International Journal of u- and e- Service, Science and Technology
Vol.9, No. 1 (2016), pp.63-72
http://dx.doi.org/10.14257/ijunesst.2016.9.1.07

# A Novel Decision Tree Framework using Discrete Haar Wavelet Transform

Bhanu Prakash Battula[1], KVSS Rama Krishna[1], Debnath Bhattacharyya[2] and Tai-hoon Kim[3]

*[1]Department of Computer Science and Engineering, Vignan'sNirula Institute of Technology and Science, Guntur, Andhra Pradesh, India.*
*[2]Department of Information Technology, BharatiVidyapeeth University College of Engineering, Pune-411043, India*
*[3]Department of Convergence Security, Sungshin Women's University, 249-1, Dongseon-dong 3-ga, Seoul, 136-742, Korea*
*battulaphd@gmail.com, ksai.mb@gmail.com, debnathb@gmail.com, taihoonn@daum.net (Corresponding Author)*

## *Abstract*

*Data Mining is a popular knowledge discovery technique. In data mining decision trees are of the simple and powerful decision making models. One of the limitations in decision trees is towards the data source which they tackle. If data sources which are given as input to decision tree are of imbalance nature then the efficiency of decision tree drops drastically, we propose a decision tree structure which uses discrete haar wavelet transformation technique along with a filter. In this paper, we propose a novel method WT Tree based on above strategy. Extensive experiments, using C4.5 decision tree as base classifier, show that the performance measures of our method is comparable to state-of-the-art methods.*

***Keywords****: Knowledge Discovery, Data Mining, Classification, Decision Trees, Discrete haar*

## 1. Introduction

In Machine Learning community, and in data mining works, classification has its own importance. Classification is an important part and the research application field in the data mining [1].

A decision tree gets its name because it is shaped like a tree and can be used to make decisions. ―Technically, a tree is a set of nodes and branches and each branch descends from a node to another node. The nodes represent the attributes considered in the decision process and the branches represent the different attribute values. To reach a decision using the tree for a given case, we take the attribute values of the case and traverse the tree from the root node down to the leaf node that contains the decision." [2]. A critical issue in artificial intelligence (AI) research is to overcome the so-called ―knowledge-acquisition bottleneck" in the construction of knowledge-based systems. Decision tree can be used to solve this problem. Decision trees can acquire knowledge from concrete examples rather than from experts [3]. In addition, for knowledge-based systems, decision trees have the advantage of being comprehensible by human experts and of being directly convertible into production rules [4].

A decision tree not only provides the solution for a given case, but also provides the reasons behind its decision. So the real benefit of decision tree technology is that it avoids the need for human expert. Because of the above advantages, there are many successes in applying decision tree learning to solve real-world problems.

ISSN: 2005-4246 IJUNESST

The paper is organized as follows. In Sect. 2 we present the recent advances in decision tree learning. This will directly motivate the main contribution of this work presented in Sect. 3, where we propose a new framework using discrete wavelet haar transform. Evaluation criteria's for decision tree learning is presented in section 4. Experimental results are reported in Sect. 5. Finally, we conclude with Sect. 6 where we discuss major open issues and future work.

## 2. Recent Advances in Decision Trees

In Data mining, the problem of decision trees has also become an active area of research. In the literature survey of decision trees we may have many proposals on algorithmic, data-level and hybrid approaches. The recent advances in decision tree learning have been summarized as follows:

A parallel decision tree learning algorithm expressed in MapReduce programming model that runs on Apache Hadoop platform is proposed by [5]. A new adaptive network intrusion detection learning algorithm using naive Bayesian classifier is proposed by [6]. A new hybrid classification model which is established based on a combination of clustering, feature selection, decision trees, and genetic algorithm techniques is proposed by [7]. A novel roughest based multivariate decision trees (RSMDT) method in which, the positive region degree of condition attributes with respect to decision attributes in rough set theory is used for selecting attributes in multivariate tests is proposed by [8].

