Dimensionality Reduction with Ensemble Learning using Growing Hierarchical Adaptive Self-Organizing Map

Chiraz Jlassi
Ecole Nationale d'Ingénieurs de Tunis
University of Tunis El Manar
BP 37, Belvédère 1002 Tunis

Ameni Filali
Department of Computer Engineering,
College of Computer Science
and Engineering, University of Hail

Najet Arous
Higher Institute of Computer Science
University of Tunis El Manar
BP 37, Belvédère 1002 Tunis

ABSTRACT
Feature selection is a technique designed to reduce the complexity of learning models by selecting the most relevant features, thereby enhancing the interpretability of the models while maintaining strong generalization performance. This study addresses the challenge of choosing a subset of the most important features for each cluster within a dataset. The proposed method extends the Random Forests approach by incorporating growing hierarchical adaptive self-organizing maps (GH_AdSOM) variant for unlabelled data. It assesses the out-of-bag feature importance across multiple partitions, each generated through various bootstrap samples and a random subset of features. The GH_AdSOM represents a neural network architecture that synergizes the benefits of two key enhancements to the self-organizing map: dynamic growth and hierarchical structure. This approach allows for adaptability in map size as well as a layered organization, resulting in a powerful and flexible neural network model.

General Terms
Dimensionality reduction; Unsupervised learning; Ensemble learning;

Keywords
Growing hierarchical adaptive self-organizing map; random forest; feature selection; recursive feature elimination.

1. INTRODUCTION
Selecting relevant features in high-dimensional data poses a significant challenge. This process is crucial for reducing the dimensionality of the original data, leading to enhanced learning performance by eliminating irrelevant and redundant features. The immediate outcomes include speeding up data mining algorithms and improving overall performance. The process of choosing the optimal subset of features is referred to as feature selection, and it stands out as one of the most intriguing and contemporary research subjects in the field of artificial intelligence [1], [2]. A unique clustering method provides generating very bad groupings as the learning algorithms decompose with high dimensional data. Learning ensembles is an efficient solution to surmount the dimensionality problem and to enhance the robustness of the clustering [3], [4], [5]. The concept involves merging outcomes from various clustering analyses into a unified data partition without direct access to the original features.

Several feature selection algorithms have been studied and proposed for machine learning applications.

These methods may be categorized as supervised methods [6], semi-supervised methods [7], and unsupervised methods, according to the way of employing label information.

Indeed, the absence of labels poses a significant challenge in the pursuit of discriminative variables, making unsupervised feature selection a more complex problem [8]. In this context, unsupervised methods evaluate the relevance of features based on their ability to retain specific properties of the data without relying on labelled information. In many real-world scenarios where high-dimensional data accumulates rapidly, there is often an abundance of unlabelled data. Consequently, the demand and potential for developing unsupervised feature selection methods are high, as they offer a promising avenue for efficiently handling data with limited or no labelled information. In this paper, we investigate some intrinsic properties of unsupervised feature selection methods discover their different connections and establish a unified approach that will permit us together to study unsupervised feature selection methods, gain a deeper comprehension of some existing successful methods, and obtain new approaches with better performance.

The proposed model, which make at the same time dimensionality reduction and Learning ensembles using growing hierarchical SOM variant (GH_AdSOM) with locally adapting neighborhood radii in order to characterize clusters. The central concept of the Growing Hierarchical Self-Organizing Map (GHSOM), is to employ a hierarchical arrangement comprising multiple layers, where each layer contains several independent Self-Organizing Maps (SOMs), also known as many SOMs. Initially, only one SOM, following Kohonen's framework is utilized at the first layer of the hierarchy. For each unit within this map, a SOM might be appended to the subsequent layer of the hierarchy. The GHSOM evolves in two dimensions: horizontally, by enlarging the size of each SOM, and vertically, by augmenting the number of layers. In the proposed model the RF approach to unlabelled data is broadened through the incorporation of an ensemble learning technique known as Random Forests Feature Selection (RFS). RFS integrates both bagging and random subspaces strategies to construct an ensemble of component learning models. In these experiments, the evidence accumulation technique proposed by [9] is adopted. This method facilitates the gathering of co-occurrences of object pairs within the observed cluster, treating them as evidence for their potential connection. Subsequently, the co-association matrix of objects introduces a new similarity measure among objects. Consensus learning is achieved by applying a classic average-link hierarchical agglomerative method to the matrix obtained from the evidence accumulation technique. During the generation of consensus clustering, we locally select relevant features for each final cluster using the out-of-bag (OOB) importance measure within the random forest framework. Next, RFS is boosted with popular recursive feature elimination (RFE) [10], method which recursively eliminates the features

with the lowest significance measure. The structure of the remaining paper is as follows:

Section 2 discusses recent studies on unsupervised feature selection and consensus learning approaches. And Growing Hierarchical self-organizing Kohonen Maps (GHSOM) and another variant are reviewed. Following that, we introduce the proposed approach, aimed at enhancing the final feature ranking. Additional experiments supporting the relevance of the method are discussed in Section 3. Section 4 presents experiments conducted on benchmark datasets, along with a comparison of the new proposed method. Finally, Section 5 concludes this work.

