Ingénierie des Systèmes d'Information
Vol. 30, No. 9, September, 2025, pp. 2339-2346
Journal homepage: http://iieta.org/journals/isi

Enhancing Relief Feature Selection via Uncertainty Sampling in Active Learning

Chourouk Elokri*, Tayeb Ouaderhman, Hasna Chamlal

Computer Science and Systems Laboratory (LIS), Department of Mathematics and Informatics, Faculty of Sciences Ain Chock, Hassan II University of Casablanca, Casablanca 20470, Morocco

Corresponding Author Email: chourouk.elokri-etu@etu.univh2c.ma

https://doi.org/10.18280/isi.300911

Received: 30 March 2025
Revised: 14 July 2025
Accepted: 4 August 2025
Available online: 30 September 2025

Keywords:
active learning, uncertainty sampling, feature selection, relief algorithm, classification performance

ABSTRACT

Feature selection represents a critical pretreatment step that allows data size reduction, thereby decreasing the computational cost of applying classification models. However, a lack of examples can adversely impact this phase, as the chosen features may lack significance or relevance. Frequently, we encounter unlabeled and expensive-to-label data, rendering it unsuitable for model training or feature selection. Active learning strategies can handle this by selecting the most relevant examples from unannotated data to optimize performance, reduce computational time and cost. In this study, we evaluate Relief feature selection with linear and non-linear classifiers. The performance is assessed using Accuracy, F1-Score, and AUC metrics. The results demonstrate that optimal performance is achieved with 70–140 examples, and feature selection significantly improves after adding the optimal number of examples. To build upon this approach, we leverage three distinct datasets.

1. INTRODUCTION

In recent years, data mining has attracted considerable attention as the need to extract meaningful insights from vast and complex datasets has become increasingly important across various industries. The rapid expansion of digital information has created new opportunities for organizations to leverage data mining techniques, uncovering hidden patterns, trends, and relationships that inform decision-making and drive innovation.

One crucial preprocessing step in data mining is feature selection, which aims to reduce dataset dimensionality by identifying the most relevant features or variables [1]. This process is especially valuable when dealing with datasets with a large number of features, as including irrelevant or redundant variables can degrade the performance of machine learning models [2]. Feature selection methods generally fall into three categories: filter methods [3], embedded methods [4, 5], and hybrid methods [6].

When working with limited datasets, selecting the most significant features presents a challenge, as the scarcity of information can make it difficult to identify meaningful correlations or trends. This issue is particularly relevant in feature selection, where insufficient data may lead to the inclusion of redundant or unnecessary attributes, ultimately affecting model performance. Additionally, small sample sizes can increase a model's susceptibility to noise, reducing its ability to generalize effectively to new data [7, 8]. To address this, active learning has emerged as an effective strategy for enhancing both the quantity and quality of labeled data by selectively acquiring the most relevant examples from an unlabeled dataset [9]. This technique involves training a model on existing labeled data, using that model to identify the most informative unlabeled instances, and then obtaining labels for those instances to refine the training set. By focusing on the most valuable data points [10], active learning can improve model performance with fewer labeled samples compared to conventional passive learning approaches [11].

Active learning techniques have been widely applied across various fields, demonstrating their adaptability and effectiveness. In text classification [12], these methods have been used to enhance tasks such as document categorization, sentiment analysis, and named entity recognition. In image processing [13], active learning has been employed to improve object detection [14], credit scoring [15], facial recognition [16], and scene understanding, allowing for more efficient use of human expertise in labeling complex visual data. The medical field has also embraced active learning, particularly in improving diagnostic accuracy and accelerating drug discovery. For example, active learning has been applied to medical image analysis [17], enabling experts to prioritize the most challenging or ambiguous cases, leading to more precise and efficient diagnoses.

Existing studies have generally treated active learning as a set of strategies to achieve increasing performance or as a means to verify and demonstrate its role in making models more effective. Most of this research has focused on evaluating active learning as a tool to boost predictive accuracy or enhance model robustness, without fully considering its potential impact on preprocessing steps such as feature selection. In particular, Relief, one of the most widely used filter-based feature selection methods, remains underexplored in the context of active learning integration.

This study addresses this gap by exploring the effectiveness

of active learning techniques, specifically in the domain of feature selection for classification tasks. The primary objective is to investigate how active learning can strengthen the Relief method in identifying the most significant features within a dataset. By iteratively selecting and labeling the most informative examples, active learning not only enriches the quality of the training data but also guides Relief toward more relevant and stable feature subsets. This approach has the potential to reduce redundancy, mitigate the effect of limited data, and ultimately lead to improved classification outcomes. In this context, our method has been evaluated on several classifiers, and it can be extended in future work to other models, including decision trees and random forests [18].

