Ensemble Feature Selection for Heart Disease Classification

Houda Benhar[1], Ali Idri[1,2] and Mohamed Hosni[1,3]
[1]*Software Project Management Research Team, ENSIAS, Mohammed V University, Rabat, Morocco*
[2]*Complex Systems Engineering and Human Systems, Mohammed VI Polytechnic University, Ben Guerir, Morocco*
[3]*Laboratory of Mathematical Modeling, Simulation and Smart Systems, ENSAM-Meknes, Moulay ISMAIL University, Meknes, Morocco*

Keywords: Heart Disease, Classification, Feature Selection, Ensemble Learning, Ensemble Feature Selection, Univariate Filter.

Abstract: Feature selection is a fundamental data preparation task in any data mining objective. Deciding on the best feature selection technique to use for a specific context is difficult and time-consuming. Ensemble learning can alleviate this issue. Ensemble methods are based on the assumption that the aggregate results of a group of experts with average knowledge can often be superior to those of highly knowledgeable individual ones. The present study aims to propose a heterogeneous ensemble feature selection for heart disease classification. The proposed ensembles were constructed by combining the results of five univariate filter feature selection techniques using two aggregation methods. The performance of the proposed techniques was evaluated with four classifiers and six heart disease datasets. The empirical experiments showed that applying ensemble feature ranking produced very promising results compared to single ones and previous studies.

1 INTRODUCTION

Heart disease (HD) is considered the principal cause of death worldwide and is, therefore, one of the main priorities in medical research (Benhar et al., 2019). Early and accurate diagnosis of cardiac disease is crucial to start appropriate treatment immediately and prevent early death. Data mining (DM) tools have been of great help to researchers to assist physicians and support patients in regard to heart disease diagnosis (Kadi et al., 2017). DM is the mathematical core in the process of knowledge discovery in databases (KDD) which offers powerful tools that allow the extraction of meaningful information, patterns, associations, or relationships from huge amounts of data. Classification is the DM task most frequently used by researchers to diagnose heart disease (Benhar et al., 2019). Classification and other DM techniques are usually hindered by some data imperfections such as missing values, outliers, noise, imbalanced data, and high dimensionality (Benhar et al., 2020). A data preprocessing step is, therefore, mandatory to prepare data for the KDD process. According to the systematic literature review conducted in (Benhar et al., 2019), researchers were mainly interested in feature selection (FS) as a preprocessing task in order to improve the performance of their DM techniques in HD prediction. Researchers made use of different types of feature selection techniques such as filters, wrappers, embedded, and hybrids. However, according to the authors' knowledge, no work has investigated the use of ensemble FS to predict HD. Ensemble methods are based on the assumption that combining the outputs of multiple learners can be significantly more accurate than the output of a single one (Zhou, 2012). In addition to classification problems, ensemble learning can be applied to improve other machine learning tasks such as FS (Seijo-Pardo et al., 2017). Ensemble FS techniques can be classified as: (1) heterogeneous ensembles which consist of using different FS techniques (or base selectors) and the same training data, and (2) homogeneous which consist of using the same base selector and different data subsets.

The present study aims to propose an heterogeneous ensemble FS for heart disease classification by combining the results of five univariate filter FS techniques namely Linear Correlation (Gooch, 2011), ReliefF (Urbanowicz et al., 2018), Information Gain (Quinlan, 1986), Symmetrical uncertainty (Hall & Smith, 1998), and Chi-square (Jin et al., 2006). Univariate filters, also known as feature rankers, consist of ranking features

Benhar, H., Idri, A. and Hosni, M.
Ensemble Feature Selection for Heart Disease Classification.
DOI: 10.5220/0010800500003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 369-376
ISBN: 978-989-758-552-4; ISSN: 2184-4305

