

A Novel Feature Selection Algorithm with Dempster Shafer Fusion Information for Medical Datasets

R.Naveen Kumar¹ and Dr. M. Anand Kumar²

¹Department of Computer Science, Karpagam University, Coimbatore, Tamil Nadu, India.

²Department of Information Technology, Karpagam University, Coimbatore, Tamil Nadu, India.

¹Orcid: 0000-0001-9033-9400

Abstract

In database classification one of the essential processes is Feature Selection. In order to measure the relevance of features in a data set there are several existing algorithm are available. An appropriate feature selection algorithm in general makes the most of the relevancy and minimizes the redundancy of the chosen features. The mutual information calculates can effectively approximation the reliance of features on the complete case space, but it cannot precisely correspond to the redundancies in the midst of features. In this paper, a new feature selection algorithm is projected based on highest relevance and least redundancy criterion. The dempster shafer based mutual information is used to compute the relevancy of every feature with class variable and estimate the redundancy by exploiting the association between candidate, chosen and class variables. The efficiency is experienced with ten benchmarked datasets available in the UCI Machine Learning Repository. The results show better performance of proposed approach when compared with some existing algorithms

Keywords: Feature selection, Mutual Information, Dempster – Shafer, Classification

INTRODUCTION

Feature selection acting a significant role in data preprocessing techniques for data mining. It is a procedure of ruling a division of features from the innovative set and appearing patterns in a given dataset to gain the most favorable feature subset according to the given objective of dispensation and condition. It diminishes the number of features, eliminates immaterial, unneeded or raucous and brings instant property for applications: rating up a data mining algorithm, civilizing mining concert such as organization exactness, and improving consequences expansively. The feature collection attempts to choose the tiny extent division of features according to the following criteria.

- The organization precision does not appreciably

reduce.

- The ensuing class sharing, given only the principles for the chosen features, is as seal as potential to the innovative class division given all feature morals.

A. Components in feature selection

Feature selection algorithms to investigate during the splits of features and attempt to locate the finest in the midst of the opposing 2^N applicant subsets according to a few assessment functions. On the other hand this method is complete and it may be too expensive yet for a intermediate sized feature set(N). Other techniques based on heuristic or accidental search methods effort to decrease computational difficulty by negotiation performance [11]. There are four fundamental steps in a characteristic feature assortment method [15, 16] such as subset generation. Feature evaluation function, stopping criterion and a validation procedure as depicted in Figure 1.

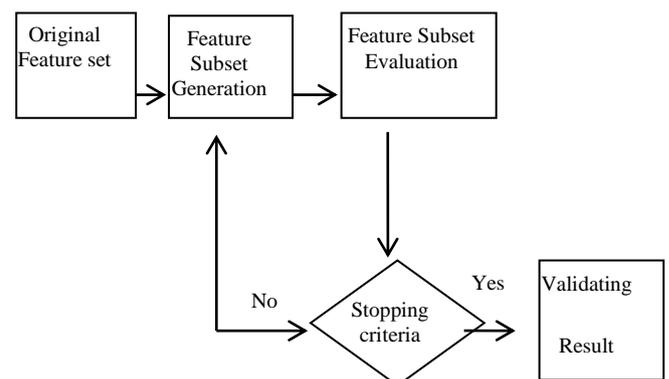


Figure 1: General Feature Selection process

The paper systematized as follows:

- Related works are briefly explained in Unit 2
- The Proposed feature selection Algorithm is Discussed in Unit 3
- Experimental result are shown and discussed in Unit 4.
- Future works and conclusion are suggested in Unit 5

RELATED WORK

In General, Redundancy, Complimentarily and Feature relevance are the basic concept followed by Filter criteria (Vergara & Estévez, 2014) [1]. Information theory method is divided into two group's 1.Linear criteria and 2.Nonlinear criteria. Linear criteria: It is a combination of MI terms. Non linear criteria: It uses Max or minimum operation in their Goal function or It uses Normalization MI function (Brown et al., 2012) [2].