2. METHOD

This section describes the active learning approach used to identify the most informative examples from the unlabeled dataset, along with the feature selection techniques applied to determine the most relevant features. These methods were implemented to assess the impact of integrating carefully chosen examples into the labeled dataset.

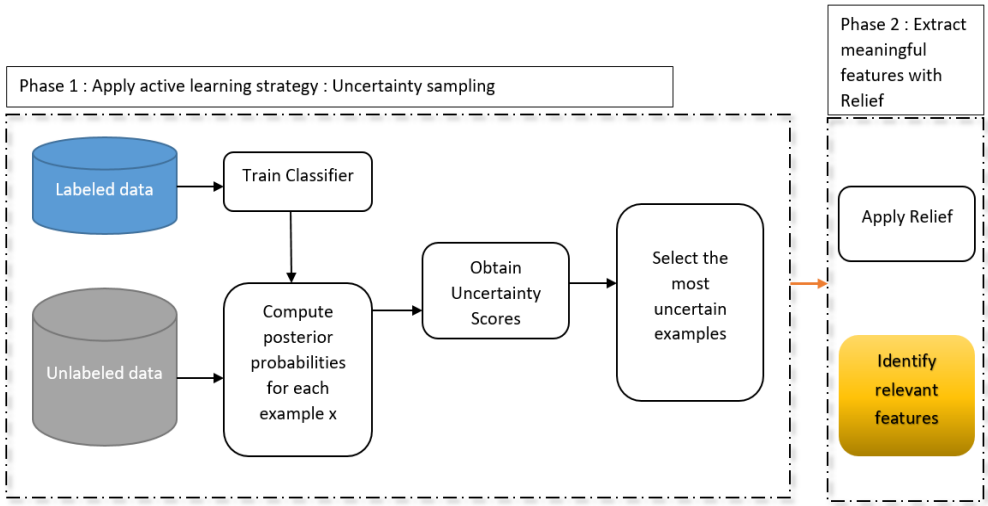

Figure 1. Overview of the proposed methodology

Figure 1 shows the proposed method, which operates in two phases. In the first phase, uncertainty sampling is applied by training a classifier, computing posterior probabilities for unlabeled instances, and selecting the most uncertain examples for labeling. In the second phase, the Relief algorithm is applied to the enriched labeled set to identify the most relevant features for classification.

2.1 Active learning method: Uncertainty sampling

Active learning techniques for query selection aim to identify the most informative data points to enhance model efficiency. The process begins with training the model on labeled data, after which it analyzes the available unlabeled examples to select the most relevant ones based on calculated information. Among the various active learning strategies, uncertainty sampling stands out as the most widely recognized technique in machine learning that aims to choose the most informative examples for labeling. This approach operates on the principle that a model learns most effectively from data points where its predictions are highly uncertain [19]. This approach includes three main variants:

- **Least confident approach** [20]**:** This method is especially useful for stochastic model in binary supervised learning. It involves selecting the example for which the model's posterior probability for a positive label is close to 0.5. In multi-class classification, this strategy translates to choosing the example with the highest posterior probability that is still near 0. This technique emphasizes information from the most likely class while ignoring the others, the uncertainty score is:

$$u_{LC}(x) = 1 - max_{c \in C} P(c|x)$$

where, $P(c \mid x)$ is the posterior probability of class c given example x. The closer the maximum probability is to 0.5 (binary) or low values (multi-class), the higher the uncertainty.

- **Margin sampling** [21]: This approach selects instance that give the small difference between the predicted probabilities of the two most likely categories. Nevertheless, in a multi-class context, this method may neglect the other classes. The margin is defined as:

$$u_{MS}(x) = P(c_1|x) - P(c_2|x)$$

where, $P(c_1 \mid x)$ and $P(c_2 \mid x)$ are the highest and second-highest class probabilities. A smaller margin indicates higher uncertainty, hence the instance is selected.

- **Entropy sampling** [22]**:** This approach considers all classes within the dataset. It utilizes the entropy function to assess the information content of class probability predictions, selecting examples that maximize this value. This approach helps ensure a thorough exploration of the dataset. The entropy-based uncertainty is defined as:

$$u_{ENT}(x) = -\Sigma_{c \in C} P(c|x)\ log\ P(c|x)$$

A higher entropy value indicates greater uncertainty in classification, making such examples more informative for labeling.

