European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 08, 2020

Wrapper Based Feature Selection Techniques On EDHS-HIV/AIDS Dataset

Daniel Mesafint Belete[1], Dr. Manjaiah D.H [2]

[1] Research Scholar,
[1] Department of Computer Science, Mangalore University, Mangalore 574199, India/Faculty Member of Information Technology Department, Debretabor University, Ethiopia
[2] Department of Computer Science, Mangalore University, Mangalore 574199, India

[1]danielmesafint1985@gmail.com, [2]mrmdhmu@gmail.com

Abstract
Selection of features is the mechanism that recognizes the most appropriate attributes and elimination of the redundant and insignificant attributes. These research focuses on a feature selection approach conducted using wrapper methods to predict the individual status/test outcome of the Ethiopian Demographic and Health Survey (EDHS-HIV/AIDS) data set for HIV / AIDS. The study uses three widely employed wrapper-based methods of feature selection to validate the efficacy of the proposed methods namely: Forward Feature Selection (FFS), Backward Feature selection (BFS) and Recursive Feature selection (RFS). We used seven classification algorithms for the purpose of testing selected feature performance, and each classifier output is evaluated using accuracy, precision, recall, f1-score, and ROC. Among the algorithms, the classifiers namely Random Forest, K-Nearest neighbors and Gradient Boosting classifiers achieve higher accuracy levels on the EDHS-HIV/AIDS dataset than others after wrapper method applied. In our research, we have proved that the importance of specified feature selection methods is improving the learning algorithm performance.

Keywords: Feature Selection; Wrapper Methods; EDHS; HIV/AIDS Status

1. INTRODUCTION

Feature is a measurable and single property of the mechanism to be studied. Any machine learning algorithm will use a series of features to perform classification. Over the last few years in machine learning or pattern detection application, the feature scope has increasingly grown. To tackle the issue of raising needless and repetitive variables that are a burden on demanding tasks, many strategies are developed[1]. The process of minimizing large number of features is called user selection by choosing only useful features from the initial dataset to eliminate unnecessary or redundant features.[1-2]

Feature selection will minimize the amount of unnecessary features or variables from input data and can explain input data effectively as well as minimize noises or insignificant attributes and have yet to strong results for prediction.[1]

Selection of features is an excellent approach for simplifying or speeding up operations as well as for enhancing the usefulness of classification and efficiency in perfect cases where attributes are chosen on the basis of class details. Selection of features reduces the excessive dimensionality of features spaces and also provides a deeper interpretation of details, which improves clustering results[3]. The attribute selection method boosts particular instance learning.[1-2,]

[4]The rationale behind the selection of attributes is the improvement in accuracy rates, decrease in dimensionality and decrease in training times as well as improved generalization by eliminating over-fitting. Methods in selection of features are a branch of the more common field of extraction of features.

This paper presents wrapper based methods namely (I) forward feature selection (SFS), (II) backward feature selection (SBS) and (III) recursive feature elimination (RFS). These are employed with seven classifiers namely Random forest, k-nearest, Support vector, Naive Bayes, Logistic Regression, Ada-boost and Gradient Boosting. This study therefore compared performance in seven classification algorithms mentioned above with wrapper based feature selection techniques.

The remaining of the paper is outlined as follows; Sect. 2 deals with related work reported in the related areas. Sect. 3 discusses the wrapper methods used in this paper, Sect. 4 presents the classification techniques, Sect. 5 explains the description of the data variables for EDHS-HIV/AIDS data set, and the experimental results of each selection method are discussed in Section 6 with the classification algorithms and finally, remarking conclusion is obtain in Sect. 7.

2. Related Work

In this section, we have presented the review works related to feature selection and classification algorithms essential in data analytic. The recent literature includes several works incorporating methods for selecting features, including methods for wrapper methods.

