International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

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

Dr. Swapnali N Jagtap[1], Dr. Saiprasad Potharaju [2], Dr. Sagar B Tambe [3], Dr. Shanmuk Srinivas Amiripalli [4],

[1] Department of Information Technology, K. K.Wagh Institute of Engineering Education & Research, Nashik, MH, India
[2,3] Department of CSE, School of Computing, MIT Art, Design and Technology University, Pune, MH, India
[4] Department of CSE, GST, GITAM University, Visakhapatnam, AP, India

E-mail address: snjagtap@kkwagh.edu.in, saiprasad.potharaju@mituniversity.edu.in, sagar.tambe@mituniversity.edu.in , samiripa@gitam.edu

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Feature selection (FS) is a crucial preprocessing step in Data Mining, aimed at enhancing classification performance by identifying the most relevant features. While numerous techniques for FS exist in the literature, there remains a continuous need to develop novel methods to achieve superior results. This research article introduces a novel framework designed to form clusters of features based on user choice and Symmetrical Uncertainty (SU). The framework creates 'N' clusters, from which one dominant cluster is selected based on the performance of a Multi-Layer Perceptron (MLP) applied to each cluster. Each cluster contains a unique set of features. The dominant cluster's features are then evaluated using Jrip, J48, and K-Nearest Neighbour (KNN) classification algorithms, combined with ensembling methods like bagging and boosting. In the proposed methodology features are grouped in 2 clusters, 3 clusters and 4 clusters. Additionally, the features identified by the proposed method are compared against those derived from traditional filter-based techniques. The proposed method demonstrates superior performance in most cases. The effectiveness of this method is validated using a well-known dataset comprising 60 features, highlighting its potential to outperform conventional FS techniques. This innovative approach addresses the ongoing demand for effective FS methods, contributing to improved classification accuracy and efficiency in data mining tasks.

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

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 the various sources(Web, survey, interviews, etc.) it needs to be preprocessed. In the preprocessing stage, there is a need for checking noisy, outliers, imbalanced class labels, high dimensionality). After this stage, Intelligent DM techniques(Regression, Classification, Clustering, Association Rule Mining) can be applied. 80 % of 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 comes under the high dimensionality issue of preprocessing. Then applied some of the classification methods(Jrip, J48, KNN) with ensemble approaches like bagging and boosting.

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

E-mail:author's email

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].

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 filter. In addition to these two methods of feature selection 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 THEORY

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 criteria along with correlation coefficient for constructing minimum spanning tree[7].Other than SU, in literature other filter based techniques are used 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. 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, 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 of microarray analysis is highly complex and typically necessitates the application of a feature-selection process. This process is essential for reducing the complexity of the feature space and identifying a subset of distinctive features. By selecting the most relevant features, we can improve the performance of the learning model and gain better insights into the underlying biological processes[11]. 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[12].

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[13]. The Multi-Layer Perceptron (MLP) has been successfully applied for classifying 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[14]. MLP was applied to know the trends of coal prices in China[15]. 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[16]. The authors proposed a deep learning-based educational user profile and user rating recommendation system for eLearning. This system uses a hybrid approach that combines collaborative filtering with deep learning techniques to provide 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.17]

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[18]. 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[19].

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[20]. The authors applied these techniques on a kidney disease dataset[21]. 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[22]. 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[23]. 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 Alzheimer's disease dataset[24].

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: Symmetric Uncertainty (SU) and Multi Layer Perceptron (MLP).

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.

$SU= 2*\text{Information Gain}/ (H(A_1)+H(A_2))$

$H(A_1)$: Entropy of A_1

$H(A_2)$: Entropy of A_2

The Symmetrical Uncertainty (SU) value ranges from 0 to 1. An SU value of 1 indicates that one feature can completely predict another, while an SU value of 0 indicates that the two features are uncorrelated. In our proposed approach, any feature with an SU score of 0 is discarded from the final dataset, as it does not contribute to the predictive power of the learning model.

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.

