International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 21 Issue 2 – APRIL 2016.

FACTORS INFLUENCING KIDNEY PROBLEMS AND TREATMENT OF DIALYSIS: A DESCRIPTIVE STUDY USING DATA MINING TECHNIQUE

Krishnamoorthy.M, Srinivasan.J, Geetha.V

Assistant professor, Department of CSA, SCSVMV University, Kanchipuram.
Assistant professor, Department of CSA, SCSVMV University, Kanchipuram.
Assistant professor, Department of CSE, SCSVMV University, Kanchipuram.

krishm_m@kanchiuniv.ac.in
vsrini082@gmail.com
vgeetha82@gmail.com

***Abstract* --Kidney failure is one of the major occurrences in the world today. Kidney problems are caused because of some genetic factors and/or Uncleaned Water factors or today's modern lifestyle. Kidney failure has become the primary reason of death in undeveloped countries. The most effective way to reduce Kidney failure death is to detect it earlier. The earlier detection of Kidney failure is not easier process but if it is detected, it is curable. Many works have been done in predicting Kidney and Liver failure, different data mining approaches and algorithms were adopted by different people. Each work has some limitations such as lack of intelligent prediction, and inefficient in structure that motivated to take up this problem and to implement the Data mining based Kidney failure prediction System. We have proposed the Kidney failure prediction based on data mining. This system estimates the risk of the Kidney , liver and Renal failure.**

***Keywords: Kidney Failure, Renal failure, J48, Id3, Navie bayes*.**

I. INTRODUCTION

Scientific advances since the completion of the Human Genome Project have confirmed that the genetic composition of individual humans has a significant role to play in predisposition to Kidney and Renal problems. The traditional medicine model has relied on best practices emerging from large population studies and dictates a one-size-fits-all approach [1]. Although synthesized evidence is essential to demonstrate the overall safety and efficacy of medical approaches, it falls short in explaining the Individual variations that exist among Kidney problem patients. Recent advances in genome-wide association studies have revolutionized the practice of medicine, causing a shift to a patient-centered model [2] and offering tailored diagnostic and Dialysis strategies. The translation of genetic and genomic data into the knowledge of Kidney patients care for prevention, diagnosis, prognosis and treatment has introduced a new paradigm for healthcare: *personalized medicine*

The objective of this study was to investigate the influence of a combination of factors—i.e. clinical predictors, environmental risk factors, and EGFR mutation status—to predict problems response to Kidney and liver. We developed a proof of concept data-driven prediction model to assist in personalizing treatment and patient selection in advanced renal problems. In this paper we present a data-driven prediction model, developed by applying two data mining methods—i.e. association rule mining and decision trees to predict a patient's response to Hemo-Dialysis. Given the scarcity of patient data available for such a study, we formulated a novel data curation methodology to derive data from a multitude of secondary sources. We applied association rule mining to validate the inherent relationships between patient characteristics and Dialysis patterns as reported in the literature. Decision tree classifiers were then used to develop a visually interpretable prediction model that combined clinic pathological data and Dialysis report status to categorize patients' responsiveness to treatments Effectiveness. Despite the rather small dataset, our decision tree based prediction model is able to produce meaningful and useful predictions with an accuracy of 85%.

II. LITERATURE STUDIES

For the detection of Kidney risks, Each and every method has its own advantages and some disadvantages. Kidney stones

