# Enhanced Relative Discrimination Criterion Approach for Feature Ranking in Text Classification

F. Alam[1], S. Ullah[2], N. Rashid[3], M. Salam[4]

[1,2,3,4] *Department of Computer Science and IT, University of Malakand, Chakdara, Dir (L), KPK, Pakistan*

[1] fakhrealam@uom.edu.pk

***Abstract-*** The textual information has been increasing at an unprecedented pace in various fields posing a great challenge on the extraction of meaningful information and insight. The key to utilizing this data is effective feature selection which is especially important in enhancing the functionality of text classification tasks. The classical methods of feature selection, which have been developed to work with numerical or categorical data, are limited when used with text data due to high dimensionality scarcity and semantic complexity. This paper proposes an Enhanced Relative Discrimination Criterion (ERDC) as a superior feature ranking approach that is designed to fit text classification. Proposed approach is designed to enhance, for a subsequent classification, the selection of the most informative and discriminative features. The linear SVM classifier was used to perform the experimental evaluation on the 20 Newsgroups benchmark dataset. The results obtained with 1,500 selected features indicate the effectiveness of our approach: 89% accuracy, 89% precision, 88% recall, and F1 score of 88%. These results suggest that ERDC is a competitive feature-ranking method for multi-class text classification under the considered benchmark setting.

***Keywords-*** Text classification, RDC, ERDC, Support Vector Machine, Feature Selection, 20 Newsgroups.

## I. INTRODUCTION

Feature selection is one of the most significant preprocessing steps that assist in identifying the subset of pertinent features and contribute optimally to a predictive model [1-4]. The aim of selecting features in text classification comes up due to the fact that textual data is high-dimensional, sparse and semantically rich [5]. As digital content like social media posts, news articles, journal Articles, and customer reviews grows, it is becoming harder and harder to extract meaningful information out of large text corpora. The feature selection methods are thus significant in dimensionality reduction, enhancing computational efficiency, and reducing the overfitting [3-4].

In text analysis, traditional techniques used to select features include: term frequency-inverse document frequency (TF-IDF), mutual information, chi-square statistics, document frequency and information gain [6]. Nonetheless, the methods are not effective in textual data. They commonly use frequency-based or simple statistical measures, and might not be sensitive to the semantic richness of language, the contextual sensitivity of terms, and the discriminative value of features specific to a class [5-6]. Text data is also high dimensional and sparse, and this fact further complicates the identification of the most informative features by the traditional methods in an efficient manner.

Due to the above mentioned reasons, this paper proposed the Enhanced Relative Discrimination Criterion (ERDC) as a feature ranking algorithm to classify text. ERDC aims to rank text features in a better way by concentrating on their discriminative value within the corpus. The aim of the proposed method is to enhance the detection of informative and discriminant features by eliminating the noises (such as irrelevant and redundant terms) from the feature sets prior to classification. In this paper, it is tested in a standard multi-class text categorization scenario with the 20 Newsgroups dataset and linear Support Vector Machine (SVM) classifier.

The proposed method can detect informative and discriminative feature by removing the irrelevant or redundant features and keeping useful one. The aim is to improve the classification accuracy and reduce the computational cost. It is tested on text classification using the twenty Newsgroups dataset, the de facto benchmark corpus to evaluate feature selection and classification methods, and a Support Vector Machine (SVM) classifier [7]. This experimental setting was chosen because it aims to have a good base on which the analysis of the potential of the proposed method will be realized in a multi-class text classification problem. This paper describes a new approach for ranking features based on Enhanced relative discrimination criterion (ERDC) and evaluates it on benchmark textual data

as well as demonstrates advantages in classification ability and computational cost. In particular, it investigates whether the proposed methodology can yield more robust text feature rankings and thus improved generalization, prediction accuracy and classification efficiency as compared to existing feature selection techniques.