2.2 Feature selection technique: Relief

Feature selection is a crucial technique used to identify the most relevant attributes in a dataset to enhance the performance of machine learning models. This approach helps reduce data dimensionality, eliminate unnecessary or redundant features, and improve both model interpretability and computational efficiency. Generally, feature selection methods are categorized into three main types: filter methods, wrapper methods, and embedded methods [23].

- **Filter methods:** These methods are commonly used as a preliminary step to select relevant features before applying more advanced techniques. They assess the intrinsic properties of the data, sometimes independently of the specific problem, using statistical metrics [24].
- **Wrapper methods:** These approaches use a learning algorithm to evaluate different feature subsets and select those that maximize model performance [25].
- **Embedded methods:** Unlike wrapper approaches, embedded methods perform feature selection during the training process. They are known for their speed and efficiency [26].

Among feature selection algorithms, Relief stands out for its ability to identify the most influential attributes in classification or regression tasks [27]. Its core principle is to assign weights to features, reflecting their importance in distinguishing between different classes [28]. The algorithm for selecting relevant features is described as follows:

Algorithm 1: Relief Algorithm

Input: X: Training Set
Output: F_s: *Selected features*

Initialize the weight vector to $W = 0$
for $j = 1$ to p **do**
  for $i = 1$ to n **do**
    $wi,j = diff(X^j, x_i, M(x_i)) - diff(X^j, x_i, H(x_i))$
  end for
  $W_j = \sum w_{i,j}$
end for
Establish the vector W.
Select the features with the top values of W.
end function

The difference between two examples is calculated using the diff function based on the variable category X^j, where:

- If X^j is non-numeric, the diff function is defined as:

$$diff(X^j, x_i, x_k) = \begin{cases} 0 & if \quad x_i^j = x_k^j \\ 1 & if \quad x_i^j \neq x_k^j \end{cases}$$

- If X^j is numeric then:

$$diff(X^j, x_i, x_k) = \frac{|x_i^j - x_k^j|}{\nu}$$

where,

– ν normalizes the diff function values to the range [0,1];

– x_i^j represents the value of the j^{th} feature for the i^{th} example.

2.3 Proposed method

The proposed method combines uncertainty sampling with Relief-based feature selection to enhance the quality of selected features under limited labeled data. The process begins by initializing a small labeled subset and treating the remaining data as unlabeled. At each iteration, a classifier is trained on the labeled set, and uncertainty sampling is applied to identify the most informative examples among the unlabeled instances. These uncertain examples are queried and added to the labeled set, progressively enriching it with relevant data. Once the query budget is reached, the Relief algorithm is applied to the final labeled set to compute feature weights and generate a ranked list of features. This integration ensures that the most informative instances guide the selection process, leading to more stable and relevant feature subsets.

The following pseudocode outlines the proposed method:

Algorithm 2: Uncertainty Sampling with Relief Feature Selection

Input: Unlabeled dataset $\boldsymbol{U}$, small labeled set L, batch size b.
Output: Ranked list of features by importance

1. Initialize L with a small labeled subset (e.g., a few examples per class).
2. Set $U \leftarrow$ remaining unlabeled data.
3. Train a model M on L.
4. **For each** $x \in U$ **do**
5. Compute an uncertainty score u(x):
   - Least-Confident: $1 - \max_c P(c|x)$
   - Margin: $P(c_1|x) - P(c_2|x)$
   - Entropy $\sum_c P(c|x) log P(c|x)$

   End For
6. Select top b most uncertain samples $S \subset U$
7. Query the true labels for S.
8. Update the sets: $L \leftarrow L \cup S, \quad U \leftarrow U \setminus S$
9. Apply Relief on L to compute feature weights.
10. Rank features according to their weights.
11. Return: Ranked features.

3. EXPERIMENTATION

This section presents the preliminary methods used to evaluate the proposed approach, including the datasets, metrics, and classifiers employed.

To pinpoint areas for improvement in active learning, we use three distinct binary datasets from the UCI Machine Learning Repository [29]. Table 1 provides a detailed overview of the properties of each dataset.

Table 1. The data utilized in the experimental investigations

Dataset	Classes	Instances	Attributes
Breast Cancer	2	569	30
Ionosphere	2	351	34
Sonar	2	207	60

We use 5-fold cross-validation, where $\frac{4}{5}$ of the dataset is used for training and the remaining 20% is reserved for testing. This well-established technique in machine learning helps evaluate and select models [30]. The process involves dividing

the data into 5 distinct subsets, using 4 for training and 1 for testing. This cycle is repeated five times, each time with a different subset used as the test set, while the other four subsets are used for training. The overall performance is then determined by averaging the results from all 5 test sets. Cross-validation is a robust method for assessing a classifier's ability to generalize, as it tests the model on data that was not included in the training phase [31].

