

International Journal of Knowledge Processing Studies (KPS)



Homepage: <http://kps.artahub.ir/>



ORIGINAL RESEARCH ARTICLE

A Weighted Multi-K Extension of ReliefF for Robust Feature Selection

Alican Doğan^{1,*}, Mansur Toçoğlu²

¹ Assistant Professor, Department of Management Information Systems, Bandırma Onyedi Eylül University, Balıkesir, Turkey. alicandogan@bandirma.edu.tr, 0000-0002-0553-2888.

² Assistant Professor D Department of Computer Engineering, Katip Çelebi University, İzmir, Turkey. mansuralp.tocoglu@ikc.edu.tr, 0000-0003-1784-9003.

ARTICLE INFO

Article History:

Received: 2025-11-25

Revised: 2026-01-22

Accepted: 2026-02-18

Published Online: 2026-03-01

Keywords:

Feature selection, ReliefF, Classification, Data mining, Machine learning.

Number of Reference: 30

Number of Figures: 4

Number of Tables: 9

DOI: 10.22034/kps.2026.588386.1283



ABSTRACT

This research proposes an improved version of ReliefF method for feature selection tasks in data mining. ReliefF-based algorithms are versatile and effective feature evaluators that identify conditional dependencies among features across instances, and they are commonly used in the preprocessing stage for tasks such as classification and regression. Since its introduction, numerous extensions of the ReliefF approach have been proposed to address redundancy, irrelevance, and noisy features, as well as the inherent limitations of the model in handling both binary and multi-class datasets. In this paper, we examine the effect of the usage of multiple number of neighbors in ReliefF algorithm concurrently with the help of standard majority voting and weighted majority voting strategies. The proposed method, is called Multi-k Weighted ReliefF (MW-ReliefF). It requires a set of different number of neighbors to calculate near hit and near miss values for each. Then it updates those values with respect to the weights of the predetermined number of nearest neighbors. The method is implemented as a preprocessing step for feature reduction before classifying 20 benchmark binary and multi-class datasets and classification performance results are obtained and compared to the counterparts of ReliefF method and its traditional variants in order to prove the validity of MW-ReliefF. In fact, the experimental findings demonstrated that the proposed algorithm outperformed the original ReliefF methods across almost all datasets.

► **Citation:** Doğan, A., & Toçoğlu, M. (2026). A Weighted Multi-K Extension of ReliefF for Robust Feature Selection. *International Journal of Knowledge Processing Studies (KPS)*, 6(1): 65-89. Doi: 10.22034/kps.2026.588386.1283

1. Introduction

Feature selection is a fundamental step in data mining and machine learning that reduces dimensionality, improves model generalization, and lowers computational cost (Park & Kwon, 2007). By selecting the most informative subset of features, it becomes possible to alleviate overfitting, shorten training time, and improve classification accuracy, particularly when dealing with high-dimensional datasets. The visual representation of a feature selection process is given in Figure 1.

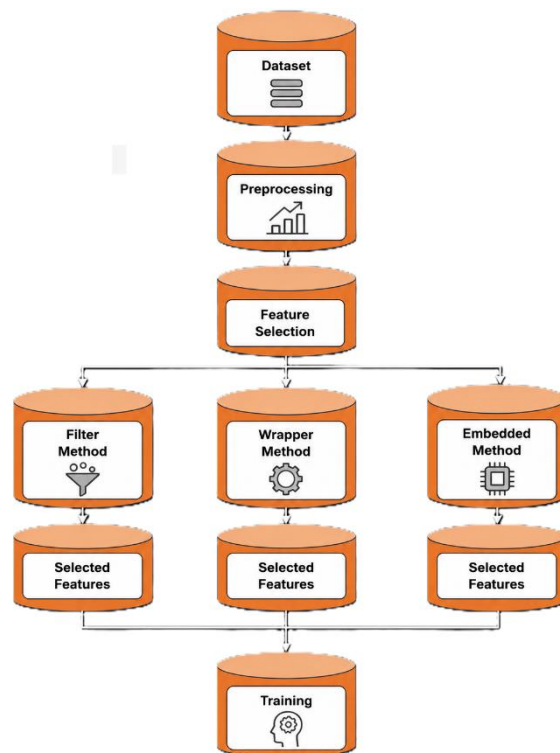


Figure 1. A simple representation of feature selection process

Feature selection techniques developed under various evaluation criteria are typically grouped into three main categories: filter, wrapper, and embedded methods (McKinney et al., 2007). Filter methods first choose a subset of features from the dataset before training the model, making the selection process independent of the learning phase. Examples include Relief and ReliefF. Wrapper methods, on the other hand, evaluate subsets directly based on model performance, such as the Las Vegas Wrapper (LVW). While these methods often achieve

better results than filters, they require repeatedly training the model, which makes them computationally expensive. The embedded approach integrates feature selection into the training process itself, as seen in methods like Least Absolute Shrinkage and Selection Operator (LASSO). In this study, we adopt the filter-based framework because of its simplicity, scalability, and independence compared to wrapper and embedded methods. The flowchart of a feature selection process is shown in Figure 2.

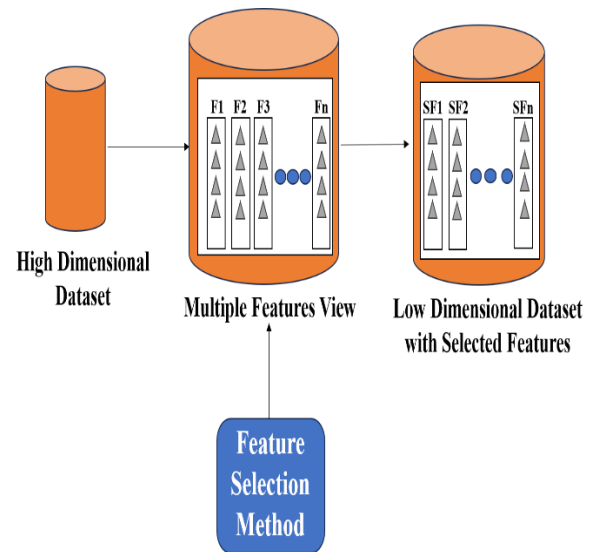


Figure 2. The flowchart of a feature selection process

Among numerous feature selection approaches, the Relief family of algorithms has attracted significant attention. Relief and its extensions are known as successful feature evaluators because they estimate the importance of features by considering conditional dependencies between instances (Sun et al., 2010). They have been widely applied in preprocessing tasks such as document classification, regression, and biomedical data analysis. Despite these advantages, the original Relief algorithm suffers from several limitations, including sensitivity to noisy and redundant features, and challenges in effectively handling both binary and multi-class problems. A number of Relief-based variants have been proposed to address these shortcomings. Algorithms such as SURF (Greene et al., 2009), TURF (Moore & White, 2007), and MultiSURF

(MacKenzie & Moore, 2013) extend the original framework by modifying neighborhood definitions, incorporating statistical thresholds, or expanding the sampling strategy. These extensions have demonstrated improved performance across different application domains. However, they still rely on a fixed neighborhood size parameter and do not explicitly exploit the benefits of combining multiple neighborhood settings simultaneously. Furthermore, the weighting of neighbor contributions remains an underexplored dimension in Relief-based feature evaluation.

Despite the progress achieved by these Relief-based extensions, an important methodological gap remains. Existing approaches generally modify the neighborhood definition or determine neighborhood relationships using a single neighborhood configuration at a time. Consequently, the feature relevance estimation may remain dependent on the characteristics of a particular neighborhood scale. Although methods such as ReliefSeq introduce adaptive neighborhood selection and STIR incorporates a statistical inference framework, these approaches do not provide a unified mechanism for aggregating feature relevance information obtained from multiple neighborhood sizes through an explicit weighting scheme. Similarly, existing weighted Relief-based approaches primarily modify the contribution of individual instances or feature scores rather than jointly combining the information obtained from multiple neighborhood scales. Therefore, there is a need for a feature selection framework that can simultaneously consider different neighborhood granularities and systematically aggregate their contributions. This constitutes the research gap addressed by the proposed MW-Relief algorithm.

Existing ReliefF extensions have primarily focused on modifying the neighborhood definition, adapting the number of neighbors, introducing statistical inference, or changing the spatial sampling strategy. However, these approaches generally evaluate feature relevance within a particular neighborhood configuration rather

than explicitly integrating feature relevance information obtained from multiple neighborhood sizes into a unified scoring mechanism. Consequently, the sensitivity of Relief-based feature evaluation to the choice of a single neighborhood size remains insufficiently addressed. A second related gap concerns the aggregation of information obtained at different neighborhood scales. Although weighting strategies have been investigated in several Relief-based approaches, the joint use of multiple neighborhood sizes and an explicit weighting mechanism to aggregate their feature-level contributions has not been systematically investigated as a unified feature selection framework. This gap motivates the development of MW-ReliefF, which simultaneously evaluates feature relevance at multiple neighborhood sizes and integrates the resulting contributions through a predefined weighting scheme.

In this study, we propose the Multi-k Weighted ReliefF (MW-ReliefF) algorithm, an ensemble extension of ReliefF that integrates multiple neighborhood sizes together with majority and weighted majority voting strategies. Unlike traditional approaches that use a single k value, MW-ReliefF computes near-hit and near-miss contributions across several neighborhood sizes, and then aggregates them using either uniform or inverse distance-based weighting schemes. This enables the algorithm to capture a more comprehensive view of local and global feature relevance, making it more robust to class imbalance and varying data distributions.

The proposed method is validated on 20 benchmark binary and multi-class datasets. In extensive experiments, MW-ReliefF is applied as a preprocessing feature selection step and compared against ReliefF and its common extensions including SURF (Greene et al., 2009), TURF (Moore & White, 2007), and MultiSURF (MacKenzie & Moore, 2013). The results show that MW-ReliefF improves classification accuracy over the competing methods on most of the evaluated benchmark datasets, demonstrating its effectiveness and practical value for

feature selection in diverse machine learning tasks.

Although MW-ReliefF retains the fundamental neighbor-based principle of the original ReliefF algorithm, its methodological contribution lies in extending the feature-weighting process from a single neighborhood scale to a multi-neighborhood framework. Conventional ReliefF assigns feature relevance according to a predefined neighborhood size, which provides a single-scale characterization of feature–class relationships. In contrast, MW-ReliefF evaluates feature relevance across multiple neighborhood sizes and integrates the resulting contributions through a weighted aggregation scheme. This formulation enables feature importance to be determined from complementary local neighborhood structures rather than from a single neighborhood configuration. Therefore, the contribution of MW-ReliefF is not a replacement of the underlying ReliefF mechanism, but a systematic extension of its feature-weighting framework that incorporates information obtained at multiple neighborhood scales.

The resulting framework provides a unified mechanism for integrating neighborhood-dependent feature relevance into a single feature ranking. In this respect, MW-ReliefF differs from approaches that primarily modify the neighborhood search, adapt the number of neighbors, or introduce alternative weighting strategies within a single Relief-based formulation. The proposed method instead focuses on the explicit aggregation of relevance information obtained from multiple neighborhood configurations, thereby providing a multi-scale feature selection strategy while preserving the original ReliefF learning principle.

