2025 IEEE 6th India Council International Subsections Conference (INDISCON) | 979-8-3315-1504-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/INDISCON66021.2025.11251795

A Hybrid Mutual Information-Based Feature Selection Framework with Redundancy Reduction through Clustering and Voting

Anubha Prajapati
Department of Computer Science and Applications
Doctor Harisingh Gour Vishwavidyalaya, Sagar
Madhya Pradesh, India
anubhaprajapati2@gmail.com

Sounak Sadhukhan
School of Computer Science Engineering and Technology
Bennett University, Greater Noida
Uttar Pradesh, India
sounaks.cse@gmail.com

Pangambam Sendash Singh
Department of Computer Science and Applications
Doctor Harisingh Gour Vishwavidyalaya, Sagar
Madhya Pradesh, India
sendashpangambam@gmail.com

Abstract—Feature selection takes a very pertinent role in machine learning by enhancing model performance through dimensionality reduction, preventing overfitting, and increasing interpretability. In this current work, we present an innovative feature selection method that integrates Feature-Label Mutual Information (FLMI) and Feature-Feature Mutual Information (FFMI) with clustering and voting techniques. The effectiveness of the proposed feature selection method is demonstrated on eight different datasets across various domains and feature types. The proposed approach is further compared with four established feature selection methods based on Mutual Information, Minimum Redundancy Maximum Relevance, ReliefF and Chi-Squared Test. Using Random Forest as the classifier, we analyze performance using accuracy, precision, recall, and F_1-score. Experimentation results have validated the consistent superiority of the proposed method over the other methods in terms of different classification metrics. Although minor performance drops were observed for certain datasets, the method achieves a balanced trade-off between dimensionality reduction and model effectiveness. These results suggest that the proposed framework can significantly enhance classification performance across various datasets.

Index Terms—Feature Selection, Mutual Information, Clustering, Voting Mechanism, Machine Learning

I. Introduction

In today's machine learning era, especially with datasets that contain numerous features, selecting the most relevant ones is essential. Effective feature selection not only contributes to the improvement of the accuracy of prediction algorithms and at the same time improves interpretability and reduces computational complexity [1], [2]. The primary goal is to determine a reduced feature set that carry the most significant information regarding the target/dependent variable, while reducing the unnecessary redundancy among them. This ensures that the resulting model remains efficient, avoiding the inclusion of irrelevant or repetitive data.

Mutual Information (MI) stands as a commonly employed criterion for feature selection because of its capacity to identify both linear and non-linear relevance between variables [3]. Popular filter methods [4], [5], use MI as a criterion to select the most revealing features for the target, while simultaneously reducing the level of correlation among the chosen features. This balance is vital for building models with relevant, clean, non-overlapping inputs, ensuring both performance and simplicity.

However, traditional MI-based approaches do have their drawbacks. They can be sensitive to noises and may struggle to identify complex patterns of redundancy, especially when dealing with high-dimensional datasets [3]. These approaches also often fall short in effectively separating informative features from redundant ones, particularly in high-dimensional spaces with complex inter-feature dependencies. Moreover, they typically lack mechanisms to group similar features or consolidate decisions across multiple feature subsets.

To address these challenges, a hybrid feature selection framework that leverages both Feature-Label Mutual Information (FLMI) and Feature-Feature Mutual Information (FFMI) is proposed in this current work. Our approach proceeds in two stages: first, we identify a core set of target relevant features using FLMI. Then, we address redundancy in the remaining features by clustering them and evaluating intra-cluster mutual information. A voting mechanism is applied across clusters to robustly select important non-redundant features.

This work makes the following key contributions:

- We introduce a two-stage mutual information-based feature selection framework that decouples relevance evaluation from redundancy filtering.

- We employ clustering to localize redundancy analysis within structurally similar feature groups, thereby improving redundancy detection accuracy.
- We propose a voting strategy that improves the stability and robustness of selected features across clusters.

Empirical evaluations on different datasets demonstrate the consistent improvements in model performance with reduced feature sets.

The rest of this paper is divided as follows. Section II outlines the prior research with respect to the current work. Section III describes the proposed methodology in detail. Section IV details the experimental configuration along with results and discussions. Finally, Section V concludes the proposed method along with future work directions.

