

An Enhanced Weighted Associative Classification Algorithm Without Preassigned Weight based on Ranking Hubs

Siddique Ibrahim S P¹

Research Scholar
School of Computing Science and Engineering
VIT University, Chennai, India

Dr. Sivabalakrishnan M²

Associate Professor
School of Computing Science and Engineering
VIT University, Chennai, India

Abstract—Heart disease is the preeminent reasons for death worldwide and in excess of 17 million individuals were kicked the bucket from heart disease in the past years and the mortality rate will be increased in upcoming years revealed by WHO. It is very tough to diagnose the heart problem by just observing the patient. There is a high demand in developing an efficient classifier model to help the physician to predict such threatening disease to recover the human life. Now a day, many researchers have focused novel classifier model based on Associative Classification (AC). But most of the AC algorithm does not consider the consequence of the attribute in the database and treat every itemsets equally. Moreover, weighted AC ignores the significance of the itemsets and suffering the rule evaluation due to support measure. In this proposed method we have introduced attribute weight, which does not require manual assignment of weight instead the weight would be calculated from link based model. Finally, the performance of the proposed algorithm is verified on different medical datasets from UCI repository with classical associative classification.

Keywords—Association rule mining; hub weight; classification; heart disease; attribute weight; associative classification

I. INTRODUCTION

Heart disease is also called as Cardio Vascular Disease (CVD) that occurs by various factors like high blood pressure, obesity, high cholesterol in the blood, high sugar level, and food habits. These conditions lead to formation of plaque in the blood path that slow down the blood flow to the heart muscle results in heart failure. One in four persons dies from heart disease in developed countries and the count will more in low and middle income countries throughout the world [1, 2]. Every day medical industries will generate large amount of patient's treatment data. This data can be analyzed to prevent the risky condition of heart disease by providing prediction at early stage and this will reduce the death rate. The data mining has applications in many fields, like education, production, inventory and bio medical which use techniques such as Regression, Clustering, Classification, Prediction, and Associative Classification (AC) plays a vital role in identifying the hidden knowledge for precautionary decision making. This motivated the researchers to develop computer aided heart disease prediction system to reduce the risk factors of the deadly disease.

Data mining is the process of learning new knowledge in terms of pattern from wide availability of database and help the information industries to make better decision on data which is based on the variety of techniques. It has wide variety of applications including banking, education, retail, e-commerce, bioinformatics, insurance and communication. Association Rule Mining (ARM) [3] is a data mining method that helps to find the frequent pattern between two or more items. It discovers a correlation among data repositories from datasets that satisfy the minimum support and confidence threshold using simple if then statements which is very easy to understand by normal users.

Classification technique is another widely used method in data mining technique. The important role of classification algorithm predict categorical class label for each given instance after construction of model in the database. For example, a classification build a model used to identify the potential customers of loan application risk level such as low, medium, and high. Several models have been projected by different researchers over the years to exactly guess the intention class. This task includes statistical [4], neural networks [5], divide and- conquer [6], decision tree [7, 8], PART [9], PRISM [10] and Naïve Bayes [7]. Association Classification (AC) has been successfully integrates association rules for classification that yield specific subset association rule whose consequent will be a class label [11]. Classification construct classifier from which new class label could be assigned to new records. A new Class based Association Rule (CBA) that enhances the learning phase to provide future data samples. Thus, the AC technique is the hybrid version of the association rule mining. For example, as the class rule R: AT1, AT2 -> class is inferred as follows: If the itemsets combination occurs together in a particular instance with Ci, then the right hand side Ci can be inferred as class label then AT1, AT2 are attribute values. Two basic metrics are in place to help each class association rules namely support and confidence user specified constraints. For an instance, the support of CBA A1, A2 ->C1 is described as number of occurrence that itemsets rule A1, A2 matches belongs to a class C1 in dataset D.

The paper is organized as follows: next section describes the preliminaries done for the proposed work. In Section 3 represent proposed weighted associative classification method which is explained along with sample computation. In

Section 4 presents with different analysis on heart disease and other medical and non-medical datasets. The final section concludes our work followed by future work.

II. RELATED WORK

Liu et al. [11] was introduced Class Association Rule (CBA) the first algorithm that incorporate the well-known Apriori algorithm [12] for class rule generation. It has rule generation stage uses well-known that scan multiple times in dataset to bring the frequent Class Association Rules (CARs). After rule generation the CBA uses minconf measure for selecting the best rules for classifier construction. Finally, the algorithm performs prediction in order to guess the unknown instance.