Battiti (1994) [3] Introduce (MIFS) Mutual Information Feature Selection called as First order Incremental Search Algorithm for selecting the most relevant K features from an Initial set of n features .To expand the subset , Greedy selection method is used Instead of calculating 1. Joint MI between selection features and class labels 2.Batti studies MI between Candidate features and the class 3.The Relationship between Already selected features and the candidate Kwok and Choi (2002) [4] The Mutual Information MI between Input features and the class label are used basically to enrich the function. The MIFS-U is introduced to improve the performance of MIFS. mRMR method is another method is introduced by Peng , Long, and Ding (2005) [5].

The mRMR redundancy term is used to divide the cardinality $|S|$ of selected subset S. To avoid large growing and to balance the magnified of this term mRMR is used .In existing literature survey (Brown et al., 2012; Peng et al., 2005) [2][5] as reported mRMR outperform the conventional MIFI and MIFS-U method Estévez, Tesmer, Perez, and Zurada (2009),[6].

NMIFS -Normalized Mutual Information Feature Selection is an Enhanced proposed version introduced instead of MIFS, MIFS-U and mRMR. Instead of Mutual Information MI the Normalized MI is used in the redundancy term. The normalization of MI prevents bias towards multivalued features and limits the value of MI to the range of zero to unity (Estévez et al., 2009) [6].MIFS-ND is a method proposed by - Hoque et al. (2014) [7] this method calculates the candidate features, class label ,Average MI between candidate and the features within the selected subsets.

Genetic algorithm is used in two steps 1. It used to maximize the Mutual Information with the class 2.It used to minimize the MI in other selected features. Mutual Information between candidate features and the Subset features in the selected context are other proposed work done by (Yang & Moody, 1999; Fleuret, 2004; Meyer & Bontempi, 2006; Vidal-Naquet & Ullman, 2003) [8, 9, 10, and 11]. The conditional mutual information method are used to minimize the complexity level of the subsets and the joint mutual information method are used to maximize the complexity level of the subsets .The cumulative mutual information methods are used in selected subsets (Yang & Moody, 1999; Meyer&Bontempi,2006) [8][10]. The proposed feature selection method introduces the

selected subsets in the maximum of the minimum' criterion (2004 Fleuret , 2003 Vidal-Naquet & Ullman) [9] [11].JMI is another proposed selection method introduced by Yang and Moody (1999) [8].Joint mutual information is used to select the subsets of the class label function and selects the max features candidates in the class label. The maximum cumulative subsets are selected from Mutual information features and the classification accuracy and the constancy are calculated. (Brown et al., 2012) [2].

DISR (Double Input Symmetrical Relevance) is a method introduced by Meyer and Bontempi (2006) [10] the main goal of this function is Joint mutual information (JMI) the substituted subset are symmetrically relevance to the class label and the features subsets are directly relevance to the maximum of the minimum criteria proposed by Vidal-Naquet and Ullman (2003) [9] [11].

Information Fragment (IF) is a proposed method introduced by Fleuret (2004) [9].This method is used to integrate the subset into features class label. Comparatively the Conditional Mutual Information method gives higher performance in the later KNN and SVM classifiers works (Freeman, Kuli ´c, & Basir, 2015) [12].The maximum – Minimum feature Interaction method is introduced by Jakulin (2005) [13]. The maximum Number of features is integrated into the minimum number of class labels and the selection criteria can be derived by the formula to find the minimum –maximum number of features .The Interaction Capping method is used to derive the linear combination of the proposed method El Akadi, El Ouardighi, and Aboutajdine (2008) [14]. There are number of feature interaction method, where JMI is typically same as IGFS Interaction Gain Based Feature Selection method. In general the likelihood has been proposed by Brown et al. (2012) [2].This features is used to derive complex problem based on the feature selection criterion .In the filter criteria this method is used to drive all the linear combination of MI and the Non linear techniques cannot be generated by this method. All multi label dataset are used in feature selection techniques Lee and Kim (2015) [15].

In information theory, multiple – label feature selection method are used. Each features in the multi label are measured individually by the filters. There are two methods in filter feature selection method 1. Rough set theory (Liang et al., 2014) [16] 2. Data Envelopment Analysis (Zhang et al., 2014) [17].In fuzzy-rough theory the time inefficiency is a major problem in the existing research. To enhance the time complexity and to improve the efficiency of fuzzy theory a proposed method is given (Qian, Wang, Cheng, Liang, & Dang, 2015) [18].The feature selection method used in DEA concept are lacking in problem solving. The systematic problem solving concept having a huge computational cost (Zhang et al., 2015) [19].