For implementing the active learning algorithm, we randomly select 10 examples from each class to serve as labeled data. The remaining examples have their labels removed and are treated as unlabeled. The data distribution for each dataset used is provided in Table 2.

Table 2. Partitioning the datasets to implement active learning techniques

Dataset	Size of Labeled Data	Size of Unlabeled Data	Size of Test Dataset
Breast Cancer	20	435	114
IONOSPHERE	20	261	70
SONOR	20	146	41

We evaluate performance using four classifiers:

- Ordinary Least Squares Linear Regression,
- Linear Support Vector Machine Classifier,
- Polynomial Support Vector Machine Classifier of degree 3,
- A Multi-Layer Perceptron classifier featuring two hidden layers (with 20 and 3 neurons) that employs an LBFGS optimization solver and includes a regularization parameter of $\alpha = 10^{-5}$.

Additionally, three performance metrics were utilized to evaluate the effectiveness of the classifiers [32]:

- **Accuracy:** This metric assesses performance in balanced datasets, gauging the model's efficacy in accurately discriminating between positive and negative examples.
- **F1-Score:** This metric evaluates performance in imbalanced datasets, balancing precision and recall to offer a thorough assessment of model effectiveness.
- **AUC:** This metric measures performance in balanced datasets, evaluating e model's ability to differentiate between positive and negative examples across varied classification thresholds.

4. RESULTS

The performance metrics for uncertainty sampling active learning, aimed at enhancing Relief feature selection, are presented in Table 3. These metrics include the highest values for Accuracy, F1-Score, and AUC, as well as the number of relevant features identified by the Relief algorithm and the important instances selected through the entropy-based Uncertainty Sampling method.

Using Accuracy, F1-Score, and AUC provides a comprehensive evaluation of the model's performance. While Accuracy offers a general overview, it can be misleading in cases of imbalanced datasets. The F1-Score, on the other hand, gives a clearer view of the balance between precision and recall, which is especially useful for imbalanced data. The AUC metric reflects the model's ability to distinguish between classes across various decision thresholds, regardless of class distributions or threshold choices.

Evaluating active learning strategies with Logistic Regression, Support Vector Machines, and Neural Networks offers valuable insights into how well these methods adapt to different decision boundaries and learning paradigms. By combining both linear and nonlinear classifiers, this approach ensures that the active learning strategy remains versatile, effective, and applicable to a wide range of classifier complexities, from basic to advanced. This highlights the flexibility of active learning techniques in handling datasets with varying levels of complexity and structure, ensuring their usefulness in a wide array of real-world applications.

Table 3 provides a comprehensive summary of the peak performance of the examined models: Linear Regression, Support Vector Machine with both linear and polynomial kernels, and a Neural Network. For each classifier, we report its highest values of accuracy, F1-score, and area under the ROC curve (AUC), together with the corresponding number of training examples and features that led to these optimal outcomes. This overview offers a detailed comparison of the classifiers' behavior under different data conditions and highlights the circumstances under which each model achieved its best performance. Figure 2 illustrates the average performance of each classifier on k-fold cross-validated data, using 0 to 200 instances in 5-instance increments.

Table 3. Optimal outcomes for evaluated models

Dataset	Classifier	Accuracy (Max value)	# Selected Examples	# Selected Features	F1-Score (Max value)	# Selected Examples	# Selected Features	AUC (Max value)	# Selected Examples	# Selected Features
Breast Cancer	LR	0.9472	70	20	0.9555	70	20	0.9457	70	20
	SVCL	0.9320	140	25	0.9748	140	25	0.9696	140	20
	SVCP	0.9139	125	15	0.9264	125	15	0.9125	125	15
	NN	0.9244	130	15	0.9344	130	15	0.9251	130	15
Sonar	LR	0.7983	120	33	0.7904	125	39	0.8095	120	35
	SVCL	0.8082	100	31	0.8011	105	31	0.8223	100	37
	SVCP	0.8562	130	51	0.8369	130	51	0.8540	130	51
	NN	0.8608	135	51	0.8473	135	31	0.8693	135	39
Ionosphere	LR	0.8775	180	31	0.9063	180	31	0.8486	180	31
	SVCL	0.8890	175	17	0.9163	180	17	0.8582	175	17
	SVCP	0.8888	170	19	0.9168	170	19	0.8632	170	19
	NN	0.9287	205	23	0.9462	205	23	0.9173	200	23

