International Journal of Computing and Digital Systems
2025, VOL. 17, NO. 1, 1–8

http://dx.doi.org/10.12785/ijcds/1571111337

Hybrid Ensemble Feature Selection Using Symmetrical Uncertainty and Multi-Layer Perceptron

Swapnali N Tambe[1], Saiprasad Potharaju [2], Sagar B Tambe [3], Ravi Kumar Tirandasu [4], Shanmuk Srinivas Amiripalli[5] and Rajana Kanaka Raju[6]

[1] *Department of Information Technology, K. K.Wagh Institute of Engineering Education Research, Nashik, MH, India*
[2] *Department of CSE, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India*
[3] *Department of CSE, School of Computing, MIT Art, Design and Technology University, Pune, MH, India*
[4] *Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India*
[5] *Department of CSE, GST, GITAM University, Visakhapatnam, AP, India*
[6] *Department Of Computer Science and Engineering,Gayatri Vidya Parishad College For Degree and PG Courses (Autonomous),Visakhapatnam, AP, India*

Received 1 July 2024, Revised 30 January 2025, Accepted 31 January 2025

Abstract: Feature selection (FS) is an essential preprocessing step in Data Mining. In literature, there are few techniques that exist for FS. Still, there is a need to propose novel methods for the best FS in order to get better classification performance. In this research article, we tried to present a framework by which a cluster of features can be formed. Our framework forms ‘N’ clusters based on the user choice and Symmetrical Uncertainty(SU). Out of ‘N’ clusters, one dominant cluster is selected based on the result of Multi-Layer Perceptron(MLP) on each cluster. Each such cluster contains unique features in it. These features in the dominant cluster are tested with Jrip, J48, and K-Nearest Neighbour (KNN) classification algorithms with ensembling bagging and boosting techniques. Also, features derived by the proposed method are compared with few of the traditional filter-based techniques. The proposed method outperforms some of the traditional methods in the majority cases. This method is tested with a well-known dataset which consists of 60 features.

Keywords: Feature Selection, Symmetrical Uncertainty, Multi-Layer Perceptron, Ensembling

1. INTRODUCTION

Data Mining(DM) has been a booming area of research for many years, as it is the best technique for drawing more insights from large amounts of data collected from diverse sources. It can be applied in many fields like health care, sports, education, social media, marketing, etc [1]. DM is not at all a straightforward approach for drawing interest patterns from the collected data. After gathering the data from various sources(Web, survey, interviews, etc.) it needs to be preprocessed. In the preprocessing stage, there is a need for checking noise, outliers, imbalanced class labels, and high dimensionality). After this stage, Intelligent DM techniques(Regression, Classification, Clustering, Association Rule Mining) can be applied. 80 % of the total cost/time can be spent on addressing these preprocessing issues in the whole DM process [2]. In this paper, we focused mainly on presenting a framework for FS, which is concerned with the high dimensionality issue of preprocessing. Then applied some of the classification methods(Jrip, J48, KNN) with ensemble approaches like bagging and boosting.

A. Need of FS

FS has a significant role in DM for better classification results. Generally, if collected data has ‘N’ features, all those features are not required for classification model generation. Some of them may be highly correlated and few may be unnecessary [3]. It is always advisable to discard those unnecessary and duplicate features and select unique and strong features. For example, Date of birth and Age are correlated features. It is not required to select both these features. Because, from Date of Birth feature Age can be derived. Sometimes, we may have a serial number feature in the dataset, it is an unnecessary feature, which can be discarded . If all the features in a dataset are considered, what could be the problems ?. It consumes more memory for generating the model. It may deviate or divert the learning model because of those duplicate and unwanted features. Learning model performance may be decreased. Because of these reasons, it is advisable to select limited and strong features for better results [4].
This research proposes a Hybrid Ensemble Feature Selec-

E-mail address: snjagtap@kkwagh.edu.in, psaiprasadcse@gmail.com, sagar.tambe@mituniversity.edu.in, ravi.tirandasu@gmail.com,shanmuk39@gmail.com,rkanakaraju@gvpcdpgc.edu.in

tion Framework that integrates Symmetrical Uncertainty (SU) and Multi-Layer Perceptron (MLP) for optimal feature selection. The objective is to enhance classification accuracy by clustering features based on SU scores and selecting the dominant cluster using MLP evaluation. The contributions include a novel feature clustering method, performance validation using ensemble classifiers, and comparative analysis with traditional filter-based approaches, demonstrating superior classification accuracy

