of feature selection is an extensive but developing part of high-dimensional data analysis. Although notable achievements have been achieved in terms of superior statistical, machine, and hybrid methods, various methodological and practical issues persist. This part overviews the major weaknesses that have been noted throughout the studies reviewed and the research directions that are emerging in the future of feature selection.

6.1. Standardised Benchmarking

Future research should use benchmark suites that vary systematically in sample size, dimensionality ratio, predictor correlation, noise, imbalance, and interaction structure. Reporting only the best-performing dataset or classifier combination should be avoided.

6.2. Leakage-Controlled Validation

Feature selection, preprocessing, imputation, resampling, and hyperparameter tuning must be conducted exclusively within training folds. Applying selection to the complete dataset before cross-validation introduces information leakage and overestimates generalisation.

6.3. Stability Reporting

Studies should report the frequency with which each feature is selected across repeated resamples, together with an overall subset-stability measure such as the Jaccard index, Kuncheva index, or related agreement statistic.

6.4. Causal Feature Selection

Only one included study explicitly focused on causal or treatment-effect objectives. Future work should investigate how causal discovery, domain knowledge, structural assumptions, and predictive selection can be combined without conflating association with relevance.

6.5. External and Temporal Validation

Many studies rely on public benchmarks or single datasets. Independent external, temporal, multicentre, and cross-domain validation should receive greater emphasis.

6.6. Reproducibility

Authors should release code, data-access instructions, random seeds, search-space definitions, stopping criteria, preprocessing steps, feature-selection timing, and complete hyperparameters.

6.7. Interpretability

Studies should distinguish intrinsic transparency from post-hoc explanation. A model does not become intrinsically interpretable merely because it operates on fewer variables or because SHAP values are calculated after trainings.

6.8. Research Gaps

Despite extensive research, there remain a few gaps in the existing literature on feature selection. The absence of standard benchmarking frameworks is one of the primary weaknesses. All research studies use various datasets, measures and experimental designs which makes it hard to directly compare the methods. Despite the fact that a few works have tried to cross-dataset validate such as those reported by [2, 8], there is no single benchmark similar to ImageNet that features selection researchers can use.

Another such area of weakness is the thin domain feature selection frameworks. Although generic approaches work fairly, domain-neutral adaptations can be much better. Medical and genomic data are one such example, medical needs approaches that are sensitive to biological importance and clinical meaning, whereas financial needs are more sensitive to withstand noise and time decay. Research conducted by [30, 35] reveal the necessity of tailored feature

selection strategies that would be based on the domain knowledge in the selection process.

Also, the combination of causal inference in the process of selecting features is under-researched. The relevance of most of the current methods is based on correlation and not causality, and this reduces the potential use of the methods in making critical decisions. Research is also emerging indicating that causal discovery plus feature selection would enhance model robustness and interpretability especially in the healthcare and policy-focused fields.

6.9. Emerging Research Trends

Future research directions reflect recent advances and include several promising directions. A key trend is the combination of feature selection with the use of deep learning where neural networks are directly trained to be feature relevant either by direct attention or by regularization techniques. Research works like [12, 18] prove that deep learning models are able to find complex nonlinear correlations in high-dimensional data.

The other developing field is the transfer learning and meta-learning to select features where information earned in one task or domain are used to enhance feature selection in another. The method is especially helpful in the data-sparse fields, *e.g.* the diagnosis of rare diseases or industry-specific tasks. Early indications indicate that transfer-based feature selection can be very effective in terms of minimizing the time of training and still achieving accuracy.

Lastly, a new and important line of research is beginning to emerge, that of multi-modal and multi-task feature selection. The present-day usage is frequently characterized by the heterogeneity of data sources, including a combination of clinical records, data of imaging, and genomic profiles. The methods which combine the selection of features between two or more different modalities or tasks, like the ones suggested by [4, 18], can provide promising directions in the creation of more holistic and robust prediction models.

