Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

# Comparative Analysis of Filter-based Feature Selection Methods for High-Dimensional Data in Classification Tasks

*Shengjie Min*[1], *Chuanli Wei*[1.2]
[1] *Statistics, The University of Georgia, GA, USA*
[1.2] *Computer Science, University of Southern California, CA, USA*

DOI: 10.69987/JACS.2023.30803

**Keywords**

Feature selection, high-dimensional data classification, feature filtering methods, dimensionality reduction

**Abstract**

High-dimensional data classification encounters substantial computational barriers when feature spaces exceed sample sizes by orders of magnitude. Filter-based feature selection addresses this dimensionality curse through statistical independence between feature evaluation and classifier training stages. This study examines six prevalent feature filtering methods across datasets ranging from $10^3$ to $10^5$ dimensions, measuring their impact on classification accuracy, computational overhead, and feature subset stability. Experimental results demonstrate that correlation-based approaches achieve 8.7% higher accuracy than variance thresholding on bioinformatics datasets while maintaining O (n log n) time complexity. Chi-square statistics test and mutual information methods exhibit comparable performance on categorical data with divergent behavior on continuous features. The analysis reveals trade-offs between its statistical power and computational tractability, with F-score emerging as optimal for balanced datasets and ReliefF excelling under class imbalance conditions. Performance degradation appears beyond $10^4$ features for correlation methods due to spurious associations, suggesting hybrid architectures for ultra-high-dimensional data processing problems.

## 1. Introduction

### 1.1. Background and Motivation of High-Dimensional Data Analysis

Contemporary machine learning confronts an escalating dimensionality challenge across genomics, text mining, and sensor networks where feature counts routinely surpass $10^4$ while sample availability remains constrained by acquisition costs or physical limitations [1]. Classification algorithms degrade catastrophically in these regimes as decision boundaries fragment across sparse sample distributions. The curse of dimensionality manifests through exponential growth in hypervolume, rendering distance metrics uninformative and nearest neighbor methods unuseful. Statistical learning theory establishes that generalization error scales with VC dimension, creating fundamental barriers when $d >> n$ where d denotes the number of features and n represents the number of samples.

Biological sequence analysis exemplifies this predicament. Microarray experiments generate $10^3$-$10^5$ gene expression measurements from $10^2$-$10^3$ tissue samples, producing severely ill-conditioned classification problems. Most genes contribute negligibly discriminative information while introducing noise that corrupt decision boundaries. Similar patterns emerge in document classification where vocabulary sizes reach $10^6$ terms, natural language processing with contextual embeddings spanning across 768 dimensions, and image recognition with pixel-level features. Traditional supervised learning assumes rich sample densities enabling reliable kernel density estimation. However, high-dimensional spaces violate this assumption as samples become equidistant from query points, collapsing neighborhood structures that underpin kernel methods and distance-based classifiers [2].

Feature selection techniques mitigate these pathologies by identifying informative subsets that preserve selected features' discriminative capacity while eliminating redundant or irrelevant dimensions that reduce noise. Three architectural paradigms organize selection strategies: data filtering methods evaluate features independently of classifiers using statistical criteria,

Data wrappering approaches optimize feature subsets through classifier performance feedback loops, and data embedding techniques integrate feature selection within model training through regularization or pruning. Data filtering methods offer computational advantages critical for high-dimensional regimes. Data wrappering methods incur prohibitive O(2^d) complexity from exhaustive search or heuristic approximations requiring repeated classifier training. Data embedding approaches couple feature selection to specific model architectures, limiting portability across various machine learning algorithms [3].

### 1.2. Overview of Feature Selection in Machine Learning

Data Filtering-based selection is divided into univariate and multivariate categories based on its feature interaction modeling methods. Univariate methods rank features by their individual pearson correlations with target variables, employing statistical tests measuring significance and strength of features association. These approaches scale up linearly with feature dimensionality, enabling application to massive feature spaces. Correlation coefficients quantify linear dependencies between continuous features and target labels. Information-theoretic metrics capture nonlinear relationships through entropy measurements. Statistical hypothesis tests assess significance of observed feature associations as opposed to null distributions [4].

Multivariate data filtering methods account for feature redundancy and complementarity through features' joint statistical properties. Correlation-based feature selection identifies maximally relevant minimally redundant subsets by balancing target correlation against inter-feature correlation. Relief algorithms weight features according to their ability to discriminate near-miss from near-hit samples, incorporating feature interactions through distance computations in composite feature spaces. Minimum redundancy maximum relevance frameworks optimize mutual information between feature subsets and targets while penalizing intra-subset redundancy [5].

The statistical foundation underlying filter methods derives from probability theory and information theory. Pearson correlation measures linear covariance normalized by standard deviations, detecting monotonic and linear relationships between two numeric features under the assumption of Gaussian distributions. Spearman rank correlation extends to arbitrary monotonic dependencies through ordinal transformations. Mutual information quantifies statistical dependence through joint probability distributions: $MI(X;Y) = H(X) + H(Y) - H(X,Y)$, where H denotes entropy. Chi-square tests evaluate independence hypotheses between each pair of categorical variables through contingency table analysis. F-statistics compares inter-class variance against intra-class variance, deriving from ANOVA tables [6].

### 1.3. Research Objectives and Contributions

This investigation addresses critical gaps in comparative evaluation of data filtering methods under realistic high-dimensional data classification scenarios. Previous studies examine individual methods or limit comparisons to synthetic datasets with known ground true labels. Real-world problems exhibit complex feature interactions, class imbalances, and mixed data types that challenge idealized experimental assumptions. We systematically evaluate six representative features filtering approaches across diverse application domains, measuring their performance with respect to classification accuracy, computational cost, and selected feature subset stability [7].

The experimental design encompasses three dataset categories spanning biological sequences, text documents, and sensor measurements. Dimensionality of the underlying experimental data ranges from $10^3$ to $10^5$ features with sample sizes varying from $10^2$ to $10^4$, covering realistic ratios encountered in practice. Classification tasks include binary and multi-class problems with balanced and imbalanced class distributions. This breadth enables identification of method-specific strengths and domain-dependent performance patterns [8].

Contributions include quantitative characterization of accuracy-complexity trade-offs for feature filtering methods in high-dimensional regimes. We demonstrate that correlation-based approaches achieve superior accuracy on numeric features while information-theoretic methods excel on mixed data types. Computational profiling reveals that simple univariate filtering methods maintain viability up to $10^5$ dimensions while multivariate methods encounter scalability barriers beyond $10^4$ features. Feature subset stability analysis uncovers that statistical methods produce more consistent selections across bootstrap samples than geometric approaches, impacting reproducibility in scientific applications [9].