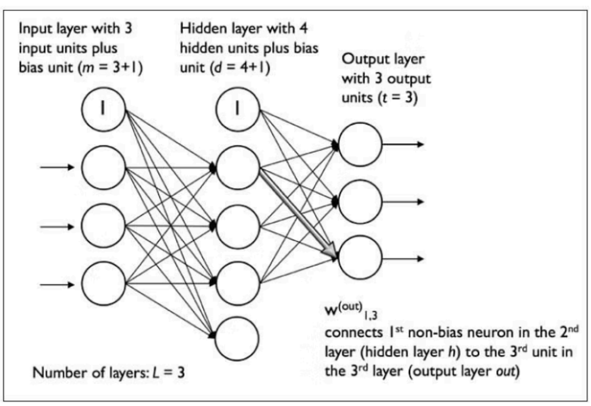

Figure 1. Three layered MLP example

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.

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={a_1,a_2,...a_n} (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 list L in a right-to-left direction, so that the first feature is inserted into the last cluster, the second feature into the second-last cluster, and so on.

Step 4. Repeat the step 3 (a) then 3 (b) until all features are placed.
Step 5. Merging the features into various clusters, merge the all vertically first level attributes or features into first group or cluster (c_1), second level attributes or features into second group or cluster (c_2), and so on till the last group or cluster (c_n).
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.

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 1.

Table 1. Cluster of features

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

LR : Left to Right, RL: Right to Left.
Group one C1 : { v1, v6, v7 }, Group Two C2: { v2, v5, v8 }, Group V3: { v3, v4, v9 }
Note : V10 and V11 are discarded per step 6 to maintain the cluster cardinality.

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 2 provides the information scores of each feature as derived from various traditional filter methods, including Symmetrical Uncertainty (SU).

Table 2 . The Score of each feature given by various methods including 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	0.137	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	0.088	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

The Column SU,CH,IG, GR has the feature number of the dataset.

If there is a need of selecting 'N' strong features, generally any of the filter based methods can be employed on the dataset, and top 'N' features can be chosen for creating a learning model for classification.

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 3. From table 2, we can understand the S =25.

Table 3. 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		12, 13, 44, 47, 51, 21, 46, 8	
	G33		9, 10, 49, 54, 4, 5, 58, 20	
4	**G41**	6	11, 44, 49, 5, 21, 43	G41 (74.77)
	G42		12, 48, 54, 4, 36, 8	
	G43		9, 45, 47, 51, 2, 20,	
	G44		10, 13, 28, 52, 46, 58	

#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 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.

4. Reseuts

In this section, we discussed the implementation of various groups or clusters of features created by the proposed approach and the top 'N' features drawn by existing filter methods using ensembling techniques. The results for clusters with sizes 2, 3, and 4 are detailed in Table 4, Table 5, and Table 6, respectively. These tables illustrate the effectiveness of our proposed feature grouping method compared to traditional filter methods.

Table 4. Performance with Two clusters.

CID	Bagging			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

From the above table 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.

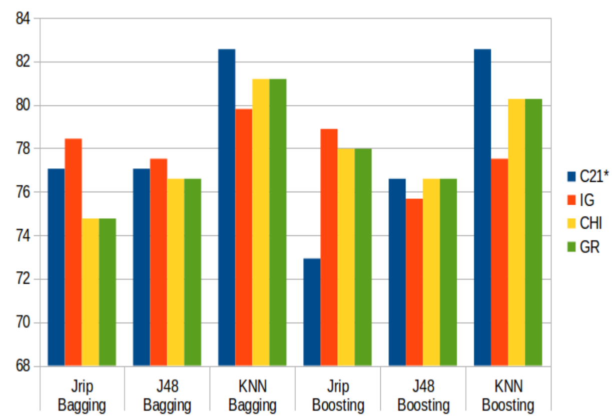

Fig 2. Performance with 2 clusters.

From the below Table 5, we can understand that the proposed method recorded the best performance than existing methods in all cases when applied bagging and boosting with all classifiers. Visualization of the same result analysis can be found 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 5.

Table 5. Performance with Three clusters..

CID	Bagging			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	71.1

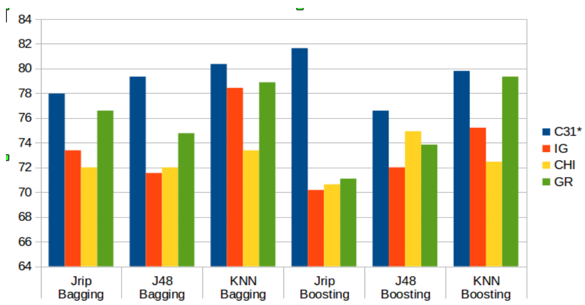

Fig 3. Performance with 3 clusters.

