International Journal of Engineering &Technology, 7 (3.12) (2018) 754 -758

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Implicating Machine Learning towards Development of Intelligent System for Wart Treatment Therapy Identification

Sabita Khatri[1], Deepak Arora[2], Anil Kumar[3]

Dept. of Computer Science & Engineering, Amity School of Engineering & Technology
Amity University, Lucknow, India
**Corresponding Author Email: [1]sabitaarora@gmail.com*

Abstract

Warts are produced on the human body because of contamination caused by Human Papillomavirus (HPV). The most affected areas of warts are hands and feet specifically, which is bit annoying and hard to recover in later stages. After massive literature survey, it is found that different treatments have been suggested for treating this illness. The basic problem found while treating that the treatment given to one patient may or may not be suitable for another patient, so it is hard to detect specific kind of treatment to be followed for faster recovery and effective customized treatment against this viral disease. The objective of this research work is to identify the ideal treatment method for both particular plantar and normal warts, between immunotherapy and cryotherapy treatment methods. Implications of machine learning techniques are now playing a vital role specifically in clinical diagnosis toward identifying different clinical patterns, disease classification and its predictions. In this research, work authors have implemented classifiers like Bayes Net, SVM, Multi-Layer Perceptron, k-NN, FURIA, Random Forest with the help of WEKA tool. The experimentation has been performed on data sets obtained from UCI Machine Learning Repositories. The experimentation was performed with total 180 patient instances having wart illness present in immunotherapy and cryotherapy datasets respectively. The result outcomes have been discussed and compared with existing methodologies mentioned in the literature. It was observed that the decision tree based classifier random forest is having the best classification accuracy among the chosen set of classifiers. The result shows highest classification accuracy in case random forest, 86% and 93% was noted for immunotherapy and cryotherapy treatment methods datasets. This research work is helpful for physicians in selecting the best treatment method for their patient suffering from wart illness in order to reduce overall treatment cost incurred and also improving the quality of treatment given to the patients.

Keywords: *Warts, Immunotherapy, Cryotherapy, Machine Learning.*

1. Introduction

Warts are noncancerous skin developments that grow on a various piece of the body in different forms and can go up against different structures. There are hundred different types of HPV identified reported in the literature till date with the help of clinical trials. Warts are infectious events that disrupt the normal epithelial barrier increase the likelihood of developing warts, so the treatment of this illness might be troublesome for some patients. Warts automatically can also get resolved inside a couple of years however some high-risk HPV subtypes are related to malignancies. The verrucous carcinoma is a moderate developing tumor and frequently confused for a typical wart. Despite the fact that it can happen on any piece of the body, it is most normal on the plantar surface. A verrucous disease once in a while spreads; however, it is locally ruinous. The dominant part of warts doesn't cause side effects but in some cases, they do cause corrective deformation and in the rare patient may be painful. Wart can be diagnosed based on patient examination. Once the conclusion is made, the treatment relies on symptoms, patient preferences, and cost of treatment.

Despite the fact that there are numerous medications for warts, but none is extremely powerful and repetition of treatment is also required with each of them. Cryotherapy, retinoic corrosive, podophyllin, topical 5-fluorouracil, interferon, and imiquimod are additionally utilized approaches for wart treatment. Better clinical results can be assured by early disease prediction and the selection of best suitable treatment for this disease. The artificial intelligence system is at present well appropriate for analytical studies. Specifically, a great deal of work has been done in therapeutic conclusion in most analytical cases. So usage of such frameworks in clinical diagnosis is growing day by day. Through a massive literature survey, it is found that substantial work on data mining has been performed and remarkable knowledge generated over the past decades, and implications of machine learning techniques towards early medical diagnosis can help in making best clinical decisions. In this paper, selected algorithms have been considered for classification purpose for immunotherapy and cryotherapy treatment dataset. The algorithms like Bayes Net, SVM, Multi-Layer Perceptron, k-NN, FURIA, Random Forest are major classifiers used to perform this study on WEKA.