This systematic review makes five contributions:

1. It establishes an explicit boundary between feature selection, which retains original variables and feature transformation, which constructs latent representations.
2. It integrates conventional algorithmic categories with an author-proposed analytical framework based on stability awareness, causal awareness, and adaptation to the data regime.
3. It reclassifies the 40 included studies using this framework to determine which methodological objectives are well represented and which remain under-investigated.
4. It applies structured quality appraisal, direction-of-effect vote counting, and thematic synthesis rather than interpreting heterogeneous accuracy values as directly comparable effect estimates.
5. It develops practitioner-oriented guidance linking common data conditions and decision requirements to appropriate starting methods, validation procedures, and implementation cautions. These contributions shift the review from a catalogue of algorithms toward a critical framework for method selection, evidence interpretation, and future research design.

CONCLUSION

This systematic review synthesised 40 empirical studies of feature selection for high-dimensional predictive modelling published between 2020 and 2025. The evidence indicates that the effectiveness of a feature-selection method depends on the interaction among sample size, dimensionality, predictor dependence, noise, outcome structure, model capacity, computational resources, validation design, and the intended selection objective.

Filter methods remain useful when scalability, simplicity, and transparency are priorities. Wrapper, embedded, evolutionary, reinforcement-learning, ensemble, and hybrid approaches can model more complex relationships but require stronger protection against overfitting, information leakage, instability, and excessive tuning. Because the included studies used non-exchangeable datasets, predictive learners, metrics, and validation procedures, no pooled averages, normalised performance scores, ANOVA tests, or category-level statistical rankings were produced. Reported numerical findings were interpreted only within their original study contexts

The proposed framework of stability awareness, causal awareness, and data-regime adaptivity provides a complementary interpretation of the field. The reclassification indicates that data-regime adaptation is widely represented and stability-related considerations are moderately represented, whereas explicit causal feature selection remains rare. Future research should prioritise leakage-controlled and nested validation, independent external testing, standardised stability reporting, reproducible implementation, transparent computational reporting, and clearer separation between predictive association and causal relevance.

Practitioners should therefore select feature-selection methods according to their data conditions, validation resources, interpretability needs, computational constraints, and decision objectives rather than relying on isolated accuracy values or broad claims of methodological superiority.

LIST OF ABBREVIATIONS

BSSA	=	Binary Salp Swarm Algorithm
IoT	=	Internet of Things
MCC	=	Matthews Correlation Coefficient
PCA	=	Principal Component Analysis
XAI	=	Explainable AI

AUTHOR'S CONTRIBUTION

The author is solely responsible for conceptualization if the study design, analysis, interpretation and manuscript writing.

REPORTING GUIDELINES

PRISMA guideline has been followed for this study.

AVAILABILITY OF DATA AND MATERIALS

The data underlying the findings of this study are available from the author upon reasonable request.

FUNDING

No research grant was provided for conducting this research.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest regarding the publication of this article.

ACKNOWLEDGEMENTS

Declared none.

DECLARATION OF AI

The author used the AI tool ChatGPT for final editing of the manuscript and takes responsibility of the published content.

REFERENCES

[1] K. G. Reddy, and D. Mishra, "Enhancing feature selection in high-dimensional data with fuzzy fitness-integrated memetic algorithms" *IEEE Access*., vol. 12, pp. 130675-130692, 2024, https://doi.org/10.1109/ACCESS.2024.3459390

[2] W. Yu, H. Kang, J. Xu, J. Li, H. Li, and G. Sun, "Enhancing evolutionary multitasking for high-dimensional feature selection through task relevance evaluation and knowledge transfer," *Knowl.-Based Syst.*, vol. 326, pp. 114076, 2025, https://doi.org/10.1016/j.knosys.2025.114076

[3] P. V. A. Venâncio and L. S. Batista, "A self-tuning decomposition strategy in cooperative co-evolutionary algorithms for high-dimensional feature selection," *Knowl.-Based Syst.*, vol. 316, pp. 113327, 2025, https://doi.org/10.1016/j.knosys.2025.113327

