Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.078927

ARTICLE

An Efficient Feature Selection with an Enhanced Supervised Term-Weighting Scheme in Multi-Class Text Classification

Osamah Mohammed Alyasiri[1,2] **and Yu-N Cheah**[1,*]

[1]School of Computer Sciences, Universiti Sains Malaysia, George Town, Penang, Malaysia

[2]Information Technology Engineering Department, Polytechnic College of Karbala, Al-Furat Al-Awsat Technical University, Karbala, Iraq

*Corresponding Author: Yu-N Cheah. Email: yncheah@usm.my

Received: 10 January 2026; Accepted: 12 March 2026; Published: 09 April 2026

ABSTRACT: Term weighting scheme and feature selection are two fundamental components in text classification (TC) systems, particularly in high-dimensional, multi-class, and imbalanced settings. Term weighting schemes aim to improve document representation by emphasizing discriminative terms across classes, while feature selection (FS) seeks to reduce dimensionality, eliminate irrelevant and redundant features, and enhance classification efficiency and effectiveness. However, most existing studies focus on FS independently of the term-weighting strategy used during document representation, thereby limiting the potential benefits of their interaction. This study addresses this gap by pursuing two main objectives. First, it employs an enhanced supervised term-weighting scheme, namely MTF-MICF, to construct a more stable and class-discriminative document representation, especially for imbalanced data. Second, it investigates the effectiveness of integrating this scheme with a filter-based FS approach using Information Gain (IG) at various levels of dimensionality reduction to assess the contribution of enhanced term weighting to the FS process. Extensive experiments were conducted across 19 benchmark multi-class text datasets. The performance was evaluated using F1-score and classification accuracy with the three prominent classifiers (MNB, SVM, and LR). The experimental results demonstrate that the proposed approach consistently outperforms conventional methods, achieving significant and stable improvements in both representation quality and classification performance. These findings confirm that enhanced supervised term weighting can serve as an effective supporting mechanism for FS in high-dimensional TC tasks.

KEYWORDS: Text classification; text categorization; feature selection; term weighting scheme; supervised term-weighting scheme; filter feature selection

1 Introduction

In real-world scenarios, textual data is inherently diverse, inconsistent, and unstructured, which makes its analysis particularly challenging for traditional computational methods. Moreover, most textual datasets are extremely high-dimensional, as each document may generate thousands or even tens of thousands of unique terms [1,2]. This high dimensionality, combined with the presence of irrelevant, redundant, and noisy features, severely affects the efficiency and generalization capability of machine learning algorithms. The abundance of uninformative features not only obscures meaningful patterns in data but also increases the risk of overfitting and degrades classification performance.

High-dimensional feature spaces also impose substantial computational burdens. Irrelevant features distract the learning process from informative patterns, redundant features unnecessarily enlarge the feature space without adding new information, and noisy features introduce random fluctuations that mislead learning algorithms, resulting in unstable models and higher error rates [3–5]. These challenges become even more critical in multi-class TC settings, where documents must be assigned to one of many possible categories under complex and often imbalanced data distributions.

TC has therefore become an essential supervised learning paradigm for organizing and managing the rapidly growing volume of unstructured textual data. Its main objective is to automatically and systematically assign documents to predefined categories or topics [6,7]. A typical TC pipeline consists of several fundamental stages: text preprocessing, feature extraction (document representation), dimensionality reduction through FS, classification, and evaluation [8–11].

Among these stages, feature extraction and FS are widely recognized as two of the most critical components, as they directly determine the quality of the input representation provided to the learning algorithm. Feature extraction in TC is commonly performed using the bag-of-words (BoW) and the Vector Space (VSM) models, where documents are represented as vectors of term frequencies [12]. To enhance this representation, term-weighting schemes are employed to assign importance scores to terms, thereby reflecting their contribution to distinguishing between classes [13–15].

Term-weighting schemes can be broadly categorized into unsupervised and supervised approaches [16–18]. Unsupervised schemes, such as Term Frequency-Inverse Document Frequency (TF-IDF) or their variants, assign weights based solely on term distribution within documents and across the corpus, without exploiting class label information. While computationally simple, these methods often fail to adequately capture the discriminative power of terms in multi-class and imbalanced settings. In contrast, supervised term-weighting schemes explicitly incorporate class information to emphasize terms that are more discriminative across categories, yielding more informative, task-oriented document representations.

Motivated by this observation, this study adopts an enhanced supervised term-weighting scheme, namely Modified Term Frequency-Modified Inverse Class Frequency (MTF-MICF), which extends the traditional Term Frequency-Inverse Class Frequency (TF-ICF) formulation by improving term frequency normalization and class-discriminative weighting. This scheme is specifically designed to mitigate document-length bias, preserve frequency information, and enhance inter-class separability, making it particularly suitable for multi-class imbalanced TC.

Despite the benefits of advanced term-weighting schemes, the resulting feature space often remains extremely large, containing tens or hundreds of thousands of features, many of which are still irrelevant or redundant for the learning task. Such non-informative features introduce noise, increase computational cost, and degrade model performance [19,20]. To address this issue, FS is commonly employed as a dimensionality reduction technique to identify the most informative subset of features.

FS methods are generally classified into filter, wrapper, embedded, and hybrid approaches [8]. Filter methods evaluate features using statistical or information-theoretic criteria independently of the classifier, offering high computational efficiency and scalability to large datasets [10]. Wrapper methods utilize a learning algorithm to evaluate feature subsets, often achieving higher accuracy but at the cost of significantly higher computational complexity [21]. Embedded methods integrate FS into the classifier's training process, while hybrid methods aim to combine the advantages of filter and wrapper methods [8,22].

In this study, a filter-based approach using IG is adopted due to its simplicity, efficiency, independence from classifiers, and proven effectiveness in high-dimensional TC problems. Moreover, IG provides a stable and interpretable criterion for ranking features based on their discriminative power across classes, making it particularly suitable for large-scale multi-class datasets.

Unlike most existing studies, which treat term weighting and feature selection as independent stages, this work explicitly investigates their interaction by positioning the enhanced supervised term-weighting scheme as a supporting mechanism for FS. Where term-weighting schemes assign importance scores to all terms, while FS methods further rank and filter them to retain only the most discriminative terms for document representation. Specifically, this study aims to demonstrate that improving document representation quality via MTF-MICF can significantly enhance the effectiveness of subsequent filter-based FS across different levels of dimensionality reduction. The proposed framework is extensively evaluated on multiple benchmark multi-class datasets using F1-score and classification accuracy, with results confirming its effectiveness and generalization. The contributions of this study are as follows:

- Introduces a unified framework that explicitly integrates an enhanced supervised term-weighting scheme (i.e., MTF-MICF) with filter-based FS, allowing term weighting to function as a supporting mechanism for improving FS effectiveness rather than treating the two stages independently.
- Employs the MTF-MICF scheme to enhance document representation quality by strengthening class discrimination, mitigating document-length bias, and preserving meaningful term frequency information, particularly in multi-class imbalanced settings.
- Provides a comprehensive experimental investigation of how enhanced supervised term weighting influences the effectiveness of IG-based FS across multiple dimensionality levels, while further validating the results using additional filter-based methods, namely ReliefF and Chi-square.
- Conducts an extensive, large-scale evaluation on 19 widely recognized multi-class benchmark text datasets using three prominent classifiers (MNB, SVM, and LR) to ensure the robustness, consistency, and generalizability of the proposed framework.

The remainder of this paper is organized as follows. Section 2 reviews recent studies on term-weighting schemes and feature selection methods for text classification. Section 3 presents the proposed methodology and provides a comprehensive and systematic investigation of the interdependent effects of term weighting schemes and feature selection. Section 4 reports and discusses the experimental results. Finally, Section 5 concludes the paper and outlines directions for future work.

2 Related Works

This section reviews related studies on term-weighting schemes and feature selection methods for text classification and discusses the main limitations of existing approaches in handling unstructured and high-dimensional vector spaces.

2.1 Variants of the Term Weighting Schemes

Term-weighting schemes used in TC are generally divided into unsupervised and supervised approaches, depending on whether class-label information is incorporated into the weighting process [16–18]. Unsupervised schemes estimate term importance solely from term-document and term-corpus statistics and assume that distributional properties such as frequency and dispersion are sufficient for classification [23,24]. While computationally efficient, these methods often fail to capture class-discriminative information, especially in multi-class and imbalanced settings.

Among unsupervised schemes, Term Frequency (TF) and TF-IDF remain the most widely used. TF emphasizes frequently occurring terms but tends to overweight non-informative words, is sensitive to document length, and cannot distinguish between locally frequent but globally uninformative terms [14,25–28]. TF-IDF mitigates some of these issues by penalizing ubiquitous terms, yet it has been shown to favor majority-class terms because global term distributions reflect corpus-level imbalance, which degrades minority-class performance [25,28–30]. Several enhanced unsupervised variants, such as TFmeanIDF, mTF, and mIDF, attempt to improve stability and balance between local and global statistics [16,31], but they remain fundamentally class-agnostic and thus inherit similar structural limitations.

To overcome these drawbacks, supervised term-weighting schemes incorporate class-label information to improve term discrimination [23,32]. Early work replaced the IDF component with FS criteria, such as Chi-squared (CHI), IG, Mutual Information (MI), and Odds Ratio (OR), leading to schemes like TF-CHI and TF-IG [33], followed by other variants such as TF-OR and TF-MI [34,35]. Class-indexing approaches, notably TF-ICF, further extended IDF to the class level by emphasizing terms concentrated in fewer categories [36]. However, TF-ICF and its extensions, such as TF-IDF-ICF and TF-IDF-ICSδF, still rely on relatively coarse class modeling and may penalize representative terms that appear across multiple classes [37–39].

More recent studies explored alternative class-aware criteria, including Term Frequency-Relevance Frequency (TF-RF), Term Frequency-Inverse Gravity Moment (TF-IGM), and their improved variants, as well as entropy-based schemes such as Cumulative Residual Entropy (CRE) and Distorted Cumulative Residual Entropy (DCRE), to better characterize term distribution across categories [15,31,37,39,40]. Another direction integrates FS criteria into term weighting, leading to Modified Distinguishing Feature Selector (MDFS) and MTF-based schemes that alleviate document length bias and improve weighting stability [14,41].

In summary, although unsupervised term-weighting schemes have played a foundational role in vector-space text representation, their inability to exploit class-distribution information fundamentally limits their discriminative capability in modern multi-class and imbalanced TC tasks. This shortcoming has driven the development of supervised and class-aware weighting strategies.

Nevertheless, despite notable advances, existing supervised term-weighting methods still exhibit several inherent limitations, including coarse modeling of class information, sensitivity to class imbalance, and limited ability to capture fine-grained term-class interactions in high-dimensional, sparse feature spaces. These challenges continue to motivate the pursuit of more accurate, adaptive, and highly discriminative supervised term-weighting schemes.

