A review on various classification algorithm for Acute Kidney Injury diagnostic method

Brinda M. Shah[1], Ishan K. Rajani[2]

[1] PG student, Department of Computer Engineering, Silver Oak college of Engineering and Technology, Gujarat, India

[2] Professor, Department of Computer Engineering, Silver Oak college of Engineering and Technology, Gujarat, India

---***---

Abstract - *Over the most recent few decades, data mining has produced numerous headways in different regions. Data and Web technologies are only two noteworthy words essentially utilized for data mining. They have been useful in an assortment of orders as customer relationship management, Market basket analysis, telecommunication industry and web mining, healthcare domain, finance sectors, and so on. In the healthcare industry, data mining procedure and algorithms are used to help the doctors in the identification of any disorder along with decision making procedure. Information Discovery in Database (KDD) [7] introduce chief technique in data mining. By employing this technique, we could detect and extract the real data from a huge data set. There are a lot of algorithms designed for mining the data. Tests demonstrate aside from all algorithms; classification algorithms are much far better compared to clustering algorithms concerning predictors in healthcare domain. Classification methods in data mining generally utilized to predict that the value out of a previously reviewed variable. (Size 10 & Italic , cambria font)*

Key Words: **Data Mining in Healthcare, Classification, Acute Kidney Injury, J48, SVM, C5.0**

1. INTRODUCTION

The medical environment can give us lots of information, but we cannot get useful knowledge from that. For this individual ought to have a few tools, that may identify prognosis and diagnosis from healthcare domain [5]. Data-mining possess that capacities, it provides better and earlier identification to get some disorder like cancer, kidney and heart. By sooner identification we can provide accurate and far better treatment for patients.

Well understood data mining Methods that are successfully employed in medical domain [5] names are Artificial Neural Networks, Nearest Neighbour system, genetic algorithms and Decision tree [19]. The information that's recovered from datamining procedure is going to be utilized for improve clinical practice, decision making process, prognoses and provide treatment recommendation for patients from healthcare [7] associations.

Classification is Used method in healthcare domain. Main goal of employing classification algorithm would be always to create perfect prediction every and every time with high accuracy. By employing classification technique, we are able to predict potential target from previous data set.

Classification works on two Different types of data sets: (1) Training data set, and (2) Testing data set [3]. A model is assembled using training data set, and performs with the prediction by simply employing the model on testing data set. Class label of training data set are known for us.

CART and C-5.0 [3] are data Mining algorithms utilized on selection of application like opinion classification, spam detection, etc. CART [4] uses GINI index impurity measures to make a decision tree while C5.0 [3] build a decision tree with maximum information gained. Sampling methods are utilised to create training samples and testing samples [3] for both CART and C5.0.

Kidney disorder [1] is now A popular disorder in around the world. The prediction of kidney disorder [1][2] is Exceptionally intricate task while managing enormous data set. The Data Set include patients Information like blood pressure grades, sugar, age, counts of blood Cells, albumin etc. that are utilised to predict the disorder. Using CART and C-5.0 That the computation time may be lowered.

2. RELATED WORK

Data mining in healthcare Has been created in a variety of application. Kidney injury is just one of these. AKI has been analysed by many researchers. There's broad field of research for example its risk, ideal definition, tool to implemented and its efficiency. KDIGO guideline [1][20] was chosen to give decision-making and recommendation process. Simple cart and J48 were chosen as the algorithm for this particular procedure.

In the field of cancer, Breast cancer is the most frequent cancer in the world. For breast cancer diagnosis, they've utilized Wisconsin Prognostic Breast Cancer (WPBC) dataset [6] from UCI machine learning repository. After all trials for breast cancer experiments, C5.0 and SVM have demonstrated 81% accuracy because of the recurrence of this illness.

For heart disease Patients it's complicated for medical professionals to predict the heart attack [4] since it's an intricate task which needs expertise and knowledge. An experiment on program of mining algorithm for example simple CART [19] so as to predict the heart attacks and also to evaluate the best available system of prediction.

Weighted Average Ensemble Learning Imputation algorithm [2] predicts that the missing value imputation and the final value is called by calculating weighted average of every version. Thus, the accuracy of kidney disorder prediction is going to be made better by utilizing WAELI.

3. EXISTING WORK

Acute Kidney Injury [1] is harmful and common disorder for the patients. It's connected with reduction chance of survival; even more hospital stays and maximize development of Chronic Kidney Disease [2]. To identification AKI, KDIGO clinical practice guideline [1][19] was released for providing standardized criteria of AKI recommendation and definition in medical pathway.

The entire strategy is split in to 3 segments. First section gives a procedure for discover AKI risk factors by employing classification technique such as simple cart [3] and also J48 algorithm for use for patient data set processing. It features all simple steps of datamining including and collect the data, selection of this data. Data pre-processing [7] steps consist of data cleaning and integrate the data. Then data needs to be modify at a certain degree. To recognize the risk factors of AKI, classification data mining procedure including J48 and cart are all employed in data mining procedure.