2. RELATED WORKS

This section discusses studies on unsupervised feature selection and consensus learning approaches.

2.1 Unsupervised feature selection

Unsupervised feature selection is indeed more challenging because it lacks label information, and its objective is to identify a subset of features that preserves the inherent clusters in the data based on a specified clustering criterion. Different strategies for unsupervised feature selection can be broadly categorized into embedded, filter, and wrapper methods [11].

Embedded methods [12], integrate feature selection in the model training process. The feature selection is inherent in the learning algorithm. For example, LASSO (Least Absolute Shrinkage and Selection Operator), Elastic Net, Decision Trees with feature importance measures, Regularized regression models (Ridge regression). Filter methods [13] evaluate the relevance of each feature independently of the learning algorithm. They are applied before the actual learning process. For example: Mutual Information, Correlation-based feature selection, Chi-squared test, Information Gain. Wrapper methods [14] use a specific learning algorithm to evaluate different feature subsets. The performance of the model guides the feature selection process. For example: Recursive Feature Elimination (RFE), Forward Selection, Backward Elimination, Genetic Algorithms for feature selection.

2.2 Ensemble learning

Ensemble learning has emerged as a powerful approach to enhance the robustness and stability of clustering solutions. Similar to its application in the supervised setting, ensemble clustering involves two main steps: the generation of multiple clustering solutions and their combination. The method used to combine individual clusterings is commonly referred to as the consensus function. Several approaches for building ensemble clustering methods include: Bagging involves generating multiple clustering solutions by creating random subsets (bootstrap samples) of the dataset and applying the clustering algorithm to each subset independently [15]. Random Subspaces methods generate diverse clustering solutions by considering only random subsets of features (subspaces) for each clustering algorithm iteration. This introduces variability and helps capture different aspects of the data. Ensembles can be formed by applying multiple clustering algorithms independently and combining their results. This allows the ensemble to capture diverse patterns that different algorithm may be sensitive to. These ensemble methods aim to improve the overall performance and generalization of clustering solutions by reducing the sensitivity to the initial conditions, dataset variations, or algorithm-specific biases. The combination of multiple clustering helps create a more robust and stable solution, often resulting in improved clustering quality.

Ensemble clustering is particularly beneficial when dealing with complex and high-dimensional datasets, where individual clustering algorithms may struggle to capture all relevant patterns. The consensus function plays a crucial role in combining the multiple clustering into a cohesive and representative solution.

Random Forests [16] stand out as highly effective techniques for constructing classifier ensembles. Their robust performance is derived from two key components: the utilization of random subsamples of the training data, akin to bagging, and the introduction of randomness in the method employed for grouping base-level classifiers, which are typically decision trees. In the initial phase, random subsamples of the learning data are leveraged, resembling the bagging approach. This enhances the diversity of the individual trees within the ensemble, leading to a more robust and generalized model. The distinctive aspect of Random Forests lies in the randomization injected into the process of constructing decision trees. At each stage of tree building, the algorithm arbitrarily selects a subset of features, and from this subset, it identifies the most influential ones. This introduces an element of unpredictability, preventing the ensemble from overfitting to specific features and improving its ability to handle diverse datasets.

2.3 Selection methods

Various techniques exist for determining the importance of local features in their assessment, including statistical tests such as the Scree test proposed by Cattell in 1966. Other methods involve choosing a predefined percentage or number of features or selecting variables whose significance exceeds a user-specified cutoff threshold. The Random Forests Feature Selection (RFS) method, a form of integration method with a locally adaptive measure of relevance, is proposed as one of these techniques.

The RFS method employs a locally adaptive measure of pertinence, offering a dynamic approach to feature selection. While numerous stop criteria have been proposed in the literature, a common challenge is the reliance on a threshold that is dataset dependent.