A novel splitting criteria which chooses the split with maximum similarity and the decision tree is called mstree is proposed by [9]. An improved ID3 algorithm and a novel class attribute selection method based on Maclaurin-Priority Value First method is proposed by [10]. A modified decision tree algorithm for mobile user classification, which introduced genetic algorithm to optimize the results of the decision tree algorithm, is proposed by [11]. A new parallelized decision tree algorithm on a CUDA (compute unified device architecture), which is a GPGPU solution provided by NVIDIA is proposed by [12]. A Stochastic Gradient Boosted Decision Trees based method is proposed by [13]. A modified Fuzzy Decision Tree for the fuzzy rules extraction is proposed by [14].

Obviously, there are many other algorithms which are not included in this literature. A profound comparison of the above algorithms and many others can be gathered from the references list.

## 3. Proposed Method

### 3.1. Wavelet Transform

A wave is a periodic oscillating disturbance that propagates through space and time, usually with transference of energy. In contrast, wavelets are localized waves that have their energy concentrated in time or space and are suited to analysis of transient signals. A wavelet is a localized change of a sound signal in 1-D or localized variations of detail in an image in 2-D. Wavelets are mathematical functions that divide continuous time signal data into different frequency components. If parts of a signal are rapidly changing, they are of a high frequency, and on the other hand, slowly smoothly changing pieces are of low frequency. A portion that does not change has zero, or no frequency.

The fundamental idea behind wavelets is to analyze according to scale. One of the easiest ways to envision scale is to use the analogy of the piano. The eight notes in an octave occur repetitively. As each of the octaves move from the left of the keyboard or to the right of the ivories, this is the translation of each of the notes. There is a low C, a high C and a middle C, but they each represent the concept of a tone, C. There is also a relationship between a musical composition which shows sets of notes, each note is a frequency component, that are played in certain order, at a certain speed, often in combination with one another. A chord is a global view that is made up of the individual notes. The succession of chords is the original signal. This illustrates the idea of scale. The large scale is the big picture, while the small scale shows the details. The reference information in a library can be viewed as an example. Note that a subject area of research, such as digital image processing, is of significantly larger scale than the course of study of wavelets. The branch of knowledge can even be further refined if one only wishes to view information about the Discrete Wavelet Transform. Once we have decomposed an equation for the DWT, and extracted a single fundamental detail, does it have any meaning in the context of the whole of area of digital image processing? This is a question; a question that can be asked as well of a signal. How far can a signal be decomposed, and does it still have value? Scale can also be seen as analogous to viewing a country map (large scale) versus a street map (details) on Yahoo maps. It is equivalent to using the +/- zoom capability to go from the country view to the street view. Is one view superior to the other? It depends on what knowledge we are trying to assimilate. We naturally use a process of scale to refine, understand and digest information. High scales correlate with global information, and low scales correspond with the detailed information.

In the case of wavelets we normally do not speak about time-frequency representations but about time-scale representations. How are mathematical structures created that are varying in scale? Assemble a function, shift it by some amount, and change its scale. In many ways, the wavelet transform is similar to the Fourier transform. Like the Fourier transform, wavelet transforms satisfy specified mathematical criteria to represent the input data. However, while the Fourier transform uses sinusoids to analyze signals, the wavelet transform uses wavelets of finite energy.

### 3.2. Wavelet Transform Decision Tree Framework

In this section, the proposed approach is presented.

In the framework of WT Tree a base algorithm is used in the implementation and the efficiency of the WT Tree will also depend on the fine tuning of the parameters and the base algorithms etc. As to find the efficiency of WT Tree for different parameters, we have designed different variations of WT Tree by varying the type of base algorithm and fine tuning parameters in implementation.