The main contributions of this study are summarized as follows:

- We propose a novel feature selection algorithm, Multi-k Weighted ReliefF (MW-ReliefF), which extends the classical ReliefF by incorporating multiple neighborhood sizes simultaneously.
- The algorithm leverages both standard majority voting and weighted majority

voting strategies to integrate information from diverse neighborhood structures.

- MW-ReliefF enhances feature selection by combining near-hit and near-miss information obtained from multiple neighborhood sizes through a weighted aggregation strategy.

- Extensive experiments were conducted on 20 benchmark binary and multi-class datasets, including statistical analysis with accuracy comparisons.

- The results demonstrate that MW-ReliefF achieves higher classification performance and stronger feature reduction capability compared to ReliefF, SURF, TURF, and MultiSURF on most of the evaluated benchmark datasets.

The methodological contribution of MW-ReliefF lies not in replacing the fundamental ReliefF mechanism, but in introducing a systematic multi-neighborhood aggregation framework for Relief-based feature relevance estimation. While conventional ReliefF determines feature relevance according to a single neighborhood size, MW-ReliefF independently obtains feature relevance contributions at multiple k values and subsequently combines these contributions through a predefined weighting scheme. This design allows information from different neighborhood scales to contribute jointly to the final feature ranking. Therefore, MW-ReliefF can be viewed as an ensemble-inspired extension of ReliefF at the feature-scoring level, where the individual neighborhood configurations provide complementary relevance estimates rather than independent predictive models. The contribution is consequently centered on the integration and weighting of multi-scale ReliefF information within a unified feature selection procedure.

The remainder of this paper is organized as follows. Section 2 reviews related work on feature selection, including Relief-based algorithms. Section 3 presents the proposed MW-ReliefF method in detail. Section 4 reports experimental settings and results, along with comparative evaluations against baseline methods. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2. Literature Review

Feature selection has drawn significant research attention over the last years. In this section, we briefly review existing approaches and outline the main challenges that our algorithm seeks to overcome.

The challenge of feature estimation has received significant attention in previous studies. Feature quality is commonly evaluated through multiple assessment measures (Sun, 2007). If the target variable is categorical, common evaluation measures include information gain, Gini index, distance measures, J-measure, ReliefF, minimum description length (MDL) as well as χ^2 and G statistics. For continuous target variables, widely used evaluation measures include correlation coefficient, mutual information, mean squared error (MSE), coefficient of determination (R^2), and statistical dependency measures such as maximal information coefficient (MIC). Many evaluation metrics rely on the assumption of conditional independence of features with respect to the target variable, which makes them unsuitable for capturing complex feature interactions. In contrast, the Relief, ReliefF, and RReliefF algorithms avoid this assumption and are capable of accurately assessing feature quality even under strong attribute dependencies. Therefore, Relief-based approaches remain among the most effective preprocessing techniques available.

ReliefF method has been a very important and one of the most frequently used feature selection models. In 1992, the algorithm was firstly proposed by Kira and Rendell (Kira & Rendell, 1992). Relief identifies relevant features through a statistical approach. Unlike heuristic-based methods, it maintains accuracy even when features interact and demonstrates robustness to noise. Moreover, its computational cost grows linearly with both the number of features and training instances, independent of the complexity of the target concept. Afterwards, Kira and Rendell show its advantages in terms of learning time and the accuracy of the learned concept, suggesting Relief's practicality (Kira & Rendell, 1992). Kononenko analyzes and extends it to deal with noisy,

incomplete, and multi-class data sets (Kononenko, 1994). Additionally, they propose ReliefF which is an extension of Relief (Kira & Rendell, 1992). Kohavi and John Compares the wrapper method with both inductions performed without feature subset selection and with the Relief algorithm (Kohavi & John, 1997).

There are some studies inspired by ReliefF. For instance, Park and Kwon introduces advanced Relief algorithms and their application in instance-based filtering for selecting document features (Park & Kwon, 2007). McKinney et al. propose an Evaporative Cooling (EC) feature selection algorithm designed to identify a subset of attributes that achieve the optimal information temperature (McKinney et al., 2013). Sun et al. presents local-learning-based feature selection for high-dimensional data analysis which is called I-RELIEF (Sun et al., 2019). Draper et al. introduce an enhanced variant of ReliefF, Iterative Relief, and demonstrates on synthetic datasets that it eliminates the bias inherent in the method (Draper et al., 2003). Similarly, Sun extends I-RELIEF to handle multiclass problems through the introduction of a novel multiclass margin definition (Sun et al., 2019).

The relief-based methods are still a fundamental research area among researchers in the last years. McKinney et al. develops a gene wise adaptive k nearest neighbor feature selection tool which is called ReliefSeq (McKinney et al., 2013). Eppstein and Haake propose a population-based variant dubbed very large scale ReliefF called VLSReliefF for genome-wide association analysis (Eppstein & Haake, 2008). Le et al. redefine the Relief-based feature selection algorithm to develop a new class of Statistical Inference Relief (STIR) estimators that preserve the capacity to detect interactions while integrating the sample variance of nearest neighbor distances into the assessment of attribute importance (Le et al., 2019). Greene et al. introduce Spatially Uniform ReliefF (SURF) for computationally-efficient filtering of gene-gene interactions (Greene et al., 2009). Stokes and Wisweswaran develop a spatially

weighted variant of Relief, termed Sigmoid Weighted ReliefF Star (SWRF*) (Stokes & Visweswaran, 2012). Moore and White present Tuning ReliefF (TURF) for genome-wide genetic analysis (Moore & White, 2007). Mackenzie and Moore propose a new heuristic, MultiSURF, capable of accurately detecting such interactions while maintaining linear time complexity with respect to the number of genes (MacKenzie & Moore, 2013). Apart from research studies, Urbanowicz et al. complete a literature

review study about Relief-based feature selection methods in literature (Urbanowicz et al., 2018). The recently proposed relief-based feature selection models and their application domains are given in Table 1.

Table 1. The recently proposed cross validation models in literature

Author	Year	Method	Domain	Data	Performance Metrics
Kılıç and Yalçın	2025	Transfer Learning and ReliefF	Agriculture	Chickpea Seed Data	Accuracy
Zhang et al.	2022	RG-NBEO	Various	18 Standard Datasets	Accuracy
Zhang et al.	2023	ReliefF-SFA	Process Monitoring	heat treatment furnace process data	Accuracy
Uzun and Ballı	2022	ReliefF	Intrusion Detection	NSL-KDD	Chi-Square
Zhang et al.	2022	RBEFF	Various	28 Datasets	Accuracy
Liu et al.	2020	ReliefF-SFA	Process Monitoring	Production process data	Fault Detection Rate
Karizaki and Tavassoli	2019	ReliefF-BDA	Various	5 Datasets	Accuracy
Yinghua et al.	2019	ReliefF-PCA	Fault Detection	Tennessee Eastman Process Data	T-Square
Peker et al.	2019	ReliefF+RF	Medical	Anesthesia Data	Accuracy
Chen et al.	2011	ReliefF+FCBF	Genetic	DNA Data	Accuracy
Xin et al.	2013	ReliefF-GA-Wrapper	Image	Chinese characters Dataset	Accuracy
Kononenko et al.	1997	ReliefF+DT	Various	4 Datasets	Accuracy
Park and Kwon	2007	Extended ReliefF	Web Document	Reuters Dataset	Precision
McKinney et al.	2007	Evaporative Cooling ReliefF	Genetic	Simulated Genotypic Data	Chi-Square
Sun et al.	2010	Iterative ReliefF	Various	11 Datasets	Classification Error Rate
Draper et al.	2003	Iterative ReliefF	Image	Animal Image Dataset	Recognition Rate
Sun	2007	Iterative ReliefF	Various	Microarray data sets	ROC Curve
MacKenzie and Moore	2013	MultiSURF	Medical	Human Disease Data	Accuracy
Greene et al.	2009	SURF	Medical	Human Disease Data	Success Rate
Moore and White	2007	TURF	Genetic	DNA Data	Accuracy
Epsstein and Haake	2008	VLS-RELIEF	Medical	Health Data	Percentile Rank
Le et al.	2019	STIR	Genetic	Gene Expression Data	T-Test

2.1 Distinction of MW-ReliefF from Existing Relief-based Methods

Although numerous extensions of the Relief family have been proposed over the past three decades, most existing methods focus on improving a single aspect of the original algorithm, such as adaptive neighborhood selection, statistical significance estimation, interaction

detection, or integration with optimization techniques. Consequently, the neighborhood size used during feature weight estimation is still determined by a single predefined parameter or an adaptive mechanism that ultimately produces one effective neighborhood for each instance.

The proposed MW-ReliefF follows a fundamentally different strategy. Instead of relying on a single neighborhood size, the

proposed algorithm simultaneously evaluates multiple neighborhood configurations and integrates their individual feature importance estimates through an ensemble-based weighted aggregation mechanism. This allows MW-ReliefF to exploit complementary information obtained from both local and broader neighborhood structures within the same feature evaluation process. Consequently, the proposed method becomes less sensitive to the selection of a single k value while producing more stable feature importance estimates.

Unlike previous weighted Relief variants, where weighting is generally applied to neighboring instances, feature distances, or adaptive neighborhood construction, MW-ReliefF introduces weighting at the decision fusion level by assigning different contributions to the feature importance vectors generated from multiple ReliefF executions. Therefore, the proposed approach represents a neighborhood-level ensemble strategy rather than a modification of the original ReliefF distance computation.

Table 2. Comparison of MW-ReliefF with representative Relief-based algorithms

Method	Multiple k values	Weighted aggregation	Adaptive neighborhood	Statistical inference	Ensemble strategy
ReliefF	X	X	X	X	X
ReliefSeq	✓ (Adaptive k)	X	✓	X	X
STIR	X	X	X	✓	X
SURF	X	X	✓	X	X
MultiSURF	X	X	✓	X	X
Iterative Relief	X	X	✓	X	X
RBEFF	X	X	Random subspaces	X	Partial
MW-ReliefF (Proposed)	✓	✓	Multiple neighborhood scales	X	✓

Recent feature selection research has increasingly explored methods beyond the Relief family, including information-theoretic filters, mutual-information-based approaches, optimization-driven feature selection, and embedded methods based on regularized or tree-based learning models. These approaches provide alternative mechanisms for estimating feature relevance and selecting informative subsets. However, direct comparison with all such methods is beyond the primary scope of the present study. The objective of MW-ReliefF is specifically to investigate whether combining multiple neighborhood scales with a weighted aggregation mechanism can improve Relief-based feature evaluation. Accordingly, the experimental study focuses on ReliefF and its established neighborhood-based extensions, namely SURF, TURF, and MultiSURF, which provide methodologically relevant baselines for isolating the contribution of the proposed multi- k strategy. Nevertheless, comparisons with a broader set of contemporary filter, wrapper, and embedded feature selection methods would provide an additional

perspective on the general competitiveness of MW-ReliefF and are therefore identified as an important direction for future investigation.