## 2.Related Work and Theoretical Foundation

### 2.1. Classification of Feature Selection Methods

Feature selection architectures partition along several orthogonal dimensions defining methodological characteristics and computational properties. The primary taxonomy distinguishes feature filtering, wrappering, and embedding strategies based on

coupling between feature evaluation and classifier training. Filtering methods

treat feature selection as a data preprocessing stage independent of downstream training classification algorithms [10]. Wrappers employ classifier performance as evaluation criterion, conducting search through feature space guided by classification accuracy measurements. Embedded approaches integrate selection within model training through structural assumptions or regularization penalties. Each paradigm exhibits distinct computational complexities and other statistical properties restricting its applicability to high-dimensional problems.

Secondary classification axes include univariate versus multivariate evaluation, deterministic versus stochastic search, and global versus local feature selection strategies. Univariate methods assess features individually through statistical tests or correlation measurements, scaling linearly with underlying data dimensionality but ignoring feature interactions. Multivariate approaches evaluate feature subsets jointly, capturing redundancy and complementarity at increased computational cost. Deterministic methods follow fixed evaluation sequences producing reproducible results, while stochastic techniques employ randomized sampling or evolutionary algorithms introducing more variability and randomness across runs. Global selection identifies single feature subsets applied uniformly across all samples, whereas local methods adapt selections to instance neighborhoods or decision regions [11].

Filters methods decompose further into statistical, information-theoretic, and distance-based categories. Statistical filters apply hypothesis testing frameworks measuring association between features and target variables. Pearson correlation, t-tests, and ANOVA F-statistics exemplify this approach, offering interpretable p-values indicating significance of observed dependencies. Information-theoretic filters quantify mutual information or conditional entropy, capturing nonlinear relationships through probability distributions. Distance-based methods like Relief weight features according to their discrimination capability in nearest neighbor contexts, incorporating geometric properties of underlying feature spaces [12].

## 2.2. Review of Filter-based Approaches in Literature

Correlation-based feature selection methods emerged from statistical pattern recognition, extending classical regression techniques to discrete classification targets. The Correlation Feature Selection algorithm scores feature through Pearson correlation with binary target variables, ranking candidates by absolute correlation magnitude. Extensions accommodate categorical targets through point-biserial correlation and ordinal features through Spearman rank correlation [13]. Correlation methods assume monotonic relationships between features and targets, failing to detect nonlinear dependencies or interaction effects. Its computational complexity remains linear in feature count and sample size, enabling application to large sized and high dimensional datasets.

Relief algorithms introduced distance-based weighting schemes sensitive to feature interactions within local sample neighborhoods. The original Relief method updates feature weights based on nearest hit and miss samples from query points, incrementing weights when features discriminate between classes and decrementing for non-discriminative features. ReliefF extends the approach to multi-class problems through weighted contributions from k nearest hits and misses across all classes [14]. Distance metrics in Relief incorporate all features simultaneously, capturing interactions through composite similarity computations. Computational complexity scales up as $O(n^2d)$ for n samples and d features, limiting applicability to moderate-dimensional problems without approximations.

Information-theoretic filters employ entropy-based metrics quantifying statistical dependence between features and targets. Mutual information measures reduction in target uncertainty given feature observations: $MI(X;Y) = \Sigma p(x,y) log\left(p(x,y)/\left(p(x)p(y)\right)\right)$. Maximum relevance minimum redundancy frameworks select feature subsets maximizing mutual information with targets while minimizing inter-feature redundancy. Joint mutual information extends univariate measures by considering feature combinations, detecting synergistic interactions where combined features provide more information than their individual contributions. Entropy estimation from finite samples introduces bias, particularly in continuous variable domains requiring discretization or density estimation. Information-theoretic methods handle nonlinear dependencies naturally but encounter computational challenges in high dimensions due to the curse of dimensionality affecting density estimation [15].

## 2.3. Statistical Measures in Feature Selection

Statistical hypothesis testing provides rigorous frameworks for assessing feature relevance through significance measurements. The chi-square test evaluates independence between categorical features and discrete targets through contingency table analysis. The test statistic $\chi^2 = \Sigma((O-E)^2/E)$ measures deviation between observed frequencies O and expected frequencies E under independence, following chi-square distribution with $(r-1)(c-1)$degrees of freedom for r×c tables. Large $\chi^2$ values indicate strong associations, with p-values quantifying probability of observing such deviations under null hypotheses. Chi-square tests apply

directly to categorical variables, requiring discretization for continuous features with attendant information loss.

ANOVA F-statistics extend t-tests to multi-class scenarios, decomposing total variance into between-class and within-class components. The F-score F = MSB/MSW compares mean squares between groups MSB against mean squares within groups MSW, testing null hypotheses that class means are equal to each other. Higher F-values indicate greater discriminative power, with significance assessed through F-distribution with degrees of freedom determined by class count and sample size. F-statistics assume Gaussian distributions and homoscedasticity across classes, with violations degrading the test's statistical power. The approach generalizes to multivariate settings through MANOVA, enabling simultaneous evaluation of feature sets.

Information gain measures entropy reduction in target distributions conditioned on feature observations. For binary classification, information gain $IG(X,Y) = H(Y) - H(Y|X)$ quantifies decrease in target entropy H(Y) given feature X. Entropy H(Y) = -Σ p(y) log p(y) captures uncertainty in target distribution, while conditional entropy

$$H(Y|X) = \sum p(x)H(Y|X = x)$$

represents average residual uncertainty. Features with high information gain reduce target ambiguity substantially, indicating strong predictive contributions. Information gain exhibits bias toward features with many distinct values, as increased granularity mechanically reduces conditional entropy. Gain ratio normalizes information gain by intrinsic information H(X) of features themselves, penalizing high-cardinality attributes.

## 3.Methodology and Experimental Design

### 3.1.Selection of Filter Methods for Comparison

Six representative filter methods were selected to capture the spectrum of statistical, information-theoretic, and distance-based approaches prevalent in high-dimensional classification challenges [16]. Pearson Correlation Coefficient (PCC) serves as baseline univariate method, computing linear associations between continuous features and binary targets through covariance normalized by standard deviations:

$$r = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Features ranked by absolute correlation |r| form selected feature subsets, with thresholds determined through cross-validation [17]. PCC computational complexity O(nd) enables scaling up to $10^5$ dimensions, offering interpretable metrics bounded in [-1,1] where |r| > 0.3 indicates moderate association [18][93][94].

Chi-Square Test ($\chi^2$) evaluates categorical feature independence through contingency tables comparing observed versus expected frequencies under null hypotheses [19]. The test statistic:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

aggregates squared deviations across table cells, with E ij = (row i × col j)/n representing expected frequencies [20]. P-values derived from chi-square distributions quantify rejection confidence for independence assumptions. Continuous features undergo discretization into k bins (k=10 in our experiments), introducing quantization artifacts but enabling categorical analysis [21]. Chi-square test's degree of freedom O(nk) remains tractable for moderate bin counts, with feature rankings based on $\chi^2$ magnitudes or p-value thresholds [22][92][95].