II. Related Work

In the domain of machine learning, feature selection is a pivotal step that extracts a subset of significant features to improve predictive accuracy of a model, especially in high-dimensional datasets.

Various methods have been proposed, including approaches based on mutual information (MI), clustering, ensemble strategies, and hybrid techniques. The Maximal Relevance Minimum Redundancy criterion in [4], [6], is one of the most widely used MI-based methods, aiming to select features that maximize relevance while minimizing redundancy. Recent improvements, such as in [5], enhance the relevance of selected features while maintaining minimal redundancy by taking account into the unique relevance of features. A dynamic MI-based feature selection method that adapts to changing data distributions is introduced in [7], offering improved computational efficiency. In addition to MI-based methods, clustering techniques have also been integrated into feature selection. A geometry-aware MI-based method that performs clustering and feature selection simultaneously is proposed in [8]. The feature selection methods proposed in [9], [10] used k-means clustering to partition features into groups, followed by a selection process that retains the most revealing features. Ensemble methods in [11], [12], aim to enhance classification accuracy by combining multiple MI-based feature selection techniques. This approach helps address the limitations that individual filter methods may have, resulting in better overall performance. Similarly, hybrid approaches in [13], [14] blend MI-based feature ranking with Genetic Algorithms (GA), showing promising results in reducing computation time while enhancing model performance. Some works have also explored voting-based techniques for feature selection [15], [16], where clustering is used alongside MI to achieve more stable and reliable feature sets. In [17], a geometry-aware discriminative clustering framework is proposed by integrating mutual information, though it assumes feature spaces with clear geometric structure. In another line of work, MI has been combined with wavelet transforms for feature selection, as seen in [18], [19], offering a different domain specific angle for capturing relevant feature characteristics.

While the existing MI based feature selection methods have shown good results, they often face difficulty in handling both feature relevance and redundancy simultaneously. Clustering and voting approaches sometimes result in inconsistent results and often lack the full use of the information captured from mutual information. These research gaps emphasize the need for a more unified and adaptable approach that not only balances feature relevance and redundancy but also adjusts to the specific structure of the data.

III. Proposed Methodology

This section presents our proposed feature selection framework, which combines FLMI and FFMI, enhancing them with clustering and a voting strategy. The proposed method is presented in Algorithm 1.

Algorithm 1: Feature Selection using Mutual Information, Clustering, and Voting

Input: Dataset $D = \{X, y\}$, Threshold τ, Number of clusters k, Voting threshold $V_{\min}$
Output: Selected feature subset X^*
Step 1: Feature-Label Mutual Information (FLMI) Filtering
Compute $FLMI(x_i, y)$ for each feature $x_i \in X$
Rank features in descending order of $FLMI$
Select top m features with $FFMI > \tau \rightarrow X_{\text{FLMI}}$
Step 2: Feature-Feature Mutual Information (FFMI) Redundancy Filtering
$X_{\text{rem}} \leftarrow X \setminus X_{\text{FLMI}}$
Cluster X_{rem} into k clusters: $\{\zeta_1, \zeta_2, \ldots, \zeta_k\}$
foreach cluster ζ_j do
    foreach feature pair $(x_p, x_q) \in \zeta_j$ do
        Compute $FFMI(x_p, x_q)$
        if $FFMI(x_p, x_q) > \tau$ then
            Retain feature with higher $FLMI$
    Store retained features in $X^{(j)}_{\text{FFMI}}$
Aggregate all selected features: $X_{\text{FFMI}} = \bigcup_{j=1}^{k} X^{(j)}_{\text{FFMI}}$
Apply voting mechanism:
Count occurrence of each feature in X_{FFMI}
Select features with votes $\geq V_{\min} \rightarrow X_{\text{vote}}$
Step 3: Final Feature Subset Selection
$X^* \leftarrow X_{\text{FLMI}} \cup X_{\text{vote}}$
return X^*

Let the dataset be defined as $D = \{X, y\}$, where:

- $X = \{x_1, x_2, \ldots, x_n\}$ is the set of n features,
- y is the target variable.