Ozcift A.[5] presented a multi-class data collection feature selection approach utilizing wrapper methods. New wrapper technique has been proposed to pick features in multiclass classification issues, the tool being IAFN-FS. This study uses two classical algorithms, C4.5 and Naive Bayes. Direct multiclass solution and several process of binary classification were discussed in order to introduce the multi-class strategy. The study indicates that, while it has the downside of choosing a higher number of attributes, multiple binary classification methods have received great accuracy tests.

Abinash and Vasudevan[6] discussed the set of cancer detection apps for SVM-based wrappers. Two feature selection algorithms are added to the UCI registry cancer dataset, the algorithms used are correlation algorithms and SVM-based wrapper to evaluate the dataset of the leukemia gene. The result of the study show the wrapper-based SVM is better adapted for cancer diagnosis.

Panthong and Srivihok[7] focuses on utilizing wrapper based Dimension Reduction feature selection based on ensemble learning algorithms utilizing 13 (thirteen) UCI Machine Learning Repository datasets. In their work, with ensemble algorithms like AdaBoost and Bagging, the analysis uses forward selection, backward selection, and optimizing selection. The performance is measured using Naive Bayes and Decision Tree. The research reveals that Forward Selection with decision tree based on bagging algorithms obtained stronger results than other approaches. To potential research, the analysis suggests applying hybrid heuristic search and other approaches of ensembles feature selection strategies to real-world problems.

Chitra and Nasira[8] presented wrapper based for CT IMAGE. The study explores numerous feature selection algorithms and suggested a new feature selection method utilizing Bacterial Foraging algorithm Swarm Intelligence. GLCM was used to derive features in this analysis. Compared with the CBFS and OneR, the proposed strategies increase the classification accuracy. The study suggests studies should also be performed using soft computing classifiers for potential research.

Leng J et al.[9] describes a novel concept for recognizing noisy and meaningless features contained in data sets and detecting the consistency of data sets composition using a wrapper-based feature

selection from broad data sets. The paper evaluates the consistency of data sets utilizing the Genetic Algorithms and KNN to eliminate the noisy characteristics of the initial data sets. For a prospective work the research contributes to a better approach to several realistic problems about agricultural and bio-informatics applications.

Hui K.H. et al.[10] proposes an advanced wrapper-based methodology prior to incorporation with the model classification SVM as a total fault diagnostic method for rolling aspect case study. The research uses the data for this analysis from Bearing Data Centre's bearing sensation dataset available. The study findings show that the optimal WFS secure the strongest subset of features with a low computation commitment by removing re-evaluation redundancy. For a potential research the analysis proposes further development of the existing WFS approach based on selecting visual features relevant to the image and integrated with machine learning algorithms.

Karegowda, A.G et al.[11] proposed wrapper feature selection methods with genetic algorithms as random searching techniques for subset generation. In this study, numerous classifier algorithms including Naive Bayes, Bayes networks, decision tree of C4.5 and Radial base function of subset classification tool on four structured datasets including Breast Cancer, Heart Stat log, PID data set and WBC. The approaches introduced indicate the classification accuracy increased.

Hsu H.H. et al.[12] introduce a hybrid filtering approach incorporating two feature filtering approaches, filters and wrappers. This hybrid system optimizes the filters as well as the wrappers. Two bio-informatics issues investigate the process, namely the identification of protein-disordered regions and selection of gene in data on microarray cancer. The experimental result suggest that a smaller number of features will achieve equal or greater prediction precision and the findings indicate that the system is effective for these two forms of feature sets.

Backstrom and Caruana[13] presented Cascade Correlation (C2) nets with selection process named C2FS using wrapper feature selection. The research uses five datasets to test the efficiency of this process. UCI Irvine machine learning system has two issues, two are taken from the 2004 KDD-CUP and the other is from the data collection for medical risk prediction. A new internal wrapper feature selection system selects features when introducing secret units to the net architecture of C2.