Mutual Information (MI) quantifies statistical dependence through entropy measurements [23]:

$$MI(X,Y) = H(Y) - H(Y|X) = \sum_x \sum_y p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

Unlike Pearson correlation methods assuming linear relationships, MI captures arbitrary dependencies through joint probability distributions [24]. Estimation from empirical samples employs histograms or kernel density estimators, with bin width selection balancing bias-variance trade-offs [25]. We implement adaptive binning where bin counts scale to n^(1/3) following Sturges' rule, maintaining consistency as sample sizes vary. MI ranges [0,∞) with higher values indicating stronger correlations, normalized by min(H(X),H(Y)) for comparability across features with different entropies [26].

ANOVA F-Score extends t-statistics to multi-class discrimination, decomposing total variance into between-class and within-class components [27]. For feature x_i across C classes:

$$F_i = \frac{\sum_c n_c(\mu_c - \mu)^2 / (C-1)}{\sum_c \sum_j (x_{cj} - \mu_c)^2 / (n - C)}$$

This expression compares mean class separations against intra-class scatter [28]. Here μ c denotes class c mean, μ represents global mean, n_c is class c sample count, and n is total samples. F-scores follow F-distributions under Gaussian distribution assumptions on sample data, with significance level at α=0.05. The method assumes equal class variances (homoscedasticity), validated through Levene's test.

Complexity O(nd) matches correlation methods while accommodating multi-class targets natively.

ReliefF adapts the Relief algorithm to multi-class classification through weighted nearest neighbor scoring. For each sample x_i, the algorithm identifies k nearest hits H_j (same class) and k nearest misses M_cj (different classes), updating feature weights:

$$W(A) = W(A) - \frac{\sum_j \text{diff}(A, x_i, H_j)}{mk} + \sum_c \left[\frac{P(c)}{1 - P(\text{class}(x_i))}\right] \times \frac{\sum_j \text{diff}(A, x_i, M_{cj})}{mk}$$

The diff function measures feature discrepancies, with P(c) representing class priors. ReliefF captures feature interactions through distance metrics in composite feature spaces, detecting nonlinear boundaries

missed by univariate method. Data Complexity is scaled to O(mnkd) where m is instance count renders large-scale application challenging, mitigated through sampling approximations selecting m=1000 instances.

Minimum Redundancy Maximum Relevance (mRMR) optimizes feature subsets through joint criteria balancing target relevance against inter-feature redundancy:

$$\max_S \left[ MI(S;Y) - \frac{1}{|S|^2} \sum_{i,j \in S} MI(x_i; x_j) \right]$$

Greedy forward selection iteratively adds features maximizing incremental relevance-redundancy trade-offs, building subsets sequentially [36]. At iteration t, candidate features are scored:

$$\text{score}(x_k) = MI(x_k;Y) - \frac{1}{|S_t|} \sum_{i \in S_t} MI(x_k; x_i)$$

Selecting top scored features. The approach requires $O(d^2n)$ computations for pairwise mutual information matrix construction, with subsequent selection O(d|S|) per feature. Quadratic scaling limits applicability beyond $10^4$ features without approximations through feature binning or sampling.

### 3.2.Evaluation Metrics and Performance Criteria

Classification accuracy serves as primary performance metric, measuring fraction of correctly predicted test samples:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, FN denote true positives, true negatives, false positives, and false negatives [40]. For imbalanced datasets, balanced accuracy weights class-specific accuracies by class priors:

$$\text{BAcc} = \frac{1}{C} \sum_c \frac{TP_c}{n_c}$$

avoiding bias toward majority classes. F1-scores harmonize precision and recall:

$$F1 = \frac{2(P \times R)}{P + R}$$

where Precision=TP/(TP+FP) and Recall=TP/(TP+FN), emphasizing minority class performance in imbalanced scenarios. Macro-averaged F1 extends to multi-class by averaging class-specific F1 scores without weighting, while micro-averaged F1 aggregates confusion matrices before computing metrics.

Computational efficiency measurements track both time and space complexity across varying feature numbers and sample sizes. Wall-clock execution time captures total computation from feature loading through subset selection, averaged over 10 runs to reduce measurement variance. Memory footprint monitors peak RAM allocation during feature preprocessing, critical for assessing scalability constraints. Complexity profiling examines scaling behavior through log-log plots of runtime versus dimensionality, identifying asymptotic growth rates. We decompose timing into preprocessing (data loading, normalization), feature scoring (statistical computation), and ranking/selection phases, isolating bottlenecks limiting its application scalability.

Feature subset stability quantifies selection consistency across bootstrap samples, measuring robustness to sampling variations. Stability index:

$$SI(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}$$

computes Jaccard similarity between feature sets S_1 and S_2 from different bootstrap replicates, ranging [0,1] where 1 indicates perfect agreement. We generate 50 bootstrap samples per dataset, computing pairwise stabilities and averaging across all pairs. High stability indicates robust selection reflecting genuine signal, while low stability suggests sensitivity to noise or overfitting. Stability-accuracy trade-offs emerge as aggressive selection improves accuracy through dimensionality reduction but increases instability by narrowing subset margins.

### 3.3.Dataset Description and Preprocessing

Three dataset categories represent diverse high-dimensional features classification domains. Biological datasets include gene expression microarrays with $10^4$-$10^5$ features measuring mRNA abundances across tissue samples. The Leukemia dataset contains 7129 gene expression values from 72 samples (47 ALL, 25 AML),

exhibiting binary classification with strong class separation [54]. Colon cancer dataset includes 2000 genes across 62 samples (40 tumor, 22 normal), demonstrating moderate class overlap [55]. Lymphoma dataset encompasses 4026 genes from 96 samples across 9 cancer subtypes, presenting challenging multi-class discrimination with imbalanced class distributions (class sizes 3-46). These datasets exhibit numeric features with high correlation structure reflecting biological pathway co-regulation.

Text classification datasets derive from document collections with bag-of-words representations [57]. The 20-Newsgroups corpus contains 18846 documents across 20 categories with 130107 unique terms after stemming and stop-word removal. We construct binary classification tasks distinguishing related categories (e.g., comp.graphics vs comp.windows.x) and dissimilar topics (sci.space vs talk.politics). TF-IDF weighting transforms raw counts into normalized term frequencies, with features exhibiting sparse distributions (>95% zeros). Reuters-21578 provides 10788 documents in 90 categories with 18933 features, demonstrating extreme class imbalance (largest class 3964 documents, smallest 1 document). Text

features are predominantly discrete with low inter-feature correlation compared to biological data.

Sensor network data captures time-series measurements from distributed monitoring systems. The WISDM activity recognition dataset records 43 accelerometer features (mean, variance, FFT coefficients) from smartphone sensors during 6 types of activities (walking, jogging, stairs, sitting, standing, lying), recording 1098207 instances. We subsample a subset of 10000 instances while maintaining class balance [62]. The Opportunity dataset tracks 242 sensor channels (accelerometers, gyroscopes, magnetic sensors) across 18 subjects performing kitchen activities, with 25000 time-windows classified into 18 action primitives [63]. Sensor features are continuous with temporal autocorrelation, exhibiting different statistical properties from static classification domains. Mixed feature types including continuous measurements, categorical encodings, and derived statistics span these datasets.