We define:

- Mutual information $I(X;Y)$ between two discrete random variables X and Y quantifying the amount of shared information between X and Y [20], [21]. It is formally given by:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

  where, $p(x,y)$ is the joint probability of $X = x$ and $Y = y$; $p(x)$ and $p(y)$ are the respective marginal distributions. A higher $I(X;Y)$ value indicates a greater dependency between X and Y, whereas $I(X;Y) = 0$ implies statistical independence between X and Y.
- τ: Threshold to identify redundancy among features,
- k: Number of clusters for redundancy analysis,
- $V_{\min}$: Minimum number of votes required to retain a feature.

The objective is to select a subset $X^* \subseteq X$ that maximizes relevance to y while minimizing redundancy among the selected features. The integration of clustering and voting in the proposed method plays a critical role in refining the relevance of selected features. Clustering groups structurally similar or redundant features based on their mutual information profiles. This localized redundancy analysis allows us to better detect and remove highly correlated features that might not be redundant globally but are functionally overlapping within a subset. The subsequent voting mechanism reinforces features that consistently exhibit strong relevance across multiple clusters, ensuring selection stability. Together, these components improve both the precision of feature relevance assessment and the robustness of the final subset. This hybrid approach guarantees that the retained features are not only significant on an individual level but also collectively diverse, making them suitable for downstream machine learning tasks.

IV. Experimental Results and Discussions

The effectiveness of proposed feature selection method is assessed on eight different datasets taken from the UCI Dataset Repository[1], chosen to represent a variety of domains, feature sizes, and class distributions. To assess its effectiveness, we compare the framework with several well-established feature selection methods, including Mutual Information [22], Minimum Redundancy Maximum Relevance [4], ReliefF [23], and the Chi-Squared Test [24] related to classification accuracy, precision, recall, and F_1-score. All experiments are conducted using Python with the Scikit-learn library, and the Random Forest classifier [25] is used for evaluation. Each experiment is performed using a train-test split (80:20 ratio), and the results are averaged over ten runs to ensure robustness and reliability.

Table I presents a comparative analysis of model performance using the full feature set versus the reduced feature set obtained through our proposed feature selection algorithm across the datatasets. We varied the parameter values of our method, which resulted in different numbers of selected features across datasets. For each dataset, the configuration yielding the best classification performance was recorded and reported. The results demonstrate that in most cases, the reduced feature sets not only retained but often improved classification metrics. This validates that our method is effective at selecting informative and relevant features. While a few datasets, such as Prep. Handwritten Digits, Mushroom, and Statlog (Landsat Satellite), show a slight drop in some metrics after feature reduction, the trade-off is minimal and acceptable given the significant decrease in dimensionality. Overall, the method strikes a strong balance between reducing feature space and maintaining model performance.

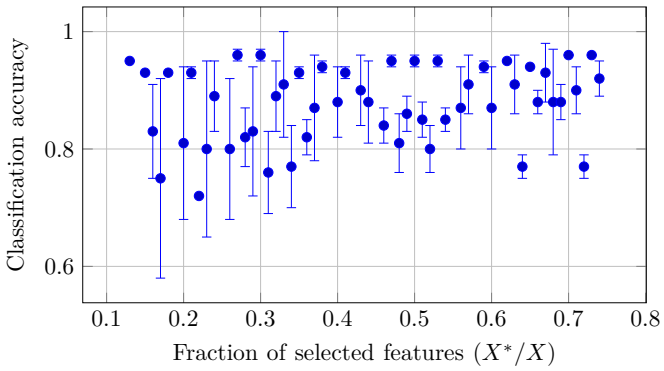

Fig. 1: Variation of classification accuracy with the fraction of selected features (X^*/X) across multiple datasets.

Figure 1 illustrates how the classification accuracy varies with the fraction of selected features across multiple datasets. The x-axis represents the ratio of retained features (X^*/X), while the y-axis shows the classification accuracy averaged over 10 independent train-test splits using a Random Forest classifier. Error bars represent the standard deviation across these runs, capturing the variability in model performance. The results reveal that high accuracy is frequently achieved even when less than 50% of the original features are retained. This suggests that many features may be redundant or irrelevant, and that our method is effective in identifying the most informative subset. The variation in accuracy, shown through standard deviation error bars, reflects differences in dataset characteristics, but the overall trend remains consistent. These results signifies the potential of the approach to reduce dimensionality without compromising performance. These trends confirm the effectiveness of the proposed method in selecting compact yet highly informative feature subsets, enabling dimensionality reduction without sacrificing classification performance.