B. How to Select Best Features

In literature, there are two basic approaches called filter and wrapper available for selecting the best features. In Filter based approaches, algorithms like Symmetrical Uncertainty(SU), Chi-Square, Information Gain, Gain Ratio, etc are used. It gives the rank to each feature in the dataset. Depending on the working mechanism of those algorithms, rank may be varied by each technique [5]. As per the user choice top ranked 'N' features can be selected for model generation. For the proposed approach, we used SU for generating the rank of feature, and other techniques are used to compare the proposed method. Wrapper method is used to derive the subset of features based on the searching criteria(Backward search, Forward Search, Genetic Algorithm, etc. [6]) This approach is time consuming compared to the filter, in addition to these two feature selection methods also becoming popular recently. In this current work, we tried to focus on drawing features other than existing techniques. For testing the performance of features, we applied Jrip-Rule based, J48- Tree based, KNN- Lazy learner with an ensemble approach. Our aim of this research is to propose a new approach for FS. For this, we formed 'N' clusters of features such that each cluster was built with finite and unrepeated features. Procedure to form the cluster is discussed in the methodology section. It is difficult to compare each cluster performance with traditional methods. So, We utilized a Multi-Layer Perceptron (MLP) for each cluster to identify the most prominent one. By training an MLP on the data within each cluster, we evaluated their performance in terms of accuracy. The cluster that achieved the highest accuracy was designated as the dominant cluster, indicating its significance in the dataset.

2. RELATED WORK

In this section, some of the related theories which can connect to the proposed methodology is discussed. Our methodology is based on SU and MLP, testing of this methodology is using ensemble approaches, comparison is using Chi-Square, Information Gain, Gain Ratio. SU is a filter based FS technique used to award the rank to each feature. It was applied by many researchers in recent literature. FAST technique is proposed by the authors for FS, they have used SU as a primary criterion along with correlation coefficient for constructing a minimum spanning tree [7]. Other than SU, other filter-based techniques are used in the literature in their research work. ReliefF and Information Gain(IG) have been applied for oil spill detection. For their research, from the year 2007 to 2011 images are collected by the Envisat satellite. The initial dataset has 52 features, After applying IG and ReliefF, 15 top-ranked features were selected and Support Vector Machine is applied later for classification [8]. To identify prominent features in the clustered dataset, the authors proposed instance based feature selection, which is based on mutual information gain [9]. Authors investigated feature selection approach to reduce the computational overhead of using API calls as features for Android malware detection, finding that the number of API calls can be reduced by 95% while maintaining high accuracy, with random forests achieving the best performance at 96.1% accuracy [10].The classification task in microarray analysis is inherently complex and often requires a feature-selection process. This process is crucial for simplifying the feature space and identifying a subset of significant features. By selecting the most relevant features, the performance of the learning model can be enhanced, providing deeper insights into the underlying biological processes. The classification task in microarray analysis is inherently complex and often requires a feature-selection process. This process is crucial for simplifying the feature space and identifying a subset of significant features. By selecting the most relevant features, the performance of the learning model can be enhanced, providing deeper insights into the underlying biological processes [11][12]. Detailed survey of FS methods on classification is discussed by the authors in their article, they have presented all major FS techniques of filter and wrapper methods [13]. MLP is a popular classification technique applied for different reasons. Prediction of moving organs during the radiation therapy of liver and lung tumors is critical. For accurate prediction of moving organs MLP using boosting has been applied by researchers, and achieved 91.43 % accuracy as a result [14]. The Multi-Layer Perceptron (MLP) has been successfully applied to classify machine-controlled software. The proposed framework, which incorporates a class balancing technique, demonstrated strong performance across all the datasets used in the study. This approach ensures that the model is robust and effective, even when dealing with imbalanced data [15]. MLP was applied to know the trends of coal prices in China [16]. A modified bio-inspired MLP algorithm proposed by the researchers enhanced the efficacy of the IDS in detecting both normal and anomalous traffic in the network [17]. The authors proposed a deep learning-based system for educational user profiling and user rating recommendations in eLearning. This system adopts a hybrid approach, integrating collaborative filtering with deep learning techniques to deliver personalized and accurate recommendations. By leveraging the strengths of both methods, the system aims to enhance the learning experience by tailoring content and resources to the individual needs and preferences of each user [18]. Only proposing a framework is not sufficient. Its strength also needs to be tested. For this, we employed KNN, J48 and Jrip classifiers with ensembling boosting and bagging [19]. These methods have been considered by many researchers in their work for different reasons. Bagging and genetic algorithm(GA) was applied by the authors for intrusion detection systems.