[4] Y. Wang, Z. Du, X. Li, W. Xiao, H. Liu, and L. Yang, "Evolving dual-directional multiobjective feature selection for high-dimensional gene expression data," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 11, pp. 8572–8581, 2025, https://doi.org/10.1109/JBHI.2025.3572310

[5] I. Shaer, and A. Shami, "WrapperRL: Reinforcement learning agent for feature selection in high- dimensional industrial data. *IEEE Access*., vol. 12, pp. 128338-128348, 2024, https://doi.org/10.1109/ACCESS.2024.3456688

[6] D. Rojas-Velazquez, A. D. Kraneveld, A. Tonda, and A. Lopez-Rincon, "Matthews correlation coefficient-based feature ranking in recursive ensemble feature selection for high-dimensional and low-sample size data," *Mach. Learn. Appl.*, vol. 22, pp. 100757, 2025, https://doi.org/10.1016/j.mlwa.2025.100757

[7] A. Bhattacharjee, G. K. Vishwakarma, S. Banerjee, and A. F. Pashchenko, "highMLR: An open-source package for R with machine learning for feature selection in high dimensional cancer clinical genome time to event data," *Expert Syst. Appl.*, vol. 210, pp. 118432, 2022, https://doi.org/10.1016/j.eswa.2022.118432

[8] A. Salhi, R. Alshamrani, A. Althbiti, A. Ismail, M. Abd-ElRahman, and B. M. Hassan, "Optimizing high dimensional data classification with a hybrid AI driven feature selection framework and machine learning schema," *Sci. Rep.*, vol. 15, pp. 35038, 2025, https://doi.org/10.1038/s41598-025-08699-4

[9] I. A. Baba, M. B. Mohammed, K. B. Jillahi, A. Umar, and H. T. Hendi, "Robust correlation feature selection based support vector machine approach for high dimensional datasets," *Results Control Optim.*, 2025, vol. 21, pp. 100609, https://doi.org/10.1016/j.rico.2025.100609

[10] B. Wei, J. Huang, L. Deng, S. Yang, J. Zheng, and Y. Huang, "Reinforcement learning-based particle swarm optimization with adaptive scoring mechanism for high-dimensional feature selection," *Swarm Evol. Comput.*, vol. 98, pp. 102104, 2025, https://doi.org/10.1016/j.swevo.2025.102104

[11] C. Cao, Q. Zhang, and Y. Deng, "A contrast based feature selection algorithm for high-dimensional datasets in machine learning," *Inf. Sci.*, vol. 717, pp. 122308, 2025, https://doi.org/10.1016/j.ins.2025.122308

[12] R. Rossi, A. Murari, and M. Gelfusa, "A deep learning framework for feature selection and dimensional analysis: Variational explainable neural networks," *Knowl.-Based Syst.*, vol. 324, pp. 113940, 2025, https://doi.org/10.1016/j.knosys.2025.113940

[13] H. Wu, Z. Chen, J. Gu, Y. Jiang, S. Gao, W. Chen, *et al.*, "Predicting chronic pain and treatment outcomes using machine learning models based on high-dimensional clinical data from a large retrospective cohort," *Clin. Ther.*, vol. 46, no. 6, pp. 490-498, 2024, https://doi.org/10.1016/j.clinthera.2024.04.012

[14] A. Belenguer-Llorens, C. Sevilla-Salcedo, J. Tohka, and V. Gómez-Verdejo, "Unified Bayesian representation for high-dimensional multi-modal biomedical data for small-sample classification," *Eng. Appl. Artif. Intell.*, vol. 160, pp. 111887, 2025, https://doi.org/10.1016/j.engappai.2025.111887

[15] I. M. Zubair and B. Kim, "A group feature ranking and selection method based on dimension reduction technique in high-dimensional data," *IEEE Access*, vol. 10, pp. 125136–125147, 2022, https://doi.org/10.1109/ACCESS.2022.3225685