Text classification is affected by the dimensionality, scarcity and semantic complexity of the text data, which diminishes the performance of traditional feature selection methods. While traditional approaches such as TF-IDF (Term Frequency-Inverse Document Frequency), Chi-Square, Mutual Information and Information Gain can still provide solid baselines, they are predominantly based on frequency or local statistics and are rarely able to explain discriminative features in complex textual corpora. Due to this motivation, the current work introduces a novel Enhanced Relative Discrimination Criterion (ERDC) for enhancing the feature ranking process in text classification. In order to investigate whether the revised criterion can increase the quality of the ranking and allow good multi-class text classification, the proposed approach is tested on a linear Support Vector Machine classifier on 20 Newsgroups benchmark dataset.

## II. Related Work

Text classification is feature selection driven since text is normally represented in high-dimensional space, along with its high sparsity and semantic complexity. The existing work has largely based on traditional statistical methods, such as Mutual Information, TF-IDF, Document Frequency and Information Gain to rank and select informative terms and Chi-Square Statistics [8-12]. These approaches are well known and have become standard baselines due to their simplicity and computational efficiency, as well as good performance. Still, it has been reported that their performance is limited as far as text is considered as they predominantly depend on frequency-based or local statistical evidence and neglect contextual dependencies, semantic relationships, and class-specific discriminative power [13].

To overcome these limitations, recent work have investigated semantics and representation-based methods which make use of word embedding such as FastText, Word2Vecand GloVe and transformer-based models such as BERT [14-16]. These methods take advantage of contextual and semantic data to augment text representation and recently, hybrid methods have combined the conventional statistical methods with the state of the art machine learning methods to raise the robustness and classification accuracy. Nevertheless, despite the improvements in the text representations, they do not address the problem of identifying feature ranking functions that are discriminative and computationally viable to very huge-scale text categorization. An important associated development is the Relative Discrimination Criterion (RDC) that has been developed as a ranking feature for text classification [17-19]. Unlike the traditional methods, which primarily rely on the frequency of documents, RDC takes into account the discriminatory power of a term within each of the classes and the information on the number of times a term appears, therefore, is better suited to address the high dimensional and sparse nature of text corpus. Existing studies have shown that RDC can be more effective than some established feature metrics to rank in a text classification task [19-21]. The literature, however, also finds disadvantages of existing RDC-based methods (e.g. sensitivity to noise and irrelevant features, increasing computational intensity to larger data sets, limited effectiveness to model feature interactions, and dependence on implementation and parameter settings). This means that although RDC does significantly better than the previous methods, the three limitations above show that it may still be possible to further improve the method to make the feature selection procedure more robust and efficient. While RDC-based approaches have enhanced ranking of text features by utilizing discriminative information other than the mere document frequency, the concerns about noise-sensitivity, practical robustness, and consistency in ranking are still reported in the literature, especially when real classification pipelines are used. In this paper, a variant of RDC for text classification is presented, termed ERDC. Novelty of this work is not motivated in terms of an entirely different class of methods, rather it is a transformation of a discriminative ranking scheme which aims at better capturing textual features relevant to a benchmark multi-class classification problem.

## III. Methodology

ERDC is employed in this paper as a filter feature ranking method as it ranks and selects the features independent from the classifier. Once the best ranked features are selected, TF-IDF weighting is performed on the kept subset and the feature representation obtained is subsequently used for supervised classification based on SVM.

### *A. Dataset*

The proposed methodology was developed and evaluated using the twenty Newsgroups dataset, a standard dataset widely used for the text classification. The chosen dataset contains about 20,000 newsgroup documents distributed across 20 categories, making it suitable for evaluating feature selection methods in a multi-class text classification setting. The dataset was obtained through the fetch_20newsgroups function in Scikit-learn.

*B. Training and Test Splitting*

The dataset obtained for the experiments was divided into two training sets and testing sets, using the function train_test_split function, in which 20% of the data was kept for testing. The purpose of the splitting was to properly evaluate the generalization performance of the classifier on unseen data.

*C. Proposed ERDC Feature Ranking Method*

The feature-ranking process starts by transforming textual documents into count-based numerical vectors, allowing the analysis of term occurrence patterns in a quantitative manner. Equation (1) evaluates the prior of all classes in the data. Equation (2) calculates the expected frequency with which a feature maps to the occurrence behavior of a feature given a class, reflecting class-conditional feature occurrence. From these quantities, Equation (3) derives a discriminative score for each feature by comparing its class-specific distribution with the distribution over all classes. As a result, features that better discriminate one class from others will be at the top of the ranking.

