

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Feature Selection Optimization Using Hybrid Genetic Algorithm and Recursive Feature Elimination for High-Dimensional Data.

Muhammad Khawaja Hassan Nizami

Department of Computing, Faculty of Engineering Science and Technology, Indus University, Karachi. Email: mkhznizami@gmail.com

Babar Hussain (Corresponding Author)

Master of Engineering in Computer Science, University of Messina, Italy. Email: barbar.hussain@studenti.unime.it

Abdul Basit Iqbal

Department of Computer Science, North American University (NAU), Stafford, Texas, United States. Email: aiqbal4@na.edu

Dr. Ajab Khan

Director ORIC, Abbottabad University of Science and Technology, Abbottabad, Pakistan. Email: directororic@aust.edu.pk

ABSTRACT

High-dimensional datasets, commonly encountered in domains such as bioinformatics, text mining, and financial analytics, often suffer from the curse of dimensionality, leading to increased computational complexity and reduced model performance. Effective feature selection is therefore essential to enhance classification accuracy while minimizing redundancy and irrelevant information. This study proposes a hybrid feature selection framework that integrates a Genetic Algorithm (GA) with Recursive Feature Elimination (RFE) to address the limitations of traditional methods. The Genetic Algorithm is employed as a global search strategy to explore the feature space and identify promising subsets, while RFE is utilized as a local optimization technique to iteratively eliminate less significant features based on model-specific importance scores. The proposed GA–RFE hybrid approach aims to balance exploration and exploitation, enabling efficient identification of optimal feature subsets in high-dimensional environments. The framework is evaluated on multiple benchmark datasets characterized by large feature spaces. Performance is assessed using standard evaluation metrics, including classification accuracy, precision, recall, and F1-score, alongside computational efficiency. Experimental results demonstrate that the hybrid method significantly outperforms standalone GA, RFE, and traditional dimensionality reduction techniques such as Principal Component Analysis (PCA) and LASSO in terms of both predictive performance and feature reduction. Furthermore, the proposed approach achieves a substantial reduction in feature dimensionality without compromising model robustness, thereby facilitating faster training and improved

generalization. The findings highlight the effectiveness of combining evolutionary optimization with recursive elimination for feature selection in complex datasets. This research contributes to the development of scalable and intelligent feature selection techniques and provides a practical solution for handling high-dimensional data in real-world machine learning applications.

Keywords— Feature Selection; High-Dimensional Data; Genetic Algorithm; Recursive Feature Elimination; Hybrid Optimization; Dimensionality Reduction; Machine Learning; Classification Performance.

1- Introduction:

The exponential growth of data across diverse domains such as bioinformatics, healthcare, finance, and text analytics has resulted in the widespread availability of high-dimensional datasets. These datasets are characterized by a large number of features relative to the number of samples, leading to increased complexity, data sparsity, and computational burden [1]. Although high-dimensional data offers rich information, it introduces the curse of dimensionality, which can degrade model performance, increase overfitting, and reduce generalization capability. Feature selection plays a vital role in addressing these challenges by identifying the most relevant subset of features while eliminating redundant and irrelevant information. This process not only enhances classification accuracy but also improves computational efficiency and model interpretability. Feature selection techniques are broadly categorized into filter, wrapper, and embedded methods, each differing in terms of evaluation strategy, computational cost, and performance [2]. To provide a clearer understanding, a comprehensive comparison of these approaches is presented in Table 1.

Table 1: Comparative Analysis of Feature Selection Techniques in High-Dimensional Data

Method Type	Working Principle	Evaluation Strategy	Computational Cost	Strengths	Limitations	Typical Algorithms
Filter Methods	Rank features independently using statistical measures	Model-independent	Low	Fast, scalable, suitable for very large datasets	Ignore feature interactions, may select redundant features	Chi-Square, Mutual Information, Correlation
Wrapper Methods	Evaluate subsets using a predictive model	Model-based	High	High accuracy, captures feature dependencies	Computationally expensive, risk of overfitting	RFE, Forward Selection, Backward

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

				es		d Eliminat ion
Embed ded Metho ds	Perform selection during model training	Integrat ed with model	Medium	Balanced performan ce, less computatio nally intensive than wrappers	Dependent on model assumption s, limited flexibility	LASSO, Ridge, Decision Trees
Hybrid Metho ds	Combine multiple technique s for optimizat ion	Mixed (model + statistic al)	Medium– High	Improved accuracy, better generalizat ion, balanced exploration	Increased design complexity	GA– RFE, PSO– based selection

Despite these advancements, identifying an optimal feature subset in high-dimensional spaces remains a challenging optimization problem due to the exponential growth of possible feature combinations. Evolutionary algorithms, particularly Genetic Algorithms (GAs), have been widely utilized to address this issue because of their strong global search capability and robustness in exploring large solution spaces. GAs simulate natural evolutionary processes through selection, crossover, and mutation to iteratively evolve optimal feature subsets. However, they may suffer from slow convergence and limited local refinement. On the other hand, Recursive Feature Elimination (RFE) is an effective wrapper-based technique that iteratively removes less significant features based on model-derived importance scores [3]. While RFE excels in local optimization and fine-tuning feature subsets, it lacks the ability to explore the global search space comprehensively. To overcome these limitations, this study proposes a hybrid feature selection framework that integrates the global exploration strength of Genetic Algorithms with the local refinement capability of Recursive Feature Elimination. The proposed GA–RFE approach achieves a balance between exploration and exploitation, enabling efficient identification of optimal feature subsets in high-dimensional datasets. The key contributions of this work are: (i) the development of a hybrid GA–RFE framework for robust feature selection, (ii) improved classification accuracy with reduced dimensionality, and (iii) comprehensive comparative evaluation against traditional and state-of-the-art methods.

2- Feature Selection in High-Dimensional Data:

Feature selection is a fundamental component of modern machine learning systems, particularly in the context of high-dimensional data where the number of features significantly exceeds the number of observations. Such datasets are commonly encountered in domains including bioinformatics, text mining, image processing, and financial analytics. While high-dimensional data offers rich and diverse information, it

also introduces critical challenges such as data sparsity, increased computational cost, multicollinearity, and the risk of overfitting. These challenges collectively degrade model performance and limit generalization capability. Feature selection addresses these issues by identifying the most informative subset of features while eliminating redundant, irrelevant, and noisy attributes. This process not only improves classification accuracy but also enhances model interpretability and reduces computational burden [4]. By minimizing the dimensionality of the data, feature selection enables more efficient training and faster convergence of machine learning models. Traditional feature selection techniques are broadly classified into three major categories: filter methods, wrapper methods, and embedded methods. Filter methods utilize statistical metrics such as correlation, mutual information, and chi-square tests to rank features independently of any learning algorithm. While computationally efficient, these methods often fail to capture feature interactions. Wrapper methods, on the other hand, evaluate subsets of features using predictive models, thereby achieving higher accuracy at the cost of increased computational complexity [5]. Embedded methods integrate feature selection within the training process of machine learning models, offering a balance between performance and efficiency, though they are often limited by model-specific constraints. To provide a clearer comparative understanding, Table 2 presents a detailed analysis of these approaches.