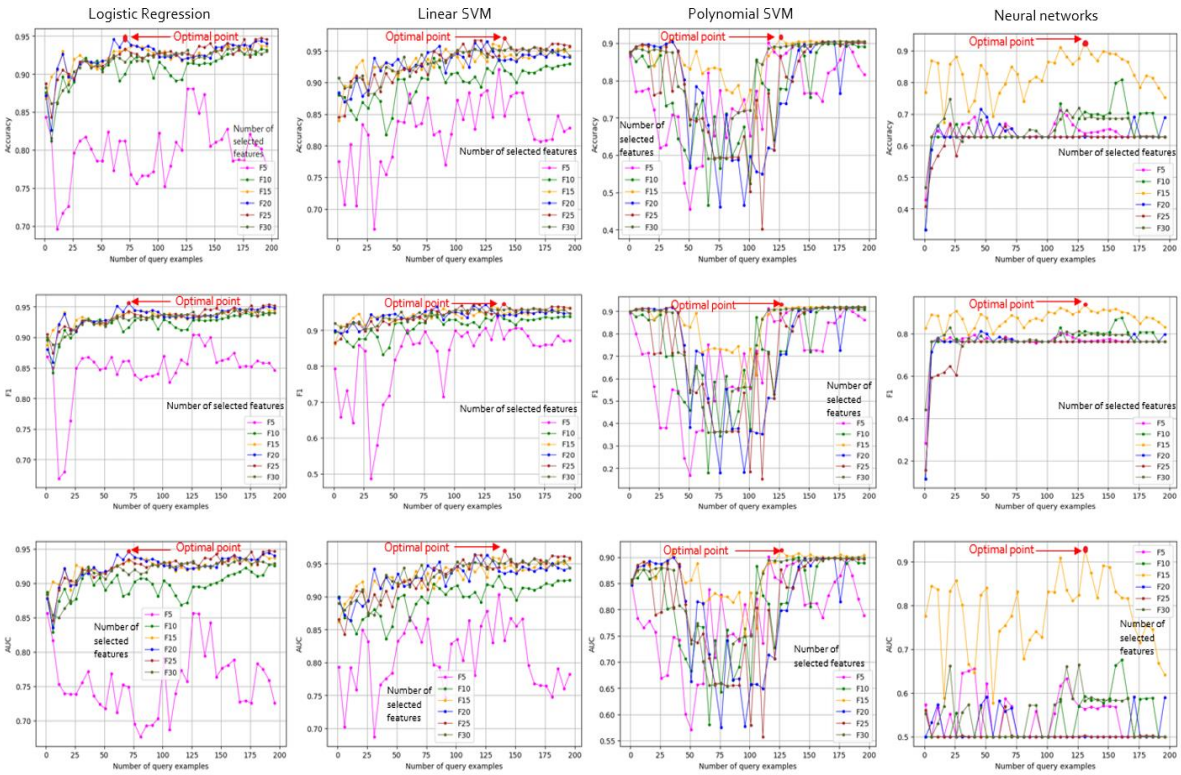

Figure 2. Impact of feature and sample selection on Breast Cancer classification

Figure 2 illustrates the impact of selecting varying numbers of features, represented by distinct colors, after different quantities of training examples have been chosen through the uncertainty sampling strategy. Using the Relief feature selection method, the results on the Breast Cancer dataset show that classifiers often reach their optimal performance before all unlabeled examples are utilized. This demonstrates the importance of focusing on the most informative examples, as Relief effectively identifies the most relevant features and enhances predictive performance. By jointly considering both the number of labeled samples and the dimensionality of the feature space, the analysis highlights how strategic feature selection can improve classifier accuracy, F1-score, and AUC, even without relying on the full dataset.

Figure 3. Evaluation of relevant features selected using the Relief algorithm before selection, after evaluating n=170 examples, and selection of the entire unlabeled dataset n= 265 using F1-Score metric

Figure 3 demonstrates the average F1-Score of various classifiers over varying numbers of the most relevant features, selected using the Relief algorithm. This includes the performance without any feature selection (Blue radar), as well as after selecting a subset of 170 features (Orange radar), and after utilizing the entire unlabeled dataset for feature selection (Grey Radar).