2.2 Feature Selection for Text Document Classification

Following the discussion on term weighting schemes, it is important to emphasize that weighting alone is insufficient to address the high intrinsic dimensionality of BoW text representations. Text documents typically contain tens of thousands of features, many of which are irrelevant, redundant, or only weakly correlated with class labels [42]. Such nondiscriminative features not only increase computational cost and memory consumption but also amplify noise and often degrade classification performance, particularly in large-scale, imbalanced text corpora [11].

FS plays a complementary and indispensable role by identifying a compact subset of the most informative and discriminative features, thereby reducing dimensionality while preserving essential semantic and discriminative content [43]. By eliminating noisy and redundant terms, FS improves learning efficiency, enhances generalization, and stabilizes classifier performance [6,44,45]. Based on the interaction between the selection process and the learning model, FS methods for TC are commonly categorized into three main paradigms: filter, wrapper, and hybrid approaches [8,21,46,47].

Filter methods operate independently of any classifier and rely on statistical, information-theoretic, or correlation-based criteria to estimate feature relevance [11]. Typical measures include distance, consistency, dependency, information, and correlation metrics [46]. These methods are further divided into univariate and multivariate approaches [3,48,49]. Univariate methods evaluate each feature independently and thus ignore redundancy among features, whereas multivariate methods jointly consider both feature relevance and inter-feature dependencies, enabling them to better handle redundant and interacting features [46,50].

Due to their simplicity, scalability, and low computational cost, filter methods have been widely adopted in TC [51,52], and numerous effective techniques have been proposed in the literature [53–57]. However, despite their efficiency, filter methods suffer from a fundamental limitation: they evaluate features in isolation from the classifier and therefore cannot guarantee optimal performance for a specific learning algorithm.

To alleviate the instability and bias of single-criterion filters, recent studies have increasingly explored ensemble-filter-based FS strategies that aggregate multiple ranking criteria to improve robustness, consistency, and selection reliability [58,59]. By leveraging the complementary strengths of diverse selection metrics, ensemble filters have demonstrated improved stability and more reliable performance across different datasets [60,61]. Nevertheless, even ensemble filters remain fundamentally disconnected from the classifier's actual decision boundaries.

Wrapper methods overcome this limitation by directly integrating a learning algorithm into the FS process. Candidate feature subsets are evaluated using a classifier-driven fitness function, and the search process explicitly accounts for feature interactions and model-specific performance. This tight coupling often yields more compact and highly discriminative feature subsets, leading to superior classification performance compared to filter methods [1,8]. However, this advantage comes at the cost of substantially higher computational complexity, since wrapper methods require repeated model training and validation, especially when cross-validation is employed [19,22]. Consequently, classical exhaustive or greedy wrapper searches become impractical for high-dimensional text spaces, where the FS problem is inherently NP-hard.

To address the combinatorial complexity of wrapper-based FS, metaheuristic optimization algorithms have become the dominant paradigm. These methods are particularly suitable for large, nonlinear, and high-dimensional search spaces due to their stochastic, population-based, and adaptive exploration mechanisms [62]. Recent research has increasingly focused on improved, hybridized, and problem-adapted metaheuristic wrappers to enhance convergence speed, stability, and solution quality in high-dimensional text FS problems [21].

Despite their effectiveness, pure wrapper-metaheuristic approaches remain computationally expensive due to repeated classifier evaluations over large feature spaces. To mitigate this limitation, hybrid filter-wrapper FS frameworks have emerged as a pragmatic and powerful solution. In these approaches, a filter stage (e.g., IG, CHI, or ensemble filters) is first used to aggressively reduce dimensionality, followed by a metaheuristic wrapper stage that refines the reduced feature subset and exploits feature interactions [63].

This strategy achieves a better balance between efficiency and effectiveness. It has gained increasing attention in recent years, as evidenced by a growing body of work, including in [64–67] and the recent comprehensive review by Alyasiri et al. [8,21]. More recently, ensemble filter and hybrid wrapper metaheuristic frameworks have been proposed to simultaneously address (i) the instability of single-filter methods and (ii) the premature convergence of wrapper metaheuristic optimization algorithms [63,68].

Despite these advances, several critical research gaps remain. First, most existing FS frameworks implicitly assume a generic classification setting and rarely consider the specific characteristics of multi-class imbalanced text data. Second, contemporary studies predominantly focus on improving the search strategy or subset evaluation mechanism while overlooking the foundational role of document representation quality. In most frameworks, term weighting is treated as a fixed preprocessing step, and FS is optimized independently of representation-level improvements. Third, limited attention has been given to systematically analyzing how supervised, task-aware term-weighting schemes can influence feature relevance estimation, ranking stability, and dimensionality reduction effectiveness, particularly in high-dimensional, sparse vector spaces.

These gaps are especially critical in multi-class imbalanced TC scenarios, where representation quality directly affects class separability and the statistical reliability of feature evaluation metrics such as Information Gain. Ignoring this interaction may limit the true potential of both term weighting and FS.

Motivated by these limitations, the present study proposes an integrated and task-aware framework that explicitly links enhanced supervised term weighting (i.e., MTF-MICF) with filter-based FS using Information Gain, ReliefF, and Chi-square. Unlike conventional approaches that treat representation and FS as isolated stages, this work positions supervised term weighting as a supporting mechanism that enhances feature relevance estimation prior to dimensionality reduction. Through systematic experimentation across multiple dimensionality reduction levels and extensive evaluation on benchmark multi-class datasets, this study provides a principled and experimentally validated analysis of the interaction between representation quality and feature selection effectiveness.

3 Methodology

This section presents a comprehensive and systematic investigation of the interdependent effects of term weighting schemes and FS, both of which play crucial roles in optimizing classification performance in multi-class TC scenarios. In doing so, this section outlines and summarizes the overall experimental methodology adopted to achieve the research objectives.

As illustrated in Fig. 1, the proposed framework follows a structured pipeline that begins with text data preparation and preprocessing. Subsequently, documents are represented using the BoW and VSM models. Within this representation stage, the impact of term weighting schemes, particularly when combined with FS, is systematically examined. The proposed methodology comprises two main components: (i) introducing an enhanced supervised term weighting scheme, namely Modified Term Frequency-Modified Inverse Class Frequency (MTF-MICF), to improve the quality and discriminative power of feature representation; and (ii) applying the IG filter as a FS mechanism to perform dimensionality reduction and to assess the effectiveness of MTF-MICF when integrated with FS in enhancing classification efficiency.

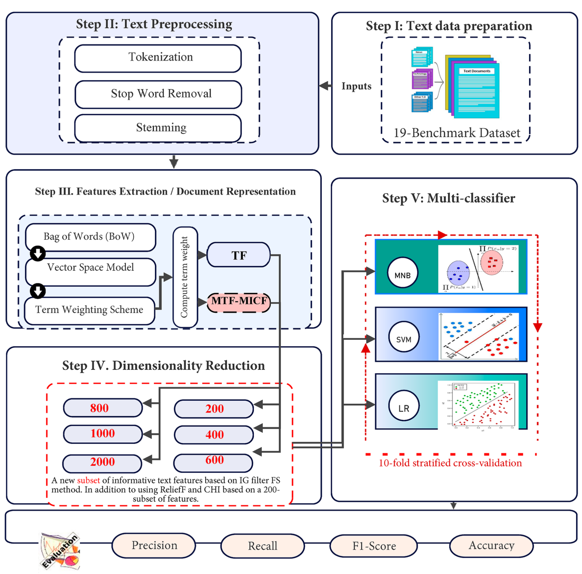


Figure 1: The research methodology phases.

After data preprocessing, feature extraction, and FS, the datasets are partitioned using a holdout strategy, with 80% allocated to training and validation, and the remaining 20% reserved as an independent test set. To address class imbalance and preserve the original class distribution, stratified 10-fold cross-validation is applied to the training portion during model learning and hyperparameter optimization. This validation strategy ensures reliable model selection, reduces performance variance, and guarantees fair representation of minority classes across all folds [69]. Finally, the selected model is evaluated on the unseen test set, which is kept completely isolated from both training and validation processes. This protocol provides an unbiased estimate of the model's generalization performance and minimizes the risk of overfitting.

3.1 The Benchmark Datasets

This study employs nineteen text collections extracted from several well-established benchmark corpora, including Reuters-21578, TREC, OHSUMED, and the WebACE project (WAP), as reported in [70]. These datasets cover a wide range of textual domains, including news articles, medical documents, web pages, and TREC collections, thereby enabling a comprehensive evaluation of the proposed approach across diverse application contexts and supporting the development of an accurate, generalizable model.

The selected datasets exhibit substantial variability in size, dimensionality, and class composition, reflecting varying levels of classification difficulty. In many cases, the datasets are characterized by very large vocabularies relative to the number of documents, leading to extremely high feature-to-instance ratios and highly sparse document representations. Such properties are well known to pose significant challenges for learning algorithms, as excessive sparsity can hinder effective pattern learning and degrade classification performance [71,72].

The summary of the characteristics of the nineteen text collections is reported in Table 1.

Table 1: The characteristics of the benchmark datasets used in the experiments.

Dataset	#Doume-nts	#Features	#Clases	#Min Class	#Max Class	#Avg Class	Matrix Sparsity	Domain
fbis	2463	2000	17	38	506	144.9	92.04%	News articles
la1s	3204	13,196	6	273	943	534.0	98.90%	
la2s	3075	12,433	6	248	905	512.5	98.84%	
new3s	9558	26,832	44	104	696	217.2	99.13%	
oh0	1003	3182	10	51	194	100.3	98.35%	Medical Documents
oh10	1050	3238	10	52	165	105.0	98.28%	
oh15	913	3100	10	53	157	91.3	99.97%	
oh5	918	3012	10	59	149	91.8	98.19%	
ohscal	11,162	11,465	10	709	1621	1116.2	99.47%	
re0	1504	2886	13	11	608	115.7	98.21%	News articles
re1	1657	3758	25	10	371	66.3	98.60%	
tr11	414	6429	9	6	132	46.0	95.62%	TREC Documents
tr12	313	5804	8	9	93	39.1	95.29%	
tr21	336	7902	6	4	231	56.0	94.05%	
tr23	204	5832	6	6	91	34.0	93.39%	
tr31	927	10,128	7	2	352	132.4	97.35%	
tr41	878	7454	10	9	243	87.8	99.74%	
tr45	690	8261	10	14	160	69.0	96.60%	
wap	1560	8460	20	5	341	78.0	98.33%	Web pages

3.2 Feature Extraction/Documents Representation

Among the most widely adopted document representation models in TC are BoW and VSM. These approaches transform each document into a numerical vector of terms (features), where each dimension corresponds to a distinct term in the vocabulary and is typically associated with its frequency of occurrence within the document. To further refine this representation and enhance its discriminative capability, term weighting schemes assign importance scores to features based on their contribution to distinguishing documents and classes. In this study, an enhanced supervised variant of the Term Frequency-Inverse Class Frequency (TF-ICF) scheme, referred to as MTF-MICF, is adopted from our previous work to improve document representation quality in multi-class imbalanced TC scenarios.