In their research, out of 41 features, 15 relevant features were selected using GA, and C4.5 tree based algorithm was applied with bagging, with this they secured 99.71 % accuracy [20]. Boosting and Bagging methods for handling imbalanced datasets have been discussed by various researchers. These methods were applied on cardiac surgery dataset to improve the classification results [21]. The authors applied these techniques on a kidney disease dataset [22]. They applied various ensemble(bagging and boosting) learning techniques and found that the model template could minimize the problem of misclassification of imbalanced data. The researchers presented J48 Classifier for predictive analytics study to identify the most common diseases among university students in Selangor, Malaysia, using data mining techniques such as decision tree and rule induction [23]. The authors compared the performance of various discretization methods on decision tree(J48) and decision rule classifiers (Jrip), and found that discretization techniques can improve the performance of these classifiers [24].The study investigates how high dimensionality and imbalanced data affect predictive models for colon cancer detection. It evaluates tree-based classifiers, rule-based classifiers, lazy learning techniques like K-nearest neighbors (KNN), and support vector machines (SVMs) to improve performance [25]. For predicting white matter hyperintensities in Alzheimer's patients during magnetic resonance images scan, authors considered KNN, decision trees, boosting and bagging techniques for evaluating lung cancer and heart disease. The authors examine various algorithms, including decision trees, support vector machines, k-nearest neighbors, random forests, and neural networks, assessing their performance in diagnosing heart conditions [26].The authors explore the application of machine learning techniques to diagnose lung cancer using computed tomography (CT) images. The authors discuss how these techniques can enhance early detection by analyzing CT scans. They also address challenges such as feature extraction and model accuracy.[27].

3. PROPOSED METHODOLOGY

Our methodology is based on the assumption that, if there is a requirement of selecting 'N' features , which features have to be selected ?. For this, apply any filter based mechanism then find out the feature rank, then choose Top 'N' features as per its rank. In this current work, we have presented a new approach to select the features other than features derived by traditional filter methods.However, this current study is based on MLP and SU. SU

It is an important measurement derived by applying below statistical formulas. The measurement is nothing but a value assigned to each feature. The higher the score is the strongest feature and lowest score indicates the weaker feature. Basically SU score is used to know the relation or association between any two features or an association between features with its target variable. Based on the SU score the features will be selected for further classification.

SU score can be defined as below.

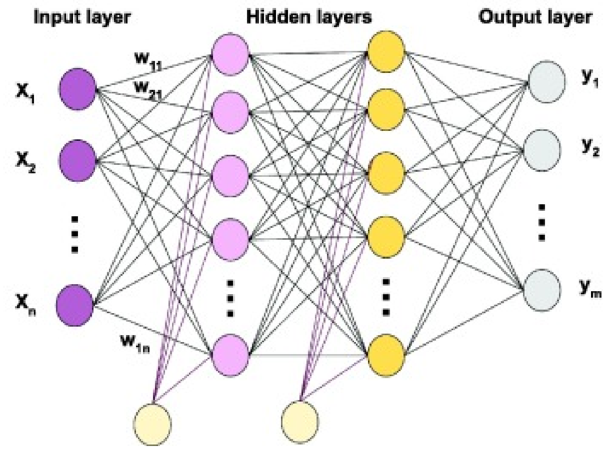

Figure 1. Three layered MLP example

SU= 2*Information Gain/ (H(A1)+H(A2))

H(A1) : Entropy of A1

H(A2) : Entropy of A2

The Symmetric Uncertainty (SU) value ranges from 0 to 1, where an SU value of 1 indicates that one feature can fully predict another. In contrast, an SU value of 0 signifies no correlation between the two features. In the proposed approach, features with an SU score of 0 are excluded from the final dataset, as they do not contribute to the predictive performance of the learning model.