In most of the studies, Bayes classifier is found to be quite effective when used with the large datasets. It uses probabilistic approach in order to classify data based on available attributes.
Another classifier support vector machine uses the concept of decision boundaries. The isolation is performed with the help of set of objects, belongs to particular class. Bayes algorithm and

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

rule-based algorithm and decision tree have also been evaluated. k-NN is useful where the instances are unknown and do not have similar behavior. In k-NN instances are classified based on relation similarity function to the known objects. Random forest follows supervised learning methodology in order to perform classification. This algorithm is based on divide and conquers approach. Khozeimeh et al. discussed the applicability of fuzzy rules for selecting the best suitable method for treatment of plantar and common type of warts. The study was performed on 180 patients, out of which cryotherapy method along with liquid nitrogen has been implemented on 90 patients whereas other 90 patients were given immunotherapy. The authors found that fuzzy logic rule-based system as most suitable algorithm for wart treatment identification method having classification accuracy of 83.33% and 80.7% for immunotherapy and cryotherapy method respectively.

In the current research work Bayes Net, SVM, Multi-Layer Perceptron, k-NN, FURIA, Random Forest classifiers have been considered for the experiment purpose. All selected classifiers have been examined on immunotherapy and cryotherapy datasets taken from UCI learning repositories. During experimentation, the random forest has given better performance over other machine learning algorithms. The efficiency of these algorithms is compared based on performance measures like MAE, Root Mean Square Error MSE, Kappa values and ROC statistics. In this research work, recall, precision and F-Measures have also been computed and considered for performance comparison. It has been found, that tree-based approach such as Random Forest classification algorithm is giving classification results with a higher degree of accuracy and prediction rate among all chosen classifiers overtaken wart treatment method dataset.

2. Literature Survey

A huge amount of heterogeneous clinical data is being generated through various sources. These huge repositories are very helpful in extracting the meaningful knowledge patterns to construct a prediction model for the various life threatening disease at the early stages of its development. Data mining is a powerful tool to extract the knowledge from raw data. In this context, various challenges and techniques have been proposed by different researchers [1, 9]. A predictive approach has been developed for human breast and colon cancer with two-stage strategy namely preprocessing stage MT is used to increase the samples of a minority class and balancing the dataset. In predictor class, KNN and SVM machine learning algorithms are used to develop hybrid MTD-SVM and MTD-KNN models [10-11]. By taking the advantages of available technology prediction model can be developed for breast cancer, through recent computational methods and tools [3]. In this regard, data mining is one of the most viable solutions, to deal with the analysis of such type of clinical data and successive prediction models [4]. The voluminous amount of medical data is available for construction of different prediction model but one has to deal with heterogeneity and privacy aspects present in the data [2].

Attribute selection plays a most vital feature in cardiovascular disease, feature selection. Six different classifiers are implemented and found that use of MFS noticeably improved the performance and accuracy [6, 7]. A fuzzy rule, based clinical expert system is also found more significant specifically for heart disease and lung cancer prediction model in comparison to the neural network-based system [5, 13]. Association rule is also found to be the best approach for designing computational intelligence system to predict heart disease [8]. A hybrid method can also improve accuracy levels for coronary artery disease [12]. In Chronic Kidney Disease predictions can also be made by implicating machine learning algorithms. The machine learning models may use both approaches of human physiology and drug pharmacology for predicting most suitable treatment method. Machine learning approach has been introduced to predict the response of immunotherapy and cryotherapy methods on 180 patient instances and found a fuzzy rule-based approach to be the best with a higher degree of accuracy [14].

3. Datasets and Experimental Setup

For their experimentation, authors have considered Immunotherapy and cryotherapy datasets collected from UCI [14]. Detail information about input variables of this dataset including 08 attributes is shown in table 1. The total 90 patient instances are present on immunotherapy datasets and 90 instances available in cryotherapy dataset. There are total 08 and 07 features are available with immunotherapy and cryotherapy datasets respectively. The information regarding attributes and its type in immunotherapy and cryotherapy datasets have been mentioned table 1-2 respectively:

Table 1: Attribute detail in immunotherapy dataset

S.No.	Attribute	Type of Attribute
1	Sex	Numeric
2	Age	Numeric
3	Time	Numeric
4	Number_ of_warts	Numeric
5	Type	Numeric
6	Area	Numeric
7	Induration_diameter	Numeric
8	Result of treatment	Nominal

Table 2: Attribute detail in cryotherapy dataset

S.No.	Attribute	Type of Attribute
1	Sex	Numeric
2	Age	Numeric
3	Time	Numeric
4	Number_ of_warts	Numeric
5	Type	Numeric
6	Area	Numeric
7	Result of treatment	Nominal

As in both datasets, there is no any missing value, so all instances have been considered. The attributes like patient's sex, age, time, number of warts present, type of wart, area, initial induration diameter test and the result of treatment for immunotherapy and cryotherapy have been considered for the experimentation purpose. The last class attribute defines the response of the patient to the particular treatment method applied in the form of yes or no. Authors have used WEKA 3.8 tool for their experimentation purpose. In this experiment, authors have used eight different classifiers named as Bayes Net, SVM, Multi-Layer Perceptron, k-NN, FURIA, Random Forest. These classifiers have been implemented on wart treatment method dataset with 10 fold cross-validation approach. Performance has been evaluated on the basis of standard performance measures and classification characteristics like accuracy, Kappa statics, ROC, MAE, RMSE as well as TP rate, TN rate, precision, recall, and F-measure and discussed in this paper.