individually based on some performance measures and the final features subset can be determined by setting a cutoff threshold or specify how many features to retain. The proposed ensemble combines the features' scores obtained with base rankers using mean and median combination methods and the final feature subsets are obtained by selecting 40% of the top ranked features. The subsets selected with ensemble rankers as well as single ones were evaluated using four classifiers: K-Nearest Neighbors (KNN) (Han et al., 2012), Decision Trees (DTs) (Han et al., 2012), Support Vector Machines (SVM) (Vapnik, 2000), and Multilayer Perceptron (MLP) for heart disease diagnosis (Gardner & Dorling, 1998). The motivation behind the choice of the aforesaid classifiers is that they are the most frequently used classifiers by researchers to predict heart disease (Benhar et al., 2019)(Hosni et al., 2020). Moreover, the reason for choosing the abovementioned FS techniques is their popularity among researchers in heart disease classification (Benhar et al., 2020) and several other fields such as bioinformatics, software development effort estimation, network intrusion detection, and educational data mining. The experiments were performed using Python's Scikit-learn and ITMO-FS libraries (Schlemmer et al., 2014)(Pilnenskiy & Smetannikov, 2020). The classifiers were evaluated using a 10-fold cross validation method and accuracy rate. Overall, this study evaluates 192 variants of classifiers: 192 = (4 classifiers) * (5 univariate-filters + 2 ensembles + original features set) * (6 datasets); and aims at addressing the following research questions:

RQ1: Is there any single ranking technique that distinctly outperform other single ranking techniques?

RQ2: Do ensemble feature rankers (EFR) outperform single ones when used for heart disease classification? Is there a combination method that resulted in better ensembles?

The remainder of this paper is organized as follows: Section 2 presents a brief review of ensemble approaches and related work. The experimental design is described in Section 3. Results are presented and discussed in Sections 4. Finally, the conclusions and future work are presented in Section 5.

2 ENSEMBLE LEARNING AND RELATED WORK

It is well known that a machine learning technique can perform well on some data and less accurately on others. Ensemble methods were introduced to overcome the weaknesses of single techniques and consolidate their advantages (Zhou, 2012). Ensemble learning has become a hot topic for the last three decades and has been successfully applied to various fields including heart disease classification (Hosni et al., 2021).

According to the results of the systematic map conducted in (Hosni et al., 2021), most of the studies state that ensemble methods are able to perform better than single ones. An overview of a set of selected studies in (Hosni et al., 2021) is presented below.

Bashir and al. (Bashir, Qamar, & Javed, 2015) developed a heterogeneous ensemble classification technique by combining three base classifiers: Naïve bayes (NB), SVM, and DT. The classifiers were combined using majority vote aggregation rule. The proposed technique achieved and accuracy of 81.82% on Cleveland heart disease dataset and outperformed single techniques. LO and al. (Lo et al., 2016) proposed a majority voting heterogeneous ensemble classifier by combining several base classifiers such as SVM, KNN, NB, and DT, among others. The proposed technique was evaluated using six heart disease datasets and achieved an accuracy which slightly outperformed those of single base classifiers. Jadhav and al. (Jadhav et al., 2014) proposed a feature selection-based homogeneous ensemble classification technique to diagnose arrhythmia. The proposed approach based on random subspace and PART tree achieved an accuracy of 91.11%. In (Qin et al., 2017), the authors suggested a novel ensemble algorithm by combining seven classifiers to predict arrhythmia. The proposed technique is based on multiple feature selection techniques and a bagging approach to increase data diversity which is an important criterion to construct ensemble techniques. The approach achieved an accuracy of 93.7%.

Although some of the studies applied feature selection, the focus was on applying ensemble learning during the classification phase. This motivates us to conduct the present study.

3 EXPERIMENTAL DESIGN

This section describes the heart disease data used and the methodology followed to conduct the experiments.

3.1 Heart Disease Dataset

Table 1 summarizes the number of features (the class attribute is not included) with their types, the number of instances, number of classes, and missing values for each dataset.

Our purpose is to distinguish between the absence and presence of a heart disease, and thus, all class values indicating the presence of heart disease in the multi-class datasets were replaced by 1 while class 0 indicates the absence of heart disease.

3.2 Methodology Used

The aim of this work is to apply ensemble FS on heart disease datasets for the classification task. The heterogeneous ensembles will combine different feature ranking techniques based on different measures for diversity. In this study the 10-fold cross validation strategy is used (Witten et al., 2011). KNN, SVM, MLP and DT classifiers were applied using the default parameters of the Scikit-learn library.

