Optimizing Document Classification Using Modified Relative Discrimination Criterion and RSS-ELM Techniques

Muhammad Anwaar[1], Ghulam Gilanie[2], Abdallah Namoun[3], Wareesa Sharif[4]
Department of Artificial Intelligence-Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur, Pakistan[1, 2, 4]
AI Center-Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia[3]

***Abstract*—Internet content is increasing daily, and more data are being digitized due to technological advancements. Ever-increasing textual data in words, phrases, terms, sentences, and paragraphs pose significant challenges in classifying them effectively and require sophisticated techniques to arrange them automatically. The vast amount of textual data presents an opportunity to organise and extract valuable insights by identifying crucial pieces of information using feature selection techniques. Our article proposes "a Modified Relative Discrimination Criterion (MRDC) Technique and Ringed Seal Search-Extreme Learning Machine (RSS-ELM) to improve document classification", which prioritizes key data and fits corresponding documents into appropriate classes. The proposed MRDC and RSS-ELM techniques are compared with several existing techniques, such as the Relative Discrimination Criterion (RDC), the Improved Relative Discrimination Criterion (IRDC), GA-EM, and CS-ELM. The MRDC technique produced superior classification results with 91.60% accuracy compared to existing RDC and IRDC for feature selection. Moreover, the RSS-ELM optimization technique improved predictions significantly, with 98.9% accuracy compared to CS-ELM and GA-ELM on the Reuter21578 dataset.**

***Keywords*—Feature selection; relative discrimination criterion; ring seal search; extreme learning machine; metaheuristic algorithms; document classification; optimization**

I. Introduction

The quantity of textual content available on the internet is enormous and continuously growing. In recent years, technological evolution rapidly increased data drastically and significantly attracted researchers to find an optimum solution for text classification. The large amount of text data might make it challenging to efficiently organize and extract knowledge that is pertinent to our needs [1], [2]. Additionally, documents frequently have many features that add complexity to the process of developing a document classifier [3]. These characteristics negatively impact classification results. In addition, it can render classifiers ineffective [4]. The medium-sized dataset may contain more than 10,000 words with ease. The procedure of choosing features through a pre-processing phase helps to reduce the number of features and speeds up the classification process [5, 6]. The text data is frequently multiclass, and researchers frequently use it to help them choose pertinent features. An efficient feature ranking system is essential to improve the accuracy and performance of text categorization. Additionally, feature selection methods remove words from the corpus that are superfluous or unimportant [7]. Furthermore, the ability to choose elements in combination during the assessment of the classification process is a prerequisite for differentiation [8]. The wrapper, filter, and embedding methods are popular feature selection strategies. Since the filter methods just choose a feature subset that contains the most important information, they are independent of classification approaches. In the wrapper approach, a particular algorithm is used to choose features throughout the classification process, whereas in the embedded approach, a selected model is determined by integrating a particular feature selection methodology into the text classification process [9]. The filter approach is the best strategy among the three for classifying text. No algorithm is needed to choose a filter mechanism [10]. The existing RDC algorithm uses feature frequency to arrange text, giving high ranking to frequently appearing terms while ignoring rare terms that are equally relevant for categorization, which produces biased terms in feature selection [11]. There is a need for an algorithm that keeps the balance of frequent and rare terms to avoid these biased terms in the final feature selection.

The main data mining approaches are supervised and unsupervised [12]. Information in the supervised classification strategy is supported by outside sources, including class labels [13]. On the other hand, an unsupervised technique, also known as clustering, requires the system to execute classification without external sources. Numerous methods, including Naive Bayes, Multilayer Perceptron, Neural Networks, K-means, and ELM, are used in both approaches.

ELM yields excellent performance results and has a substantially faster convergence rate than previous approaches [14, 15]. The ELM method introduces biases in hidden layers and permits random values to be used to weigh them. Furthermore, the parameters don't alter during the training procedure. The weight between the output and hidden layers is the parameter that needs to be learned. Due to its lack of iteration, ELM has a high convergence rate [16, 17].

Metaheuristics are rules that enhance the likelihood of the best-optimized solution and aid in determining the best ideal solution to any given situation [18]. Furthermore, existing metaheuristic approaches such as Cuckoo Search (CS), Ringed Seal Search (RSS), and Genetic Algorithms (GA) are becoming more and more popular because they use several fields to get the best answer.

Furthermore, investigation and exploitation are crucial in resolving optimization issues. The exploitation refers to identifying a better solution or enhancement over the current one while exploring a region to find the optimal answer internationally [19]. Existing CS and GA methods cannot maintain an equilibrium between exploration and exploitation, but RSS performs better due to its two search states behavior. Furthermore, when compared to GA and CS methods, RSS employs fewer parameters.

II. Related Work

Text categorization is grouping texts into a predefined set of groups. For instance, categories like "politics, and sports" may be applied to incoming news stories. D = (d1, d2,..., dn) represents the training set, which has already been given class names like C1, C2, etc. (e.g., "sports," "politics") [20]. The stages of document categorization include pre-processing, indexing, feature selection, classifier, and performance measurement [21]. Feature selection is one of these, and it is crucial for increasing classification accuracy. Moreover, finding related characteristics or terms that set different document classes apart is one of its benefits. Classifiers can more precisely allocate new documents to the relevant categories when they have discriminative properties. Feature selection is an important phase in document classification since it can greatly increase classification accuracy [22].

As opposed to traditional feature ranking methods such as RDC, which tend to prioritize terms that occur frequently. Furthermore, by prioritizing rare phrases, the IRDC technique [23] seeks to improve classification performance. The proposed technique, a Modified Relative Discrimination Criterion (MRDC), is an improved methodology of RDC and IRDC to enhance classification accuracy. The parameterization of ELM presents another difficulty in document classification, which may impact classification performance.

Metaheuristic methods such as GA and CS are commonly used to modify the parameters of ELM. Even so, CS and GA are algorithms that search the world and are used extensively in various applications. Slow convergence is a potential limitation of the CS technique, particularly for difficult optimization problems. They might find it difficult to balance exploration and exploitation when looking for a novel solution, leading to decreased accuracy. Furthermore, large populations and high-dimensional search spaces are problems for GA. GA also causes delays in processing and is computationally expensive. Whereas GA and CS struggle to find the ideal value for parameter settings, RSS has proven to be more successful.

To adjust to dynamic changes and improve performance, optimization methods are essential. An optimization method needs to balance intensification with diversity to be considered robust. While intensification focuses on finding better solutions in a more limited local search region, diversification covers a wider search arena. The ELM technique is designed for SLFNs that exhibit faster convergence than traditional methods for promising performance. In addition, ELM operates without the need for gradient-based back propagation. Moreover, the ELM mechanism has the number of neurons in the input layer, hidden layer, and output layer are n, L, and m, respectively.