In the pursuit of identifying pertinent features, the Scree Test, a statistical test, is applied. This technique enables the automatic selection of relevant variables without the need for a predefined stop threshold. Additionally, the Recursive Feature Elimination (RFE) approach is utilized, adding another layer to the feature selection process. RFE iteratively removes the least significant features, contributing to the identification of the most relevant variables for the given problem without the need for explicit threshold specifications.

The Recursive Feature Elimination (RFE) selection method [17], operates as a recursive procedure, ranking variables based on a designated measure of their significance. In the RFE algorithm in each iteration, the significance of variables is calculated, and the least pertinent ones are systematically removed. Notably, the elimination process occurs in groups of variables to expedite the overall procedure. The recursive nature of RFE is essential because, for certain measures, the significance of each variable can undergo substantial changes when assessed over various subsets of variables during the stepwise elimination process, especially when dealing with highly correlated variables. The order in which variables are eliminated throughout the recursive process is then utilized to generate a final ranking. The variable selection procedure allows for the extraction of the top n variables from this ranking, providing a subset of the most relevant variables for the given problem. This iterative and recursive approach of

RFE ensures a thorough exploration of the variable space, accounting for the potential fluctuations in significance that may arise when dealing with correlated variables.

The Recursive Feature Elimination (RFE) approach is commonly employed in conjunction with Support Vector Machines (SVMs) to iteratively build a model while eliminating variables with low weights [18], [19]. This method can be perceived as falling between filters and wrappers regarding computational complexity. Within the Recursive Feature Elimination-Support Vector Machine (RFE-SVM) framework, an external estimator, typically a Support Vector Machine (SVM) assigns weights to variables, represented as coefficients of a linear model. RFE then recursively selects variables based on successively smaller sets of variables. The initial estimator is trained on the entire set of variables, and weights are assigned to each variable. Subsequently, variables with the smallest total weights are pruned from the current set of variables. This pruning process is iteratively applied to the reduced ensemble until the desired number of variables is finally achieved. In this context, the RFE method is effective for feature selection by leveraging the capability of an external estimator, such as an SVM, to assign weights to variables. This approach systematically identifies and eliminates non-pertinent and redundant features, contributing to the construction of a model with a reduced set of relevant variables. In this paper, an alternative approach is proposed where RFE is used in conjunction with a Growing Hierarchical Adaptative Self-Organizing Map (GH_AdSOM) to achieve similar goals of constructing a model while refining and selecting pertinent variables.

2.4 Growing Hierarchical Adaptive Self-Organizing Maps: GH_AdSOM

The fundamental concept behind the Growing Hierarchical Self-Organizing Model (GHSOM) [20], is to employ a hierarchical arrangement with multiple layers, where each layer comprises several independent Self-Organizing Maps (SOMs), often referred to as many SOMs. Initially, only one SOM is utilized at the first layer of the hierarchy. For each unit within this map, the possibility exists to add a SOM to the subsequent layer of the hierarchy. This underlying principle is iteratively applied to the third and any subsequent layers of the GHSOM. The growth of the GHSOM occurs in two dimensions: width expansion, achieved by increasing the size of each SOM, and depth expansion, accomplished by adding more layers to the hierarchical structure.

In the GH_AdSOM, each neuron (denoted as i) possesses an individual neighborhood width parameter, σi. When the training process commences, the initial value of σi is set to half the diameter of the lattice for all neurons. Throughout training, sample vectors are introduced to the GH_AdSOM, like the basic GHSOM procedure. The weights undergo adjustments using the well-known update rule, characterized by a diminishing learning rate. Notably, the width of the neighbourhood during training is dictated by the σ_i value associated with the best matching unit. The parameter σ_i is determined by the local topographic errors. If for a sample vector x the two nearest weight vectors are w_j and w_k, and the corresponding best matching units n_j and n_k are not adjacent, there is local topographic error. Then, for units n_i that are near the two BMUs, a new value for the neighbourhood radius σ_i is calculated Eq 2.

$$\sigma_i = \begin{cases} \left\| n_j - n_k \right\| & \text{if } \max \{\|n_i\text{-}n_j\|, \|n_i\text{-}n_k\|\} \le \|n_j\text{-}n_k\| \\ \left\| n_j - n_k \right\| - s & \text{if } s = \min \{\|n_i\text{-}n_j\|, \|n_i\text{-}n_k\|\} \le \|n_j\text{-}n_k\| \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

Thus, for units in the middle of the two BMUs, the new σ_i is the same to the distance on the map between the BMUs; outside that area, the σ_i decreases linearly to one.