In this work we presented a proposed feature selection

algorithm for redundancy between features with respect to class variables.

PROPOSED SUPERVISED MUTUAL INFORMATION (DEMPSTER SHAFER) ALGORITHM

In supervised MI the features gives a higher classification performance. Individual best features unless they are combined they will give a good classification performance 5.8.13.All the features having high of relevancy with each classes. If the value of the features are inappropriately combined then the redundancy between the features will be inappropriate.

Based on the above observations, a new algorithm is proposed to determine the redundancy between the selected features and the candidate features.

A. THE FRAME OF DISCERNMENT (Θ)

The Hypothesis space is describing all the sets in the complete (exhaustive) space. Θ symbol denotes the frame and the elements are mutually exclusive

If the number of the elements in the set is n, then the power set (set of all subsets of (Θ)) will have 2^n elements.

B. BPA (BASIC PROBABILITY ASSIGNMENT)

In BPA the theory of evidence assigns a Belief mass. Each Belief mass is assigned to a subset. All the subset is called power set. 0 and 1 are the positive no given for the sets. In this form the probability value exists.

If Θ is the frame of discernment, then a function

$m: 2^\Theta \rightarrow [0, 1]$ is called a bpa, whenever

$m(\emptyset) = 0$ and

$\sum m(A) = 1$ and

$A \subseteq \Theta$

C. BELIEF (BEL)

Given a frame of discernment Θ and a body of empirical evidence $\{m(B_1), m(B_2), m(B_3), \dots\}$, the belief committed to $A \in \Theta$ is

$Bel(A) = \sum m(B_i)$

$B \subseteq A$

Also, $Bel(\Theta) = 1$

D. PLAUSIBILITY FUNCTION (PL)

The plausibility is the sum of all the masses values. The sum of all mass value is Set B.

Set B intersect the set A

$Pl(A) = \sum m(B_i), B \cap A \neq \emptyset$

E. BELIEF RANGE

The interval $[Bel(A), Pl(A)]$ is called belief range.

Plausibility (Pl) and Belief (Bel) are related as follows

$Pl(A) = 1 - Bel(\bar{A})$

F. DEMPSTER 'S COMBINATION RULE

The combination called the joint mass (m_{12}) is calculated from the two sets of masses m_1 and m_2

$B \cap C = A, \sum m_1(B) m_2(C)$

$m_{12}(A) = \frac{\sum m_1(B) m_2(C)}{1 - [B \cap C = \emptyset, \sum m_1(B) m_2(C)]}$

G. DEMPSTER-SHAFER SENSOR FUSION ALGORITHM

The generalization Bayesian statistical theory is essential in Dempster-Shafer algorithm. Here all the new feature allows and supports the distributing propositions Eg. "this is Client-A"). Dempster-Shafer reasoning system is enumerated by "frame-of-discernment" system. All the mutual exclusive context interpretation are enumerated and denoted by symbol Θ . E.g.: "A person in a conference room" for a reality constrains this person can normally be "Client A" or "Client B". The task specifies the user identity as one of the four possibilities.

Described by: $\Theta = \{A, B, \{A, B\}, \phi\}$

Significance: 1." Client A" 2."Client B"

3. "Either Client - A or Client-B" actually an indication of ignorance or "neither Client - A nor Client - B" is an indication of exceptional situation.

The Sensor S_i would contribute its observation by assigning its "belief" with the form of discernment symbol Θ . Here, S_i denotes M_i . The Probability mass function is called the assignment function. The probability are indicated by "confidence interval"

S_i observation is calculated by Lower bound and Upper bound method.

The probability of Si observation is “the detected person is the Client-A”. According to confidence interval the lower bound is “Belief” and Upper bound is "plausibility"

[Belief_i(A), Plausibility_i(A)]

Belief_i(A) is quantified by all the pieces of evidence E_k that support proposition “User-A”:

$$Belief_i(A) = \sum_{E_k \subseteq A} m_i(E_k)$$

Plausibility_i(A) is quantified by all the pieces of evidence E_k that do not rule out the proposition "User-A":

$$Plausibility_i(A) = 1 - \sum_{E_k \cap A = \phi} m_i(E_k)$$

For each proposition in Θ, e.g., "User-A", Dempster-Shafer theory gives a rule of combining sensor Si’s observation m_i and sensor S_j’s observation m_j:

$$(m_i \oplus m_j)(A) = \frac{\sum_{E_k \cap E_{k'} = A} m_i(E_k) m_j(E_{k'})}{1 - \sum_{E_k \cap E_{k'} = \phi} m_i(E_k) m_j(E_{k'})}$$

This rule can be bounded directly, if it is viewed m_j not as sensor S_j’s observation, instead of the previously combined observations of sensors S_k and S_l.