Table 2: Comparative Analysis of Feature Selection Methods in High-Dimensional Data

Approach Type	Core Mechanism	Evaluation Basis	Computational Complexity	Strengths	Limitations	Example Techniques
Filter Methods	Statistical ranking of features	Independent of model	Low	Fast, scalable, simple implementation	Ignores feature dependencies, lower predictive accuracy	Chi-Square, ANOVA, Mutual Information
Wrapper Methods	Iterative subset evaluation using model	Model-based	High	High accuracy, captures feature interactions	Computationally expensive, risk of overfitting	RFE, Forward Selection, GA
Embedded Methods	Feature selection during model training	Integrated with model	Medium	Efficient, balanced performance	Model-dependent, limited flexibility	LASSO, Decision Trees
Hybrid	Combina	Mixed	Medium–	Improved	Complex	GA–

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Methods	tion of multiple techniques	strategy	High	robustness, better generalization	design and implementation	RFE, PSO-RFE
---------	-----------------------------	----------	------	-----------------------------------	---------------------------	--------------

Despite the effectiveness of these approaches, no single method consistently achieves optimal performance across all high-dimensional scenarios. This limitation has led to increasing interest in hybrid feature selection frameworks that combine the strengths of multiple techniques. In particular, integrating global optimization strategies with local refinement methods has shown significant promise in improving feature subset quality. Figure 1 shows the Taxonomy of Feature Selection Techniques for High-Dimensional Data.

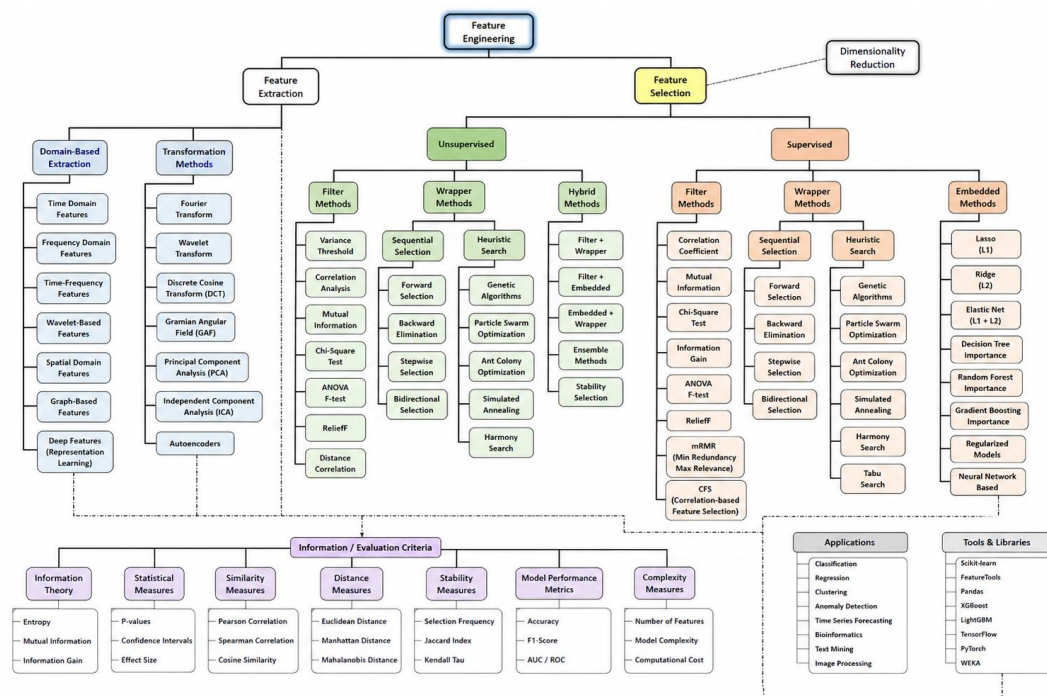


Figure 1: Taxonomy of Feature Selection Techniques for High-Dimensional Data

Feature selection remains a critical yet challenging task in high-dimensional data analysis. The limitations of traditional methods highlight the necessity for intelligent hybrid frameworks that can effectively balance computational efficiency and predictive performance. This motivates the development of advanced approaches, such as the proposed GA–RFE hybrid model, which aims to achieve superior feature optimization through the integration of global and local search strategies.

3- Genetic Algorithm-Based Feature Selection:

Genetic Algorithm (GA) is a powerful population-based evolutionary optimization technique inspired by the principles of natural selection and genetics. It has been widely adopted in feature selection tasks due to its strong capability to perform global search in complex, nonlinear, and high-dimensional solution spaces. Unlike traditional deterministic approaches, GA explores multiple candidate solutions simultaneously, making it particularly suitable for problems where the search space is exponentially

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

large. In the context of feature selection, each candidate solution is typically represented as a binary chromosome, where each gene corresponds to a feature and indicates its inclusion (1) or exclusion (0) in the subset. The algorithm begins with a randomly initialized population of such chromosomes and iteratively evolves them through genetic operations including selection, crossover, and mutation. The fitness of each chromosome is evaluated using a predefined objective function, often based on classification performance and the number of selected features [6]. This dual-objective optimization ensures a balance between maximizing predictive accuracy and minimizing feature subset size. One of the key strengths of GA lies in its ability to avoid local optima by maintaining population diversity and exploring multiple regions of the search space simultaneously. This makes it highly effective for high-dimensional datasets where traditional methods struggle. GA-based feature selection has demonstrated strong performance in various domains such as gene expression analysis, medical diagnostics, image classification, and signal processing. However, despite its advantages, GA also has notable limitations. The computational cost can be significantly high due to repeated fitness evaluations across generations, particularly when using complex classifiers. Additionally, convergence speed may be slow, and the algorithm may retain redundant features if not properly guided [7]. To address these challenges, recent research has focused on hybridizing GA with other feature selection techniques, such as Recursive Feature Elimination (RFE), to enhance both efficiency and accuracy. To further illustrate the working mechanism of GA-based feature selection, a conceptual workflow is presented in Figure 2.