Liu, Y. et al.[14]builds an advanced SVM model for demand forecasting. Firstly, genetic algorithm dependent wrappers are employed to evaluate a product's sales results. Then the product of the selection is added to create a regression model for SVM. Often used for comparison and validation are numerous other methods such as Radius Basis Feature Neural Network (RBFNN), Winter Model and SVM without function selection.

Wang A. etal.[15] proposed a strategy for optimizing subset selection techniques using embedded KNN based on the wrapper. The researchers have proposed that a classifier distance matrix be built and retained dynamically, which is the distance between instances and the chosen subset of features. This suggested solution refers to three forms of feature selection to optimize the methods like SFS, IWSS, and IWSSr wrapper methods. Eight microarray datasets with KNN classifiers embedded used to evaluate the method's output. The suggested solution accelerates with high classification precision the wrapper-based feature selection procedure. Regarding potential research they proposed researching backward sequential selection and floating sequential selection to test other learning algorithms with related properties.

3. Wrapper Methods

In wrapper method [16], the selection algorithm for the subset feature occurs as a wrapper around the algorithm for induction or classification. The subset of the feature selection algorithm utilizes the inference or classification algorithm itself as part of the feature evaluation subset principle to perform a search for an suitable subset. The concept under the wrapper method shown in Figure 1 is straightforward: a black box is assumed to be the induction or classification algorithm. From

the data, usually partitioned into internal training and holdout groups, the induction or classification process is done with different sets of features excluded from the results. The subset of the feature in the highest performance is selected as final subset, induction algorithm can be conducted. The classifier result would then be evaluated on a specific data range not encountered during the search.

These approaches are based on greedy search algorithms as they analyze all possible combinations of features and pick the combination which produces the best results for a particular machine learning algorithm. In our work, we focus on the wrapper based methods with forward feature selection, backwards feature selection and Exhaustive or Recursive feature selection.

Forward Feature Selection (FFS):
This algorithm begins its search with an empty set, then searches for the feature which helps to achieve the highest classification accuracy, when found, this attribute is simply applied to the empty set which forms the subset of the searched element. This process is replicated as needed until the classification precision cannot be further enhanced by including all the remaining elements.[18]
This algorithm does return a solution, though within a suitable time span, it is assumed that the accuracy of the solution given would be low because the scope is very narrow and a chosen feature cannot be excluded in more iteration. In this work we use these feature selection criteria to select best features from the total original features from EDHS data set. For this study we select 20, 15 and 10 features from EDHS data set using FFS. And here we presents the selected 20 features namely: Sex, Reg, Res_P, Rel, C_Wor, R_SeA, N_ S_Part, Had_Sex, Con_Use, R_Use_Con, R_Have_1SP, R_Nhave_Sex, HIV_Mosq, H_STI, H_O_STI, H_AIDS, E_T_HIV, P_T_HIV, S_Test, T_in_LAB.

Backward Feature Selection (BFS):
This approach works similarly to the FFS, but the operation goes in a certain direction. In reality, the algorithm begins with in a group containing all the accessible features instead it mostly excludes the removal of the feature which improves the accuracy of the classification. This task is replicated again until no classification efficiency can be increased by deleting all of the remaining characteristics.[18]

As FFS, BFS are used in this work to evaluate the performance of the classification algorithms in the experiment. The features are selected after applying these methods from the original data set.

Recursive Feature Selection (RFE):
Recursive elimination of features carries out greedy quest to determine the highest achieving subset of features. This builds the subsequent pattern with the left features before it examines the whole features. This then lists the features according to the elimination order. In the worst case, if a set of data includes N number of features RFE can perform a greedy quest for 2N feature compound. In this study, we are using these methods and using its best features for experiment to evaluate the performances of the algorithms used in this work.

4. Classification Algorithms

This part summarizes briefly the algorithm used in this paper. There are a broad variety of classification algorithms with its strengths and weakness. There is no single learning algorithm which works best on all problems of supervised learning. For the purpose of selecting features and testing the accuracy of each feature selection, we used seven classification algorithms for each selected features. The classification algorithms which tested in this work are RF, KNN, SVM, AdaBoost, Logistic Regression, Gradient Boosting and Naive Bayes.