The methodology performed on each dataset is as follows:

Step 1: Apply single feature ranking techniques

Step 2: Combine the results of the 5 rankers using mean and median combination methods

Step 3: Apply the 40% threshold for single and ensemble rankers. This will result in 7 subsets for each dataset in addition to the original feature set.

Step 4: Classify the 8 obtained subsets using KNN, SVM, MLP and DT classifiers. Evaluate, by means of accuracy score, the four classifiers using a 10-fold cross validation method. In total we obtain 32 classifiers for each dataset.

Step 5: Cluster the classifiers using Scott-Knott test (Scott & Knott, 1974) based on their accuracy scores to assess the statistical significance of the classification results.

For the sake of simplicity, we used the following abbreviations to name the constructed classifiers:

LC, RF, **IG, SU, and CHI2** denote Linear Correlation, ReliefF, Info gain, Symmetrical uncertainty, and Chi-square univariate filter FS techniques respectively. EME and EMD are the abbreviations of the ensemble rankers constructed with mean and median combination methods respectively. Furthermore, the entire feature set was denoted ORG.

Example: SVMEME refers to SVM classifier trained on a subset selected with the ensemble ranker using mean combination method.

4 EMPIRICAL RESULTS AND DISCUSSION

The results of the empirical experiments are presented and discussed in this section. Feature selection and classification were performed using ITMO-FS and Scikit-learn python libraries respectively, while the Scott-Knott (SK) statistical test was performed using R Software. Thereafter, we present a comparison of our results with those from the literature.

4.1 Data Cleaning and Transformation

Before tackling the feature selection process, the datasets were checked for missing values and irrelevant features. Therefore, a total of thirty-eight attributes of the Unprocessed Cleveland dataset were removed since they contained high percentages of missing values (more than 20%), were irrelevant, or had the same values over all instances. Moreover, one attribute containing 83% of missing values was deleted from the Arrhythmia dataset. Thereafter, instances containing missing values were deleted. Afterwards, all attributes were transformed using the Min-Max normalization technique. The performance of the four classifiers before and after applying normalization was verified. The transformation process did not hurt the classification accuracy; on the contrary, it significantly improved it in the majority of cases.

4.2 Single and Ensemble Feature Selection Results

The application of single and ensemble feature selection resulted in the selection of different feature subsets with the sizes of 5, 4, 5, 14, 22, and 111 features for processed Cleveland, Hungarian, Statlog, unprocessed Cleveland, Z-Alizadeh Sani, and Arrhythmia datasets respectively.

4.3 Classification Results

For each dataset, a total of 32 classifiers were evaluated. The SK test results in terms of accuracy score for the six selected datasets are illustrated in Fig. 1.

The SK test identified two clusters for the processed Cleveland dataset. The best cluster contains 23 classifiers. All SVMs, MLPs, and KNNs appeared in the best cluster, with the exception of those based on SU single ranker. On the contrary, all

DTs belonged to the second cluster with the exception of the one based on SU single ranker.

It can be noticed that the best SK cluster of Statlog dataset contains three clusters. A total of 18 classifiers belong to the best cluster. With the exception of MLPRF, all MLP classifiers appear in the best cluster. No DT classifier appears in the best cluster, except for DTEMD. Furthermore, the best cluster include all SVM and KNN classifiers trained with the original feature set and subsets selected with CHI2, LC, EME, and EMD.

The SK test for Hungarian dataset identified three clusters. With the exception of SVMRRF, DTIG, DTORG, DTRF, and KNNSU, all the classifiers belong to the best SK cluster.

A total of 22 classifiers are present in the best SK cluster for the unprocessed Cleveland dataset. As can be observed, with the exception of SVMORG, SVMCHI2 and DTIG, all DTs, SVMs, and MLPs belong to the best cluster. Moreover, only one KNN classifier is present in the best cluster (KNNIG).

For Z-Alizadeh Sani dataset, the SK test identified two clusters. The best SK cluster includes a total of 19 classifiers. With the exception of MLPSU and SVMSU, all SVMs and MLPs are present in the best cluster. None of DT classifiers appear in the best cluster while for KNN, only three appeared in the best cluster.