3. Method

3.1 Notations

A dataset suitable for a classification task is denoted by D . Each attribute of this dataset is represented by f and n is the total number of samples. d is the total number of features. Each sample is denoted by x . The selected features after applying the feature selection algorithm are represented as f' and n is the number of selected features where $n < d$. Finally, $F = \{f_1, f_2, \dots, f_d\}$ is the set of selected features as a result of the feature selection process.

Let a dataset for classification be denoted by $D = \{x_i, y_i\}_{i=1}^n$ where:

n : total number of samples (instances)

d : total number of original features

$F = \{f_1, f_2, \dots, f_d\}$: the complete set of features

$x_i = \{x_{i1}, x_{i2}, \dots, x_{id}\} \in \mathbf{R}^d$: the i -th sample described by d features

$y_i \in Y$: class label of the i -th sample, where $y = \{1, 2, \dots, C\}$ and C is the number of classes

$f' \subseteq F$: the selected subset of relevant features.

$n = |f'|$: the number of selected features, where $n < d$.

$\delta(x_i, x_j)$: distance function between instances x_i and x_j .

$NH_k(x_i)$: the set of k nearest neighbors of x_i belonging to the same class (near-hit set).

$NM_{k,c}(x_i)$: the set of k nearest neighbors of x_i from class $c \neq y_i$ (near-miss set).

$W(f)$: importance (weight) assigned to feature f .

k : the neighborhood size parameter in ReliefF.

$K = \{k_1, k_2, \dots, k_m\}$: the set of multiple neighborhood sizes in MW-ReliefF.

ω_k : weight coefficient assigned to neighborhood size k , either uniform ($\omega_k = 1/m$) or inverse proportional ($\omega_k \propto 1/k$).

3.2 ReliefF Method

The fundamental concept behind the original Relief algorithm is to assess the relevance of attributes, those with weights exceeding a defined threshold, by measuring the differences in attribute values between a given instance and its two nearest neighbors, namely the Hit and the Miss. Based on the degree of relevance between features and class labels, distinct weights are assigned to each feature, and those whose weights fall below a specified threshold are eliminated. In the Relief algorithm, two nearest neighbors are identified for each randomly selected instance: one from the same class, referred to as the nearest Hit instance (H), and another from a different class, referred to as the nearest Miss instance (M). After computing the final feature weights, the features are ranked in descending order and the top- n features are selected. The larger the feature's weight, the stronger its ability to discriminate between classes. Conversely, a smaller weight indicates a weaker contribution to classification performance.

For each randomly sampled instance $x_i \in D$, the algorithm finds:

$NH_k(x_i)$: k nearest neighbors from the same class (near-hits).

$NM_{k,c}(x_i)$: k nearest neighbors from each of the other classes $c \in Y, c \neq y_i$ (near-misses).

The weight of feature $f_j \in F$ updated as:

(1)

$$W(f_j) \leftarrow W(f_j) - \frac{1}{m} \sum_{x_h \in NH_k(x_i)} \text{diff}(f_j, x_i, x_h) + \frac{1}{m} \sum_{c \neq y_i} \frac{P(c)}{P(y_i)} \sum_{x_m \in NM_{k,c}(x_i)} \text{diff}(f_j, x_i, x_m)$$

$\text{diff}(f_j, x_i, x_h)$ measures the difference between two instances on feature f_j .

For continuous features,

(2)

$$\text{diff}(f_j, x_i, x_h) = \frac{|x_{ij} - x_{hj}|}{\max(f_j) - \min(f_j)}$$

$P(c)$ is the prior probability of class c . ReliefF repeats this process for m iterations (usually $m=n$) and produces a ranking of features based on their final weights. In summary, the idea of this algorithm is that the weight of a feature is determined by its ability to discriminate the instances with respect to their classes without using the values of the other features. For this reason, the values of the selected feature of k nearest neighbors of a randomly chosen sample from the same class should be close while the values of the neighbors from different classes should be highly different in order to increase the weight of the selected feature.

Figure 3 demonstrates weight update process of ReliefF method for a dataset including two distinct classes. In this visual representation, k value for nearest neighbor selection is supposed to be 1. Weights of each feature in the total feature set of any experimented data ($F_1, F_2, \dots, F_n$) are initialized first and then updated with respect to calculated near-hit and near-miss values for each randomly selected instances. This process is reiterated until all randomly selected samples are analysed. Finally, all features are ordered with respect to their weight values and the first determined number of features are selected and feature selection task using ReliefF method is completed.

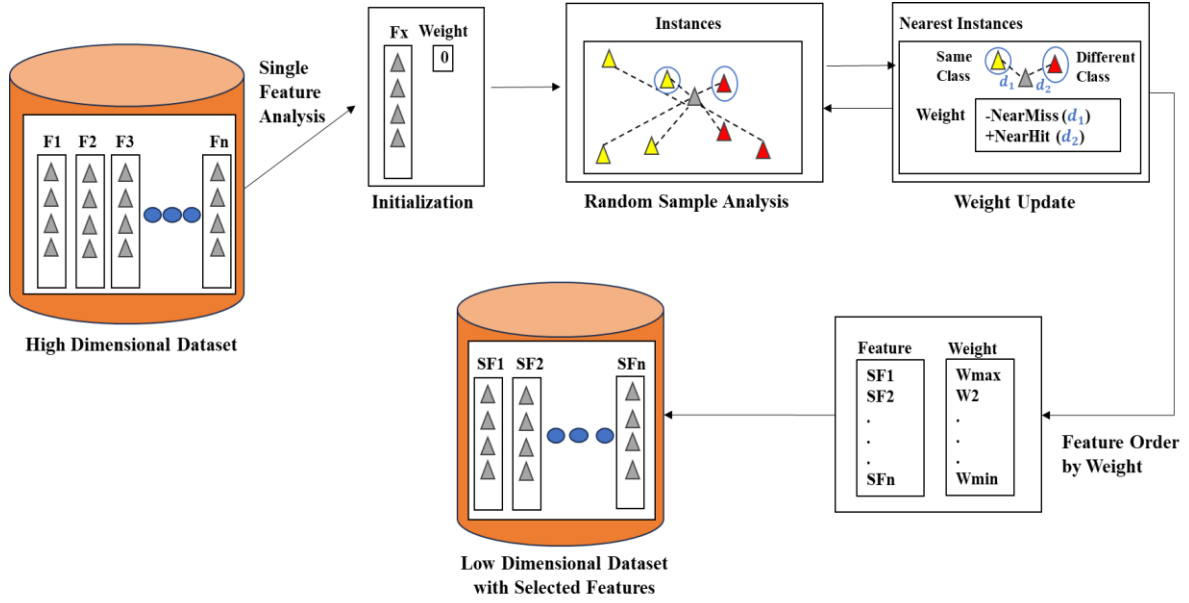


Figure 3. The sample flowchart of ReliefF method

3.3 The Proposed Approach: MW-ReliefF Method

The classical ReliefF algorithm utilizes a fixed number of nearest neighbors k to estimate the relevance of each feature by analyzing the differences between an instance and its nearest hits and misses. Although this approach effectively identifies conditionally dependent features, its performance is sensitive to the selection of the k parameter. A small k may fail to capture global patterns, while a large k may dilute local relationships, leading to reduced discriminative power. To overcome this limitation, the Multi- k Weighted ReliefF (MW-ReliefF) algorithm extends ReliefF by integrating multiple k -values simultaneously within a unified framework.

Let $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{N \times d}$ be a dataset with N instances and d features, and $Y = [y_1, y_2, \dots, y_N]$ be the corresponding class labels. For each instance x_i , the MW-ReliefF algorithm considers a set of neighborhood parameters $K = \{k_1, k_2, \dots, k_m\}$. For each $k_j \in K$, the algorithm identifies:

- The k_j nearest hits, $H_i^{(k_j)}$, from the same class as x_i and
- The k_j nearest misses, $M_{i,c}^{(k_j)}$, from each different class as x_i and $c \neq y_i$

For a feature f , the difference function is defined as:

(3)

$$\text{diff}(f, x_i, x_j) = \begin{cases} |x_{i,f} - x_{j,f}|, & \text{if } f \text{ is continuous} \\ 0, & \text{if } f \text{ is discrete and } x_{i,f} = x_{j,f} \\ 1, & \text{if } f \text{ is discrete and } x_{i,f} \neq x_{j,f} \end{cases}$$

For each k in the set K , ReliefF computes a local importance contribution for feature f as:

$$\Delta W_f^{(k)} = \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{k} \sum_{x_h \in H_i^{(k)}} \text{diff}(f, x_i, x_h) - \sum_{c \neq y_i} p(c) \cdot \frac{1}{k} \sum_{x_m \in M_{i,c}^{(k)}} \text{diff}(f, x_i, x_m) \right]$$

where $p(c)$ is the prior probability of class c . The MW-ReliefF algorithm aggregates these multiple k -level contributions through a weighted ensemble strategy:

(5)

$$W_f = \sum_{j=1}^m \omega_j \cdot \Delta W_f^{(k_j)}$$

where ω_j represents the weight assigned to the k_j level. Two weighting schemes are considered:

(6)

$$\omega_j = \begin{cases} \frac{1}{m}, & \text{(equal weighting, "same")} \\ \frac{1/k_j}{\sum_{t=1}^m (1/k_t)}, & \text{(inverse weighting, "inverse")} \end{cases}$$

This ensemble design allows MW-ReliefF to integrate both local and global neighborhood information, thus balancing sensitivity to small-scale and large-scale data structures. After aggregating contributions across all k -levels, features with weights lower than a predefined threshold θ are

eliminated, and the remaining top- d' features are selected:

(7)

$$S = \{f \mid W_f > \theta\}, |S| = d_t$$

Algorithmically, MW-ReliefF can be described in the following steps:

1. Initialization: Define $K = \{k_1, k_2, \dots, k_m\}$ and select the weighting scheme ω_j .
2. Neighborhood Search: For each instance x_i , find hit and miss sets for each $k_j \in K$.
3. Feature Scoring: Compute $\Delta W_f^{(k_j)}$ for each feature f and each k_j .
4. Aggregation: Combine contributions using the ensemble weighting function.
5. Feature Selection: Rank features by W_f and retain the top- n features.

In summary, MW-ReliefF not only generalizes the traditional ReliefF by considering multiple neighborhood granularities concurrently, but also enhances robustness by integrating adaptive weighting strategies. This enables the algorithm to capture a richer spectrum of feature relevance across varying data complexities, leading to superior classification performance as demonstrated in the experimental results. Figure 3 shows the flow chart of this algorithm with 3 different k values which are 1, 2, and 3. The method takes a list of k values as input with different numbers. These values do not have to be sequential. For example, a sample input for this parameter could be $k\text{-list} = \{3, 7, 11, 19\}$. The ensemble weighting function could consider either proportional or inversely proportional approach with respect to k values since the effect of the calculation with smaller number of neighbors is supposed to be higher.

In certain cases, it may be more desirable for the effects of feature selection derived from smaller k values to have a stronger influence than those obtained from larger k values. This rationale arises from the intuition that the nearest neighbors with smaller k capture the most local and fine-grained relationships between instances, providing more discriminative information about subtle class boundaries. Conversely, larger k values tend to introduce a broader

neighborhood that can dilute the sensitivity of the feature evaluation by incorporating less similar or even noisy instances. Therefore, assigning higher weights to smaller k values within the inverse weighting mechanism allows the algorithm to emphasize highly relevant, instance-specific dependencies while still preserving the stability contribution of larger neighborhoods.