Random Forest (RF):

RF algorithms construct a class of classification techniques that depend on the multiple decision trees combined. This Classifier Ensembles have a peculiarity in that a certain degree of randomness from their tree-based components. Based on that concept, RF is characterized as a collection of randomized decision trees ensembles generic principles.[19] RF's core unit the so-called root learner is a binary tree constructed using recursive partitioning. RF suits a variety of decision trees on different dataset sub-samples and uses average to improve predictive precision modeling and over-fit controls. In this paper, we are using this classifier to perform the performance of the selected features for prediction.

K-Nearest neighbors (KNN):
KNN is supervised learning technique and one of the main algorithms.[20]The classification rules are produced without additional data by the training samples themselves. KNN does not seek to construct an internal base model, but merely stores instances of training data. Classification is decided by a simple majority vote of the nearest k-neighbors to each point. We use KNN to measure the output of selected feature expected from the initial data set.

Support Vector Machine (SVM):
SVM is an algorithm which has multiple kernel options depending on the fashion of the distribution of data. It can classify data in multiple linear ways but SVM gives us the optimal among all the possible options. Types of kernel in SVM are linear, rbf, poly, sigmoid. These studies use SVM as a classifier to perform the performances of the feature which is selected with wrapper methods.

Naive Bayes (NB):
NB is a basic model of the concept of generative probabilistic classification that implies equality of entity characteristics to be categorized.[21] The Naive Bayes classifier then applies Bayes theorem assuming the existence or absence is irrelevant to certain features. Given its assumption of independence, its designation effectiveness has been proved.[22] In addition, Naive Bayes only requires a limited amount of training data to estimate the parameters needed for classification. We are using these classification methods on EDHS real data set for evaluate the features.

Logistic Regression (LR):
LR is statistical approach for evaluating a data collection where one or even more independent variables are present which determine the result. The effect is calculated using a dichotomous equation (only two outcomes are possible). A Dependent variable in logistic regression is dichotomous or binary, which contains only data coded as either 1 (TRUE) or 0 (FALSE).[23] The probabilities of using a logistic function in this approach that describe the possible outcomes of the single experiment are modeled.

AdaBoost (AB):
AB performs the classification by selecting only those discrete features that can best be distinguished between the classes.[24]The most influential algorithm within the Boosting family is AdaBoost. It preserves the distribution of probabilities of training sample and changes the distribution of probabilities during each iteration for each study. The member classifier is developed using a specific learning algorithm, and its error rate is calculated on the training. AdaBoost uses the error rate to change the distribution of training samples in probability.[24]

Gradient Boosting (GB):
GB is an algorithm that iteratively builds and improves a set of decision trees, each one conditioned and pruned on instances that previously learned trees have passed through. The previous trees wrongly labeled instances are re-sampled with higher likelihood to give a new distribution of likelihood for the next iteration.

5. Dataset Description

To perform the performance of the proposed methods, we are using HIV/AIDS dataset obtained from Ethiopian Demographic and Health Survey (EDHS) dataset from Central Statistics Agency (CSA).[25] The sample size of the EDHS-HIV/AIDS dataset includes 78,877 instances, out of which 55,209 instances belong to one class and 23,668 instances are another class. In the dataset 26 define features are available, some of which are numerical and some are nominal and the predicted class is negative or positive. Table I presents the description.

6. Experimental Results

EDHS-HIV / AIDS was used to evaluate with wrapper based feature selection methods for predicting individual test status. To evaluate the performances the classification accuracy, seven classification algorithms mentioned above were considered. The methods of classification of the feature implemented in this paper are FFS, BFS and RFS. During feature selection task 20, 15, and 10 features were selected by the feature selection algorithms. After selecting the required features same experiment was repeated for seven classifiers.