Furthermore, RSS offers a more effective parameter optimization method by continuously alternating between the normal and urgent search phases until the best answer is discovered. Additionally, RSS uses fewer parameters than GA and CS and performs better for global optima with a balance of exploration and exploitation. The hybrid RSS-ELM technique, which combines RSS and ELM parameters for text data categorization, is also explored in the proposed work.

III. Proposed Methodology

For the proposed techniques, text datasets Reuters-21578, 20newsgroups, and TDT-2 are taken from the UCI repository[24]. The Reuter21578 dataset with 10,788 documents is divided into two subsets: a training set and a testing set. There are 135 classes in the corpus that correspond to various categories. All balanced and imbalanced datasets (Reuter21578, TDT2, and 20newsgroup) are preprocessed to remove unnecessary content and to improve accuracy. Various preprocessing steps, such as tokenization, lower casing, stopword removal, and normalization, are performed to make datasets more concise and to increase accuracy. Moreover, to extract punctuation, spaces, and other non-alphanumeric characters from the text, parsing is utilized. The study framework and stages are depicted in Fig. 1.

The Proposed MRDC method is implemented, where keywords are used wisely to classify documents into appropriate classes. In addition, the proposed MRDC is compared with existing RDC and IRDC techniques. The results clarified that the proposed MRDC technique showed better results than the existing techniques. Moreover, another proposed RSS-ELM technique in which ELM parameters are optimized with the RSS technique. The suggested RSS-ELM's performance is compared with other methods, including CS-ELM and GA-ELM. The suggested RSS-ELM demonstrated more significant findings than the current methods because it had two search states and fewer RSS parameters. The success of the suggested technique is evaluated using four measurement criteria: accuracy, precision, recall, and F1-measure.

A. *The Proposed Modified Relative Discrimination Criterion Technique*

The MRDC selects important features from text documents in a dataset. The MRDC technique uses modified document frequency to count the number of documents that contain the term "t" and whose term count is tc. For positive and negative classes, the normalized document frequency is represented by the True Positive Rate (TPR) and False Positive Rate (FPR), respectively. TPR and FPR are calculated by the MRDC for each term count for RDC and IRDC. Furthermore, MRDC, in contrast to RDC, considers the frequency of word counts in a single class, rather than merely the quantity of documents in a dataset. Just like RDC assigns value only to frequently occurring terms and IRDC is more focused on rarely occurring terms in each class. Neither existing RDC nor IRDC techniques could create tradeoffs frequently, nor do they rarely occur with a term count. We improvised the legacy of existing techniques. To create a tradeoff, a log transformation is used for both frequent and rare terms, which reduces the dominant high terms with the balance of small terms. Moreover, it improves stability and interoperability.

The Proposed MRDC assigns a value to significant characteristics based on tprtc in the positive class and the fprtc in the negative class for each word count. In the proposed MRDC, a tradeoff is created for both RDC and IRDC techniques, shown in equations 1 and 2. As a result, the MRDC technique did not disregard a term's worth regardless of how often or infrequently it appears shown in Eq. (3). Additionally, both short and long documents are handled easily using MRDC. The steps in Algorithm 1 illustrate a holistic diagram of the MRDC approach as shown in Fig. 1.

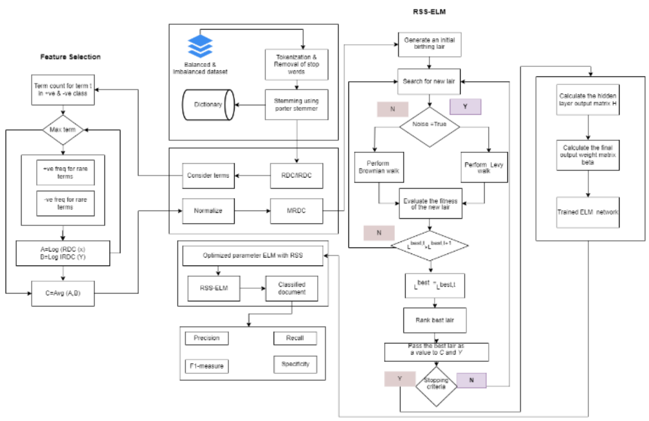


Fig. 1. A holistic diagram of MRDC and RSS-ELM techniques for text classification.

$$Ad\_RDC = log\left(\frac{|tpr_{tc} - fpr_{tc}|}{(min(tpr_{tc} - f\rho r_{tc})) * tc}\right) \quad (1)$$

$$Ad\_IRDC = log\left(\frac{|TPR_{tc} - FPR_{tc}|}{min\ (TPR_{tc}, FPR_{tc})}\right) * tc \quad (2)$$

$$MRDC = (Ad\_RDC + Ad\_IRDC)/2 \quad (3)$$

There is a need for a normalized term for *tc* where *tc* is high and the difference is low for increased term count. Fig. 2 shows modified relative discrimination criterion technique.

- Input: Text dataset
- Output: 1500 top-selected features
- Pos_frequency = Total frequency of term t in positive class
- Neg_frequency = Total frequency of term t in negative class
- Tc_{max} = maximum term count for term t
- for Tc = 1 to Tc_{max} do
- Tp = term t appears in positive documents
- Fp = term t appears in negative documents
- $t_{prtc} = \frac{Tp}{documents\ in\ POS}$
- $f_{prtc} = \frac{Fp}{documents\ in\ Neg}$
- $Ad_{RDC} = \log(\frac{|TPR_{tc}FPT_{tc}}{\min(TPR_{tc}\ ,\ FPR_{tc})*Tc})$
- $tp_{tc} = \frac{T_p}{Pos_{frequency}}$
- $fp_{tc} = \frac{F_p}{Neg_{frequency}}$
- $TPR_{tc} = \frac{tp_{tc}}{\sum_{i=0}^{n} tp_{tc}}$
- $FPR_{tc} = \frac{fp_{tc}}{\sum_{i=0}^{n} fp_{tc}}$
- Ad_IRDC= $\log(\frac{|TPR_{tc}FPT_{tc}}{\min(TPR_{tc}\ ,\ FPR_{tc})})$*Tc
- MRDC= (Ad_RDC , Ad_IRDC)/2
- End for
- $AUC_t = 0$
- For Tc=1 to $Tc_{max} do$
- $AUC_t = AUC_t + \frac{ERDC_{Tc} + ERDC_{Tc+1}}{2}$
- end

Fig. 2. Modified relative discrimination criterion technique.