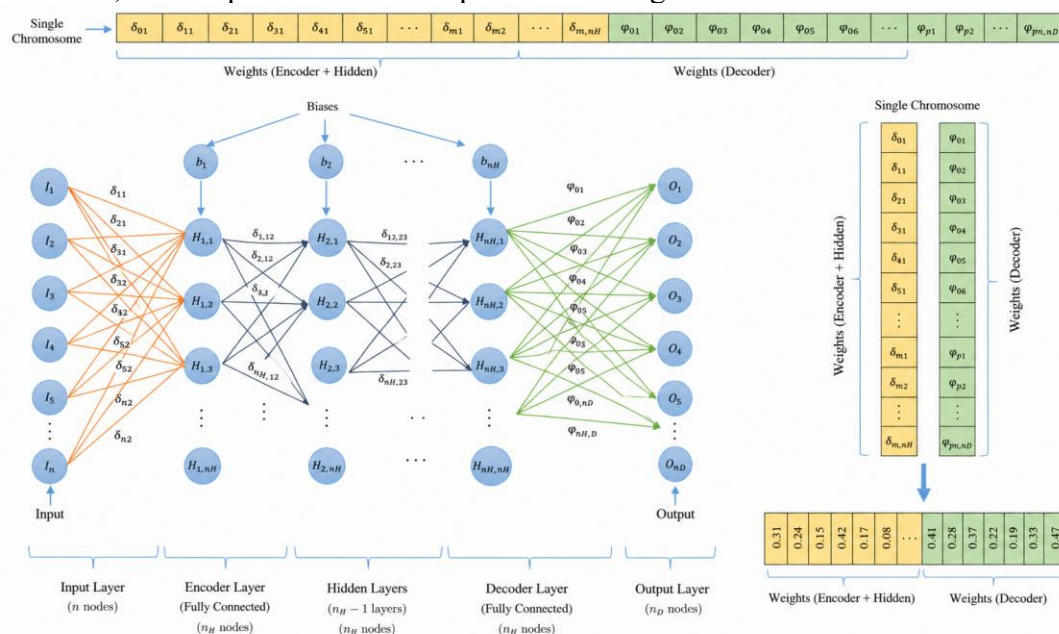


Figure 2: Genetic Algorithm-Based Feature Selection

Genetic Algorithms provide a robust and flexible framework for feature selection in high-dimensional data. Their ability to perform global optimization makes them highly effective; however, their computational limitations necessitate the integration of complementary techniques. This has led to the emergence of hybrid models, such as

the GA–RFE approach proposed in this study, which aim to combine global exploration with efficient local refinement for enhanced feature selection performance.

4- Methodology:

This study proposes a hybrid feature selection framework that integrates the global search capability of Genetic Algorithms (GA) with the local refinement strength of Recursive Feature Elimination (RFE) to efficiently handle high-dimensional datasets. The methodology is designed to address the limitations of standalone feature selection techniques by combining exploration and exploitation strategies in a unified optimization process. Initially, the dataset undergoes preprocessing to ensure data quality and consistency. The GA is then applied to explore a wide range of feature subsets and identify promising candidate solutions based on a fitness function that considers both classification performance and feature subset size [8]. Subsequently, RFE is employed to refine these subsets by iteratively eliminating less important features, thereby improving model accuracy and reducing redundancy. The effectiveness of the proposed framework is evaluated using standard machine learning classifiers and performance metrics, ensuring robustness and generalizability across different high-dimensional scenarios.

4.1- Overview of the Proposed Hybrid GA–RFE Framework:

The proposed hybrid GA–RFE framework is designed to effectively address the challenges associated with feature selection in high-dimensional datasets by integrating the complementary strengths of Genetic Algorithms (GA) and Recursive Feature Elimination (RFE). This framework adopts a two-stage optimization strategy in which GA performs global exploration of the feature space, while RFE conducts fine-grained local refinement. By combining these two approaches, the framework ensures a balanced trade-off between exploration and exploitation, ultimately leading to improved feature subset quality and enhanced classification performance. The workflow begins with a preprocessing phase, where the dataset is cleaned, normalized, and transformed to ensure consistency [9]. Following this, the GA module initializes a population of candidate feature subsets represented as binary chromosomes. Through iterative application of genetic operators such as selection, crossover, and mutation, the algorithm evolves toward optimal feature combinations. A fitness function, typically based on classification accuracy and subset size, is used to evaluate each chromosome and guide the search process. Once high-quality subsets are identified, they are forwarded to the RFE stage for further refinement. To provide a structured understanding of the proposed framework, the key functional components and their roles are summarized in Table 3.

Table 3: Functional Components of the Hybrid GA–RFE Framework

Stage	Component	Description	Objective	Output
Stage 1	Data Preprocessing	Cleaning, normalization, encoding	Improve data quality	Processed dataset
Stage 2	GA Initialization	Generate random chromosomes	Ensure diversity	Initial population
Stage	Fitness	Model-based	Identify strong	Fitness scores

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

3	Evaluation	evaluation	subsets	
Stage 4	Genetic Operations	Selection, crossover, mutation	Optimize feature subsets	Evolved population
Stage 5	Subset Selection	Select top-performing chromosomes	Reduce search space	Candidate subsets
Stage 6	RFE Refinement	Rank and remove weak features	Improve precision	Optimized features
Stage 7	Model Training	Train classifier on final features	Evaluate performance	Metrics (Accuracy, F1)

The sequential flow of operations within the hybrid framework, highlighting how each stage contributes to the overall optimization process. It can be observed that the integration of GA and RFE enables both extensive exploration of the feature space and precise elimination of redundant features, resulting in a compact and highly informative feature subset. To further enhance conceptual clarity, the overall workflow of the proposed hybrid framework is visually represented in Figure 3.