For the proposed GHSOM variant, during the training phase, when a sample input vector is assigned to a Best Matching Unit (BMU) unit, we store its associated value, label, and update its activation frequency. Each unit in the proposed variant is characterized by:

- A general centroid prototype vector: determined through the Kohonen update rule [21].
- Information specific to each phoneme class assigned to a unit: including the mean prototype vector, label, and activation frequency.

3. PROPOSED METHODOLOGY

This section provides a detailed explanation of clustering ensembles based on unsupervised feature selection using GH_AdSOM variant as the base learning approach, boosted with a Random Forest Feature Selection with Recursive Feature Elimination (RFS-RFE).

The synergy between GH_AdSOM variants and RFS-RFE offers a dual-layered approach. The GH_AdSOM variants excel at capturing complex patterns in the data, providing a robust foundation for clustering. Meanwhile, RFS-RFE contributes by iteratively selecting and discarding features, ensuring that only the most discriminative and informative attributes are retained.

3.1 Merging multiple clustering solutions

The proposed methodology aimed at assessing the significance of variables. Given that ensemble learning is crafted to enhance the performance of both unsupervised and supervised classification algorithms, it is plausible to explore its applicability in addressing the unsupervised feature selection (FS) challenge. This methodology suggests integrating random subspaces and bagging to generate a collection of component clustering solutions. RFS-RFE produces a series of clustering outcomes through the following steps: a novel training set is constructed, with replacement, from the initial dataset. Additionally, variables are randomly selected from the entire feature set, and a clustering solution is derived by applying the base clustering method to the chosen variables. This process is reiterated t times. The process of combining these clusters is referred to as the cluster ensemble problem. Various methodologies have been suggested in existing literature to address this challenge. In this study, the average-link is employed as a consensus function, which is based on co-association values between data. Subsequently, the similarity between pairs of cases is computed by directly assessing the percentage of clusters that these patterns have in common in the initial partitions. Various similarity-based clustering techniques can be employed on the similarity matrix to formulate the final partition. In the proposed approach, Agglomerative Hierarchical Classification is used. The determination of the optimal number of clusters is based on maximizing the average mutual information criterion.

3.2 Measure feature importance

The RFS-RFE algorithm provides a framework for addressing the issue of feature relevance evaluation. Within RFS-RFE, the combination of bagging and feature subspace enables continuous assessments of the significance of features within the aggregated clustering set. Similar to the methodology in random forest, these evaluations are conducted out-of-bag (oob). In RFS-RFE, each clustering iteration involves a subset of the clustering ensemble that is excluded during the learning process. This exclusion because each clustering is developed solely on a bootstrap sample. The out-of-bag (oob) subsets serve as unbiased measures of variable importance. In assessing the importance of variables within the clustering process, RFS-RFE adopts the following approach. After each clustering iteration, every variable undergoes arbitrary permutation within the oob examples, and the oob data are then reassigned into clusters. The oob cluster assignments, with the perturbed *f*-th feature, are compared with the initial cluster assignment. Features that are not relevant will typically not alter the classification when modified in this manner. The importance score for feature-f is determined by calculating the average number of times the feature-f-permuted oob example is misclassified across all clustering runs in the set T. Additionally, the importance of the *f*-th variable for a particular cluster in the final agreement clustering is calculated as the sum of all the significance values over all the objects falling into that particular cluster.

As advocated by Breiman, precision in evaluating the model should ideally match that of using an assessment set of the same size as the dataset. Consequently, the utilization of the out-of-bag (oob) error estimate obviates the need for a separate ensemble test set. In each bootstrap dataset, approximately one-third of the instances are left out. As a result, oob assessments involve the combination of only about one-third as many clustering methods as in the ongoing essential combination. The incorporation of local feature selection enables the characterization of each object cluster.

The unsupervised feature selection method, RFS-RFE with GH_AdSOM, is illustrated in Figure.1. It's worth noting that various approaches can be employed to select the most crucial local features based on their significance assessments. These approaches may include statistical tests, the Scree test (Kotsiantis and Kanellopoulos 2012) selecting a predefined percentage or number of features, choosing features whose significance exceeds a user-defined cutoff threshold, and employing the same recursive feature elimination (RFE) scheme used with Self-Organizing Map (SOM).

4. EXPERIMENTAL VALIDATION

This section analyzes the experimental results.

4.1 Datasets

Nineteen UCI datasets were selected to assess the effectiveness of RFS-RFE using the GH AdSOM as the base clustering algorithm in performance evaluation. These datasets are widely employed to test unsupervised feature selection methods. Table1 provides a comprehensive overview of the details pertaining to these benchmark datasets utilized in the experiments.