In this work evaluation metrics were using to evaluate each the algorithm performance using selected features. The most widely used evaluation metrics are accuracy, precision, recall, and confusion matrix. The next part will show each evaluation metrics as follows:

Accuracy:
is the amount of accurate predictions determined by the overall number of predictions multiplied by hundred to give it a percentage.

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (1)$$

Recall:
the number of True Positives (TP) divided by the number of True Positives (TP) and the number of False Negatives (FN). Another way to express is the number of positive predictions divided by the number of positive class values in the test data.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (2)$$

Precision:
is calculated based on the number of True Positives (TP) divided by the number of True Positives (TP) and False Positives (FP). In another way the number of positive predictions divided by the total number of positive class values predicted.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

F1-measure: is calculated based on precision and recall

$$F1 - Measure = 2\frac{Precision*Recall}{Precision+Recall} \quad (4)$$

ROC: is commonly employed when determining statistical outcomes. Those are the instances with true positive situations for the false positive figure on the X and Y axes.

Confusion Matrix:
is a metric shows correctly classified and Miss-classified samples from a given test data. Table II shows the confusion matrix used in this work.

Table III shows the result of proposed method. The highest accuracy obtained from EDHS-HIV/AIDS data set using wrapper method with Random Forest, Gradient Boosting, K-Nearest Neighbor, Support Vector Machine, AdaBoost, Naive Bayes and Logistic Regression classifier is 0.889, 0.888, 0.868, 0.832, 0.796, 0.767 and 0.746 respectively. The above accuracy score are obtained from the selected feature with 20 features from

	Variables	Values	Description
	Sex	M, F	Gender
	Age	Continues values	Age of Individuals
	Reg	1,2,3,4,5,6,7,8, 9,10,11	Region where the individual is living
	Res_P	1,2	Place where the individual is living
	Rel	1-15	Religion of Individual
	Edu_Lvl	0,1,2,3,4	An Educational Level of Individual
	Edu_Ata	0,1,2,3,4,5	Educational Attainment of the Individual
	M_Sta	0,1,2,3,4,5	Marital Status
	C_Wor	0,1	Is the person is Currently working?
	W_Ind	1,2,3,4,5	Wealth Index
	R_SeA	0,1	Resent Sexual Activity
	N_ S_Part	0,1	Number of Sex partner
	Had_Sex	0,1	Early Sexual Intercourse
	Con_Use	0,1	Condom Usage
	R_Use_Con	0,1	Reducing Condom Usage
	R_Have_1SP	0,1	Reducing Sexual partner in to One
	R_Nhave_Sex	0,1	Reducing HIV without having Sex
	HIV_Mosq	0,1	Can Mosquito Transfer HIV?
	H_STI	0,1	Hearing of Sexual Transmitted Infection

	H_O_STI	0,1	Hearing Other Sexual Transmitted Infection
	H_AIDS	0,1	Hearing of HIV
	E_T_HIV	0,1	Ever Tested HIV Before
	P_T_HIV	0,1	Place to Test HIV
	S_Test	0,1	Sample Test result of HIV
	T_in_LAB	0,1	Laboratory Test
	F_T_Resu	0,1	Final Test Result

the original features compared with selected features of Forward Feature Selection (FFS) with feature 15 and 10.

Similarly, in the Backward Feature Selection methods, the experimental result is obtained with same classifiers which are used in FFS methods and the result with Random Forest, Gradient boosting, K-Nearest neighbors, Ada-boost, Support Vector Machine, Logistic Regression and Naive Bayes are 0.907, 0.895, 0.875, 0.756, 0.748, 0.701 and 0.701 respectively. This accuracy score is obtained from the selected features of 20 features compared with 15 and 10 selected features from original features of the EDHS-HIV/AIDS data set.
In this experiment, we also test Recursive feature selection methods with 20, 15 and 10 features and experiments are done with same classifier as taken from FFS and BFS methods. The experimental result are obtained with Random Forest, Gradient Boosting, K-Nearest neighbors, Ada-boost, Support Vector Machine, Logistic Regression and Naive Bayes is 0.948, 0.909, 0.903, 0.821, 0.815, 0.714 and 0.713 respectively.