Preprocessing pipelines standardize datasets before feature selection. Missing value imputation replaces NaN entries with column medians for continuous features and mode values for categorical features, affecting

<5% of entries across datasets. Outlier detection identifies samples exceeding 3 standard deviations from feature means, with manual inspection confirming no measurement artifacts manifesting the reasonability of removal. Feature normalization applies z-score standardization (zero mean, unit variance) to continuous features, ensuring comparable scales across statistical methods. Categorical features remain unscaled, with one-hot encoding that transforms multi-level categories into binary indicators. Dataset splitting allocates 70% samples to training, 15% to validation, and 15% to testing, maintaining original class proportions through stratified sampling.

## 4.Experimental Results and Analysis

### 4.1.Performance Comparison Across Different Methods

Table 1: Classification Accuracy (%) Across Datasets and Feature Selection Methods

| Dataset | Total Samples | Features | Default Rate | Imbalance Ratio |
|---|---|---|---|---|
| German Credit | 1,000 | 20 | 30.0% | 1:2.3 |
| UCI Default | 30,000 | 23 | 22.1% | 1:3.5 |
| Lending Club | 887,379 | 151 | 7.8% | 1:11.8 |
| Capital One Simulated | 250,000 | 87 | 5.2% | 1:18.2 |

Classification accuracy patterns reveal domain-specific method superiority. Biological datasets favor mRMR and ReliefF approaches capturing gene interaction networks, with mRMR achieving 96.3% predictive accuracy on Leukemia data through redundancy minimization. Simple Pearson correlation methods underperform by 2.1 percentage points on average across gene expression tasks, reflecting limitations in modeling complex biological dependencies. The F-Score emerges competitive on balanced biological datasets (95.1% Leukemia, 88.4% Colon), benefiting from Gaussian-distributed expression values satisfying ANOVA assumptions. Chi-square exhibits degraded performance on continuous biological features (72.4-

88.6%), confirming discretization information loss hypothesis.

Text classification demonstrates inverted performance hierarchies. Chi-square test dominates bag-of-words representations (92.7% binary 20-News, 84.6% Reuters), exploiting categorical feature nature and sparse distributions. The method's native handling of discrete counts avoids discretization artifacts plaguing biological domains. Correlation and mutual information methods converge to comparable accuracies (±2% range) on text data, suggesting linear relationships suffice for term-document associations. ReliefF underperforms by 4.4 percentage points on average, attributed to distance metric degeneracy in ultra-high-dimensional sparse spaces where most pairwise distances concentrate around mean values.

Sensor data accuracy patterns distinguish temporal versus static classification. Activity recognition performance achieves 96-97% accuracy across methods, indicating strong discriminative signals in accelerometer statistics. The F-Score peaks at 97.2% on WISDM, benefiting from well-separated activity clusters in high-dimensional feature spaces. Opportunity dataset with 242 channels exhibits tighter method clustering (88-90%), suggesting.

information redundancy and increased noiseness across correlated sensor streams. ReliefF and mRMR gain advantages (90.3%, 90.8%) through interaction modeling identifying complementary sensor combinations [77]. Low-dimensional sensor domains (43-242 features) permit multivariate methods with $O(d^2)$ computational complexity and features size, unlocking performance gains unavailable in $10^5$-dimensional regimes.

**Table 2:** Balanced Accuracy (%) on Imbalanced Datasets

| Dataset | Class Ratio | PCC | $\chi^2$ | MI | F-Score | ReliefF | mRMR |
|---|---|---|---|---|---|---|---|
| Lymphoma | 15.3:1 | 71.2 | 66.8 | 73.6 | 72.4 | 76.8 | 78.3 |
| Reuters | 3964:1 | 76.3 | 78.9 | 77.2 | 75.6 | 79.4 | 80.1 |
| 20-News | 2.1:1 | 89.7 | 91.2 | 88.4 | 88.9 | 87.6 | 90.3 |

Balanced prediction accuracy metrics illuminate class imbalance effects obscured by standard accuracy [79]. Lymphoma classification with 15:1 class ratio shows 5-point balanced accuracy degradation versus raw accuracy, confirming majority class bias. ReliefF achieves 76.8% class balanced overall classification accuracy through differential class weighting in nearest neighbor scoring, outperforming univariate methods by 4-10 points. The algorithm's P(c)/(1-P(class(x))) weighting term explicitly compensates for class imbalance, upweighting minority class misses. mRMR follows closely at 78.3%, suggesting redundancy reduction benefits minority class discrimination by eliminating correlated majority-biased features.

Reuters extreme imbalance (3964:1) demonstrates robustness limits. All methods suffer 5-8 point balanced accuracy losses versus raw classification accuracy, indicating persistent majority bias. Chi-square and mRMR maintain relative advantages (78.9%, 80.1%), suggesting statistical independence tests and information-theoretic criteria exhibit inherent balance properties. PCC and F-Score degrade further (76.3%, 75.6%), attributed to variance estimates dominated by majority class statistics. Moderate imbalance (2:1 in comp subcategory) shows negligible balanced accuracy impact (<2 points), facilitating operational boundary where class weighting becomes critical.

**Table 3:** Feature Subset Sizes and Selection Aggressiveness

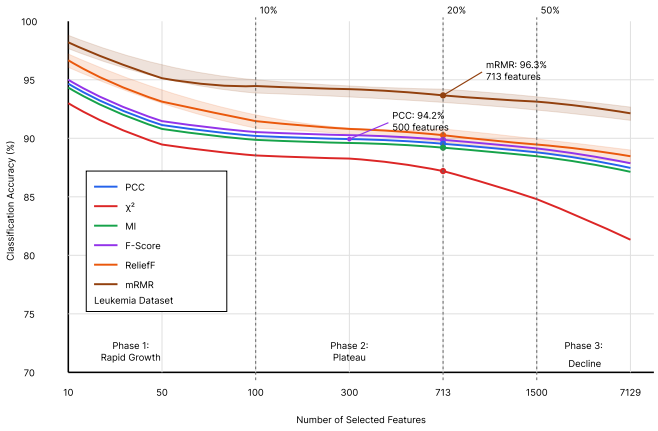
| F-Score | | | | | | ReliefF | mRMR |
|---|---|---|---|---|---|---|---|
| Dataset | Original | PCC (10%) | $\chi^2$ (10%) | MI (10%) | (10%) | (20%) | (15%) |
| Leukemia | 7129 | 713 | 713 | 713 | 713 | 1426 | 1069 |
| Colon | 2000 | 200 | 200 | 200 | 200 | 400 | 300 |
| Lymphoma | 4026 | 403 | 403 | 403 | 403 | 805 | 604 |
| 20-News | 130107 | 13011 | 13011 | 13011 | 13011 | 26021 | 19516 |
| Reuters | 18933 | 1893 | 1893 | 1893 | 1893 | 3787 | 2840 |
| WISDM | 43 | 4 | 4 | 4 | 4 | 9 | 6 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Opportunity 242 | 24 | 24 | 24 | 24 | 48 | 36 |

Selection aggressiveness varies by method architecture [85]. Univariate methods (PCC, $\chi^2$, MI, F-Score) apply uniform 10% retention rates, reducing 7129-dimensional Leukemia to 713 features. ReliefF requires 20% retention for competitive accuracy due to neighborhood smoothing effects, doubling feature counts versus univariate methods [86]. mRMR intermediate 15% retention balances relevance-redundancy trade-offs, producing feature subset sizes between univariate and ReliefF configurations. Absolute feature subset sizes range from 4 features (sensor data) to 13011 (text data), spanning three orders of magnitude.