Furthermore, we compared RFS-RFE based on GH_AdSOM as a clustering method with RFS-RFE based on the AdSOM variant as the clustering method, which has proven its effectiveness compared to other SOM variants.

4.2 Evaluation metrics

Calculating the Normalized Mutual Information (NMI) for Growing Hierarchical Self-Organizing Maps (GHSOM) involves comparing the clustering results of your algorithm with some ground truth or reference clustering. The NMI is a measure of the mutual information between the true labels and the predicted labels, normalized by the entropy of the individual label sets [22].

Calculate the mutual information I(X; Y) between the true labels and predicted clusters :

$$I(X;Y) = \sum_{i=1}^{k}\sum_{j=1}^{n} P(X_i\, Y_j)\,.\,log_2\left(\frac{P(X_i \cap Y_i)}{P(X_i).P(Y_i)}\right) \qquad (3)$$

Where $P(Xi \cap Yj)$ is the joint probability of true label i and predicted cluster j.

Normalize the mutual information to obtain the NMI:

$$NMI(X,Y) = \frac{I(X;Y)}{\sqrt{H(X).H(Y)}} \qquad (4)$$

NMI ranges from 0 to 1, where 1 indicates perfect clustering agreement.

Table1. The used datasets

Datasets	Instances	Features	Labels
Breast tissue	106	9	7
Ionosphere	351	34	2
Isolet	1559	617	26
Leukemia	73	7129	2
Lung	32	56	3
Madelon	2600	500	3
Multiple features	2000	216	10
Ovarian	54	1536	2
Parkinson	195	22	2
Pima	768	8	2
Promoters	106	57	2
Robot	88	90	4
Segmentation	2310	19	7
Soybean	266	35	15
Spect	267	22	2
Wdbc	569	30	2
Wine	178	13	3
Wisconsin	699	9	2
Waveform	5000	40	3

4.3 Experimental results

In these experiments, the Recursive Feature Elimination (RFE) selection method is executed as an iterative process that arranges variables based on a significance measure in a descending order. During each iteration, the least important variable is removed, continuing until the desired number of features is achieved.

The primary idea behind testing the proposed approach is to evaluate the clustering quality by utilizing an external criterion. This criterion assesses how effectively the clustering aligns with the known class labels. The clustering performance on a validation test set serves as an indication of the importance of the selected variables.

In this section, the GH_AdSOM variant was utilized to generate the partitions for the combination.

Next, the performance of Recursive Feature Selection (RFS) is enhanced by incorporating the Recursive Feature Elimination (RFE) strategy, as previously discussed. Then the clustering performance is assessed in relation to the variables ranked in decreasing order of significance. The base clustering algorithm, once again, was the variant of hierarchical Self-Organizing Maps (GH_AdSOM).

To gauge the effectiveness of the feature selection process, the clusters formed by GH_AdSOM using the variables selected by RFS are compared against clusters obtained by employing GH_AdSOM on all features. Additionally, these results are compared with clusters formed using the features selected by RFS, further boosted with the RFE method using SOM variant AdSOM as the base classifier. The method of enrichment of information explained in section 2 is ensured in all the clustering method used.

Experimental results are shown in Table 2.

The normalized mutual information (NMI) of clustering methods AdSOM and GH_AdSOM on subsets of attributes selected by RFS-RFE are much higher than those obtained by these clustering methods on all attributes.

So, the hierarchical aspect of GH_AdSOM enhances the clustering method, as demonstrated by comparing it with the AdSOM variant. The NMI of clustering solutions generated by GH_AdSOM on features selected by RFS-RFE is higher than those generated by AdSOM maps on features selected by the feature selection method based on Random Forest applying the recursive elimination process on the nineteen databases. For example, the GH_AdSOM clustering method achieved an NMI of 91.75% for the Waveform database defined on features selected by RFS-RFE, which is higher than the 83.15% obtained by AdSOM on features selected by RFS-RFE. Knowing that the results of AdSOM variant surpass those generated by SOM-E, SEQ-OPT and DeSieno clustering methods were used on the variables selected by RFS algorithm with a RFE [23]. It's important to note that the NMI is influenced by the number of selected features for each dataset.