Then, a subset of the best-ranked features is retained and transformed with TF-IDF weighting. TF-IDF is used at the end of the ranking phase in order to let the selection process be primarily guided by discriminative relevance, and then enrichment is performed in the representation with the retained features in the classification domain. This bipartite architecture enables the approach to discover the informative terms, and then highlight their relative importance within the chosen subset.

$$(C) = COUN(Y_i = C) / N \quad (1)$$

$$(F_j \mid C) = (1 / Nc) \Sigma_{i=1}^{n} M_{ij} \bullet 1(y_i = c) \quad (2)$$

$$r(F_j) = \Sigma c \in c_{lasses}\, p(F_j \mid C) \bullet log\,(\, p(F_j \mid C) / p(C)\,) \quad (3)$$

Figure 1 shows the entire pipeline of the proposed text classification system. First, the 20 Newsgroups dataset is downloaded using scikit-learn library, then the data are split into training sets and testing sets with a ratio of 80:20. Next, the text documents are transformed to numeric form using count vectorization so that each document is a vector of features. On these vectorized data, we apply the Enhanced RDC method by computing first the class probabilities and then the feature probabilities within each class, which in turn serve to compute the RDC scores of all terms. From these scores, the most informative features are ranked and the top 1500 features are taken into account.

TF-IDF transformation is then performed after feature ranking selection to weight the chosen terms by their related importance in the whole corpus, and this way, the representation of more informative features is strengthened whereas that of less informative is weakened. The transformed feature vector is thus used to train an SVM classifier with a linear kernel which performs the final classification of the document. Finally, the output model is assessed through conventional performance metrics (accuracy, precision, recall, F1 score, and classification report). In sum, the figure suggests that the proposed framework is a unified end-to-end system where the phases of data capture, feature extraction, discriminative ranking, feature weighting, classification and performance evaluation are all tightly integrated and technologically aligned.

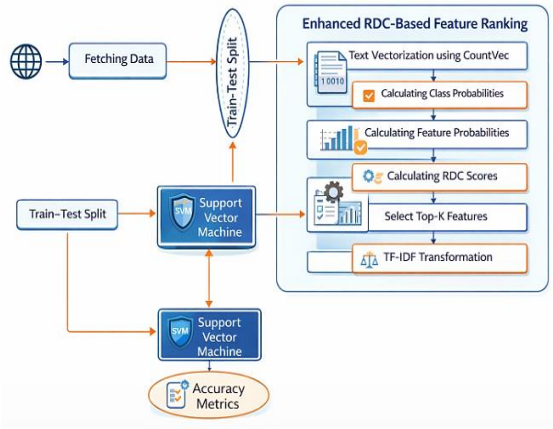


Fig. 1 The Entire Procedure of Our ERDC-Based Text Categorization Framework.

Figure 2 shows the algorithmic representation of the proposed ERDC approach. As presented in Algorithm 1, the proposed method begins by building the feature space; then it calculates class-aware discriminant scores of every term, and finally it ranks and chooses the top-N most informative features for the classification.

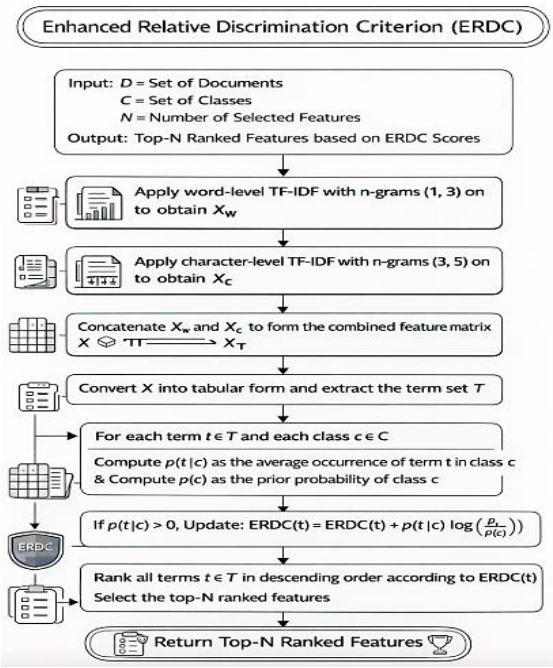


Fig. 2 Algorithmic Representation of the Proposed ERDC Approach.