Wenmin et al. [13] introduced the concept of producing FP-tree based CAR. This method was attained maximum precision than CBA algorithm. Cowling et al. [14] was proposed multi-class and, multi-label algorithm to solve multiple scanning issue by introduced three different measures for evaluating the algorithm efficiency. The authors have used 28 different UCI datasets shows that MMAC performed better than traditional CBA algorithm.

The problem of generating Multi-Class Association Rule (MCAR) was proposed by Thabtah et al. [15] that maintains item occurrence position during the rule generation phase instead of rules. The author proved that the MCAR has limited number of rule generation in the initial stage.

Alwidian et al. [16] developed statistical ranking measure for enhancing the accuracy of CBA algorithm. Various algorithms were compared in performance evaluation section that showed the ECBA algorithm could provide better performance in terms of accuracy. Hadi et al. [17] developed fast associative classification based on Diffset method to improve the class rule generation. This algorithm used sorted order rule evaluation using confidence and support measure. Moreover, this algorithm formed rule based cluster from which the rule matching has been performed. The author showed the better result with other traditional AC algorithms.

Over the past few years, many researchers have been focused to finding efficient solution for heart disease prediction. Anbarasi et al. [18] executed the decision tree method of popular data mining techniques to obtain trusted value of accuracy in diagnosis of patients. Similarly, this was acquired previously by more number of attributes. The author uses genetic algorithm that plays a major contribution towards the diagnosis of heart disease by determining the correct attributes which will help both patients as well as physicians to reduce the number of test taken for diagnosing the disease. By using local search algorithm the author have reduced to 6 attributes from 13 attributes used in traditional algorithms.

Harleen et al. [19] inspired the important and essential part of classification technique such as Artificial Neural Network (ANN), Rule induction and decision tree to diagnosing the patient's health. Automated prediction system was proposed by Srinivas et al. [20] to enhance the diagnosis system and to reduce the medical expenses by implementing decision tree, Naive Bayes and neural network.

Latha et al. [21] developed a prototype for heart disease prediction system based on neuro-fuzzy system combined with Genetic Algorithm (GA) that were improved the prediction. Nidhi et al. [22] described decision tree algorithm that contains only 4 attributes for heart disease prediction. Here the author converted all the data into categorized form in order to reduce mean absolute error. Pethalakshmi et al. [23] projected the genetic algorithm with few numbers of features to increase classification accuracy. The author projected the fuzzy logic method that reducing the execution time and improving accuracy.

Alladoubaye et al. [24] model an intelligent heart disease prediction system using data mining techniques to predict heart disease from real datasets with 14 clinical features. This algorithm achieved 91.42% accuracy over SVM algorithm. Ankita et al. [25] used improved k-means and ANN algorithms to attained greater accuracy of 99%. Mohamed et al [26] had applied different data mining algorithm to predict medical diseases using five different algorithms, namely, C4.5, SVM, logistic regression, Neural network and Naive Bayes.

A. Weighted Association Rule Mining (WARM)

First we have discussed WARM in this section, further we explained the proposed EWAC using Ranking Hubs with suitable illustrations.

The traditional frequent itemset mining methods do not consider the importance of the individual items in the data set. The mining process is based on the count of its itemset in a database. But various real applications such as the business market and the clickstream require the significance of itemsets in the database. Various algorithms were presented in recent years to describe the quantitative aspects representing the knowledge for the itemset in a given database. Few among them are discussed below.

Wang et al. [27] proposed the WAR algorithm. They introduced weighted_support which represents the significance of the item in the database. Here, domain experts assign weights to every individual item and mining interesting rules form weighted_support measure. Cai et al. [28] break the traditional downward closure property by introducing distinct weight property. But this method consumes more time and expensive when the database is outsized. Wang et al. [29] eliminated the concept of assigning weights to the items by human experts. The author proposed a link-based ranking model that represents the association rules. The major drawback of this algorithm is, it finds only the rank and it does not attempt to calculate any measure like weighted_support to evaluate the mining process.

This is impressive as the domain expert assigns a higher weight to an uninteresting item in the transaction. But sometimes it leads to provide incorrect knowledge and misses the valuable itemsets. Moreover, it is impractical to manually assign the weights to the items when the dataset has a massive number of items with different fields. Hence, it becomes necessary to introduce a method to overcome the above drawbacks. Based on the above reasons, we proposed the enhanced weight calculation strategy which does not require domain expert interaction on datasets. It derives the weights

completely from the internal structure of the database and based on the assumption of good items consisting of higher weights.