4. Results and Discussion

The result of this analysis for individual metrics was found very close for all classifiers applied during the experimentation using ten-fold cross-validation on weka. The result of the comparison is represented in table 3 and 4. It can easily be noted that random forest is the best suitable classifier as it shows a high degree of accuracy and kappa values. The value of kappa statistics varies from zero or one, the classifier having K value more than zero, shows good chance of agreement, as well as ROC value is also on higher side and error parameter like MAE and RMSE must be on the lower side as shown in comparison graphs of all classifier in figure 1-8 for immunotherapy and cryotherapy datasets.

Table 3: Classification Results for Immunotherapy Dataset

Algorithm	Accuracy	Kappa Statistics	MAE	RMSE	ROC
Bayes Net	85.5556	0.5578	0.226	0.3433	0.706
SVM	78.8889	0	0.2111	0.4595	0.5
ML Perceptron	80	0.321	0.1964	0.388	0.759
k-NN	70	0.0441	0.3048	0.5412	0.521
FURIA	82.2222	0.4217	0.1804	0.4066	0.755
Random Forest	86.667	0.5267	0.238	0.353	0.777

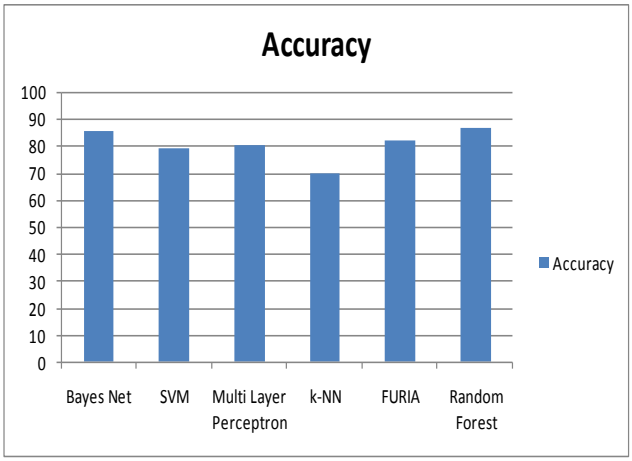

Fig. 1: Accuracy for Immunotherapy Dataset

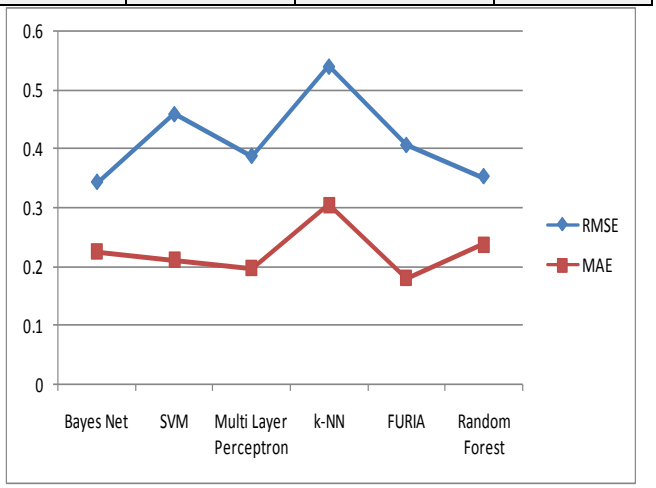

Fig. 3: MAE and RMSE for Immunotherapy Dataset

In the case of immunotherapy treatment dataset, the value of kappa statistics value 0.5578 as in table 3 is higher for Bayes Net whereas ROC value 0.777 is found higher for Random Forest algorithm in figure 2. The MAE value as in table 3 for immunotherapy treatment shows lower for FURIA and RMSE also minimum for Random Forest, as depicted in table 3. The Precision, Recall, True Positive Rate, F-Measure values for Random Forest is shown in table 4 and found better than over other chosen set of classifiers.