B. The Proposed RSS-ELM Optimization Technique

The ELM technique randomly generates input weights and thresholds, only by setting the number of hidden nodes and acquiring unique ideal solutions. Compared with traditional neural network algorithms, ELM has the advantages of fast learning and good performance for a single-hidden-layer neural network. Furthermore, ELM determines output weights based on randomly generated input weights and biases before training. It also configures the activation function type, number of neurons in the hidden layer, and ultimately determines the optimal solution. The setting of the activation function as g(x), the network model of ELM is expressed. where i=[i1,i2,..., in] is the input weight, bi is the bias of the ith hidden neuron, xj=[x1j,x2j,..., xnj]T, i=[i1,i2,..., im], the output weight, uj=[u1j,u2j,..., umj] T is the network output [16].

The training goal of ELM is to minimize training errors. When the activation function is infinitely differentiable, and input weights and biases can be randomly selected, ELM training is equivalent to obtaining the output weights by solving the least squares solution in Eq. (4). The solution of an equation is Eq. (5) where H+ is the Moore-Penrose generalised inverse of H.

One kind of feed-forward neural network with a single hidden layer is called an Extreme Learning Machine (ELM) [25]. The single hidden layer of ELM presumes that the output function. For both classification and regression issues, an ELM is defined as a least squares-based single hidden layer feed-forward neural network (SLFN)[26]. The following is an ELM representation with training data N, hidden neurons H, and activation function f(x).

$$e_j = \sum_{i=1}^{H} \propto_i f(W_i, C_i, X_i) \quad J = 1,2,3, \ldots . . N \quad (4)$$

where $w_i \propto_i$ are the weight vectors that connect the input layer with the output layer, respectively. The input variables are shown by x_i. For the data points, j, the output from ELM is represented by e_j, and C_i is the hidden bias of the i^{th} hidden neuron. Eq. (2) is used for calculating the output weights and is as follows

$$\beta = A\dagger Y \quad (5)$$

A† is the Moore-Penrose generalized inverse of A where Y shows the targeted value of ELM.

$$A = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} f(W_1, C_1 X_1) & \cdots & f(W_H, C_H, X_i) \\ \vdots & \vdots & \vdots \\ f(W_1, C_1, X_j & \cdots & f(W_H, C_H, X_j) \end{bmatrix}, \alpha = \begin{bmatrix} \propto_1^T \\ \vdots \\ \propto_H^T \end{bmatrix},$$

$$\text{and } Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix} \quad (6)$$

ELM is a kind of regularization neural network, and its algorithm output is mainly based on matrix A. To optimize ELM parameters, optimization algorithms like GA, CS, and

RSS can successfully handle the broad search area and fine-tuned steps. The parameters of the ELM classification technique are optimized with the Ringed Seal Search (RSS) technique to improve results. The RSS method is inspired by how seal pups search for the best hiding spot from predators. The proposed RSS algorithm offers a sensitive search strategy that takes seal movement into account. These lairs safeguard against predators by offering thermal insulation against cold air and severe wind chills. A seal may have several lairs in one location.

A series of actions occurs while a seal pup explores a lair with multiple chambers or looks for new lairs. The process of evolution involves changing a random value. The ideal combination of parameters for each iteration is determined by evolving the selected parameters into a vector form using a matrix representing a starting population of ELM parameters.

Motivated by nature with default settings, RSS always starts to solve an optimization problem. In all optimization methods, the initial solution is represented by a vector of values, L-I, where i = 1, 2, 3,...n, in Eq. (7). Consisting of several chambers, the RSS algorithm always starts with an initial number of birthing lairs n. To find a new refuge of higher quality, the pups go into the search area. Finding a better search space necessitates creating an array of these initial values in the search space.

$$L = [i * m] \tag{7}$$

$$L_i, i = 1,2,3 \ldots \ldots n$$

There are multiple chambers in each lair, arranged haphazardly. For instance, the array of [i×m] for lair i represents the current lair I of the habitat. These values are uniformly and randomly distributed between the lower bound Lbj and the upper bound Ubj at the search space, as Eq. (8) illustrates.

$$L_i = Lb + (Ub - Lb)rand(size(Lb)) \tag{8}$$

Where I= 1,2,3,…n

The number of initialized lairs is n, whereas I indicates the number of lairs. The seal follows a specific search pattern as it moves from one lair to another, leading to fresh discoveries. (New layers) x^{t+1} for a seal i, a new layer is found in Eq. (9).

$$X_i^{(t+1)} = x_i^t + \propto \odot \, \Delta x \tag{9}$$

Where α is linked to the search pattern and denotes the step size in the normal or urgent state.

$$\Delta x = \lambda_{levy} \quad where\ w = 1 \tag{10}$$

Conversely, ω stands for a uniform discrete distribution. Eq. (10) calculates the step size of the Levy walk random walk, which is characterized by a probability distribution with an inverse power-law tail.

$$\text{Levy} \sim \text{u} = t^{-\lambda} \tag{11}$$

In contrast, t is the flight length and 1 < λ < 3. A random direction is chosen using the uniform distribution approach, and the step size is determined using the Levy distribution. If λ is greater than or equal to 3, the distribution does not have a heavy tail, and the total lengths converge to a Gaussian distribution [27, 28, 29]. An anomalous diffusion occurs when the mean squared displacement of the Levy walk increases more quickly than linearly with time. A Brownian walk, on the other hand, is characterized by a normal diffusion with a linear increase in mean squared displacement shown in Eq. (12)

The Levy walk is one way that animals find supplies that are dispersed throughout several different locations. Animals commonly employ two techniques: intensive (exploitation) and extensive (exploration). When an animal is investigating, it switches between extensive and intensive modes, concentrating on the search within the patch while moving from patch to patch.

$$\Delta x = \lambda_{browni} \quad where\ w = 0 \tag{12}$$

Eq. (12) illustrates the Brownian walk search for a new chamber inside a multi-chambered lair construction.

$$\text{S} = \text{K} * \text{rand}(\text{d}, N_{dot}) \tag{13}$$

K is the standard deviation of the normal distribution for diffusion rate, N is the number of Brownian particles in the search space, and d indicates the problem's dimensions presented in Eq. (13). The proposed RSS-ELM method has been put into practice using Python, and the outcomes of both the suggested and current methods are assessed using evaluation criteria (precision, accuracy, recall, and F-measure).

When compared to the GA-ELM and CS-ELM optimization approaches currently in use, the proposed RSS-ELM strategies produced notable results. Reuters21578, 20Newsgroup, and TDT2 are the three benchmark text datasets used in the research. These are typical text datasets for the experimental settings downloaded from the UCI repository. Fig. 1 illustrates RSS-ELM approaches used to examine these datasets. In addition, RSS-ELM's algorithm is presented in Fig. 3.