III. PROPOSED EWACRH ALGORITHM

A. Enhanced Weighted Associative Classification with Ranking Hubs (EWACRH)

In this section, at first we describe the weighted associative classification and its significance followed by details of Hub based weighted scheme of HITS Model. Consider the Fig. 1 shows the proposed architecture.

B. Problem Definition

Let consider a database $D = \{At_1, At_2, At_3, \dots, At_n, C\}$ be a list of attributes and C is a class label in the database. If a rule states that $A_2 \rightarrow C_1$, while A_2 is attribute value then C_1 is the class attribute.

Algorithm: Enhanced Weighted Associative Classification

```

Dataset D with n Instances
Divide T {Traning_data T, Test_data T1}
Training_data T
{
S1=Empty_set, F1=Empty_set
K=1
Do{
For all transaction  $\epsilon$  T
Initialize auth(i) to 1
Hub(t)=
Auth(i)= hub(t)
S1=Generate all candidate k itemset rule(Training data T,
n, min_support
Weighted_support)
For each Itemset r1 in s1
Support(r1)=supportcount(r1)->class/N
weight_support=Authority_weight(r1)+support/k
If weighted_support(r1) min weighted_support
F1=S1+r
End if
Find all combination ruleset(rk)
End for
K=k+1
}
While Sk
Return F1'
}
    
```

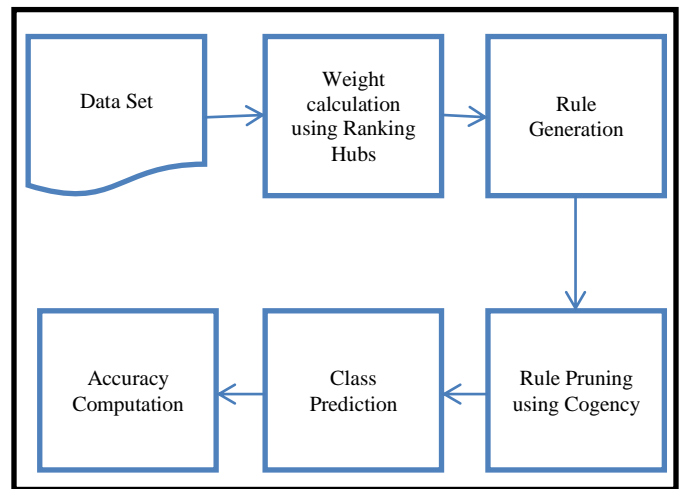


Fig. 1. Generating Enhanced Weighted Associative Classification using Ranking Hubs.

Example 1. Let consider a sample Heart Disease data description shown in below Table I.

C. Hub-based Weighting Scheme

The proposed Hub based weighting scheme constructs a bipartite graph from the values of the data set. It is known that the bipartite graph consists of two disjoint, independent set of vertices. These two sets are interconnected with the edges. In the proposed scheme, the unique row values (V) and all possible instances of attributes (U) form the two independent vertices. An edge is placed between any two vertices belonging to these subsets of vertices V and U, if and only if $U_i \subset V_i$. The degree of the vertices in set U is considered as the authority weight

This bipartite graph gives us an idea of applying link based approach for classification attributes. Here the AC takes the support values proportional to its degree of appearance in the database. Classical CBA [11] algorithm does not consider these different transaction significance during the support calculation. Moreover, the proposed method invoking the concept of hub and authority while calculating the itemset weight. The below equations are applied to the bipartite graph for computing authority weight in each iteration. In Fig. 2 presented the bipartite graph equivalent of Table I.



Fig. 2. Bipartite Graph.

TABLE. I. SAMPLE HEART DISEASE DATASET

Age	Restwm	Sex	Recurrence_Event
70..79	Mild	M	Yes
20..30	Severe	F	No
10..20	Dys	F	Yes
20..30	Mild	F	Yes
70..79	Dys	M	Yes
10..20	Severe	M	No
70..79	Mild	M	No

$$auth(i) = \sum_{T:i \in T}^n hub(T) \quad (1)$$

$$hub(T) = \sum_{i:i \in T}^n auth(i) \quad (2)$$

Example 2. Let consider the Table II for automated weight calculation using HITS model. The authority weight is listed for corresponding to the itemset.