The SK test for the Arrhythmia dataset resulted in two clusters. It is to be noted that, with the exception of MLPSU and SVMSU, all SVMs and MLPs belong to the best cluster. None of KNN classifiers appear in the best SK cluster for this dataset. For DT classifiers, only DTRF, DTORG, DTIG and DTLC are present in the best cluster.

4.4 Discussion

The empirical results are discussed in this section according to the RQs from Section 1.

RQ1: Is there any single ranking technique that distinctly outperform other single ranking techniques? The SK test results are summarized in Table 2. to answer this RQ. Table 2. presents the number of occurrences of each feature selection technique present in the best SK cluster for each dataset regardless of the classifier used. We can conclude that LC gives very satisfying results over different datasets since in total 19 out of 24 LC techniques were present in the best SK clusters. The number of occurrences of RF, IG, and CHI2 is acceptable over different datasets. Nonetheless, SU single ranker seem to perform worse than other single rankers and fail to select the most relevant features since its total number of occurrence in the best clusters is very low. In fact, the main difference between LC, which seems to be the best performing single ranker, and SU, the worst performing one, is that LC is based on linear relationships while SU is based on non-linear ones (Saikhu et al., 2019). This suggests that the most relevant features to predict heart disease have a linear relationship with the class attribute and SU failed to identify them.

RQ2: Do ensemble feature rankers (EFR) outperform single ones when used for heart disease classification? Is there a combination method that resulted in better ensembles?

Taking into consideration the initial number of single and ensemble rankers used, 61% of single rankers and 81% of ensemble rankers were present in the best SK clusters over all datasets. This shows that promising results can be achieved by applying ensemble feature selection for heart disease classification. However, some poor performing single techniques such as SU in this case, may influence the performance of ensemble techniques, and thus, investigating multiple ensembles of different sizes might be required. Besides, using the features ranks instead of their scores should be investigated.

As regards the combination methods, there is only a difference of three occurrences between the presence frequency of ensembles constructed with mean and those constructed with median, therefore, it is difficult to draw conclusions.

4.5 Accuracy Comparison with Previous Studies

Compared to previous works, the classification results achieved in our study are very encouraging as shown in Table 3. For example, the accuracy rate achieved for Cleveland dataset with MLP and five attributes selected with ensemble ranking feature selection is very promising compared to that of more complex models such as: (1) BagMOOV (Bashir, Qamar, & Hassan, 2015), an ensemble technique based on five heterogeneous classifiers, or (2) RF ensemble based on CFS and PCA (Ozcift & Gulten, 2011).

For Hungarian dataset, it can be noticed that there is not a significant difference between the accuracy achieved in our study and that achieved in (Kadam & Jadhav, 2020) which used ensemble classification, hyper-parameter optimization and the entire feature set.

Very competitive results are achieved for Statlog and unprocessed Cleveland datasets compared with the previous studies, with only 5 and 14 attributes.

Table 1: Datasets descriptions.

Dataset	No. of instances	No. of features	Types of features	No. of missing values	No. of classes
Processed Cleveland dataset	303	13	6 numeric, 7 nominal	6	5
Hungarian dataset	294	13	6 numeric, 7 nominal	782	5
Statlog Heart data	270	13	6 numeric, 7 nominal	0	2
Unprocessed Cleveland dataset	282	75	42 numeric, 33 nominal	5968	5
Z-Alizadeh Sani dataset	303	55	22 numeric, 33 nominal	NA	2
Arrhythmia dataset	452	279	206 numerical, 73 nominal	407	16