For example, on the Wdbc dataset, RFS-RFE identified features {f3, f7, f28} that yield significant clusters with an NMI of 95.67% using GH_AdSOM, while AdSOM only achieves 90.86% for the same number of features. This same trend is expected to be observed with the Madelon, waveform, and wine databases except for Pima. For Madelon, features {f29, f337} are the most important attributes selected by RFS-RFE, resulting in a rate of 70.45% with GH-AdSOM, which is notably better than the 65.57% accuracy generated with AdSOM on the same attributes. Figure2 shows a clear performance improvement when feature selection is applied. Across almost all datasets, RFS_RFE GH_AdSOM achieves the highest NMI values, indicating its robustness and effectiveness. Both RFS-based methods consistently outperform the models using all features, demonstrating that feature selection significantly enhances clustering quality. The performance gain is particularly noticeable on datasets such as Wdbc, Waveform, Soybean, and Wisconsin.

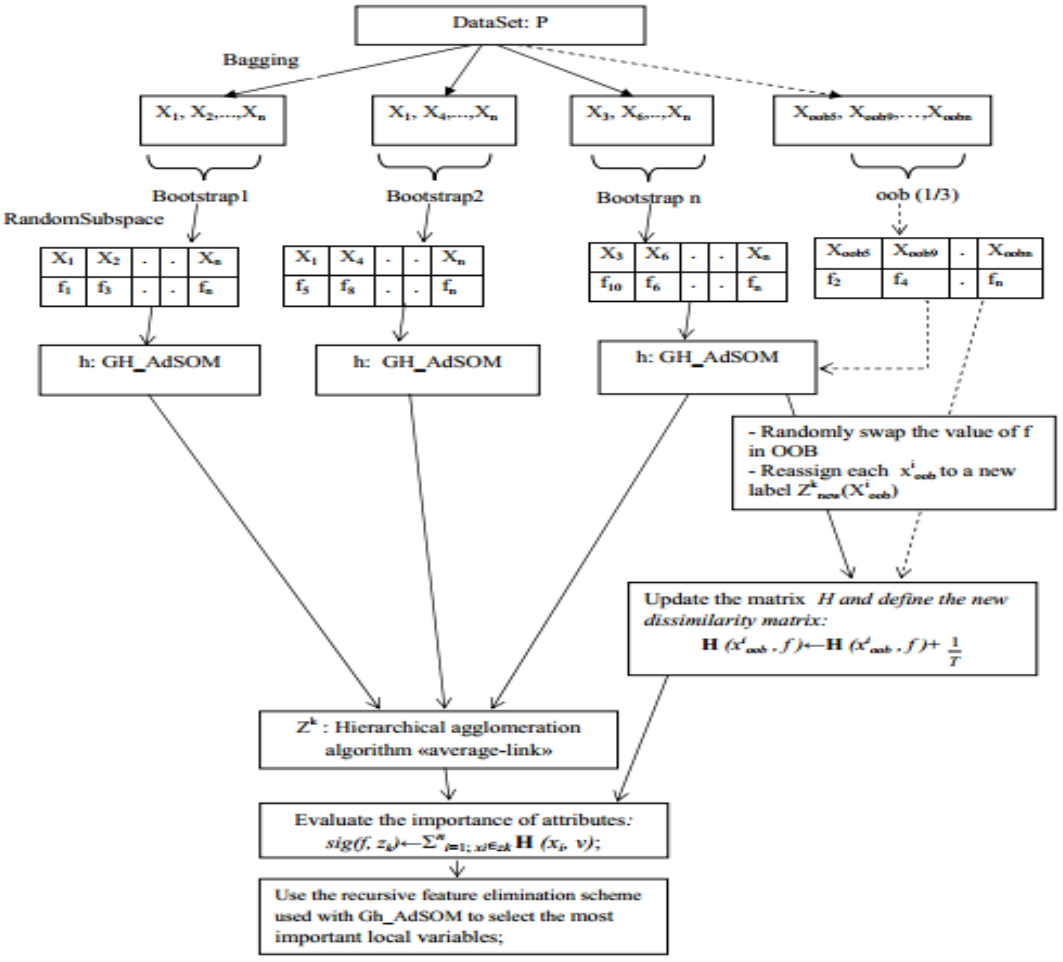

Fig1: Proposed Methodology

Table2. Experimental results on UCI datasets

Dataset	All Features AdSOM	All Features GH_AdSOM	RFS_RFE AdSOM	RFS_RFE GH_AdSOM	Number of Selected Features
Breast tissue	60,48	66,12	71,32 ± 3,3	80,11 ± 1,25	6
Ionosphere	18,85	20,45	22,33 ± 2,17	32,45 ± 2,76	8
Isolet	62,51	70,11	75,45 ± 0,88	75,97 ± 3,11	26
Leukemia	15,88	19,87	21,44 ± 2,75	30,88 ± 1,56	9
Lung	27,36	32,65	38,95 ± 6,11	40,15 ± 3,66	7
Madelon	55,66	60,17	65,57 ± 0,00	70,45 ± 1,45	2
Multiple features	49,04	66,64	79,25 ± 3,10	83,12 ± 2,56	9
Ovarian	57,74	59,77	64,76 ± 2,98	67,95 ± 1,95	6
Parkinson	25,98	30,12	32,45 ± 5,5	39,75 ± 5,03	2
Pima	13,84	14,67	17,44 ± 2,32	28,11 ± 0,66	3
Promoters	32,51	41,50	49,22 ± 1,51	52,75 ± 1,13	4