The proposed MTF-MICF scheme is specifically designed to enhance class discrimination and provide more informative input representations for subsequent FS and classification stages. By producing a more discriminative, class-aware term weighting, it facilitates the removal of non-informative features during FS and consequently improves overall classification performance.

The MTF-MICF scheme integrates two complementary components. First, it incorporates a Modified Term Frequency (MTF) factor to preserve informative term frequencies while mitigating bias introduced by variations in document length [14], calculated as:

$$MTF\left(t_i, d_k\right) = \sqrt{f_{ik} \cdot \log_2\left(1 + \frac{avg\_l}{l\left(d_k\right)}\right)} \tag{1}$$

where f_{ik} is the raw term frequency of term t_i in d_k, and $l(d_k)$ is the document length of d_k, which is defined as the sum of all term frequencies, and avg_l is the average document length of all training documents.

Second, it introduces a Modified Inverse Class Frequency (MICF) component that jointly considers: (i) the number of categories in which a term appears, and (ii) the average frequency of that term within each category. This dual modeling strategy enables a more balanced, accurate, and class-aware estimation of term importance, particularly in multi-class and imbalanced data settings. The MICF component is computed as follows: (i) The average frequency of each term t_i within each category j is calculated as:

$$\text{avg\_freq}_{t_i}^{\left(\text{category }_j\right)} = \frac{1}{N_j} \sum_{k \in D_j} x_{ik} \tag{2}$$

where $\text{avg\_freq}_{t_i}^{\left(\text{category }_j\right)}$ is the average frequency of the term t_i in category j. N_j is the number of documents in category j. D_j is the set of document indices in category j. x_{ik} is the frequency of the term t_i in document k.

Next, (ii) the average frequency of each term across all categories is computed as:

$$\text{avg\_ freq } t_{t_i} = \frac{1}{C} \sum_{j=1}^{C} \text{avg\_ freq}_{t_i}^{\left(\text{category }_j\right)} \tag{3}$$

where avg_freq_{t_i} is the average frequency of a term t_i in category j. C is the total number of categories. $\text{avg\_freq}_{t_i}^{\left(\text{category }_j\right)}$ is the average frequency of a term t_{i} in category j.

(iii) to balance the influence of class dispersion and frequency distribution, a weighting parameter α is introduced. This parameter controls the trade-off between the number of categories in which a term appears and its average frequency across categories. Accordingly, the proposed Modified Inverse Class Frequency (MICF) is defined as:

$$\text{MICF}_{(t_i,d_j,c_k)} = 1 + \log\left(\alpha \cdot \frac{C}{c_{t_i}} + (1-\alpha) \cdot avgfreq_{t_i}\right) \tag{4}$$

where α is a weight parameter (0.5 by default). C denotes the total number of predefined categories in the collection. c_{t_i} is the count of categories in which the term t_i appears. $avgfreq_{t_i}$ is the average frequency of a term t_i across all categories.

Finally, the complete MTF-MICF weighting scheme is formulated as:

$$\text{MTF} - \text{MICF}\left(t_i, d_j, c_k\right) = MTF\left(t_i, d_k\right) * MICF\left(t_i, d_j, c_k\right) \tag{5}$$

where $MTF\left(t_i, d_k\right)$ is Modified Term Frequency calculated as in Eq. (1). $MICF\left(t_i, d_j, c_k\right)$ is the proposed modified inverse class frequency calculated as in Eq. (4).

3.3 Dimensionality Reduction Technique

The dimensionality reduction experiment in this study serves a twofold purpose. First, it aims to eliminate irrelevant and weakly informative features, producing compact, discriminative feature subsets that reduce noise, computational complexity, and the risk of overfitting in high-dimensional TC tasks. Given the large variability in vocabulary size and feature dimensionality across benchmark datasets, a fixed set of top-N ranked features is employed to ensure a controlled, dataset-independent evaluation framework.

Since there is no universally standardized range of feature dimensionalities in the literature, many previous studies have explored different dimensionalities to analyze model behavior and performance [61,73–75]. Following this common practice, the number of selected features K is set to range from 200 to 1000 features, with increments of 200 (200, 400, 600, 800, 1000), and an additional setting with 2000 features is included as a reference, corresponding to the maximum dimensionality used in the fbis dataset. This configuration enables a systematic investigation of the impact of feature dimensionality on classification performance and provides a consistent basis for evaluating the effectiveness of the proposed MTF-MICF scheme across datasets with diverse attribute spaces and class distribution.

Second, the experiment is designed to investigate the interaction between FS and the proposed MTF-MICF term weighting scheme under varying reduction ratios using a filter-based approach. By evaluating performance across multiple feature subset sizes, the study assesses its strength and discriminative capacity under both aggressive and moderate feature reduction settings, thereby assessing its ability to preserve informative features as dimensionality decreases.

For this purpose, IG is adopted as the filter-based FS method due to its proven effectiveness in high-dimensional TC. IG measures the reduction in class uncertainty contributed by each feature and ranks them, accordingly, retaining only the most informative terms [66]. This process significantly reduces dimensionality while maintaining strong discriminative power and computational efficiency. Additionally, two widely used filter-based FS techniques, namely Chi-Square (CHI) and ReliefF, were employed to further evaluate the generalizability of the proposed model across different feature ranking strategies.

3.4 Machine Learning Algorithm

To ensure a comprehensive, consistent, and unbiased evaluation, this study adopts the three best-performing classifiers identified as statistically significant in our previous work [76].

3.4.1 Multinomial Naive Bayes (MNB)

MNB is chosen for its simplicity, computational efficiency, and well-established effectiveness in TC tasks, particularly in scenarios where term frequency information plays a dominant role. Despite its strong conditional independence assumption, MNB remains a competitive and reliable baseline for textual data. In this study, MNB is implemented using its standard formulation with default parameter settings.

3.4.2 Support Vector Machine (SVM)

SVMs are particularly well-suited for high-dimensional, sparse feature spaces, which are characteristic of TC problems. By constructing an optimal separating hyperplane, SVM effectively handles large-scale feature spaces while maintaining strong generalization performance. In this study, a linear SVM is employed using the LIBLINEAR library with L2-regularized L2-loss support vector classification (dual formulation) to ensure both scalability and robustness.

3.4.3 Logistic Regression (LR)

LR is a widely used discriminative linear classifier in TC. Unlike linear regression, LR models class membership probabilities through a sigmoid function, thereby providing valid probabilistic outputs. In this study, LR is implemented using LIBLINEAR with L2-regularized logistic regression (dual formulation) to achieve stable, efficient, and scalable optimization.

3.5 Evaluation Metrics

In TC, particularly in topic classification tasks involving multi-class and imbalanced datasets, selecting appropriate evaluation metrics is crucial for obtaining reliable and meaningful performance estimates. Relying solely on classification accuracy is often inadequate, as it can be biased toward majority classes and may therefore fail to reflect true predictive performance in minority classes [8,76].

For this reason, more informative metrics, such as the F1-score, which combines Precision (P) and Recall (R), are widely recommended [11]. The F1-score provides a balanced trade-off between False Positives (*FP*) and False Negatives (*FN*) and thus offers a more robust and reliable assessment of classifier performance under class imbalance conditions. Therefore, in this study, the proposed models are evaluated using both classification accuracy and the weighted-average F1-score, computed from the standard confusion matrix components: True Positives (*TP*), True Negatives (*TN*), *FP*, and *FN*.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{F1} - \text{score} = \frac{2\,(\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \tag{8}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

4 Results and Discussions

4.1 Experimental Evaluation of MTF-MICF with IG Based on Different Dimension Ratios

Tables 2–4 present a comparative evaluation between the baseline term-weighting scheme (standard Term Frequency—TF) and the enhanced supervised weighting scheme (MTF-MICF), both integrated with IG-based FS across multiple dimensionality levels (200–2000 features). Across the majority of datasets and dimensionality levels, integrating MTF-MICF with IG consistently improves both F1-score and classification accuracy for MNB, SVM, and LR. The improvements are particularly pronounced for MNB and LR, confirming that enhanced supervised weighting strengthens discriminative term representation prior to feature ranking.

Table 2: F1-score and classification accuracy comparisons of IG filter FS based on TF and the proposed MTF-MICF based on MNB.

Dataset	FS = 200				FS = 400				FS = 600			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
fbis	72.36	71.81	78.10	77.89	77.03	76.06	79.03	78.90	75.61	74.85	79.16	78.90
la1s	84.27	84.24	84.77	84.71	86.30	86.27	89.72	89.70	87.15	87.21	89.69	89.70
la2s	84.20	84.23	87.63	87.48	84.61	84.72	88.69	88.62	84.79	84.88	87.98	87.97
new3s	69.67	70.40	73.56	74.01	73.83	74.37	77.57	77.93	74.44	74.84	78.86	79.29
oh0	86.85	86.57	92.55	92.54	87.59	87.56	94.11	94.03	88.65	88.56	96.02	96.02
oh10	78.10	78.10	83.94	84.29	79.13	79.05	82.72	82.86	79.14	79.05	83.72	83.81
oh15	83.09	83.61	87.06	87.43	85.50	85.79	86.73	86.89	84.73	85.25	86.78	86.89
oh5	83.48	83.70	89.30	89.67	85.69	85.87	91.07	91.30	88.43	88.59	90.65	90.76
ohscal	72.94	72.68	73.05	72.86	73.14	73.09	75.01	74.92	73.08	73.04	75.02	74.97
re0	81.38	80.40	79.56	79.07	81.35	80.73	79.56	79.40	81.03	80.73	79.31	79.07
re1	82.68	82.83	83.47	83.13	82.35	83.13	84.99	84.94	82.30	83.73	83.44	83.73
tr11	91.04	90.36	89.54	89.02	89.52	89.16	94.08	93.90	88.42	87.95	94.23	93.90
tr12	82.75	82.54	88.81	88.89	82.66	82.54	91.87	92.06	84.43	84.13	91.69	92.06
tr21	73.76	69.12	88.70	88.06	69.03	66.18	88.95	88.06	70.45	67.65	90.20	89.55
tr23	64.65	60.98	87.95	87.80	69.21	65.85	90.32	90.24	68.87	65.85	90.32	90.24
tr31	92.26	91.40	98.39	98.38	91.91	90.86	97.41	97.30	92.68	91.40	95.77	95.68
tr41	91.88	91.48	93.31	93.10	91.77	91.48	93.85	93.68	93.36	93.18	96.06	95.98
tr45	81.05	78.99	94.18	94.20	84.32	84.06	95.03	94.93	84.87	84.78	93.45	93.48
wap	79.32	79.81	80.65	82.32	79.39	80.45	82.12	83.60	81.88	82.69	81.42	82.96
Average	80.83	80.17	**86.03**	**86.05**	81.81	81.43	**87.52**	**87.54**	82.33	82.02	**87.57**	**87.63**