Table II. Confusion Matrix

		Predicted	
		Positive	**Negative**
Actual	**Positive**	True Positive	False Positive
	Negative	False Negative	True Negative

Table III. Classification result of wrapper based feature selection Methods on EDHS-HIV/AIDS Dataset

Classification Algorithms	Feature Selection Methods	No. Of Selected Features	Evaluation Metrics				
			Accuracy	Precision	Recall	F1-Score	ROC
Random Forest (RF)	FFS	20	0.889	0.833	0.974	0.898	0.89
		15	0.870	0.899	0.833	0.864	0.87
		10	0.857	0.898	0.807	0.850	0.86
	BFS	20	0.907	0.860	0.971	0.912	0.91
		15	0.875	0.896	0.849	0.871	0.88
		10	0.800	0.800	0.800	0.800	0.80
	RFS	20	0.948	0.918	0.984	0.950	0.95
		15	0.923	0.887	0.970	0.926	0.92
		10	0.859	0.803	0.951	0.871	0.86
K-	FFS	20	0.8	0.83	0	0.8	0

nearest Neighbor (KNN)			68	5	.916	74	.87
		15	0.852	0.862	0.841	0.851	0.85
		10	0.848	0.910	0.778	0.839	0.85
	BFS	20	0.875	0.837	0.932	0.882	0.88
		15	0.857	0.880	0.830	0.854	0.86
		10	0.761	0.768	0.753	0.760	0.76
	RFS	20	0.903	0.872	0.944	0.907	0.90
		15	0.885	0.848	0.938	0.891	0.89
		10	0.848	0.805	0.919	0.858	0.85
Support Vector Machine (SVM)	**FFS**	20	0.832	0.775	0.936	0.848	0.83
		15	0.830	0.774	0.932	0.846	0.83

		10	0.830	0.775	0.929	0.845	0.83
	BFS	20	0.748	0.677	0.949	0.790	0.75
		15	0.739	0.675	0.923	0.780	0.74
		10	0.739	0.675	0.920	0.779	0.74
	RFS	20	0.815	0.754	0.932	0.834	0.82
		15	0.826	0.778	0.912	0.840	0.83
		10	0.821	0.778	0.909	0.835	0.82
Naive Bayes (NB)	**FFS**	20	0.767	0.815	0.862	0.838	0.77
		15	0.759	0.756	0.770	0.763	0.76
		10	0.756	0.750	0.773	0.761	0.76
	BFS	20	0.701	0.682	0.75	0.715	0.70

					2		
		15	0.666	0.628	0.817	0.710	0.67
		10	0.645	0.559	0.877	0.712	0.65
	RFS	20	0.713	0.670	0.853	0.750	0.71
		15	0.714	0.665	0.868	0.753	0.71
		10	0.713	0.665	0.876	0.756	0.71
Logistic Regression (LR)	**FFS**	20	0.746	0.726	0.805	0.764	0.75
		15	0.757	0.755	0.765	0.760	0.78
		10	0.755	0.753	0.773	0.763	0.76
	BFS	20	0.701	0.684	0.746	0.714	0.70
		15	0.689	0.713	0.642	0.676	0.69
		10	0.661	0.656	0.6	0.673	0.6

					91		6
	RFS	20	0.714	0.671	0.851	0.751	0.71
		15	0.715	0.667	0.870	0.755	0.71
		10	0.715	0.669	0.876	0.759	0.71
AdaBoost (AB)	**FFS**	20	0.796	0.779	0.838	0.807	0.80
		15	0.794	0.775	0.832	0.802	0.79
		10	0.776	0.796	0.752	0.774	0.78
	BFS	20	0.756	0.755	0.767	0.761	0.76
		15	0.736	0.743	0.733	0.738	0.74
		10	0.679	0.671	0.709	0.689	0.68
	RFS	20	0.821	0.799	0.863	0.830	0.82
		15	0.803	0.786	0.	0.812	0.