To make a fair comparison, we ensured that each baseline method: Mutual Information (MI), mRMR, ReliefF, and the Chi-Squared Test selected the same number of features as our method did for each dataset. Figures 2, 3,

[1]https://archive.ics.uci.edu/

TABLE I: Classification performance with random forest classifier on different datasets; X is the original dataset and X* is the reduced dataset.

Dataset	No. of Features		Accuracy (%)		Precision		Recall		F_1-Score	
	X	X*	X	X*	X	X*	X	X*	X	X*
Automobile	25	09	75.01	90.62	0.759	0.908	0.751	0.906	0.754	0.907
Breast Cancer	30	07	96.49	99.12	0.965	0.991	0.964	0.991	0.965	0.991
Glass	09	05	83.72	90.69	0.866	0.991	0.837	0.906	0.851	0.911
Soyabean	35	19	87.03	92.59	0.890	0.943	0.870	0.925	0.880	0.934
Ionosphere	34	16	94.36	95.77	0.944	0.960	0.943	0.957	0.944	0.959
Prep. Handwritten digits	64	16	97.95	97.24	0.979	0.972	0.979	0.972	0.979	0.972
Mushroom	22	07	99.99	99.44	0.999	0.994	0.999	0.972	0.996	0.995
Statlog (Landsat Satellite)	36	20	91.76	90.98	0.916	0.908	0.917	0.909	0.916	0.908

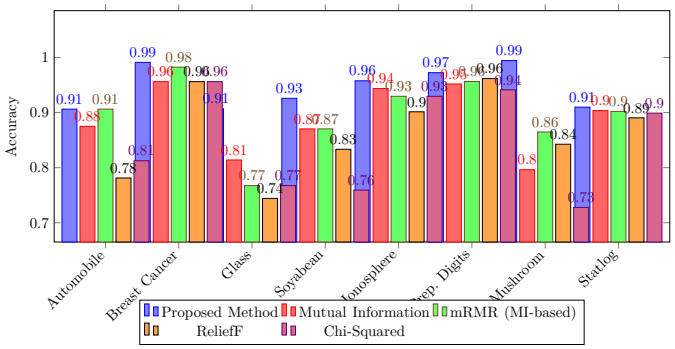

Fig. 2: Classification accuracy of the proposed method and baseline feature selection techniques.

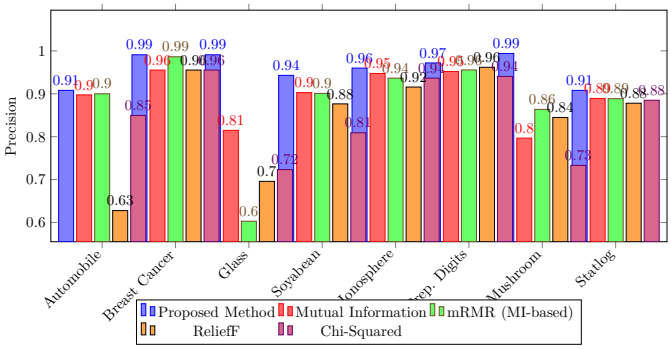

Fig. 3: Precision comparison of the proposed method and baseline feature selection techniques.

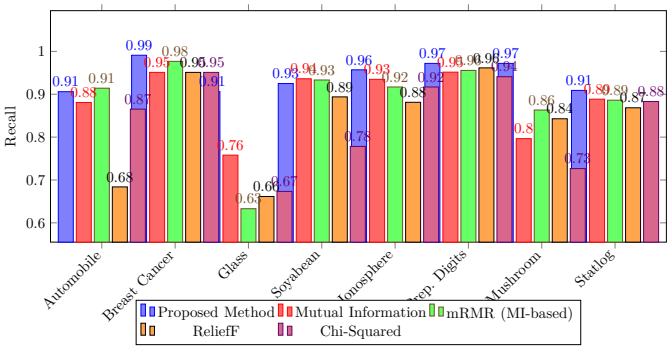

Fig. 4: Recall comparison of the proposed method and baseline feature selection techniques.