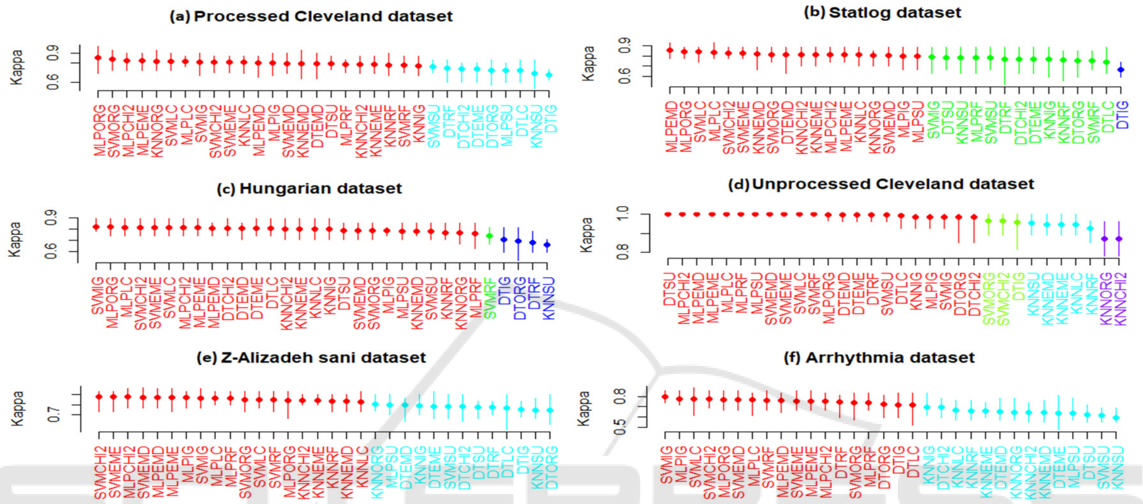

Figure 1: SK test results on each dataset. The x-axis represents the classifiers generated where the better positions start from the left side. The y-axis represents the accuracy values. Each vertical line represents the 10-fold cross validation values for each variant and the small dots represent the mean accuracy values. Lines (classifiers) with the same color belong to the same cluster.

Table 2: Number of occurrence for each FS technique present in the best cluster regardless of the classifier used over all datasets.

	Single rankers						Ensemble rankers		
Dataset	**LC**	**RF**	**IG**	**SU**	**CHI**	**Total**	**EME**	**EMD**	**Total**
Processed Cleveland dataset	3	3	3	1	3	**13**	3	4	**7**
Hungarian dataset	4	2	3	3	4	**16**	4	4	**8**
Statlog Heart data	3	0	1	1	3	**8**	3	4	**7**
Unprocessed Cleveland dataset	3	4	3	3	2	**15**	3	3	**6**
Z-Alizadeh Sani dataset	3	3	2	0	3	**11**	3	4	**7**
Arrhythmia dataset	3	3	3	0	2	**11**	2	2	**4**
Total	**19**	**15**	**15**	**8**	**17**	**74**	**18**	**21**	**39**

For Z-Alizadeh Sani dataset, SVMEMD and KNNEME achieved good results compared to the HE classification technique proposed by Cuvitoglu and al. (Cüvitoğlu & Işik, 2018). Moreover, although the Bagging SMO-SVM outperformed our classifiers, it used a higher number of features (Alizadehsani et al., 2013).

While the classifiers constructed in this study showed promising results for Arrhythmia dataset compared with those proposed in (Xu et al., 2017), the ensemble technique proposed by Jadhav and al. (Jadhav et al., 2014) achieved a remarkably higher accuracy. Nevertheless, we believe that our results can be improved by building ensembles of different sizes, using other combination methods, and optimizing the hyper-parameters of the classifiers.

Table 3: Accuracy comparison with previous works.

Dataset	Study	Technique	No. of features	Accuracy
Processed Cleveland dataset	**Our study**	**MLPEME**	**5**	**82.17%**
	(Bashir, Qamar, & Hassan, 2015)	BagMOOV	_	84.16%
	(Ozcift & Gulten, 2011)	CFS + PCA + RF	7	80.49%
Hungarian dataset	Our study	SVMEME	4	81.48%
		MLPEME	4	81.11%
		KNNEME	4	80%
	(Kadam & Jadhav, 2020)	DT- based AdaBoost + RS	13	83%
Statlog Heart data	**Our study**	**MLPEMD**	**5**	**85.55%**
		SVMEME	**5**	**82.96%**
		KNNEMD	**5**	**81.85%**
	(Kadam & Jadhav, 2020)	DT- based AdaBoost BO	13	84.81%
	(Bashir, Qamar, & Hassan, 2015)	BagMOOV	_	84.07%
Uprocessed Cleveland dataset	**Our study**	**MLPEMD and MLPEME**	**14**	**100%**
		SVMEMD and SVMEME	**14**	**100%**
		DTEMD and DTEME	**14**	**99.66%**
	(H. et al., 2016)	AdaBoost	29	80.14%
	(Gárate-Escamila et al., 2020)	Gradient-boosted Tree	75	98.7%
Z-Alizadeh Sani dataset	Our study	**SVMEMD**	**22**	**87%**
		KNNEME	**22**	**84.17%**
	(Cüvitoğlu & Işik, 2018)	t-test + PCA + HE	25	86%
	(Alizadehsani et al., 2013)	Bagging SMO	33	92.74%
Arrhythmia dataset	**Our study**	**SVMIG**	**111**	**80%**
		SVMEMD	**111**	**76.66%**
	(Xu et al., 2017)	FDR + DNN	236	80.64%
	(Jadhav et al., 2014)	Random supspace PART tree	_	91.11%