Human perception-Reasoning Process is captured by Dempster –Shafer approach. Where the lower end “belief “ and the upper end “plausibility” are the two different key features captured by Dempster – Shafer reasoning process. To compare in order to show the difference, The Bayesian algorithm is a essential subset of Dempster – Shafer approach. Here No mechanism is provided for the quantitative dealings. Quantitative dealings are the range of "belief" and "plausibility". These are humans characteristic attach to the likelihood[20].

In some specific problem the desired features of k varies. If DSI (f_i) with DSI (f_{i-1}) is lower than the small value then the process will stop. Here f_i and f_{i-1} are the selected features. The iteration process for the selected features is ith and i-1th respectively. S is the selected feature and f_i fetches small information into S which has already selected. This process leads to a slight change in the classification where the result

increases slightly [21].

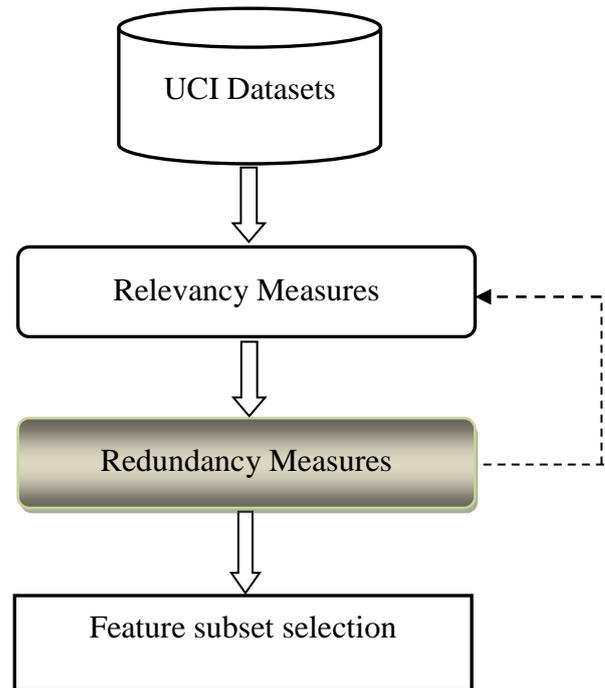


Figure 2: Proposed Framework

EXPERIMENTAL SETUP

This segment demonstrates the investigational outcome and examination of DEMPSTER SHAFER algorithm in conditions of classification correctness and the total number of features preferred. The DEMPSTER SHAFER algorithm is applied in Matlab 2007 on a Pentium IV with 6.4 GHz processor and I GB RAM. Ten datasets are used in this research where these datasets are taken from the well known machine learning engines. The following description lists the total number of features, occurrences and modules used in every dataset.

The implication of the preferred features is estimated by means of two classifiers. The Support Vector Machine¹⁹ (SVM) classifier is based on numerical learning theory. It recognizes a linear or non-linear unraveling exterior in the contribution space of a dataset using a sequential nominal optimization algorithm. The key aspect in the hold vector machine is to use kernels to build a non-linear conclusion boundary. In the current work, radial basis function is used as a kernel. J4.8 classifier is based on conclusion tree developed by Quinlan²⁰. The algorithm construct decision trees from a set of instruction data using entropy based on principle. Hence, each aspect of the data can be used to make a conclusion that splits the data into two subsets. The algorithm inspects the standardized information gain to make the result. The above two classifiers algorithm was passed out in the Weka workbench²¹.

The classification accuracies are the typical value over 10-fold cross justification. All the datasets are accidently divided into 10 subsets of approximately equivalent size. By chance chosen one part of dataset is used for testing and the tingeing is for training. In order to use every part of testing, this method is frequent by ten times the average of ten individual accuracies is the classification precision.