The contribution of each neighborhood size is controlled by a weighting coefficient that determines how strongly its feature relevance estimate contributes to the final MW-ReliefF score. Two weighting strategies are considered in the proposed framework. In the same-weight setting, all neighborhood sizes receive an equal contribution. In the inverse-weight setting, the contribution assigned to a neighborhood is inversely related to its neighborhood size, such that smaller k values receive larger weights than larger k values. More specifically, for a neighborhood size k_i , the inverse weighting coefficient is defined as

(8)

$$a_i = \frac{\frac{1}{k_i}}{\sum_{j=1}^m \frac{1}{k_j}}$$

Where m denotes the number of neighborhood sizes in the selected k -list. The normalized coefficients satisfy $\sum_{i=1}^m a_i = 1$. The final feature score is then obtained by aggregating the relevance contributions obtained at the different neighborhood sizes:

(9)

$$W_f = \sum_{i=1}^m a_i \Delta W_f^{(k_i)}$$

where $\Delta W_f^{(k_i)}$ represents the feature-weight contribution estimated at neighborhood size k_i . Thus, the weighting mechanism allows MW-ReliefF to control the relative contribution of different neighborhood scales while preserving information from all evaluated neighborhood sizes.

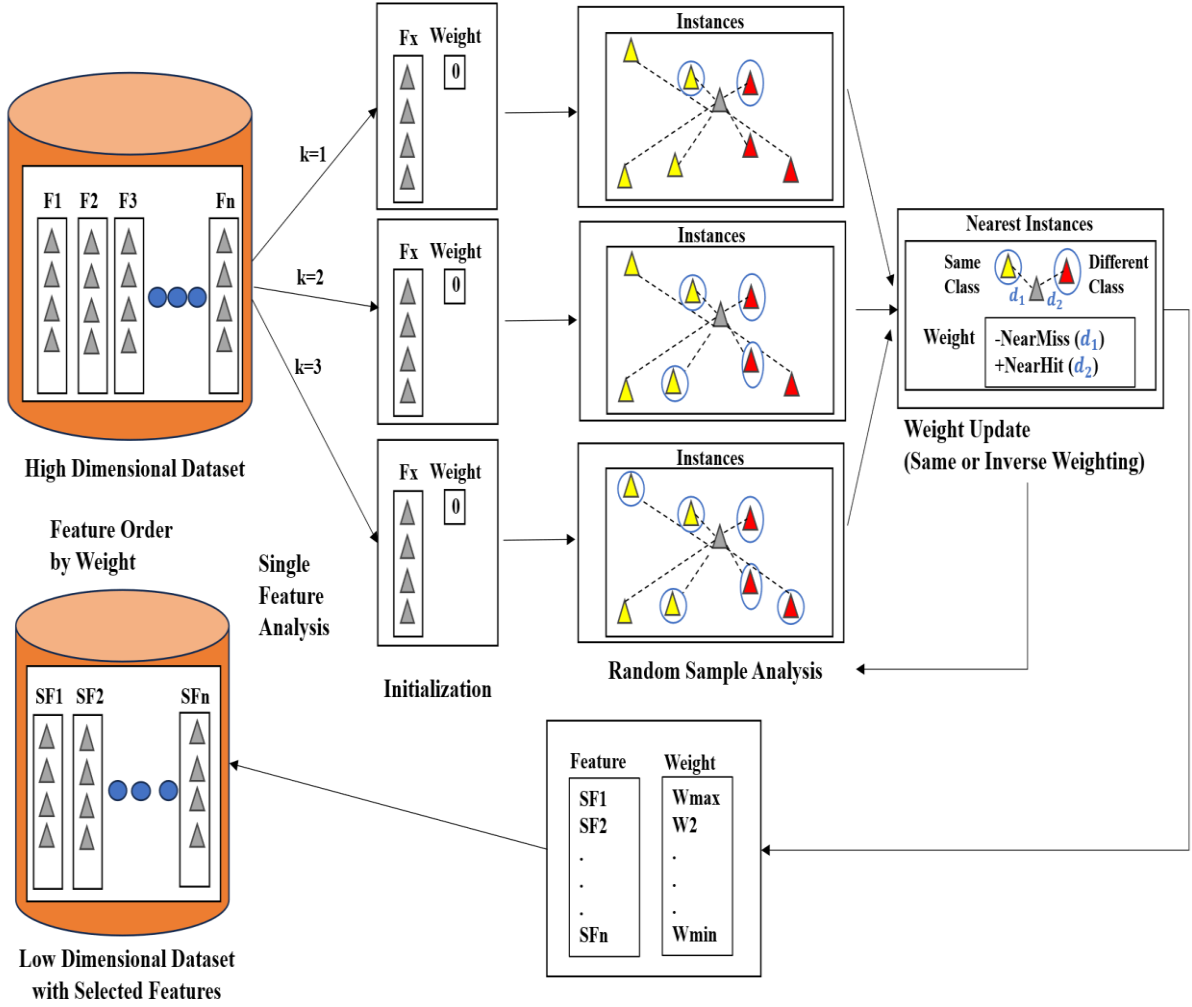


Figure 4. The sample flowchart of MW-ReliefF method for k values 1,2, and 3

Algorithm 1 summarizes the overall procedure of the proposed Multi-k Weighted ReliefF (MW-ReliefF) feature selection method. The algorithm begins by initializing the feature weights and normalizing the dataset to ensure comparability across continuous and discrete attributes. Next, it determines the weighting coefficients corresponding to each selected number of neighbors k_j according to the chosen weighting scheme (“same” or “inverse”). For every instance in the dataset, the algorithm identifies its nearest hit and miss neighbors for each k_j value, computes the average attribute differences, and updates the feature weights by aggregating the near-hit and near-miss contributions across all neighborhood sizes. This process enables MW-ReliefF to capture both local and global relevance patterns by integrating multiple neighborhood scales. Finally, the accumulated weights are normalized and

ranked, and the top-weighted features are selected to form the optimal subset T . Overall, Algorithm 1 illustrates how MW-ReliefF extends the classical ReliefF by incorporating a multi-k ensemble mechanism and adaptive weighting to achieve more robust and discriminative feature evaluation.

Algorithm 1: The Multi-k Weighted ReliefF (MW-ReliefF) for Feature Selection

Input:

The training dataset $D = (X, C)$ where $X \in \mathbb{R}^{n \times d}$ contains n samples and d features, and $C \in \mathbb{R}^n$ is,

the class label vector,

the set of neighbor sizes $K = \{k_1, k_2, \dots, k_m\}$,

the weighting strategy $W_{type} \in \{\text{'same'}, \text{'inverse'}\}$,

the number of selected features d' .

Output:

The subset of selected features T .

Initialization

1. Initialize the feature weight vector $W = [0, 0, \dots, 0]_{1 \times d}$
2. Compute the prior probability of each class $P(c) = \frac{|\{y_i=c\}|}{n}$ for all $c \in C$
3. Normalize continuous features in X using min-max scaling; for discrete features, encode categorical values.

Determine weighting coefficients

4. If $W_{type} = \text{"same"}:$

$$\text{Set } \omega_j = \frac{1}{m} \text{ for all } j = \{1, 2, \dots, m\}$$

Else if $W_{type} = \text{"inverse"}:$

$$\text{Compute } \omega_j = \frac{1/k_j}{\sum_{t=1}^m (1/k_t)} \text{ for all } j = \{1, 2, \dots, m\}$$

End if.

Iterative weight update for each instance

5. For each instance $x_i \in X$
6. For each $k_j \in K$
7. Find k_j nearest **Hit** neighbors $H_{kj}(x_i)$ (same class as x_i)
8. For each other class $c \notin C_i$ Find k_j nearest **Miss** neighbors $M_{kj}^{(c)}(x_i)$.
9. For each feature f_t where $t = \{1, 2, \dots, d\}$
10. Compute the average difference to hits:

$$\Delta_{hit}(f_t) = \frac{1}{|H_{kj}(x_i)|} \sum_{h \in H_{kj}(x_i)} |x_{i,t} - x_{h,t}|$$

11. Compute the average difference to misses:

$$\Delta_{miss}(f_t) = \sum_{c \neq C_i} \frac{P(c)}{1 - P(C_i)} \times \frac{1}{|M_{kj}^{(c)}(x_i)|} \sum_{m \in M_{kj}^{(c)}(x_i)} |x_{i,t} - x_{m,t}|$$

12. Update the feature weight increment for the current k_j :

$$\Delta W_{kj}(f_t) = -\Delta_{hit}(f_t) + \Delta_{miss}(f_t).$$

13. Accumulate the contribution according to weighting strategy:

$$W(f_t) = W(f_t) + \alpha_j \times \Delta W_{kj}(f_t).$$

14. End for

15. End for

16. End for

/ Feature ranking and selection /

17. Normalize all weights $W(f_t)$ into the range $[0, 1]$.
18. Rank all features in descending order of $W(f_t)$.

19. Select the top d' features with the highest weights to form the subset T .

20. Return T .

The proposed MW-ReliefF algorithm is not restricted to a specific set of neighborhood sizes. Instead, it is designed as a general ensemble framework capable of integrating feature importance estimates obtained from any predefined collection of neighborhood values. The neighborhood set used in this study was selected to cover both local and relatively broader neighborhood structures while avoiding excessive redundancy between consecutive neighborhood sizes. Consequently, progressively increasing odd neighborhood values were adopted to provide diverse neighborhood representations throughout the feature evaluation process.

MW-ReliefF does not require the selection of a single optimal neighborhood size prior to feature evaluation. Instead, the method uses a predefined -list to obtain feature relevance estimates at multiple neighborhood scales. In the present study, progressively increasing odd neighborhood sizes were employed to provide sufficiently distinct local representations while limiting redundancy between consecutive neighborhood sizes. The selection of the neighborhood set was further examined through a sensitivity analysis in which different -list configurations were evaluated. The resulting average classification accuracies are reported in Table 3. This analysis was performed to determine whether the effectiveness of MW-ReliefF depends strongly on a particular neighborhood configuration. The results indicate that the proposed aggregation framework remains effective across the evaluated -list configurations, suggesting that the method is not excessively sensitive to the precise choice of neighborhood sizes.

It should be emphasized that the effectiveness of MW-ReliefF does not depend on a particular combination of neighborhood sizes. Rather, its performance originates from the ensemble aggregation of multiple neighborhood scales, which reduces the sensitivity associated with selecting a single neighborhood parameter in

conventional ReliefF. Table 3 shows the obtained average accuracy with respect to different k-list combinations when MW-ReliefF method was applied to all of the experimented datasets. For example, when k-list is selected as [1-5], the selected k values become 1,2,3,4, and 5.