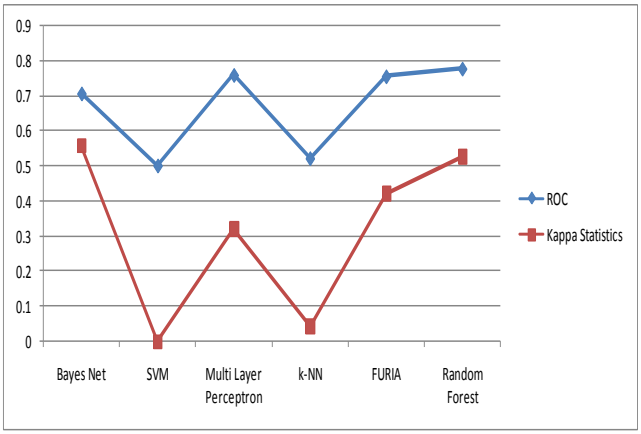

Fig. 2: ROC & Kappa Statistics for Immunotherapy Dataset

Table 4: Classification Accuracy Measures for Immunotherapy Dataset

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure
Bayes Net	0.856	0.308	0.853	0.856	0.854
SVM	0.789	0.789	0.622	0.789	0.696
Multi Layer Perceptron	0.8	0.516	0.78	0.8	0.785
k-NN	0.7	0.658	0.682	0.7	0.69
FURIA	0.822	0.433	0.81	0.822	0.814
Random Forest	0.867	0.421	0.862	0.867	0.852

Table 4 shows that Random Forest algorithm has the highest value of accuracy along with high kappa statistics and ROC value. This can also be drawn from results mentioned in table 5-6 and figure 5 and 6 for cryotherapy treatment method. Figure 7 depicted that MAE is found lower for fuzzy logic (FURIA) however the RMSE value is found, minimum for Random Forest for cryotherapy. Figure 8 is representing the classification measures values for True Positive, False Positive, Recall, Precision and F-measure for cryotherapy dataset.

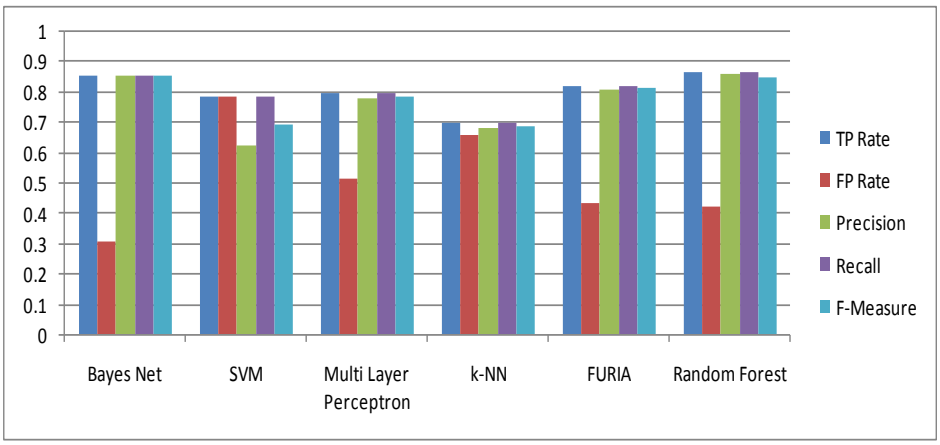

Fig.4 : Performance Measures Comparison for Immunotherapy Dataset

Table 5: Classification Results for Cryotherapy Dataset

Algorithm	Accuracy	Kappa Statistics	MAE	RMSE	ROC
Bayes Net	84.4444	0.6856	0.1897	0.3168	0.941
SVM	82.2222	0.6353	0.1778	0.4216	0.811
M L Perceptron	87.778	0.7556	0.1365	0.304	0.926
k-NN	90	0.7994	0.1096	0.3126	0.902
FURIA	91.1111	0.8225	0.1003	0.3045	0.911
Random Forest	93.3333	0.8669	0.143	0.2346	0.981