TABLE. II. AUTHORITY WEIGHT OF THE EXAMPLE DATASET

Sl.No.	Attribute	Authority Weight
1	70..79	3
2	20..30	2
3	10..20	2
4	Mild	3
5	Severe	2
6	Dys	2
7	M	4
8	F	3

D. Weighted_Support Calculation

The following section illustrates how the weighted AC works based on automated HITS algorithm in step by step manner. Here, the weight does not assign by the subject experts.

S1. Generate single candidate weight

Computing support count and weighted_support of each itemset rules we will generate candidate single itemset in Table III. Here the min. weighted_support =4. Min. weighted_support is a predefined parameter for selecting active rulesets. Assigning larger value of this parameter leads to move restriction number of rule generation. One of the main advantages of the proposed methodology is to control the number of rule generation in AC by increasing min_weight parameter in each iteration. The following example shows the evaluation of weighted_support for ruleset r1.

$$\text{Weighted_support} = \frac{\text{Authority_weight}(r1) * \text{Support}}{K} \quad (3)$$

$$\text{Weight}(r1) = \frac{\text{Authority_weight}(r1)}{K} \quad (4)$$

For sample calculation: Let us consider first rule in Table XIII:

$$\text{Weight}(70..79) \rightarrow \text{yes} = 3/1$$

$$\text{Weighted_support}(70..79 \rightarrow \text{yes}) = 3/1 * 2 = 6$$

TABLE. III. SINGLE CANDIDATE RULESET

No. of Rules	Single Itemset	Support	Weighted_support
1	70..79->yes	2	3*2=6
2	70..79->no	1	3*1=3
3	20..30->yes	1	2*1=2
4	20..30->no	1	2*=2
5	10..20->yes	1	2*1=2
6	10..20->no	1	2*1=2
7	Mild->yes	2	3*2=6
8	Mild->no	1	3*1=3
9	Severe->yes	0	2*0=0
10	Severe->no	2	2*2=4
11	Dys->yes	2	2*2=4
12	Dys->no	0	2*0=0
13	M->yes	2	4*2=8
14	M->no	2	4*2=8
15	F->yes	2	3*=6
16	F->no	1	3*1=3

S2. Frequent single class ruleset will be generated which contains the ruleset has a weighted_support greater than or equal to the specified minimum weighted_support as given in Table IV.

S3. Find the two candidate itemset based on the support and weighted_support as shown in below Table V. The frequent itemset for data's in Table V is given in Table VI.

TABLE. IV. FREQUENT SINGLE RULESET

No. of Rules	Frequent Single Itemset	Weighted support
1	70..79->yes	6
2	Mild->yes	6
3	Severe->no	4
4	Dys->yes	4
5	M->yes	8
6	M->no	8
7	F->yes	6

TABLE. V. TWO CANDIDATE ITEMSET

No. of Rules	Two Itemset rules	Support	Weighted support
1	70..79, Mild->yes	1	3*1=3
2	70..79, Dys->yes	1	3*1=3
3	70..79, M->yes	2	4*2=8
4	70..79, F->yes	0	3*0=0
5	Mild, Dys->yes	1	3*1=3
6	Mild, M->yes	1	4*1=4
7	Mild, F->yes	1	3*1=3
8	Dys, M->yes	1	3*1=3
9	Dys, F->yes	1	3*1=3
10	M, F->yes	0	Ignore same instances
11	Severe, M->no	1	3*1=3

TABLE. VI. FREQUENT TWO ITEMSET

1	70..79, Mild->yes	1	3*1=3
2	70..79, Dys->yes	1	3*1=3

S4. Find the three candidate itemset based on the min_support and weighted_support as shown in Table VII.

In the above Table VII there is no ruleset has sufficient support count that satisfy the min_weighted_support. Hence the ruleset process is stops by the algorithm and start for preparing useful ruleset.

S5. Find the useful ruleset based on minimum confidence (i.e ≥ 0.5) in the frequent CAR and remove all other ruleset as shown in Table VIII.

Cogency is posteriori based mathematics [30, 31] that conclude the confabulation if the assume facts being true.

S6. For sample cogency computation: Let consider the first rule in the below Table IX.