Table 3. Parameter Sensitivity Analysis

k-list	Average Accuracy
[1-5]	0.7648
[1-10]	0.7762
[1-20]	0.8033
[1-30]	0.7946
[1-40]	0.8183
[1-50]	0.8097

The proposed MW-ReliefF exhibited relatively stable performance across different neighborhood configurations. Although slight variations were observed, the overall classification accuracy remained consistent, demonstrating that the proposed ensemble mechanism is not highly sensitive to the exact choice of neighborhood sizes. This indicates that the improvement achieved by MW-ReliefF mainly originates from integrating multiple neighborhood scales rather than relying on a specific neighborhood configuration.

The use of multiple neighborhood sizes is motivated by the scale-dependent nature of local feature relevance. In ReliefF, the contribution of a feature is estimated from differences between an instance and its nearest hit and miss neighbors. Therefore, the selected value of determines the local scale at which feature relevance is evaluated. A small emphasizes highly local relationships and can provide a more sensitive representation of fine-grained class structure, whereas larger values incorporate a broader portion of the feature space and can capture more global relevance patterns. Consequently, relying on a single neighborhood size may make the resulting feature ranking dependent on a particular local scale. MW-ReliefF addresses this scale dependency by obtaining feature relevance estimates at multiple neighborhood sizes and aggregating these estimates into a unified feature score. From this perspective, the different values can be regarded as complementary views of the same feature space rather than independent feature

selection procedures. The aggregation therefore provides a mechanism for integrating relevance information across local scales and reducing the sensitivity of feature ranking to the choice of a single neighborhood parameter.

3.4 Computational Complexity Analysis

In this study, experiments were performed on twenty benchmark datasets containing varying numbers of features to evaluate the proposed method against several existing feature selection techniques.

The computational complexity of MW-ReliefF is primarily determined by the repeated execution of the ReliefF algorithm over multiple neighborhood sizes. Let n denote the number of instances, d the number of features, and m the number of different neighborhood sizes contained in the predefined k -list.

For a single neighborhood size, ReliefF computes the distances between instances, identifies the nearest hit and nearest miss neighbors, and updates the feature weights accordingly. Assuming a straightforward implementation based on exhaustive distance computation, the time complexity of one ReliefF execution is approximately

$$(10) \quad O(n^2 d)$$

where the pairwise distance computation dominates the overall execution time.

Unlike the conventional ReliefF algorithm, MW-ReliefF repeats the feature weighting process independently for each neighborhood size in the predefined set

$$(11) \quad K = \{k_1, k_2, \dots, k_m\}$$

Consequently, the overall computational complexity becomes

$$(12) \quad O(mn^2 d)$$

where m denotes the number of neighborhood sizes evaluated during the ensemble process. Since the weighted aggregation of the resulting feature importance vectors requires only a linear traversal of all features,

$$(13) \quad O(md)$$

its computational cost is negligible compared with the repeated nearest-neighbor search.

Similarly, the memory complexity remains largely unchanged from that of ReliefF. Besides storing the original dataset requiring

(14)

$$O(nd)$$

MW-ReliefF maintains an additional weight vector for each neighborhood size. Therefore, the total memory requirement becomes

(15)

$$O(nd + md)$$

Because the number of evaluated neighborhood sizes is typically much smaller than both the number of samples and features ($m \ll n$, $m \ll d$), the additional memory overhead introduced by MW-ReliefF is relatively small.

Although MW-ReliefF requires approximately m times more feature-weight estimation operations than the original ReliefF algorithm, this additional computation enables the algorithm to exploit complementary neighborhood information obtained from multiple scales. As demonstrated by the experimental results, the resulting improvement in feature selection robustness and classification performance compensates for the moderate increase in computational cost.

4. Findings

In this study, experiments were performed on twenty benchmark datasets containing varying numbers of features to evaluate the proposed method against several existing feature selection techniques.

4.1 Datasets

Twenty datasets obtained from the UCI Machine Learning Repository were utilized to validate the effectiveness of the proposed method. These datasets were chosen from diverse domains and exhibit varying numbers of features, instances, and classes. A detailed summary of their characteristics, including the number of features, samples, and class labels, is provided in Table 2.

4.2 Baseline and Comparative Feature Selection Methods

We compared Multi-k Weighted ReliefF (MW-ReliefF) algorithm against the baseline model that uses all available features, as well as several widely adopted Relief-based feature selection methods to evaluate its performance. All Relief-based algorithms were implemented using the scikit-rebate (skrebate) library in Python to ensure consistency and fairness in comparison.

Table 2. List of datasets

Dataset	Number of features	Number of samples	Number of Classes	Domain
Automobile	25	205	7	Vehicle
Breast Cancer	9	286	2	Health
Breast Cancer Wisconsin (Prognostic)	33	198	2	Health
Car Evaluation	6	1728	4	Vehicle
Chess (King-Rook vs. King-Pawn)	35	3196	2	Entertainment
Cylinder Bands	39	512	2	Material
Dermatology	34	366	6	Health
Glass Identification	9	214	7	Material
Hayes-Roth	4	160	3	Social Science
Horse Colic	27	368	2	Health
Ionosphere	34	351	2	Radar
Lung Cancer	56	32	3	Health
Page Blocks Classification	10	5473	5	Document
Post-Operative Patient	8	90	3	Health
SPECT Heart	22	267	2	Health
SPECTF Heart	44	267	2	Health
Tic-Tac-Toe Endgame	9	958	2	Entertainment
Congressional Voting Records	16	435	2	Voting
Statlog (German Credit Data)	20	1000	2	Finance
Statlog (Landsat Satellite)	32	6435	7	Environment

The details of the compared methods are summarized below:

- **Baseline:** In the baseline approach, all available features are used for training without any feature selection. This serves as a reference point to assess the impact of the feature selection algorithms.
- **ReliefF (Kononenko, 1994):** ReliefF is an extension of the original Relief algorithm designed to handle multi-class datasets. It estimates feature importance by considering the average difference between instances and their nearest neighbors of the same and different classes.
- **SURF (Spatially Uniform Relief) [33]:** The SURF algorithm improves upon ReliefF by eliminating the fixed parameter k for nearest neighbors. Instead, it uses all instances within a dynamically defined distance threshold to compute near-hit and near-miss scores, enhancing robustness against noisy data.
- **TuRF (Tuned ReliefF) [34]:** TuRF iteratively applies the ReliefF algorithm while progressively removing features with the lowest weights at each iteration. This iterative refinement helps to reduce redundancy and improve the accuracy of selected features.
- **MultiSURF [35]:** MultiSURF further enhances SURF by introducing a more sophisticated distance threshold that adapts to the overall distribution of samples. It better detects complex feature interactions and dependencies in high-dimensional data.

These methods collectively represent state-of-the-art Relief-based feature selection techniques, allowing for a comprehensive comparison of the proposed MW-ReliefF algorithm in terms of its ability to identify informative and non-redundant features across multiple benchmark datasets.

4.3 Performance Metrics

In this study, the performance evaluation of all feature selection methods was conducted under a unified classification framework to ensure fairness and comparability. After applying each feature selection technique, the reduced feature sets were used to train a Ridge Classifier, which was selected for its robustness against

multicollinearity and its ability to generalize effectively on high-dimensional datasets.

The classification results were assessed using three widely adopted quantitative measures: Accuracy, Precision, and Recall. Accuracy represents the overall correctness of the classifier, Precision measures the proportion of correctly predicted positive instances among all predicted positives, and Recall evaluates the ability of the model to identify all relevant positive instances.

By comparing these metrics across all evaluated feature selection approaches, the effectiveness and stability of the proposed Multi-k Weighted ReliefF (MW-ReliefF) method were systematically examined against the baseline and competing algorithms.

Table 3. Summary of experimental settings

Parameter	Value
Weight Type	Inverse
Number of Features to Select	Ranging from 2 to 30
K for ReliefF	5
k-list for MW-ReliefF	[1-40]
Distance Metric	Euclidean

Table 3 give explanation about the used parameters in the experimental setup. Each parameter is mapped to its exact value or the range of values implemented in the experiments for the datasets included in the given dataset list. In our implementations, weight type of MW-ReliefF method is selected as inverse so that the effects of feature selection iteration is inversely proportional to its k value input. The number of features to be selected among the whole features set vary in terms of dataset size. The smallest number of selected features is 2 and the highest one is 30 in our experiments using the 20 datasets. Another important input parameter is k value for ReliefF model and it is set to 5 for all experiments when applying this algorithm and its traditional derivatives (Surf, Multi-Surf, and Turf).

The experimental parameters were selected based on widely adopted practices in the ReliefF literature and to ensure a fair comparison among the evaluated methods. For the conventional ReliefF algorithm and its variants, the neighborhood parameter was fixed at $k = 5$, which is one of the most commonly adopted default settings in previous ReliefF-based studies. Using a

fixed neighborhood size also provides a consistent baseline for evaluating the proposed MW-ReliefF algorithm. The inverse weighting strategy was selected to emphasize the contributions of smaller neighborhood sizes, since closer neighbors generally provide more discriminative local information than distant neighbors. Finally, the number of selected features was adjusted according to the dimensionality of each dataset, ranging from 2 to 30 features, in order to avoid selecting an excessively small or unnecessarily large subset for datasets with substantially different numbers of attributes.

The classifier used in the experiment is Ridge classifier. This classifier initially maps binary targets to the set $\{-1, 1\}$ and formulates the task as a regression problem, optimizing the same objective function as previously described. The final class prediction is determined by the sign of the regressor's output. In the case of multiclass classification, the task is framed as a multi-output regression problem, and the class associated with the highest output value is selected as the prediction.

Although employing a (penalized) Least Squares loss for classification may appear unconventional compared to the commonly used logistic or hinge losses, in practice, their performance is often comparable. Models based on penalized least squares can achieve similar cross-validation results in terms of accuracy, precision, and recall, while also benefiting from access to alternative numerical solvers that provide

different computational efficiency characteristics.

4.4 Experimental Results

The feature selection performance of the proposed MW-ReliefF method and its comparison with other algorithms are presented in this section. Table 4 summarizes the average classification accuracies obtained using the Ridge Classifier after applying each feature selection technique. As illustrated in Table 4, Table 5, Table 6, and Table 7, MW-ReliefF achieves superior classification accuracy compared to the other evaluated methods across the majority of datasets.

Table 4 presents the classification accuracies obtained using different feature selection methods across 20 benchmark datasets. The results clearly demonstrate that the proposed Multi-k Weighted Relief (MW-ReliefF) algorithm achieved superior performance than the traditional ReliefF method and its established variants, including SURF, TURF, and MultiSURF, across most datasets.