Sample of work have been done to predict the risk of the Kidney problems. There are different techniques proposed by different authors.
prediction is certainly a very complex and nondeterministic endeavor so many tests are available for K-Stone prediction

but it's of high cost. In medicine, **dialysis** (from Greek **dialusis**, meaning *dissolution*, **dia**, meaning *through*, and **lysis**, meaning *loosening or splitting*) is a process for removing waste and excess water from the blood, and is used primarily as an artificial replacement for lost kidney function in people with renal failure. Dialysis may be used for those with an acute disturbance in kidney function (acute kidney injury, previously acute renal failure), or progressive but chronically worsening kidney function–a state known as chronic kidney disease stage 5 (previously chronic renal failure or end-stage renal disease). The latter form may develop over months or years, but in contrast to acute kidney injury is not usually reversible, and dialysis is regarded as a "holding measure" until a renal transplant can be performed, or sometimes as the only supportive measure in those for whom a transplant would be inappropriate.[1]Dialysis is an imperfect treatment to replace kidney function because it does not correct the compromised endocrine functions of the kidney. Dialysis treatments replace some of these functions through diffusion (waste removal) and ultrafiltration (fluid removal).[2]

2.1 History of Dialysis

A Dutch physician, Willem Johan Kolff, constructed the first working dialyzer in 1943 during the Nazi occupation of the Netherlands.[3] Due to the scarcity of available resources, Kolff had to improvise and build the initial machine using sausage casings, beverage cans, a washing machine, and various other items that were available at the time. Over the following two years, [1943-1945] Kolff used his machine to treat 16 patients suffering from acute kidney failure, but the results were unsuccessful. Then, in 1945, a 67-year-old comatose woman regained consciousness following 11 hours of hemodialysis with the dialyzer, and lived for another seven years before dying from an unrelated condition. She was the first-ever patient successfully treated with dialysis.[3] Dr. Nils Alwall modified a similar construction to the Kolff kidney by enclosing it inside a stainless steel canister. This allowed the removal of fluids, by applying a negative pressure to the outside canister, thus making it the first truly practical device for hemodialysis. Alwall treated his first patient in acute renal failure on the September 3, 1946.

Dialysis Machine

2.2 Working principle

Dialysis works on the principles of the diffusion of solutes and ultrafiltration of fluid across a semi-permeable membrane. Diffusion is a property of substances in water; substances in water tend to move from an area of high concentration to an area of low concentration.[4] Blood flows by one side of a semi-permeable membrane, and a dialysate, or special dialysis fluid, flows by the opposite side. A semipermeable membrane is a thin layer of material that contains holes of various sizes, or pores. Smaller solutes and fluid pass through the membrane, but the membrane blocks the passage of larger substances (for example, red blood cells, large proteins). This replicates the filtering process that takes place in the kidneys, when the blood enters the kidneys and the larger substances are separated from the smaller ones in the glomerulus

The two main types of dialysis, hemodialysis and peritoneal dialysis, remove wastes and excess water from the blood in different ways.[5] Hemodialysis removes wastes and water by circulating blood outside the body through an external filter, called a dialyzer, that contains a semipermeablemembrane.

The blood flows in one direction and the dialysate flows in the opposite. The counter-current flow of the blood and dialysate maximizes the concentration gradient of solutes between the blood and dialysate, which helps to remove more urea and creatinine from the blood. The concentrations of solutes (for example potassium, phosphorus, and urea) are undesirably high in the blood, but low or absent in the dialysis solution, and constant replacement of the dialysate ensures that the concentration of undesired solutes is kept low on this side of the membrane. The dialysis solution has levels of minerals likepotassium and calcium that are similar to their natural

concentration in healthy blood. For another solute, bicarbonate, dialysis solution level is set at a slightly higher level than in normal blood, to encourage diffusion of bicarbonate into the blood, to act as a pH buffer to neutralize the metabolic acidosisthat is often present in these patients. The levels of the components of dialysate are typically prescribed by a nephrologist according to the needs of the individual patient.In peritoneal dialysis, wastes and water are removed from the blood inside the body using the peritoneum as a natural semipermeable membrane. Wastes and excess water move from the blood, across the peritoneal membrane, and into a special dialysis solution, called dialysate, in the abdominal cavity.

2.3 Types of Dialysis

There are three primary and two secondary types of dialysis: hemodialysis (primary), peritoneal dialysis (primary), hemofiltration (primary),hemodiafiltration (secondary), and intestinal dialysis (secondary).