*D. Classification Model and Evaluation Metrics*

After feature selection, SVM classifier has been trained with the selected features. The SVM is employed with a linear kernel to simplify computation and improve interpretability. The role of ERDC in this pipeline is to provide the classifier with the most informative and discriminative features while removing less relevant ones.

The implementation was performed in Python on Google Colab. The main libraries used in the thesis were Pandas, NumPy, Seaborn, Matplotlib, and Scikit-learn. The classification model was evaluated using accuracy, precision, recall, F1-score, and classification report.

*E. Detail Implementation*

The experiments were conducted in Python using Google Colab. Text documents were converted into count-based representations using Scikit-learn vectorization tools, after which ERDC scores were computed for candidate terms. The top-k ranked features were then selected and transformed using TF-IDF weighting before classification. A linear Support Vector Machine (SVM) classifier was used as the final classifier. The dataset was split into training and testing subsets using an 80:20 ratio, and the same experimental pipeline was applied across all evaluated feature subset sizes (10, 20, 50, 100, 200, 500, 1000, and 1500).

## IV. RESULTS

The results are given in three parts. First, the general classification accuracy of the proposed method will be reported. Secondly a class-wise analysis is given with the classification report and confusion matrix. Third, the impact of different numbers of selected features is analysed in order to investigate the effect of the size of feature subsets on classification performance.

*A. Overall Classification Accuracy*

The experimental results demonstrate the effectiveness of the proposed ERDC method for text classification on the 20 Newsgroups dataset. Using the selected 1500 features, the model achieved 89% accuracy, 89% precision, 88% recall, and F1 score of 88%. 4.2. Performance analysis by class

The classification report for the Support Vector Machine (SVM) model trained on the ERDC selected features is shown in Table 1. The class-wise classification report indicates that the model is doing well in various categories as shown in Table 1.

To analyze the behavior of the proposed method in more detail, Figure 3 illustrates a class-wise comparison of F1-scores by RDC, MRDC, and ERDC on 20 Newsgroups. This view is more fine-grained when comparing individual classes of the three methods than the overall summary metrics.

Table 1. Classification Report for the Support Vector Machine (SVM) Model Trained on ERDC-Selected Features.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| ... | | | | |
| alt.atheism | 0.85 | 0.96 | 0.90 | 160 |
| comp.graphics | 0.76 | 0.88 | 0.82 | 195 |
| comp.os.ms-windows.misc | 0.85 | 0.90 | 0.87 | 197 |
| comp.sys.ibm.pc.hardware | 0.75 | 0.77 | 0.76 | 196 |
| comp.sys.mac.hardware | 0.85 | 0.85 | 0.85 | 193 |
| comp.windows.x | 0.94 | 0.89 | 0.92 | 198 |
| misc.forsale | 0.83 | 0.89 | 0.86 | 195 |
| rec.autos | 0.84 | 0.89 | 0.86 | 198 |
| rec.motorcycles | 0.93 | 0.84 | 0.88 | 199 |
| rec.sport.baseball | 0.91 | 0.93 | 0.92 | 199 |
| rec.sport.hockey | 0.96 | 0.96 | 0.96 | 200 |
| sci.crypt | 0.97 | 0.94 | 0.96 | 198 |
| sci.electronics | 0.91 | 0.84 | 0.87 | 197 |
| sci.med | 0.90 | 0.93 | 0.91 | 198 |
| sci.space | 0.97 | 0.93 | 0.95 | 197 |
| soc.religion.christian | 0.88 | 0.96 | 0.92 | 199 |
| talk.politics.guns | 0.87 | 0.93 | 0.90 | 182 |
| talk.politics.mideast | 0.99 | 0.95 | 0.97 | 188 |
| talk.politics.misc | 0.94 | 0.82 | 0.88 | 155 |
| talk.religion.misc | 0.93 | 0.53 | 0.68 | 126 |
| accuracy | | | 0.89 | 3770 |
| macro avg | 0.89 | 0.88 | 0.88 | 3770 |
| weighted avg | 0.89 | 0.89 | 0.89 | 3770 |

