

# Renal Disease Prediction By Feature Extraction Techniques Using CT Scan Images

M.Fathima<sup>1</sup>, Mrs.R.karthiyayini<sup>2</sup>

<sup>1</sup>Department of computer applications, Anna University, BIT campus, Tiruchirappalli

<sup>2</sup>Assistant Professor, Department of computer applications, Anna University, BIT campus, Tiruchirappalli

**Abstract** - Renal disease is the very critical disease spreading nowadays in the whole world due to the change in our life style, including food habits, environment, etc.. Kidney Disease is currently considered as the general problem in recent days. Data mining is an interdisciplinary field used to extract the hidden patterns for critical resolution making in any discipline. This field consists of plenty of techniques and algorithms for developing data mining systems for any specific domain. The scope of this project is to anticipate Chronic renal Disease prediction using the techniques of classification algorithms SVM and Naïve Bayes. Final stage of renal disease and kidney transplantation remain the most definitive renal end points, thus the project contribution is focused on quality of human life. Several computer scientists have been proposing new algorithms, combination of certain methods and system based steps for the medical image processing, although many ways are suggested that accuracy of the prediction still requires modifications. The Image preprocessing are done in MATLAB using Gray conversion, BW image processing, Gaussian filtering and Morphological operations in the CT Scan images for renal disease analysis.

**Keywords** - Image segmentation, Chronic Kidney Disease (CKD), SVM and Naïve Bayes.

## I. INTRODUCTION

Data mining is an analytical process to investigate definite data from large size of data. It is a process that finds previously unknown patterns and trends in databases. This information is further used to design a predictive models. Large size of data which is generated for the prediction of renal disease is analysed traditionally and is too complicated and very large to be processed.

MATLAB has an greatest collection of graphic tools and it's Plotting a provided data set or the conclusion of estimation is possible with very less commands. In this Paper highly encouraged to plot mathematical functions and results of analysis as often as possible. Trying to understand mathematical equations with graphics is an enjoyable and very efficient way of learning mathematics

## II. LITERATURE SURVEY

[1] Dr.S.Vijayarani et al., Discusses in Data Mining approach to elicit knowledge about the interaction between many of measured Parameters and patient survival. Two different data mining algorithms were employed for extracting knowledge in the form of decision rules. Decision-making Algorithms has discriminant rules, which predicts survival of new unseen patients.

[2] S.Dilli Arasu, et al., discussed the prediction of kidney diseases can be done efficiently using Priority assigning algorithm. It is done with the help of Attribute selection method. But existing algorithm does not suit for imbalanced dataset or large number of distinct values. This makes the classification process efficient and kidney disease can be predicted efficiently.

[3] Jyoti Saini, et al., suggested that in large amounts of data which is generated for the prediction of kidney disease is analysed traditionally and is too complicated and voluminous to be processed. Data mining is effectively related with data science that involves classification and manipulation of data by applying mathematical and statistical concepts. Data mining is a significant phase in discovery of knowledge and comprises application of discovery and analytical techniques on data to create particular models across data.

[4] Guneet Kaur et al., presented in their work in various data mining classification approaches are used for machine learning and decision making in a way to extract some meaningful information from the data. Research area in 'Data Mining' is well-known for the extraction process. Currently, Big data is the developing technology. The data can be structured, Semi-structured and unstructured.

[5] Sahil Sharma et al., discussed in sub-domain of AI has been widely used in order to assist medical experts and doctors in the prediction, diagnosis and prognosis of various diseases and other medical disorders. In this manuscript the authors applied various machine learning algorithms to a problem in the domain of medical diagnosis and analysed efficiency in predicting the results.

## III. Proposed system

The scope of this paper is to examine CKD and the time how fast it occurs or arises in humans. We have used SVM and Naive Bayes classifiers algorithms for the time improves and some vast feature extraction approaches for our experiments. All are attained using Data mining algorithms in MATLAB environment.