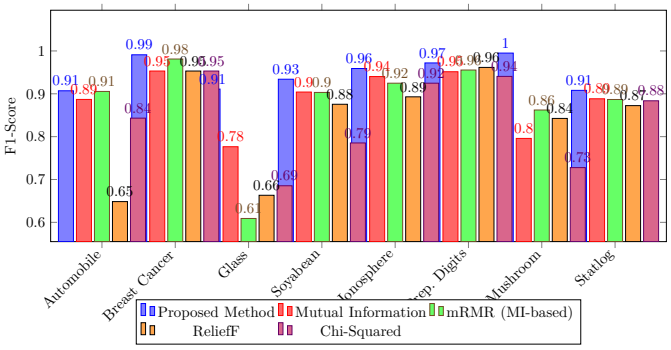

Fig. 5: F_1-Score comparison of the proposed method and baseline feature selection techniques

4, and 5 illustrate the comparative performance analysis of the presented method against the four baseline approaches across the datasets used. As shown in Figure 2, the proposed method consistently outperforms the baseline methods, especially in datasets like Automobile, Glass, and Soyabean. Similarly, Figure 3 highlights that our approach generally achieves higher precision, particularly on Glass and Soyabean, reducing false positives more effectively than the baselines. In terms of recall, depicted in Figure 4, our method shows significant improvements in capturing true positives, especially in datasets like Automobile and Glass, which are critical for imbalanced classifications. Finally, Figure 5 reveals that our method achieves strong F_1-scores across all datasets, providing an optimal trade-off between precision and recall, with particularly notable performance on Automobile, Glass, and Soyabean. These results validates the credibility of the presented feature selected strategy in consistently improving classification performance across multiple metrics.

However, it can be noted that for certain datasets such as Prep. Handwritten Digits and Statlog (Landsat Satellite), there is a slight drop in some metrics after feature reduction. While these decreases are marginal, they may be attributed to the nature of the feature set in these

datasets or the inherent characteristics of the features that our method selected. For example, the Prep. Handwritten Digits dataset, which involves image-like features, could experience minor drops due to the nature of dimensionality reduction in image data. Despite these slight decreases, the overall trade-off between dimensionality reduction and performance is acceptable, as the reduction in feature space significantly contributes to lower computational costs and potentially better generalization. Thus, the results still validate the credibility of our method in most cases, maintaining strong classification performance while reducing the feature set to a manageable size.

Compared to recent hybrid techniques such as [8], [19], the proposed method avoids heuristic domain specific tuning and instead uses a deterministic pipeline that integrates local redundancy filtering and voting for better stability across datasets. Its low computational overhead and flexibility across domains address the generalizability limitations of many recent state-of-the-art methods.

A. Computational Complexity Analysis

The computational complexity of the proposed algorithm is as follows:

- FLMI Calculation: For n features, computing feature-label mutual information is $O(n)$.
- FFMI + Clustering: For $m = |X \setminus X_{FLMI}|$ remaining features, FFMI requires $O(m^2)$ pairwise computations. Clustering using k-means incurs $O(kmdi)$, where d is feature dimensionality and i is the number of iterations.
- Voting Mechanism: Counting votes is $O(m)$.

Compared to traditional mRMR, which computes global pairwise MI for all features ($O(n^2)$), our method benefits from localized redundancy filtering within clusters and selective voting, which scales better in practice.

B. Parameter Sensitivity and Robustness Evaluation

This subsection presents a comprehensive evaluation of our method with respect to both parameter sensitivity and robustness under noisy and imbalanced conditions. These experiments are conducted to demonstrate the adaptability, generalization, and resilience of the proposed framework across a range of scenarios.

Parameter Sensitivity Analysis: We examined the effect of three key parameters: redundancy threshold (τ), number of clusters (k), and voting threshold ($V_{\min}$) on the classification performance of the proposed method using three representative datasets: Automobile, Glass, and Soybean. The observations are summarized as follows:

- Varying τ in the range [0.1, 0.9] revealed that values between 0.3 and 0.7 provided consistently stable performance. Extremely low values retained redundant features, whereas very high values risked discarding informative features that were correlated.
- When varied within the range [2, 10], moderate values of k (typically between 4 and 6) achieved a good balance between redundancy grouping and selection granularity. Too few clusters led to broad grouping, while too many reduced the benefits of cluster-level redundancy filtering.
- Setting $V_{\min} \geq 2$ was found effective in filtering out noisy features. However, excessively high thresholds reduced the number of retained features, occasionally affecting classification performance.