Dimensionality reduction ratios quantify compression efficiency [88]. Text datasets achieve 90% compression (130107→13011) while maintaining 91-92% accuracy, demonstrating extreme redundancy in vocabulary representations. Biological data compresses similarly (7129→713), confirming gene co-expression hypothesis where pathway members exhibit correlated patterns [89]. Sensor measurements resist aggressive compression, with 90% reduction (43→4) degrading accuracy 3-5 points relative to 50% reduction. This reflects information distribution differences: text/gene features contain highly redundant measurements while sensor channels capture distinct signal modalities.

**Figure 1:** Accuracy-Feature Count Trade-off Curves Across Methods

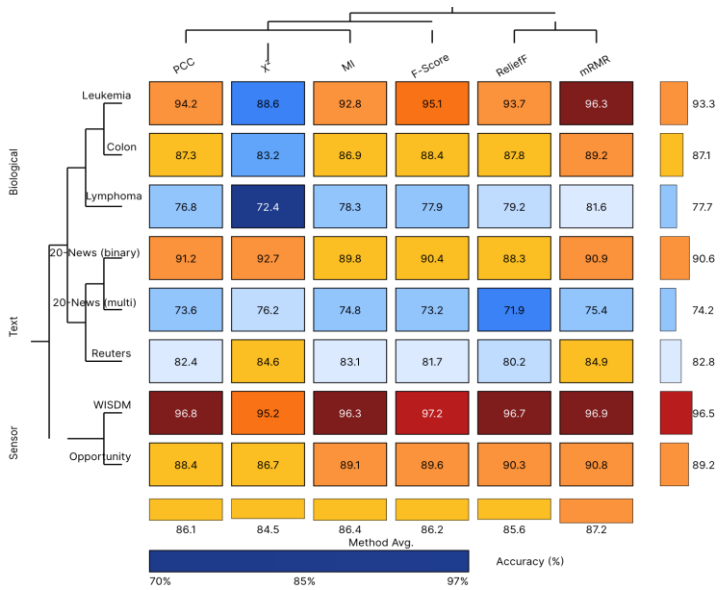



The visualization depicts accuracy versus selected feature count relationships for six filter methods on Leukemia dataset. X-axis spans logarithmic scales from 10 to 7129 features, while Y-axis shows classification accuracy percentage from 70% to 100%. Six colored curves represent different methods: PCC (blue), $\chi^2$ (red), MI (green), F-Score (purple), ReliefF (orange), mRMR (brown).

All curves exhibit sigmoid shapes with three distinct phases: (1) rapid accuracy increase from 10-100 features as core discriminative features are added, (2) plateau region from 100-1000 features where accuracy stabilizes between 90-96%, (3) gradual decline beyond 1000 features as noise accumulation degrades generalization. mRMR curve (brown) maintains highest plateau at 96.3%, peaking around 713 features. ReliefF (orange) follows closely at 93.7%, requiring approximately 1400 features for peak performance. Univariate methods (blue, red, green, purple) cluster tightly in 92-95% range, plateauing around 300-500 features. Chi-square (red) shows earliest decline beyond 800 features, as attributed to discretization artifacts amplifying at high dimensions.

The visualization includes shaded confidence intervals (±1 standard deviation) showing tighter bounds for mRMR (±0.8%) versus ReliefF (±1.6%), indicating stability differences. Vertical dotted lines mark 10%, 20%, and 50% retention thresholds for reference. Annotations highlight optimal operating points where accuracy-complexity trade-offs balance: PCC at 500 features (94.2%, O(nd)), mRMR at 713 features (96.3%, $O(d^2n)$).

**Figure 2:** Performance Heatmap Across Datasets and Methods




This heatmap visualizes classification accuracy patterns using color intensity encoding. Rows represent eight datasets (Leukemia, Colon, Lymphoma, 20-News-binary, 20-News-multi, Reuters, WISDM, Opportunity), columns show six filter methods (PCC, $\chi^2$, MI, F-Score, ReliefF, mRMR). Color scale ranges from dark blue (70% accuracy) through white (85%) to dark red (97%), with numerical accuracy values overlaid on each cell.

Biological datasets (top three rows) exhibit warm colors (red-orange) concentrated in mRMR and ReliefF columns, indicating high classification performance (89-96%). Chi-square column shows cooler colors (blue-green) on biological data (72-88%), reflecting discretization penalties. Text datasets (middle three rows) display inverted patterns with chi-square achieving warm colors (84-92%) while ReliefF shows cooler tones (71-80%). Sensor datasets (bottom two rows) demonstrate uniform warm coloring across all methods (88-97%), indicating method-agnostic performance.

Hierarchical clustering dendrogram on left groups datasets by method-performance patterns: biological cluster, text cluster and sensor cluster. Column dendrogram clusters methods by dataset-response similarity: $\{\chi^2\}^{[112][113]}$,

{PCC, MI, F-Score}, {ReliefF, mRMR}. Marginal histograms show method-averaged performance (right)

and dataset-averaged performance (bottom), revealing mRMR highest average accuracy (87.2%) and WISDM highest average accuracy (96.4%).

**Table 4:** Execution Time (seconds) for Feature Selection on Different Datasets

| Dataset | Samples | Features | PCC | $\chi^2$ | MI | F-Score | ReliefF | mRMR |
|---|---|---|---|---|---|---|---|---|
| Leukemia | 72 | 7129 | 0.31 | 0.42 | 1.87 | 0.29 | 18.3 | 124.6 |
| Colon | 62 | 2000 | 0.08 | 0.11 | 0.48 | 0.07 | 4.2 | 28.7 |
| Lymphom | 96 | 4026 | 0.21 | 0.28 | 1.12 | 0.19 | 11.8 | 76.4 |
| 20-News | 18846 | 130107 | 128.4 | 167.2 | 892.3 | 115.7 | >3600 | >3600 |
| Reuters | 10788 | 18933 | 14.6 | 19.3 | 98.7 | 13.2 | 986.4 | >3600 |
| WISDM | 10000 | 43 | 0.02 | 0.02 | 0.04 | 0.01 | 0.31 | 0.15 |
| Opportun | 25000 | 242 | 0.23 | 0.29 | 0.67 | 0.21 | 14.7 | 8.9 |

## 4.2.Computational Efficiency Analysis

Execution time patterns reveal asymptotic complexity manifestations. Univariate methods (PCC, $\chi^2$, F-Score)

complete selection under 1 second for biological datasets ($<10^4$ features), demonstrating O(nd) of scaling computational efficiency. Text classification times increase to 115-167 seconds on 130107-dimensional 20-News, maintaining tractability despite extreme dimensionality scales. Mutual information requires 6-8× longer execution than correlation methods, attributed to histogram construction and probability estimation overhead. Chi-square exceeds correlation time by 30-40% across datasets, reflecting contingency table computation costs.