CFS: Correlation based feature selection, **PCA:** Principal component analysis, **HE:** Heterogeneous Ensemble, **RF:** Rotation Forest, **SMO:** Sequential Minimal Optimization, **AdaBoost:** Adaptive boosting, **BO:** Bayesian Optimization, **RS:** Random search, **DNN:** Deep neural networks, **FDR:** Fisher discriminant ratio

5 CONCLUSION

The aim of this study was to investigate the performance of ensemble feature ranking techniques compared to single ones for heart disease prediction To this, the relevant features of six heart disease datasets were selected using five single and two ensemble ranking techniques constructed using mean and median combination methods. The subsets selected with ensemble rankers as well as single ones were evaluated KNN, SVM, MLP and DT classifiers. The results of the empirical experiments showed that linear correlation seem to be the best performing single univariate filter while symmetrical uncertainty is the worst performing one. Moreover, the results obtained with ensemble feature ranking techniques are very promising.

We believe that our results can still be improved by building ensembles of different sizes, using feature ranks instead of feature scores, using other combination methods, and optimizing the hyper-parameters of the constructed classifiers. These aspects will be taken into consideration in future work. Moreover, other missing data handling strategies and multi-class classification will be investigated.

REFERENCES

Alizadehsani, R., Habibi, J., Hosseini, M. J., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., Bahadorian, B., & Sani, Z. A. (2013). A data mining approach for diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine*. https://doi.org/10.1016/j.cmpb.2013.03.004

Bashir, S., Qamar, U., & Hassan, F. (2015). Bagmoov: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting. *Australasian Physical and Engineering Sciences in Medicine*. https://doi.org/10.1007/s13246-015-0337-6

Bashir, S., Qamar, U., & Javed, M. Y. (2015). An ensemble based decision support framework for intelligent heart disease diagnosis. *International Conference on Information Society, i-Society 2014*. https://doi.org/10.1109/i-Society.2014.7009056

Benhar, H., Idri, A., & Fernández-Alemán, J. L. (2019). A Systematic Mapping Study of Data Preparation in Heart Disease Knowledge Discovery. *Journal of Medical Systems*, *43*(1), 17. https://doi.org/10.1007/s10916-018-1134-z

Benhar, H., Idri, A., & L Fernández-Alemán, J. (2020). Data preprocessing for heart disease classification: A systematic literature review. In *Computer Methods and Programs in Biomedicine*. https://doi.org/10.1016/j.cmpb.2020.105635

Cüvitoğlu, A., & Işik, Z. (2018). Classification of CAD dataset by using principal component analysis and machine learning approaches. *2018 5th International Conference on Electrical and Electronics Engineering, ICEEE 2018*. https://doi.org/10.1109/ICEEE2.2018.8391358

Gárate-Escamila, A. K., Hajjam El Hassani, A., & Andrès, E. (2020). Classification models for heart disease prediction using feature selection and PCA. *Informatics in Medicine Unlocked*. https://doi.org/10.1016/j.imu.2020.100330

Gardner, M. ., & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0

Gooch, J. W. (2011). Pearson Product-Moment Correlation Coefficient. In *Encyclopedia of Measurement and Statistics*. Sage Publications, Inc. https://doi.org/10.4135/9781412952644.n338

H., K., H., J., & J., G. (2016). Diagnosing Coronary Heart Disease using Ensemble Machine Learning. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/ijacsa.2016.071004

Hall, M. a., & Smith, L. a. (1998). Practical feature subset selection for machine learning. *Computer Science*.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. https://doi.org/10.1016/C2009-0-61819-5