The WT Tree follows a feature conversion and instance misclassification detection approach for continuous improvement. The above said strategy is implemented in the WT system. In the initial stage the decision tree learning process will initiate with the identification of influential features or attributes from the data source. The selected or identified prominent features from the data source are grouped together and again used for conversion by using wavelet transform. In this case, we have used discrete wavelet transform. In the discrete wavelet transform, one of the best applicable approaches for the real world data is Haar wavelet transform. The so called process of features conversion is done. After the conversion of the influential features, the learning process is initiated for identification of misclassified instances by using a base algorithm. The converted “n” features are used to build the model for that particular data source and the efficiency of the model is evaluated. In the next, phase the improved data source is applied to the base algorithm (C4.5) and the learning process is continued for the evaluation of efficiency of the model.

The algorithm for WT Tree approach is given below,

**Algorithm: Wavelet Transform Decision Tree**

```
Input:Tr: a training sample set
La: Label for Tr
M: The number of selected features
Output Measures: Accuracy, Tree size.
Procedure:
Phase I:
Finding important features using filter
Features = Apply Filter (Data set, CFS)
Phase II:
Conversion of the features
for feature i = 1 to M//M represents # features selected
    forTr (j) = 1 to N//N represents # Transform the features
        Transform (WT, j-1)
        j=j-1
    end for
end for
Phase III:
Training on Converted features
For Tr (i,j) = 1 to O//O represents # converted features
    Learn Model = Buildc(Tr (i.j), C4.5)
End for
Predict (Learning Mode, Measure)
```

**The Algorithm: Wavelet Transform Decision Tree can be Explained as Follows**

The inputs to the algorithm are training sample set ‘‘Tr’’ and label of training sample set ‘‘La.’’ The output of the algorithm will be the average measures such as accuracy and tree size produced by the WT Tree method. The algorithm begins with the initial stage of identifying important features M, where M is the number of features extracted by applying correlation based feature subset filter on the data set. The ‘‘M’’ value will change from one data set to other, and depending upon the unique properties of the data set, the number of features can be more or less. In the next stage, features are converted and consecutively trained and decremented on the base algorithm.

## 4. Experimental Design and Evaluation Criteria's

We used the open source tool Weka [16] and implemented our proposed model. In order to test the robustness of our method it is compared to existing methods C4.5 [17], Reduced Error Pruning Tree (REP), Classification and Regression Trees (CART) [18] and naive bayes Trees [NB Tree] in our experiments.

In order to compare the classifiers, we use 10-fold cross validation. In 10-fold cross validation, each dataset is broken into 10 disjoint sets such that each set has (roughly) the same distribution. The classifier is learned 10 times such that in each iteration a different set is withheld from the training phase, and used instead to test the classifier. We then compute the measures as the average of each of these runs. To assess the classification results we used the performance evaluation measures.

The Accuracy (ACC) measure is computed by equation (1),

$$AUC = \frac{1 + TP_{RATE} - FP_{RATE}}{2}$$

The tree size of the decision tree is measured by the number of leaves and depth of the tree.

Table 1.summarizes the datasets used in the proposed study from UCI [15]. The details of the datasets are given in table 1. For each data set, S.no., name of the dataset, number of instances, Missing values, numeric attributes, nominal attributes and Classes are described in the table for all the datasets.

**Table 1.Summary of Imbalanced Datasets**

| S.No. | Dataset | Instances | Missing values | Numeric attributes | Nominal attributes | Classes |
|---|---|---|---|---|---|---|
| 1. | Balance-scale | 625 | No | 4 | 0 | 3 |
| 2. | Breast-cancer | 286 | Yes | 0 | 9 | 2 |
| 3. | Pima diabetes | 768 | No | 8 | 0 | 2 |
| 4. | Glass | 214 | No | 9 | 0 | 6 |
| 5. | Heart-statlog | 270 | No | 13 | 0 | 2 |
| 6. | Ionosphere | 351 | No | 34 | 0 | 2 |
| 7. | Iris | 150 | No | 4 | 0 | 3 |
| 8. | Sonar | 208 | No | 60 | 0 | 2 |
| 9. | Vehicle | 846 | No | 18 | 0 | 4 |
| 10. | Waveform | 5,000 | No | 41 | 0 | 3 |

## 5. Results

We experimented with 10 standard datasets from the UCI repository (Breast_w, Diabetes, Hepatitis, Sonar, Ionosphere, Vote, Colic, labor, Breast and Sick); these datasets are standard benchmark imbalanced datasets used in the context of supervised learning. The goal is to examine whether the WT Tree achieve better predictive performance than a number of existing standard learning algorithms. We compared the above methods with the C4.5, REP, CART and NB Tree state-of-the-art metric learning algorithms. In all the experiments we estimate accuracy using 10-fold cross-validation and control for the statistical significance of observed differences using t-test (sig. level of 0.05).