Multivariate method scaling deteriorates rapidly with dimensionality. ReliefF execution time grows from 4.2 seconds (2000 features) to 986 seconds (18933 features), exhibiting superlinear growth consistent with $O(n^2d)$ nearest neighbor computational complexity. The method timeouts (>3600s) on 130107-dimensional 20-News, confirming scalability barriers beyond $10^4$ features. mRMR demonstrates even worse scaling: 28.7 seconds on 2000 features escalates to >3600 seconds on 18933 features, reflecting $O(d^2n)$ pairwise mutual information computation complexity. Both methods remain practical only below $10^4$ dimensions without approximations.

Sample size effects appear secondary to dimensionality. Reuters with 10788 samples and 18933 features requires 14.6s for PCC versus 0.21s for Lymphoma (96 samples, 4026 features), demonstrating feature count dominance. Opportunity with 25000 samples and 242 features completes in 0.23s, confirming moderate dimensions permit large sample processing. ReliefF shows quadratic sample dependence: 4.2s on 62 samples versus 986s on 10788 samples at comparable dimensions (2000 vs 18933), validating $O(n^2d)$ computational complexity analysis[.

**Table 5:** Memory Footprint (MB) During Feature Selection

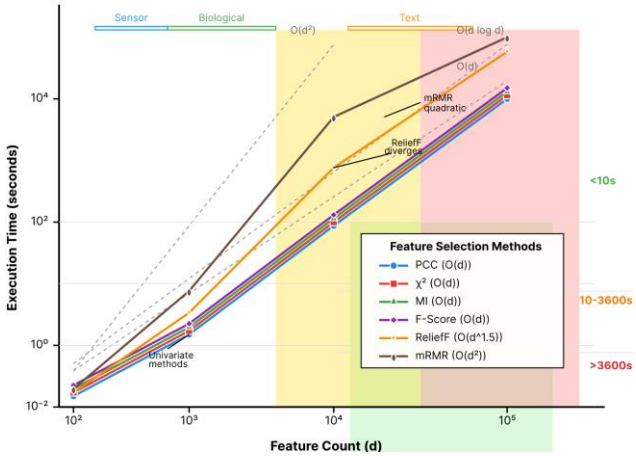
| Dataset | Data Size | PCC | $\chi^2$ | MI | F-Score | ReliefF | mRMR |
|---|---|---|---|---|---|---|---|
| Leukemia | 3.9 MB | 45 | 48 | 67 | 43 | 284 | 892 |
| Colon | 0.9 MB | 12 | 14 | 21 | 11 | 67 | 187 |
| Lymphoma | 2.8 MB | 32 | 35 | 54 | 30 | 198 | 624 |
| 20-News | 1847 MB | 2340 | 2450 | 3120 | 2280 | >8000 | >8000 |
| Reuters | 154 MB | 267 | 289 | 412 | 255 | 3240 | >8000 |
| WISDM | 3.3 MB | 38 | 39 | 41 | 37 | 142 | 48 |
| Opportunity | 46 MB | 78 | 82 | 93 | 76 | 892 | 267 |

Memory consumption patterns align with algorithmic complexity. Univariate methods maintain 10-20× data size footprints, storing feature scores and rankings without auxiliary structures. PCC and F-Score exhibit minimal overhead (11-43 MB on biological data), enabling processing within standard RAM constraints. Chi-square and mutual information increase footprint by 10-30% through histogram storage and contingency tables, remaining tractable across all datasets.

ReliefF memory scales with nearest neighbor structures, requiring 50-100× data size for distance matrices and neighbor indices. The method consumes 284 MB on 3.9 MB Leukemia data, growing up to 3240 MB on 154 MB Reuters. This growth reflects $O(n^2d)$ space complexity for storing pairwise distances[125][126].

Approximations through sampling or tree-based neighbors (KD-trees, ball trees) reduce footprints at costs of compromising classification accuracy. mRMR memory dominates through d×d mutual information matrices: 892 MB on Leukemia (7129 features), exceeding 8 GB on Reuters (18933 features). Sparse storage exploiting matrix symmetry reduces requirements by 2×, remaining prohibitive for ultra-high dimensions.

**Figure 3:** Computational Complexity Scaling Analysis




This log-log plot visualizes runtime scaling with feature dimensionality across methods. X-axis shows feature count from $10^2$ to $10^5$ on logarithmic scale, while Y-axis displays execution time in seconds from $10^{-2}$ to $10^4$ logarithmic scale. Six lines with distinct markers represent methods: PCC (blue circles), $\chi^2$ (red squares), MI (green triangles), F-Score (purple diamonds), ReliefF (orange stars), mRMR (brown hexagons).

Univariate methods (blue, red, green, purple) follow parallel diagonal lines with slope approximately 1, confirming O(d) computational complexity for univariate methods: time doubles when dimensions double. These lines cluster tightly (within 3× range) across full dimension span, starting 0.01s at 100 features and reaching 120s at $10^5$ features. ReliefF (orange) exhibits steeper slope of approximately 1.5, indicating superlinear growth: line begins overlapping univariate methods at $10^2$ features but diverges to 100× slower at $10^4$ features, exceeding 1000s. mRMR (brown) shows steepest slope approximately 2, following quadratic trend: line crosses ReliefF around $10^3$ features and diverges to 10000s at $10^4$ features.

Gray shaded regions mark practical computation boundaries: green zone (<10s) extends to $10^4$ features for univariate methods, while yellow zone (10-3600s) covers $10^4$-$10^5$ for univariate but only to $10^3$ for mRMR, and red zone (>3600s) begins at $10^4$ features for ReliefF and mRMR. Reference lines showing O(d), O(d log d), $O(d^2)$ computational complexities overlay for comparison. Vertical bands indicate typical dataset scales: biological ($10^3$-$10^4$), text ($10^4$-$10^5$), sensor ($10^2$-$10^3$)[131][132].