Table 4. Classification accuracy results with base classifier as Ridge

Dataset	# of Selected Features	Base (Ridge)	ReliefF	MW-ReliefF	Surf	Turf	Multisurf
Automobile	5	0.7142	0.7619	0.8095	0.666667	0.52381	0.66
Breast Cancer	5	0.5517	0.5862	0.6551	0.551724	0.517241	0.5862
Breast Cancer Wisconsin (Prognostic)	5	0.8	0.85	0.9	0.9	0.9	0.75
Car Evaluation	3	0.6358	0.6531	0.6647	0.6531	0.6531	0.6473
Chess (King-Rook vs. King-Pawn)	5	0.9281	0.9375	0.9343	0.9343	0.9343	0.9187
Cylinder Bands	5	0.8	0.8181	0.8363	0.6363	0.6727	0.7818
Dermatology	5	0.9729	0.8918	1	0.8648	0.8648	0.8378
Glass Identification	5	0.6818	0.7272	0.7272	0.8378	0.5454	0.5454
Hayes-Roth	2	0.3125	0.375	0.4375	0.375	0.375	0.375

Horse Colic	5	0.8378	0.8108	0.8648	0.8378	0.8378	0.8378
Ionosphere	5	0.8055	0.7777	0.8333	0.8611	0.7222	0.8055
Lung Cancer	5	0.9729	0.8918	1	0.25	0.5	0.75
Page Blocks Classification	5	0.9434	0.9416	0.9489	0.9324	0.9324	0.9105
Post-Operative Patient	4	0.7777	0.8888	0.8888	0.8888	0.8888	0.8888
SPECT Heart	5	0.7037	0.7777	0.8148	0.8148	0.6666	0.8148
SPECTF Heart	5	0.7407	0.7777	0.8148	0.8148	0.8148	0.8148
Tic-Tac-Toe Endgame	7	0.6354	0.625	0.6458	0.5937	0.625	0.5937
Congressional Voting Records	10	0.9772	0.9545	1	1	1	0.9545
Statlog (German Credit Data)	5	0.77	0.81	0.83	0.77	0.77	0.77
Statlog (Landsat Satellite)	30	0.7577	0.7562	0.7608	0.7531	0.7546	0.5326
Average	-	0.7659	0.7806	0.8183	0.7468	0.7249	0.7387

Table 5. Classification accuracy results with base classifier as Support Vector Machine

Dataset	# of Selected Features	Base (SVM)	ReliefF	MW-ReliefF	Surf	Turf	Multisurf
Automobile	5	0.5714	0.7142	0.8095	0.8571	0.7619	0.7619
Breast Cancer	5	0.5517	0.6206	0.6551	0.5862	0.6206	0.6206
Breast Cancer Wisconsin (Prognostic)	5	0.8	0.8	0.8	0.8	0.8	0.8
Car Evaluation	3	0.9479	0.7803	0.8150	0.8150	0.7803	0.8150
Chess (King-Rook vs. King-Pawn)	5	0.9343	0.9281	0.9343	0.9375	0.9062	0.9375
Cylinder Bands	5	0.5272	0.7272	0.8545	0.7090	0.6727	0.7090
Dermatology	5	1	0.8918	1	0.9189	0.8378	0.8378
Glass Identification	5	0.4090	0.4090	0.4090	0.3636	0.4090	0.3636
Hayes-Roth	2	0.625	0.5	0.5	0.3125	0.5	0.3125
Horse Colic	5	0.5675	0.8108	0.8108	0.8108	0.5945	0.8108
Ionosphere	5	0.5833	0.6944	0.7222	0.6666	0.7222	0.7222
Lung Cancer	5	0.5	0.75	1	0.25	0.75	0.5
Page Blocks Classification	5	0.9105	0.9105	0.9489	0.9105	0.9105	0.9105
Post-Operative Patient	4	0.8888	0.8888	0.8888	0.8888	0.8888	0.8888
SPECT Heart	5	0.7037	0.8148	0.8148	0.8148	0.6296	0.6296
SPECTF Heart	5	0.8148	0.8148	0.8148	0.8148	0.8148	0.8148
Tic-Tac-Toe Endgame	7	0.8229	0.7604	0.7916	0.7916	0.7604	0.7604
Congressional Voting Records	10	0.9545	0.9545	0.9772	0.9772	0.9545	0.9772
Statlog (German Credit Data)	5	0.71	0.76	0.8	0.75	0.76	0.75
Statlog (Landsat Satellite)	30	0.2810	0.3105	0.3312	0.2872	0.2919	0.3105
Average	-	0.7051	0.7420	0.7838	0.7131	0.7182	0.7089

Table 6. Classification accuracy results with base classifier as K Nearest Neighborhood

Dataset	# of Selected Features	Base (KNN)	ReliefF	MW-ReliefF	Surf	Turf	Multisurf
Automobile	5	0.8095	0.7619	0.8571	0.8095	0.8571	0.7619
Breast Cancer	5	0.5517	0.6206	0.6551	0.6206	0.6551	0.6206
Breast Cancer Wisconsin (Prognostic)	5	0.8	0.8	0.85	0.75	0.9	0.8

Car Evaluation	3	0.8728	0.7803	0.9017	0.7456	0.7803	0.8150
Chess (King-Rook vs. King-Pawn)	5	0.9531	0.8906	0.9187	0.9375	0.9093	0.9375
Cylinder Bands	5	0.7454	0.7272	0.8363	0.6363	0.7090	0.7090
Dermatology	5	0.9189	0.8918	1	0.7837	0.8648	0.8378
Glass Identification	5	0.8636	0.6816	0.9090	0.7727	0.8636	0.3636
Hayes-Roth	2	0.5625	0.5625	0.6875	0.3125	0.5625	0.3125
Horse Colic	5	0.6486	0.8108	0.8108	0.8108	0.8108	0.8108
Ionosphere	5	0.8055	0.7777	0.8611	0.8333	0.75	0.7222
Lung Cancer	5	0.5	0.5	0.75	0.5	0.75	0.5
Page Blocks Classification	5	0.9726	0.9177	0.9762	0.9671	0.9762	0.9105
Post-Operative Patient	4	0.7777	0.5555	0.8888	0.8888	0.5555	0.8888
SPECT Heart	5	0.6666	0.6296	0.8148	0.5925	0.7037	0.6296
SPECTF Heart	5	0.7037	0.8148	0.8518	0.8518	0.7777	0.8148
Tic-Tac-Toe Endgame	7	0.8229	0.7604	0.875	0.7812	0.75	0.7604
Congressional Voting Records	10	0.9318	0.9545	1	0.9772	0.9545	0.9772
Statlog (German Credit Data)	5	0.64	0.64	0.78	0.73	0.74	0.75
Statlog (Landsat Satellite)	30	0.9021	0.8928	0.9037	0.8959	0.9021	0.3105
Average	-	0.7724	0.7485	0.8563	0.7598	0.7886	0.7116

Table 7. Classification accuracy results with base classifier as Random Forest

Dataset	# of Selected Features	Base (RF)	ReliefF	MW-ReliefF	Surf	Turf	Multisurf
Automobile	5	0.8571	0.7619	0.9047	0.8571	0.8095	0.8095
Breast Cancer	5	0.6896	0.6206	0.7241	0.6206	0.6206	0.6551
Breast Cancer Wisconsin (Prognostic)	5	0.85	0.8	0.9	0.85	0.85	0.85
Car Evaluation	3	0.9537	0.7514	0.9306	0.8381	0.7514	0.8381
Chess (King-Rook vs. King-Pawn)	5	0.9718	0.9281	0.9281	0.9375	0.9093	0.9375
Cylinder Bands	5	0.8727	0.7272	0.9090	0.6727	0.7272	0.7818
Dermatology	5	1	0.8648	1	0.7567	0.8378	0.8378
Glass Identification	5	0.9545	0.8636	0.9090	0.8181	0.8181	0.7727
Hayes-Roth	2	0.5625	0.4375	0.5625	0.3125	0.5	0.3125
Horse Colic	5	0.8378	0.8378	0.8648	0.8378	0.8108	0.8108
Ionosphere	5	0.9166	0.8888	0.9444	0.8611	0.8055	0.9166
Lung Cancer	5	0.75	0.75	1	0.25	0.5	0.25
Page Blocks Classification	5	0.9762	0.9762	0.9799	0.9762	0.9762	0.9781
Post-Operative Patient	4	0.6666	0.4444	0.8888	0.6666	0.6666	0.5555
SPECT Heart	5	0.7407	0.6296	0.7037	0.5925	0.6296	0.6296
SPECTF Heart	5	0.8148	0.7777	0.8888	0.8148	0.7777	0.7777
Tic-Tac-Toe Endgame	7	0.9583	0.8229	0.9270	0.8437	0.75	0.8437
Congressional Voting Records	10	0.9772	0.9090	1	0.9772	0.9090	0.9545
Statlog (German Credit Data)	5	0.8	0.79	0.83	0.72	0.76	0.71
Statlog (Landsat Satellite)	30	0.9177	0.9130	0.9239	0.9177	0.9145	0.9114
Average	-	0.8533	0.7747	0.8859	0.7560	0.7661	0.7566

Table 5 presents the classification accuracy results obtained using Support Vector Machine (SVM) as the base classifier. The results indicate that the proposed MW-ReliefF method maintains its effectiveness when the underlying classifier is changed from Ridge Classifier to SVM. MW-ReliefF achieves the highest average accuracy among all evaluated feature selection approaches, with an accuracy of 0.7838, compared with 0.7051 for the SVM baseline, 0.7420 for ReliefF, 0.7131 for SURF, 0.7182 for TuRF, and 0.7089 for MultiSURF. MW-ReliefF obtains the best accuracy on several datasets, including Breast Cancer, Cylinder Bands, Ionosphere, Lung Cancer, Page Blocks Classification, Congressional Voting Records, Statlog (German Credit Data), and Statlog (Landsat Satellite). Particularly notable improvements are observed for Cylinder Bands, where the accuracy increases from 0.5272 for the baseline to 0.8545 with MW-ReliefF, and for Lung Cancer, where MW-ReliefF reaches an accuracy of 1.00 compared with 0.50 for the baseline. More importantly, the improvement in the mean accuracy demonstrates that aggregating feature relevance information across multiple neighborhood sizes can remain beneficial when the selected feature subset is subsequently evaluated by an SVM classifier.

Table 6 reports the results obtained when K-Nearest Neighbors (KNN) is employed as the base classifier. The results provide further evidence that the performance of MW-ReliefF is not dependent on the Ridge Classifier used in the initial experiments. MW-ReliefF achieves the highest mean classification accuracy of 0.8563, substantially exceeding the KNN baseline (0.7724) and the other feature selection methods, including ReliefF (0.7485), SURF (0.7598), TuRF (0.7886), and MultiSURF (0.7116). MW-ReliefF obtains the highest accuracy on 18 of the 20 datasets, while TuRF achieves the highest result on Breast Cancer Wisconsin (Prognostic) and the baseline KNN performs best on Chess (King-Rook vs. King-Pawn). This distribution of the best-performing results provides a particularly strong indication of

the consistency of MW-ReliefF under a distance-based classifier.

The improvements are also substantial on several individual datasets. For example, MW-ReliefF increases the accuracy on Glass Identification from 0.8636 for the baseline to 0.9090, on Hayes-Roth from 0.5625 to 0.6875, on SPECT Heart from 0.6666 to 0.8148, and on Tic-Tac-Toe Endgame from 0.8229 to 0.8750. On Congressional Voting Records, MW-ReliefF reaches perfect accuracy (1.00), compared with 0.9318 for the baseline and 0.9545 for ReliefF. The improvement in average accuracy, together with its strong performance across datasets with different characteristics, indicates that the proposed multi-neighborhood weighting strategy can provide informative reduced feature representations for KNN classification. Nevertheless, the results also show that feature selection does not universally improve every dataset, as demonstrated by the Chess dataset, where the baseline remains superior. This observation supports a more cautious interpretation that MW-ReliefF generally provides a favorable feature subset rather than guaranteeing improvement for every individual dataset.