A. *MLP*

In short, the Multi layer perceptron is called MLP. It is a class of feed forward artificial neural network. Basically it is a neural network. We know everything about the perceptron. Perceptron is a single unit, if We combine these perceptrons to perform the complicated task that is called a neural network or multi layer perceptron. So in MLP, multiple layers of perceptrons will be placed . In the multi layer perceptron there can be more than one linear layer which are combined together. In this there will be one input layer and one output layer in between these that will be thousands of hidden layers. If we take the simple example of the three layered network, the first layer will be the input layer, the last layer will be the output layer and the middle layer will be the hidden layer. These hidden layers can be extended depending on the problem statement. If we want a more accurate model we can place multiple hidden players. If we have more hidden layers, a lot of competitions are required to perform the calculations. An example of the same is given in Figure 1 with calculations and formulas. As proposed work, split the primary space into a set of clusters(groups) to guess the strong cluster , The MLP is applied initially on each cluster formed by the current approach . The proposed method is inspired from the ensemble approaches, in which two or more classifiers can be clubbed for the classification, so that weak learners can get useful knowledge from the strong one, and overall strength can become strong. In the similar fashion, instead of selecting all top features, we have mixed the strong, medium, and weak features in an organized manner, so that we could form the proper clusters.

Based on the above concepts our methodology is proposed, which is articulated below.

B. Proposed Algorithm

Input: DBL, G, S, LTF
DBL: Balanced Dataset
G: Groups or Cluster to be generated
S: The count of features with SU value greater than 0
L : List of features with SU value greater than 0

Output: MF=a1,a2,...an (Minimized Feature set)

Step 1: Imbalance Check, if the initial data frame is not balanced , get the DBL after employing SMOTE.

Step 2: Get the Count of S. Apply SU on DBL to get S, then arrange them in L as per its score in such a way that the highest Score feature will be Positioned first.

Step 3: Get MF, in such a way that,
a. Place the first or next 'G' features or attributes from list L in a left-to-right direction, so that the first feature is inserted into cluster number 1, the second feature into cluster number 2, and so on. Continue this process by reading the next 'G' features from list L.
b. Place the first or next 'G' features or attributes from the list LL in a right-to-left manner, such that the first feature is assigned to the last cluster, the second feature to the second-last cluster, and so forth, continuing this pattern sequentially.

Step 4: Repeat step 3(a), followed by step 3(b), iteratively until all features are assigned to their respective clusters.

Step 5: Merging the features into various clusters, merge the all vertically first level attributes or features into first group or cluster (c1), second level attributes or features into second group or cluster (c2), and so on till the last group or cluster (cn).

Step 6: Check the cluster cardinality. Calculate the number of features formed in each cluster. If any cluster or group has any extra features, discard them from the group to maintain the cluster balance.

Step 7: Decide the best cluster. For this, apply MLP on each balanced cluster of features. Based on the highest accuracy given by a strong cluster will be decided.

Step 8: Get the topmost 'N' features from the balanced dataset after applying the filter based methods. Here N is the number of features available in a strong cluster derived by the proposed method.

Step 9: Compare the performance of proposed methods with existing approaches with classifiers.

C. Example

For example there are 13 attributes or variables in the original balanced dataset. Assume S (The count of features with SU value greater than 0) is 11. G: Groups or Cluster to be generated is 3 L : List of features with SU value greater than 0 are [v1, v2, v3, v4, v5, v6, v7, v8, v9, v10, v11]

According to the proposed method the features are grouped in various clusters as given in Table I.

4. EXPERIMENT

The proposed approach is tested with the SONAR dataset, which is collected from a popular UCI machine learning repository. Initial SONAR dataset has 60 features and 2 classes (Rock and Mine), 208 records. Rock has 97 records and Mine has 111 records. Initial dataset is a little imbalanced, In order to get the DBL that is a balanced dataset employ the SMOTE. As, SMOTE is on the basis of K-Nearest Neighbour, for balancing the dataset K=5 is considered. After balancing, 218 instances are generated. In the Balanced set, Rock has 107 records, and Mine has 111 records. After this, Symmetrical Uncertainty, which one of the core components in this contribution is applied on the balances dataset (DBL) then recorded the S which is the total number of features whose SU score is greater than zero. Below, Table II provides the information scores of each feature as derived from various traditional filter methods, including Symmetrical Uncertainty (SU).