Robot	44,55	48,90	58,45 ± 3,12	60,88 ± 3,12	10
Segmentation	69,13	77,12	79,65 ± 1,75	82,23 ± 1,25	13
Spect	18,36	20,22	25,75 ± 1,32	33,78 ± 2,56	7
Soybean	77,57	78,89	81,35 ± 1,75	88,65 ± 5,1	28
Wine	73,38	75,56	88,74 ± 2,75	90,11 ± 3,24	5
Wdbc	88,85	90,01	90,86 ± 2,11	95,67 ± 1,78	3
Wisconsin	73,48	75,44	81,95 ± 2,75	89,77 ± 3,45	6
Waveform	62,50	80,70	83,15 ± 2,45	91,75 ± 3,40	4

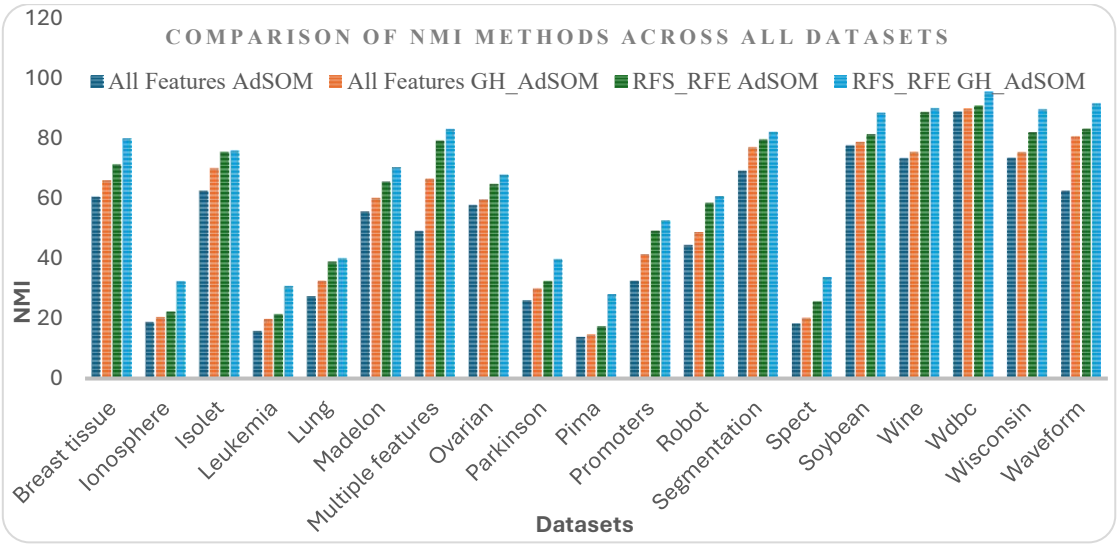

Fig 2: Comparison of NMI Methods Across All Datasets

5. CONCLUSION

This paper proposes a novel unsupervised feature selection algorithm utilizing ensemble learning, named RFS_RFE, employing SOM variant AdSOM and GH_AdSOM. This method is elaborated upon extensively within this study. And significant improvements in clustering accuracy can be achieved using a highly constrained subset of features.

The combination of GH_AdSOM variants and RFS-RFE presents a dual-layered approach. GH_AdSOM variants are proficient in capturing intricate patterns within the data, thereby establishing a strong basis for clustering. On the other hand, RFS-RFE iteratively selects and eliminates features, guaranteeing the retention of only the most discriminative and informative attributes. This synergy enhances the overall performance of the clustering process by leveraging the strengths of both methodologies.

As future work, the proposed RFS-RFE framework can be extended to high-dimensional big data scenarios and real-world applications such as bioinformatics and image analysis. Further investigation could explore its integration with deep learning-based clustering models to enhance representation learning and scalability. Additionally, developing adaptive or automated parameter selection strategies may further improve robustness and generalization performance across diverse datasets.