Dataset	FS = 800				FS = 1000				FS = 2000			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
fbis	76.28	75.86	79.02	78.70	77.60	77.08	80.16	79.72	75.92	75.66	80.26	80.12
la1s	87.90	87.99	89.87	89.86	88.41	88.46	90.98	90.95	90.74	90.80	91.93	91.89
la2s	85.79	85.85	88.49	88.46	86.66	86.67	89.77	89.76	88.64	88.62	90.94	90.89

(Continued)

Table 2 (continued)

Dataset	FS = 800				FS = 1000				FS = 2000			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
new3s	74.52	74.90	79.83	80.13	75.08	75.26	81.08	81.17	76.33	76.57	81.38	81.43
oh0	89.15	89.05	96.54	96.52	89.16	89.05	96.54	96.52	88.23	88.06	95.04	95.02
oh10	78.01	78.10	82.97	83.33	80.57	80.48	84.56	84.76	78.19	78.57	84.68	84.76
oh15	84.11	84.70	86.66	86.89	85.27	85.79	87.54	87.98	86.63	86.89	88.38	88.52
oh5	86.83	86.96	88.53	88.59	87.85	88.04	90.11	90.22	84.19	84.24	91.73	91.85
ohscal	72.80	72.86	75.02	75.01	73.02	73.09	75.65	75.64	73.91	74.07	75.54	75.55
re0	81.99	81.73	80.27	80.07	81.53	81.40	80.34	80.07	81.47	81.40	78.67	78.41
re1	79.79	81.63	84.83	85.24	79.28	81.02	84.83	85.54	81.57	83.73	85.55	86.14
tr11	91.92	91.57	93.03	92.68	90.81	90.36	94.23	93.90	88.69	87.80	94.18	93.90
tr12	83.02	82.54	91.69	92.06	88.42	88.89	95.09	95.24	87.56	87.30	91.69	92.06
tr21	71.57	69.12	88.73	88.06	71.20	69.12	88.73	88.06	72.48	70.15	91.74	91.04
tr23	73.92	70.73	90.32	90.24	71.47	68.29	92.68	92.68	73.04	70.73	88.28	87.80
tr31	93.14	92.47	96.25	96.22	92.58	92.47	96.28	96.22	93.65	93.51	97.35	97.30
tr41	93.33	93.18	96.10	95.98	93.89	93.75	96.10	95.98	95.46	95.40	96.10	95.98
tr45	85.53	85.51	94.20	94.20	84.80	84.78	94.92	94.93	86.92	86.96	94.92	94.93
wap	81.47	82.69	81.86	83.28	79.74	81.41	81.16	82.96	79.90	81.67	79.10	80.71
Average	82.69	82.50	**87.59**	**87.66**	83.02	82.92	**88.46**	**88.54**	83.34	83.27	**88.29**	**88.33**

Note: The bold values indicate the best performance (i.e., the highest average score) achieved among the compared methods in all datasets.

Table 3: F1-score and classification accuracy comparisons of IG filter FS based on TF and the proposed MTF-MICF based on SVM.

Dataset	FS = 200				FS = 400				FS = 600			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
fbis	78.79	79.11	78.96	79.31	76.60	76.67	80.77	80.73	78.23	78.50	82.63	83.16
la1s	85.62	85.65	87.14	87.21	85.63	85.49	88.00	87.99	83.36	83.15	86.87	86.90
la2s	85.54	85.53	89.42	89.43	85.38	85.37	88.52	88.62	85.69	85.85	86.48	86.50
new3s	75.13	75.26	78.49	78.92	77.30	77.35	81.55	81.59	79.93	79.92	82.02	82.11
oh0	83.61	83.58	86.63	86.57	85.17	85.07	90.18	90.05	84.05	84.08	89.67	89.55
oh10	76.12	75.71	78.57	78.10	76.52	76.19	83.65	83.33	73.95	73.33	80.53	80.00
oh15	79.70	79.78	82.07	81.97	82.92	83.06	84.72	84.70	81.79	81.97	85.23	85.25
oh5	79.31	79.89	81.69	82.07	85.08	85.33	87.32	87.50	82.28	82.61	87.42	87.50
ohscal	77.06	77.07	78.31	78.24	76.74	76.76	79.10	79.09	74.59	74.56	78.10	78.10
re0	85.25	85.38	86.65	86.71	85.87	86.05	86.03	86.05	85.13	85.05	85.17	85.05
re1	82.85	83.13	85.23	85.54	82.26	83.43	85.62	85.84	82.32	82.83	84.62	84.94

(Continued)

Table 3 (continued)

Dataset	FS = 200				FS = 400				FS = 600			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
tr11	85.69	85.54	89.18	89.02	93.90	93.98	86.56	86.59	95.20	95.18	85.66	85.37
tr12	87.60	87.30	84.59	84.13	84.95	84.13	84.72	84.13	93.60	93.65	87.70	87.30
tr21	87.36	88.24	98.47	98.51	86.20	88.24	98.47	98.51	86.20	88.24	98.47	98.51
tr23	87.51	87.80	95.23	95.12	87.76	87.80	92.99	92.68	87.76	87.80	92.99	92.68
tr31	97.85	97.85	98.91	98.92	98.42	98.39	97.87	97.84	98.42	98.39	97.83	97.84
tr41	94.85	94.89	95.83	95.98	97.03	97.16	95.84	95.98	96.46	96.59	94.69	94.83
tr45	94.16	94.20	92.51	92.75	93.38	93.48	91.98	92.03	91.88	92.03	92.73	92.75
wap	75.34	75.64	77.27	77.49	77.33	77.56	83.59	83.92	78.38	78.53	86.05	86.17
Average	84.18	84.29	**86.59**	**86.63**	85.18	85.34	**87.76**	**87.74**	85.22	85.38	**87.63**	**87.61**

Dataset	FS = 800				FS = 1000				FS = 2000			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
fbis	79.74	79.72	83.05	83.37	81.00	81.14	85.32	85.60	83.13	83.37	85.73	86.00
la1s	83.11	83.00	88.30	88.30	83.63	83.46	89.14	89.08	85.17	85.18	89.87	89.86
la2s	85.56	85.69	87.32	87.48	86.53	86.67	89.35	89.43	89.21	89.27	91.41	91.38
new3s	80.11	80.07	82.82	82.79	81.82	81.80	84.68	84.68	84.73	84.68	86.60	86.51
oh0	85.98	86.07	90.60	90.55	88.01	88.06	92.59	92.54	86.54	86.57	92.67	92.54
oh10	73.13	72.86	81.09	80.48	74.24	73.81	82.04	81.43	77.13	77.62	80.56	80.48
oh15	86.22	86.34	86.88	86.89	85.49	85.79	86.33	86.34	84.09	84.15	90.15	90.16
oh5	85.14	85.33	88.39	88.59	87.96	88.04	89.64	89.67	87.88	88.04	90.70	90.76
ohscal	72.30	72.41	77.43	77.47	70.56	70.62	75.76	75.73	72.33	72.32	76.18	76.13
re0	82.75	82.72	86.12	86.05	83.26	83.06	86.74	86.71	83.45	83.39	88.23	88.37
re1	83.30	84.34	84.81	84.94	84.02	84.64	84.75	84.64	84.99	85.84	87.54	87.95
tr11	94.19	93.98	88.97	89.02	93.10	92.77	90.22	90.24	92.98	92.68	92.55	92.68
tr12	92.16	92.06	87.70	87.30	85.72	85.71	90.88	90.48	90.29	90.48	90.29	90.48
tr21	86.23	86.76	98.47	98.51	88.81	89.71	98.47	98.51	89.63	91.04	98.47	98.51
tr23	87.51	87.80	90.58	90.24	87.51	87.80	92.99	92.68	87.51	87.80	92.31	92.68
tr31	98.42	98.39	98.37	98.38	98.40	98.39	98.91	98.92	98.93	98.92	98.91	98.92
tr41	96.46	96.59	95.27	95.40	96.46	96.59	95.90	95.98	96.50	96.55	95.32	95.40
tr45	91.83	92.03	92.00	92.03	91.83	92.03	92.00	92.03	91.83	92.03	92.00	92.03
wap	79.78	80.45	86.44	86.50	79.65	80.13	86.80	87.14	81.12	81.67	87.35	87.78
Average	85.47	85.61	**88.14**	**88.12**	85.68	85.80	**89.08**	**89.04**	86.71	86.93	**89.83**	**89.93**

Table 4: F1-score and classification accuracy comparisons of IG filter FS based on TF and the proposed MTF-MICF based on LR.

Dataset	FS = 200				FS = 400				FS = 600			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
fbis	80.50	81.14	83.41	83.98	81.20	81.34	84.33	84.79	83.56	83.77	86.51	87.02
la1s	86.24	86.27	89.23	89.24	87.19	87.21	89.18	89.24	86.32	86.27	89.18	89.24
la2s	87.31	87.48	89.49	89.59	88.09	88.13	91.01	91.06	87.58	87.64	90.66	90.73
new3s	78.26	78.56	81.54	81.85	82.28	82.43	86.04	86.19	84.00	84.05	86.31	86.45
oh0	87.12	87.06	92.07	92.04	90.63	90.55	93.62	93.53	89.71	89.55	93.65	93.53
oh10	80.42	80.48	81.57	81.43	77.22	77.14	84.29	84.29	77.83	77.62	84.65	84.29
oh15	83.94	84.15	84.97	85.25	85.19	85.25	87.47	87.43	85.09	85.25	87.99	87.98
oh5	83.85	84.24	88.35	88.59	88.44	88.59	88.90	89.13	88.36	88.59	91.13	91.30
ohscal	86.11	86.38	77.98	77.92	77.61	77.65	79.77	79.76	77.34	77.34	80.23	80.25
re0	77.29	77.30	88.15	88.37	86.92	87.04	88.81	89.04	86.59	86.71	88.50	88.70
re1	84.35	85.24	86.54	87.35	84.12	85.54	87.28	88.25	84.47	85.54	86.68	87.65
tr11	90.19	90.36	91.67	91.46	94.13	93.98	88.99	89.02	95.20	95.18	88.97	89.02
tr12	87.44	87.30	90.22	90.48	87.65	87.30	92.00	92.06	90.46	90.48	90.46	90.48
tr21	87.36	88.24	98.47	98.51	87.56	89.71	98.47	98.51	87.56	89.71	98.47	98.51
tr23	92.17	92.68	97.52	97.56	89.88	90.24	92.68	92.68	87.51	87.80	92.99	92.68
tr31	97.84	97.85	100	100	98.42	98.39	99.46	99.46	98.37	98.39	98.92	98.92
tr41	95.45	96.02	96.40	96.55	97.00	97.16	96.47	96.55	96.44	96.59	95.32	95.40
tr45	93.44	93.48	94.04	94.20	92.65	92.75	94.07	94.20	92.63	92.75	93.37	93.48
wap	80.06	81.09	82.52	83.60	80.35	81.73	84.28	85.21	81.76	83.01	87.59	88.42
Average	86.28	86.60	**89.17**	**89.37**	87.19	87.48	**89.85**	**90.02**	87.41	87.70	**90.08**	**90.21**