RESULTS AND DISCUSSION

It has been shown that the performance of the Dempster Shafer algorithm is evaluated using ten benchmark datasets. To evaluate and find the classification accuracy features selection algorithms are used. Original features (Unselect), MIFS-U and NMIFS feature selection algorithms are compared. The Table and the Figure shows the classification accuracies of j4.8. Different methods of selected features are used in classification accuracy. The datasets description is as follows.

Dataset description

The experimental result consists of ten different dataset. Each dataset is described as follows

- The BUPA Liver dataset consist of 345 instances with 6 features and 2 classes for distinguishing records.
- Cleveland heart dataset consists of 296 instances with 13 features and 5 classes for distinguishing records.
- Dermatology dataset consists of 358 instances with 34 features and 6 classes for distinguishing records.
- Geman dataset consists of 1000 instances with 20 features and 2 classes for distinguishing record.
- Heart-statlog dataset consists of 270 instances with 13 features and 2 classes for distinguishing record.
- Ionosphere dataset consists of 351 instances with 34 features and 2 classes for distinguishing record.
- Lymphography dataset consists of 148 instances with 18 features and 4 classes for distinguishing record.
- Parkinsons dataset consists of 195 instances with 22 features and 2 classes for distinguishing record.
- Pima Indians Diabetes dataset consists of 768 instances with 8 features and 2 classes for distinguishing record
- WDBC dataset consists of 569 instances with 30 features and 2 classes for distinguishing record

Table 1. Number of Features selected by different methods with j4.8 classifier

Dataset	Whole features	JMIFS	NMIFS	DSFS
BUPA Liver	6	5	5	3
Cleveland heart	13	4	5	4
Dermatology	34	21	22	12
German	20	18	19	7
Heart-statlog	13	7	8	10
Ionosphere	34	9	12	9
Lymphography	18	10	11	8
Parkinsons	22	6	12	7
Pima Indians Diabetes	8	4	6	4
WDBC	30	23	27	19

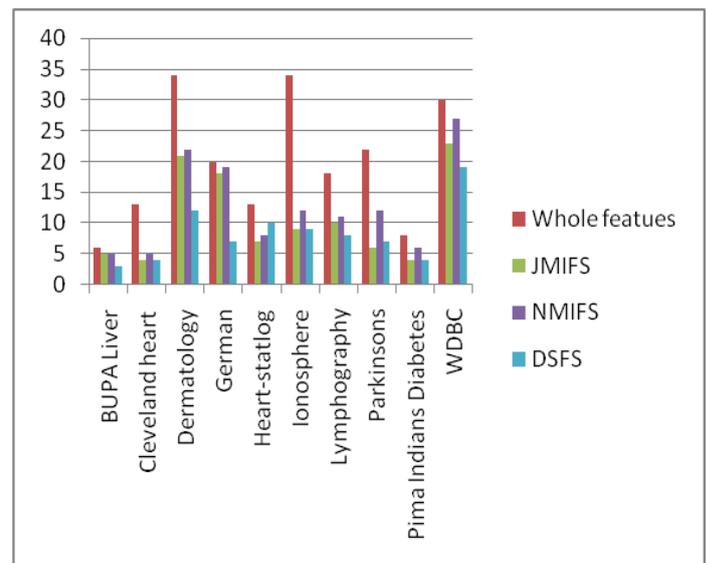


Figure 3: Performance comparison of three approaches with the number of features selected with j4.8 classifier

The table 1 and the figure 3 shows the performance comparison of three methods which have produced their selected features based on J4.8 The bupa liver dataset totally consist of 6 features the JMIFs and NMIFs selects 5 attributes while DSFS selects 3 attributes. The Cleveland heart dataset totally consist of 13 features the JMIFs selects 4 attributes and NMIFs selects 5 attributes while DSFS selects 4 attributes. The Dermatology dataset totally consist of 34 features the JMIFs selects 21 attributes and NMIFs selects 22