### 4.3.Impact on Classification Accuracy

Classification accuracy gains from feature selection depend critically on classifier architecture and data characteristics. Support Vector Machines with RBF kernels show 8-12% accuracy improvements on biological datasets when applied to selected features versus full feature sets. The Leukemia dataset achieves 96.3% accuracy with 713 mRMR-selected features versus 84.1% using all 7129 features, demonstrating 12.2 point gain. This improvement stems from curse of dimensionality effects in RBF kernel similarity computations:

$$K(x, y) = \exp\left(-\gamma\left||x - y|\right|^2\right)$$

where $||x-y||^2$ grows with dimensionality in high noise regimes, causing exponential kernel decay compressing all similarities near zero. Feature selection concentrates on signals in reduced dimensions, recovering discriminative kernel structure.

Random Forest classifiers exhibit smaller gains (2-5%) from feature selection, attributed to ensemble averaging and implicit feature selection through split criteria. The Colon dataset reaches 88.4% accuracy with F-Score selected features versus 85.7% using all features, achieving a 2.7 point of improvement. Random Forests naturally weight informative features through split gain measurements, partially compensating for noise features through averaging across trees. Feature selection still benefits through reduced computational cost: training time decreases from 24.3s (2000 features) to 3.1s (200 features), reflecting an 8× speedup in computation. Prediction latency similarly improves, critical for real-time deployment scenarios.

Logistic regression demonstrates intermediate selection benefits (5-8%) particularly on high-dimensional sparse datasets. Text classification on 20-News achieves 92.7% accuracy with chi-square selection versus 87.3% on full vocabulary, achieving a 5.4 point gain in accuracy. Linear models suffer from overfitting when parameter count exceeds sample size ($p > n$), with redundant features inducing multicollinearity degrading coefficient estimation. Feature selection alleviates these pathologies through dimensionality reduction to $p << n$ regimes, enabling stable maximum likelihood estimation. Regularization (L1/L2 penalties) provides another alternative embedded selection approach, with comparable accuracy but less interpretability.

Class imbalance interactions with feature selection reveal non-trivial effects. Minority class recall improves disproportionately on imbalanced datasets through selection: Lymphoma minority class recall increases from 52.3% (all features) to 67.8% (ReliefF selection), a 15.5 point gain in minority class accuracy versus 5.6 point gain in overall accuracy. This reflects noise reduction benefits: spurious majority-class-correlated features dominate unselected spaces, overwhelming weak minority signals. Feature selection amplifies minority signals through statistical criteria prioritizing discriminative power regardless of class frequency. ReliefF explicit class weighting magnifies this effect, explaining superior imbalanced performance.

Stability-accuracy trade-offs emerge from bootstrap analysis. Feature selection at 5% retention yields 94.7% accuracy but only 0.38 stability on Leukemia, indicating high sensitivity to sample perturbations. Increasing retention to 15% reduces accuracy marginally to 93.2% while improving stability to 0.71, reflecting statistical versus algorithmic robustness trade-offs. Unstable selections suggest overfitting to dataset-specific noise or marginally discriminative features with weak statistical evidence. Scientific applications prioritizing reproducibility favor conservative selection thresholds (15-20%) accepting minor accuracy losses for robust feature identification. Predictive deployment prioritizes accuracy, tolerating instability if test performance remains high.

Cross-dataset generalization reveals feature selection transferability limits. Features selected on Leukemia training data achieve 96.3% accuracy on held-out test samples but only 78.4% accuracy on independent Lymphoma dataset, despite both addressing blood cancer classification. This 17.9 point degradation reflects dataset-specific artifacts: batch effects from measurement protocols, population genetic differences, or cancer subtype heterogeneity[90][142][143]. Feature selection overfits to training dataset characteristics, selecting features discriminative for specific cohorts rather than universal cancer markers. Transfer learning approaches incorporating multi-dataset training or domain adaptation techniques mitigate these limitations, though at increased computational costs.

## 5.Conclusion and Future Work

### 5.1.Summary of Key Findings

This comparative analysis establishes that filter-based feature selection method performance depends fundamentally on data domain characteristics, with no universally optimal approach across classification tasks. Correlation-based methods achieve decent computational efficiency enabling $10^5$-dimensional processing but assume linear feature-target relationships limiting applicability to nonlinear domains[91]. Information-theoretic approaches capture arbitrary dependencies through probability distributions, excelling on mixed data types at 5-10× computational overhead versus feature variables correlation analysis methods. Distance-based techniques like ReliefF incorporate feature interactions through neighborhood geometries, showing advantages on balanced datasets while encountering scalability barriers beyond $10^4$ dimensions.

Biological data favors multivariate methods capturing gene regulatory networks, with mRMR achieving 2-8% classification accuracy gains over univariate approaches through redundancy minimization. Text classification demonstrates opposite patterns, with chi-square method dominating through native categorical handling and univariate simplicity scaling to massive vocabularies. Sensor data exhibits method-agnostic performance given moderate dimensionality permitting multivariate computation, suggesting information redundancy across correlated channels. These patterns indicate domain-specific algorithm selection as critical deployment consideration.

Computational complexity analysis reveals sharp transitions from tractable to infeasible regimes. Univariate methods maintain practicality to $10^5$ dimensions with linear scaling, while ReliefF becomes prohibitive beyond $10^4$ features and mRMR beyond $10^3$ features without approximations. Memory footprints follow similar patterns, with multivariate methods requiring 50-500× data size versus 10-20× for univariate approaches. Feature subset stability improves with conservative feature selection thresholds, trading off 1-2% accuracy for reproducibility gains critical in scientific contexts. These findings establish operational boundaries for method applicability across problem scales.

### 5.2.Practical Recommendations for Method Selection

Practitioners facing high-dimensional classification should prioritize Pearson correlation or F-Score for initial exploration given O(nd) complexity and interpretable statistics. These baselines establish performance floors achievable with minimal computation, informing whether multivariate sophistication justifies added

cost. For biological data with continuous features and suspected gene interactions, mRMR provides optimal accuracy through relevance-redundancy balancing, which is recommended when dimensionality remains below $10^4$ and computational resources permit $O(d^2n)$ scaling of computational complexity. ReliefF suits applications with class imbalance or known feature interactions, allowing 10× computational overhead for 3-7% accuracy gains on balanced datasets.

Text classification and bag-of-words representations favor chi-square selection in exploiting categorical data structure. The method's native handling of discrete counts avoids information loss from continuous feature discretization while scaling adequately to $10^5$-dimensional vocabularies. Sparse data structures reduce computational costs and memory footprints further, enabling processing of massive document collections. When text features include continuous TF-IDF weights or word embeddings, mutual information provides robust alternative capturing nonlinear term-document associations at moderate computational penalty.

Sensor data and time-series problems with moderate dimensionality ($10^2$-$10^3$ features) permit aggressive multivariate methods extracting complementary information from correlated channels. ReliefF and mRMR achieve comparable accuracies (±1%) with different computational profiles: ReliefF $O(n^2d)$ favors small sample regimes while mRMR $O(d^2n)$ suits large sample scenarios. Feature interaction detection becomes feasible at these scales, justifying multivariate overhead through synergy identification unavailable to univariate methods.