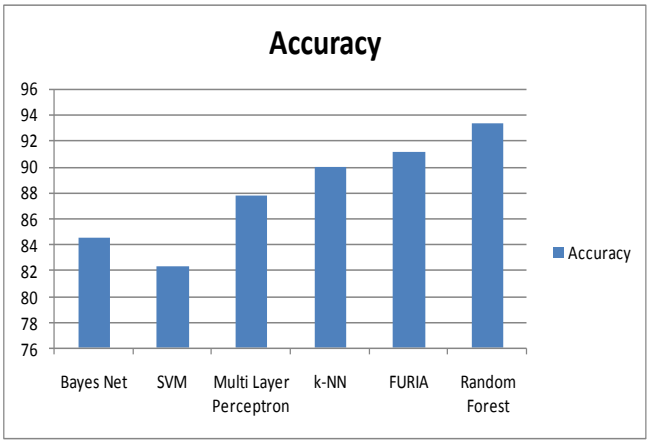

Fig. 5: Accuracy for Cryotherapy Dataset

If we compare the experimental results of this study with earlier results mentioned in the literature, authors found 86% and 93% as highest classification accuracy levels for immunotherapy and cryotherapy treatment methods respectively. It shows the increase of 3% in case of immunotherapy and 13 % in case of cryotherapy treatment method with decision tree based classifiers for wart treatment method identification.

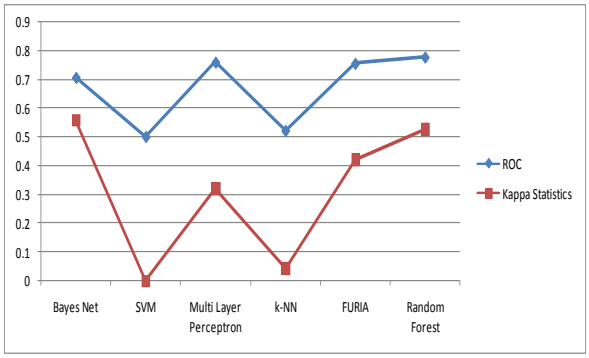

Fig. 6: ROC & Kappa Statistics for Cryotherapy Dataset

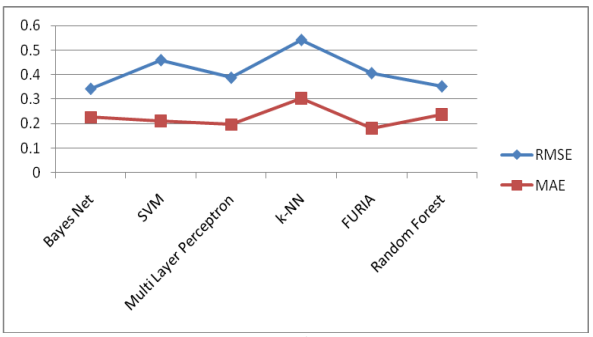

Fig. 7: MAE and RMSE for Cryotherapy Dataset

Table 6: Classification Accuracy Measures for Cryotherapy Dataset

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure
Bayes Net	0.844	0.163	0.846	0.844	0.844
SVM	0.822	0.2	0.854	0.822	0.816
ML Perceptron	0.878	0.119	0.88	0.878	0.757
k-NN	0.9	0.099	0.9	0.9	0.8
FURIA	0.911	0.084	0.915	0.0911	0.911
Random Forest	0.933	0.061	0.937	0.933	0.933

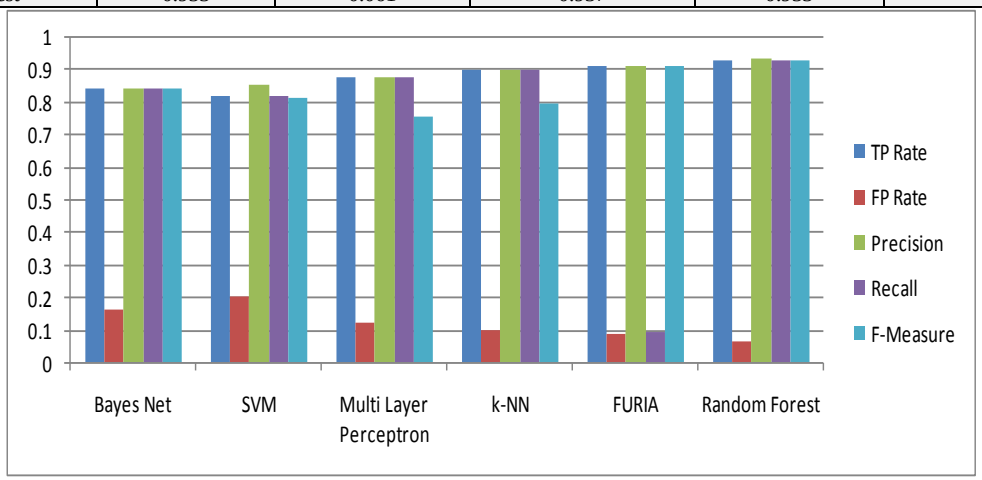

Fig. 8: Performance Measures Comparison for Cryotherapy Dataset

The objective of this research work is to identify the best possible treatment method, which leads to the fact that further advanced algorithm can be developed on the basis of the decision tree in order to get better treatment method with higher accuracy rate. This research work establishes the fact that Random Forest classifiers having better accuracy results over others. In total eight different classifiers Bayes Net, SVM, Multi-Layer Perception, k-NN, Fuzzy Logic (FURIA), Random Forest, have been analyzed and implemented in this research work.