The Column SU,CH,IG, GR has the feature number of the dataset. To select 'N' strong features, filter-based methods can be applied to the dataset. These methods rank the features based on their relevance, allowing the top 'N' features to be chosen for building a classification model. As per the proposed algorithm, a remaining process such as forming the clusters and balancing the cluster is performed. In order to test the performance of the proposed method the features are formed with the 2, 3 and 4 clusters. Then out of those clusters to decide the best MLP is applied . As per the accuracy given by the MLP , the best cluster is decided. The accuracy with those clusters of features is given in Table III. From Table II, we can understand the S =25.

#G: Number of groups or Clusters
#Gid: Group or Cluster ID
#N: Size of each cluster or group

For the further analysis of the proposed method, Top 'N' number of features derived by the methods listed in Table 2 are considered. For example, to test the features

TABLE I. Cluster of features

$I^{st} Order(Group - C1)$	$2^{nd} Order(Group - C2)$	$3^{rd} Order(Group - C3)$	Direction of Feature Placement
v1	v2	v3	LR
v6	v5	v4	RL
v7	v8	v9	LR
	v11	v10	RL

TABLE II. Scores assigned to each feature by various methods, including Symmetric Uncertainty (SU)

Rank	SU Score	SU	IG	CHI	GR
1	0.2242	11	11	11	11
2	0.2007	12	12	12	12
3	0.1636	9	9	9	58
4	0.1518	10	10	10	44
5	0.1370	13	13	13	9
6	0.1167	45	48	48	54
7	0.1153	48	49	49	45
8	0.1116	44	45	52	13
9	0.1109	49	52	51	10
10	0.1006	54	51	47	2
11	0.0983	47	47	21	28
12	0.0973	28	21	4	48
13	0.0912	52	4	45	49
14	0.0907	51	44	5	47
15	0.0880	4	28	28	5
16	0.0867	5	5	36	52
17	0.0858	21	36	20	51
18	0.0774	36	54	46	4
19	0.0758	2	46	44	21
20	0.0749	46	20	8	36
21	0.0729	58	8	54	46
22	0.0712	20	43	1	20
23	0.0655	8	1	43	43
24	0.0636	43	2	2	8
25	0.0604	1	58	58	1

TABLE III. Various features formed by the proposed algorithm

G	Gid	N	Features in it	Best Cluster (Accuracy)
2	G21	12	11, 10, 13, 44, 49, 28, 52, 5, 21, 46, 58, 43	G21 (82.56)
	G22	12	9, 45, 48, 54, 47, 51, 4, 36, 2, 20, 8	
3	G31	8	11, 45, 48, 28, 52, 36, 2, 43	G31 (81.65)
	G32	8	12, 13, 44, 47, 51, 21, 46, 8	
	G33	8	9, 10, 49, 54, 4, 5, 58, 20	
4	G41	6	11, 44, 49, 5, 21, 43	G41 (74.77)
	G42	6	12, 48, 54, 4, 36, 8	
	G43	6	9, 45, 47, 51, 2, 20	
	G44	6	10, 13, 28, 52, 46, 58	

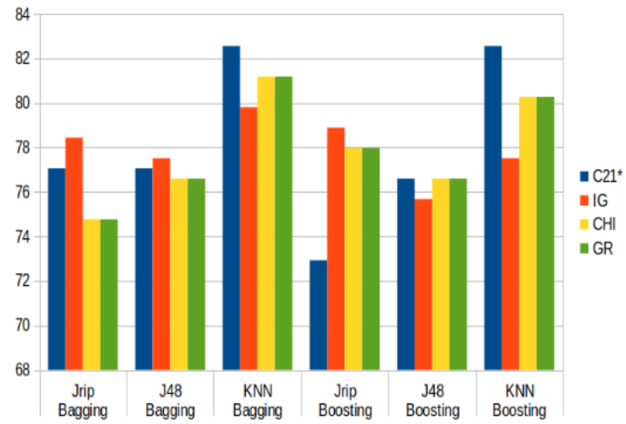

Figure 2. Performance with 2 clusters.

formed with G=2, top 12 features of the existing method are considered. Similarly for G=3, top 8 and for G=4, top 6 features of the existing method are considered. The strength of each is tested with ensembling techniques like boosting and bagging. For the ensemble, KNN, Jrip, J48 classifiers are considered. The same is implemented with python as well WEKA with default setting. The results of WEKA are considered in this paper.