attributes while DSFS selects 12 attributes. The German dataset totally consist of 20 features the JMIFs selects 18 attributes and NMIFs selects 19 attributes while DSFS selects 7 attributes. The Heart-statlog dataset totally consist of 13 features the JMIFs selects 7 attributes and NMIFs selects 8 attributes while DSFS selects 10 attributes. The Ionosphere dataset totally consist of 34 features the JMIFs selects 9 attributes and NMIFs selects 12 attributes while DSFS selects 9 attributes. The Lymphography dataset totally consist of 18 features the JMIFs selects 10 attributes and NMIFs selects 11 attributes while DSFS selects 8 attributes. The Parkinsons dataset totally consist of 22 features the JMIFs selects 6 attributes and NMIFs selects 12 attributes while DSFS selects 7 attributes. The Pima Indians Diabetes dataset totally consist of 8 features the JMIFs selects 4 attributes and NMIFs selects 6 attributes while DSFS selects 4 attributes. The WDBC Diabetes dataset totally consist of 30 features the JMIFs selects 23 attributes and NMIFs selects 27 attributes while DSFS selects 19 attributes.

Table 2: Number of Features selected by the different methods with SVM classifier

Dataset	Whole features	JMIFS	NMIFS	DSFS
BUPA Liver	6	5	6	5
Cleveland heart	13	9	8	7
Dermatology	34	16	24	16
German	20	18	17	12
Heart-statlog	13	8	10	5
Ionosphere	34	24	18	12
Lymphography	18	12	14	10
Parkinsons	22	10	16	9
Pima Indians Diabetes	8	5	7	6
WDBC	30	12	14	8

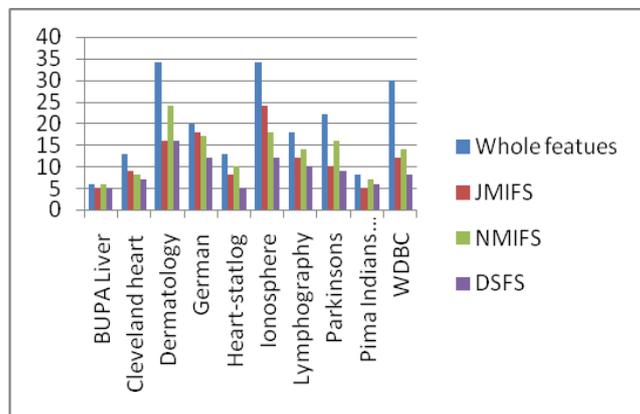


Figure 4: Performance comparison of three approaches with the number of features selected with SVM classifier.

The table 2 and the figure 4 shows the performance comparison of three methods which have produced their selected features based on SVM. The bupa liver dataset totally consist of 6 features the JMIF and DSFS selects 5 attributes while NMIF selects 6 attributes. The Cleveland heart dataset totally consist of 13 features the JMIFs selects 9 attributes and NMIFs selects 8 attributes while DSFS selects 7 attributes. The Dermatology dataset totally consist of 34 features the JMIFs selects 21 attributes and DSFS selects 16 attributes while NMIFs selects 24 attributes. The German dataset totally consist of 20 features the JMIFs selects 18 attributes and NMIFs selects 17 attributes while DSFS selects 12 attributes. The Heart-statlog dataset totally consist of 13 features the JMIFs selects 8 attributes and NMIFs selects 10 attributes while DSFS selects 5 attributes. The Ionosphere dataset totally consist of 34 features the JMIFs selects 24 attributes and NMIFs selects 18 attributes while DSFS selects 12 attributes. The Lymphography dataset totally consist of 18 features the JMIFs selects 12 attributes and NMIFs selects 14 attributes while DSFS selects 10 attributes. The Parkinsons dataset totally consist of 22 features the JMIFs selects 10 attributes and NMIFs selects 16 attributes while DSFS selects 9 attributes. The Pima Indians Diabetes dataset totally consist of 8 features the JMIFs selects 5 attributes and NMIFs selects 7 attributes while DSFS selects 6 attributes. The WDBC Diabetes dataset totally consist of 30 features the JMIFs selects 23 attributes and NMIFs selects 27 attributes while DSFS selects 19 attributes.