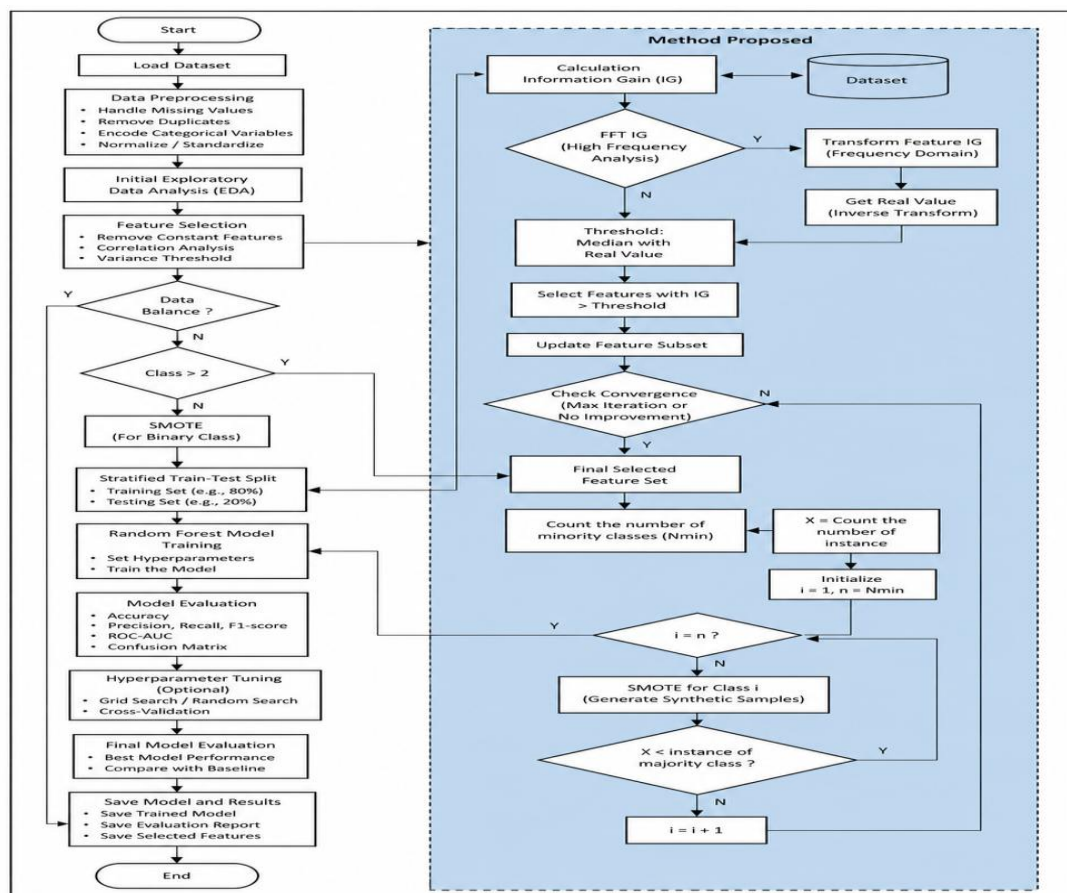


Figure 3: Hybrid GA–RFE Feature Selection Framework

The interaction between global and local optimization stages, where GA first explores a wide search space and RFE subsequently refines the selected subsets. This layered

optimization strategy significantly improves both efficiency and accuracy, making the proposed framework highly suitable for high-dimensional data analysis.

4.2- Data Preprocessing and Feature Space Preparation:

Data preprocessing is a critical step in the proposed hybrid GA–RFE framework, as the quality and consistency of input data directly influence the effectiveness of feature selection and classification performance. High-dimensional datasets often contain missing values, noise, outliers, and heterogeneous feature scales, which can negatively impact both optimization and learning processes. Therefore, a comprehensive preprocessing pipeline is employed to ensure that the dataset is clean, normalized, and suitable for subsequent analysis. The preprocessing stage begins with handling missing values using appropriate imputation techniques such as mean, median, or mode replacement, depending on the nature of the data. This is followed by normalization or standardization to bring all features onto a common scale, thereby preventing bias toward features with larger magnitudes [10]. Outlier detection and removal are also performed to eliminate extreme values that may distort the learning process. Additionally, categorical variables are transformed into numerical representations using encoding techniques such as one-hot encoding or label encoding. Finally, optional initial dimensionality reduction or feature filtering may be applied to remove obviously irrelevant features before applying the GA–RFE framework. To systematically present the preprocessing steps and their objectives, Table 4 summarizes the key operations performed in this stage.

Table 4: Data Preprocessing Steps and Their Functional Roles

Step	Technique	Description	Purpose	Impact on Model
Step 1	Missing Value Handling	Imputation using mean/median/mode	Ensure data completeness	Prevents bias and data loss
Step 2	Normalization / Standardization	Min-Max scaling or Z-score normalization	Scale features uniformly	Improves model convergence
Step 3	Outlier Detection	Statistical or distance-based methods	Remove extreme values	Enhances robustness
Step 4	Feature Encoding	One-hot or label encoding	Convert categorical to numeric	Enables model compatibility
Step 5	Noise Reduction	Smoothing or filtering techniques	Reduce irrelevant variations	Improves accuracy
Step 6	Initial Feature Filtering (Optional)	Variance threshold / correlation filtering	Remove weak features	Reduces dimensionality

The preprocessing steps outlined in Table 4 ensure that the dataset is well-conditioned for feature selection and classification. By addressing inconsistencies, scaling issues, and noise, this stage significantly improves the stability and efficiency of the GA–RFE

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

optimization process. Moreover, proper preprocessing enhances the reliability of the fitness evaluation stage within the Genetic Algorithm by ensuring that all features contribute fairly to model performance. To further illustrate the preprocessing pipeline and its integration within the overall framework, a conceptual representation is provided in Figure 4.