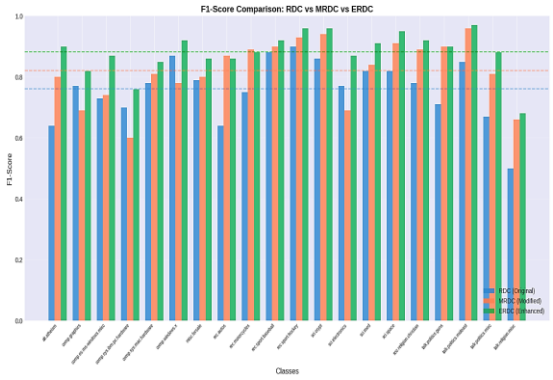


Fig. 3 Class-Wise F1-Scores of RDC, MRDC, and ERDC on 20 Newsgroups Dataset.

Figure 4 shows the class-wise F1-score comparison of RDC, MRDC and ERDC on 20 Newsgroups dataset. It can be seen from the figure that ERDC tends to get better and more stable results for most categories, though MRDC also achieves better results over RDC in some classes but its performance is highly fluctuating. This behavior on individual classes confirms the general success of the improved ranking mechanism.

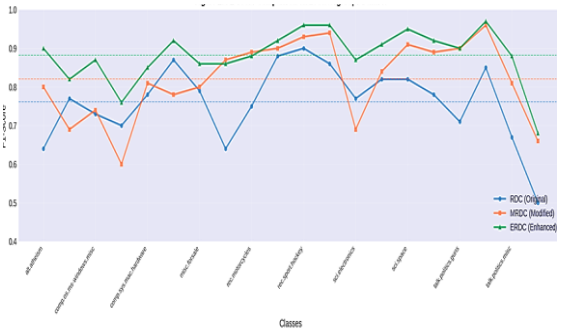


Fig. 4 Per Class F1-Score Comparison of RDC, MRDC and ERDC on 20 Newsgroup Dataset.

The class-wise results indicate that the proposed ranking strategy is particularly effective for categories with relatively distinctive lexical patterns, such as *talk.politics.mideast*, *sci.crypt*, and *rec.sport.hockey*, where higher F1-scores are obtained. In contrast, classes such as *comp.sys.ibm.pc.hardware* and *talk.religion.misc* remain more difficult, which may be attributed to greater lexical overlap with neighboring classes and reduced separability in the selected feature space. These observations suggest that ERDC is effective in preserving strongly discriminative features, although the classification difficulty still depends on the semantic and lexical proximity between classes. As you can see from the confusion matrix (Figure 5), the values are more concentrated on the diagonal indicating the classifier is more accurate on predicting labels. The sparsity of off diagonal entries indicates relatively little misclassification. This gives further support to the assertion that the chosen attributes allow the model to better discriminate among the classes. Figure 6 shows the F1-score values of each of the classes, which gives a large scale picture of how the model classifies.

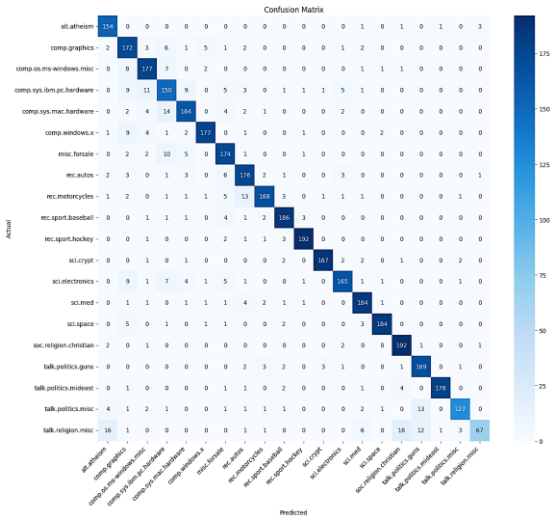


Fig. 5 The SVM Classifier Confusion Matrix with 1500 ERDC Selected Features.