5. Conclusion and Future Scope

In this paper, authors have used majorly six classifiers from different classification families for selecting best warts treatment

method. This approach has been compared with other approaches mentioned in the existing literature and found most significant improvements over them. The immunotherapy and cryotherapy dataset from UCI has been chosen in order to perform experimentation for this study. In this experimentation, authors have considered significantly six distinct classifiers including Bayes Net, SVM, Multi-Layer Perceptron, k-NN, FURIA, Random Forest. The two distinctive warts treatment method datasets have been considered with all most equivalent occurrences of total 180 instances of patient responses. The present research work uncovers that choice of tree-based classifiers is performing great on account of considered datasets from UCI repository. Random forest tree is demonstrating better ROC trends and class forecast precision, however, the MAE and RMSE values are bit higher side indicating some less encouraging outcomes. The Kappa statistics are concern random forest is demonstrating better outcomes. This experimentation shows that decision tree based classifiers to be considered as a best to build prediction models for ailment related therapeutic datasets. Further, this research work can be extended towards the development of hybrid algorithms which performs classification using decision tree through ensemble learning approach.

References

[1] I.Kononenko, Machine learning for medical diagnosis: history,state of the art and perspective, Artif. Intell.Med23(2001) 89-109.

[2] K.J Cios, G.W. Moore, Uniqueness of medical data mining. Artif. Intell. Med. 26(2002)1-24.

[3] D.Delen, G.Waller, A.Kadam, Prediction breast cancer survivability a comparison of three data mining methods, Artif.Intell.Med 34(2006) 113-127.

[4] R.Bellazzi, B.Zupan, Predicting data mining in clinical medicine: current issues and guidelines, Int. J. Med.Inform. 77(2008), 81-97.

[5] Anooj P.K, Clinical decision support system:Risk level prediction of heart disease using weighted fuzzy rules.J.King Saud Univ.Comput.Inf.Sci.24(2012)27-40.

[6] S.Swati, G.Ashok,Feature selection for medical diagnosis: Evalution for cardiovascular disease. Expert Syst.Appl.40 (2013)4146-4153.

[7] N. Jesmin, K.Tasadduq lmam,TS. Kevin, PC.Yi-Ping, Computation intelligence for heart disease diagnosis: A medical knowledge driven approach. Expert Syst.Appl.40(2013)96-104.

[8] N. Jesmin, K.Tasadduq lmam,TS. Kevin, PC.Yi-Ping, Association rule mining to detect factors which contribute to heart disease in males and females. Expert Syst.Appl.40(2013)1086-1093.

[9] N.Esfandiary, M.Babavalian, A.Moghadam, V.Tabar, Knowledge discovery in medicine: Current issue and future trend.Expert Syst.Appl.41(2014)4434-4463.

[10] A.Majid, S.Ali, M.Iqbal, N.Kausar,Prediction of human breast and colon cancers from imbalanced data using neareat neighbor and support vector machines.Comput.Methods Prog. Biomed.113(2014) 792-808.

[11] C. Barbieri, F.Mari, A.Stopper, E.Gatti, P.Escandell-Montero, JM.Martinez-Mzrtinez, JD.Martin-Guerrero, A new machine learning approach for predicting the response to anemia treatment in a large cohort of End Stage Renal Disease patients undergoing dialysis.Comput Biol Med.61(2015)56-61.

[12] L.Verma, S.Srivastava , P.C.Negi, A hybrid data mining model to predict coronary artery disease cases using non- invasive clinical data, J.Med.Syst.40(2016)1-7.

[13] T.Manikandan, N.Bharathi, Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier, J.Med.Syst.40(2016)

[14] F.Khozeimeh,R.Alizadehsani, M.Roshanzamir, A.Khosravi, P.Layegh, S.Nahavandi,An expert system for selecting wart treatment method.Comput Biol Med.81(2017)167-175.