- Begin
- Initialized ELM parameter and structure
- Generate an initial number of birthing lairs,
  L1 = (f = 1, 2, 3, . . . , n)
- while stopping criterion is not met do
  if noise = false then
  Search in the proximity for a new lair by using a Brownian walk;
- Else
- Expand the search for a new lair by using a Levy walk;
- end if
- Evaluate the fitness of each new lair and compare it with the previous;
- If $L_{best,t} > L_{best,t+1}$ $then$
- Choose the new lair
- $L_{best} = L_{best,t}$
- Go to step 4
- End if
- Rank the lairs;
- End while
- Return the best lair;
- The global best lair is fed to ELM classifier for training
- Training the ELM classifier
  End=0

Fig. 3. Algorithm of proposed ring seal search-extreme learning machine technique.

IV. File Evaluation Measuring Criteria

In text classification, datasets are skewed in size. Three text datasets (Reuters, 20newsgroups, TDT2) are evaluated with various measuring criteria. Accuracy is not the only criterion for measuring the performance of the algorithm. Precision, recall, and F-measure are used to evaluate the above-mentioned text datasets. Moreover, F-measure is the harmonic mean of precision and recall [28] shown in Eq. (14), (15), (16), and (17) respectively.

$$Precision = \frac{tp}{tp+fb} \tag{14}$$

tp denotes the true positive rate and *fp* shows the false positive rate in precision.

$$Accuracy = \frac{tp+tn}{tp} \tag{15}$$

tp denotes the true positive rate and *tn* shows the true negative rate in accuracy.

$$Recall = \frac{tp}{tp+tn} \tag{16}$$

tp describes the true positive rate and *fn* denotes the false negative rate in recall.

$$F1 = \frac{2\ .\ Precision\ .\ Recall}{Precision+Recall} \tag{17}$$

V. Results of Proposed MRDC Feature Selection Technique

The experiment of the proposed MRDC and existing IRDC, RDC feature selection technique has been implemented in Python for three different text datasets (Reuters21578, 20newsgroup, TDT2). The proposed MRDC and the existing techniques' results are measured through various performance metrics (e.g., Accuracy, Precision, Recall, F-measure) shown in Table I.

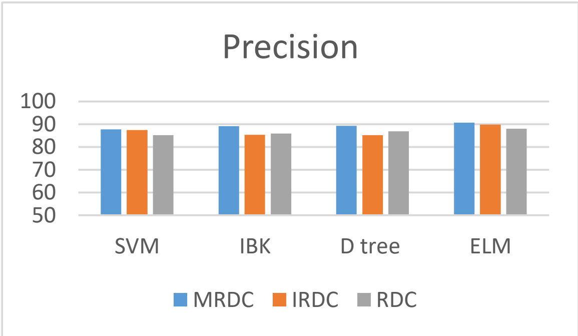
A. *Results of the Reuter21578 Dataset for Feature Selection Techniques*

In this section, the results of reuter21578 text dataset are presented. In addition, the results of the proposed MRDC and existing RDC, IRDC feature selection techniques are evaluated with various classifiers (SVM, IBK, Decision Tree, ELM), as elaborated in Table I and Fig. 4.

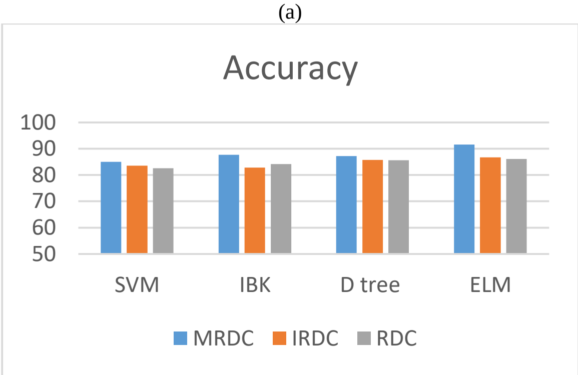
The Reuters dataset is assessed through the precision of MRDC, RDC, and IRDC, which is shown in Fig. 4(a). The precision values (87.80%, 89.24%, 89.26%, and 90.66%) for MRDC are higher than those for IRDC (87.520%, 85.42%, 85.23%, and 89.82%) and RDC (85.26%, 85.98%, 86.99%, and 88.02%) techniques. Accuracy for reuter21578 dataset is shown in Fig. 4 (b) which elaborate that proposed MRDC technique (85.00%, 87.64%, 87.20%, 91.60%) performs better than IRDC (83.60%, 82.80%, 85.70%, 86.70%) and RDC (82.60%, 84.10%, 85.66%, 86.10%) techniques. The resilience of various classifiers, such as SVM [30], IBK, ELM, and Decision Tree, for the proposed MRDC feature selection technique showcases a significant accuracy, higher than the percentages of IRDC and RDC techniques.

TABLE I. Results of Proposed MRDC Feature Selection Technique for Reuters21578 Dataset

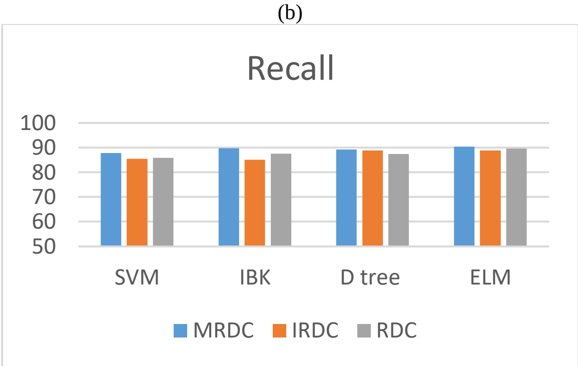
Classifier		SVM	IBK	D tree	ELM
MRDC (%)	Pre	87.8	89.24	89.26	90.66
	Acc	85.0	87.64	87.2	91.6
	Rec	87.8	90.13	89.26	90.05
	F M	87.8	89.68	89.26	90.36
IRDC (%)	Pre	87.52	85.42	85.23	89.82
	Acc	83.6	82.8	85.7	86.7
	Rec	85.53	85.13	88.87	88.82
	F M	86.51	85.27	87.01	89.32
RDC (%)	Pre	85.26	85.98	86.99	88.02
	Acc	82.6	84.1	85.66	86.1
	Rec	85.83	87.57	87.39	89.59
	F M	85.55	86.77	87.19	88.8



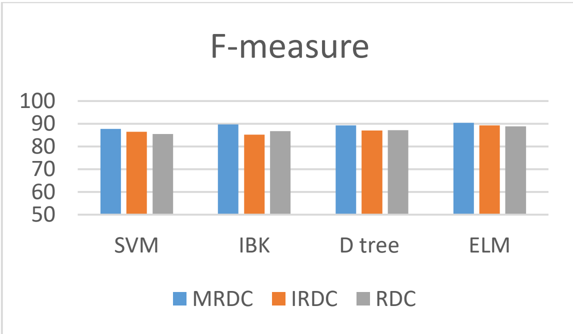
(a)



(b)



(c)



(d)

Fig. 4. Results of proposed MRDC feature selection technique for the Reuters21578 dataset.