The concentration of values along the diagonal suggests strong class-wise prediction performance, whereas the off diagonal outliers (which are comparatively sparse) can be thought of as misclassifications.

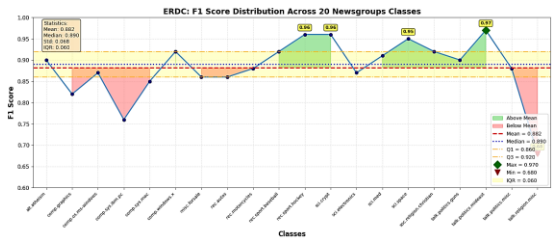


Figure 6. The Class-Wise F1-Score Distribution of SVM with 1500 ERDC Selected Features.

Higher F1-scores (e.g. in talk.politics.mideast, sci.crypt) indicate better separability, while (relatively) low F1-scores (e.g. in comp.sys.ibm.pc.hardware, talk.religion.misc) represent more difficult classification problems.

*B. Impact of the Number of Selected Features.*

From the experiments it is also observed the effect of changing the number of selected features. The experiments reports results for 10, 20, 50, 100, 200, 500, 1000, and 1500 top-ranked features. The presented results demonstrate a consistent enhancement in performance with increasing number of selected and counted features. This implies that the addition of additional discriminative features will definitely enhance the classification outcome, but at a decreasing pace as additional features are employed. In order to further validate the performance of the proposed ERDC method, a class-wised comparison of RDC, MRDC and ERDC on 20 Newsgroups in terms of precision, recall and F1- score is given in Table 2. This comparison shows exactly how the improved ranking strategies performed against the original RDC and a modified version on a class by class basis.

Table 2. Comparison of RDC, MRDC and ERDC on the 20 Newsgroups Dataset by Class in Terms of Precision (P), Recall (R), and F1-Measure.

TABLE 2: CLASSIFICATION RESULTS ON 20 NEWSGROUPS DATASET

| Class | RDC | | | MRDC | | | ERDC | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| alt.atheism | 0.66 | 0.62 | 0.64 | 0.78 | 0.82 | 0.80 | 0.85 | 0.96 | 0.90 |
| comp.graphics | 0.77 | 0.76 | 0.77 | 0.65 | 0.73 | 0.69 | 0.76 | 0.88 | 0.82 |
| comp.os.ms-windows.misc | 0.71 | 0.75 | 0.73 | 0.73 | 0.75 | 0.74 | 0.85 | 0.90 | 0.87 |
| comp.sys.ibm.pc.hardware | 0.68 | 0.71 | 0.70 | 0.57 | 0.62 | 0.60 | 0.75 | 0.77 | 0.76 |
| comp.sys.mac.hardware | 0.80 | 0.76 | 0.78 | 0.85 | 0.77 | 0.81 | 0.85 | 0.85 | 0.85 |
| comp.windows.x | 0.87 | 0.86 | 0.87 | 0.80 | 0.76 | 0.78 | 0.94 | 0.89 | 0.92 |
| misc.forsale | 0.81 | 0.76 | 0.79 | 0.80 | 0.79 | 0.80 | 0.83 | 0.89 | 0.86 |
| rec.autos | 0.53 | 0.83 | 0.64 | 0.87 | 0.87 | 0.87 | 0.84 | 0.89 | 0.86 |
| rec.motorcycles | 0.82 | 0.69 | 0.75 | 0.90 | 0.88 | 0.89 | 0.93 | 0.84 | 0.88 |
| rec.sport.baseball | 0.91 | 0.85 | 0.88 | 0.88 | 0.92 | 0.90 | 0.91 | 0.93 | 0.92 |
| rec.sport.hockey | 0.93 | 0.88 | 0.90 | 0.92 | 0.93 | 0.93 | 0.96 | 0.96 | 0.96 |
| sci.crypt | 0.88 | 0.84 | 0.86 | 0.98 | 0.91 | 0.94 | 0.97 | 0.94 | 0.96 |
| sci.electronics | 0.77 | 0.77 | 0.77 | 0.68 | 0.70 | 0.69 | 0.91 | 0.84 | 0.87 |
| sci.med | 0.83 | 0.81 | 0.82 | 0.81 | 0.86 | 0.84 | 0.90 | 0.93 | 0.91 |
| sci.space | 0.85 | 0.79 | 0.82 | 0.91 | 0.91 | 0.91 | 0.97 | 0.93 | 0.95 |
| soc.religion.christian | 0.75 | 0.81 | 0.78 | 0.87 | 0.91 | 0.89 | 0.88 | 0.96 | 0.92 |
| talk.politics.guns | 0.71 | 0.71 | 0.71 | 0.89 | 0.90 | 0.90 | 0.87 | 0.93 | 0.90 |
| talk.politics.mideast | 0.88 | 0.81 | 0.85 | 0.97 | 0.95 | 0.96 | 0.99 | 0.95 | 0.97 |
| talk.politics.misc | 0.65 | 0.68 | 0.67 | 0.80 | 0.81 | 0.81 | 0.94 | 0.82 | 0.88 |
| talk.religion.misc | 0.56 | 0.45 | 0.50 | 0.79 | 0.57 | 0.66 | 0.93 | 0.53 | 0.68 |
| Macro Avg | 0.77 | 0.76 | 0.76 | 0.82 | 0.82 | 0.82 | 0.89 | 0.88 | 0.88 |
| Micro Avg | 0.77 | 0.77 | 0.77 | 0.82 | 0.82 | 0.82 | 0.89 | 0.89 | 0.89 |