Dataset	FS = 800				FS = 1000				FS = 2000			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
fbis	83.91	83.98	85.49	85.80	83.46	83.57	85.88	86.21	86.14	86.41	86.83	87.22
la1s	87.01	87.05	89.97	90.02	86.71	86.74	89.50	89.55	87.55	87.52	91.86	91.89
la2s	89.03	89.11	91.78	91.87	90.67	90.73	91.99	92.03	91.68	91.71	92.86	92.85
new3s	85.10	85.15	86.70	86.77	86.30	86.40	88.39	88.39	87.82	87.76	89.56	89.49
oh0	89.15	89.05	93.10	93.03	89.12	89.05	94.07	94.03	89.20	89.05	94.10	94.03
oh10	78.83	78.57	83.83	83.33	78.73	78.57	82.89	82.38	77.73	78.10	82.78	82.86
oh15	86.68	86.89	89.06	89.07	85.46	85.79	89.01	89.07	85.26	85.25	90.69	90.71
oh5	87.82	88.04	91.13	91.30	90.60	90.76	90.71	90.76	90.11	90.22	92.93	92.93
ohscal	76.75	76.76	79.85	79.89	76.11	76.13	80.30	80.34	75.88	75.91	80.20	80.21
re0	86.59	86.71	86.83	87.04	85.90	86.05	87.85	88.04	86.79	86.71	87.18	87.38
re1	83.95	85.54	86.14	87.05	85.18	86.45	86.75	87.65	86.67	87.65	86.33	87.35

(Continued)

Table 4 (continued)

Dataset	FS = 800				FS = 1000				FS = 2000			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
tr11	95.20	95.18	90.22	90.24	94.11	93.98	91.37	91.46	91.83	91.46	91.37	91.46
tr12	90.70	90.48	91.97	92.06	85.72	85.71	93.46	93.65	90.56	90.48	93.14	93.65
tr21	91.60	92.65	98.47	98.51	91.60	92.65	98.47	98.51	89.63	91.04	98.47	98.51
tr23	87.51	87.80	92.99	92.68	87.51	87.80	92.68	92.68	87.51	87.80	92.31	92.68
tr31	98.37	98.39	99.45	99.46	98.36	98.39	99.45	99.46	98.92	98.92	99.45	99.46
tr41	96.44	96.59	95.90	95.98	96.44	96.59	95.90	95.98	96.50	96.55	95.90	95.98
tr45	92.58	92.75	94.11	94.20	92.58	92.75	93.37	93.48	92.58	92.75	94.07	94.20
wap	81.65	82.69	87.94	88.42	81.37	82.37	87.30	87.78	81.99	82.64	87.54	88.10
Average	87.84	88.07	**90.26**	**90.35**	87.68	87.92	**90.49**	**90.60**	88.12	88.31	**90.92**	**91.10**

As shown in Table 2, the MNB classifier exhibits consistent and substantial improvements in both average F1-score and classification accuracy across all feature subset sizes when the proposed MTF-MICF scheme is applied. For instance, at FS = 200, the average F1-score increases from 80.83 to 86.03, while classification accuracy improves from 80.17% to 86.05%. Similar performance gains are observed with larger feature subsets, indicating that MNB, which is inherently dependent on accurate term-frequency statistics, benefits strongly from the more stable, class-aware frequency modeling introduced by MTF-MICF. This confirms that enhancing term-frequency representations is particularly crucial for probabilistic classifiers whose decisions are directly driven by word-occurrence statistics.

For the SVM classifier, which constructs optimal separating hyperplanes in high-dimensional feature spaces, the integration of the proposed MTF-MICF weighting scheme generally yields consistent and meaningful performance improvements, as shown in Table 3. For instance, at FS = 400, the average F1-score increases from 85.18 to 87.76, while classification accuracy improves from 85.34% to 87.74%. Although these improvements are more moderate compared to those observed with MNB, they remain stable across most feature subset sizes. This behavior indicates that MTF-MICF enhances the discriminative structure of the feature space, thereby improving class separability and enabling SVM to estimate more reliable decision boundaries, even under reduced-dimensional settings.

Nevertheless, despite the overall positive trend, certain configurations exhibit localized performance degradation relative to the TF baseline. For example, the tr11 dataset with SVM at FS = 400 shows a noticeable decrease in both F1-score and accuracy, with minor fluctuations also observed at FS = 800. These isolated cases can be attributed to the interaction between dataset-specific characteristics, feature subset size, and the sensitivity of margin-based classifiers to feature scaling. Specifically, the tr11 dataset exhibits relatively balanced class distributions and moderately discriminative term patterns, where IG-based FS under the TF representation may already preserve highly informative features at intermediate dimensionality levels. In such cases, introducing additional class-sensitive scaling through MICF may slightly alter the geometric distribution of feature vectors, which can affect the margin optimization process of SVM. This effect is particularly evident when the feature space is neither sufficiently reduced to benefit fully from enhanced weighting nor large enough to absorb scaling variations smoothly. Importantly, this degradation is localized rather than systematic, as performance improvements resume at higher dimensionality levels, confirming the overall robustness and effectiveness of the proposed weighting scheme across diverse experimental conditions.

Similarly, for the LR classifier (Table 4), the performance gains introduced by MTF-MICF are particularly pronounced. For instance, at FS = 600, the F1-score increases from 87.41 to 90.08, and classification accuracy improves from 87.70% to 90.21%. Comparable trends are observed across other feature subset sizes, indicating that LR benefits significantly from the improved feature scaling and class-aware weighting provided by MTF-MICF, especially in moderately to highly dimensional settings where representation quality plays a dominant role.

The experimental results clearly demonstrate that the effectiveness of FS is strongly dependent on the quality of the underlying document representation, particularly in high-dimensional TC problems characterized by extreme sparsity and class imbalance, as reported in Table 1. In sparse text matrices, where sparsity often exceeds 99%, most terms occur infrequently and contribute limited statistical evidence for reliable feature evaluation. Under such conditions, conventional TF-based representations tend to assign similar importance to both discriminative and non-discriminative rare terms, thereby weakening the reliability of entropy-based feature ranking methods such as IG. Since IG estimates feature importance based on the reduction of class uncertainty, its effectiveness is fundamentally constrained by the statistical fidelity of the underlying term weights. If term weights do not adequately reflect class-specific distributions, the estimated entropy reduction becomes less reliable, particularly in extremely sparse and imbalanced settings.

By integrating the MICF component into the MTF formulation, the proposed MTF-MICF scheme introduces an explicit class-discriminative scaling factor that adjusts term importance according to its distribution across classes rather than solely its frequency within documents. Mathematically, MICF amplifies terms that are concentrated in a limited number of classes while suppressing terms that are uniformly distributed across many classes. This mechanism effectively increases the signal-to-noise ratio in sparse feature spaces by strengthening the contribution of class-indicative terms and reducing the influence of globally frequent but non-discriminative terms. As sparsity increases, this class-sensitive scaling becomes increasingly important because the discriminative information is concentrated in a small subset of informative features. Consequently, the proposed representation improves the separability between class-conditional feature distributions, which directly enhances the reliability of entropy-based ranking and enables IG to select more informative features even under aggressive dimensionality reduction.

The benefits of MTF-MICF are particularly evident in datasets with severe class imbalance, where conventional TF weighting tends to be biased toward majority-class term distributions. In imbalanced settings, terms associated with minority classes often have lower absolute frequencies and may be undervalued or discarded during FS. The MICF component mitigates this limitation by assigning higher weights to terms that exhibit strong class specificity, regardless of their global frequency. This improves the relative representation of minority-class features and leads to more balanced entropy estimates during IG-based ranking. From a theoretical perspective, this improves the estimation of class-conditional probability distributions, resulting in more accurate posterior probability estimation for probabilistic classifiers (e.g., MNB) and improved margin separability for discriminative classifiers (e.g., SVM and LR).

As sparsity approaches extreme levels, the discriminative power of representation becomes the dominant factor controlling classification performance. In such regimes, FS alone cannot compensate for weak or noisy representations, because entropy-based ranking relies entirely on the statistical structure encoded in the term weights. By improving the alignment between term weights and class-conditional distributions, MTF-MICF enhances the intrinsic quality of the feature space prior to FS. This leads to more stable feature importance estimation, promotes the earlier selection of truly discriminative features, and improves classification performance across different dimensionality levels. Although localized underperformance is observed in a limited number of configurations (e.g., tr11 and tr12), these cases reflect dataset-specific interactions between feature distribution, sparsity structure, and classifier sensitivity, rather than a systematic limitation of the proposed weighting scheme.