Data source contains all types of CT scan abnormal images. Thus abnormal images are retrieved from database and constitute to form a database to store retrieved images, then these images are preprocessed on using support vector machine and Naïve Bayes classifiers where the images are prepared for mining. Image processing, Kidney image Segmentation, Feature set extraction are the techniques adopted to perform decision making. Each technique has several processes like Grayscale Conversion, Gaussian Medium Filtering, Binary Image Conversion, and Morphology Custom Operation are implemented by a using support vector machine and Naïve Bayes classifiers. After getting the classification result based to further operations are performed. Kidney Segmentation operation performed using operations like Histogram Evaluation, Dilation, Resize Intensity Based Access Split, and DCT Based Segmentation. After segmenting kidney images the relevant parameters are selected for Feature set extraction using SRAD,GAC, Content Access Marking operations.

#### IV. Methodology

In this system for Renal Disease Prediction, feed forward Support Vector Machine (SVM) and Naïve Bayes Algorithms are used.

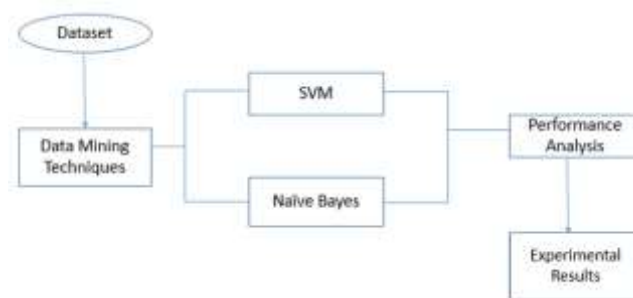


Figure 1 shows the system process Architecture.

#### Data mining techniques

There are many different methods used to perform data mining tasks. Data mining techniques not only require specific types of data structures, but also imply certain types of algorithmic approaches. Some of the common data mining techniques are classification, clustering, prediction etc.

**Support Vector Machine (SVM):** These classifiers are based on structural risk minimization principal and statistical learning theory with an aim of determining the hyperplanes (decision boundaries) that produce the efficient separation of classes. The underlying algorithm is Support Vector Classification (SVC). Support Vector Machines, a new method for the classification of both linear and nonlinear data. In a casing, a support vector machine (SVM) is an algorithm that works as follows. It uses a nonlinear mapping to renovate the exclusive training data into a higher dimension

**Naïve Bayes:** Proposed framework makes use of "Naïve Bayes Algorithm" This algorithm predicts whether the patient is pathologic or normal by ckd.

Step 1: Scan the dataset (storage servers)

Step 2: Calculate the probability of each attribute value. [n, n<sub>c</sub>, m, p]

Step 3: Apply the formulae

$$P(\text{attribute value}(a_i)/\text{subject value}(v_j)) = (n_c + mp)/(n+m)$$

Where:

- n = the number of training examples for which v = v<sub>j</sub>
- n<sub>c</sub> = number of examples for which v = v<sub>j</sub> and a = a<sub>i</sub>
- p = a priori estimate for P(a<sub>ij</sub>v<sub>j</sub>)
- m = the equivalent sample size

Step 4: Multiply the probabilities by p

Step 5: Compare the values and classify the attribute values to one of the predefined set of class.

#### V. Architecture

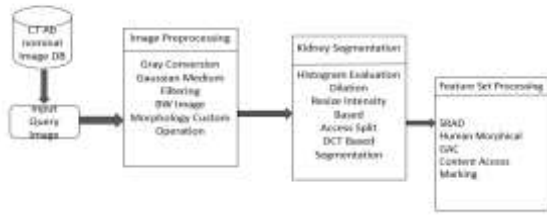


Figure 1.1 shows the flow of CT SCAN Image Feature Extraction process.

- The System is Classified into Three phases

**i. Image Pre-processing**

The Image Preprocessing contains the main fields like Gray Conversion, Gaussian Medium Filtering, Binary Image and Morphology Custom Operation. It prepared the images for mining.

**ii. Kidney Segmentation**

The Kidney Segmentation contains the main fields like Histogram Evaluation, Dilation, Resize Intensity Based Access Split, and DCT Based Segmentation.