These results confirm that the proposed method is robust to parameter variations and performs reliably across a broad configuration space.

Robustness to Noise.: To evaluate noise robustness, Gaussian noise was added to the feature values of the Glass and Breast Cancer datasets at levels of 5%, 10%, and 20% standard deviation. The performance of the proposed method was recorded in terms of mean classification accuracy and is shown in Table II. The results indicate that the method maintains strong performance even with increasing noise levels, reflecting its stability under feature corruption.

TABLE II: Classification Accuracy (%) of the Proposed Method under Different Noise Levels

Noise Level (%)	Accuracy (%)
0	96.02
5	95.27
10	93.20
20	89.01

Robustness to Class Imbalance.: To assess performance under class imbalance, the Soybean dataset was artificially modified to create class distributions with minority-to-majority ratios of 1:2, 1:3, and 1:4. The results presented in Table III demonstrate that the method maintains high classification accuracy, underscoring its robustness to skewed class distributions.

TABLE III: Classification Accuracy (%) of the Proposed Method under Class Imbalance (Soybean Dataset)

Imbalance Ratio	Accuracy (%)
Original (Balanced)	92.59
1:2	91.26
1:3	90.32
1:4	89.26

These results provide empirical support to our claim that the proposed method is more robust to feature noise than traditional MI-based approaches. The clustering stage effectively localizes redundancy, minimizing the influence of spurious correlations caused by noise, while the voting mechanism ensures that only consistently strong features are retained. This dual-layer filtering suppresses the effect of noisy features, leading to more stable and accurate models under perturbation. Overall, the results across all scenarios validate the proposed method's effectiveness, parameter robustness, and resilience to data distortions in practical settings.

V. Conclusion and Future Directions

In this study, an innovative feature selection method is proposed by combining two different types of mutual information with clustering and voting mechanisms to improve classification performance. The use of clustering and voting not only enables better control over feature redundancy but also stabilizes feature selection outcomes, especially in high-dimensional or noisy environments. Experimental results demonstrate the superiority over several well-established feature selection techniques, including Mutual Information, mRMR, ReliefF, and the Chi-Squared Test, across a variety of benchmark datasets in terms of different classification metrics, while reducing dimensionality. Despite some minor drops in performance on certain datasets, the method achieves a strong balance between reduced feature space and model's predictive capability, thereby serving as a useful technique for real-world classification tasks. Overall, the results highlight the effectiveness and robustness of our feature selection framework in improving classification outcomes while managing high-dimensional data.

While our proposed feature selection method shows promising results, there are still the rooms for further exploration. Future work could involve extending the approach to handle multi-class and multi-label classification problems, as well as testing its scalability on larger, high-dimensional datasets. Additionally, integrating our method with deep learning models and exploring its impact on neural network performance could provide valuable insights. Another potential direction is the incorporation of other techniques, such as Principal Component Analysis or autoencoders, to enhance the method's efficiency and applicability. Furthermore, applying our approach to practical applications in domains such as bioinformatics, finance, and image analysis could validate its practical utility and robustness in complex, noisy environments.

References

[1] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," ACM computing surveys (CSUR), vol. 50, no. 6, pp. 1–45, 2017.

[2] J. Dunn, L. Mingardi, and Y. D. Zhuo, "Comparing interpretability and explainability for feature selection," arXiv preprint arXiv:2105.05328, 2021.

[3] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," Neural computing and applications, vol. 24, pp. 175–186, 2014.

[4] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," Journal of bioinformatics and computational biology, vol. 3, no. 02, pp. 185–205, 2005.

[5] S. Liu and M. Motani, "Improving mutual information based feature selection by boosting unique relevance," Journal of Artificial Intelligence Research, vol. 82, pp. 1267–1292, 2025.

[6] S. Ramírez-Gallego, I. Lastra, D. Martínez-Rego, V. Bolón-Canedo, J. M. Benítez, F. Herrera, and A. Alonso-Betanzos, "Fast-mrmr: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data," International Journal of Intelligent Systems, vol. 32, no. 2, pp. 134–152, 2017.