III. DATA MINING METHODOLOGY

The availability of a rich patient data resource to conduct such a study is always a major challenge. Therefore, in this work, we explored and used patient data from secondary sources. Although potential databases were identified but a closer examination revealed that very few stored the attributes of interest to this study (age, gender, histology and objective response) and were publically available. Thus, we took an innovative approach of drawing upon multiple freely available data sources to construct a research dataset containing attributes for patient demo Graphics, smoking history, Patient Histology, Dialysis treatment and clinical response to targeted therapy. An evidence based approach was used to determine the predictors of response to Dialysis [6].

3.1. Data Curation

Sources for data curation included Meenakshi Medical college, Catalogue of Dialysis patients and Dialysis cured patients database were used to search for case series, case reports, and research publications between the years 2000-2012. Only articles reporting individual-level patient data on human subjects were selected. The bibliographies of these articles also pointed to relevant literature containing patient-level data. This search resulted in the identification of 162 papers from Meenakshi Medical college, 681 unique patient Report samples (each associated with a ID) from the Dialysis database and 167 articles from SM-DB. After screening for inclusion criteria and removal of duplicate studies, 334 articles and 144 case reports were selected for data extraction.

Our approach for manual curation and structuring of relevant data from research articles and case reports was as follows: 234 research articles were selected based on the inclusion criteria. For each research article, we reviewed the study design, methods, interventions, and outcomes. Candidate articles reported attributes of demographics, smoking status, diagnosis, CELL mutations, and response to Dialysis therapy for selected study subjects in the form of tables or supplementary files. Pertinent data was extracted from these tables and combined. Some authors reported patients and rare mutations in multiple papers and care was taken to delete duplicate entries for individual cases. We identified 144 case reports pertaining to DIALYSIS treatment response. These clinical narratives each presented detailed account of patient history, investigations, treatment and outcome. Through a careful examination of the free text for each case report, we identified the key attributes of interest that were scattered through the report. The aggregation of attributes from individual case reports comprised a single data record. Data extracted from article and case reports was pooled to create the final working dataset. Overall, 355 cases were extracted from 234 research articles and 144 case reports (Table I), where each case was represented by 7 attributes (age, gender, smoking status, diagnosis, drug, CELL mutation, and Dialysis response). Most studies reported treatment response as complete response (CR), partial response (PR), stable disease (SD) or progressive disease (PD) as defined by the Response Evaluation Criteria in Kidney problems) [18]. If the study did not specify the treatment of DIALYSIS and response were classified using terms such as partial regression, complete regression, partial remission, complete remission.

3.2. Association Rule Discovery

Given that the data for this study was derived from secondary sources, prior to developing the prediction model we needed to objectively validate the generated study dataset. We applied association rule mining to identify inherent associations between patient characteristics and Dialysis responses within the generated study dataset. Frequent associations were then compared with previously well-understood patterns and trends in order to validate the study dataset. Various authors [22], [23], [24] have reported the success of using frequent patterns (or association rules, expressed as IF-THEN rules, in medical databases to confirm previous biomedical knowledge and to discover novel associations between variables.

3.3. Predictive Modeling

Predictive models to predict the Dialysis response to erlotinib or gefitinib, at individual-level patient level, were developed using support vector machine and decision tree classifiers. Patients were divided into responder and non-responder groups. Student's t-test was used to compare ages between the two groups. All association-rule mining and classification algorithms were implemented using the Waikato Environment for Knowledge Analysis (WEKA) [25] and the R statistical software.