Table 3: Performance of three approaches based on accuracy using j4.8 Classifier

DATASET	WHOLE FEATURE S	JMIFS	NMI FS	DSFS
Bupa liver	62	66	63	68
Cleveland heart	77	82	84	89
Dermatology	94	93	92	91
German	70	75	76	78
Heart-statlog	79	82	84	89
Ionosphere	88	90	92	94
Lymphography	79	79	81	85
Parkinsons	84	87	89	92
Pima indians diabetes	65	66	70	72
Wdbc	94	93	95	97
Average percentage	79.2	81.3	82.6	85.5

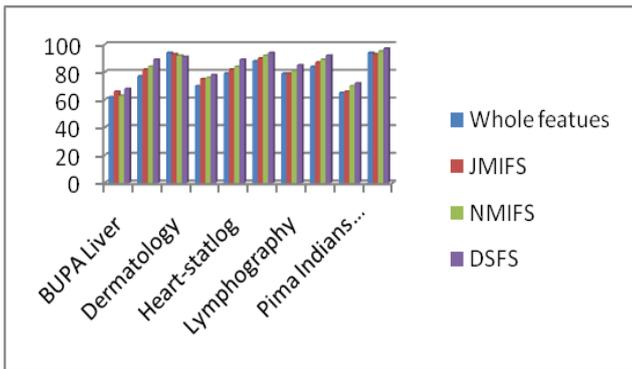


Figure 5: Performance of three approaches based on accuracy using j4.8 Classifier

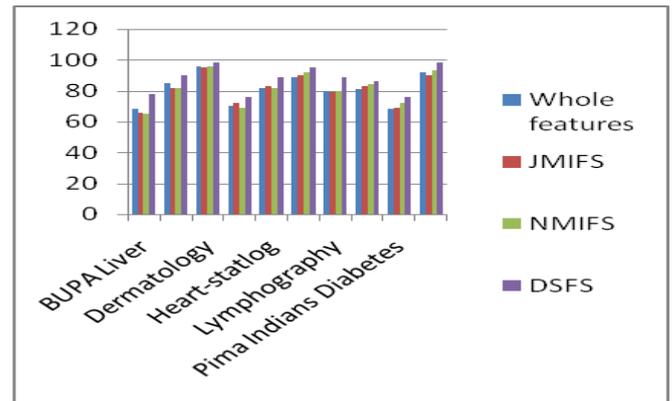


Figure 6: Performance of three approaches based on accuracy using SVM Classifier

The Table 3 and The Figure 5 show the classification accuracy of J4.8. Ten different datasets are classified. In connection with ten datasets, the proposed Dempster-Shafer algorithm shows better performance. The performance are compared with other algorithm. Some of them are German, Heart-statlog, Parkinsons and WDBC datasets. The same accuracy is obtained by Cleveland heart and Ionosphere datasets which are the best approaches in classification theory. Among these classification the Dempster-shafer algorithm performs better with a accuracy of 85.5% where JMIFS algorithm yields 81.3% and NMIFS algorithm provides 82.6%.

Table 4: Performance of three approaches based on accuracy using SVM Classifier

DATA SET	WHOLE FEATURES	JMIFS	NMIFS	DSFS
BUPA Liver	68	66	65	78
Cleveland heart	85	82	82	90
Dermatology	96	95	96	98
German	70	72	69	76
Heart-statlog	82	83	82	89
Ionosphere	89	90	92	95
Lymphography	80	79	80	89
Parkinsons	81	83	84	86
Pima Indians Diabetes	68	69	72	76
WDBC	92	90	93	98
Average Accuracy	81.1	80.9	81.5	87.5

The Table 4 and The Figure 6 show the classification accuracy of SVM. Ten different datasets are classified. In connection with ten datasets the proposed Dempster-Shafer algorithm shows better performance. The performance are compared with other algorithm such as German, Cleveland heart and Lymphography datasets. MIFS_U algorithm gives the equal accuracy in Heart-statlog dataset. Among these classifications the Dempster-shafer algorithm gives better result with accuracy of 87.5% where JMIFS algorithm provides 80.9% and NMIFS algorithm yields 81.5%.

The emissions results for the each algorithm are obtainable. The results are presented in graph and table form that has been shown that the proposed algorithm proves that less redundant features help to improve the classification accuracy. We showed that our propose model significantly perform well and its gives better result when compared to Existing algorithm.

CONCLUSION

This work centers on exploit the maximum relevancy and Lessing the redundancy of features selected. The proposed dempster shafer algorithm calculates the redundancy and relevancy among candidate and selected features with respect to class variable. The classification performance of the algorithm is evaluated by J4.8 and SVM using well known datasets from machine learning repository. The outcome of the algorithm (Results) is compared with the JMIFS and NMIFS feature selection algorithms. It was identified that the proposed DEMPSTER SHAFER results are better with J4.8 and SVM classifiers.