Hosni, M., Carrillo de Gea, J. M., Idri, A., El Bajta, M., Fernández Alemán, J. L., García-Mateos, G., & Abnane, I. (2020). A systematic mapping study for ensemble classification methods in cardiovascular disease. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-020-09914-6

Hosni, M., Carrillo de Gea, J. M., Idri, A., El Bajta, M., Fernández Alemán, J. L., García-Mateos, G., & Abnane, I. (2021). A systematic mapping study for ensemble classification methods in cardiovascular disease. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-020-09914-6

Jadhav, S., Nalbalwar, S., & Ghatol, A. (2014). Feature elimination based random subspace ensembles learning for ECG arrhythmia diagnosis. *Soft Computing*. https://doi.org/10.1007/s00500-013-1079-6

Jin, X., Xu, A., Bie, R., & Guo, P. (2006). Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/11691730_11

Kadam, V. J., & Jadhav, S. M. (2020). Performance analysis of hyperparameter optimization methods for ensemble learning with small and medium sized medical datasets. *Journal of Discrete Mathematical Sciences and Cryptography*. https://doi.org/10.1080/09720529.2020.1721871

Kadi, I., Idri, A., & Fernandez-Aleman, J. L. (2017). Systematic mapping study of data mining–based empirical studies in cardiology. *Health Informatics Journal*, *1*. https://doi.org/10.1177/1460458217717636

Lo, Y. T., Fujita, H., & Pai, T. W. (2016). Prediction of coronary artery disease based on ensemble learning approaches and co-expressed observations. *Journal of Mechanics in Medicine and Biology*. https://doi.org/10.1142/S0219519416400108

Ozcift, A., & Gulten, A. (2011). Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Computer Methods and Programs in Biomedicine*. https://doi.org/10.1016/j.cmpb.2011.03.018

Pilnenskiy, N., & Smetannikov, I. (2020). Feature selection algorithms as one of the python data analytical tools. *Future Internet*. https://doi.org/10.3390/fi12030054

Qin, C.-J., Guan, Q., & Wang, X.-P. (2017). Application Of Ensemble Algorithm Integrating Multiple Criteria Feature Selection In Coronary Heart Disease Detection. *Biomedical Engineering: Applications, Basis and Communications*, *29*(06). https://doi.org/10.4015/S1016237217500430

Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*. https://doi.org/10.1023/A:1022643204877

Saikhu, A., Arifin, A. Z., & Fatichah, C. (2019). Correlation and symmetrical uncertainty-based feature selection for multivariate time series classification. *International Journal of Intelligent Engineering and Systems*. https://doi.org/10.22266/IJIES2019.0630.14

Schlemmer, A., Zwirnmann, H., Zabel, M., Parlitz, U., & Luther, S. (2014). Evaluation of machine learning methods for the long-term prediction of cardiac diseases. *2014 8th Conference of the European Study Group on Cardiovascular Oscillations, ESGCO 2014*. https://doi.org/10.1109/ESGCO.2014.6847567

Scott, A. J., & Knott, M. (1974). A Cluster Analysis Method for Grouping Means in the Analysis of Variance. *Biometrics*. https://doi.org/10.2307/2529204

Seijo-Pardo, B., Porto-Díaz, I., Bolón-Canedo, V., & Alonso-Betanzos, A. (2017). Ensemble feature selection: Homogeneous and heterogeneous approaches. *Knowledge-Based Systems*, *118*, 124–139. https://doi.org/10.1016/j.knosys.2016.11.017

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. In *Journal of Biomedical Informatics*. https://doi.org/10.1016/j.jbi.2018.07.014

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer New York. https://doi.org/10.1007/978-1-4757-3264-1

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier. https://doi.org/10.1016/C2009-0-19715-5

Xu, S. S., Mak, M. W., & Cheung, C. C. (2017). Deep neural networks versus support vector machines for ECG arrhythmia classification. *2017 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2017*. https://doi.org/10.1109/ICMEW.2017.8026250

Zhou, Z. H. (2012). Ensemble methods: Foundations and algorithms. In *Ensemble Methods: Foundations and Algorithms*. https://doi.org/10.1201/b12207