Recall of proposed MRDC technique (87.80%, 89.68%, 89.26%, and 90.36%) for SVM, IBK, Decision Tree, and ELM which is higher than those of IRDC (85.53%, 85.13%, 88.87%, and 88.82%) and RDC (85.83%, 87.57%, 87.39%, and 89.59%) respectively. It is evaluated that the recall of the proposed MRDC feature selection technique is accurate and significant compared to existing techniques.

Regarding F-Measure scores for SVM, IBK, Decision Tree, and ELM, the proposed MRDC routinely outperforms IRDC and RDC. MRDC performs better than RDC (85.55% to 88.80%) and IRDC (86.51% to 89.32%). This demonstrates MRDC's dominance in achieving a balanced trade-off between recall and precision.

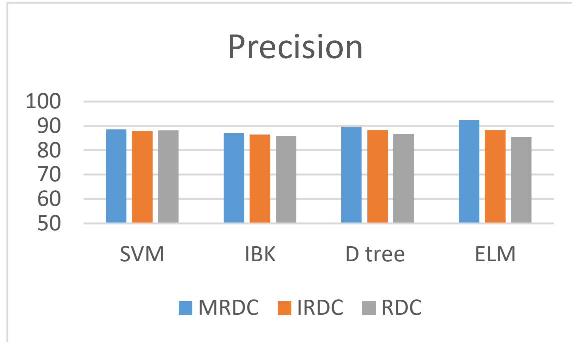
B. Results of 20 Newsgroup Datasets for Feature Selection Techniques

The result of the 20newsgroup text dataset for the proposed MRDC and existing RDC, IRDC feature selection technique with various classifiers (SVM, IBK, Decision Tree, ELM) is demonstrated in this section, shown in Table II.

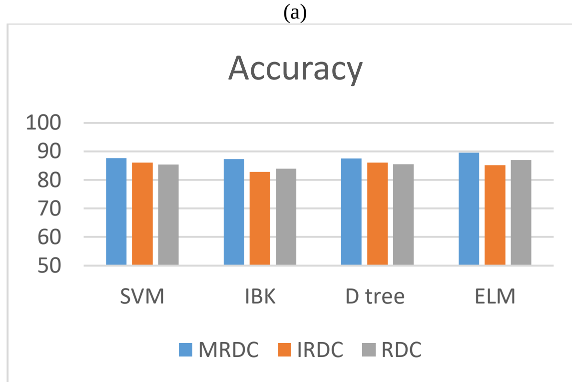
TABLE II. RESULTS OF PROPOSED MRDC FEATURE SELECTION TECHNIQUE FOR 20NEWSGROUP DATASET

Classifier		**SVM**	**IBK**	**D tree**	**ELM**
MRDC (%)	Pre	88.57	86.92	89.66	92.32
	Acc	87.65	87.28	87.5	89.5
	Rec	89.45	86.78	88.89	90.02
	F M	89.01	86.85	89.27	91.15
IRDC (%)	Pre	87.93	86.41	88.33	88.33
	Acc	86	82.77	86	85.11
	Rec	87.93	83.61	88.33	87.04
	F M	87.93	84.99	88.33	87.68
RDC (%)	Pre	88.21	85.83	86.71	85.47
	Acc	85.4	83.89	85.53	87
	Rec	86.71	87.31	88.06	89.37
	F M	87.46	86.56	87.38	87.38

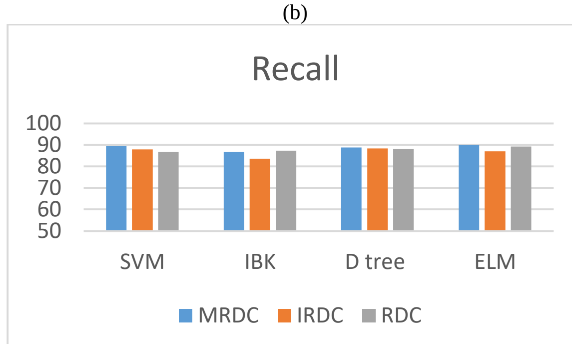
The analysis of the suggested MRDC's performance for 20 newsgroup datasets uses F-measure, precision, accuracy, and recall. Experimenting with MRDC methodologies, which yield superior outcomes than the current feature ranking methods, RDC, and IRDC.



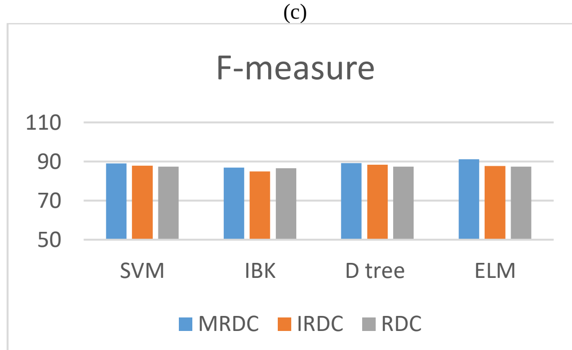
(a)



(b)



(c)



(d)

Fig. 5. Results of the proposed MRDC feature selection technique for 20 newsgroup dataset.

When the 20newsgroup dataset was tested using the various classifiers displayed in Table II, the proposed MRDC outperformed the current IRDC and RDC methods. In Fig. 5 (a), MRDC outperformed IRDC (87.93%, 86.41%, 88.33%, and 88.33%) and RDC (88.21%, 85.83%, 86.71%, and 85.47%) in terms of precision (85.57%, 86.92%, 89.66%, and 92.32%).

The accuracy results for the machine learning models (SVM, IBK, Decision Tree, and ELM) in Fig. 5(b) shows that the MRDC performs better than the alternative methods. The accuracy values for RDC are 85.40%, 83.89%, 85.53%, and 89.00%, MRDC is 87.65%, 87.28%, 87.50%, and 89.50%, and IRDC is 86.00%, 82.77%, 86.00%, and 85.11%. MRDC is shown to have a high score and to produce better outcomes than earlier methods that were tested on SVM, IBK, Decision Tree, and ELM. Additionally, the results indicate that MRDC is compatible with the ELM classifier displayed in Fig. 5 and Table II.

These results suggest that the proposed MRDC technique outperforms the other two feature selection techniques in terms of prediction accuracy and dependability. Fig. 5(c) illustrates that the recall for MRDC is 89.45%, 86.78%, 88.89%, and 90.02%, which is superior to that of IRDC, which ranges from 87.93% to 87.04%, and RDC, which ranges from 86.71% to 89.37%. Furthermore, as shown in Fig. 5(d), the MRDC feature selection technique's F-measure (89.01%, 86.85%, 89.27%, and 91.15%) is significant compared to the IRDC, which ranges from 87.93% to 87.04%) and the RDC, which ranges from 87.46% to 87.38%).