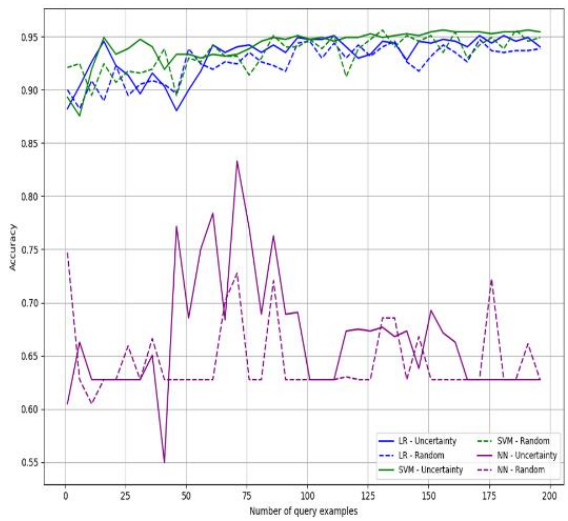

Figure 4. Performance comparison of Uncertainty Sampling and Random Selection using multiple classifiers on the Breast Cancer dataset, with feature selection fixed at 20 features

It can be observed from the Figure 4 that Logistic Regression (LR) achieves a clear benefit from uncertainty sampling, with accuracy steadily improving beyond 0.95 after approximately 120 query examples, compared to around 0.93 for random selection. The improvement of ~2% is statistically

meaningful ($p < 0.05$, paired t-test). For SVM, uncertainty sampling consistently outperforms random selection, reaching values close to 0.96 at its peak, while random selection remains slightly lower around 0.94, showing a gain of ~1.5% that is also significant ($p < 0.05$, paired t-test). In the case of the Neural Network (NN), performance is generally lower, fluctuating between 0.60 and 0.75, yet the uncertainty sampling curve remains above the random one by approximately 3–4% on average, especially after 50 queries. These results, obtained with an enhanced feature selection where 20 features were retained for the Breast Cancer dataset, confirm that uncertainty-based active learning yields more informative labeled sets and significantly improves classification performance across models when compared to random selection.

5. DISCUSSION

The proposed strategy intends to initiate by implementing the most prevalent method in the area of active learning, uncertainty sampling, with the purpose to substantially increase the data size and evaluate this increase on feature selection, particularly for the Relief filtering approach.

By incorporating additional data in general, the selected features will be more extensive, as the distribution is no longer confined to a small data size. However, this study observes that maximum performance is attained prior to the full incorporation of unlabeled data and with an optimal number of features, which underscores the necessity of employing active learning to minimize the cost of labeling.

Carefully selecting a subset of training examples can refine the interpretation of feature selection algorithms. The Breast Cancer data analysis, as shown in Figure 2, reveals that after evaluating 70 examples, most classifiers attain their maximum predictive capacity, underscoring the importance of active learning strategies that identify the most informative examples to improve the relevance of the selected features. Selecting only 5 relevant features selected by Relief approach consistently leads to very poor performance across all classifiers. This is likely coming as a result of the complex constitution of the Breast Cancer dataset, which necessitates a more comprehensive set of features to accurately capture the underlying patterns. However, when 10 or more features are selected, the performance becomes more reliable, suggesting that a larger feature set is necessary to achieve better classification results.

Each classifier exhibits varying levels of success depending on the number of selected features and examples as shown in Table 3, highlighting the importance of careful model selection and hyperparameter tuning for this problem domain.

In cases where the initial dataset is limited to just 10 instances, the performance of the chosen features tends to be suboptimal. However, by strategically incorporating the most informative instances through uncertainty sampling, we can enhance classification results across different numbers of selected features. As illustrated in Figure 3, selecting the right number of examples is vital for effective feature selection. The orange radar, representing 170 examples out of a total of 255, achieves the best performance among the classifiers. The graph also illustrates the range of features identified as most relevant by the Relief algorithm.

The results in Figure 4 demonstrate that uncertainty sampling consistently outperforms random selection across classifiers on the Breast Cancer dataset with 20 selected features. This confirms that active learning provides a more effective strategy for guiding the feature selection process and improving classification performance compared to random querying.

This approach addresses the challenge of small datasets by strategically selecting a few key examples, enhancing the model's ability to recognize underlying patterns and produce a more meaningful subset of features.

By leveraging performance metrics such as Accuracy, F1-score, and Area Under the Curve (AUC), we obtain a comprehensive evaluation of the model's effectiveness, regardless of dataset imbalance. These metrics provide an overall assessment of how well the model performs after example selection and preprocessing.

Integrating both linear and nonlinear classifiers allows for a more thorough assessment of this method across different datasets. This strategy ensures a more well-rounded evaluation of model performance, as linear and nonlinear classifiers capture distinct aspects of the data and offer complementary insights.

The study's findings emphasize the advantages of combining active learning with feature selection techniques to enhance classification performance across diverse datasets.

6. CONCLUSION

Active learning is a highly adaptable and efficient approach, particularly valuable in situations where labeled data is scarce. By strategically selecting the most informative examples for labeling, active learning helps overcome the challenges posed by limited labeled data and significantly enhances the effectiveness of machine learning models. This method optimizes the use of available labeled data, resulting in improved model performance compared to traditional supervised learning techniques. Among the various active learning strategies, uncertainty sampling is the most commonly applied in machine learning tasks.