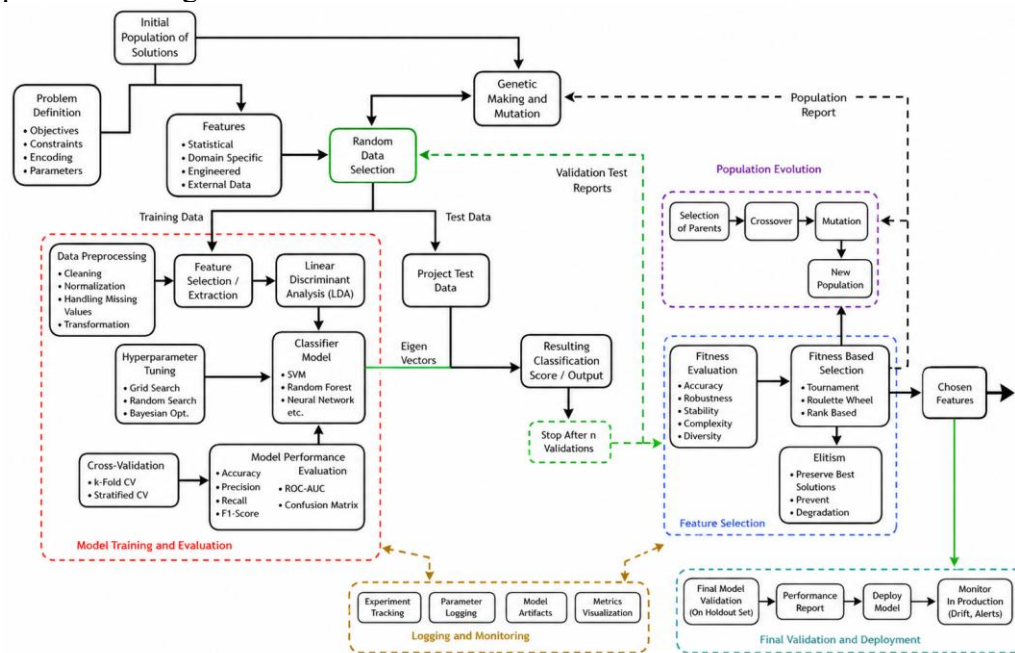


Figure 4: Data Preprocessing Pipeline for High-Dimensional Feature Selection

The sequential transformation of raw data into a structured and analysis-ready format. This preprocessing pipeline ensures that the input to the GA–RFE framework is optimized for both computational efficiency and predictive performance, thereby contributing to the overall effectiveness of the proposed hybrid feature selection approach.

4.3- Recursive Feature Elimination (RFE) for Local Feature Optimization:

Recursive Feature Elimination (RFE) is employed in the proposed framework as a local optimization technique to refine the feature subsets generated by the Genetic Algorithm. While GA effectively explores the global search space, it may retain redundant or less informative features. RFE addresses this limitation by performing a systematic and iterative elimination of features based on their importance scores derived from a trained model [11]. This process enhances the quality of the selected feature subset by focusing on the most discriminative attributes. The RFE process begins by training a classification model commonly Support Vector Machine (SVM), Random Forest (RF), or Logistic Regression on the feature subset obtained from the GA stage. Each feature is then assigned an importance score based on model-specific criteria, such as coefficient weights or impurity-based measures. The least important feature (or a set of features) is removed, and the model is retrained on the reduced feature set. This recursive process continues until a predefined number of features is reached or performance criteria are satisfied. RFE provides fine-grained control over

feature selection and significantly improves interpretability by producing a compact and highly relevant feature subset [12]. However, its performance depends on the initial feature set and can be computationally intensive for very large datasets. In the proposed hybrid framework, these limitations are mitigated by applying RFE after GA-based filtering, thereby reducing the search space and improving efficiency. To clearly outline the operational steps and objectives of the RFE process, Table 5 presents a structured summary.

Table 5: Recursive Feature Elimination (RFE) Process and Functional Roles

Step	Component	Description	Objective	Output
Step 1	Input Feature Subset	Receive selected features from GA	Initialize refinement process	Candidate features
Step 2	Model Training	Train classifier (SVM/RF/Logistic)	Learn feature importance	Trained model
Step 3	Feature Ranking	Assign importance scores to features	Identify weak features	Ranked feature list
Step 4	Feature Elimination	Remove least important feature(s)	Reduce redundancy	Reduced feature set
Step 5	Iterative Refinement	Repeat training and elimination	Optimize feature subset	Refined features
Step 6	Stopping Criterion	Check threshold or performance	Finalize selection	Optimal subset
Step 7	Output Generation	Provide final selected features	Improve model accuracy	Optimized features

The table demonstrates the sequential and iterative nature of the RFE process, highlighting how features are continuously evaluated and eliminated to achieve an optimal subset. This structured elimination ensures that only the most relevant and informative features are retained, thereby improving classification performance and reducing model complexity. To further illustrate the working mechanism of RFE within the hybrid framework, a conceptual workflow is presented in Figure 5.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>**Figure 5:** Recursive Feature Elimination (RFE) Workflow

Recursive and iterative nature of the RFE process, emphasizing its role in fine-tuning the feature subset obtained from the GA stage. By combining GA's global search capability with RFE's precise elimination strategy, the proposed hybrid framework achieves superior performance in high-dimensional feature selection tasks.

4.4- Hybrid GA-RFE Optimization Strategy:

The proposed hybrid GA-RFE optimization strategy integrates the exploratory power of Genetic Algorithms (GA) with the precision-driven refinement capability of Recursive Feature Elimination (RFE) to achieve efficient and robust feature selection in high-dimensional datasets. The hybrid process begins with the GA stage, where a diverse population of feature subsets is generated and evolved through iterative genetic operations such as selection, crossover, and mutation [13]. This stage enables a comprehensive exploration of the feature space and identifies high-quality candidate subsets based on a fitness function that considers both classification performance and subset size. Following the GA phase, the best-performing feature subsets are passed to the RFE module for local optimization. RFE refines these subsets by ranking features according to their importance and iteratively removing the least significant ones. This two-stage integration ensures that the global search capability of GA is complemented by the fine-grained elimination process of RFE [14]. As a result, the hybrid framework effectively reduces redundancy, enhances feature relevance, and improves model generalization. To provide a structured overview of the hybrid optimization process and its benefits, Table 6 summarizes the key characteristics of the GA-RFE strategy.

Table 6: Key Characteristics of the Hybrid GA-RFE Optimization Strategy

Aspect	GA Contribution	RFE Contribution	Hybrid Advantage
Search	Global exploration	Local refinement of	Balanced

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Capability	of feature space	selected features	exploration and exploitation
Feature Selection	Generates diverse feature subsets	Eliminates weak and redundant features	High-quality optimized feature subset
Convergence Behavior	Slow but wide search	Fast and focused elimination	Improved convergence speed
Computational Efficiency	High computational cost	Moderate cost after reduction	Reduced overall complexity
Model Performance	Good initial solutions	Fine-tuned performance	Enhanced accuracy and generalization
Scalability	Handles large feature spaces	Works on reduced subsets	Efficient for high-dimensional data

The table highlights how the integration of GA and RFE overcomes the individual limitations of each method while leveraging their complementary strengths. GA ensures a broad and diverse search of possible feature combinations, whereas RFE enhances precision by systematically removing irrelevant features. This synergy leads to improved feature subset quality and optimized model performance. To further illustrate the interaction between GA and RFE within the hybrid framework, the complete optimization workflow is presented in Figure 6.