(70..79->yes)

$1 * \text{weighted_support} * \text{confidence} / \text{count of yes}$

$1 * 6 * 0.67 / 4 = 1$

TABLE. VII. THREE CANDIDATE ITEMSET

No. of Rules	Frequent Two Itemset rules	Support	Weighted support
1	70..79, Mild, M->yes	1	3*1=3

TABLE. VIII. RULESET SATISFY THE CONFIDENCE MEASURE

No. of Rules	Itemset rules	Support	Confidence
1	70..79->yes	2	0.67
2	Mild->yes	2	0.67
3	Severe->no	2	0.5
4	Dys->yes	2	0.5
5	M->yes	2	2
6	M->no	2	2
7	F->yes	2	0.67
8	70..79->yes	2	0.67
9	Mild->yes	1	0.5

TABLE. IX. COGENY COMPUTATION

No. of Rules	Frequent Two Itemset rules	Confidence	Weighted support	Cogency
1	70..79->yes	0.67	6	1
2	Mild->yes	0.67	6	1
3	Severe->no	0.5	4	0.66
4	Dys->yes	0.5	4	0.5
5	M->yes	2	8	4
6	M->no	2	8	5.33
7	F->yes	0.67	6	1
8	70..79, M->yes	0.67	8	1.34
9	Mild, M->yes	0.5	4	0.5

TABLE. X. SORTED RULESET

No. of Rules	Frequent Two Itemset rules	Confidence	Weighted support	Cogency	Rank
1	M->no	2	8	5.33	1
2	M->yes	2	8	4	2
3	70..79, M->yes	0.67	8	1.34	3
4	70..79->yes	0.67	6	1	4
5	Mild->yes	0.67	6	1	5
6	F->yes	0.67	6	1	6
7	Severe->no	0.5	4	0.66	7
8	Dys->yes	0.5	4	0.5	8
9	Mild, M->yes	0.5	4	0.5	9

S7. Table X shows the sorted CAR rules based on cogency measure. If more than one rule has the same cogency value then the rule will be sorted based on confidence measure. If both the rules have the same cogency and confidence then considers the rules weighted_support followed by length of the rule for further decision.

S8. Suppose one can predict the class of unknown rule with following values {70..79, Mild, Y}. The new rules selects based on cogency of (Ri) > Rj elseif $\text{conf}(ri) > \text{conf}(Rj)$ or $\text{conf}(Ri) = \text{conf}(Rj)$ and $w.\text{sup}(Ri) > w.\text{sup}(Rj)$ or $\text{conf}(Ri) = \text{conf}(Rj)$ and $w.\text{sup}(ri) = w.\text{sup}(Rj)$ and $\text{sup}(Ri) > \text{sup}(Rj)$ [22]. The algorithm have found the following potential rules for given new instances: 70..79->yes, Mild->yes, Y->yes. Hence the rule can be assigned with 'yes' class.

IV. EVALUATION OF WEIGHTED ASSOCIATIVE CLASSIFICATION

The proposed system was tested using heart disease dataset and several benchmark UCI repository [32] data sets and brief description is presented in Table XI. The performance of the proposed hub based weighted AC algorithm was evaluated by comparing with three well known algorithms are CBA, CMAR, and MCAR. All the datasets were implemented using JAVA programming. All the experiments were carried out on a computer with Windows 7 OS 64 bit, AMD A4-3330 with clock rate of 2.40GHz with 8GB Physical memory. In the dataset, Holdout approach [33] was used 70% of the datasets were used as training data and the remaining 30% of the datasets were used as testing data. The system performance was examined by different metrics such as accuracy, precision and recall for heart disease and other datasets. The dynamic minimum threshold set for weighted support and confidence parameters were used for testing all the algorithms. Tables XIII and XIV representing the results of precision and recall on heart disease dataset. Tables XV and XVI shows the accuracy for the different medical datasets and non-medical datasets using our proposed approach.

The proposed algorithm is applied to the heart disease dataset and obtained the results. From results in Tables XII, XIII and XIV, it is inferred that the proposed algorithm, EWACRH outperformed the other with an accuracy of 97%, precision of 95%, and recall of 93% respectively.

TABLE. XI. DATASET DESCRIPTION

Dataset	No. of Transaction	No. of Classes
Heart Disease	303	2
Breast Cancer	286	2
Breast -W	699	2
Diabetes	768	2

TABLE. XII. ACCURACY OF HEART DISEASE DATASET

Algorithm	Accuracy
CBA	87
CMAR	88
MCAR	91
EWACRH	97

TABLE. XIII. PRECISION OF HEART DISEASE DATASET

Algorithm	Precision
CBA	83
CMAR	82
MCAR	87
EWACRH	95

TABLE. XIV. RECALL OF HEART DISEASE DATASET

Algorithm	Recall
CBA	80
CMAR	81
MCAR	85
EWACRH	93

TABLE. XV. ACCURACY OF MEDICAL DATASET

Datasets	CBA	CMAR	MCAR	EWACRH
Heart Disease	87	88	91	97
Breast Cancer	77	75	80	91
Breast-w	80	77	75	90
Diabetes	78	77	79	94
Avg. Accuracy	81	79	81	93