C. *Results of the TDT2 Dataset for Feature Selection Techniques*

Another text dataset, TDT2, is used for feature selection techniques, MRDC, IRDC, and RDC. Results of the above-mentioned techniques are compared with each other and verified by well-known classifiers, SVM, IBK, Decision tree, and ELM, which are shown in Table III and Fig. 6.

TABLE III. RESULTS OF PROPOSED MRDC FEATURE SELECTION TECHNIQUE FOR TDT-2 DATASET

Classifier		SVM	IBK	D tree	ELM
MRDC (%)	Pre	86.81	89.5	88.14	89.93
	Acc	85.9	86.65	85.18	87.25
	Rec	88.79	87.76	89.2	90
	F M	87.79	88.62	88.67	89.97
IRDC (%)	Pre	86.4	87.26	85.08	86.6
	Acc	84.12	85.91	84.8	84.01
	Rec	86.81	89.02	87.57	86.9
	F M	86.6	88.13	86.31	86.31
RDC (%)	Pre	84.6	86.53	87.93	88.83
	Acc	83	85.02	85.1	87.16
	Rec	86.36	86.31	87.5	89.08
	F M	85.47	86.42	87.72	88.95

MRDC outperforms IRDC and RDC approaches in terms of precision values when measuring the outcome. The precision scores of the classifiers SVM, IBK, Decision Tree, and ELM in MRDC are 86.81%, 89.50%, 88.14%, and 89.93%, respectively. These values are higher than those of IRDC (86.40% to 86.60%) and RDC (84.60% to 88.83%), as illustrated in Fig. 6(a). The accuracy of MRDC is consistently higher than that of the current IRDC and RDC feature selection methods in Fig. 6(b) across SVM, IBK, Decision Tree, and ELM models. For SVM, IBK, Decision Tree, and ELM classifiers, the MRDC technique performs better, with values ranging from 85.90%, 86.65%, 85.18%, and 87.25 percent, respectively, whereas IRDC and RDC had values ranging from 84.12% to 84.01% and 87.00% to 87.16%, respectively.

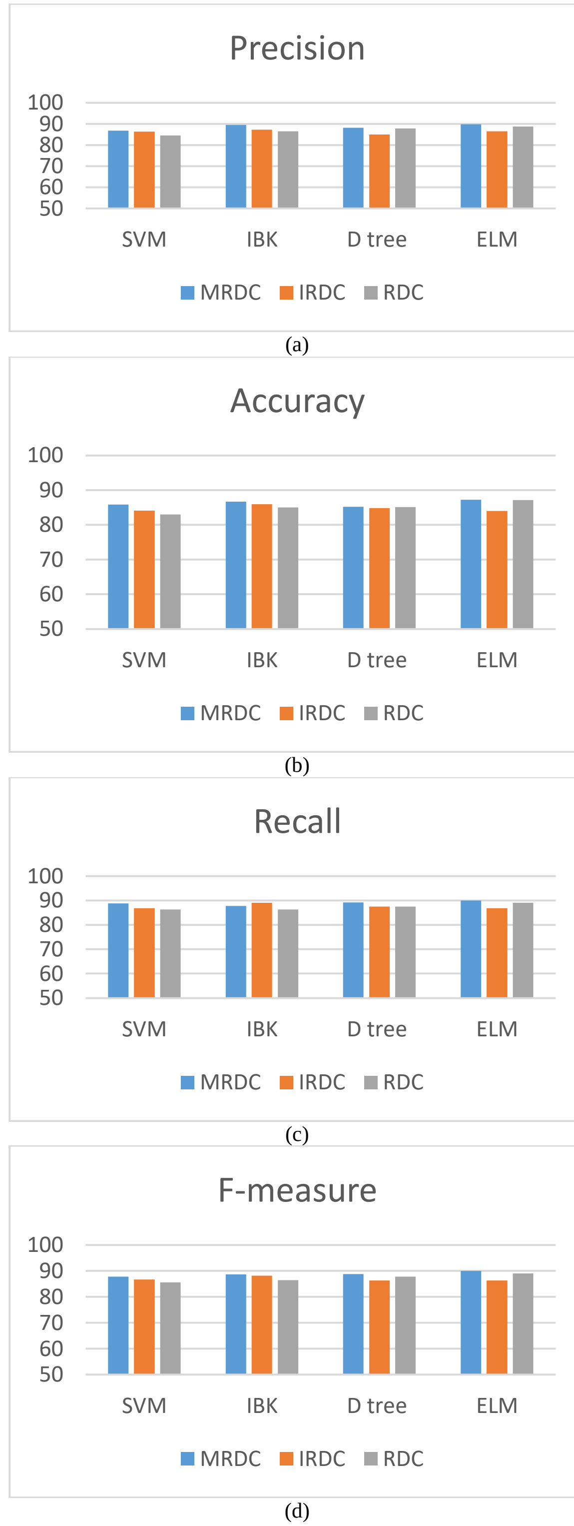


Fig. 6. Results of proposed MRDC feature selection technique for TDT-2 dataset.

Regarding recall metrics, SVM, IBK, Decision Tree, and ELM models similarly outperform MRDC. As illustrated in Fig. 6(c), the MRDC values for the SVM, IBK, Decision Tree, and ELM classifier models range from 87.79% to 87.76%, 89.20% to 90.00%, respectively, and are superior to the IRDC and RDC, which range from 86.81% to 86.90% and RDC (85.47% to 88.95%), respectively.

Fig. 6(d) shows that MRDC performed better than the classifiers mentioned earlier. The suggested MRDC feature selection method yielded significant results (87.79%, 88.62%, 88.67%, and 89.97%), outperforming SVM, IBK, and decision trees. The F measure is also computed for MRDC, IRDC, and RDC feature selection strategies.

VI. The Proposed RSS-ELM Optimization Technique

The experiment is conducted on three different optimization techniques, which are RSS-ELM, GA-ELM, and CS-ELM. The proposed RSS-ELM optimization technique is compared with GA-ELM and CS-ELM techniques with three text datasets such as reuter21578, 20newsgroup and TDT2. To evaluate these results, various evaluation metrics are used (Accuracy, Precision, Recall, F-measure).

A. Results of the Reuter21578 Dataset for Proposed RSS-ELM Techniques

In this section, the result of the reuter21578 text dataset for the proposed RSS-ELM and existing GA-ELM and CS-ELM optimization techniques is shown in Table IV and Fig. 7.