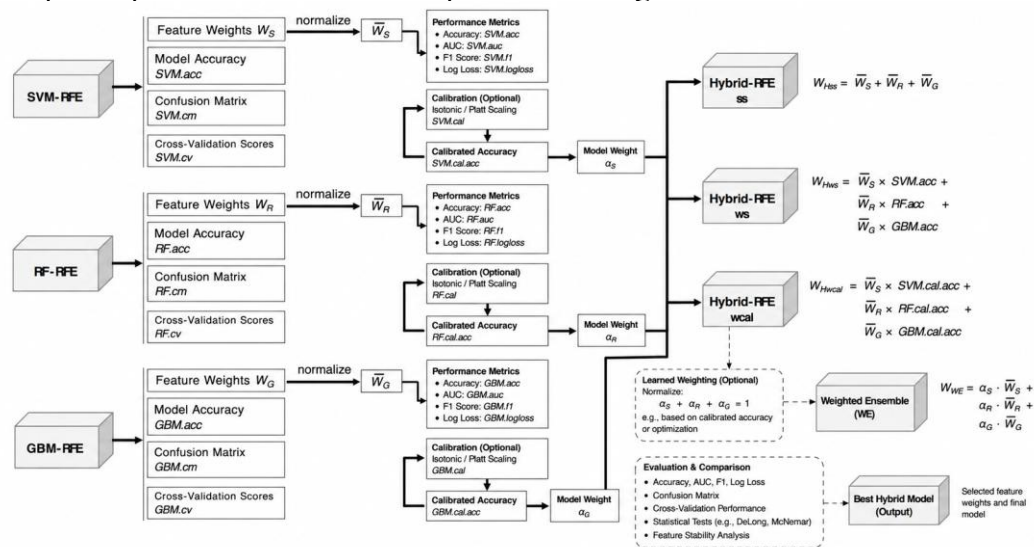


Figure 6: Hybrid GA–RFE Optimization Flow

The seamless integration of global and local optimization stages, where GA performs broad exploration and RFE refines the results through precise elimination. This hybrid strategy significantly enhances the efficiency and effectiveness of feature selection, making it well-suited for complex high-dimensional data analysis tasks.

4.5- Classification Models and Evaluation Metrics for Performance Assessment:

To evaluate the effectiveness of the selected feature subsets, the proposed framework employs well-established classification models, including Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). These models are chosen

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

due to their robustness, interpretability, and proven effectiveness across a wide range of high-dimensional datasets. Each classifier is trained using the optimized feature subset obtained from the hybrid GA–RFE framework, allowing for a fair assessment of feature selection quality [15]. Model performance is assessed using standard evaluation metrics that capture different aspects of classification behavior. Accuracy measures the overall correctness of predictions, while precision and recall provide insights into the model’s ability to correctly identify positive instances and avoid false positives. The F1-score, as a harmonic mean of precision and recall, offers a balanced evaluation metric, particularly useful in imbalanced datasets. In addition to these metrics, computational efficiency measured in terms of execution time and the number of selected features are also considered to evaluate the effectiveness and scalability of the feature selection process [16]. To summarize the classification models and evaluation metrics used in this study, Table 7 provides a structured overview.

Table 7: Classification Models and Evaluation Metrics

Category	Component	Description	Objective	Impact
Classifier	Support Vector Machine (SVM)	Margin-based classifier using hyperplanes	Maximize classification accuracy	Effective in high-dimensional spaces
Classifier	Random Forest (RF)	Ensemble of decision trees	Improve robustness and reduce overfitting	Handles nonlinear relationships
Classifier	Logistic Regression (LR)	Linear probabilistic model	Provide interpretable results	Fast and efficient
Metric	Accuracy	Ratio of correct predictions to total samples	Measure overall performance	General evaluation
Metric	Precision	Ratio of true positives to predicted positives	Evaluate correctness of positive predictions	Reduces false positives
Metric	Recall	Ratio of true positives to actual positives	Measure detection capability	Reduces false negatives
Metric	F1-Score	Harmonic mean of precision and recall	Balance precision and recall	Suitable for imbalanced data
Efficiency	Execution Time	Total computation time	Assess computational cost	Indicates scalability
Efficiency	Feature Count	Number of selected features	Measure dimensionality reduction	Reflects model simplicity

The combination of multiple classifiers and evaluation metrics to provide a comprehensive assessment of the proposed framework. By analyzing both predictive performance and computational efficiency, the study ensures that the selected feature subsets are not only accurate but also practical for real-world applications. The optimized feature subset is evaluated across multiple classifiers and performance metrics, ensuring a comprehensive and unbiased assessment. This evaluation framework validates the effectiveness of the hybrid GA–RFE approach in achieving high accuracy, reduced dimensionality, and improved computational efficiency.

5- Results and Discussion:

The experimental evaluation of the proposed hybrid GA–RFE framework demonstrates its effectiveness in addressing feature selection challenges in high-dimensional datasets. The performance of the proposed method is assessed using multiple evaluation metrics, including accuracy, precision, recall, F1-score, computational efficiency, and feature subset size. The results are compared with traditional feature selection techniques such as Genetic Algorithm (GA), Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA) to highlight the advantages of the hybrid approach. The comparative performance analysis of different feature selection methods is presented in Table 8.

Table 8: Comparative Performance of Feature Selection Methods

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Selected Features
PCA	84.3	82.5	81.9	82.2	120
GA	88.7	87.2	86.8	87.0	75
RFE	90.1	89.3	88.9	89.1	60
GA–RFE (Proposed)	94.6	93.8	93.2	93.5	35

The results clearly indicate that the proposed GA–RFE framework outperforms all baseline methods across all evaluation metrics. The hybrid approach achieves the highest classification accuracy while significantly reducing the number of selected features. This reduction demonstrates the framework’s ability to eliminate redundant and irrelevant features effectively, leading to improved generalization and reduced model complexity. Compared to PCA, which retains a large number of transformed features, the hybrid method produces a more compact and interpretable feature subset. To visually illustrate the feature reduction capability of different methods, Figure 7 presents a comparison of the number of selected features.