5. RESULTS

This section presents the implementation of various groups or clusters of features generated by the proposed approach, alongside the top 'N' features selected using existing filter methods combined with ensembling techniques. The results for clusters of sizes 2, 3, and 4 are provided in Table 4, Table 5, and Table 6, respectively. These tables highlight the effectiveness of the proposed feature grouping method in comparison to traditional filter-based methods.

From the Table IV we can interpret that, the second cluster of features (G21) produced better results than existing CHI, GR and IG when bagging is applied with Jrip, J48 and KNN. The same results are good when compared with CHI and GR. The same is true when Boosting is applied with J48 and KNN. Visualization of this result analysis can be found in Figure 2.

From Table V, it is evident that the proposed method achieved superior performance compared to existing methods across all cases when bagging and boosting techniques were applied with various classifiers. A visual representation of this performance analysis is provided in Figure 3. Bagging+ Jrip secured 77.98% which is higher than all existing methods. Boosting + Jrip produced 81.65 which is also higher than all. The remaining results can also be interpreted in the same way as per the Table V.

From Table VI , we can observe that the features formed

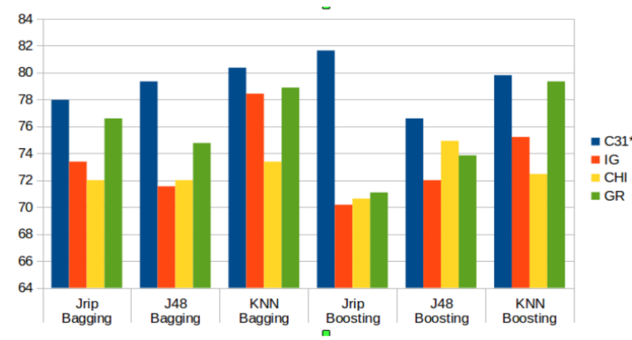

Figure 3. Performance with 3 clusters.

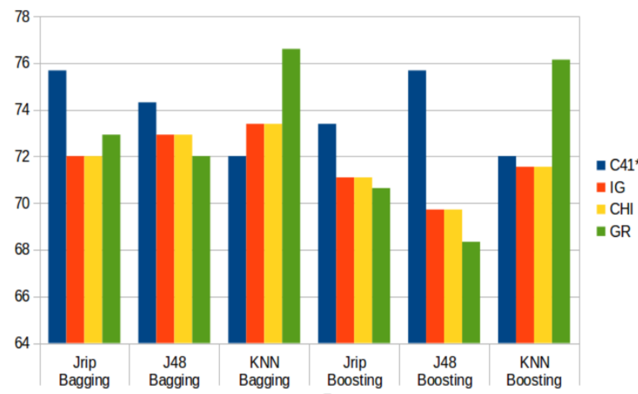

Figure 4. Performance with 4 clusters.

by the proposed method (G41) have recorded the best performance compared to existing methods when using Jrip and J48 classifiers with bagging and boosting techniques. Visualization of this result analysis can be found in Figure 4.

* The cluster of features formed by proposed method

Here instead of MLP any other strong classifiers can be applied on each cluster to evaluate its strength.

A. Discussion

The novelty of the approach lies in integrating Symmetrical Uncertainty (SU) for ranking features and using ensemble techniques like bagging and boosting to validate cluster effectiveness. These steps provide a systematic and reproducible method for feature selection.

The proposed methodology of merging strong, medium, and weak features into organized clusters is a deliberate approach inspired by ensemble techniques, where weaker elements can enhance overall performance when combined with stronger ones. This method preserves the potential interactions between features, as strong features may overshadow valuable weak features if considered independently. By evaluating these clusters using Multi-Layer Perceptron (MLP), the framework identifies the combination of features that yields the highest accuracy, ensuring that only the

TABLE IV. The performance of the proposed approach using two clusters

CID	Ensemble Bagging			Ensemble Boosting		
	KNN	J48	Jrip	KNN	J48	Jrip
G21*	82.56	77.06	77.06	82.56	76.60	72.93
IG	79.81	77.52	78.44	77.52	75.68	78.89
CHI	81.19	76.60	74.77	80.27	76.60	77.98
GR	81.19	76.60	74.77	80.27	76.60	77.98

TABLE V. The performance of the proposed approach using Three clusters

CID	Ensemble Bagging			Ensemble Boosting		
	KNN	J48	Jrip	KNN	J48	Jrip
G31*	80.37	79.35	77.98	79.81	76.60	81.65
IG	78.44	71.55	73.39	75.22	72.01	70.18
CHI	73.39	72.01	72.01	72.47	74.93	70.64
GR	78.89	74.77	76.60	79.35	73.85	71.10