TABLE IV. Results on Proposed RSS-ELM Optimization Technique for Reuters21578-Dataset

Algorithm	Precision	Accuracy	Recall	F Measure
RSS-ELM	99.1	98.9	98.7	98.9
CS-ELM	67	66	66	66
GA-ELM	58	60	59	58

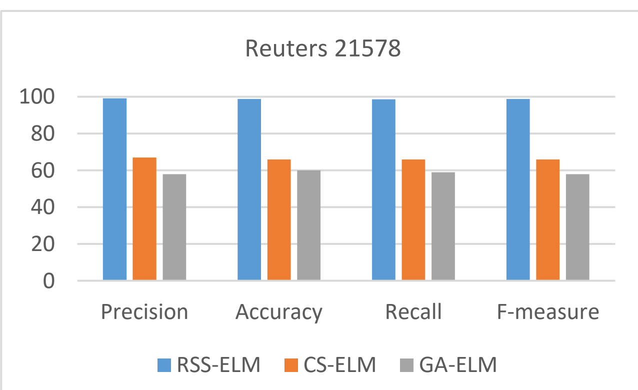


Fig. 7. Results on proposed RSS-ELM optimization technique for Reuters 21578-dataset.

Performance of RSS-ELM, CS-ELM, and GA-ELM optimization techniques is evaluated for the Reuters21578 text dataset. The results revealed that RSS-ELM technique performed better than other existing techniques, achieving the highest precision (99.1%) for RSS-ELM than CS-ELM (67%) and GA-ELM (58%). Furthermore, CS-ELM and GA-ELM improved by 66% and 60%, respectively, while the RSS-ELM technique reached a noteworthy accuracy of 98.9%. This demonstrates its excellent capacity to recognize pertinent facts and generate precise forecasts. Additionally, the RSS-ELM technique demonstrated the highest recall (98.9%), demonstrating its ability to capture relevant data, whereas CS-ELM and GA-ELM received 66% and 59%, respectively.

Furthermore, it's F-measure of 98.9%, in conjunction with CS-ELM (66%) and GA-ELM (58%), demonstrates a noteworthy performance. On the other hand, CS-ELM and GA-ELM produced lower values for every metric, indicating that they performed worse in this task. These results show that RSS-ELM is the best option for this dataset, highlighting its potential use in practical applications where data categorization accuracy and precision are essential.

B. Results of 20newsgroup Dataset for Proposed RSS-ELM Techniques

Experiments of proposed RSS-ELM and existing GA-ELM, CS-ELM optimization techniques conducted in another 20newsgroup text dataset. These experiments are also conducted in the Python language. Four evaluation metrics (Precision, Accuracy, Recall, and F-measure) are used to verify these significant results. Detailed results are presented in Table V and Fig. 6.

TABLE V. Results on Proposed RSS-ELM Optimization Technique for 20newsgroup Dataset

Algorithm	Precision	Accuracy	Recall	F-Measure
RSS-ELM	96	97.2	97	97
CS-ELM	78	77	77	76
GA-ELM	79	78	78	79

The proposed RSS-ELM outperforms other optimization methods like CS-ELM (78%) and GA-ELM (79.0%) in terms of precision (96%). Additionally, the suggested RSS-ELM obtained 97.2% accuracy, but CS-ELM and GA-ELM improved by 77% and 78%, respectively. While the CS-ELM generated Recall (77%) and F-measure (76%), and the GA-ELM produced Recall 78% and F-Measure 79%, the RSS-ELM underperforms in both Recall (97%) and F-Measure (97%), as presented in Fig. 8 and Table V. Compared to other optimization methods, the suggested RSS-ELM performance is superior.

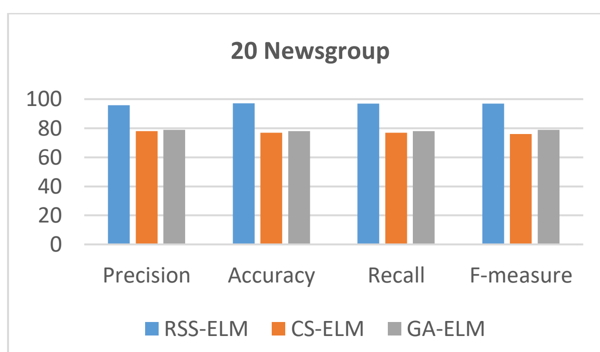


Fig. 8. Results of proposed RSS-ELM optimization technique for 20 newsgroup dataset.

C. Results of the TDT2 Dataset for Proposed RSS-ELM Techniques

Another text dataset is utilized for experiments of the proposed RSS-ELM and existing GA-ELM, CS-ELM techniques shown in Table VI and Fig. 9.

TABLE VI. RESULTS ON PROPOSED RSS-ELM OPTIMIZATION TECHNIQUE FOR TDT2 DATASET

Algorithm	Precision	Accuracy	Recall	F-measure
RSS-ELM	97	97.5	97	97
CS-ELM	62	64	65	62
GA-ELM	59	58	58	58

Three optimization techniques—RSS-ELM, CS-ELM, and GA-ELM—were evaluated for performance in our study using the TDT-2 dataset. According to the results, RSS-ELM performed better than CS-ELM and GA-ELM, producing 62% and 59% of the total, respectively, with the highest precision (97%). The suggested RSS-ELM optimization method outperformed the current CS-ELM and GA-ELM (64%, 58%), with a noteworthy accuracy of 97.5%, demonstrating its capacity to identify pertinent data and generate accurate forecasts precisely.

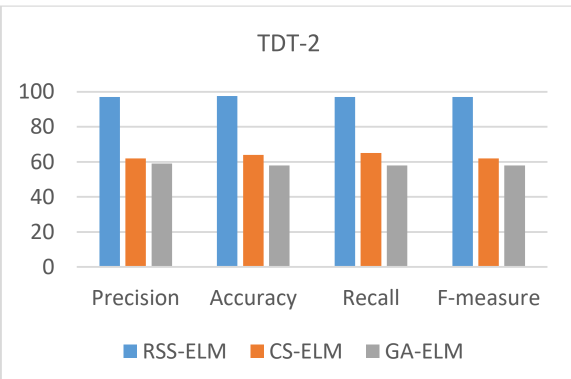


Fig. 9. Results on proposed RSS-ELM optimization technique for TDT2 dataset.

In comparison to CS-ELM and GA-ELM (65%, 58%), it also demonstrated a respectable recall of RSS-ELM (97%), demonstrating its efficacy in capturing a significant amount of relevant information. Additionally, the 97% F-measure demonstrates that RSS-ELM achieved an impressive balance between recall and precision, making it a reliable option for data categorization tasks. However, CS-ELM and GA-ELM displayed lower values for all measures (62%, 58%), indicating that they performed less well on this specific dataset. These results highlight the potential usefulness of RSS-ELM in applications that need to classify data using the TDT-2 dataset with both accuracy and precision.