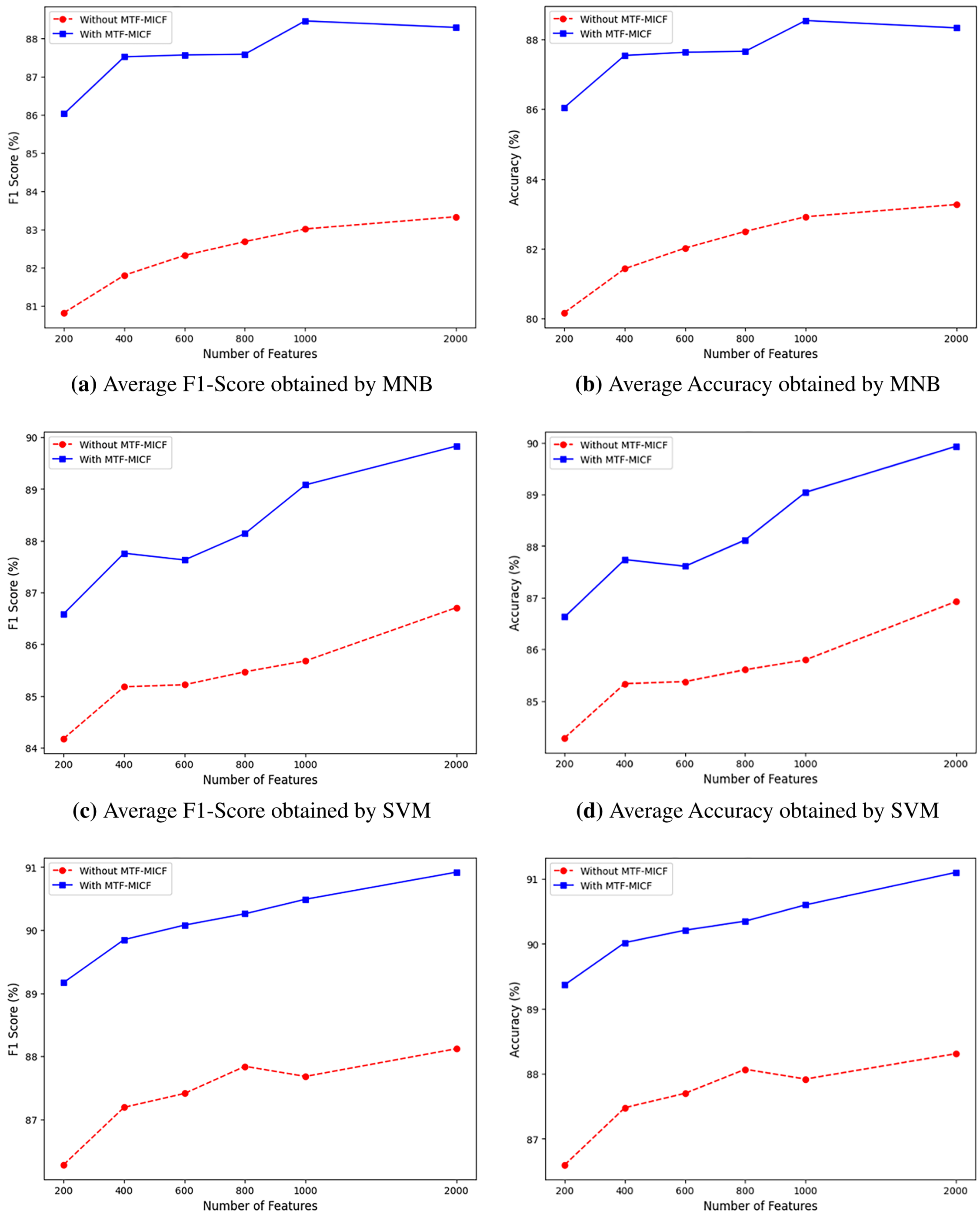
Across the comprehensive experimental setting, including 19 benchmark datasets with varying sparsity levels and imbalance ratios, three fundamentally different classifiers, and six dimensionality levels, the proposed MTF-MICF scheme consistently achieves statistically meaningful improvements in both average F1-score and classification accuracy compared to the TF baseline. Notably, the performance gains are more pronounced in datasets exhibiting higher sparsity and stronger imbalance, confirming that the proposed class-sensitive weighting mechanism becomes increasingly beneficial as the statistical estimation problem becomes more challenging. These findings confirm that MTF-MICF plays a central role in improving classification performance by enhancing representation quality prior to FS. By tightly integrating class-aware term weighting with IG-based FS, the proposed framework produces more informative, stable, and discriminative feature subsets while maintaining computational efficiency.

A more comprehensive view of these trends is illustrated in Fig. 2a–f, which depicts the performance of IG filter-based FS with MTF-MICF and without MTF-MICF (using baseline TF) across different dimensionality reduction ratios using MNB, SVM, and LR classifiers. Across all three classifiers, MTF-MICF consistently improves both average F1-score and accuracy for all feature subset sizes (FS = 200, 400, 600, 800, 1000, and 2000). However, the magnitude and dynamics of these improvements vary with the dimensionality level and the learning model. For MNB, the performance gains become more pronounced as the number of retained features increases, indicating that MTF-MICF effectively mitigates information loss induced by aggressive dimensionality reduction while preserving the discriminative power of frequent and class-indicative terms. For SVM, the improvements are more gradual but steady, reflecting the fact that better term weighting primarily refines the geometry of the decision boundary rather than radically altering the feature space structure. For LR, the gains are especially evident in higher-dimensional settings, where MTF-MICF provides better-scaled, more informative inputs, thereby facilitating more accurate probability estimation and improved class separation.

Overall, these results demonstrate that MTF-MICF not only enhances feature representation quality but also improves learning robustness under varying dimensionality constraints. More importantly, the consistent performance gains across fundamentally different classifiers confirm that the proposed scheme captures intrinsic, classifier-independent properties of term-class relevance. From a broader perspective, these findings indicate that carefully designed, class-aware term weighting is not merely a preprocessing refinement, but a central component in controlling the trade-off between dimensionality reduction and information preservation in high-dimensional, multi-class, and imbalanced TC problems. Consequently, MTF-MICF provides a principled and effective mechanism for managing high-dimensional textual data while maintaining strong and stable classification performance, thereby validating its role as a key building block in the proposed FS and learning framework.

4.2 Experimental Evaluation of MTF-MICF with ReliefF and Chi-Square

The effectiveness of the proposed MTF-MICF term-weighting scheme was further evaluated using additional filter-based FS methods to examine whether its impact is specific to IG or generalizable across different feature ranking strategies. In particular, two widely used filter techniques, Chi-Square (CHI) and ReliefF, were employed to assess the interaction between the enhanced document representation and alternative FS mechanisms. For computational efficiency and fair comparison, the MTF-MICF weighting scheme was first applied to the full feature space, after which the FS methods were used to extract the top 200 most informative features. This subset size was selected to significantly reduce the computational cost associated with repeated training and evaluation while preserving a representative set of discriminative features for reliable comparison. By fixing the dimensionality level, the experiments focus specifically on analyzing how the enhanced term representation produced by MTF-MICF influences the effectiveness of different FS strategies in ranking and selecting informative terms.



(a) Average F1-Score obtained by MNB

(b) Average Accuracy obtained by MNB

(c) Average F1-Score obtained by SVM

(d) Average Accuracy obtained by SVM

(e) Average F1-Score obtained by LR

(f) Average Accuracy obtained by LR

Figure 2: Performance comparison of filter-based FS with and without MTF-MICF across varying feature dimensions.

The results reported in Tables 5 and 6 demonstrate that integrating the proposed MTF-MICF weighting scheme consistently improves classification performance compared with the TF baseline when combining with the two widely used filter-based FS methods, namely ReliefF and Chi-Square. Based on the average F1-score and classification accuracy across the 19 benchmark datasets, MTF-MICF achieves noticeable improvements for all three classifiers (MNB, SVM, and LR). For instance, under the ReliefF-based FS (Table 5), the average F1-score increases from 77.53 to 82.41 for MNB, from 79.96 to 84.52 for SVM, and from 82.22 to 86.69 for LR. Similarly, under the Chi-Square-based FS (Table 6), the average results also confirm consistent gains across most datasets and classifiers. Moreover, dataset-level analysis indicates that the proposed representation frequently achieves higher performance on the majority of datasets, further confirming that the enhanced MTF-MICF representation improves the effectiveness of feature ranking and selection. These results indicate that the proposed weighting scheme functions as a robust supporting mechanism for different FS strategies, rather than being limited to a specific FS method.

Table 5: F1-score and classification accuracy comparisons of ReliefF filter FS based on TF and the proposed MTF-MICF based on MNB, SVM, and LR.

Dataset	MNB				SVM				LR			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
fbis	56.87	57.20	76.24	75.86	63.14	63.46	76.16	76.18	61.95	63.08	80.70	81.14
la1s	75.59	75.35	77.09	77.22	77.35	77.44	81.69	81.81	80.18	80.34	84.15	84.24
la2s	80.68	80.65	83.25	83.09	77.86	77.90	84.55	84.62	81.67	81.79	86.56	86.67
new3s	59.95	60.98	72.80	73.27	66.09	66.98	76.01	76.43	67.29	68.25	79.76	80.23
oh0	80.78	80.60	89.24	89.05	80.95	81.06	84.88	85.05	81.48	81.09	86.75	86.57
oh10	75.77	75.71	82.00	82.38	72.24	72.38	78.94	78.73	77.85	78.10	81.63	81.90
oh15	80.73	81.42	80.54	80.87	75.31	75.55	77.20	77.37	81.26	81.42	80.71	80.87
oh5	84.61	84.78	88.67	89.13	79.55	79.71	85.68	85.87	86.24	86.41	88.40	88.59
ohscal	70.38	70.26	67.29	67.22	73.50	73.45	73.00	72.95	74.02	73.89	73.12	73.00
re0	81.33	80.73	77.74	77.41	83.69	83.85	83.45	83.41	84.82	85.05	86.35	86.71
re1	82.26	82.23	85.88	86.75	81.10	81.53	86.01	86.35	84.36	85.84	86.75	87.95
tr11	79.89	79.52	91.57	91.46	84.32	84.80	88.28	88.62	86.45	86.75	90.49	90.24
tr12	88.73	88.89	82.57	82.54	86.72	87.23	86.43	86.17	90.14	90.48	88.59	88.89
tr21	83.36	83.82	74.76	73.13	82.40	83.17	92.59	93.00	82.65	83.82	92.47	92.54
tr23	62.11	58.54	79.07	75.61	83.07	83.87	88.32	88.71	85.10	85.37	89.78	90.24
tr31	88.83	87.63	92.71	92.43	95.85	96.06	97.82	97.84	96.21	96.24	98.92	98.92
tr41	90.62	89.77	90.48	90.23	93.31	93.94	96.57	96.55	94.88	95.45	95.90	95.98
tr45	79.92	81.88	91.83	92.03	89.63	89.86	90.30	90.34	89.61	89.86	93.41	93.48
wap	70.68	72.44	82.01	83.60	73.17	73.50	78.01	78.37	76.06	76.92	82.68	83.92
Average	77.53	77.50	**82.41**	**82.28**	79.96	80.30	**84.52**	**84.65**	82.22	82.64	**86.69**	**86.95**

Table 6: F1-score and classification accuracy comparisons of Chi-square filter FS based on TF and the proposed MTF-MICF based on MNB, SVM, and LR.