TABLE VI. The performance of the proposed approach using Four clusters

CID	Ensemble Bagging			Ensemble Boosting		
	KNN	J48	Jrip	KNN	J48	Jrip
G41*	72.01	74.31	75.68	72.01	75.68	73.39
IG	73.39	72.93	72.01	71.55	69.72	71.10
CHI	73.39	72.93	72.01	71.55	69.72	71.10
GR	76.60	72.01	72.93	76.14	68.34	70.64

most effective group is selected. This approach provides flexibility, as weak features are naturally excluded from the dominant cluster if they do not contribute positively to predictive performance. Empirical results from the study demonstrate that this method outperforms traditional filter-based techniques, highlighting the advantage of considering diverse feature combinations rather than relying solely on strong features. Using 2, 3, and 4 clusters provides a systematic way to evaluate and leverage feature interactions. These clusters balance feature diversity and model complexity, ensuring effective noise reduction, enhanced generalization, and data-driven feature selection. By combining empirical validation with theoretical rationale, the approach demonstrates that clustering is a critical component of its success. The findings highlight that integrating Symmetrical Uncertainty (SU) with Multi-Layer Perceptron (MLP) enhances feature selection efficiency, outperforming traditional methods. The proposed clustering approach preserves feature diversity, improving classification accuracy across different ensemble classifiers. However, the study is limited to a single dataset (SONAR), and results may vary with high-dimensional and real-time data. Future research could explore adaptive feature selection, deep learning integration, and applications in diverse domains like medical diagnostics and cybersecurity.

6. Conclusions and Future Work

In this research article, we introduce a novel feature selection framework titled ”Symmetrical Uncertainty (SU) and Multi-Layer Perceptron (MLP)-Based Feature Selection Framework: An Ensembling Approach.” The framework centers on Symmetrical Uncertainty and Multi-Layer Perceptron as key measures of feature relevance. Features are grouped into clusters using ensemble techniques, and MLP is utilized to identify the optimal cluster based on the highest achieved accuracy. We compared the selected cluster of features with fee filter-based feature selection methods. To test the performance of our proposed method, we considered bagging and boosting ensembles with various classifiers, such as J48, JRip, and KNN. In the majority of cases, our proposed method outperformed existing methods. We tested our method using the SONAR dataset with 50%, 33%, and 25% feature sizes, as well as multiple dimensions on various datasets. Our proposed method produced significant results in those cases as well.

References

[1] E. N. Ekwonwune, C. I. Ubochi, and A. E. Duroha, “Data mining as a technique for healthcare approach,” *International Journal of Communications, Network and System Sciences*, vol. 15, no. 9, pp. 149–165, 2022.

[2] S. M. Dol and P. M. Jawandhiya, “Classification technique and its combination with clustering and association rule mining—a survey,” *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106071, 2023.

[3] N. Krishnaveni and V. Radha, “Feature selection algorithms for data mining classification: a survey,” *Indian Journal of Science and Technology*, vol. 12, no. 6, 2019.

[4] N. S. ElZawawi, H. G. Saber, M. Hashem, and T. Gharib, “An efficient hybrid approach for diagnosis high dimensional data for alzheimer’s diseases using machine learning algorithms,” *Interna-*

tional Journal of Intelligent Computing and Information Sciences, vol. 22, no. 2, pp. 97–111, 2022.

[5] A. Sikri, N. P. Singh, and S. Dalal, "Analysis of rank aggregation techniques for rank based on the feature selection technique," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, pp. 95–108, 2023.

[6] R. Jain and W. Xu, "Artificial intelligence based wrapper for high dimensional feature selection," *BMC Bioinformatics*, vol. 24, pp. 392–399, 2023.

[7] R. S. Epstein E, Nallapareddy N, "On the relationship between feature selection metrics and accuracy," *Entropy*, vol. 25, no. 12, pp. 16–46, 2023.

[8] Q. Song, J. Ni, and G. Wang, "A fast clustering-based feature subset selection algorithm for high-dimensional data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 1–14, 2023.

[9] D. Mera, V. Bolon-Canedo, J. M. Cotos, and A. Alonso-Betanzos, "On the use of feature selection to improve the detection of sea oil spills in sar images," *Computers & Geosciences*, vol. 100, pp. 166–178, 2017.