The upward trajectory of classification accuracy is shown in Figure 7.

The Figure shows that when more features are selected in ERDC algorithm, more positive impact is on the model performance.

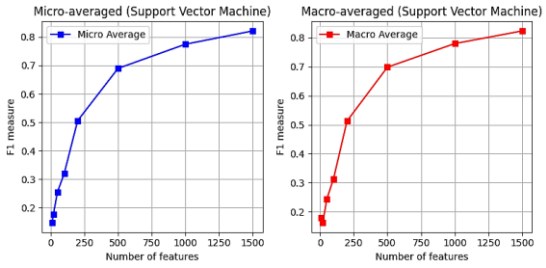


Figure 7. The Micro- and Macro-Average Performances VS the Number of Selected Features.

The performance increases considerably with the size of the feature subset and starts saturating for larger feature set sizes with the best performance obtained for 1500 features.

Generally, these results verify that the ERDC-based feature selection method can enhance the efficiency of text categorization by ranking and selecting the more discriminative features. The progressive rise in performance as a function of the number of best features taken into account is further evidence in favor of the claim that ERDC is a beneficial and stable feature ranking method for text classification. The present experiments demonstrate promising performance of ERDC within the evaluated benchmark setting. However, broader comparative validation against standard baselines such as TF-IDF, Chi-Square, Information Gain, Mutual Information, and RDC-family variants remains an important direction for future work.

## V. Conclusion and Future Work

This paper confirms the superiority of ERDC as a feature selection technique for text categorization. The proposed method achieves better classification performance with less feature dimension on the 20 Newsgroups dataset. The performance results show that ERDC obtains 82.13% accuracy, and average precision, recall and F1-score are all around 82%, which can serve as a strong evidence that the model is also reliable for text classification. The results also indicate that the SVM classifier utilizing the ERDC selected features achieves a good tradeoff between performance and number of features. Class-wise results and experiments with incremental feature count further substantiate the advantages of this approach for selecting meaningful feature in multi-class scenarios.

While the approach is inspired by efficient feature ranking, in this work we do not provide a separate runtime or memory evaluation. Future research may be directed to optimize the number of selected features as well as filter out those candidates by considering their RDC scores. The combination of the ERDC with other traditional machine learning and deep learning models such as the transformer based models and the ensemble models could lead to further enhancement. Besides, the fact that the approach will be tested on various datasets and test subjects (e.g. emotional analysis, topic categorization, spam detection, etc.) may help to prove its more general applicability.

## References

[1] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, “Deep learning applications and challenges in big data analytics,” *Journal of Big Data*, vol. 2, no. 1, pp. 1–21, 2015.