Dataset	MNB				SVM				LR			
	TF		MTF-MICF		TF		MTF-MICF		TF		MTF-MICF	
	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.
fbis	77.13	76.47	80.60	80.53	78.86	78.89	78.92	79.30	80.54	80.73	83.38	84.18
la1s	84.53	84.56	85.52	85.34	84.13	84.10	87.65	87.73	85.69	85.65	88.28	88.30
la2s	84.45	84.55	87.16	86.99	84.41	84.40	87.55	87.54	87.20	87.32	89.18	89.27
new3s	72.36	72.33	76.03	75.99	76.74	77.06	80.42	80.68	78.58	78.82	81.95	82.11
oh0	88.24	88.06	92.05	92.04	86.59	86.38	87.81	87.71	87.74	87.56	90.58	90.55
oh10	80.99	80.95	82.37	82.86	79.43	80.00	80.24	80.00	79.20	79.52	82.89	82.86
oh15	84.57	84.70	87.96	87.98	83.15	83.21	82.48	82.12	82.63	82.51	87.98	87.98
oh5	85.22	85.33	88.90	89.13	82.61	82.61	85.01	85.14	85.69	85.87	87.26	87.50
ohscal	73.44	73.22	73.74	73.58	78.09	78.14	79.54	79.55	78.40	78.41	78.99	78.91
re0	82.04	81.06	80.15	79.73	84.00	84.29	84.40	84.51	85.09	85.38	87.80	88.04
re1	85.48	85.54	83.92	83.73	86.25	85.60	85.54	86.35	88.91	89.16	86.67	87.65
tr11	85.48	85.54	91.94	91.46	86.25	85.60	89.64	89.43	88.91	89.16	90.26	90.24
tr12	88.70	88.89	88.34	88.89	88.72	88.30	83.26	82.98	85.42	85.71	87.13	87.30
tr21	71.74	67.65	90.18	89.55	88.24	89.11	94.99	95.00	90.17	91.18	97.01	97.01
tr23	72.01	68.29	93.02	92.68	83.02	83.87	91.52	91.94	89.90	90.24	97.52	97.56
tr31	95.04	95.16	98.92	98.92	96.96	97.13	97.84	97.84	98.89	98.92	100	100
tr41	92.74	92.61	93.79	93.68	96.14	96.21	97.32	97.32	96.44	96.59	97.04	97.13
tr45	83.76	83.33	93.53	93.48	91.44	90.82	94.14	94.20	92.84	92.75	95.62	95.65
wap	74.66	75.64	77.25	79.10	75.00	75.00	77.35	77.94	77.11	77.88	78.43	79.74
Average	77.13	76.47	**86.60**	**86.61**	84.74	84.77	**86.61**	**86.70**	86.28	86.49	**88.84**	**89.05**

5 Conclusion and Future Work

This study presents a comprehensive investigation into the joint impact of supervised, class-aware term weighting (i.e., MTF-MICF) and filter-based FS (i.e., IG, CHI, and ReliefF) on classification performance for multi-class and imbalanced TC tasks. Rather than treating these two stages as independent preprocessing steps, the proposed framework explicitly integrates them into a unified pipeline, where improved representation quality directly facilitates more effective, stable, and discriminative feature subset selection. Extensive experimental evaluations across nineteen benchmark datasets demonstrate that filter-based FS, combined with MTF-MICF, consistently improves both F1-score and classification accuracy across different classifiers, feature subset sizes, and dimensionality reduction ratios. The results confirm that the proposed approach is not only accurate and scalable, but also highly effective in preserving discriminative information under both aggressive and moderate feature reduction settings. In particular, the analysis across varying feature dimensionalities shows that MTF-MICF maintains stable performance gains, highlighting its generalizability and suitability for small and medium-sized training datasets where feature sparsity and class imbalance are more pronounced. These findings provide strong evidence that class-aware term weighting, when tightly coupled with FS, should be regarded as a core component of modern TC systems rather than merely a preprocessing refinement. Despite the consistent performance of the proposed framwork, several promising research directions remain. Future work will focus on integrating MTF-MICF with ensemble multi-filter and metaheuristic optimization frameworks to further enhance FS constancy and performance, and on extending it to deep learning, cross-domain, and low-resource TC settings. In addition, a systematic sensitivity analysis of the balancing parameter α used in the MICF formulation will be conducted to further investigate its influence on classification performance under different data characteristics, such as varying levels of sparsity, class imbalance, and vocabulary size.

Acknowledgement: We would like to express our sincere gratitude to the School of Computer Sciences, Universiti Sains Malaysia (USM), for their encouragement, support, and the conducive academic environment provided throughout this work.

Funding Statement: This work was supported by the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education Malaysia, under Grant FRGS/1/2023/ICT02/USM/02/4.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Osamah Mohammed Alyasiri; data collection: Osamah Mohammed Alyasiri; analysis and interpretation of results: Osamah Mohammed Alyasiri, Yu-N Cheah; writing—original draft preparation, Osamah Mohammed Alyasiri; writing—review and editing, Osamah Mohammed Alyasiri, Yu-N Cheah; supervision, Yu-N Cheah. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in [WEKA platform website] at [https://sourceforge.net/projects/weka/files/datasets/text-datasets/19MclassTextWc.zip/download].

Ethics Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ayash L, Algarni A, Alqahtani O. Advancements in feature selection and extraction methods for text mining: a review. Discov Appl Sci. 2025;7(8):914. doi:10.1007/s42452-025-07587-w.
2. Allam H, Makubvure L, Gyamfi B, Graham KN, Akinwolere K. Text classification: how machine learning is revolutionizing text categorization. Information. 2025;16(2):130. doi:10.3390/info16020130.
3. Islam MR, Lima AA, Das SC, Mridha MF, Prodeep AR, Watanobe Y. A comprehensive survey on the process, methods, evaluation, and challenges of feature selection. IEEE Access. 2022;10(8):99595–632. doi:10.1109/access.2022.3205618.
4. Dhal P, Azad C. A comprehensive survey on feature selection in the various fields of machine learning. Appl Intell. 2022;52(4):4543–81. doi:10.1007/s10489-021-02550-9.
5. Song X, Zhang Y, Zhang W, He C, Hu Y, Wang J, et al. Evolutionary computation for feature selection in classification: a comprehensive survey of solutions, applications and challenges. Swarm Evol Comput. 2024;90(4):101661. doi:10.1016/j.swevo.2024.101661.
6. Pintas JT, Fernandes LAF, Garcia ACB. Feature selection methods for text classification: a systematic literature review. Artif Intell Rev. 2021;54(8):6149–200. doi:10.1007/s10462-021-09970-6.
7. Dhar A, Mukherjee H, Dash NS, Roy K. Text categorization: past and present. Artif Intell Rev. 2020;54(4):3007–54. doi:10.1007/s10462-020-09919-1.
8. Alyasiri OM, Cheah YN, Zhang H, Al-Janabi OM, Abasi AK. Text classification based on optimization feature selection methods: a review and future directions. Multimed Tools Appl. 2025;84(15):14187–233. doi:10.1007/s11042-024-19769-6.
9. Nafis NSM, Awang S. The impact of pre-processing and feature selection on text classification. In: Advances in electronics engineering. Singapore: Springer; 2019. p. 269–80. doi:10.1007/978-981-15-1289-6_25.

10. Hong M, Wang H. Filter feature selection methods for text classification: a review. Multimed Tools Appl. 2024;83(1):2053–91. doi:10.1007/s11042-023-15675-5.
11. Theng D, Bhoyar KK. Feature selection techniques for machine learning: a survey of more than two decades of research. Knowl Inf Syst. 2024;66(3):1575–637. doi:10.1007/s10115-023-02010-5.
12. HaCohen-Kerner Y, Miller D, Yigal Y. The influence of preprocessing on text classification using a bag-of-words representation. PLoS One. 2020;15(5):e0232525. doi:10.1371/journal.pone.0232525.
13. Samant SS, Bhanu Murthy NL, Malapati A. Improving term weighting schemes for short text classification in vector space model. IEEE Access. 2019;7:166578–92. doi:10.1109/ACCESS.2019.2953918.
14. Chen L, Jiang L, Li C. Using modified term frequency to improve term weighting for text classification. Eng Appl Artif Intell. 2021;101(2–3):104215. doi:10.1016/j.engappai.2021.104215.
15. Okkalioglu M. TF-IGM revisited: imbalance text classification with relative imbalance ratio. Expert Syst Appl. 2023;217(6):119578. doi:10.1016/j.eswa.2023.119578.
16. Alsaeedi A. A survey of term weighting schemes for text classification. Int J Data Min Model Manag. 2020;12(2):237. doi:10.1504/ijdmmm.2020.106741.
17. Mazyad A, Teytaud F, Fonlupt C. A comparative study on term weighting schemes for text classification. In: Machine learning, optimization, and big data. Cham, Switzerland: Springer; 2018. p. 100–8. doi:10.1007/978-3-319-72926-8_9.
18. Rathi RN, Mustafi A. The importance of Term Weighting in semantic understanding of text: a review of techniques. Multimed Tools Appl. 2023;82(7):9761–83. doi:10.1007/s11042-022-12538-3.
19. Thirumoorthy K, Muneeswaran K. Optimal feature subset selection using hybrid binary Jaya optimization algorithm for text classification. Sādhanā. 2020;45(1):201. doi:10.1007/s12046-020-01443-w.
20. Çekik R. Effective text classification through supervised rough set-based term weighting. Symmetry. 2025;17(1):90. doi:10.3390/sym17010090.
21. Alyasiri OM, Cheah YN, Abasi AK, Al-Janabi OM. Wrapper and hybrid feature selection methods using metaheuristic algorithms for English text classification: a systematic review. IEEE Access. 2022;10(1):39833–52. doi:10.1109/ACCESS.2022.3165814.
22. Farek L, Benaidja A. A hybrid feature selection method for text classification using a feature-correlation-based genetic algorithm. Soft Comput. 2024;28(23):13567–93. doi:10.1007/s00500-024-10386-x.
23. Attieh J, Tekli J. Supervised term-category feature weighting for improved text classification. Knowl Based Syst. 2023;261(1):110215. doi:10.1016/j.knosys.2022.110215.
24. Aditya CSK, Sumadi FDS. Combination of term weighting with class distribution and centroid-based approach for document classification. Kinetik. 2023;8(4):781–8. doi:10.22219/kinetik.v8i4.
25. Okkalioglu M. A novel redistribution-based feature selection for text classification. Expert Syst Appl. 2024;246(2):123119. doi:10.1016/j.eswa.2023.123119.
26. Shehzad F, Rehman A, Javed K, Alnowibet KA, Babri HA, Rauf HT. Binned term count: an alternative to term frequency for text categorization. Mathematics. 2022;10(21):4124. doi:10.3390/math10214124.
27. Das M, Kamalanathan S, Alphonse P. A comparative study on TF-IDF feature weighting method and its analysis using unstructured dataset. arXiv:2308.04037. 2023.
28. Jiang Z, Gao B, He Y, Han Y, Doyle P, Zhu Q. Text classification using novel term weighting scheme-based improved TF-IDF for internet media reports. Math Probl Eng. 2021;2021(1):6619088. doi:10.1155/2021/6619088.
29. Naderalvojoud B, Akcapinar Sezer E. Term evaluation metrics in imbalanced text categorization. Nat Lang Eng. 2020;26(1):31–47. doi:10.1017/s1351324919000317.
30. Sabbah T, Selamat A, Selamat MH, Al-Anzi FS, Viedma EH, Krejcar O, et al. Modified frequency-based term weighting schemes for text classification. Appl Soft Comput. 2017;58(1):193–206. doi:10.1016/j.asoc.2017.04.069.
31. Tang Z, Li W, Li Y. An improved supervised term weighting scheme for text representation and classification. Expert Syst Appl. 2022;189(24):115985. doi:10.1016/j.eswa.2021.115985.
32. Debole F, Sebastiani F. Supervised term weighting for automated text categorization. In: Text mining and its applications. Berlin/Heidelberg, Germany: Springer; 2004. p. 81–97. doi:10.1007/978-3-540-45219-5_7.