Secondly section presents a notion for developing diagnostic tool-using KDIGO guideline [19]. Data-set of patients comprise a few vital data like date admission, creatinine level and urine outcome. AKI severity staging model [12] detect the stage of AKI also it exhibits instantly into an individual. The risk factors are accessed from the end result recovered from datamining procedure along with stores within the application for risk factors seeing.

Last section refers to the way of evaluating the procedure utilizing 2*2 statute. We can quantify sensitivity, specificity and accuracy [1] utilizing this table. One major constraint of this procedure is it may simply be utilized at the section of identification procedure for sculpting and supply guide line advice just, it's unable to handle and manage most the clinical procedure. Doctor's attention and decision on the individual treatment will probably soon be required.
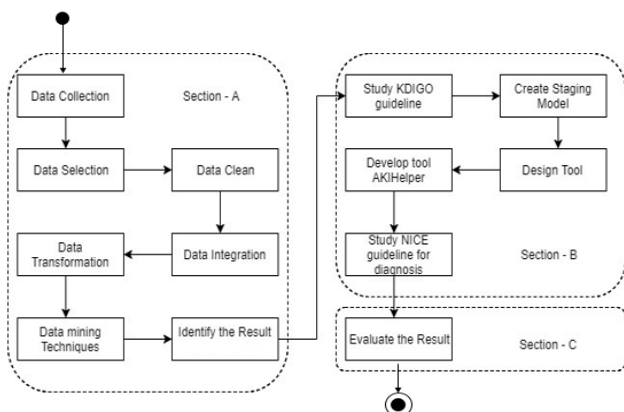
4. PROPOSED WORK FLOW

Fig -1: Proposed work flow for AKI diagnosis

Process of Acute kidney injury Identification Will Be split in to 3 segments. The first step is really for Identify the Risk factors of AKI [1]. There'll soon be couple actions in section, that can soon be achieved as data pre-processing [7]. It's going to include Data set from lab or famous hospital. From then on, data need to be selected since it'll soon be listed to your first seven days of a calendar month or two past a month's data. Initial data set will incorporate patient's basic information such as their characteristics, laboratory results, etc.

Data Pre-Processing will comprise some fundamental Performance for data cleaning and integrating. Data redundancy and data complexity is going to be lowered in such steps. To identify the risk factors of AKI [1][12] predicated on such features, Classification datamining technique C5.0 [3][13] or SVM is going to be properly used. By employing this sort of algorithm, we could possibly secure more accuracy without time complexity.

Secondly Section is going to be for Developing Diagnostic Tool. Where KDIGO will provide the guidelines for clinical practice along with recommendations to assist decision making, looking after patient in risk and picking therapy to increase survival and also sustain kidney function. AKI staging model is going to be generated based on principle, which supply the meaning and staging of Severe Kidney Injury. To get far better recommendation NICE guideline [21] system is going to be utilised in hospital therapy and diagnosis [17].

Third Section is really for Evaluating the end result of several measures such as sensitivity, specificity and accuracy.

Table -1: Evaluation Result

| Test Result		Actual Result	
		Disease	*Non-Disease*
	Positive	a True Positive	b False Positive
	Negative	c False Negative	d True Negative

Sensitivity = a / (a + c)

Specificity = d / (b + d)

Accuracy = (a + d) / (a + b + c + d)

Sensitivity value Can Be used to Quantify the Percentage of advantages which are accurately identified. Specificity value is Used to quantify the ratio of downsides which are accurately identified. Truth significance is employed the accuracy of this system.

5. COMPARISON
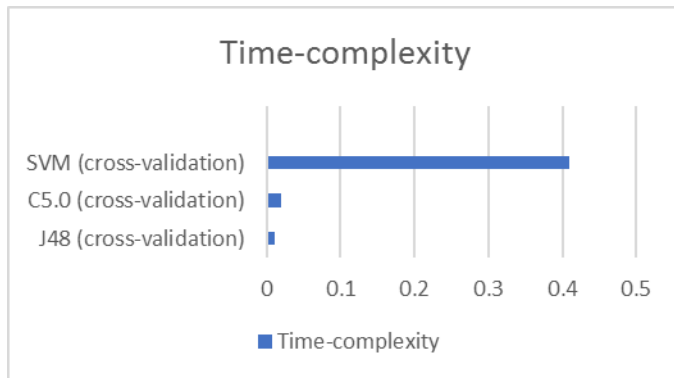
Chart -1: Comparison

As we have discussed above, existing system for AKI diagnosis [12] used J48 [1] algorithm. While in proposed system SVM or C5.0 [3][18] will be used for the same dataset. Thus, we can calculate the actual time complexity of those 3 algorithms to compare which one is better with their accuracy. Here, result shows that J48 has lesser time taken while C5.0 [13] and SVM has taken more time to perform same task. But in terms of more accurate result, we can say that C5.0 and SVM has more accurate result. So, we can take C5.0 algorithm for the further process.