In Table 2-3, we present the results of the comparison between C4.5, REP, CART, NB Tree and WT Tree. From these results we can make several observations. The developed WT Tree based on C4.5, CART and REP generally given competitive results for C4.5, REP and traditional benchmarks; the advantage of our methods is most visible in the breast_w, diabetes, vehicle, ionosphere and sonar datasets. Finally, the method that most often win is WT Tree.

**Table 2.Summary of Tenfold Cross Validation Performance for Accuracy on All the Datasets**

| S. No | Dataset | C4.5 | REP | CART | NB Tree | WT Tree |
|---|---|---|---|---|---|---|
| 1. | Balance-scale | 77.82● | 78.54● | 78.73● | 75.96● | 97.24 |
| 2. | Breast-cancer | 74.28● | 69.35● | 70.22● | 70.99● | 98.74 |
| 3. | Pima_diabetes | 74.49● | 74.46● | 74.56● | 74.96● | 95.25 |
| 4. | Glass | 67.63● | 65.54● | 71.26● | 69.84● | 76.48 |
| 5. | Heart-statlog | 78.15● | 76.15● | 78.07● | 80.93● | 91.51 |
| 6. | Ionosphere | 89.74● | 89.46● | 88.87● | 90.03● | 93.16 |
| 7. | Iris | 94.73● | 93.87● | 94.20● | 93.47● | 95.19 |
| 8. | Sonar | 73.61● | 72.69● | 70.72● | 77.11● | 82.82 |
| 9. | Vehicle | 72.28● | 70.18● | 69.91● | 70.98● | 86.64 |
| 10. | Waveform | 75.25● | 76.57● | 76.65● | 79.84● | 89.66 |
| Win/Tie/Loss | | (10/0/0) | (10/0/0) | (10/0/0) | (10/0/0) | |

●Bold dot indicates the win of proposed method;

○ Empty dot indicates the loss of proposed method.