From Table 6 below, we can observe that the features formed 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.

Table 6. Performance with Four clusters.

CID	Bagging			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	71.1
CHI	73.39	72.93	72.01	71.55	69.72	71.1
GR	76.6	72.01	72.93	76.14	68.34	70.64

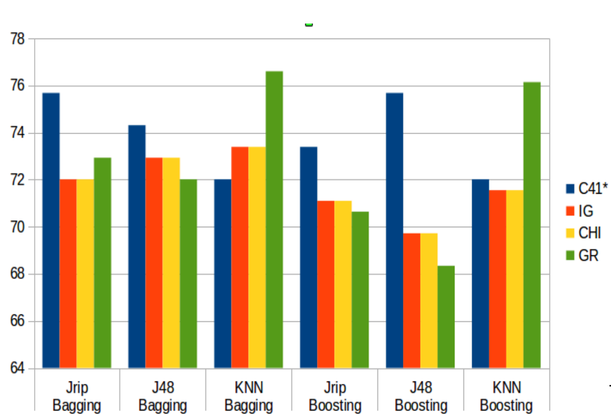

Fig 4. Performance with 4 cluster

* 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.

5. Conclusion

In this research article, we propose a novel feature selection framework titled "Symmetrical Uncertainty (SU) and Multi-Layer Perceptron (MLP)-Based Feature Selection Framework: An Ensembling Approach". The core of this paper is Symmetrical Uncertainty and Multi-Layer Perceptron, which are measures of feature strength. We formulated the features into various clusters using ensemble techniques and employed MLP to nominate the best cluster based on its highest 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. Ekwonwune, E. N., Ubochi, C. I., & Duroha, A. E. (2022). Data Mining as a Technique for Healthcare Approach. *International Journal of Communications, Network and System Sciences*, *15*(9), 149-165.
2. Dol, S. M., & Jawandhiya, P. M. (2023). Classification technique and its combination with clustering and association rule mining in educational data mining—A survey. *Engineering Applications of Artificial Intelligence*, *122*, 106071.
3. Krishnaveni, N., & Radha, V. (2019). Feature selection algorithms for data mining classification: a survey. *Indian J Sci Technol*, *12*(6).
4. ElZawawi, N. S., Saber, H. G., Hashem, M., & Gharib, T. (2022). An efficient Hybrid approach for diagnosis High dimensional data for Alzheimer's diseases Using Machine Learning algorithms. *International Journal of Intelligent Computing and Information Sciences*, *22*(2), 97-111.
5. Sikri, A., Singh, N. P., & Dalal, S. (2023). Analysis of Rank Aggregation Techniques for Rank Based on the Feature Selection Technique. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*, 95-108.
6. Saqib, P., Qamar, U., Aslam, A., & Ahmad, A. (2019). Hybrid of filters and genetic algorithm-random forests based wrapper approach for feature selection and prediction. In *Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 2* (pp. 190-199). Springer International Publishing.
7. Malji, P., & Sakhare, S. (2017, January). Significance of entropy correlation coefficient over symmetric uncertainty on FAST clustering feature selection algorithm. In *2017 11th International Conference on Intelligent Systems and Control (ISCO)* (pp. 457-463). IEEE.
8. Song, Q., Ni, J. and Wang, G., 2023. A fast clustering-based feature subset selection algorithm for high-dimensional data. IEEE transactions on knowledge and data engineering, 25(1), pp.1-14.
9. Mera, D., Bolon-Canedo, V., Cotos, J.M. and Alonso-Betanzos, A., 2017. On the use of feature selection to improve the detection of sea oil spills in SAR images. Computers & Geosciences, 100, pp.166-178.
10. Ahmad Mantoo, B., & Ali Khan N, Z. (2024). Impact of Feature Selection Algorithms in Detecting Android Malware Using Machine Learning Over Permissions and API's. *International Journal of Computing and Digital Systems*, *16*(1), 189-200.
11. Zaky Hakim Akmal, M., & Fitrianah, D. (2024). Exploring Feature Selection for Microarray Classification. *International Journal of Computing and Digital Systems*, *16*(1), 1-10.
12. Khtoom, A. A., & Wedyan, M. (2020). 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), Istanbul, Turkey, October 30-31, 2019 1* (pp. 247-257). Springer International Publishing.
13. Sun, W. et al.2017, Respiratory Signal Prediction Based on Adaptive Boosting and Multilayer Perceptron Neural Network, International Journal of Radiation Oncology • Biology • Physics , Volume 96 , Issue 2 , E702
14. Iqbal, A., & Aftab, S. (2020). A Classification Framework for Software Defect Prediction Using Multi-filter Feature Selection Technique and MLP. *International Journal of Modern Education & Computer Science*, *12*(1).
15. Xu, X., & Zhang, Y. (2022). Thermal coal price forecasting via the neural network. *Intelligent Systems with Applications*, *14*, 200084.
16. *Laftah Al-Yaseen, W., & Abdullah Abed, Q. (2024). Using a Grey Wolf Optimization and Multilayer Perceptron Algorithms for an Anomaly-Based Intrusion Detection System. International Journal of Computing and Digital Systems, 16(1), 1-10.*
17. *Kulkarni, P. V., Rai, S., Sachdeo, R., & Kale, R. (2023). Deep Learning-based Educational User Profile and User Rating Recommendation System for E-Learning. Journal of Information Systems and Telecommunication (JIST), 3(43), 185.*
18. *Loukili, M., Messaoudi, F., & El Youbi, R. (2023). Implementation of Machine Learning Algorithms for Customer Churn Prediction. Journal of Information Systems and Telecommunication (JIST), 3(43), 196.*
19. Gaikwad, D.P. and Thool, R.C., 2015. Intrusion detection system using bagging with partial decision treebase classifier. Procedia Computer Science, 49, pp.92-98.
20. Kaur, P., & Gosain, A. (2019). Empirical assessment of ensemble based approaches to classify imbalanced data in binary classification. *International Journal of Advanced Computer Science and Applications*, *10*(3).
21. Potharaju, S.P. and Sreedevi, M., 2016. Ensembled Rule Based Classification Algorithms for predicting Imbalanced Kidney Disease Data. Journal of Engineering Science and Technology Review, 9(5), pp.201-207.
22. Aziz, M. A., Jasri, A., Shamsudin, M., Maskat, R., Noordin, N., & Ninggal, M. (2021). Predicting common diseases among students using decision tree (j48) classification algorithm. *International Journal of Academic Research in Business and Social Sciences*, *11*(9), 469-478.
23. Kaya, Y., & Tekin, R. (2022). Comparison of discretization methods for classifier decision trees and decision rules on medical data sets. *European Journal of Science and Technology*, (35), 275-281..
24. Mahsa Dadar,Josefina Maranzano,Karen Misquitta,Cassandra J, et al. 2017, Performance comparison of 10 different classification techniques in segmenting white matter hyperintensities in aging. NeuroImage, Elsevier, Vol: 157, Page: 233-249