Table 7 presents the results obtained using Random Forest (RF) as the base classifier. The RF experiments further demonstrate the robustness of the proposed method across different learning paradigms. The RF results also reveal an important distinction between the proposed method and the competing approaches. While RF itself is already a strong classifier and therefore produces high baseline accuracies on several datasets, MW-ReliefF is still able to provide additional gains on many of these datasets. For example, the accuracy increases from 0.8571 to 0.9047 on Automobile, from 0.8727 to 0.9090 on Cylinder Bands, from 0.9166 to 0.9444 on Ionosphere, and from 0.9177 to 0.9239 on Statlog (Landsat Satellite). On the other hand, the baseline remains superior for Car Evaluation, Chess, and SPECT Heart, illustrating that the effect of feature selection can depend on the interaction between the selected feature subset and the classifier. Overall, the highest mean accuracy among

all methods and the preservation of strong performance with RF provide additional empirical support for the classifier-independent usefulness of MW-ReliefF. Taken together with the SVM and KNN experiments, these findings indicate that the proposed method can produce competitive feature subsets across linear, distance-based, and ensemble-based classification settings.

Compared to the baseline (using all features without feature selection), MW-ReliefF achieves a notable increase in predictive accuracy while utilizing a significantly reduced subset of features. This improvement indicates that MW-ReliefF effectively identifies and preserves the most informative attributes, reducing data dimensionality without compromising and often enhancing classification performance. For example, in datasets such as Dermatology, Congressional Voting Records, and Lung Cancer, the MW-ReliefF method achieves a perfect or near-perfect classification accuracy (1.0 or close), surpassing all other competing algorithms. Similarly, in Automobile, Cylinder Bands, and SPECT Heart datasets, the proposed approach yields substantial performance gains over both ReliefF and SURF-based models.

The superior performance of MW-ReliefF can be attributed to its multi-scale neighborhood exploration mechanism, which integrates information from multiple k -values simultaneously. Unlike ReliefF, which operates under a fixed neighborhood size, MW-ReliefF adaptively balances the influence of local and global structures within the data. This enables the algorithm to better capture both fine-grained feature dependencies and broader class separability patterns. Furthermore, the inverse weighting mechanism ensures that closer neighbors (smaller k -values) have a stronger influence on feature importance estimation, enhancing robustness in datasets with overlapping or imbalanced classes.

In contrast, conventional algorithms such as SURF and TURF rely on static distance thresholds or uniform neighborhood weighting, which may overlook subtle but significant inter-feature relationships. While

MultiSURF improves upon SURF by extending the neighborhood definition, it often exhibits instability in high-dimensional or noisy datasets, leading to fluctuating accuracy levels, as observed in Glass Identification and Statlog (Landsat Satellite) datasets. MW-ReliefF, by incorporating weighted multi-scale analysis, mitigates these issues and yields more stable and generalizable feature rankings across diverse domains.

The superior performance of MW-ReliefF can be attributed to its neighborhood-level ensemble mechanism. Existing Relief variants generally estimate feature importance using a single neighborhood configuration, making their performance dependent on the selection of an appropriate k value. In contrast, MW-ReliefF aggregates feature importance information obtained from multiple neighborhood scales. This design reduces the sensitivity to neighborhood selection and enables the algorithm to capture both local and relatively global feature interactions simultaneously. As demonstrated by the experimental results across twenty benchmark datasets, this strategy improves the robustness of feature ranking without modifying the fundamental ReliefF framework.

Overall, the experimental outcomes confirm that MW-ReliefF achieves the highest average accuracy across the 20 benchmark datasets, demonstrating its effectiveness as a robust, adaptive, and scalable feature selection framework. The method's ability to integrate local and global neighbor information through multi- k weighting not only enhances classification accuracy but also contributes to better feature interpretability and model efficiency, establishing MW-ReliefF as a strong candidate for real-world high-dimensional data applications.

The observed performance improvement of MW-ReliefF can be explained by the way feature relevance is estimated across different neighborhood scales. In conventional ReliefF, the relevance score of a feature is determined using a single predefined value of k , which makes the resulting feature ranking dependent on one

particular neighborhood configuration. A small k emphasizes highly local relationships and can better preserve fine-grained class-boundary information, whereas a larger k incorporates a broader set of neighboring instances and therefore provides a less localized representation of feature relevance. MW-ReliefF combines the feature relevance contributions obtained at multiple k values instead of relying on a single neighborhood scale. Consequently, a feature that remains informative across different neighborhood configurations can accumulate support from multiple local structures, while a feature whose relevance is strongly dependent on a particular neighborhood scale has a reduced influence on the aggregated ranking. The weighting mechanism further controls the contribution of each neighborhood scale, with inverse weighting assigning greater importance to smaller k values while retaining information from broader neighborhoods. Thus, the improvement observed in the experiments can be interpreted as a consequence of reducing the dependence of feature ranking on a single neighborhood configuration and integrating complementary local and broader neighborhood information. This interpretation is also consistent with the experimental results obtained using different classifiers, where MW-ReliefF achieved the highest average accuracy for SVM, KNN, and Random Forest, with accuracies of 0.7838, 0.8563, and 0.8859, respectively.

The effectiveness of a feature selection method should not be inferred from a single dataset or a single classification model although accuracy is used as the primary performance measure in the experimental comparison. Therefore, the evaluation in this study was conducted across 20 benchmark datasets and repeated using four different base classifiers, namely Ridge Classifier, Support Vector Machine, K-Nearest Neighbors, and Random Forest. In addition, non-parametric statistical analysis was performed using the Wilcoxon signed-rank test to determine whether the observed differences between MW-ReliefF and the competing feature selection methods were statistically significant. Thus, the

effectiveness of MW-ReliefF was assessed through its performance consistency across heterogeneous datasets, classifier settings, and statistical comparisons rather than solely through an isolated difference in Accuracy.

Non-parametric statistical tests were performed for comparing multiple algorithms over multiple datasets to determine whether the observed improvements of MW-ReliefF over the competing feature selection methods are statistically significant. Since the null hypothesis was rejected, pairwise comparisons between MW-ReliefF and each competing method were subsequently conducted using the Wilcoxon signed-rank test. A significance level of $\alpha = 0.05$ was adopted throughout all statistical analyses.

The obtained p-values indicate whether the observed improvements of MW-ReliefF are statistically significant rather than being caused by random variations among datasets as can be seen in Table 8. Wilcoxon signed-rank tests demonstrated that MW-ReliefF significantly outperformed the baseline ReliefF algorithm as well as the remaining competing methods on the majority of benchmark datasets. These results support that the observed performance improvements are not due to random variations but originate from the proposed multi-neighborhood weighted aggregation mechanism.

Table 8. Wilcoxon signed-rank test comparison.

Compared Models	p-value
MW-ReliefF vs ReliefF	0.012
MW-ReliefF vs SURF	0.001
MW-ReliefF vs TuRF	0.0003
MW-ReliefF vs MultiSURF	0.004
MW-ReliefF vs Baseline	0.009

The effectiveness of a feature selection method cannot be assessed solely from the absolute classification accuracy obtained by a single classifier. Therefore, the evaluation in this study considers multiple complementary aspects of feature selection performance. First, classification accuracy is used as the primary predictive criterion to quantify whether the selected feature subsets preserve or improve the discriminative information contained in the original feature space. Second, the robustness of the feature selection method is examined across

different classification algorithms, including Ridge Classifier, Support Vector Machine, K-Nearest Neighbors, and Random Forest, to determine whether the observed improvements are dependent on a particular classifier. Third, the generality of the proposed method is evaluated across 20 datasets with different characteristics and numbers of selected features. Finally, statistical significance is assessed using the Wilcoxon signed-rank test to determine whether the performance differences between MW-ReliefF and the competing feature selection methods are statistically significant. In addition, a sensitivity analysis of different neighborhood configurations is used to examine the stability of MW-ReliefF with respect to the selected k -lists. These complementary evaluations provide a broader assessment of predictive effectiveness, classifier independence, cross-dataset robustness, statistical significance, and parameter sensitivity.

Quantitative comparison of the obtained results was performed at both the dataset level and the aggregate level. For each of the 20 benchmark datasets, the classification accuracy obtained after feature selection by MW-ReliefF was directly compared with

those obtained using ReliefF, SURF, TURF, and MultiSURF. In addition to individual dataset results, the mean accuracy across all datasets was calculated for each feature selection method, providing an overall measure of comparative performance. The experiments were also repeated using different base classifiers to examine whether the observed performance differences were dependent on the underlying classifier. This reporting strategy enables the effectiveness of MW-ReliefF to be evaluated not only from individual accuracy values but also from its aggregate performance across heterogeneous datasets and classification models.

The Wilcoxon signed-rank test was employed for pairwise comparisons between MW-ReliefF and each competing feature selection method to complement the descriptive comparison. The test was applied across the datasets using a significance level of $\alpha = 0.05$. A p -value below 0.05 was considered evidence of a statistically significant difference between the paired methods. Thus, the evaluation considers both the magnitude of the observed classification accuracies and the statistical significance of the differences.

Table 9. Classification F-Score results with base classifier as Ridge

Dataset	# of Selected Features	Base (Ridge)	ReliefF	MW-ReliefF	Surf	Turf	Multisurf
Automobile	5	0.6930	0.7407	0.7983	0.6454	0.5026	0.6388
Breast Cancer	5	0.5132	0.5477	0.6266	0.5132	0.4787	0.5477
Breast Cancer Wisconsin	5	0.7680	0.8180	0.8780	0.8680	0.8680	0.7180
Car Evaluation	3	0.6078	0.6251	0.6467	0.6251	0.6251	0.6193
Chess (King-Rook vs. King-Pawn)	5	0.9134	0.9228	0.9296	0.9196	0.9196	0.9040
Cylinder Bands	5	0.7853	0.8034	0.8316	0.6216	0.6580	0.7671
Dermatology	5	0.9612	0.8801	0.9983	0.8531	0.8531	0.8261
Glass Identification	5	0.6458	0.6912	0.6912	0.8018	0.5094	0.5094
Hayes-Roth	2	0.2845	0.3470	0.3470	0.4195	0.3470	0.3470
Horse Colic	5	0.8066	0.7796	0.8436	0.8066	0.8066	0.8066
Ionosphere	5	0.7949	0.7671	0.8049	0.8505	0.7116	0.7949
Lung Cancer	5	0.9338	0.8527	0.9709	0.2109	0.4609	0.7109
Page Blocks Classification	5	0.9084	0.9066	0.9239	0.8974	0.8974	0.8755
Post-Operative Patient	4	0.7613	0.8724	0.8724	0.8724	0.8724	0.8724
SPECT Heart	5	0.6882	0.7622	0.8093	0.7993	0.6511	0.7993
SPECTF Heart	5	0.7252	0.7622	0.8093	0.7993	0.7993	0.7993
Tic-Tac-Toe Endgame	7	0.6163	0.6059	0.6367	0.5746	0.6059	0.5746
Congressional Voting Records	10	0.9515	0.9288	0.9743	0.9743	0.9743	0.9288
Statlog (German Credit Data)	5	0.7470	0.7870	0.7470	0.8070	0.7470	0.7470
Statlog (Landsat Satellite)	30	0.7390	0.7375	0.7521	0.7344	0.7359	0.5139
Average	-	0.7422	0.7569	0.7963	0.7290	0.7012	0.7150

Table 9 presents the F-Score results obtained using Ridge as the base classifier. Overall, MW-ReliefF demonstrates strong and competitive performance, achieving the highest average F-Score of 0.7963, compared with 0.7422 for the baseline Ridge classifier, 0.7569 for ReliefF, 0.7290 for SURF, 0.7012 for TURF, and 0.7150 for MultiSURF. MW-ReliefF achieves the highest F-Score on most datasets, although SURF performs better on a few datasets such as Glass Identification, Hayes-Roth, and Ionosphere. These results provide additional quantitative evidence that the performance advantage of MW-ReliefF is not limited to accuracy alone and is also reflected in the balance between precision and recall.