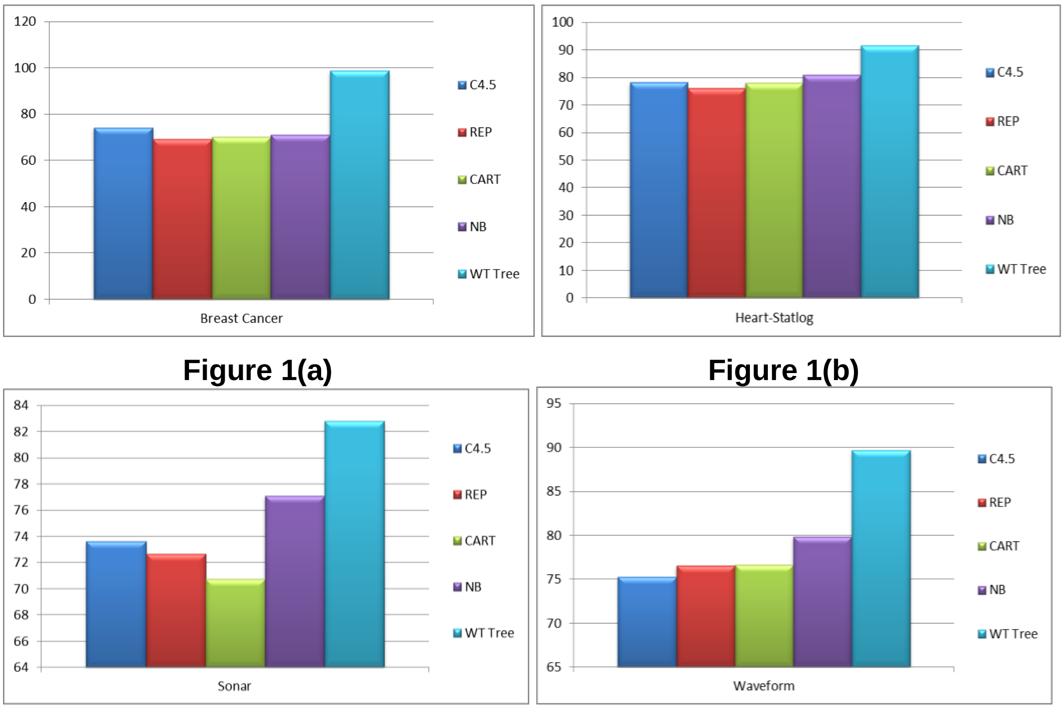

**Figure 1(a)**

**Figure 1(b)**

**Figure 1(c)**

**Figure 1(d)**

**Figure 1. Test Results on Accuracy between the C4.5, REP, CART, NB Tree and WT Treeon 1(a) Breast Cancer 1(b) Heart-Statlog 1(c) Sonar and 1(d) Waveform Dataset**

**Table 3. Summary of Tenfold Cross Validation Performance for Tree Size on All the Datasets**

| S. No | Dataset | C4.5 | REP | CART | NB Tree | WT Tree |
|---|---|---|---|---|---|---|
| 1. | Balance-scale | 82.20● | 42.36● | 55.28● | 17.38○ | 30.64 |
| 2. | Breast-cancer | 23.46● | 13.76● | 15.90● | 5.68○ | 9.58 |
| 3. | Pima_diabetes | 43.40● | 30.98● | 17.36○ | 5.18○ | 27.58 |
| 4. | Glass | 46.16● | 19.70○ | 21.16○ | 10.0○ | 39.98 |
| 5. | Heart-statlog | 34.64● | 14.78○ | 15.36○ | 9.62○ | 19.38 |
| 6. | Ionosphere | 26.74● | 8.76○ | 8.42○ | 16.20○ | 18.42 |
| 7. | Iris | 8.28○ | 5.84○ | 7.40○ | 4.38○ | 10.90 |
| 8. | Sonar | 27.90● | 10.20○ | 10.50○ | 13.74○ | 26.46 |
| 9. | Vehicle | 138.0● | 58.52○ | 92.54● | 57.70○ | 84.62 |
| 10. | Waveform | 591.94● | 167.24○ | 98.32○ | 94.48○ | 290.44 |
| Win/Tie/Loss | | (9/0/1) | (3/0/7) | (3/0/7) | (0/0/10) | |

●Bold dot indicates the win of proposed method;
○ Empty dot indicates the loss of proposed method.