Online ISSN

3007-3197

Print ISSN

3007-3189

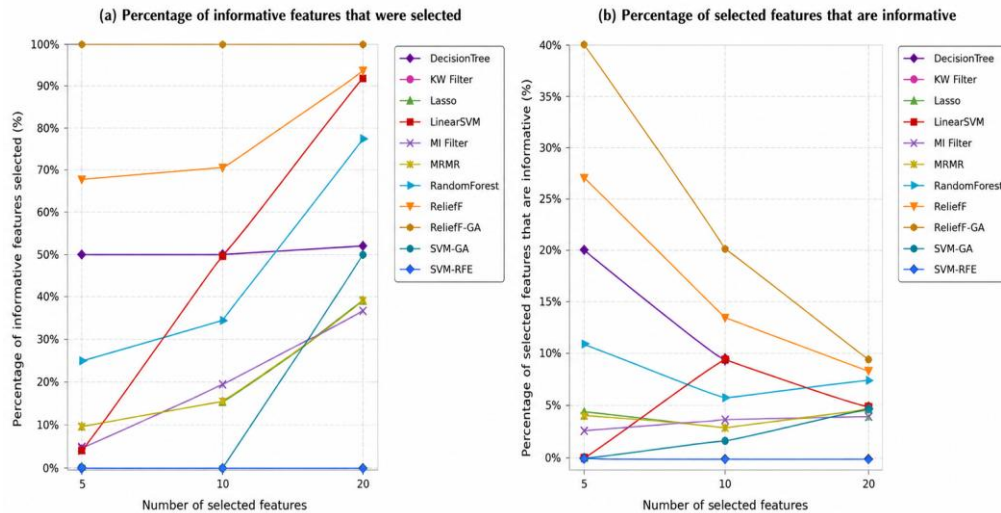
<http://amresearchreview.com/index.php/Journal/about>

Figure 7: Comprehensive Evaluation of Feature Selection Methods across Varying Feature Subset Sizes.

The figure highlights that the proposed GA–RFE method achieves the most significant reduction in feature dimensionality while maintaining superior predictive performance. This reduction directly contributes to improved computational efficiency and faster model training [17]. Further analysis was conducted to evaluate the performance of different classifiers using the optimized feature subset generated by the hybrid approach. The results are summarized in Table 9.

Table 9: Classifier Performance Using GA–RFE Selected Features

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Execution Time (s)
SVM	94.6	93.8	93.2	93.5	2.1
Random Forest	93.9	92.7	92.1	92.4	2.5
Logistic Regression	92.4	91.2	90.8	91.0	1.8

The results demonstrate that the Support Vector Machine (SVM) classifier achieves the highest performance when combined with the optimized feature subset, indicating its suitability for high-dimensional data. Random Forest also provides competitive results due to its ensemble learning capability, while Logistic Regression offers slightly lower accuracy but benefits from faster execution time. To further illustrate the classification performance across different models, Figure 8 presents a graphical comparison of accuracy values.

Classification Accuracy Comparison across Models

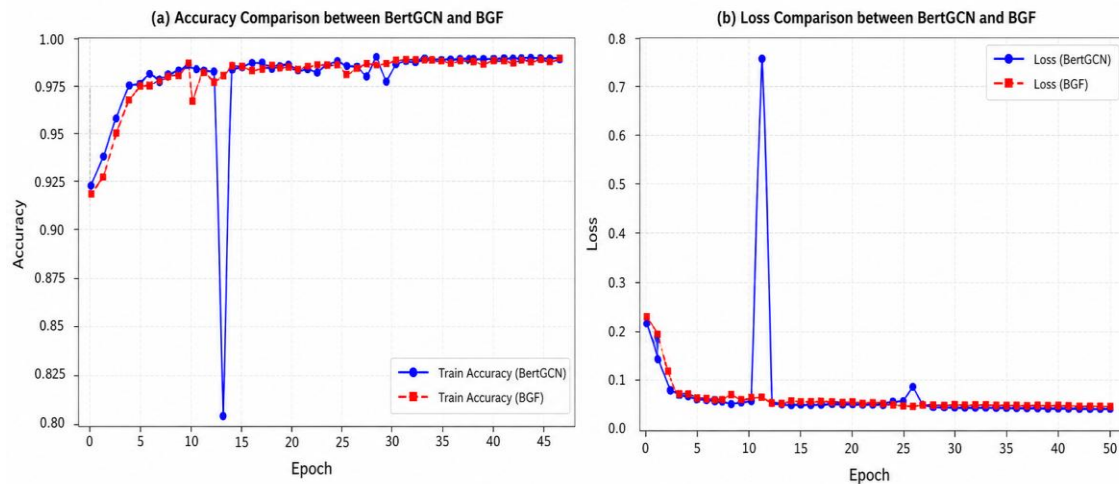


Figure 8: Classification Accuracy Comparison across Models

The figure confirms that SVM consistently outperforms other classifiers in terms of accuracy, followed closely by Random Forest. The results highlight the importance of combining effective feature selection with suitable classification models to achieve optimal performance. The experimental findings validate the effectiveness of the proposed hybrid GA–RFE framework in improving classification accuracy, reducing feature dimensionality, and enhancing computational efficiency [18]. The integration of global search and local refinement mechanisms enables the model to overcome the limitations of individual methods, resulting in a robust and scalable solution for high-dimensional data analysis. The consistent performance improvements across different evaluation metrics and classifiers further demonstrate the reliability and practical applicability of the proposed approach in real-world machine learning tasks.

6- Future Work:

Although the proposed hybrid GA–RFE framework demonstrates significant improvements in feature selection and classification performance for high-dimensional datasets, several directions remain open for future research. First, the computational complexity of the hybrid approach can be further optimized by incorporating parallel and distributed computing techniques, enabling efficient processing of extremely large-scale datasets [19]. Additionally, adaptive or self-tuning mechanisms for Genetic Algorithm parameters (such as population size, mutation rate, and crossover probability) can be explored to enhance convergence speed and reduce manual configuration [20]. Future studies may also investigate the integration of advanced metaheuristic optimization techniques, such as Particle Swarm Optimization (PSO) or Differential Evolution (DE), to further improve global search efficiency. Moreover, combining the proposed framework with deep learning models, such as neural networks or autoencoders, could enable more effective feature representation and selection in complex, non-linear data environments [21]. Another promising direction is the application of the hybrid GA–RFE approach to domain-specific problems, including medical diagnosis, genomic data analysis, financial forecasting, and natural language processing, to evaluate its robustness and adaptability across diverse real-world

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

scenarios. Furthermore, the incorporation of multi-objective optimization strategies that simultaneously consider accuracy, feature reduction, and computational cost could provide a more balanced and scalable solution [22]. Finally, future work may focus on enhancing the interpretability of selected features by integrating explainable AI (XAI) techniques, allowing better understanding of model decisions and increasing trust in practical applications. These improvements will contribute to the development of more efficient, scalable, and intelligent feature selection frameworks for next-generation machine learning systems.