Feature selection aims to identify the most relevant features rather than relying on the entire feature set. Relief-F, a widely used feature selection algorithm, has been applied across multiple domains, demonstrating its ability to enhance model performance by reducing data dimensionality.

When working with extremely small datasets, feature selection techniques may struggle to identify the most meaningful features. To address this issue, an active learning strategy based on uncertainty sampling is employed to significantly expand the dataset. This expansion is then followed by an evaluation of its impact on the performance of the Relief feature selection algorithm.

Experiments Conducted on the Three Datasets in this research indicate that optimal performance can be attained by choosing an appropriate number of informative examples through an uncertainty sampling strategy.

Moreover, the selection of features becomes even more critical after establishing the optimal number of examples. This study can be relevant for high-dimensional data, as the data volume is often slight in relation to the number of features, which can negatively influence the identification of the most relevant features. Additionally, the evaluation can be broadened to include other feature selection methods, such as filtering, wrapper, and embedded techniques, to significantly lower labeling costs and improve the identification of the most

relevant features.

Additionally, using different active learning strategies can highlight the efficacy of the recommended approach. The main goal of this study is to create a thorough methodology for optimizing the balance between the number of examples and features in classification tasks.

Increasing the dataset size can incur higher computational costs, but this investment may be justified by the resulting performance improvements observed in the classification task.

REFERENCES

[1] Wan, J., Chen, H., Li, T., Huang, W., Li, M., Luo, C. (2022). R2CI: Information theoretic-guided feature selection with multiple correlations. Pattern Recognition, 127, 108603. https://doi.org/10.1016/j.patcog.2022.108603

[2] Qian, W., Huang, J., Xu, F., Shu, W., Ding, W. (2023). A survey on multi-label feature selection from perspectives of label fusion. Information Fusion, 100: 101948. https://doi.org/10.1016/j.inffus.2023.101948

[3] Janane, F.Z., Ouaderhman, T., Chamlal, H. (2023). A filter feature selection for high-dimensional data. Journal of Algorithms & Computational Technology, 17: 17483026231184171. https://doi.org/10.1177/17483026231184171

[4] Hasan, F.M., Hussein, T.F., Saleem, H.D., Qasim, O.S. (2024). Enhanced Unsupervised Feature Selection Method Using Crow Search Algorithm and Calinski-Harabasz. International Journal of Computational Methods and Experimental Measurements, 12(2): 185-190. https://doi.org/10.18280/ijcmem.120208

[5] Aaboub, F., Chamlal, H., Ouaderhman, T. (2023). Statistical analysis of various splitting criteria for decision trees. Journal of Algorithms & Computational Technology, 17: 17483026231198181. https://doi.org/10.1177/17483026231198181

[6] Chamlal, H., Benzmane, A., Ouaderhman, T. (2024). Elastic net-based high dimensional data selection for regression. Expert Systems with Applications, 244: 122958. https://doi.org/10.1016/j.eswa.2023.122958

[7] Ali, M.Z., Abdullah, A., Zaki, A.M., Rizk, F.H., Eid, M.M., El-Kenway, E.M. (2024). Advances and challenges in feature selection methods: A comprehensive review. Journal of Artificial Intelligence and Metaheuristics, 7(1): 67-77. https://doi.org/10.54216/JAIM.070105

[8] Rietdijk, H.H., Strijbos, D.O., Conde-Cespedes, P., Dijkhuis, T.B., Oldenhuis, H.K., Trocan, M. (2024). Feature Selection with Small Data Sets: Identifying Feature Importance for Predictive Classification of Return-to-Work Date after Knee Arthroplasty. Applied Sciences, 14(20): 9389. https://doi.org/10.3390/app14209389

[9] Settles, B. (2010). Active Learning Literature Survey. University of Wisconsin, Madison, 52. http://digital.library.wisc.edu/1793/60660.

[10] Tharwat, A., Schenck, W. (2023). A survey on active learning: State-of-the-art, practical challenges and research directions. Mathematics, 11(4): 820. https://doi.org/10.3390/math11040820

[11] Settles, B. (2012). Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning.