33. Liu Y, Loh HT, Sun A. Imbalanced text classification: a term weighting approach. Expert Syst Appl. 2009;36(1):690–701. doi:10.1016/j.eswa.2007.10.042.
34. Altınçay H, Erenel Z. Analytical evaluation of term weighting schemes for text categorization. Pattern Recognit Lett. 2010;31(11):1310–23. doi:10.1016/j.patrec.2010.03.012.
35. Lertnattee V, Theeramunkong T. Analysis of inverse class frequency in centroid-based text classification. In: Proceedings of the IEEE International Symposium on Communications and Information Technology; 2004 Oct 26–29; Sapporo, Japan. p. 1171–6. doi:10.1109/ISCIT.2004.1413903.
36. Ren F, Sohrab MG. Class-indexing-based term weighting for automatic text classification. Inf Sci. 2013;236(5):109–25. doi:10.1016/j.ins.2013.02.029.
37. Chen K, Zhang Z, Long J, Zhang H. Turning from TF-IDF to TF-IGM for term weighting in text classification. Expert Syst Appl. 2016;66(11):245–60. doi:10.1016/j.eswa.2016.09.009.
38. Quan X, Liu W, Qiu B. Term weighting schemes for question categorization. IEEE Trans Pattern Anal Mach Intell. 2011;33(5):1009–21. doi:10.1109/TPAMI.2010.154.
39. Lan M, Tan CL, Su J, Lu Y. Supervised and traditional term weighting methods for automatic text categorization. IEEE Trans Pattern Anal Mach Intell. 2009;31(4):721–35. doi:10.1109/TPAMI.2008.110.
40. Dogan T, Uysal AK. Improved inverse gravity moment term weighting for text classification. Expert Syst Appl. 2019;130(2):45–59. doi:10.1016/j.eswa.2019.04.015.
41. Chen L, Jiang L, Li C. Modified DFS-based term weighting scheme for text classification. Expert Syst Appl. 2021;168(2–3):114438. doi:10.1016/j.eswa.2020.114438.
42. Wang H, Hong M. Supervised Hebb rule based feature selection for text classification. Inf Process Manag. 2019;56(1):167–91. doi:10.1016/j.ipm.2018.09.004.
43. Palacharla RK, Vatsavayi VK. A novel approach for text classification using feature selection algorithm and term weight measures. In: Accelerating discoveries in data science and artificial intelligence I. Cham, Switzerland: Springer; 2024. p. 287–301. doi:10.1007/978-3-031-51167-7_28.
44. Iqbal M, Abid MM, Khalid MN, Manzoor A. Review of feature selection methods for text classification. Int J Adv Comput Res. 2020;10(49):138–52. doi:10.19101/ijacr.2020.1048037.
45. Dogra V, Singh A, Verma S, Kavita, Jhanjhi NZ, Talib MN. Understanding of data preprocessing for dimensionality reduction using feature selection techniques in text classification. In: Intelligent computing and innovation on data science. Singapore: Springer; 2021. p. 455–64. doi:10.1007/978-981-16-3153-5_48.
46. Ashokkumar P, Siva Shankar G, Srivastava G, Maddikunta PKR, Gadekallu TR. A two-stage text feature selection algorithm for improving text classification. ACM Trans Asian Low-Resour Lang Inf Process. 2021;20(3):1–19. doi:10.1145/3425781.
47. Labani M, Moradi P, Jalili M. A multi-objective genetic algorithm for text feature selection using the relative discriminative criterion. Expert Syst Appl. 2020;149:113276. doi:10.1016/j.eswa.2020.113276.
48. Cekik R. A new filter feature selection method for text classification. IEEE Access. 2024;12(1):139316–35. doi:10.1109/ACCESS.2024.3468001.
49. Al-Shalif SA, Senan N, Saeed F, Ghaban W, Ibrahim N, Aamir M, et al. A systematic literature review on meta-heuristic based feature selection techniques for text classification. PeerJ Comput Sci. 2024;10(3):e2084. doi:10.7717/peerj-cs.2084.
50. Farek L, Benaidja A. An optimal feature selection method for text classification through redundancy and synergy analysis. Multimed Tools Appl. 2025;84(16):16397–423. doi:10.1007/s11042-024-19736-1.
51. Cascaro RJ, Gerardo BD, Medina RP. Filter selection methods for multiclass classification. In: Proceedings of the 2nd International Conference on Computing and Big Data; 2019 Oct 18–20; Taichung, Taiwan. p. 27–31. doi:10.1145/3366650.3366655.
52. Lazhar F, Amira B. Semantic similarity-aware feature selection and redundancy removal for text classification using joint mutual information. Knowl Inf Syst. 2024;66(10):6187–212. doi:10.1007/s10115-024-02143-1.
53. Farek L, Benaidja A. A non-redundant feature selection method for text categorization based on term co-occurrence frequency and mutual information. Multimed Tools Appl. 2024;83(7):20193–214. doi:10.1007/s11042-023-15876-y.

54. Thirumoorthy K, Muneeswaran K. Feature selection for text classification using machine learning approaches. Natl Acad Sci Lett. 2022;45(1):51–6. doi:10.1007/s40009-021-01043-0.
55. Badawi SS, Saeed AM, Ahmed SA, Hassan DA. Enhanced category-feature association measure: a robust approach for text classification through feature selection. Aro Sci J Koya Univ. 2025;13(2):114–23. doi:10.14500/aro.12034.
56. Farghaly HM, Abd El-Hafeez T. A new feature selection method based on frequent and associated itemsets for text classification. Concurr Comput Pract Exp. 2022;34(25):e7258. doi:10.1002/cpe.7258.
57. Singh KN, Devi HM, Mahant AK, Dorendro A. FE-TAC: an effective document classification method combining feature extraction and feature selection. Int J Appl Decis Sci. 2023;16(6):717–40. doi:10.1504/ijads.2023.134204.
58. Bolón-Canedo V, Alonso-Betanzos A. Ensembles for feature selection: a review and future trends. Inf Fusion. 2019;52(2):1–12. doi:10.1016/j.inffus.2018.11.008.
59. Seijo-Pardo B, Porto-Díaz I, Bolón-Canedo V, Alonso-Betanzos A. Ensemble feature selection: homogeneous and heterogeneous approaches. Knowl Based Syst. 2017;118(2):124–39. doi:10.1016/j.knosys.2016.11.017.
60. Fu G, Li B, Yang Y, Li C. Re-ranking and TOPSIS-based ensemble feature selection with multi-stage aggregation for text categorization. Pattern Recognit Lett. 2023;168(3):47–56. doi:10.1016/j.patrec.2023.02.027.
61. Theng D, Bhoyar KK. Stability of feature selection algorithms. In: Artificial intelligence on medical data. Singapore: Springer; 2023. p. 299–316. doi:10.1007/978-981-19-0151-5_26.
62. Kaya C, Kilimci ZH, Uysal M, Kaya M. A review of metaheuristic optimization techniques in text classification. Int J Comput Exp Sci Eng. 2024;10(2):126–33. doi:10.22399/ijcesen.295.
63. Thirumoorthy K, Britto JJ. A two-stage feature selection approach using hybrid elitist self-adaptive cat and mouse based optimization algorithm for document classification. Expert Syst Appl. 2024;254:124396. doi:10.1016/j.eswa.2024.124396.
64. Kaya C, Kilimci ZH, Uysal M, Kaya M. Migrating birds optimization-based feature selection for text classification. PeerJ Comput Sci. 2024;10(4):e2263. doi:10.7717/peerj-cs.2263.
65. Farek L, Benaidja A. An adaptive binary particle swarm optimization algorithm with filtration and local search for feature selection in text classification. Memet Comput. 2025;17(4):45. doi:10.1007/s12293-025-00481-3.
66. Alyasiri OM, Cheah YN, Abasi AK. Hybrid filter-wrapper text feature selection technique for text classification. In: Proceedings of the 2021 International Conference on Communication & Information Technology (ICICT); 2021 Jun 5–6; Basrah, Iraq. p. 80–6. doi:10.1109/ICICT52195.2021.9567898.
67. Thirumoorthy K, Muneeswaran K. Feature selection using hybrid poor and rich optimization algorithm for text classification. Pattern Recognit Lett. 2021;147(10):63–70. doi:10.1016/j.patrec.2021.03.034.
68. Ige OP, Gan KH. Ensemble filter-wrapper text feature selection methods for text classification. Comput Model Eng Sci. 2024;141(2):1847–65. doi:10.32604/cmes.2024.053373.
69. Szeghalmy S, Fazekas A. A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. Sensors. 2023;23(4):2333. doi:10.3390/s23042333.
70. Rossi RG, Marcacini RM, Rezende SO. Benchmarking text collections for classification and clustering tasks [Internet]. 2013 [cited 2026 Jan 1]. Available from: https://repositorio.usp.br/directbitstream/342060e9-eebc-4530-8074-bd60bb8b125e/Relat%C3%B3rios%20T%C3%A9cnicas_395.
71. Dönicke T, Damaschk M, Lux F. Multiclass text classification on unbalanced, sparse and noisy data. In: Proceedings of the First NLPL Workshop on Deep Learning for Natural Language Processing; 2019 Sep 30; Turku, Finland. p. 58–65.
72. Unnikrishnan P, Govindan VK, Madhu Kumar SD. Enhanced sparse representation classifier for text classification. Expert Syst Appl. 2019;129(3):260–72. doi:10.1016/j.eswa.2019.04.003.
73. Tang Z. A generic multi-level framework for building term-weighting schemes in text classification. Comput J. 2024;67(11):3042–55. doi:10.1093/comjnl/bxae068.
74. Jin L, Zhang L, Zhao L. Feature selection based on absolute deviation factor for text classification. Inf Process Manag. 2023;60(3):103251. doi:10.1016/j.ipm.2022.103251.
75. Li C, Li W, Tang Z, Li S, Xiang H. An improved term weighting method based on relevance frequency for text classification. Soft Comput. 2023;27(7):3563–79. doi:10.1007/s00500-022-07597-5.
76. Alyasiri OM, Cheah YN. Multi-class text classification using machine learning techniques. Eng Technol Appl Sci Res. 2025;15(3):22598–604. doi:10.48084/etasr.9994.