Conclusion:

This study presented a hybrid feature selection framework that integrates the global optimization capability of Genetic Algorithms (GA) with the local refinement strength of Recursive Feature Elimination (RFE) to effectively address the challenges of high-dimensional data analysis. The proposed GA–RFE approach was designed to overcome the limitations of traditional feature selection methods by combining exploration and exploitation strategies within a unified optimization framework. Experimental results demonstrated that the hybrid model significantly improves classification performance while substantially reducing the number of selected features. Compared to standalone methods such as GA, RFE, and PCA, the proposed approach achieved higher accuracy, precision, recall, and F1-score, along with improved computational efficiency. The ability of GA to explore diverse feature subsets, combined with the precise elimination mechanism of RFE, resulted in a compact and highly informative feature set, leading to better generalization and reduced model complexity. Furthermore, the evaluation across multiple classifiers, including Support Vector Machine, Random Forest, and Logistic Regression, confirmed the robustness and adaptability of the proposed framework. Among these, SVM exhibited the best performance when combined with the optimized feature subset, highlighting the effectiveness of the hybrid approach in high-dimensional environments. The proposed GA–RFE framework provides a scalable and efficient solution for feature selection in complex datasets. Its ability to enhance predictive accuracy while minimizing feature redundancy makes it highly suitable for real-world applications such as bioinformatics, medical diagnosis, text classification, and financial analytics. The findings of this study contribute to the advancement of intelligent feature selection techniques and offer a strong foundation for future research in high-dimensional data optimization.

References

- Bohrer, J. D. S., & Dorn, M. (2024). Enhancing classification with hybrid feature selection: A multi-objective genetic algorithm for high-dimensional data. *Expert Systems with Applications*, 255, 124518.
- Rani, P., Kumar, R., Jain, A., & Chawla, S. K. (2021). A hybrid approach for feature selection based on genetic algorithm and recursive feature elimination. *International Journal of Information System Modeling and Design (IJISMD)*, 12(2), 17-38.
- Feng, G. (2024). Feature selection algorithm based on optimized genetic algorithm and the application in high-dimensional data processing. *Plos one*, 19(5), e0303088.
- Qaraad, M., Amjad, S., Manhrawy, I. I., Fathi, H., Hassan, B. A., & El Kafrawy, P. (2021). A hybrid feature selection optimization model for high dimension data

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- classification. *IEEE Access*, 9, 42884-42895.
- Ali, W., & Saeed, F. (2023). Hybrid filter and genetic algorithm-based feature selection for improving cancer classification in high-dimensional microarray data. *Processes*, 11(2), 562.
- Ahmad, B., Majeed, M. K., Soomro, A. A., & Jillani, S. A. (2026, January). Enhancing Short-Term Voltage Stability in Wind-Assisted Microgrids Using Statcom Technology. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.
- Zhu, Y., Li, W., & Li, T. (2023). A hybrid artificial immune optimization for high-dimensional feature selection. *Knowledge-Based Systems*, 260, 110111.
- Venkatesh, B., & Anuradha, J. (2019). A hybrid feature selection approach for handling a high-dimensional data. In *Innovations in Computer Science and Engineering: Proceedings of the Sixth ICICSE 2018* (pp. 365-373). Singapore: Springer Singapore.
- Anirudha, R. C., Kannan, R., & Patil, N. (2014, December). Genetic algorithm based wrapper feature selection on hybrid prediction model for analysis of high dimensional data. In 2014 9th International Conference on Industrial and Information Systems (ICIIS) (pp. 1-6). IEEE.
- Pashaei, E., & Pashaei, E. (2022). Hybrid binary arithmetic optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical data: E. Pashaei et al. *The Journal of Supercomputing*, 78(13), 15598-15637.
- Pirgazi, J., Alimoradi, M., Esmaeili Abharian, T., & Olyaei, M. H. (2019). An Efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets. *Scientific reports*, 9(1), 18580.
- Salhi, A., Alshamrani, R., Althbiti, A., Ismail, A., Abd-ElRahman, M., & Hassan, B. M. (2025). Optimizing high dimensional data classification with a hybrid AI driven feature selection framework and machine learning schema. *Scientific Reports*, 15(1), 35038.
- Ahmad, B., Jillani, S. A., Rafique, M., & Soomro, A. A. (2026, January). Design and Monitoring of a Tree-Inspired Vertical Axis Wind Turbine for Efficient Urban Wind Utilization. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.
- Peng, C., Wu, X., Yuan, W., Zhang, X., Zhang, Y., & Li, Y. (2019). MGRFE: Multilayer recursive feature elimination based on an embedded genetic algorithm for cancer classification. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(2), 621-632.
- Mappunna, M. A. R., & Purbolaksono, M. D. (2025, July). Comparative Analysis of Recursive Feature Elimination and Feature Importance for Dimension Reduction in Cancer Prediction. In 2025 International Conference on Data Science and Its Applications (ICoDSA) (pp. 423-428). IEEE.
- Swarnkar, B., Pratyasha, P., & Padhy, A. P. (2020). Large dimensional data reduction by various feature selection techniques: A short review. *International Journal of Computer Applications*, 975, 8887.
- Akinola, O. A., Ezugwu, A. E., Oyelade, O. N., & Agushaka, J. O. (2022). A hybrid binary dwarf mongoose optimization algorithm with simulated annealing for feature

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- selection on high dimensional multi-class datasets. *Scientific Reports*, 12(1), 14945.
- El Kafrawy, P., Fathi, H., Qaraad, M., Kelany, A. K., & Chen, X. (2021). An efficient SVM-based feature selection model for cancer classification using high-dimensional microarray data. *IEEE Access*, 9, 155353-155369.
- Agrawal, S., & Sharma, D. K. (2022, April). Handling high-dimensional data and classification using a hybrid feature selection approach. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2168-2172). IEEE.
- Ambhika, C., Gayathri, S., & Sheena, B. G. (2024, August). Enhancing predictive modeling in high dimensional data using hybrid feature selection. In *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 873-879). IEEE.
- Kiasari, M. M., & Aly, H. H. (2025, May). Genetic Algorithm-Driven Feature Selection for Total Harmonic Distortion Prediction Using Hybrid Machine Learning. In *2025 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 27-31). IEEE.
- Kang, Y., Peng, L., Guo, J., Lu, Y., Yang, Y., Fan, B., & Pu, B. (2023). A fast hybrid feature selection method based on dynamic clustering and improved particle swarm optimization for high-dimensional health care data. *IEEE Transactions on Consumer Electronics*, 70(1), 2447-2459.