TABLE. XVI. ACCURACY OF DATASETS (NON-MEDICAL)

Data Sets	CBA	CMAR	MCAR	EMACRH
Wine	71	71	72	90
Iris	86	89	89	99
Car	73	75	75	95
Bank	71	76	84	89
Avg. Accuracy	75	78	80	93

The Hub based weighting scheme proposed in this paper aids the EWACRH algorithm in getting a better result. Due to this scheme, the attributes influencing the decision get higher priority, which leads to an accurate decision. In heart disease dataset, age, sex, chest pain, restecg, heart muscle condition, and resting blood pressure are the decision influencing attributes. Another notable advantage of the proposed algorithm is, it considers the unassociated items also in its decision. From the obtained results, rules are framed. The cogency measure [30] used in this paper helps to prune the rules.

A. Accuracy Computation

Accuracy can be distinct as the amount of uncertainty in a measurement with respect to an absolute standard. It means the proportion of correct results that a classifier achieved. If the classifier can correctly predict the class label of half of the given datasets, then the accuracy is 50%. We used the following metrics:

True Positive (TP): Number of positive samples, labeled as positive

False Positive (FP): Number of negative samples, labeled as positive

True Negative (TN): Number of negative samples, labeled as positive

False Negative (FN): Number of positive samples, labeled as negative

We can calculate the accuracy as follows:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (5)$$

Fig. 3 and 6 shows the graphical representation of proposed algorithm accuracy for heart disease and other medical datasets. Also, Fig. 4 and 5 shows the precision and recall results of our proposed algorithm on heart disease dataset. The prediction accuracy of our proposed algorithm for different non-medical datasets, compared with those of some existing models is presented in Fig. 7.

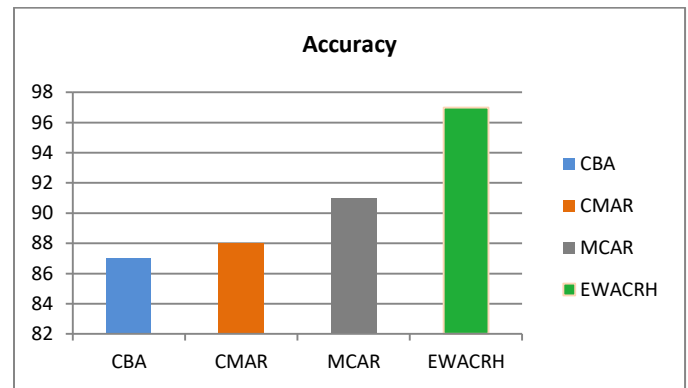


Fig. 3. Comparisons of different Rule mining Algorithms on Heart Disease Dataset.

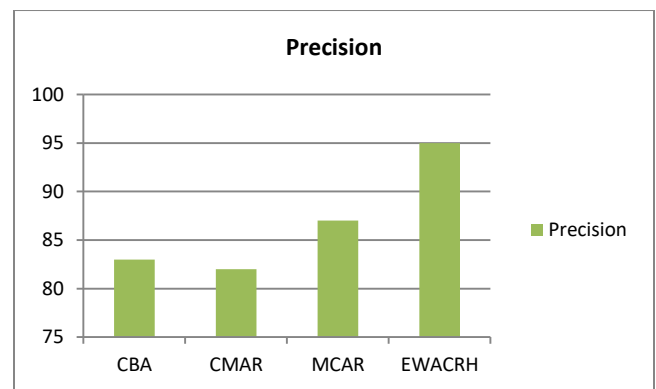


Fig. 4. Precision of Heart Disease Dataset.

V. CONCLUSION

In this paper, we proposed an enhanced weight measure scheme for heart disease prediction. This proposed algorithm uses hub ranking and the cogency measure to improve the performance. The weighted AC algorithm has applied the weight on each fields in the database to prioritize the ruleset from the least important one based on Hub ranking. In the future, this algorithm EWACRH can be evaluated by applying it to different domain data sets. It is also planned to use the memetic algorithm for investigating different weighting, pruning, and ranking.