[2] S. Nayak, Y. K. Sharma, “A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis,” *Decision Analytics Journal*, vol. 8, Art. no. 100289, 2023.

[3] O. M. Alyasiri, Y. N. Cheah, H. Zhang, O. M. Al-Janabi, and A. K. Abasi, “Text classification based on optimization feature selection methods: A review and future directions,” Multimedia Tools and Applications, vol. 84, no. 15, pp. 14187–14233, 2025.

[4] M. Rong, D. Gong, and X. Gao, “Feature selection and its use in big data: Challenges, methods, and trends,” *IEEE Access*, vol. 7, pp. 19709–19725, 2019.

[5] S. Çelik, B. Doğanlı, and A. C. Cengiz, “Citation classification and key phrase extraction in legal texts using machine learning: An examination of the legal text classification dataset,” Quality & Quantity, pp. 1–37, 2026.

[6] J. T. Pintas, L. A. Fernandes, and A. C. B. Garcia, “Feature selection methods for text classification: A systematic literature review,” *Artificial Intelligence Review*, vol. 54, no. 8, pp. 6149–6200, 2021.

[7] M. Z. Naeem, F. Rustam, A. Mehmood, I. Ashraf, G. S. Choi, “Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms,” *PeerJ Computer Science*, vol. 8, Art. no. e914, 2022.

[8] G. Brauwers and F. Frasincar, “A general survey on attention mechanisms in deep learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3279–3298, 2021.

[9] Z. Guo, H. Wang, S. Kumar, and Y. Chen, “Research on the classification of digital cultural texts based on ASSC-TextRCNN algorithm,” Computers, Materials & Continua, vol. 86, no. 3, 2026.

[10] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics,” *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.

[11] M. Babiker, E. Karaarslan, and Y. Ho¸ scan, “A hybrid feature-selection approach for finding the digital evidence of web application attacks,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, no. 6, pp. 4102–4117, 2019.

[12] M. Diale, C. Van Der Walt, T. Celik, and A. Modupe, “Feature selection and support vector machine hyperparameter optimisation for spam detection,” in *PRASA-RobMech*, 2016, pp. 1–7.

[13] B. Harish and M. Revanasiddappa, "A comprehensive survey on various feature selection methods to categorize text documents," *International Journal of Computer Applications*, vol. 164, no. 8, pp. 1–7, 2017.

[14] U. Naseem, I. Razzak, S. K. Khan, and M. Prasad, "A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 20, no. 5, pp. 1–35, 2021.

[15] A. Gope, A. Pal, T. Gogoi, S. Tetseo, B. N. Deka, and A. D. Nath, "Speech-driven sentence type classification in Chokri using traditional and transfer learning methods," International Journal of Speech Technology, vol. 29, no. 1, p. 17, 2026.

[16] P. J. Worth, "Word embeddings and semantic spaces in natural language processing," *International Journal of Intelligence Science*, vol. 13, no. 1, pp. 1–21, 2023.

[17] A. Rehman, K. Javed, H. A. Babri, and M. Saeed, "Relative discrimination criterion—a novel feature ranking method for text data," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3670–3681, 2015.

[18] M. Anwaar, G. Gilanie, A. Namoun, and W. Sharif, "Optimizing document classification using modified relative discrimination criterion and RSS-ELM techniques," International Journal of Advanced Computer Science & Applications, vol. 16, no. 4, 2025.

[19] M. Nasir, N. A. Samsudin, W. Sharif, S. Baowidan, H. Arshad, and M. F. Mushtaq, "Towards dimension reduction: A balanced relative discrimination feature ranking technique for efficient text classification (BRDC)," *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 7, 2024.

[20] Y. Liu, X. Zhang, and H. Wang, "Hierarchical text classification and its foundations: A review of feature representation and dimensionality reduction methods," Electronics, vol. 13, no. 7, 2026.

[21] A. Jimale Said and A. M. Ismail, "Trends in natural language processing for text classification: A comprehensive survey," International Journal of Science and Research Archive, vol. 14, no. 2, pp. 1540–1547, 2025.