REFERENCES

- [1] Vergara, J & Estévez, P. (2014) "A review of feature selection methods based on mutual information". *Neural Computing and Applications* 24 (1): 175-

186 (2014).

Information Science” University of Ljubljana.

- [2] Brown, G., Pocock, A., Zhao, M., & Lujan, M. (2012) “Conditional likelihood maximization: A unifying framework for information theoretic feature selection” *Journal of Machine Learning Research*, 13, 27–66
- [3] Battiti, R. (1994) “Using mutual information for selecting features in supervised neural net learning” *IEEE Transaction on Neural Networks*, 5, 537–550.
- [4] Kwok, N. & Choi, C. (2002) “Feature selection using Joint Mutual Information Maximisation Expert Systems with Applications” *Volume 42, Issue 22, 1 December 2015, Pages 8520–8532*
- [5] Peng, H., Long, F., & Ding, C. (2005) “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 27, 1226–1238.
- [6] Estévez, P.A., Tesmer, M., Perez, A., & Zurada, J.M. (2009) “Normalized mutual information feature selection” *IEEE Transaction on Neural Networks*, 20, 189–
- [7] Hoque, N., Bhattacharyya, D.K., & Kalita, J.K. (2014) “MIFS-ND: A mutual information-based feature selection method” *Expert Systems with Applications*, 41(14), 6371–6385.
- [8] Yang, H., & Moody, J. (1999) “Feature selection based on joint mutual information” In *Proceedings of international ICSC symposium on advance in intelligent data analysis* (pp.22–25).
- [9] Fleuret, F. (2004) “Fast binary feature selection with conditional mutual information” *Journal of Machine Learning Research*, 5, 1531–1555.
- [10] Meyer, P.E., & Bontempi, G. (2006) “On the use of variable complementarity for feature selection in cancer classification” In *Proceedings of European workshop on applications of evolutionary computing: Evo Workshops* (pp.91–102).
- [11] Vidal-Naquet, M., & Ullman, S. (2003) “Object recognition with informative feature and linear classification” In *Proceeding of the 10th IEEE international conference on computer vision* (pp.281–289).
- [12] Freeman, C., Kulić, D., & Basir, O. (2015) “An evaluation of classifier-specific filter measure performance for feature selection” *Pattern Recognition*, 48, 1812–1826.
- [13] Jakulin, A. (2005) “Machine learning based on attribute interactions (Ph.D.thesis), Computer and Information Science” University of Ljubljana.
- [14] ElAkadi, A., El Ouardighi, A., & About ajdine, D. (2008) “A powerful feature selection approach based on mutual information” *International Journal of Computer Science and Network Security*, 8, 116–121.
- [15] Lee, J., & Kim, D. (2015) “Fast multi label feature selection based on information-theoretic feature ranking” *Pattern Recognition*, 48, 2761–2771.
- [16] Liang, J., Wang, F., Dang, C., & Qian, Y. (2014) “A group incremental approach to feature selection applying rough set technique” *IEEE Transaction on Knowledge and Data Engineering*, 26(2), 294–308.
- [17] Zhang, Y., Yang, A., Xiong, C., Wang, T., & Zhang, Z. (2014) “Feature selection using data envelopment analysis” *Knowledge – Based Systems*, 64, 70–80.
- [18] Qian, Y., Wang, Q., Cheng, H., Liang, J., & Dang, C. (2015) “Fuzzy-rough feature selection accelerator” *Fuzzy Sets and Systems*, 258, 61–78.
- [19] Zhang, Y., Yang, C., Yang, A., Xiong, C.Y., Zhou, X., & Zhang, Z. (2015) “Feature selection for classification with class-separability strategy and data envelopment analysis” *Neuro computing*, 166, 172–184.
- [20] Naveen Kumar R, Anand Kumar M. “Medical data mining techniques for health care systems” *IJESCI*; 6(4):3498–503. DOI: 10.4010/2016.811. ISSN 2321 3361.
- [21] Naveen Kumar R, Anand Kumar M. “Enhanced Fuzzy K-NN Approach for Handling Missing Values in Medical Data Mining” *Indian Journal of Science and Technology*, 9(1):1-7. DOI: 10.17485/ijst/2016/v9iS1/94094.