**iii. Feature Set Processing**

Feature Set Processing contains the main fields like SRAD, GAC, Content Access Marking for feature set extraction to identify pathologic images.

**VI. UML DIAGRAMS**

**a. Use Case Diagram**

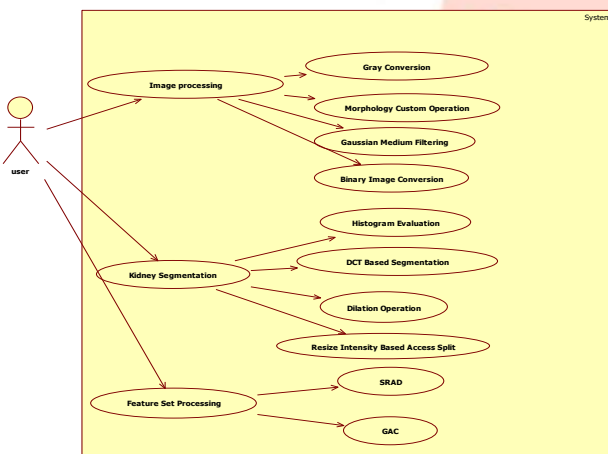


Figure 2.1 display the user accessing methods of Feature extraction.

**b. Class Diagram**

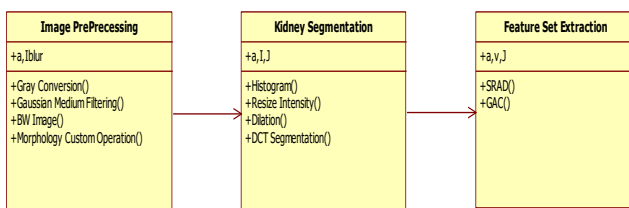


Figure 2.2 illustrate the class diagram of Feature Extraction process.

**VII. Results**



Figure 3.1 shows three phases of renal disease prediction method.



Figure 3.2 shows Image processing phase of renal disease prediction method.



Figure 3.3 shows the input phase of Morphological operations.



Figure 3.4 shows the output phase of Morphological dilation operations.



Figure 3.5 shows the result of image conversion



Figure 3.6 shows the Gaussian filtering output result.



Figure 3.7 shows the effect of resize intensity based access split



Figure 3.8 shows effect of DCT based segmentation.



Figure 3.9 shows the historical representation of Image GPU calculation.

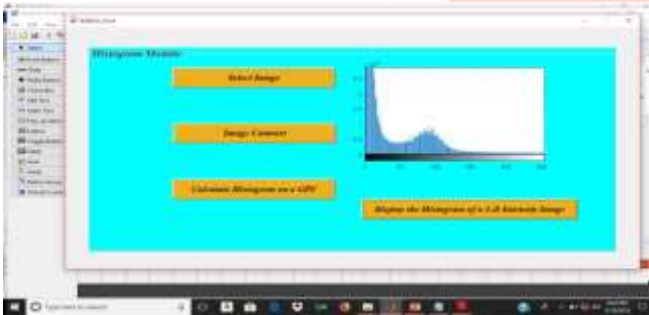


Figure 3.10 shows the histogram of 3D intensity image.



Figure 3.11 shows the effect of feature extraction of CT Scan Images.

## VIII. CONCLUSION AND FUTURE WORK



Medical related information is highly voluminous in nature in the healthcare industry. It can be derived or retrieved from various sources which are not entirely applicable in this feature. In this work, kidney disease prediction system was developed using classification algorithms (SVM&Naive Bayes) through Mat lab data mining tool to predict effective and better accurate results regarding whether the patient is suffering from kidney disease or not. As the kidney disease patients are increasing world-wide each year and huge amounts of data is available for research, where different data mining techniques are used in the diagnosis of kidney disease.