VI. DATA AVAILABILITY

<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29>

<https://archive.ics.uci.edu/ml/datasets/Diabetes>

<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

<https://archive.ics.uci.edu/ml/datasets/Iris>

<https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

REFERENCES

- [1] <https://medlineplus.gov/heartdiseases.html>.
- [2] Heart Disease and Woman : www.womenshealth.gov/files/images/hds-heart-disease-and-women-1.jpg
- [3] Rakesh A, and Tomasz I., "Mining Association Rules between sets of Items in large databases" Proceed. of the ACM SIGMOD conference, USA, 1993.
- [4] Kumar V, and Wu "Top 10 algorithms in data mining" KIS(2008), Springer series, 2007.
- [5] Yeh Ivy and Che-Hui Lien., "Cosmetics purchasing behavior-An analysis using association reasoning neural networks" Elsevier., volume.37, No. 10, PP.7219-7226, Oct. 2010.
- [6] Johannes F., "Separate and Conquer Rule Learning", Kluwer Academic publisher, PP. 3-52, 1999.
- [7] Ross Quinlan J., "C4.5: programs for Machine Learning" Artificial Intelligence – Springer, Morgan Kaufmann Publisher, 1994.
- [8] George John H, and Langlay P., "Estimating Continuous Distributions in Bayesian Classifiers" Machine Learning, PP. 338-345, Feb 2003.
- [9] EFrank E and Ian H., "Generating accurate rule sets without global optimization" - Proceeding of the Inter. Conference on machine learning (ICML 98), PP. 144-151. July 1998.
- [10] Jadzia C., "PRISM: An Algorithm for inducing modular rules" Int. journal of Man-Machine Studies, PP. 349-370, vol.27, 1987.
- [11] Bing Liu, and Hsu W., "Integrating Classification and association rule mining" (KDD), 1998.
- [12] Rakesh A, and Ramakrishnan S., "Fast Algorithms for mining association rule" Proceed. of the twentieth Int. conference on VLDB, PP. 487-499. 1994.
- [13] Wenmin Li, Jiawei H and Jilan P., "CMAR: Accurate and Efficient Classification based on Multiple-class Association Rule" IEEE Int. conference on data Mining, PP. 360-368. 2001.
- [14] Fadi G, Peter Cowling and Yonghon P., "MMAC: A New Multi-Class, Multi-Label Associative Classification Approach" IEEE Int. conference on data Mining, 2004.

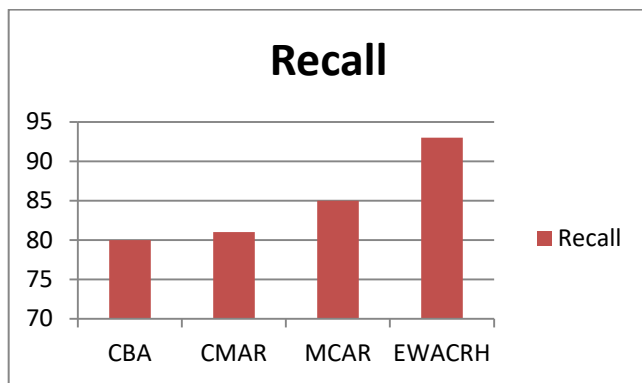


Fig. 5. Recall of Heart Disease Dataset.

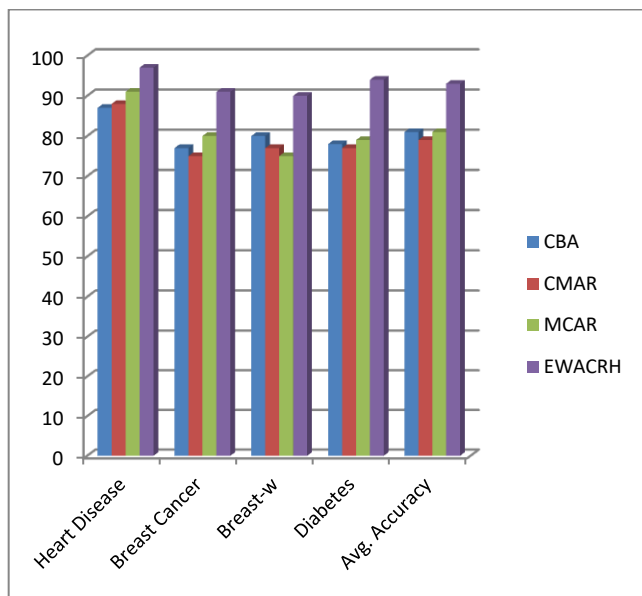


Fig. 6. Comparisons of different rule Mining Algorithms on Medical Datasets.