6. CONCLUSIONS

The datamining has played at a vital part in healthcare domain, notably in predicting a variety of sorts of diseases. The diagnosis is widely utilized in predicting ailments, and they are broadly utilised in medical care. In summary, there isn't any one data mining system to eliminate the difficulties in the health data collections. In order to acquire the greatest accuracy among classifiers that's crucial in medical diagnosing together with the characteristics of data getting cared, we will need to look for a hybrid model that may resolve the cited topics. Our prospective instructions are to boost the predictions with hybrid models.

REFERENCES

[1]   Issariya Uboltham, Nakornthip Prompoon, Wirichada Pan-ngum, "AKIHelper: Acute Kidney Injury Diagnostic Tool Using KDIGO Guideline Approach," ICIS 2016, Okayama, Japan.

[2]   S.Dilli Arasu, Dr. R.Thirumalaiselvi, "A Novel Imputation Method for Effective Prediction of Coronary Kidney Disease," 2017 Second international Conference on Computing and Communications Technologies (ICCCT'17).

[3]   M Balamurugan, S Kannan, "Performance Analysis of Cart and C5.0 using Sampling Techniques," 2016 IEEE International Conference on Advances in Computer Applications (ICACA).

[4]   Dr.Neeraj Bhargava, Sonia Dayna, Abhishek Kumar, Pramod Singh, "An Approach for Classification using Simple CART Algorithm in Weka," 2017 11th International Conference on Intelligent Systems and Control (ISCO).

[5]   Narender Kumar, Sabita Khatri, "Implementing WEKA for Medical Data Classification and Early Disease Prediction," 3rd IEEE International Conference on "Computational Intelligence and Communication Technology" (IEEE-CICT 2017).

[6]   Uma Ojha, Dr. Savita Goel, "A Study on Prediction of Breast Cancer Recurrence using Data Mining Techniques," 2017 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence.

[7]   Neesha Jothi, Nur Aini Abdul Rashid, Wahidah Husain, "Data Mining in Healthcare – A Review," 2015 Peer-review under responsibility of organizing committee of Information Systems International Conference (ISICO2015).

[8]   Rutvija Pandya, Jayati Pandya, "C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning," 2015 International Journal of Computer Applications.

[9]   Wiga Maulana Baihaqi, Noor Akhmad Setiawan, Igi Ardiyanto, "Rule Extraction for Fuzzy Expert System to Dignose Coronary Artery Disease," 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE).

[10] Thanh-Long Nguyen, Bay Vo, Loan T.T, Nguyen, "A New Method for Mining Colossal Patterns." 2016 IEEE International Conference on Systems, Man, and Cybemetics (SMC2016).

[11] Elnaz Pashaei, Mustafa Ozen, Nizamettin Aydin, "An Application of Black Hole Algorithm and Decision Tree for Medical Problem," IEEE.

[12] M. Joannidis, P. M. Honore, W. Druml, E. Hoste, "Prevention of Acute kidney injury and protection of renal function in the intensive care unit: update 2017," Expert opinion of the working group on prevention, AKI section, European Society of intensive care medicine.

[13] PANG Su-lin, GONG Ji-zhang, "C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks," 2009 PANG Su-lin al/Systems Engineering – Theory & Practice.

[14] Yanbin Guo, Jianzhong Zhang, Yu Zhang, "An algorithm for analysing the city residents' activity information through mobile big data mining," 2016 IEEE TrustCom-BigDataSE-ISPA.

[15] P. Sita Rama Murty, R. Kiran Kumar, M. Sailaja, "Exploring the Similarity / Dissimilarity Measures for Unsupervised IDS."

[16] Zhou Xu, Jifeng Xuan, Jin Liu, Xiaohui Cui, "MICHAC: Defect Pediction via Feature Selection based on Maximal Information Coefficient with Hierarchical Agglomerative Clustering," 2016 IEEE 23rd International Conference on Software Analysis, Evolution and Reengineering.

[17] Chi-Jane Chen, Tun-Wen Pai, Shih-Syun Lin, Min-Hui Liu, Chao-Hung Wang, "Application of PrefixSpan Algorithms for Disease Pattern Analysis," 2016 International Computer Symposium.

[18] Sergii Babichev, mohamed Ali Taif, Volodymyr Lytvynenko, "Inductive Model of Data Clustering based on the Agglomerative Hierarchical Algorithm," IEEE First International Conference on Data Stream Mining & Processing.

[19] Zhu Xiaoliang, Wang Jian, Wu Shangzhuo, Yan Hongcan "Research and Application of the improved Algorithm C4.5 on Decision Tree" Hebei Polytechnic University, International Conference on Test and Measurement 2009.

Web-References:

[1] "Acute   Kidney   Injury", http://kdigo.org/guidelines/acute-kidney-injury/

[2] "Acute Kidney Injury: prevention, detection and management", https://www.nice.org.uk/guidance/cg169

[3] http://ieeexplore.ieee.org

[4] https://www.rulequest.com

[5] https://www.researchgate.net