[7] I. C. Covert, W. Qiu, M. Lu, N. Y. Kim, N. J. White, and S.-I. Lee, "Learning to maximize mutual information for dynamic feature selection," in International Conference on Machine Learning. PMLR, 2023, pp. 6424–6447.

[8] L. Ohl, P.-A. Mattei, C. Bouveyron, M. Leclercq, A. Droit, and F. Precioso, "Sparse and geometry-aware generalisation of the mutual information for joint discriminative clustering and feature selection," Statistics and Computing, vol. 34, no. 5, p. 155, 2024.

[9] P. Wang, B. Xue, J. Liang, and M. Zhang, "Feature clustering-assisted feature selection with differential evolution," Pattern Recognition, vol. 140, p. 109523, 2023.

[10] F. Nie, Z. Ma, J. Wang, and X. Li, "Fast sparse discriminative k-means for unsupervised feature selection," IEEE Transactions on Neural Networks and Learning Systems, 2023.

[11] N. Hoque, M. Singh, and D. K. Bhattacharyya, "Efs-mi: an ensemble feature selection method for classification: An ensemble feature selection method," Complex & Intelligent Systems, vol. 4, pp. 105–118, 2018.

[12] A. S. Sumant and D. Patil, "Ensemble feature subset selection: integration of symmetric uncertainty and chi-square techniques with rrelieff," Journal of The Institution of Engineers (India): Series B, vol. 103, no. 3, pp. 831–844, 2022.

[13] A. Alzubaidi, G. Cosma, D. Brown, and A. G. Pockley, "Breast cancer diagnosis using a hybrid genetic algorithm for feature selection based on mutual information," in 2016 International Conference on Interactive Technologies and Games (ITAG). IEEE, 2016, pp. 70–76.

[14] R. Vijayanand, D. Devaraj, and B. Kannapiran, "A novel intrusion detection system for wireless mesh network with hybrid feature selection technique based on ga and mi," Journal of Intelligent & Fuzzy Systems, vol. 34, no. 3, pp. 1243–1250, 2018.

[15] E. Tasci, Y. Zhuge, H. Kaur, K. Camphausen, and A. V. Krauze, "Hierarchical voting-based feature selection and ensemble learning model scheme for glioma grading with clinical and molecular characteristics," International Journal of Molecular Sciences, vol. 23, no. 22, p. 14155, 2022.

[16] H. Roubhi, A. H. Gharbi, K. Rouabah, and P. Ravier, "Mutual information-based feature selection strategy for speech emotion recognition using machine learning algorithms combined with the voting rules method," Engineering, Technology & Applied Science Research, vol. 15, no. 1, pp. 19 207–19 213, 2025.

[17] L. Ohl, P.-A. Mattei, C. Bouveyron, M. Leclercq, A. Droit, and F. Precioso, "Sparse and geometry-aware generalisation of the mutual information for joint discriminative clustering and feature selection," Statistics and Computing, vol. 34, no. 5, p. 155, oct 2024.

[18] B. Li, P. Zhang, S. Liang, and G. Ren, "Feature extraction and selection for fault diagnosis of gear using wavelet entropy and mutual information," in 2008 9th international conference on signal processing. IEEE, 2008, pp. 2846–2850.

[19] G. Yuan, X. Li, P. Qiu, and X. Zhou, "Feature selection method based on wavelet similarity combined with maximum information coefficient," Information Sciences, vol. 699, p. 121801, 2025.

[20] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, vol. 69, no. 6, p. 066138, 2004.

[21] M. J. Canty, Image analysis, classification and change detection in remote sensing: with algorithms for Python. Crc Press, 2019.

[22] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, "Mifs-nd: A mutual information-based feature selection method," Expert systems with applications, vol. 41, no. 14, pp. 6371–6385, 2014.

[23] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relieff and rrelieff," Machine learning, vol. 53, pp. 23–69, 2003.

[24] Y. Wang and C. Zhou, "Feature selection method based on chi-square test and minimum redundancy," in Emerging Trends in Intelligent and Interactive Systems and Applications: Proceedings of the 5th International Conference on Intelligent, Interactive Systems and Applications (IISA2020). Springer, 2021, pp. 171–178.

[25] L. Breiman, "Random forests," Machine learning, vol. 45, pp. 5–32, 2001.