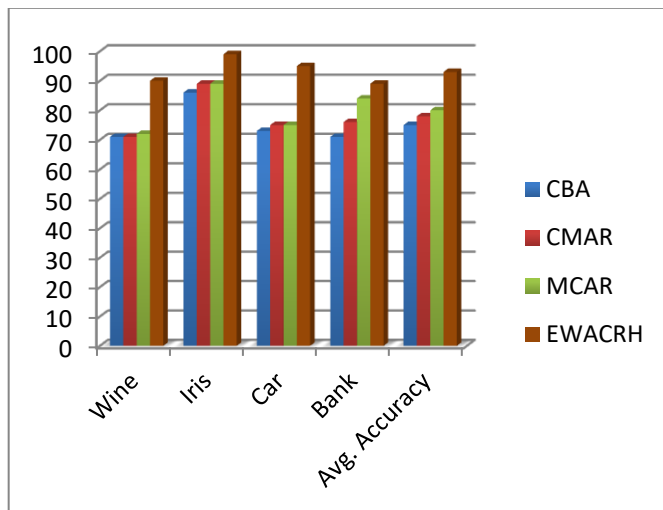


Fig. 7. Comparisons of different rule Mining Algorithms on Non-Medical Datasets.

- [15] Fadi T, and Peter I. "Multi-Class Classification based on Association Rule" IEEE Int. conference on Computer systems and Applications (AICCSA), PP. 33-40, 2005.
- [16] Jaber A, and Bassam H., "Enhanced CBA algorithm based on Apriori Optimization and statistical Ranking Measure", 28th Int. conference on vision 2020 (IBIMA), Nov. 2016.
- [17] Waai Hadi, and Faisal A., "A new fast associative classification algorithm for detecting phishing websites", vol. 48, No. 16, PP. 729-734, Nov. 2016.
- [18] Ambarasi M., "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, IJESI, Vol. 2, 2010.
- [19] Harleen K, Siri K., "Empirical study on applications of data mining techniques in healthcare", Int. journal of computer science, PP.194-200, 2006.
- [20] Srinivas K, and Kavitha R., "Applications of data mining techniques in health care and prediction of heart attack", Int. Journal of Computer Sci. and Engg, vol.02, pp.250-255, 2010.
- [21] Latha Parthiban, Subramanian R., "Intelligent heart disease prediction system using CANFIS and genetic algorithm", Int. journal of medical and health sciences, Vol.01, No:05, PP.278-281, 2007.
- [22] Nidhi B, and Kiran J., "An analysis of heart disease prediction using different data mining techniques", IJET, Vol. 1, PP. 1-4, Oct. 2012.
- [23] Pethalakshmi A., "Effective feature selection via featuristic genetic on heart data", Int. journal of computer science and engineering, Vol. 7, June 2012.
- [24] Alladombaye N, and Lin Lei., "Comparative Study of Datamining Techniques on Heart Disease Prediction System: "Republic of chad" " Int. Journal Of Sci. and Research (IJSR), PP. 6-391, May 2016.
- [25] Ankita R., "Heart disease prediction using ANN and improved k-means" IJET, Vol.19, PP. 3081-3085, Sep 2017.
- [26] Basker M., "An intelligent and electronic system based classification and prediction for heart disease diagnosis", Int. Journal of Emerging Trends and Technology(IJETT) in Computer Science, May 2016.
- [27] Wei W, and Jiong Y., "Efficient Mining of Weighted Association Rules (WAR)" Sixth Int. conference of ACM SIGKDD, PP. 270-274, Aug. 2000.
- [28] Cai CH, and AWC Fu., "Mining Association rules with weighted items", Proc. IDEAS 98, July 1998.
- [29] Wang Ke., "Item selection by "hub-authority" profit ranking", Int conference on KDD 2002, PP. 652.657.
- [30] Robert., "Confabulation Theory"Artificial Intelligence Springer Series. pp. 419-426. 2006.
- [31] Robert Hecht., "Cogent Confabulation" Neural Network in Elsevier, vol. 18, Issue. 2, PP. 111-115, 2005.
- [32] Machine-learning-databases/heart-disease UCI: <https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/>
- [33] Reitermanova Z., "Data Splitting" proceeding of Contributed Papers WDS-10, Part 1, PP. 31-36, 2010.