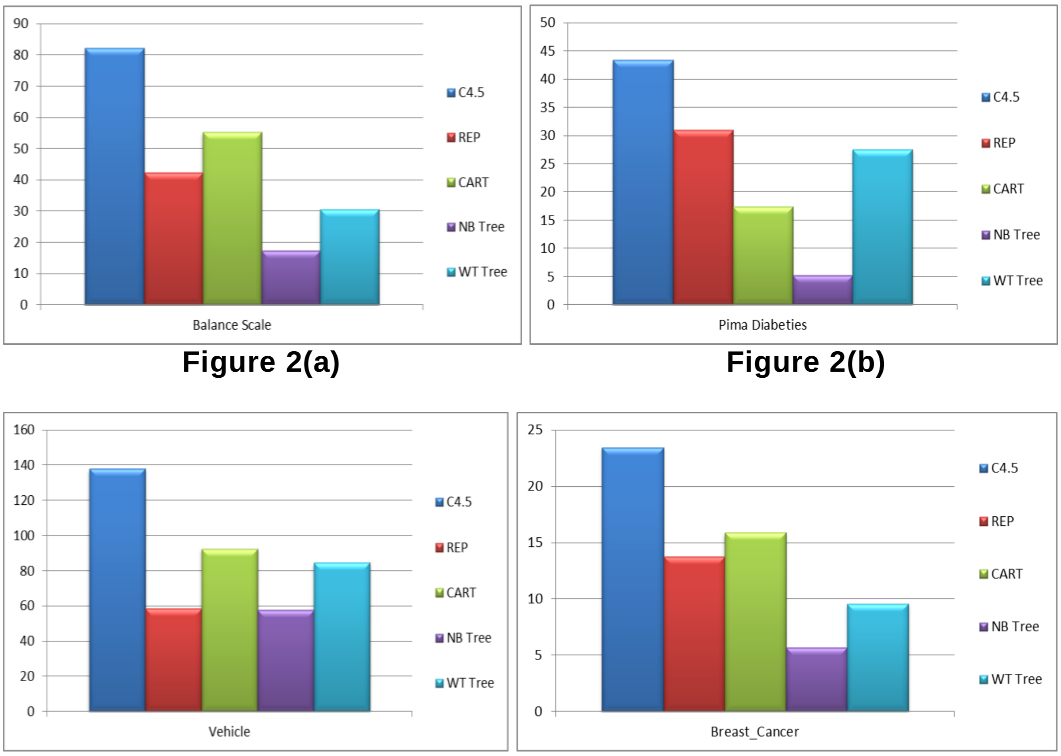

**Figure 2(a)**

**Figure 2(b)**

**Figure 2(c)**

**Figure 2(d)**

**Figure 2. Test Results on Tree Size between the C4.5, REP, CART, NB Tree and WT Tree on 2(a) Balance Scale 2(b)Pima Diabetes 2(c) Vehicle and 2(d) Breast Cancer Dataset**

These results are remarkable since WT Tree, which are based on a simple idea, performs equally well as the more elaborate standard learning algorithm that has been reported to consistently outperform other metric learning techniques over a number of non-trivial learning problems. Finally, we mention that the surprisingly poor performance of WT Tree on iris, waveform and glass datasets in tables 3, might be explained by the fact that its learning function is not convex and hence it is sensitive to the tree size.

Table 2 and 3 presents the summary of experimental results conducted using non-parametric statistical tests with Wilcoxon test for all the datasets. The last row in tables2 and 3 describes number of wins or tie of WT Tree results of accuracy and tree size for every dataset. The values M/N/O specify the number of wins, number of ties and number of losses with compared WT Tree algorithm. Tables 2 and 3 presents the comparative results of proposed algorithm WT Tree against C4.5, REP, CART and NB Tree. The value in the table; example: "(9/0/1)" specifies that the proposed algorithm has registered 9 wins, 0 ties and 1 loss against compared algorithm on that dataset for that specified measure. One can observe from the tables 2 and 3 that our proposed algorithm has registered good number of wins against the compared algorithms on all the datasets.

Fig. 2 and Fig. 3 displays the accuracy and tree size results for the different UCI datasets for each algorithm verses WT Tree. The y-axis in fig.1 to 2 represents the accuracy and tree size value respectively, with the x-axis representing the compared algorithms and WT Tree. Note that each figure is presented relative to a different scale to make the trends more visible, so care must be taken when comparing trends for different learners.

From, Fig.1 and Fig. 2, we can observe that WT Tree are the best algorithm, because on an average they had obtained the highest value in the ranking and the bar charts representing that the proposed algorithm can be seen to performing well on the chart.

In overall, from all the tables we can conclude that our proposed WT Tree has given good results when compared to benchmark algorithms. The unique properties of datasets such as size of the dataset and the number of attributes will also effect on the results of our WT Tree. The above given results are enough to project the validity of our approach and more deep analysis should be done for further analysis.

## 6. Conclusion

In this paper, we proposed a new decision tree algorithm dubbed as WT Tree for classification problem. The WT Tree assumes using a discrete haar wavelet transform for attribute transformation with eliminating mostly misclassified instances can improve all the classification measures such as accuracy and tree size.