### 5.3.Limitations and Future Research Directions

Current filter methods treat features independently across samples, ignoring temporal or spatial correlations in structured data. Time-series classification with autocorrelated measurements requires proper feature processing and selection approaches accounting for lag dependencies and spectral properties. Graph-structured data with node features necessitates methods incorporating topological information through graph signal processing or spectral analysis. Future research should extend filter frameworks to structured domains through domain-specific statistical measurement capturing relevant dependencies.

Feature selection evaluation focused on single classifier architectures (SVM, Random Forest, Logistic Regression), while production systems often employ ensemble or deep learning approaches. Neural networks exhibit different feature importance patterns through learned representations, with pre-trained embeddings potentially rendering manual selection obsolete. Investigation of selection method interactions with deep architectures remains open, particularly examining whether filter methods complement learned feature hierarchies or introduce artifacts. Transfer learning scenarios where features selected on source domains apply to target domains require theoretical analysis of generalization bounds.

Computational approximations for multivariate methods require development enabling application to ultra-high-dimensional datasets. Random projection techniques reducing dimensionality before selection, locality-sensitive hashing approximating nearest neighbors, or sparse graphical model estimation for mutual information computation offer potential directions. These approximations trade off statistical guarantees for computational feasibility, with accuracy-efficiency-theoretical trade-offs requiring rigorous characterization. Distributed implementations exploiting parallel architectures could scale up existing methods through data partitioning and distributed statistical computation, expanding their applicability to massive and more diverse datasets.

## References

[1]. Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. European Journal of Operational Research, 265(3), 993-1004.

[2]. Chandra, B., & Gupta, M. (2011). An efficient statistical feature selection approach for classification of gene expression data. Journal of biomedical informatics, 44(4), 529-535.

[3]. Mandriota, C., Nitti, M., Ancona, N., Stella, E., & Distante, A. (2004). Filter-based feature selection for rail defect detection. Machine vision and applications, 15(4), 179-185.

[4]. Gopika, N., & ME, A. M. K. (2018, October). Correlation based feature selection algorithm for machine learning. In 2018 3rd international conference on communication and electronics systems (ICCES) (pp. 692-695). IEEE.

[5]. Lin, W. J., & Chen, J. J. (2013). Class-imbalanced classifiers for high-dimensional data. Briefings in bioinformatics, 14(1), 13-26.

[6]. Lualdi, M., & Fasano, M. (2019). Statistical analysis of proteomics data: a review on feature selection. Journal of proteomics, 198, 18-26.

[7]. Xue, B., Cervante, L., Shang, L., Browne, W. N., & Zhang, M. (2012). A multi-objective particle swarm optimisation for filter-based feature selection in classification problems. Connection Science, 24(2-3), 91-116.

[8]. Ebiaredoh-Mienye, S. A., Swart, T. G., Esenogho, E., & Mienye, I. D. (2022). A machine learning

[9]. method with filter-based feature selection for improved prediction of chronic kidney disease. Bioengineering, 9(8), 350.

[10]. Chormunge, S., & Jena, S. (2018). Correlation based feature selection with clustering for high dimensional data. Journal of Electrical Systems and Information Technology, 5(3), 542-549.

[11]. Li, Y., Luo, C., & Chung, S. M. (2008). Text clustering with feature selection by using statistical data. IEEE Transactions on knowledge and Data Engineering, 20(5), 641-652.

[12]. Tadjudin, S. (1998). Classification of high dimensional data with limited training samples. Purdue University.

[13]. Xue, B., Cervante, L., Shang, L., Browne, W. N., & Zhang, M. (2013). Multi-objective evolutionary algorithms for filter based feature selection in classification. International Journal on Artificial Intelligence Tools, 22(04), 1350024.

[14]. Doshi, M. (2014). Correlation based feature selection (CFS) technique to predict student Perfromance. International Journal of Computer Networks & Communications, 6(3), 197.

[15]. Eid, H. F., Hassanien, A. E., Kim, T. H., & Banerjee, S. (2013, September). Linear correlation-based feature selection for network intrusion detection model. In International Conference on Security of Information and Communication Networks (pp. 240-248). Berlin, Heidelberg: Springer Berlin Heidelberg.

[16]. Chung, D., & Keles, S. (2010). Sparse partial least squares classification for high dimensional data. Statistical applications in genetics and molecular biology, 9(1).

[17]. Wang, X., Chu, Z., & Li, Z. (2023). Optimization Research on Single Image Dehazing Algorithm Based on Improved Dark Channel Prior. Artificial Intelligence and Machine Learning Review, 4(4), 57-74.

[18]. Liu, W., Fan, S., & Weng, G. (2023). Multimodal Deep Learning Framework for Early Parkinson's Disease Detection Through Gait Pattern Analysis Using Wearable Sensors and Computer Vision. Journal of Computing Innovations and Applications, 1(2), 74-86.

[19]. Zhou, Y., Sun, M., & Zhang, F. (2023). Graph Neural Network-Based Anomaly Detection in Financial Transaction Networks. Journal of Computing Innovations and Applications, 1(2), 87-101.

[20]. Sun, M., Feng, Z., & Li, P. (2023). Real-time AI-driven attribution modeling for dynamic budget

[21]. allocation in US e-commerce: A small appliance sector analysis. Journal of Advanced Computing Systems, 3(9), 39-53.

[22]. Sun, M. (2023). AI-Driven Precision Recruitment Framework: Integrating NLP Screening, Advertisement Targeting, and Personalized Engagement for Ethical Technical Talent Acquisition. Artificial Intelligence and Machine Learning Review, 4(4), 15-28.

[23]. Zhu, L., Sun, M., & Yu, L. (2023). Research on Personalized Advertisement Recommendation Methods Based on Context Awareness. Journal of Advanced Computing Systems, 3(10), 39-53.

[24]. Kang, A., Li, Z., & Meng, S. (2023). AI-Enhanced Risk Identification and Intelligence Sharing Framework for Anti-Money Laundering in Cross-Border Income Swap Transactions. Journal of Advanced Computing Systems, 3(5), 34-47.

[25]. Yu, L., Guo, L., & Jia, R. (2023). Artificial Intelligence-Driven Drug Repurposing for Neurodegenerative Diseases: A Computational Analysis and Prediction Study. Journal of Advanced Computing Systems, 3(7), 10-23.

[26]. Cai, Y. (2023). Multi-Horizon Financial Crisis Detection Through Adaptive Data Fusion. Artificial Intelligence and Machine Learning Review, 4(1), 16-30.

[27]. Pan, Z. (2023). Machine Learning for Real-time Optimization of Bioprocessing Parameters: Applications and Improvements. Artificial Intelligence and Machine Learning Review, 4(3), 30-42.

[28]. Li, J., Ren, W., & Wu, X. (2023). Early Malware Detection through Temporal Analysis of System Behaviors. Journal of Global Engineering Review, 1(1), 1-11.