[16] B. Al-Helali, Q. Chen, B. Xue, and M. Zhang, "Genetic programming for feature selection based on feature removal impact in high-dimensional symbolic regression," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 8, no. 3, pp. 2269–2282, 2024, https://doi.org/10.1109/TETCI.2024.3369407

[17] Y. Du, X. Zhou, Q. Gao, C. Yang, and T. Huang, "A deep reinforcement learning-based feature selection method for invasive disease event prediction using imbalanced follow-up data," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 2, pp. 1472–1483, 2025, https://doi.org/10.1109/JBHI.2024.3497325

[18] Y. Chen, W. Ding, J. Huang, W. Zhang, and T. Zhou, "Multigranularity fuzzy autoencoder for discriminative feature selection in high-dimensional data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 9, pp. 17433–17447, 2025, https://doi.org/10.1109/TNNLS.2025.3569893

[19] W. Chen, Y. Xu, Z. Yu, W. Cao, C. P. Chen, and G. Han, "Hybrid dimensionality reduction forest with pruning for high-dimensional data classification," *IEEE Access*, vol. 8, pp. 40138-40150, 2020, https://doi.org/10.1109/ACCESS.2020.2975905

[20] Y. Chen, Y. Wang, L. Cao, and Q. Jin, "CCFS: A confidence-based cost-effective feature selection scheme for healthcare data classification," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 18, no. 3, pp. 902–911, 2021, https://doi.org/10.1109/TCBB.2019.2903804

[21] S. S. Shekhawat, H. Sharma, S. Kumar, A. Nayyar, and B. Qureshi, "bSSA: Binary salp swarm algorithm with hybrid data transformation for feature selection," *IEEE Access*, vol. 9, pp. 14867–14882, 2021, https://doi.org/10.1109/ACCESS.2021.3049547

[22] Y. Zhong, P. Chalise, and J. He, "Nested cross-validation with ensemble feature selection and classification model for high-dimensional biological data," *Commun. Stat. Simul. Comput.*, vol. 52, no. 1, pp. 110–125, 2023, https://doi.org/10.1080/03610918.2020.1850790

[23] M. S. Islam, S. Shikalgar, and M. Noor-E-Alam, "A two-stage feature selection approach for robust evaluation of treatment effects in high-dimensional observational data," *IISE Trans. Healthc. Syst. Eng.*, vol. 15, no. 2, pp. 117–136, 2025, https://doi.org/10.1080/24725579.2024.2447715

[24] Y. Liu, J.-S. Wang, J.-Y. Wen, Y.-T. Li, and P.-G. Yan, "A multi-objective evolutionary algorithm for solving the feature selection problem of high-dimensional sparse data and its application in the absorption, distribution, metabolism, excretion and toxicity (ADMET) classification," *Eng. Optim.*, vol. 58, no. 1, pp. 1–32, 2025, https://doi.org/10.1080/0305215X.2025.2464861

[25] L. Sun and A. Barbu, "Stochastic feature selection with annealing and its applications to streaming data," *J. Nonparam. Stat.*, vol. 37, no. 3, pp. 1–18, 2025, https://doi.org/10.1080/10485252.2025.2456767

[26] S. Sharma, V. Singh, A. Singh, and A. Bhardwaj, "Software failure prediction using hybrid deep learning model with optimization-enabled feature selection," *Commun. Stat. Simul. Comput.*, 2025, pp. 1-22, https://doi.org/10.1080/03610918.2025.2571974

[27] W. Yu, S. Wade, H. D. Bondell, and L. Azizi, "Nonstationary Gaussian process discriminant analysis with variable selection for high-dimensional functional data," *J. Comput. Graph. Stat.*, vol. 32, no. 2, pp. 588-600, 2023, https://doi.org/10.1080/10618600.2022.2098136

[28] R. A. Rayan, A. Suruliandi, and S. Raja, "Modified mutual information feature selection algorithm to predict COVID-19 using clinical data," *Comput. Methods Biomech. Biomed. Eng.*, vol. 29, no. 6, pp. 1193-1213, 2024, https://doi.org/10.1080/10255842.2024.2429012