Dr. Swapnali N. Jagtap is working as an Assistant Professor at KK Wagh Institute, Nashik, India, in the Department of Information Technology. She received her PhD degree in the field of Computer Science and Engineering from SJJTU, Rajasthan. Her areas of research include Blockchain technology and Machine Learning. She has published various research articles in reputed journals.

Dr. Saiprasad Potharaju is working as an Associate Professor at MIT ADT University, Pune, MH, India. He received his PhD degree in the field of Computer Science and Engineering from K L University, AP. His areas of interest include Data Mining, Cyber Security, Cloud Computing, and Automation. He is the author of a book titled "Introduction to Cyber Security". He has published various research articles in reputed journals, and he has 4 patents and 1 copyright.

Dr. Sagar B. Tambe is working as an Associate Professor in the Department of Computer Science Engineering at MIT Art, Design & Technology University, Loni Kalbhore, Maharashtra, India. He completed a Postdoctoral from University of Parma, Italy & Ph.D from S.G.G.S. Institute of Engineering and Technology, Nanded. He received his B.E. and M.E. degree in Computer Engineering from Pune University, Pune in 2010 and 2013, respectively. He published many papers in reputed conferences and journals. His research interests include Machine Learning, Internet of Things, Signal Processing, Bio-medical sensor network, Wireless Communication.

Dr. A. Shanmuk Srinivas is working as an Assistant Professor at GITAM School of Technology, GITAM University, Visakhapatnam. He has done his Ph.D. in the field of IoT from K L University, Andhra Pradesh. He has more than 15 years of teaching experience. He has published 26 research papers in internationally reputed journals. His main areas of interest include Graph Analytics and Visualization, Cyber Security, Machine Learning, and Artificial Intelligence.