## IX. References

- [1] Vijayarani, S., & Dhayanand, S. (2015). Data mining classification algorithms for kidney disease prediction. *International Journal on Cybernetics and Informatics (IJCI)*.
- [2] Arasu, S. D., & Thirumalaiselvi, R. (2017). A Prediction of Chronic Kidney Disease Using Feature based Priority Assigning Algorithm. *International Journal of Applied Engineering Research*, 12(20), 9500-9505.
- [3] Jyoti Saini, R.C Gangwar and Mohit Marwaha, A Novel Detection for Kidney Disease Using Improved Support Vector Machine, 2015 *International Journal of Latest Trends in Engineering and Technology*.
- [4] Guneet Kaur, Ajay Sharma, Predict Chronic Kidney Disease Using Data Mining Algorithms In Hadoop, *International Journal of Advanced Computational Engineering and Networking*, ISSN: 2320-2106, Volume-5, Issue-6, Jun-2017.
- [5] Sharma, S., Sharma, V., & Sharma, A. (2016). Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis. *arXiv preprint arXiv:1606.09581*.
- [6] Yue Jiao, PhD, Dan Geary, MHA, Sheelal, Peter Kotanko, Development and Testing of Prediction Models for End Stage Kidney Disease Patient Nonadherence to Renal Replacement Treatment Regimens Utilizing Big Data and Healthcare Informatics, 2015 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* Pages 1721-1721.
- [7] Abeer Y. Al- Hyari, "Chronic Kidney Disease Prediction System Using Classifying Data Mining Techniques", library of university of Jordan, 2012.
- [8] P. Sinha and P. Sinha, "Comparative study of chronic kidney disease prediction using knn and svm," *International Journal of Engineering Research and Technology*, vol. 4, no. 12, 2015.
- [9] Sonali Sonavane, Alka Khade, Gaikwad. V.B, (2013) "Novel Approach for Localization of Indian Car Number Plate Recognition System using Support Vector Machine", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 8, August, page no 179-183.
- [10] B.Rakesh et al "Prediction of Slow Learners in Higher Educational Institutions Using Random Forest Classification Algorithm.", *i-Managers Journal on Cloud Computing*, Vol-2, No-3, May-July 2015.
- [11] K.R. Lakshmi, Y. Nagesh and M. Veera Krishna, "Performance comparison of three data mining techniques for predicting kidney disease survivability", *International Journal of Advances in Engineering & Technology*. Vol. 7, No. 1, pp. 242-254, (2014).
- [12] Abeer Y. Al-Hyari, "Chronic Kidney Disease Prediction System Using Classifying Data Mining Techniques", library of university of Jordan, 2012.
- [13] Dipak V. Patil, Bichkar. R.S, (2011) "An Optimistic Data Mining Approach for Handling Large Data Set using Data Partitioning Techniques", *International Journal of Computer Applications (0975 – 8887)*, Volume 24– No.3, June, page no 29- 33
- [14] Andrew Kusiak, Bradley Dixon, Shital Shaha, (2005) "Predicting survival time for kidney dialysis patients: a data mining approach", Elsevier Publication, *Computers in Biology and Medicine* 35, page no 311–327.
- [15] Ashfaq Ahmed K, Sultan Aljahdali and Syed Naimatullah Hussain, (2013) "Comparative Prediction Performance with Support Vector Machine and Random Forest Classification Techniques", *International Journal of Computer Applications* Volume 69– No.11, May page no 12-16
- [16] DSVGK Kaladhar, Krishna Apparao Rayavarapu\* and Varahalarao Vadlapudi, "Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-analysis", *Open Access Scientific Reports*, Vol. 1, No. 12, pp., 2012.
- [17] Ishani Mandli, Mahesh Panchal, (2014) "Selection of Most Relevant Features from High Dimensional Data using IG-GA Hybrid Approach", *International Journal of Computer Science and Mobile Computing*, Vol.3 Issue.2, February, pg. 827-830
- [18] Jainesh Patel, Neha R Soni, (2014) "Survey of Supervised and Unsupervised Algorithms in Email Management", *International Journal of Scientific & Engineering Research*, Volume 5, Issue 3, March, page no 1241-1243
- [19] Abhishek, G. S. M. T., & Gupta, D., 2012. Proposing Efficient Neural Network Training Model for Kidney Stone Diagnosis. *International Journal of Computer Science and Information Technologies*, 3(3), pp. 3900-3904.