[29] M. Demirarslan and A. Suner, "OCtS: An alternative of the t-score method sensitive to outliers and correlation in feature selection," *Commun. Stat. Simul. Comput.*, vol. 53, no. 3, pp. 1409–1422, 2024, https://doi.org/10.1080/03610918.2022.2046087

[30] Y. Shi, J. Fang, J. Li, K. Yu, J. Zhu, and Y. Lu, "Fracture risk prediction in diabetes patients based on Lasso feature selection and machine learning," *Comput. Methods Biomech. Biomed. Eng.*, vol. 29, no. 3, pp. 511–527, 2024, https://doi.org/10.1080/10255842.2024.2400325

[31] J. Chen, R. H. Aseltine, F. Wang, and K. Chen, "Tree-guided rare feature selection and logic aggregation with electronic health records data," *J. Am. Stat. Assoc.*, vol. 119, no. 547, pp. 1765–1777, 2024, https://doi.org/10.1080/01621459.2024.2326621

[32] C. Huang, M. Wang, H. A. Asghar, Z. Wang, and H. Chen, "Q-learning enhanced differential evolution for feature selection in high-dimensional medical data analysis," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 37, no. 9, pp. 280, 2025, https://doi.org/10.1007/s44443-025-00303-z

[33] M. Teke, T. Etem, and M. Karhan, "Enhancing anemia diagnosis using ensemble machine learning and feature selection techniques on CBC data," *Eur. Phys. J. Spec. Top.*, vol. 234, no. 15, pp. 4635–4647, 2025, https://doi.org/10.1140/epjs/s11734-025-01838-y

[34] M. Ghaffarzadeh-Esfahani, M. Ghaffarzadeh-Esfahani, A. Salahi-Niri, H. Toreyhi, Z. Atf, A. Mohsenzadeh-Kermani, *et al.,* "Large language models versus classical machine learning performance in COVID-19 mortality prediction using high-dimensional tabular data," *Sci. Rep.*, vol. 15, no. 1, no. 42712, 2025, https://doi.org/10.1038/s41598-025-26705-7

[35] S. Ayad, R. E. Mallouhy, and C. Guyeux, "A hybrid ontology-based feature selection framework for enhancing predictive accuracy in regression models," *Knowl. Inf. Syst.*, vol. 67, no. 10, pp. 9111–9145, 2025, https://doi.org/10.1007/s10115-025-02497-0

[36] S. K. R. Mallidi and R. R. Ramisetty, "Bowerbird courtship-inspired feature selection for efficient high-dimensional data analysis using a novel meta-heuristic," *Discov. Comput.*, vol. 28, no. 6, 2025, https://doi.org/10.1007/s10791-025-09497-2

[37] M. Braik, M. A. Awadallah, H. Alzoubi, and H. Al-Hiary, "Heterogeneous cognitive learning chameleon swarm algorithm for high-dimensional feature selection," *J. Supercomput.*, vol. 81, no. 5, pp. 652, 2025, https://doi.org/10.1007/s11227-025-07139-4

[38] M. Asha and T. Johnpeter, "Optimized ensemble learning framework with enhanced feature selection for efficient and accurate classification," *SN Comput. Sci.*, vol. 6, no. 6, pp. 669, 2025, https://doi.org/10.1007/s42979-025-04205-9

[39] M. Garcia-Torres, "Feature selection for high-dimensional data using a multivariate search space reduction strategy based scatter search," *J. Heuristics*, vol. 31, no. 1, pp. 10, https://doi.org/10.1007/s10732-025-09550-9

[40] C. Wang, Q. Chen, B. Xue, and M. Zhang, "Improving generalization of genetic programming for high-dimensional symbolic regression with Shapley value based feature selection," *Data Sci. Eng.*, vol. 10, no. 2, pp. 196–211, 2025, https://doi.org/10.1007/s41019-024-00270-x