					839		80
		10	0.776	0.771	0.799	0.785	0.78
Gradient Boosting (GB)	**FFS**	20	0.888	0.912	0.962	0.962	0.89
		15	0.875	0.911	0.833	0.870	0.86
		10	0.861	0.909	0.807	0.855	0.86
	BBF	20	0.895	0.914	0.875	0.894	0.90
		15	0.877	0.896	0.856	0.875	0.88
		10	0.799	0.792	0.816	0.804	0.80
	RFS	20	0.909	0.904	0.917	0.911	0.91
		15	0.897	0.901	0.894	0.897	0.90
		10	0.876	0.904	0.847	0.875	0.88

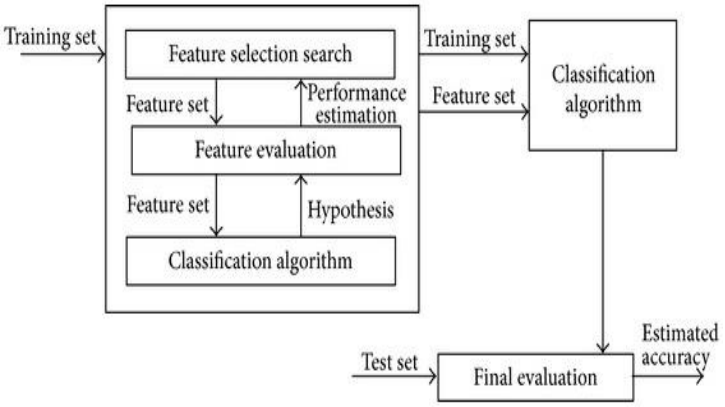

Figure. 1 The wrapper methods of selecting features[17]

7. CONCLUSIONS AND FUTURE WORK

Selection of features is an excellent approach for simplifying or speeding up operations as well as for enhancing the usefulness of text classification and efficiency in perfect cases where attributes are chosen on the basis of class data.

In this research, a proposed feature selection approach was conducted using wrapper methods to predict the individual status or test outcome of HIV / AIDS from the EDHS data set. The research uses three methods of features selection under the wrapper base to classify the data set namely FFS, BFS and RFS. In our work, we assess the selected features and test their classification performance output using Random Forest, K-Nearest neighbors, Support Vector Machine, Gradient Boosting, AdaBoost, Naive Bayes and Logistic Regression classifier. The performance was measured by five evaluation metrics namely: Accuracy, Precision, Recall, F1-measure and ROC.

From the experiment, Random Forest, K-Nearest neighbors and Gradient Boosting classifiers have higher accuracy levels on EDHS-HIV/AIDS data set than the others after the applying the wrapper based feature selection methods. This study shows that methods of selecting features are capable of improving the learning algorithms efficiency.

Finally, the output of this study can make significant contributions in the prediction of the status of HIV/AIDS result of individuals in health domain research and provide wrapper based feature selection methods for machine learning studies. As a future work, a research will be designed as a potential job to explore the other methods of selecting features which to compete with wrapper based on the efficiency of feature selection methods and classification accuracy.

Acknowledgement:
The authors are thankful for all positive reviews and suggestions to the publisher and the anonymous reviewers.

REFERENCES

[1] G. Chandrasekhar and F. Sahin, “A survey on feature selection methods,” Computers & Electrical Engineering, vol. 40, no. 1, pp. 16–28, 2014.
[2] V. Kumar and S. Minz, “Feature selection: a literature review,” SmartCR, vol. 4, no. 3, pp. 211–229, 2014.
[3] K. Mugunthadevi, S. Punitha, M. Punithavalli, and K. Mugunthadevi, “Survey on feature selection in document clustering,” International Journal on Computer Science and Engineering, vol. 3, no. 3, pp. 1240–1244, 2011.