The experiments conducted with WT Tree specify that improved classification measures can be achieved. We have conducted experiments on 10 datasets from UCI which suggest that WT Tree can quickly remove redundant, irrelevant and weak instances and attributes as long as the properties of the dataset are normal. Excellent improvement in classification measures on some natural domain datasets shows the compatibility of WT Tree approach on real-time applications. Finally, we can conclude that WT Tree can be a good contribution as a decision tree induction method for efficient learning of the datasets.

## References

[1] J. Hu, J. Deng and M. Sui, "A new approach for decision tree based on principal component analysis", International Conference on Computational Intelligence and Software Engineering, CiSE, IEEE, **(2009)**.

[2] S. Bergsma, "Large-scale semi-supervised learning for natural language processing", University of Alberta, **(2010)**.

[3] J. Durkin and J. Durkin, "Expert systems: design and development", Prentice Hall PTR, **(1998)**.

[4] J. R. Quinlan, "C4. 5: programs for machine learning", Elsevier, **(2014)**.

[5] V. Purdila and Ş. G. Pentiuc, "MR-Tree-A Scalable MapReduce Algorithm for Building Decision Trees", Journal of Applied Computer Science & Mathematics, vol. 16, no. 8, **(2014)**.

[6] D. Farid, N. Harbi and M. Z. Rahman, "Combining naive bayes and decision tree for adaptive intrusion detection", International Journal of Network Security & Its Applications, vol. 2, no. 2, **(2010)**, pp. 12-25.

[7] M. Khanbabaei and M. Alborzi, "The Use Of Genetic Algorithm, Clustering And Feature Selection Techniques In Construction Of Decision Tree Models For Credit Scoring", International Journal of Managing Information Technology (IJMIT), vol. 5, **(2013)**, pp. 13-31.

[8] D. Wang, X. Liu, L. Jiang, X. Zhang and Y. Zhao, "Rough set approach to multivariate decision trees inducing", Journal of Computers, vol. 7, no. 4, **(2012)**, pp. 870-879.

[9] X. Zhang and S. Jiang, "A Splitting Criteria Based on Similarity in Decision Tree Learning", Journal of software, vol. 7, no. 8, **(2012)**, pp. 1775-1782.

[10] Y. Wang, X. Peng and J. Bian, "Computer Crime Forensics Based on Improved Decision Tree Algorithm", Journal of Networks, vol. 9, no. 4, **(2014)**, pp. 1005-1011.

[11] D. Liu and S. Fan, "A modified decision tree algorithm based on genetic algorithm for mobile user classification problem", The Scientific World Journal, **(2014)**.

[12] W. Lo, Y. Chang, R. Sheu, C. Chiu and S. Yuan, "CUDT: a CUDA based decision tree algorithm", The Scientific World Journal, **(2014)**.

[13] T. Chopra and J. Vajpai, "Fault Diagnosis in Benchmark Process Control System Using Stochastic Gradient Boosted Decision Trees", International Journal of Soft Comput. Eng., vo. 1, **(2011)**, pp. 98-101.

[14] S. V. S. G. Devi, "Fuzzy Rule Extraction for fruit Data Classification", Compusoft, An international journal of advanced computer technology, vol. 2, no. 12, **(2013)**, pp. 400-403.

[15] A. Newman and D. Asuncion, "UCI Repository of Machine Learning Database", (School of Information and Computer Science), Irvine, CA: Univ. of California, **(2007)**.

[16] I. H. Witten and E. Frank, "Data Mining: Practical machine learning tools and techniques", Morgan Kaufmann, **(2005)**.

[17] J. R. Quinlan, "Induction of Decision Trees", Machine Learning, vol. 1, no. 1, pp. 81-106

[18] L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, "Classification and Regression Trees", Wadsworth International Group, Belmont, CA, Case Description Feature Subset Correct Missed FA Misclass, vol. 1, **(1993)**, pp. 1-3.