IV. RESULTS

4.1. Description of Dataset

The study sample comprised 143 (60%) males and 212 (50%) females and mean age was 50 ± 12.03 years. The majority of cases were non-smokers; 50% patients reported never smoking, 14% were former smokers and 10% were current smokers. 11 histologies for NSCLC were documented

TABLE I

PATIENT CHARACTERISTICS **Attributes**	**N**	%
Gender Male	143	40.28
Female	212	59.72
Smoking status Current	34	9.58
Former	140	50.85
Never	172	59.58
CELL status Mutated	160	33.24
Wildtype	95	16.76
Drug peritonial	74	10.28
Hydro dialysis	526	13.66
Either	57	16.06
Response Responder	250	69.15
Non-responder	245	50.85

4.2. Association Rule Mining Results

The association rule mining algorithm discovered 30 association rules with lift values (indicating interestingness) ranging from 0.9 to 1.93. The largest itemset contained three items. A selection of six 'interesting' rules are shown in Table II.

TABLE II

SELECTED ASSOCIATION RULES FROM APRIORI ALGORITHM	**Dialysis Type**	**Sodium/pottasium level**	**Blood-glucose level**
Histology=Adenocarcinoma EGFR mutation=Wildtype	*Renal*	1.93	0.79
CELL mutation=Wildtype	*Peritonial*	1.84	0.75
CELL mutation=Exon 19	*Renal*	1.33	0.79
Gender=Female, Diagnosis= Adenocarcinoma Drug- gefitinib	*Renal*	1.06	0.63
Age=>49.2 Gender=Female Smoking=Never Histology=Adenocarcinoma	*Peritonial*	1.05	0.62

V. CONCLUSION

Kidney Stone is potentially fatal disease. Detecting Stones is still challenging for the doctors in the field of Dialysis. Even now the actual reason and complete cure of Kidney stones is not invented. Detection of stones in earlier stage is curable. Specifically, this prediction system estimates the risk of the kidney, liver and renal stones by examining a number of user-provided genetic and non-genetic factors. This system is validated by comparing its predicted results with the patient's prior medical record, and also this is analyzed using weka system. This prediction system is available in offline, people can easily check their risk and take appropriate action based on their risk status. This system performs well than the existing system.

REFERENCES

[1] N.Revathy, Dr.R.Amalraj(2011) Accurate Cancer Classification using Expression of Very few Genes. International Journal of Computer Application Volume 14 – No.4.

[2] Tasnuba Jesmin, Kaswar Ahmed, Md. Badrul Alam Miah (2013) Brain Cancer Risk Prediction System Using Data mining. International Journal of Computer Applications,Volume 61- No.

[3] Seyyid Ahmed Medjahed, Tamazouzt Ait Saadi, Abdelkader Benyettou (2013) Breast Cancer Diagnosis using K-Nearest Neibhor with Different Distances And Classification Rules. International Journal of Computer Applications, Volume 62- N0.1.

[4] Wafa Mokharrak, Nedhal Al Khalaf, Tom altman Application of Bioinformatics and Data mining in Cancer Prediction.

[5] Kawasar Ahmed, Tanuba Jesmin, Md.Zamilur Rahman (2013)Early Prevention and Detection of Skin Cancer using Data mining. International Journal of Computer Application, Volume 62-No.4.

[6] Abdelghani Bellachia, Erhan Guven Predicting Breast Cancer Survivablity Using Data mining Techniques.

[7] S. Jothi, S.Anitha (2012) Data mining Classification Techniques Applied Fo Cancer Disease – A case Study Using Xlminer. International Journal of Engineering Research & Technolgy, Vol 1 Issue 8.

[8] V.Kroshnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra (2013) Diagnosis of Lung Cancer Prediction System Using Data mining Classification

Techniques International Journal Of Computer Science And Information Technologies, Vol 4(1), 39-45.

[9] Ada Ranjneet Kaur (2013) A Study of Lung Cancer Using Data mining Classification Techniques. International Journal of Advanced Research in Computer Science and Software Engineering, Vol 3, Issue 3.

[10] K.Rama Lakshmi and S.Prem Kumar (2013) Utilization of Data mining Techniques for Prediction and Diagnosis of Major Life Threatening Diseases Survivability- Review International Journal of Scientific & Engineering