VII. CONCLUSION

The limitations of existing feature ranking and classification techniques for high-dimensional text data are highlighted in this work. This work proposes MRDC, a dependable feature selection strategy for balanced and unbalanced text datasets. Common and uncommon terms should be considered when choosing features to normalize terms for improved categorization. The proposed MRDC method outperforms RDC and IRDC methods in classification by effectively selecting the best characteristics. To make feature ranking research more dependable and simpler, phrase count might be used to modify future studies. Furthermore, RSS-ELM for optimization is an additional contribution. Compared to GA-ELM and CS-ELM, the suggested RSS-ELM technique is important for parameter optimization in two-way state finding.

Additionally, RSS optimizes ELM with fewer parameters. Furthermore, RSS-ELM exhibits superior performance in terms of F-measure, recall, accuracy, and precision. The RSS approach can optimize several kinds of alternative and hybrid walks in optimization. Additionally, RSS can be used to assess alternative classification methods for datasets of text and images. Our proposed technique can have significant real-life applications in diverse contexts, including collaborative enterprises [31].

REFERENCES

[1] Dokeroglu, T., A. Deniz, and H.E. Kiziloz, A comprehensive survey on recent metaheuristics for feature selection. Neurocomputing, 2022. 494: p. 269-296.

[2] Zhou, X., et al. A survey on text classification and its applications. in Web intelligence. 2020. IOS Press.

[3] Sharif, W., Improved relative discriminative criterion using rare and informative terms and ringed seal search-support vector machine techniques for text classification. phd thesis 2019, Universiti Tun Hussein Onn Malaysia.

[4] Deng, X., et al., Feature selection for text classification: A review. Multimedia Tools and Applications, 2019. 78(3): p. 3797-3816.

[5] Kotsiantis, S.B., I. Zaharakis, and P. Pintelas, Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 2007. 160(1): p. 3-24.

[6] Onan, A. and S. Korukoğlu, A feature selection model based on genetic rank aggregation for text sentiment classification. Journal of Information Science, 2017. 43(1): p. 25-38.

[7] Rathi, P. and N. Singh, An Efficient Algorithm for Informational Retrieval using Web Usage Mining. International Journal of Hybrid Information Technology, 2019. 12(2): p. 13-20.

[8] Biglari, M., F. Mirzaei, and H. Hassanpour, Feature selection for small sample sets with high dimensional data using heuristic hybrid approach. International Journal of Engineering, 2020. 33(2): p. 213-220.

[9] Bashir, S., et al., A novel feature selection method for classification of medical data using filters, wrappers, and embedded approaches. Complexity, 2022. 2022(1): p. 8190814.

[10] Sharif, W., et al., Improved relative discriminative criterion feature ranking technique for text classification. Int. J. Artif. Intell, 2017. 15(2): p. 61-78.

[11] Rehman, A., et al., Relative discrimination criterion–A novel feature ranking method for text data. Expert Systems with Applications, 2015. 42(7): p. 3670-3681.

[12] Muaad, A.Y., et al., An effective approach for Arabic document classification using machine learning. Global Transitions Proceedings, 2022. 3(1): p. 267-271.

[13] Paschen, J., J. Kietzmann, and T.C. Kietzmann, Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. Journal of business & industrial marketing, 2019. 34(7): p. 1410-1419.

[14] Tang, J., C. Deng, and G.-B. Huang, Extreme learning machine for multilayer perceptron. IEEE transactions on neural networks and learning systems, 2015. 27(4): p. 809-821.

[15] Zhu, Q.-Y., et al., Evolutionary extreme learning machine. Pattern recognition, 2005. 38(10): p. 1759-1763.

[16] Wang, J., et al., A review on extreme learning machine. Multimedia Tools and Applications, 2022. 81(29): p. 41611-41660.

[17] Eshtay, M., H. Faris, and N. Obeid, Improving extreme learning machine by competitive swarm optimization and its application for medical diagnosis problems. Expert Systems with Applications, 2018. 104: p. 134-152.

[18] Nassef, A.M., et al., Review of metaheuristic optimization algorithms for power systems problems. Sustainability, 2023. 15(12): p. 9434.

[19] Morales-Castañeda, B., et al., A better balance in metaheuristic algorithms: Does it exist? Swarm and Evolutionary Computation, 2020. 54: p. 100671.

[20] Anjum, N. and S. Badugu, A comparative study on classification algorithms using different feature extraction and vectorization techniques for text. Turkish Online Journal of Qualitative Inquiry, 2021. 12(7).

[21] Sikri, A., N. Singh, and S. Dalal, Chi-square method of feature selection: Impact of pre-processing of data. International Journal of Intelligent Systems and Applications in Engineering, 2023. 11(3s): p. 241-248.

[22] Xie, W., et al., Improved multi-layer binary firefly algorithm for optimizing feature selection and classification of microarray data. Biomedical Signal Processing and Control, 2023. 79: p. 104080.

[23] Popa, D.N., et al. Implicit discourse relation classification with syntax-aware contextualized word representations. in 32nd FLAIRS Conference 2019: Sarasota, Florida, USA. 2019.

[24] Kowsari, K., et al., Text classification algorithms: A survey. Information, 2019. 10(4): p. 150.

[25] Sanjeevikumar, P., et al., Machine learning-based hybrid demand-side controller for renewable energy management, in Sustainable Developments by Artificial Intelligence and Machine Learning for Renewable Energies. 2022, Elsevier. p. 291-307.

[26] Pal, M. and S. Deswal, Extreme learning machine-based modeling of resilient modulus of subgrade soils. Geotechnical and Geological Engineering, 2014. 32: p. 287-296.

[27] Saadi, Y., et al., Ringed seal search for global optimization via a sensitive search model. PloS one, 2016. 11(1): p. e0144371.

[28] Uysal, A.K. and S. Gunal, A novel probabilistic feature selection method for text classification. Knowledge-Based Systems, 2012. 36: p. 226-235.

[29] Sharif, W., Samsudin, N. A., Deris, M. M., & Khalid, S. K. A. (2018). A technical study on feature ranking techniques and classification algorithms. J. Eng. Appl. Sci, 13(9), 7074-7080.

[30] Sharif, W. A. R. E. E. S. A., Yanto, I. T. Y., Samsudin, N. A., Deris, M. M., Khan, A. B. D. U. L. L. A. H., Mushtaq, M. F., & Ashraf, M. U. H. A. M. M. A. D. (2019). An optimised support vector machine with ringed seal search algorithm for efficient text classification. Journal of Engineering Science and Technology, 14(3), 1601-1613.

[31] Wajid, U., Namoun, A., Marín, C. A., & Mehandjiev, N. (2013). Designing and evaluating a system of document recognition to support interoperability among collaborative enterprises. Computers in industry, 64(5), 598-608.