[12] Sahan, M., Smidl, V., Marik, R. (2021). Active Learning for Text Classification and Fake News Detection. In Proceedings of the International Symposium on Computer Science and Intelligent Controls (ISCSIC), Rome, Italy, pp. 87-94. https://doi.org/10.1109/ISCSIC54682.2021.00027

[13] Sun, L., Gong, Y. (2019). Active learning for image classification: A deep reinforcement learning approach. In 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI), Xi'an, China, pp. 71-76. https://doi.org/10.1109/CCHI.2019.8901911

[14] Holub, A., Perona, P., Burl, M.C. (2008). Entropy-based active learning for object recognition. In 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, AK, USA, pp. 1-8. https://doi.org/10.1109/CVPRW.2008.4563068

[15] Mehdi, B., Hasna, C., Tayeb, O. (2019). Intelligent credit scoring system using knowledge management. IAES International Journal of Artificial Intelligence, 8(4): 391. https://doi.org/10.11591/ijai.v8i4.pp391-398

[16] Branson, S., Wah, C., Schroff, F., Babenko, B., Welinder, P., Perona, P., Belongie, S. (2010). Visual recognition with humans in the loop. In European Conference on Computer Vision, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 438-451. https://doi.org/10.1007/978-3-642-15561-1_32

[17] Budd, S., Robinson, E., Kainz, B. (2021). A survey on active learning and human-in-the-loop deep learning for medical image analysis. Medical Image Analysis, 71: 102062. https://doi.org/10.1016/j.media.2021.102062

[18] Chamlal, H., Aaboub, F., Ouaderhman, T. (2024). A preordonance-based decision tree method and its parallel implementation in the framework of Map-Reduce. Applied Soft Computing, 167: 112261. https://doi.org/10.1016/j.asoc.2024.112261

[19] Sun, L.L., Wang, X.Z. (2010). A survey on active learning strategy. In 2010 International Conference on Machine Learning and Cybernetics, Qingdao, China, pp. 161-166. https://doi.org/10.1109/ICMLC.2010.5581075

[20] Rückstieß, T., Osendorfer, C., Van Der Smagt, P. (2011). Sequential feature selection for classification. In Australasian Joint Conference on Artificial Intelligence, pp. 132-141. https://doi.org/10.1007/978-3-642-25832-9_14

[21] Scheffer, T., Decomain, C., Wrobel, S. (2001). Active hidden markov models for information extraction. In International symposium on intelligent data analysis, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 309-318. https://doi.org/10.1007/3-540-44816-0_31

[22] Liu, S., Li, X. (2023). Understanding Uncertainty Sampling. arXiv preprint arXiv:2307.02719. https://doi.org/10.48550/arXiv.2307.02719

[23] Guyon, I., Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3(Mar): 1157-1182. https://doi.org/10.1162/153244303322753616

[24] Sánchez-Maroño, N., Alonso-Betanzos, A., Tombilla-Sanromán, M. (2007). Filter methods for feature selection–a comparative study. In International conference on intelligent data engineering and automated learning, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 178-187. https://doi.org/10.1007/978.3.540-77226-2.19

[25] Patel, D., Saxena, A., Wang, J. (2024). A machine

learning-based wrapper method for feature selection. International Journal of Data Warehousing and Mining, 20(1): 1-33. https://doi.org/10.4018/IJDWM.352041

[26] Zhang, Y., Zhao, Z., Zhao, S., Liu, Y., He, K. (2020). A New Embedded Feature Selection Method using IBALO mixed with MRMR criteria. Journal of Physics: Conference Series, 1453(1): 012027. https://doi.org/10.1088/1742-6596/1453/1/012027

[27] Kamalov, F., Sulieman, H., Alzaatreh, A., Emarly, M., Chamlal, H., Safaraliev, M. (2025). Mathematical methods in feature selection: A review. Mathematics, 13(6): 996. https://doi.org/10.3390/math13060996

[28] Kira, K., Rendell, L. A. (1992). A practical approach to feature selection. In Machine Learning Proceedings 1992, Morgan Kaufmann, pp. 249-256. https://doi.org/10.1016/B978-1-55860-247-2.50037-1

[29] UCI Machine Learning Repository. (2017). University of California, Irvine.

[30] Chamorro-Atalaya, O., Arévalo-Tuesta, J., Balarezo-Mares, D., Gonzáles-Pacheco, A., et al. (2023). K-fold cross-validation through identification of the opinion classification algorithm for the satisfaction of university students. International Journal of Online & Biomedical Engineering, 19(11): 39887. https://doi.org/10.3991/ijoe.v19i11.39887

[31] Hastie, T., Tibshirani, R., Friedman, J. (2009). The elements of statistical learning. Preface to the Second Edition.

[32] Santoso, D.A., Rizqa, I., Aqmala, D., Alzami, F., Rijati, N., Marjuni, A. (2025). Performance analysis of multiple knapsack problem optimization algorithms: A comparative study for retail and SME applications. Ingénierie des Systèmes d'Information, 30(2): 533-550. https://doi.org/10.18280/isi.300224