6. REFERENCES

[1] Kaur, S., Kumar, Y., Koul, A., and Kumar Kamboj, S. 2023. A systematic review on metaheuristic optimization techniques for feature selections in disease diagnosis, open issues and challenges. Archives of Computational Methods in Engineering, 30(3), pp 1863-1895.

[2] Fan, Y., Liu, J., Tang, J., Liu, P., Lin, Y., and Du, Y. 2024. Learning correlation information for multi-label feature selection. Pattern Recognition, 145, 109899.

[3] Strehl, A., and Ghosh, J. 2002. Cluster ensembles-a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning.

[4] Dong, X., Yu, Z., Cao, W., Shi, Y., and Ma, Q. 2020. *A survey on ensemble learning*. Frontiers of Computer Science, 14, 241-258.

[5] Deng, S., Zhu, Y., Yu, Y., and Huang, X. 2024. An integrated approach of ensemble learning methods for stock index prediction using investor sentiments. Expert Systems with Applications, 238, 121710.

[6] Nie, F., Huang, H., Cai, X., and Ding, C. 2010. Efficient and robust feature selection via joint l2,1-norms minimization, in Proc. Adv. NIPS, pp. 1813–1821.

[7] Xu, Z., King, I., Lyu, M. R.-T and. Jin, R. 2010. Discriminative semisupervised feature selection via manifold regularization, IEEE Trans. Neural Netw.21(7) 1033–1047.

[8] Li, Z., Yang, Y., Liu, J., Zhou, X. and Lu, H. 2012. Unsupervised feature selection using nonnegative spectral analysis, inProc. Conf. AAAI, pp. 1026–1032.

[9] Fred, A. and Jain, A. 2008. Combining multiple clusterings using evidence accumulation, IEEE Trans. Pattern Anal. Mach. Intell.27(6) pp 835–850.

[10] Li, F., and Yang, Y. 2005. Analysis of recursive feature elimination methods. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval pp. 633-634.

[11] Nogales, R.E., Benalcázar, M.E. 2023. Analysis and Evaluation of Feature Selection and Feature Extraction Methods. Int J Comput Intell Syst 16, 153.

[12] Deschênes, T., Tohoundjona, F. W. E., Plante, P. L., Di Marzo, V., and Raymond, F. 2023. Gene-based microbiome representation enhances host phenotype classification. Msystems, 8(4), e00531-23.

[13] Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., and Lang, M. 2020. *Benchmark for filter methods for feature selection in high-dimensional classification data*. Computational Statistics & Data Analysis, 143, 106839.

[14] Liu, H., and Setiono, R. 2022. Feature selection and classification–a probabilistic wrapper approach. In Industrial and engineering applications or artificial intelligence and expert systems (pp. 419-424). CRC Press.

[15] Dudoit, S. and Fridlyand, J. 2003. Bagging to improve the accuracy of a clustering procedure, Bioinformatics, Vol. 19, No. 9, pp.1090–1099.

[16] Breiman, L. 2001. Random forests, Mach. Learn.45(1) 5–32.

[17] Guyon, I. and Elissee, A. 2003. An introduction to variable and feature selection (kernelmachines section). J. Mach. Learn. Res.31157–1182.

[18] Mundra, P. A., and Rajapakse, J. C. (2009). SVM-RFE with MRMR filter for gene selection. IEEE transactions on nanobioscience, 9(1), 31-37.

[19] Nakao, H., Imaoka, M., Hida, M., Imai, R., Nakamura, M., Matsumoto, K., and Kita, K. 2023. Determination of individual factors associated with hallux valgus using SVM-RFE. BMC Musculoskeletal Disorders, 24(1), 534.

[20] Dittenbach, M., Merkl, D., Rauber, A. 2000. The growing hierarchical self-organizing map, in proceedings of the International Joint Conference on Neural Networks (IJCNN), Vol. 6, pp 15-19.

[21] Kohonen, T. 2001. Self-organizing Maps, third edition, Springer.

[22] Meila, M. 2005. *Comparing clusterings: An axiomatic view*, inProc. 22nd Int. Conf. Machine Learning (ICML), Bonn, Germany, pp. 577–584.

[23] Filali, A., Jlassi, C., and Arous, N. 2017. Recursive Feature Elimination with Ensemble Learning Using SOM Variants. International Journal of Computational Intelligence and Applications (IJCIA), Vol. 16, No. 01, 1750004.