5. Conclusion

This study introduced Multi-k Weighted ReliefF (MW-ReliefF), an enhanced feature selection algorithm that extends the classical ReliefF method by incorporating multiple neighborhood sizes through a weighted ensemble mechanism. The core innovation of MW-ReliefF lies in its ability to capture both local and global relationships among features by simultaneously considering multiple k-values rather than relying on a single neighborhood scale. By integrating inverse or uniform weighting strategies, MW-ReliefF effectively balances the influence of smaller and larger neighborhoods, allowing for a more nuanced and stable evaluation of feature relevance.

Comprehensive experiments conducted on twenty benchmark datasets demonstrated that MW-ReliefF achieved superior performance on most of the evaluated benchmark datasets. The superior classification accuracies obtained using the Ridge Classifier validate the method's robustness and its capacity to identify highly discriminative and non-redundant features. Particularly, the inverse weighting strategy proved effective in enhancing sensitivity to locally informative features while maintaining generalization across diverse data distributions.

The results confirm that MW-ReliefF not only improves classification performance but also contributes to more interpretable and

computationally efficient models by reducing data dimensionality without sacrificing accuracy. Its adaptability across both binary and multi-class problems underscores its potential applicability to a wide range of machine learning domains, including bioinformatics, finance, and signal processing.

Although MW-ReliefF was evaluated on 20 benchmark datasets with different numbers of instances, features, and class structures, the present experimental study does not specifically target extremely large or ultra-high-dimensional datasets. The proposed method remains within the filter-based feature selection framework and therefore does not require repeated training of a downstream classifier during feature evaluation. However, unlike conventional ReliefF, MW-ReliefF evaluates neighborhood relationships for multiple k values, which introduces additional computational overhead as the number of neighborhood settings increases. Consequently, the scalability of MW-ReliefF should be considered in relation to both dataset size and the number of k values used. A systematic evaluation on substantially larger and higher-dimensional datasets, such as large-scale text, genomic, and other high-dimensional datasets, is therefore required to establish the practical scalability and broader generalizability of the proposed method.

Future research will explore the integration of MW-ReliefF into hybrid feature selection frameworks that combine filter and wrapper paradigms, as well as its adaptation for deep learning-based representations. Additionally, extending the algorithm to handle streaming or dynamic data environments represents a promising direction for further investigation.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal

relationships that could have appeared to influence the work reported in this paper.

References

- Chen, J., Song, A., & Zhang, W. (2011). A novel hybrid gene selection approach based on ReliefF and FCBF. *International Journal of Digital Content Technology and Its Applications*, 5, 404-411.
- Draper, B., Kaito, C., & Bins, J. (2003). Iterative Relief. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop* (pp. 62-62). Madison, WI, USA.
<https://doi.org/10.1109/CVPRW.2003.10040>
- Eppstein, M. J., & Haake, P. (2008). Very large scale ReliefF for genome-wide association analysis. In *2008 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology* (pp. 112-119). Sun Valley, ID, USA.
<https://doi.org/10.1109/CIBCB.2008.4681570>
- Granizo-Mackenzie, D., & Moore, J. H. (2013). Multiple threshold spatially uniform ReliefF for the genetic analysis of complex human diseases. In L. Vanneschi, W. S. Bush, & M. Giacobini (Eds.), *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (EvoBIO 2013)* (Lecture Notes in Computer Science, Vol. 7833, pp. 1-10). Berlin, Heidelberg: Springer.
https://doi.org/10.1007/978-3-642-37195-0_1
- Greene, C. S., Penrod, N. M., Kiralis, J., & Moore, J. H. (2009). Spatially uniform ReliefF (SURF) for computationally-efficient filtering of gene-gene interactions. *BioData Mining*, 2, 5.
<https://doi.org/10.1186/1756-0381-2-5>
- Karizaki, A. A., & Tavassoli, M. (2019). A novel hybrid feature selection based on ReliefF and binary dragonfly for high dimensional datasets. In *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 300-304). Mashhad, Iran.
<https://doi.org/10.1109/ICCKE48569.2019.8964852>
- Kılıç, İ., & Yalçın, N. (2025). A novel hybrid methodology based on transfer learning, machine learning, and ReliefF for chickpea seed variety classification. *Applied Sciences*, 15(3), 1334.
<https://doi.org/10.3390/app15031334>
- Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In D. Sleeman & P. Edwards (Eds.), *Machine Learning Proceedings 1992* (pp. 249-256). San Francisco, CA: Morgan Kaufmann.
<https://doi.org/10.1016/B978-1-55860-247-2.50037-1>
- Kira, K., & Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)* (pp. 129-134). Menlo Park, CA: AAAI Press.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 273-324.
[https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
- Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. In F. Bergadano & L. De Raedt (Eds.), *Machine Learning: ECML-94* (Lecture Notes in Computer Science, Vol. 784, pp. 171-182). Berlin, Heidelberg: Springer.
https://doi.org/10.1007/3-540-57868-4_57
- Kononenko, I., Šimec, E., & Robnik-Šikonja, M. (1997). Overcoming the myopia of inductive learning algorithms with RELIEFF. *Applied Intelligence*, 7, 39-55.
<https://doi.org/10.1023/A:1008280620621>
- Le, T. T., Urbanowicz, R. J., Moore, J. H., & McKinney, B. A. (2019). STatistical inference Relief (STIR) feature selection. *Bioinformatics*, 35(8), 1358-1365.
<https://doi.org/10.1093/bioinformatics/bty788>
- Liu, X., Zhang, Y., Zhang, L., & Yang, Y. (2020). A novel process monitoring method based on dynamic related ReliefF-SFA method. *IEEE Access*, 8, 41673-41683.
<https://doi.org/10.1109/ACCESS.2020.2977225>
- McKinney, B. A., Reif, D. M., White, B. C., Crowe, J. E. Jr., & Moore, J. H. (2007). Evaporative cooling feature selection for genotypic data involving interactions. *Bioinformatics*, 23(16), 2113-2120.
<https://doi.org/10.1093/bioinformatics/btm253>
- McKinney, B. A., White, B. C., Grill, D. E., Li, P. W., Kennedy, R. B., Poland, G. A., & Oberg, A. L. (2013). ReliefSeq: A gene-wise adaptive-k nearest-neighbor feature

- selection tool for finding gene-gene interactions and main effects in mRNA-Seq gene expression data. *PLoS ONE*, 8(12), e81527. <https://doi.org/10.1371/journal.pone.0081527>
- Moore, J. H., & White, B. C. (2007). Tuning ReliefF for genome-wide genetic analysis. In E. Marchiori, J. H. Moore, & J. C. Rajapakse (Eds.), *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (EvoBIO 2007)* (Lecture Notes in Computer Science, Vol. 4447, pp. 166-175). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-71783-6_15
- Park, H., & Kwon, H. C. (2007). Extended Relief algorithms in instance-based feature filtering. In *Proceedings of the Sixth International Conference on Advanced Language Processing and Web Information Technology (ALPIT 2007)* (pp. 123-128). Luoyang, China. <https://doi.org/10.1109/ALPIT.2007.18>
- Peker, M., Arslan, A., Şen, B., Çelebi, F. V., & But, A. (2015). A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF+RF). In *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)* (pp. 1-8). Madrid, Spain. <https://doi.org/10.1109/INISTA.2015.7276748>
- Stokes, M. E., & Visweswaran, S. (2012). Application of a spatially-weighted Relief algorithm for ranking genetic predictors of disease. *BioData Mining*, 5, 20. <https://doi.org/10.1186/1756-0381-5-20>
- Sun, L., Kong, X., Xu, J., Xue, Z., Zhai, R., & Zhang, S. (2019). A hybrid gene selection method based on ReliefF and ant colony optimization algorithm for tumor classification. *Scientific Reports*, 9(1), 8978. <https://doi.org/10.1038/s41598-019-45442-9>
- Sun, Y. (2007). Iterative RELIEF for feature weighting: Algorithms, theories, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 1035-1051. <https://doi.org/10.1109/TPAMI.2007.1093>
- Sun, Y., Todorovic, S., & Goodison, S. (2010). Local-learning-based feature selection for high-dimensional data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), 1610-1626. <https://doi.org/10.1109/TPAMI.2009.187>
- Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, 85, 189-203. <https://doi.org/10.1016/j.jbi.2018.07.014>
- Uzun, B., & Ballı, S. (2022). A novel method for intrusion detection in computer networks by identifying multivariate outliers and ReliefF feature selection. *Neural Computing and Applications*, 34, 17647-17662. <https://doi.org/10.1007/s00521-022-07490-3>
- Yang, Y., Chen, X., Zhang, Y., & Liu, X. (2019). A novel decentralized weighted ReliefF-PCA method for fault detection. *IEEE Access*, 7, 140478-140487. <https://doi.org/10.1109/ACCESS.2019.2943772>
- Zhang, B., Li, Y., & Chai, Z. (2022). A novel random multi-subspace based ReliefF for feature selection. *Knowledge-Based Systems*, 252, 109400. <https://doi.org/10.1016/j.knosys.2022.109400>
- Zhang, L. X., Wang, J. X., Zhao, Y. N., & Yang, Z. H. (2003). A novel hybrid feature selection algorithm: Using ReliefF estimation for GA-Wrapper search. In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics* (pp. 380-384). Xi'an, China. <https://doi.org/10.1109/ICMLC.2003.1259764>
- Zhang, M., Wang, J. S., Hou, J. N., et al. (2023). RG-NBEO: A ReliefF guided novel binary equilibrium optimizer with opposition-based S-shaped and V-shaped transfer functions for feature selection. *Artificial Intelligence Review*, 56, 6509-6556. <https://doi.org/10.1007/s10462-022-10317-0>
- Zhang, Y., Wang, S., Leng, Y., Shan, Z., Jin, A., Guo, Z., Lin, C., Zhang, J., & Zhang, S. (2023). A novel process monitoring method of heat treatment furnace based on ReliefF-SFA. *Proceedings of SPIE*, 12636, 126360H. <https://doi.org/10.1117/12.2674302>