[10] B. Ahmad Mantoo and N. Z. Ali Khan, "Impact of feature selection algorithms in detecting android malware using machine learning over permissions and api's," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 189–200, 2024.

[11] M. Zaky Hakim Akmal and D. Fitrianah, "Exploring feature selection for microarray classification," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1–10, 2024.

[12] P. K. Ram and P. Kuila, "Feature selection from microarray data: Genetic algorithm based approach," *Journal of Information and Optimization Sciences*, vol. 40, no. 8, pp. 1599–1610, 2019.

[13] A. A. Khtoom and M. Wedyan, "Feature selection models for data classification: Wrapper model vs filter model," in *Intelligent Computing Paradigm and Cutting-edge Technologies: Proceedings of the First International Conference on Innovative Computing and Cutting-edge Technologies (ICICCT 2019)*. Springer International Publishing, 2020, pp. 247–257.

[14] W. Sun *et al.*, "Respiratory signal prediction based on adaptive boosting and multilayer perceptron neural network," *International Journal of Radiation Oncology Biology Physics*, vol. 96, no. 2, p. E702, 2017.

[15] A. Iqbal and S. Aftab, "A classification framework for software defect prediction using multi-filter feature selection technique and mlp," *International Journal of Modern Education & Computer Science*, vol. 12, no. 1, 2020.

[16] X. Xu and Y. Zhang, "Thermal coal price forecasting via the neural network," *Intelligent Systems with Applications*, vol. 14, p. 200084, 2022.

[17] W. Laftah Al-Yaseen and Q. Abdullah Abed, "Using a grey wolf optimization and multilayer perceptron algorithms for an anomaly-based intrusion detection system," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1–10, 2024.

[18] P. V. Kulkarni, S. Rai, R. Sachdeo, and R. Kale, "Deep learning-based educational user profile and user rating recommendation system for e-learning," *Journal of Information Systems and Telecommunication (JIST)*, vol. 3, no. 43, p. 185, 2023.

[19] M. Loukili, F. Messaoudi, and R. El Youbi, "Implementation of machine learning algorithms for customer churn prediction," *Journal of Information Systems and Telecommunication (JIST)*, vol. 3, no. 43, p. 196, 2023.

[20] D. P. Gaikwad and R. C. Thool, "Intrusion detection system using bagging with partial decision treebase classifier," *Procedia Computer Science*, vol. 49, pp. 92–98, 2015.

[21] P. Kaur and A. Gosain, "Empirical assessment of ensemble based approaches to classify imbalanced data in binary classification," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, 2019.

[22] S. P. Potharaju and M. Sreedevi, "Ensembled rule based classification algorithms for predicting imbalanced kidney disease data," *Journal of Engineering Science and Technology Review*, vol. 9, no. 5, pp. 201–207, 2016.

[23] M. A. Aziz, A. Jasri, M. Shamsudin, R. Maskat, N. Noordin, and M. Ninggal, "Predicting common diseases among students using decision tree (j48) classification algorithm," *International Journal of Academic Research in Business and Social Sciences*, vol. 11, no. 9, pp. 469–478, 2021.

[24] Y. Kaya and R. Tekin, "Comparison of discretization methods for classifier decision trees and decision rules on medical data sets," *European Journal of Science and Technology*, no. 35, pp. 275–281, 2022.

[25] S. N. Tambe, S. Potharaju, S. S. Amiripalli, R. K. Tirandasu, and Y. Algat, "Exploring the effects of high dimensionality and imbalanced data on predictive models for colon cancer detection using tree, rule, and lazy learning techniques," *Journal of Artificial Intelligence and Technology*, vol. 4, no. 4, pp. 342–350, 2024.

[26] R. Vinoth, T. Vijayan, V. Ganesan, and B. Kalaiselvi, "Computed tomography images for the diagnosis of lung cancer through machine learning," *Computer Science Engineering*, vol. 4, pp. 327–334, 2024.

[27] P. Rani, R. Kumar, A. Jain, R. Lamba, R. K. Sachdeva, K. Kumar, and M. Kumar, "An extensive review of machine learning and deep learning techniques on heart disease classification and prediction," *Archives of Computational Methods in Engineering*, vol. 31, p. 3331–3349, 2024.