[4] A. Janecek, W. Gansterer, M. Demel, and G. Ecker, "On the relationship between feature selection and classification accuracy," in New challenges for feature selection in data mining and knowledge discovery, 2008, pp. 90–105
[5] A. Ozcift and A. Gulten, "A robust multi-class feature selection strategy based on rotation forest ensemble algorithm for diagnosis of erythematosquamous diseases," Journal of medical systems, vol. 36, no. 2, pp. 941–949, 2012.
[6] M. Abinash and V. Vasudevan, "A study on wrapper-based feature selection algorithm for leukemia dataset," in Intelligent Engineering Informatics. Springer, 2018, pp. 311–321.
[7] R. Panthong and A. Srivihok, "Wrapper feature subset selection for dimension reduction based on ensemble learning algorithm," Procedia Computer Science, vol. 72, pp. 162–169, 2015.
[8] D. Chitra and G. Nasira, "Wrapper based feature selection for ct image," ICTACT journals in/paper/IJIVP Paper 7, pp. 1096–1103, 2015.
[9] J. Leng, C. Valli, and L. Armstrong, "A wrapper-based feature selection for analysis of large data sets," 2010.
[10] K. H. Hui, C. S. Ooi, M. H. Lim, M. S. Leong, and S. M. Al-Obaidi, "An improved wrapper-based feature selection method for machinery fault diagnosis," PloS one, vol. 12, no. 12, 2017.
[11] A. G. Karegowda, M. Jayaram, and A. Manjunath, "Feature subset selection problem using wrapper approach in supervised learning," International journal of Computer applications, vol. 1, no. 7, pp. 13–17, 2010.
[12] H.-H. Hsu, C.-W. Hsieh and M.-D. Lu, "Hybrid feature selection by combining filters and wrappers," Expert Systems with Applications, vol. 38, no. 7, pp. 8144–8150, 2011.
[13] L. Backstrom and R. Caruana, "C2fs: An algorithm for feature selection in cascade neural networks," in The 2006 IEEE International Joint Conference on Neural Network Proceedings. IEEE, 2006, pp. 4748–4753.
[14] Y. Liu, Y. Yin, J. Gao, and C. Tan, "Wrapper feature selection optimized svm model for demand forecasting," in 2008 The 9th International Conference for Young Computer Scientists. IEEE, 2008, pp. 953–958.
[15] A. Wang, N. An, G. Chen, L. Li, and G. Alterovitz, "Accelerating wrapper-based feature selection with k-nearest-neighbor," Knowledge Based Systems, vol. 83, pp. 81–91, 2015
[16] R. Kohavi, "Feature subset selection as search with probabilistic estimates," in AAAI fall symposium on relevance, vol. 224, 1994.
[17] R. Kohavi, G. H. John et al., "Wrappers for feature subset selection," Artificial intelligence, vol. 97, no. 1-2, pp. 273–324, 1997.
[18] N. El Aboudi and L. Benhlima, "Review on wrapper feature selection approaches," in 2016 International Conference on Engineering & MIS (ICEMIS). IEEE, 2016, pp. 1–5.
[19] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.
[20] B. V. Dasarathy, "Nearest neighbor (nn) norms: Nn pattern classification techniques," IEEE Computer Society Tutorial, 1991.
[21] T. O. Ayodele, "Introduction to machine learning," New Advances in Machine Learning, pp. 1–9, 2010.
[22] I. Rish et al., "An empirical study of the naive bayes classifier," in IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, no. 22, 2001, pp. 41–46.
[23] S. C. Kartik Chopra, "Logistic regression and convolutional neural networks performance analysis based on size of dataset."
[24] 24. P. Wu and H. Zhao, "Some analysis and research of the adaboost algorithm," in International Conference on Intelligent Computing and Information Science. Springer, 2011, pp. 1–5.
[25] CSA, Demographic and Health Survey (2018)http://www.csa.gov.et /surveyreport/category /2-demographic-and-health-survey(accessedOctober 28, 2018).